Studien zum Physik- und Chemielernen

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

340

Arne Bewersdorff

Untersuchung der Effektivität zweier Fortbildungsformate zum Experimentieren mit dem Fokus auf das Unterrichtshandeln

λογος

Studien zum Physik- und Chemielernen

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

Diese Reihe im Logos Verlag Berlin lädt Forscherinnen und Forscher ein, ihre neuen wissenschaftlichen Studien zum Physik- und Chemielernen im Kontext einer Vielzahl von bereits erschienenen Arbeiten zu quantitativen und qualitativen empirischen Untersuchungen sowie evaluativ begleiteten Konzeptionsentwicklungen zu veröffentlichen. Die in den bisherigen Studien erfassten Themen und Inhalte spiegeln das breite Spektrum der Einflussfaktoren wider, die in den Lehr- und Lernprozessen in Schule und Hochschule wirksam sind.
Die Herausgeber hoffen, mit der Förderung von Publikationen, die sich mit dem Physik- und Chemielernen befassen, einen Beitrag zur weiteren Stabilisierung der physik- und chemiedidaktischen Forschung und zur Verbesserung eines an den Ergebnissen fachdidaktischer Forschung orientierten Unterrichts in den beiden Fächern zu leisten.

Martin Hopf, Hans Niedderer, Mathias Ropohl und Elke Sumfleth

Studien zum Physik- und Chemielernen

Band 340

Arne Bewersdorff

Untersuchung der Effektivität zweier Fortbildungsformate zum Experimentieren mit dem Fokus auf das Unterrichtshandeln

Logos Verlag Berlin

Studien zum Physik- und Chemielernen
Martin Hopf, Hans Niedderer, Mathias Ropohl und Elke Sumfleth [Hrsg.]

Bibliografische Information der Deutschen Nationalbibliothek

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

© Copyright Logos Verlag Berlin GmbH 2022

Alle Rechte vorbehalten.

ISBN 978-3-8325-5522-1
ISSN 1614-8967

Logos Verlag Berlin GmbH
Georg-Knorr-Str. 4, Geb. 10
D-12681 Berlin
Tel.: +49 (0)30 / 42 85 10 90
Fax: +49 (0)30 / 42 85 10 92
https://www.logos-verlag.de

Untersuchung der Effektivität zweier Fortbildungsformate zum Experimentieren mit dem Fokus auf das Unterrichtshandeln

Dissertation

zur Erlangung des akademischen Grades eines
Doktors der Philosophie
– Dr. phil. –

vorgelegt der Fakultät für Natur- und Gesellschaftswissenschaften
an der Pädagogischen Hochschule Heidelberg

von
Arne Bewersdorff
aus Heidenheim an der Brenz

München, den 26.07.2021

Erstgutachter: Prof. Dr. Armin Baur, Pädagogische Hochschule Heidelberg
Zweitgutachter: Prof. Dr. Markus Emden, Pädagogische Hochschule Zürich

Tag und Ort der Disputation: 10. Mai 2022 in Heidelberg

Prüfungskommission:

Prof. Dr. Armin Baur, Prof. Dr. Jens Bucksch (Vorsitzender), Prof. Dr. Markus Emden, Prof. Dr. Lutz Kasper, Prof. Dr. Manuela Welzel-Breuer

Inhaltsverzeichnis

0 Verortung im Gesamtprojekt 5
1 Einleitung und Problemstellung 7
2 Theoretischer Rahmen 9
2.1 Relevanz und Definition Experiment 9
2.1.1 Erlernen naturwissenschaftlicher Erkenntnismethoden als zentraler Beitrag zu naturwissenschaftlicher Grundbildung 9
2.1.2 Das Experiment als Unterrichtskonzept 12
2.1.3 Entdeckendes Experimentieren 13
2.2 Befunde zum aktuellen Unterrichtshandeln beim Experimentieren 13
2.2.1 Zum Begriff des Unterrichtshandelns 14
2.2.2 Aktuelle Befunde zum Unterrichtshandeln beim Experimentieren 14
2.3 Aspekte effektiven Unterrichtshandelns zum Experimentieren 15
2.4 Wirkungen von Lehrkräftefortbildungen auf das Unterrichtshandeln 32
2.4.1 Transferbegriff, -formen und -theorien 32
2.4.2 Empirische Befunde zur Wirkung von Lehrkräftefortbildungen auf das Unterrichtshandeln – eine Zusammenschau 40
2.5 Abgeleitete Forschungsfragen und Hypothesen 47
2.6 Existierende Rating- und Kodiermanuale: Potenziale und Limitationen 49
3 Design und Methodik 51
3.1 Die Lehrkräftefortbildung 51
3.1.1 Grundlage des Fortbildungsdesigns 51
3.1.2 Inhalt und Ziele der Fortbildung 53
3.1.3 Die Fortbildungsformate 54
3.2 Anlage und Ablauf der Studie 54
3.3 Die Stichprobe 55
3.4 Eignung der Manuale zur Erfassung der Änderung von Unterrichtshandeln zum (Entdeckenden) Experimentieren 56
3.5 Messinstrumente 59
3.5.1 Theoretische Herleitung und Begründung der Methode 59
3.5.2 Entwicklung und Qualität der Kategoriensysteme 69

3.6 Beschreibung des Vorgehens bei der Clusteranalyse 98
4 Ergebnisse 101
4.1 Unterrichtshandeln von Lehrkräften beim Experimentieren 103
4.1.1 Ergebnisse der Clusteranalyse 103
4.1.2 Änderungen des Unterrichtshandeln mit Fokus auf die einzelnen Lehrkräfte 118
4.1.3 Ergebnisse der Vergleichsanalyse 121
4.2 Phasen des Experimentierprozesses 133
4.2.1 Welche Teilprozesse treten auf? 133
4.2.2 Analyse der aufgewendeten Unterrichtszeit 135
4.3 Öffnungsgrade der Teilprozesse beim Experimentieren 141
4.4 Definition des Experiments 144
4.5 Zusammenfassung der Ergebnisse 148
5 Diskussion 151
5.1 Klärung der Hypothesen 151
5.2 Einordnung und Interpretation der Befunde bzgl. der Aspekte effektiven Unterrichts 151
5.3 Einordnung und Interpretation der Befunde hinsichtlich des Lerntransfers 155
5.4 Limitationen und kritische Punkte der Studie 158
5.4.1 Allgemeine Limitationen 158
5.4.2 Limitationen der eingesetzten Instrumente 160
5.4.3 Limitationen der Auswertung 161
5.4.4 Limitationen bezüglich der Interpretation 161
6 Zusammenfassung, Fazit und Ausblick 163
7 Literaturverzeichnis 167
8 Danksagung 189
9 Abbildungsverzeichnis 191
10 Tabellenverzeichnis 193
11 Anhang 197
11.1 Kategorien- und Ratingsysteme ‚Unterrichtshandeln von Lehrkräften beim Experimentieren', ‚Öffnungsgrade der Teilprozesse beim Experimentieren' und ‚Definition des Experiments' 197

11.2 Kategoriensystem ‚Phasen des Experimentierprozesses' 211
11.3 Validierungsbogen Kategoriensystem ‚Unterrichtshandeln von Lehrkräften beim Experimentieren' 215
11.4 Handlungsschritte zur Durchführung von Unterrichtsvideografien 218
11.5 Transkriptionsleitfaden 220
11.6 Clusteranalysen (Rohform) 223
11.6.1 Dimension ‚Prozessstruktur' 223
11.6.2 Oberkategorie ‚Problem und Fragestellung' 224
11.6.3 Dimension , Prozessreflexion (meso und meta)' 225
11.6.4 Dimension ‚Angemessenheit und Unterstützung' 226
11.6.5 Oberkategorie ‚Mesoreflexion' 227
11.6.6 Oberkategorie ‚L als Trainer, Ansprechpartner und Experte' 228

0 Verortung im Gesamtprojekt[1]

Die im folgenden vorgestellte Studie ist Teil eines Projekts zur umfassenden Erfassung von Wirkungen von Lehrkräftefortbildungen zum Experimentieren. Die vorliegende Studie fokussiert im Rahmen des Gesamtprojekts den Zielbereich Unterrichtshandeln.

Zentrales Ziel des Gesamtprojekts „‘Entdeckendes Experimentieren‘ – Wirksamkeit einer Lehrerinnen- und Lehrerfortbildung im Format der Schulwerkstatt“ (Emden & Baur, 2015) ist die Untersuchung der Wirksamkeit zweier Lehrkräftefortbildungsformate in den Zielbereichen Lehrkräftekognition, Beliefs und Unterrichtshandeln im Kontext eines kompetenzorientierten Unterrichts zum effektiven Experimentieren.

Die Zielbereiche leiten sich ab aus einem zugrundeliegenden Wirkungs- und Gestaltungsmodell für effektive Lehrkräftefortbildung (Emden & Baur, 2017, s. a. Kapitel 3.1). Gemäß dem Modell wurde die Fortbildung als langfristige Fortbildung (3 Fortbildungsphasen über 1,5 Schuljahre), mit dezidiertem inhaltlichem Schwerpunkt (geöffnetes Experimentieren) unter Berücksichtigung der bildungspolitischen (Lehrpläne) und lokalen Rahmensetzungen (z. B. Ausstattung) konzipiert.

Die beiden Fortbildungsformate unterscheiden sich allein in der Tiefe der kollektiven Teilnahme (Emden & Baur, 2015)[2]. In der Experimentalbedingung (sog. kokonstruktives Fortbildungsformat) wird dieser Aspekt im Vergleich zu der Kontrollbedingung (sog. individualisiert-konstruktives Fortbildungsformat) durch die Stellung und Bearbeitung kokonstruktiver Arbeitsaufträge besonders gestärkt. Die einzelnen Fortbildungsphasen sind zwischen den beiden Fortbildungsformaten vergleichbar gestaltet, um auf den Beitrag der kokonstruktiven Elemente Rückschlüsse ziehen zu können.

Im Gesamtprojekt wurden folgende Forschungsfragen untersucht:

1. Welche Wirkungseffekte zeigen die beiden Fortbildungsformate auf die Entwicklung:

    a. des Professionswissens zum effektiven Experimentieren?

    b. der Beliefs betreffend der Vermittlung eines effektiven Experimentierens im naturwissenschaftlichen Unterricht?

[1] Das Kapitel 0 wurde nach der Disputation hinzugefügt, um die Verortung im Gesamtprojekt darzustellen.
[2] In späteren Arbeiten wird diese Variation bevorzugt unter der Perspektive des ‚aktiven Lernens‘ besprochen, wobei die Argumentationen sich entsprechen. Die vorliegende Darstellung folgt dem originalen Antragstext, auf dessen Basis die berichtete Studie entstanden ist. Entsprechende Inkonsistenzen gegenüber folgenden Darstellungen sind jedoch eher terminologischer, denn substanzieller Natur.

c. des beobachtbaren Unterrichtshandelns zum entdeckenden Experimentieren?

2. Welchen differenziellen Wirkungseffekt haben die realisierten kokonstruktiven Elemente im Format der Schulwerkstatt im Vergleich zu einem individualisiert-konstruktiven Format auf die Zielbereiche?

3. Welche Wirkungseffekte zeigen die beiden Fortbildungsformate auf die Entwicklung der Schülerkognition?

In der hier vorgestellten Teilstudie wird auf Forschungsfrage 1 c) sowie Forschungsfrage 2, mit dem Fokus auf das Unterrichtshandeln, fokussiert. Erste Ergebnisse zu den weiteren Forschungsfragen finden sich bei Emden, Baur & Bewersdorff (2020).

1 Einleitung und Problemstellung

Trotz einer Vielzahl von Bemühungen seit den 1960er-Jahren haben Lernende immer noch oft ein fehlerhaftes Verständnis von *Nature of Science* und den Methoden naturwissenschaftlicher Erkenntnisgewinnung wie etwa dem Experimentieren (Abd-El-Khalick, 2006a; Lederman, 2006; McComas, Clough & Nouri, 2020).

Wie die Erfassung der Kompetenzbereiche des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) (vgl. Pant et al., 2013) auf Basis der 2004 von der KMK verabschiedeten Bildungsstandards (KMK, 2005a, KMK, 2005b, KMK, 2005c) zeigt, ist dieser Befund, bezüglich naturwissenschaftlicher Kompetenzen besonders im nichtgymnasialen Bereich, weiter gültig (Emden & Baur, 2017; Köller, 2014).

Diese unbefriedigenden Kompetenzstände gehen vermutlich auf ein limitiertes Verständnis – und damit zwangsläufig ebensolches Unterrichtshandeln – der Lehrkräfte zurück. Naturwissenschaftlicher Unterricht ist noch immer häufig vor allem an der Vorstellung von Effekten und Phänomenen und der reinen Nachahmung naturwissenschaftlicher Erkenntnisgewinnung und impliziten Vermittlung orientiert (Capps & Crawford, 2012; Tesch & Duit, 2004).

Dieser Befund führt zu den Fragestellungen, welchen im Rahmen der hier vorliegenden Arbeit nachgegangen werden soll:

1. Was sind Aspekte effektiven Unterrichtshandelns von Lehrkräften bezüglich des Experimentierens? Zur Beantwortung dieser Fragestellung werden Unterrichtsprinzipien, welche in einer vorangegangenen Arbeit abgeleitet wurden (Baur, Emden & Bewersdorff, 2019), weiter empirisch fundiert und definitorisch präzisiert.

2. Lässt sich bei Lehrkräften das Unterrichtshandeln zum Experimentieren durch effektive Lehrkräftefortbildungen weiterentwickeln? Dazu wird in der vorliegenden Arbeit der Einfluss einer Lehrkräftefortbildung auf das Unterrichtshandeln untersucht. In assoziierten Teilprojekten werden Änderungen der Wertvorstellungen, des fachdidaktischen Wissens und der Schülerkognition untersucht. Ziel ist es, einen Beitrag zum Schließen des von Capps, Crawford & Constas (2012) identifizierten Desiderats zu leisten: “We recommend that future studies be designed to investigate the connections between the design of inquiry-focused PD, teacher knowledge, changes in teacher beliefs and practice, and student knowledge” (S. 307).

Aus der ersten Fragestellung (‚Was ist Unterrichtshandeln zum Experimentieren?‘) und der zweiten Fragestellung (‚Ändert sich Unterrichtshandeln durch Fortbildungen?‘) leitet sich die dritte zentrale Fragestellung dieser Arbeit ab:

3. Wie lassen sich Änderungen im Unterrichtshandeln zum Experimentieren erfassen? Hierzu werden Ansprüche an geeignete Instrumente beschrieben und existierende Erhebungsinstrumente bezüglich ihrer Eignung gesichtet. Darauf aufbauend werden geeignete Instrumente adaptiert oder neu entwickelt, auf ihre Güte geprüft und in dieser Studie erstmalig zur Anwendung gebracht.

2 Theoretischer Rahmen

Im Theorieteil wird ausgehend von einer Begriffsverortung der in dieser Arbeit (und im Gesamtprojekt) verwendete Experimentierbegriff vorgestellt (Kapitel 2.1). Die hier vorliegende Arbeit ist dabei Teil des DFG-Projekts ‚Entdeckendes Experimentieren. Wirksamkeit einer Lehrerinnen- und Lehrerfortbildung im Format der Schulwerkstatt' in welchem die Wirkungen einer Lehrkräftefortbildung umfassend analysiert werden (Emden & Baur, 2015). Die Lehrkräftefortbildung selbst wird in Kapitel 3.1 vorgestellt. Nach der Einführung des Experimentierbegriffs erfolgt anhand der in Baur et al. (2019) abgeleiteten Unterrichtsziele und -prinzipien die Identifikation von Aspekten potenziell effektiven Unterrichtshandelns (Kapitel 2.3).

Zur Bestimmung und Klassifizierung der Wirkungen von Lehrkräftefortbildung auf das Unterrichtshandeln (Transferformen und -wirkungen, Kapitel 2.4.1) folgt eine Zusammenschau von Studien zur Wirkungen von Lehrkräftefortbildungen speziell zum Unterrichtshandeln von Lehrkräften beim Experimentieren (Kapitel 2.4.2). Diese Zusammenschau dient dazu, aufgestellte Forschungsfragen und Hypothesen zu erwarteten Änderungen des Unterrichtshandelns (Kapitel 2.5) theoretisch zu verankern und die Studie zu verorten.

Existierende Rating- und Kodiermanuale werden in Kapitel 2.6 vorgestellt. Dieser Überblick dient später im Methodik-Teil der Arbeit der begründeten Auswahl geeigneter Erhebungsinstrumente.

2.1 Relevanz und Definition Experiment

Ausgehend von einer Darstellung der Relevanz der Vermittlung naturwissenschaftlicher Erkenntnismethoden als Beitrag zu naturwissenschaftlicher Grundbildung findet eine Verortung speziell des Experiments als Unterrichtskonzept statt. Unterrichtskonzepte werden im Folgenden als Gesamtorientierungen didaktisch-methodischen Handelns, in denen ein begründeter Zusammenhang von Ziel-, Inhalts- und Methodenentscheidungen hergestellt wird verstanden (Meyer, 1987, S. 208). Anschließend wird das „Entdeckende Experimentieren" vorgestellt. Dieses Unterrichtskonzept bildet den zentralen Inhalt der Lehrkräftefortbildung, welche auf ihre Effektivität hin untersucht wurde.

2.1.1 Erlernen naturwissenschaftlicher Erkenntnismethoden als zentraler Beitrag zu naturwissenschaftlicher Grundbildung

Über die Notwendigkeit einer naturwissenschaftlichen Bildung besteht ein gesamtgesellschaftlicher und bildungspolitischer Konsens (Emden, Bewersdorff & Baur, 2019; Labudde & Möller, 2012). Bereits Dewey stellte heraus, dass naturwissenschaftliche

Bildung mehr ist als das Behalten von Fakten, vielmehr beinhaltet naturwissenschaftliche Bildung das Verständnis naturwissenschaftlicher Methoden und Konzepte und die Kompetenz, diese anzuwenden (Dewey, 1910, 1938).

Im Besonderen durch den Aufbau von Kompetenzen zum Verständnis, mit welchen Methoden naturwissenschaftliches Wissen erlangt wird, und zur Kritik von gewonnenen naturwissenschaftlichen Erkenntnissen sowie deren Implikationen auf den Menschen und die Umwelt sollen die Lernenden befähigt werden, in der Wissensgesellschaft aktiv zu partizipieren.

Durch den Erwerb von Kompetenzen der naturwissenschaftlichen Erkenntnisgewinnung sollen Schüler lernen, wie naturwissenschaftliches Wissen erlangt wird, und somit wissenschaftliche Perspektiven verstehen und hinterfragen können, um in einer wissenschaftsdominierten Welt mit diesem Wissen kritisch umgehen zu können und Entscheidungen bspw. über ihre Alltagsgestaltung, ihre Gesundheit und über menschliche Eingriffe in die Natur treffen zu können. International wurde diese Kompetenz seit dem Beginn des neuen Jahrtausends unter dem Begriff *Scientific Literacy* zusammengefasst. Bybee (2002) definierte (und differenzierte) den Begriff bezogen auf die Naturwissenschaften in vier Dimensionen:

Nominale Scientific Literacy: Lernende auf dieser Stufe können Begriffe, Ideen und Themen aus den Naturwissenschaften mit diesen Gebieten assoziieren, wobei jedoch ein naives, fehlerhaftes oder unvollständiges Verständnis zugrunde liegt.

Funktionale Scientific Literacy: Lernende verwenden naturwissenschaftliche Begriffe korrekt und angemessen. Sie können Texte mit naturwissenschaftlichem Inhalt schreiben und lesen. [3]

Konzeptionelle und prozedurale Scientific Literacy: Lernende haben ein umfassendes Verständnis für Prozesse und Konzepte, welche die Grundlage der Naturwissenschaften bilden, sie können Informationen zu konzeptionellen Ideen in Beziehung setzen und können naturwissenschaftliche Fragen und Konzepte identifizieren, welche naturwissenschaftliche Untersuchungen leiten. Sie können in naturwissenschaftlichen Wirkzusammenhängen denken und ihr Verständnis in Erkenntnisprozessen anwenden.

[3] Dieser Dimension wurde durch explizite Begriffsbildung jahrelang die zentrale Stellung bei der Vermittlung von *Scientific Literacy* zugesprochen bzw. sie wurde mit *Scientific Literacy* im Allgemeinen verwechselt (Bybee, 2002).

Multidimensionale Scientific Literacy: Lernende entwickeln vielfältige Perspektiven auf Naturwissenschaft und verstehen die Naturwissenschaft in ihren persönlichen, gesellschaftlichen und geschichtlichen Kontexten.

Um alle beschriebenen Facetten der *Scientific Literacy* zu fördern, sollte sich naturwissenschaftlicher Unterricht nach Hodson (2014) an vier grundlegenden Zielen orientieren – *learning science, learning about science, doing science* und *learning about socio-scientific issues* – und somit sowohl konzeptuelles Wissen (*learning science* und *learning about science*) als auch prozedurales Wissen (*doing science*) und integrativ über die Naturwissenschaften hinaus vernetzendes Wissen (*learning about socio-scientific issues*) vermitteln.

Die (divergierenden) Begründungen für die Vermittlung von *Scientific Literacy* sind sowohl das Heranbilden der Lernenden zu partizipierenden Staatsbürgern (*education for citizenship*) als auch die Ausbildung einer neuen Generation von Forschenden (OECD, 2017). Auch in Deutschland kam es, unter anderem ausgelöst durch die Ergebnisse der TIMSS- (Baumert, Bos & Lehmann, 2000) und im Wesentlichen durch die PISA-Studien (Baumert, 2012) zu einer Neuausrichtung des naturwissenschaftlichen Unterrichts.

Während in Deutschland prozessbezogene Kompetenzen in den Lehrplänen zuvor nur eine implizite Rolle spielten und der Vermittlung von fachlichen Inhalten nachgeordnet waren (Labudde & Möller, 2012), werden diese Ziele in den Bildungsstandards nun explizit benannt (siehe jeweils S. 6: Konferenz der Kultusminister der Länder, 2005a, Konferenz der Kultusminister der Länder, 2005b, Konferenz der Kultusminister der Länder, 2005c). International hat sich die Zielvorgabe der Vermittlung von Scientific Literacy im naturwissenschaftlichen Unterricht inzwischen ebenfalls in den Curricula und Steuerungsdokumenten etabliert (Australian Curriculum, Assessment And Reporting Authority, 2016; Department for Education, 2014; Hazelkorn et al., 2015; NRC, 2012, s. a. bei Emden 2011 und Baur et al., 2019).

Das Experimentieren als zentrale Erkenntnismethode der Naturwissenschaften (Darian, 2003; Labudde & Möller, 2012) erweitert somit seine Funktion im naturwissenschaftlichen Unterricht von einem vornehmlichen Vermittler von konzeptuellem Wissen, etwa über Demonstrationsexperimente, um den Aspekt des *expliziten* Lerngegenstands selbst zum Erlernen von prozeduralem (*learning science*) und konzeptuellem Wissen (*learning about science*).

In dem Kompetenzbereich Erkenntnisgewinnung (KMK 2005a, 2005b, 2005c) sind die Lernleistungen noch nicht zufriedenstellend (Pant et al., 2013), weswegen die Frage nach dem effektiven Ausbau von Experimentierkompetenz ein zentrales Anliegen naturwissenschaftsdidaktischer Forschung ist.

2.1.2 Das Experiment als Unterrichtskonzept

In diesem Kapitel werden etablierte Konzepte für die Vermittlung von Experimentierkompetenz schlaglichtartig vorgestellt. Die im Folgenden vorgestellten Konzepte decken dabei oft jedoch die naturwissenschaftliche Erkenntnisgewinnung im Allgemeinen ab. Der in dieser Arbeit verwendete Begriff des Experimentierens im naturwissenschaftlichen Unterricht wird definiert.

In der englischsprachigen fachdidaktischen Literatur wird meist nicht speziell das Experimentieren verwendet, sondern die breiter gefassten Begriffe *Inquiry-Based Learning* (Bell, Urhahne, Schanze & Ploetzner, 2010; Prince & Felder, 2007) oder *Inquiry-Based Teaching* (Barron & Darling-Hammond, 2008; Hodson, 2014) bzw. *Inquiry-Based Instruction* (Capps et al., 2012; Capps, Shemwell & Young, 2016) und viele mehr (siehe Cairns & Areepattamannil, 2017). Diese Begriffe werden untereinander teils synonym gebraucht (siehe Zion & Mendelovici, 2012) oder es wird auch in Bezug auf *Science Education* nur von *Inquiry* gesprochen (Banchi & Bell, 2008; Bell, Smetana & Binns, 2005). Forbes, Biggers & Zangori (2013) kommen zu dem Befund, dass "[w]ithin the field of science education, there remains little agreement as to the definition and characteristics of classroom inquiry" (Forbes et al., 2013, S. 1). Dieser Befund teils fehlender konzeptueller Schärfe kann auch für den deutschsprachigen Bereich getroffen werden (Bewersdorff, Baur & Emden, 2020).

Der im Folgenden verwendete Experimentierbegriff steht also im Spannungsfeld zwischen *Inquiry-Based Learning* bzw. *Inquiry-Based Teaching* und *Scientific Inquiry* und ist damit in diesem Verständnis sowohl Methode zum Erlernen neuen Wissens (IBL) als auch Lerngegenstand (IBT) selbst und damit jeweils auch Teilmenge des *Scientific Inquiry*:

Ein Experiment in den empirischen Wissenschaften ist ein wiederholbares, nach expliziten Regeln gestaltetes Verfahren zur Erkenntnisgewinnung (Gut-Glanzman & Mayer, 2018). Gropengießer (2018) benennt sechs Elemente vollständigen Experimentierens: 1. Eine empirisch überprüfbare Hypothese, 2. Ein Untersuchungsplan mit Variablenkontrolle, Kontrollansatz und Replikation, 3. Festlegung und Angabe der zu beobachtenden bzw. zu messenden Daten, 4. Formulierung einer Vorhersage oder Prognose, 5. Dokumentation, Aufbereitung und Auswertung der Daten, 6. Deutung der Versuchsergebnisse mit Bezug auf Prognose und Hypothese.

Experimentieren wird in dieser Arbeit als Erkenntnismethode zur Beantwortung von naturwissenschaftlichen Fragestellungen verstanden, die von Lernenden erlernt und

angewandt werden soll. Dazu wird auf Begriff und Definition des ‚Entdeckenden Experimentierens' (Emden & Baur, 2017) Bezug genommen. Das ‚Entdeckende Experimentieren' wird im Folgenden vorgestellt.

2.1.3 Entdeckendes Experimentieren

In diesem Kapitel wird das ‚Entdeckende Experimentieren' als zentraler Gegenstand der untersuchten Lehrkräftefortbildung vorgestellt (siehe Kap. 3.1.2 Inhalt und Ziele der Fortbildung).

Experimentieren, das streng epistemologisch vorgehend auf ein genuines Erkenntnisinteresse bei Lernenden zielt, bezeichnen Emden und Baur (2017) als ‚Entdeckendes Experimentieren'. Dabei verstehen sie die Vermittlung von Experimentierkompetenz als langfristige Aufgabe des naturwissenschaftlichen Unterrichts explizit über Klassenstufen und Fächergrenzen der einzelnen naturwissenschaftlichen Disziplinen hinweg. Schrittweise – aber nicht notwendigerweise linear – kommt den Lernenden dabei bei den einzelnen Teilprozessen des Erkenntnisprozesses eine zunehmende Eigenverantwortlichkeit zu (Baur, Hummel, Emden & Schröter, 2020). Bei den Teilprozessen orientieren sich Emden und Baur (2017) an der von Mayer und Ziemek (2006) vorgeschlagenen Ordnung: naturwissenschaftliche Frage stellen, Hypothese(n) aufstellen, Experiment planen, Experiment durchführen, Experiment auswerten, Schlussfolgerungen ziehen. Lernende sollen die Teilprozesse der experimentellen Erkenntnisgewinnung selbstständig und reflexiv zu einer kohärenten Untersuchung fügen können. Im Teilprozess „Experiment durchführen" sollen sie zu wissenschaftlicher Genauigkeit angehalten werden, was eine sich anschließende Reflexion der Grenzen und Unsicherheiten der vorgenommenen Untersuchung erfordert.

Konkret wird Entdeckendes Experimentieren durch sechs Merkmale charakterisiert: 1. Der Experimentierprozess wird von Lernenden ausgeführt, 2. Der Verlauf orientiert sich an den Phasen des wissenschaftlichen Experimentierprozesses, 3. Die innere Logik/Abhängigkeit des Erkenntnisprozesses wird reflektiert („Der Epistemologie verpflichtet"), 4. Der Experimentierprozess führt die Lernenden zu einer (für sie) neuen Erkenntnis, 5. Lernenden kommt eine zunehmende Eigenverantwortlichkeit zu, 6. Beim Experimentieren wahren die Lernenden wissenschaftliche Strenge.

2.2 Befunde zum aktuellen Unterrichtshandeln beim Experimentieren

In diesem Kapitel werden der Begriff des Unterrichtshandelns eingeführt und anschließend Befunde zum Unterrichtshandeln von Lehrkräften beim Experimentieren vorgestellt.

2.2.1 Zum Begriff des Unterrichtshandelns

Unterrichtshandeln wird auf unterschiedliche Art und Weise konzeptualisiert (vgl. Holzberger, Philipp & Kunter, 2016). Das Unterrichtshandeln wird von Li und Oliveira (2015) als *classroom practice* – im Kontext des Mathematikunterrichts – wie folgt charakterisiert:

> „Classroom practice, as a process, involves multiple agents and their interactions within the classroom as a system. The process can be manifested in diverse formats and structures, and its effectiveness can be influenced by numerous factors both internal and external to the classroom. Research on classroom practice can thus take different perspectives“ (Li & Oliveira, 2015, S. 489).

Diese Charakterisierung lässt Umfang sowie Abhängigkeiten, sprich die Komplexität des Unterrichtshandelns, erahnen.

Zum Unterrichtshandeln, auch speziell bezüglich des Experimentierens, scheint keine konsensuale Begriffsdefinition zu existieren, auf die rekurriert werden kann. Vielmehr ist hier der Begriff des Unterrichtshandelns oft Begriffshülle für einzelne Teilaspekte des Unterrichtshandelns oder Perspektiven auf das Unterrichtshandeln. Die Verschiedenheit jeweils einbeschriebener Aspekte des Unterrichtshandelns können der Zusammenschau von Erhebungsinstrumenten zur Erfassung von effektivem Unterrichtshandeln in Kapitel 2.6 (Tabelle 4) entnommen werden.

Unterricht ist die „Gestaltung von Lernumgebungen mit dem Ziel, optimale Gelegenheiten für die effektive Ausführung von Lernaktivitäten der Schüler/innen bereitzustellen“ (Seidel & Reiss, 2014, S. 254).

In dieser Arbeit wird unter dem Begriff Unterrichtshandeln, basierend auf der Definition des Unterrichts von Seidel & Reiss (2014), das pädagogisch-didaktische Handeln sowie die Kommunikation der Lehrperson im Lehr-Lern-Raum verstanden. Auch die Strukturierung des Unterrichts wird dem Unterrichtshandeln zugeordnet (vgl. Wagner, 2016).

2.2.2 Aktuelle Befunde zum Unterrichtshandeln beim Experimentieren

Lehramtsstudierende haben ein limitiertes Verständnis von *Nature of Science* allgemein (Abd-El-Khalick, 2006a; Höttecke & Rieß, 2007), aber auch speziell von *Scientific Inquiry* (Capps et al., 2016). Obwohl die Werthaltungen (*Beliefs*) zu *Scientific Inquiry* prinzipiell positiv sind (Engeln, Euler & Maass, 2013; Maaß & Euler), gibt es wohl Hindernisse bei der Implementation im Unterricht – möglicherweise auch allein schon, weil Lehrkräfte ein falsches Verständnis (‚Hands on‘ allein genügt schon dem Ansatz des *Scientific Inquiry*) oder nur ein rudimentäres Verständnis des Konzeptes

haben (Capps et al., 2016). Eine umfassende Umsetzung von *Scientific Inquiry* – oder nahestehenden Unterrichtskonzepten – ist keinesfalls trivial und damit für Lehrkräfte eine große Herausforderung (Engeln, Mikelskis-Seifert & Euler, 2014).

So werden praktische Arbeitsphasen vornehmlich dazu verwendet, die Lernenden ein bereits bekanntes Ergebnis mithilfe von Schritt-für-Schritt-Anleitungen verifizieren zu lassen (Capps & Crawford, 2012). Außerdem werden Experimente selten effektiv eingesetzt, viele Schülerexperimente sind trivial (Tesch & Duit, 2004), nicht zielgerichtet und beziehen sich oft nicht auf Fähigkeiten und Interessen der Lernenden (Harlen, 1999). Um den ‚Überraschungseffekt' nicht zu zerstören, verzichten viele Lehrkräfte bewusst auf eine – für ein wissenschaftsorientiertes Vorgehen zwingend notwendige – Zielbeschreibung bzw. Hypothese (Tesch & Duit, 2004). Eine eigene Planung von Experimenten oder die Auswertung von Daten wird von Schülerinnen und Schülern nur selten verlangt (Tesch & Duit, 2004). Nur wenig Zeit wird auf die Entwicklung der Planungsideen der Lernenden verwendet (Abrahams & Millar, 2008).

Aus den beschriebenen Befunden kann geschlossen werden, dass Lehrkräfte effektives Unterrichtshandeln zum Experimentieren selten zeigen. Im Folgenden sollen nun Aspekte effektiven Unterrichts zum Experimentieren vorgestellt werden.

2.3 Aspekte effektiven Unterrichtshandelns zum Experimentieren

In diesem Kapitel werden Unterrichtsprinzipien zum effektiven Unterrichtshandeln zum Experimentieren abgeleitet und empirisch fundiert. Bei der Auswahl geeigneter Rating- und Kodiermanuale zur Erfassung der Änderung von Unterrichtshandeln beim Experimentieren sollen diese Unterrichtsprinzipien als ‚Lastenheft' leiten (Kapitel 3.4).

Obwohl der in dieser Arbeit verwendete Begriff des Experimentierens enger gefasst ist als die Begriffe des *Inquiry-Based Teaching* bzw. *Inquiry-Based Learning* und *Scientific Inquiry*, in deren Spannungsfeld sich das Experimentieren verorten lässt, wird im Folgenden auch auf Forschungsliteratur aus dem gesamten Spektrum des *Inquiry*-Begriffs zurückgegriffen. Auch wenn ein Verständnis dieses Begriffs *Inquiry* in der naturwissenschaftsdidaktischen Forschung nicht klar definiert ist und je nach Kontext eine unterschiedliche Bedeutung erfährt (Crawford, 2014; Engeln et al., 2014).

Teils wird die Lernwirksamkeit der *Inquiry*-Ansätze grundlegend infrage gestellt (Kirschner, Sweller & Clark, 2006) und vereinzelt mit geringerem Naturwissenschaftsverständnis in Verbindung gebracht (Cairns & Areepattamannil, 2017). Teig, Scherer und Nilsen (2018) geben im Rahmen einer Analyse von norwegischen Daten

der TIMS-Studie einen Erklärungsansatz für die Ergebnisse der Analyse von Cairns und Areepattamannil (2017): So entwickelt sich das Naturwissenschaftsverständnis nicht linear mit der Menge der Unterrichtsstunden, welchen *Inquiry*-Ansätze zugrunde liegen. Vielmehr gibt es wohl einen ‚Deckeneffekt', welchem sich das Naturwissenschaftsverständnis der Lernenden mehr asymptotisch denn linear nähert. Wie jedes anderes Unterrichtskonzept sind *Inquiry*-Ansätze nicht an sich effektiv, der Erfolg ist abhängig von sowohl der konkreten Implementation als auch dem damit verfolgten Ziel (Vorholzer & Aufschnaiter, 2019) und unltmativ von den Lernenden selbst.

Es gibt dennoch in einer Vielzahl von Publikationen Hinweise, die die Wirksamkeit von *Inquiry*-Ansätzen auf drei der vier Wissensdimensionen (Anderson, 2009) konzeptuelles Wissen, prozedurales Wissen und metakognitives Wissen empirisch stützen.

Inquiry-Ansätze führen nach der Metastudie von Furtak, Seidel, Iverson und Briggs (2012) allgemein zu höheren Lernerfolgen im Vergleich zu Gruppen, welche traditionell unterrichtet werden. *Inquiry* hat einen positiven Einfluss auf die Entwicklung wissenschaftlicher Arbeitsweisen, konzeptuellen Verständnisses und von „*Active Thinking*" (Dimensionen: prozedurales und konzeptuelles Wissen; Minner, Levy & Century, 2010). Naturwissenschaften im Rahmen eines *Inquiry*-Prozesses zu lernen, entwickelt bei den Lernenden extensives Wissen der Prozeduren, welche notwendig sind, um neues Wissen zu generieren („zu Wissen wie": prozedurales Wissen; Osborne, 2014b). *Inquiry*-orientierte Unterrichtskonzepte werden auch als wesentlich für die Entwicklung einer Scientific Literacy (konzeptuelles Wissen) gesehen (Engeln et al., 2014). Lernende, welche *Inquiry*-Aktivitäten durchgeführt haben, haben dadurch ihre metakognitiven Fähigkeiten (Dimension: metakognitives Wissen) benutzt und diese ausgebaut (Kipnis & Hofstein, 2008).

Aufgrund der vielen ‚*Inquiry*' einbeschriebenen Aspekte existiert auch eine große Breite von postulierten Teilaspekten des *Inquiry*, welche im Hinblick auf Unterrichtshandeln als potenziell effektiv gelten können.

Aufgrund dieser Bandbreite des Begriffs *Inquiry* und der Vielzahl seiner in den unterschiedlichen Arbeiten in den Blick genommenen Facetten ist es zur Untersuchung effektiven Unterrichtshandelns sowohl zum (Entdeckenden) Experimentieren (als Teilmenge der *Inquiry*-Ansätze) als auch darauf aufbauend für die Konzeption eines Kodiermanuals zur Beschreibung der Entwicklung von Unterrichtshandeln zum Experimentieren hilfreich, eindeutige Unterrichtsziele und -prinzipien zum Aufbau experimenteller Kompetenzen theoretisch begründet abzuleiten.

Die originäre Ableitung der Unterrichtsziele und -prinzipien erfolgte bei Baur et al. (2019) multidimensional auf Basis von curricularen, fachlich-disziplinären, bildungstheoretischen und lerntheoretischen Begründungen. Aus diesen vier Perspektiven wurden in einem ersten Schritt vier Lernziele identifiziert:

1. Die Lernenden können Experimentieren als Strategie beschreiben: Die Lernenden kennen die unterschiedlichen Teilprozesse und deren Funktion.
2. Die Lernenden können die innere Logik eines experimentellen Vorgehens erklären.
3. Die Lernenden können Grenzen der Aussagekraft von Experimenten reflektieren.
4. Die Lernenden können Experimentieren zum Erkenntnisgewinn selbstständig anwenden.

Anschließend wurden den abgeleiteten Unterrichtszielen acht Unterrichtsprinzipien zugeordnet, welche zum Erreichen der (langfristig angelegten) Unterrichtsziele als notwendig erscheinen. Als Unterrichtsprinzipien werden dabei didaktische Grundsätze aufgefasst, welche das Unterrichtsplanen und -handeln leiten. Die Ableitung der Unterrichtsziele und Unterrichtsprinzipien wird im Folgenden theoretisch nachgezeichnet und durch weitere Befunde ergänzt.

In internationalen Curricula spiegelt sich die Forderung nach der Vermittlung von prozeduralem Wissen und epistemologischem Verständnis, also von dem Verständnis naturwissenschaftlichen Erkenntnismethoden, etwa dem Experimentieren, neben dem Inhaltswissen wieder (Australian Curriculum, Assessment And Reporting Authority, 2016; Department for Education, 2014; Konferenz der Kultusminister der Länder, 2005b; Ministry of Education, 2007; NRC, 2012). Die Rahmensetzungen der Kultusministerkonferenz der Länder (KMK) im Fach Chemie sollen exemplarisch aufgeführt werden:

> „Die Schülerinnen und Schüler erkennen und entwickeln Fragestellungen […].
>
> Die Schülerinnen und Schüler planen geeignete Untersuchungen zur Überprüfung von Vermutungen und Hypothesen, […].
>
> Die Schülerinnen und Schüler führen qualitative und einfache quantitative experimentelle und andere Untersuchungen durch und protokollieren diese.
>
> Die Schülerinnen und Schüler erheben bei Untersuchungen, insbesondere in chemischen Experimenten, relevante Daten oder recherchieren sie.

> Die Schülerinnen und Schüler finden in erhobenen oder recherchierten Daten Trends, Strukturen und Beziehungen, erklären diese und ziehen geeignete Schlussfolgerungen“ (Konferenz der Kultusminister der Länder, 2005b, S. 12).

Wie exemplarisch aus den hier auszugsweise dargestellten Bildungsstandards der KMK hervorgeht, sollen die Lernenden die einzelnen Teilprozesse des Experiments selbstständig gestalten. Aus der bildungspolitischen (curricularen) Perspektive lassen sich somit die Ziele „Die Lernenden können Experimentieren als Strategie beschreiben: Die Lernenden kennen die unterschiedlichen Teilprozesse und deren Funktion“ sowie „Die Lernenden können Experimentieren zum Erkenntnisgewinn selbstständig anwenden“ ableiten.

Das Lernen im Rahmen handlungsorientiert-konstruktivistischer Lehr-Lern-Umgebungen hat sich in der Naturwissenschaftsdidaktik etabliert und zahlreiche Unterrichtskonzepte hervorgebracht, welche das sogenannte. *Active Learning* fördern sollen (Barron & Darling-Hammond, 2008; Hood Cattaneo, 2017). Bei einem Verständnis von Lernen als Konstruktion sollen Lernende unter größtmöglicher Selbsttätigkeit und Selbstbestimmung eigene „Weisen der Welterzeugung“ konstruieren (Neubert, Reich & Voß, 2001, S. 260). Wissen und Kompetenzen können also nach der konstruktivistischen Lehr-Lern-Theorie nicht ‚von außen‘ oktroyiert werden, sondern entwickeln und entfalten sich im Individuum. Die Lernumgebung inklusive der Lehrpersonen bildet nach konstruktivistischer Lehre also allein einen geeigneten Rahmen für den Lernprozess des Individuums. Daraus kann gefolgert werden, dass Lernende auch beim Experimentieren möglichst selbstständig planen und Versuche durchführen sollen. Aus dieser Forderung kann folgendes Ziel formuliert werden: „Die Lernenden können Experimentieren zum Erkenntnisgewinn selbstständig anwenden.“ Bei einem Verständnis von Lernen als Dekonstruktion wird die Notwendigkeit betont, eingespielte Wirklichkeitskonstruktionen kritisch zu hinterfragen und dadurch ultimativ zu durchdringen (Neubert et al., 2001). Dies kann besonders für die oft unreflektiert ‚abgearbeiteten‘ Prozessschritte des Experiments gelten. Das kritische Hinterfragen der Notwendigkeit der einzelnen Teilprozesse wie auch ihrer Reihenfolge kann den Lernenden die Sinnhaftigkeit des Vorgehens, und damit die Akzeptanz des Handelns, deutlich machen. Aus diesen Überlegungen lässt sich folgendes Ziel ableiten: „Die Lernenden können die innere Logik eines experimentellen Vorgehens erklären“.

Unter den naturwissenschaftlichen Erkenntnismethoden nimmt das Experiment eine herausgehobene Stellung ein (Darian, 2003). Um die epistemologische Komplexität im Unterricht zu verringern, wird das Experiment oft reduziert als Folge von Teilprozessen (z. B. Gut-Glanzmann & Mayer, 2018; Mayer & Ziemek, 2006) dargestellt. Die Teilprozesse müssen dabei nicht stets linear von der Fragestellung über die Hy-

pothese, Planung, Durchführung, Auswertung zur Erkenntnis durchlaufen werden, jedoch ist ein Verständnis der Funktion und der Beziehungen der Teilprozesse untereinander notwendig, um das Experiment als naturwissenschaftliche Erkenntnismethode beschreiben zu können. Dies lässt sich als Ziel formulieren: „Lernende können Experimentieren als Strategie beschreiben: Die Lernenden kennen die unterschiedlichen Teilprozesse und deren Funktion.“ Zentrales Merkmal der Naturwissenschaften ist ein exaktes, sauberes, gewissenhaftes und nachvollziehbares Arbeiten. Diese Aspekte wissenschaftlicher Strenge (Scientific Rigor) und wissenschaftlicher Konventionen müssen mit den Lernenden explizit thematisiert und von ihnen internalisiert werden (Vorholzer, 2017). Beim Experimentieren ist daher eine Diskussion über die Aussagekraft der durch das Experiment gewonnenen Erkenntnisse auch im Hinblick auf mögliche (Mess-)Fehler und die Reproduzierbarkeit notwendig. Daraus lässt sich folgendes Ziel ableiten: „Die Lernenden können Grenzen der Aussagekraft von Experimenten reflektieren.“

Das Experiment verfügt sowohl über material wie auch formal bildende Eigenschaften (Emden et al., 2019). Aus der Perspektive materialer Bildungstheorien, welche eine Vermittlung von grundlegendem Wissen und Fertigkeiten fordern, lässt sich das Ziel „Die Lernenden können Experimentieren als Strategie beschreiben: Die Lernenden kennen die unterschiedlichen Teilprozesse und deren Funktion“ begründen. Aus der Perspektive funktionaler Bildung (einer Entwicklung von ‚Kräften‘ und Fähigkeiten zum Übertrag auf andere Problembereiche) ist es relevant, dass Lernende durch ihr Wissen über das Experiment, seine Funktion sowie seine Grenzen befähigt werden, andere naturwissenschaftliche Erkenntnisse bezüglich ihrer Aussagekraft bewerten zu können. Aus bildungstheoretischer Perspektive folgt daraus folgendes Ziel: „Die Lernenden können Grenzen der Aussagekraft von Experimenten reflektieren.“

Die abgeleiteten Unterrichtsziele sind als Leitziele zu verstehen, welche die Lernenden langfristig, etwa im Verlauf der Sekundarstufe I, erreichen sollen. Um diese Ziele zu erreichen, benötigt es geeignete Unterrichtskonzepte. Nachfolgend werden zu den abgeleiteten Unterrichtszielen Unterrichtsprinzipien aufgestellt, welche bei der Entwicklung von Unterrichtskonzepten und Unterrichtseinheiten zum Experimentieren berücksichtigt werden sollen. Auch vonseiten der Unterrichtsforschung werden von Miranda und Damico (2015) handhabbare, praxistaugliche Unterrichtsprinzipien gefordert, um Unterrichtshandeln bei naturwissenschaftlichen Untersuchungen effektiv erfassen zu können.

Der Begriff der Unterrichtsprinzipien soll sich hierbei an Glöckel (2003, S. 280) orientieren: „Prinzipien sind flexibler als Regeln und weniger zwingend als Gesetze, sie fordern ein bewegliches Entscheiden gemäß der Situation. Sie erlauben die begründete Ausnahme, schließen aber die bloße Willkür aus [...].“

Das Unterrichtsziel Z1 „Die Lernenden können Experimentieren als Strategie beschreiben: Die Lernenden kennen die unterschiedlichen Teilprozesse und deren Funktion“ wurde curricular, bildungstheoretisch sowie fachlich-disziplinär abgeleitet. Aus diesem Unterrichtsziel ergeben sich unmittelbar die Prinzipien, dass die Teilprozesse des Experimentierens berücksichtigt werden und dass das Experimentieren als Strategie der Erkenntnisgewinnung expliziter Unterrichtsgegenstand ist. Da nachhaltiges Lernen nicht ohne Üben denkbar ist (Heymann, 2012), muss auch das Experimentieren systematisch geübt werden (P3: Experimentieren üben).

Das Unterrichtsziel Z2 „Die Lernenden können die innere Logik eines experimentellen Vorgehens erklären“ wurde lerntheoretisch begründet. Da die Lernenden die Teilprozesse sowie ihre Funktion kennen müssen, um die innere Logik des Experiments erklären zu können, ergibt sich, dass die aus Z1 abgeleiteten Prinzipien ebenfalls für das Erreichen von Z2 notwendig sind (P1: Teilprozesse; P2: expliziter Unterrichtsgegenstand; P3: Üben). Um die innere Logik des Experiments zu erschließen, ist auch eine explizite Prozessreflexion beim spezifischen Experiment erforderlich (P5: Prozessreflexion).

Das Experiment ist eine Form des Problemlösens (Mayer, 2007): Um die innere Logik des Experiments zu erschließen, muss dieses auch im Kontext und mit Bezug auf das Problem geschehen (P4: Experimentieren zum Problemlösen einsetzen), da die Art der Problemstellung das experimentelle Vorgehen beeinflusst.

Damit die Lernenden die Grenzen der Aussagekraft von Experimenten reflektieren können (Z3, abgeleitet aus der fachlich-disziplinären sowie der bildungstheoretischen Perspektive), ist es geboten, die Lernenden selbst das Experimentieren zum Lösen von Problemen einsetzen zu lassen (P4: Problemlösen), um realitätsnah über Ergebnisse und ihre Limitationen sprechen zu können (P6: Grenzen von Experimenten ansprechen). Wissenschaftliche Strenge beim Experimentieren (P7) ist die Grundlage der anschließenden Diskussion über die Grenzen der Aussagekraft der Experimente.

Das Unterrichtsziel Z4 „Die Lernenden können Experimentieren zum Erkenntnisgewinn selbstständig anwenden“ wurde aus der Lerntheorie sowie curricular abgeleitet. Die Unterrichtsziele Z1–Z3 bereiten den Boden für das Erreichen von Unterrichtsziel Z4, indem sie über die jeweils artikulierten Unterrichtsprinzipien systematisch zum Verständnis der einzelnen Facetten des Experiments beitragen. Unmittelbar aus dem Unterrichtsziel Z4 ergibt sich das Unterrichtsprinzip der Selbstständigkeit der Lernenden beim Experimentieren (P8). Das Unterrichtsziel Z4 soll nicht erst abschließend nach Erreichen der Ziele Z1–Z3 forciert werden, vielmehr ist eine schrittweise Übergabe der Verantwortung über die einzelnen Teilprozesse hinweg anzustreben (Banchi

& Bell, 2008; Baur & Emden, 2020; Bell et al., 2005). Abbildung 1 zeigt eine Zusammenschau der Ziele und Prinzipien:

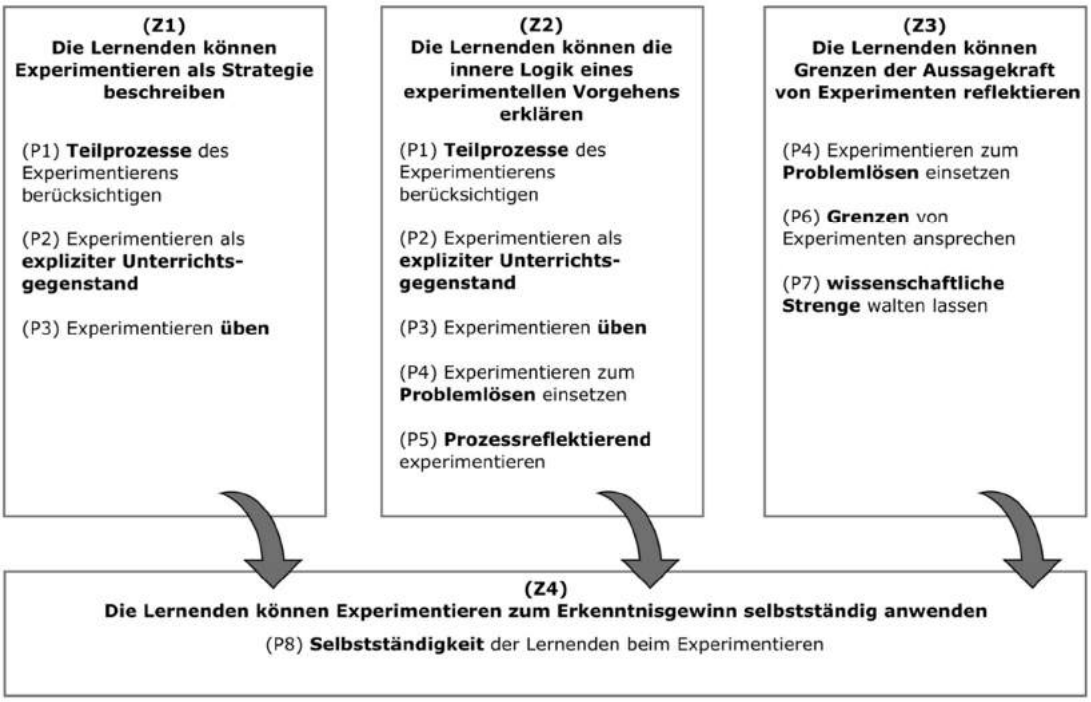


Abbildung 1: Übersicht über die abgeleiteten Unterrichtsziele und die zugehörigen Unterrichtsprinzipien (Baur et al., 2019)

Im folgenden Abschnitt sollen nun die zuvor aus den Unterrichtszielen abgeleiteten Unterrichtsprinzipien durch empirische Befunde gestützt und damit weiter legitimiert werden. Im Hinblick auf das zu entwickelnde Beobachtungsinstrument sollen die Unterrichtsprinzipien außerdem weiter ausgeschärft werden. Anhand von Literatur sollen zudem Merkmale für eine wirksame Implementation der Prinzipien im Unterricht kondensiert werden, um diese bei der Entwicklung des Manuals aufgreifen zu können.

P1: Teilprozesse des Experimentierens berücksichtigen

Inwieweit der gesamte Experimentierprozess erfolgreich verläuft, ist abhängig davon, wie erfolgreich die einzelnen Teilprozesse durchlaufen werden (Härtig, Neumann & Erb, 2017).

Wie bereits in Kapitel 2.1 dargelegt, darf nicht der Eindruck entstehen, dass es *die* experimentelle Methode gibt (‚*The Scientific Method*': Alexakos, 2010; Lederman, 2007; Windschitl, Thompson & Braaten, 2008), vielmehr ist Erkenntnisgewinnung auch ein kreativer Prozess, dessen Verlauf nicht einer starren Abfolge unterliegt, sondern welcher Sprünge und Wiederholungen erlaubt (Bell, 2012). Die Teilprozesse treten also nicht zwingend als klar voneinander abgetrennte Phasen, noch in der stets

gleichen Reihenfolge auf. Doch wird das Vorhandensein von Teilprozessen in der Didaktik als konstituierend für den Experimentierprozess angesehen (Emden, 2021; Härtig et al., 2017).

Unabhängig davon, wie im spezifischen Anwendungsfall der Prozess der Erkenntnisgewinnung ausgestaltet ist, benötigen die Lernenden das Wissen um die Funktion und Abhängigkeit der einzelnen Teilprozesse. Didaktisch leitende Idee bei der Ausdifferenzierung des wissenschaftlichen Erkenntnisprozesses in einzelne Teilprozesse ist die dadurch geschaffene Möglichkeit einer differenzierten Analyse dieser einzelnen Teilprozesse mit den Lernenden (Gut-Glanzmann & Mayer, 2018) im Sinne einer didaktischen Reduktion. In Bezug auf den Experimentierprozess gibt es jedoch in der Literatur keinen Konsens darüber, welche Teilprozesse dieser Prozess beinhaltet, alle Teilprozesse können jedoch auf einen basalen Dreischritt – 1) Formulierung von Fragen und Hypothesen, 2) Planung von Untersuchungen sowie 3) Auswertung und Interpretation von Daten (Emden, 2011; Vorholzer, Aufschnaiter & Kirschner, 2016) – zurückgeführt und in das SDDS-Modell (*Scientific Discovery as Dual Seach*: Klahr & Dunbar, 1988) einbeschrieben werden (Emden, 2011). Klahr und Dunbar (1988) beschreiben aus kognitionspsychologischer Sicht den hypothesenprüfenden wissenschaftlichen Erkenntnisprozess als Suche in zwei getrennten Räumen: dem Hypothesen-Raum und dem Experiment-Raum. Der Hypothesen-Raum wird durch das Vorwissen der Personen und die ermittelten Ergebnisse aus dem Experiment-Raum bestimmt, der Experiment-Raum von den verfügbaren Methoden. Als zeitliche Abfolge manifestiert sich die wechselseitige Suche im Hypothesen- bzw. Experiment-Raum in den drei Teilprozessen ‚*Search Hypothesis Space*', ‚*Test Hypothesis*' und ‚*Evaluate Evidence*', welche mit dem basalen Dreischritt in Deckung gebracht werden können.

Eine schematisierende Untergliederung des Experiments in Teilprozesse findet sich etwa bei Mayer und Ziemek (2006) in Problem, Planung, Durchführung, Auswertung und Deutung. Weitere Möglichkeiten der Untergliederung stellen Pedaste et al. (2015) in einem Review vor. Walpuski (2006) konnte für Strukturierungshilfen (Flussdiaramme mit abgebildeten Teilprozessen des Experiments) keine signifikanten Effekte auf die Experimentierkompetenz nachweisen. Da die Flussdiagramme in dieser Studie zuvor bei den Lernenden jedoch nicht eingeführt wurden, ist eine zusätzliche kognitive Belastung der Lernenden denkbar, welche eine Erklärung für die unbefriedigenden Leistungsdaten bietet. In einer Nachfolgestudie kommt Wahser (2007) zu dem Schluss, dass, um die Lernenden nicht zu überfordern, es sinnvoll ist zuerst einen didaktisch reduzierten Prozessablauf zu vermitteln. Es besteht, trotz vereinzelt abweichender Studien, allgemein Einigkeit bezüglich einer geringen Wirksamkeit von generell ungeleitetem Experimentieren (vgl. Kirschner et al., 2006; Sweller, Kirschner & Clark, 2007).

Somit lässt sich die Notwendigkeit der Berücksichtigung der Teilprozesse des Experimentierens mittels drei Argumentationslinien begründen:

1. Die Teilprozesse sind integraler Bestandteil eines Experiments, ohne Kenntnis ihrer Funktion und ihrer gegenseitigen Beziehungen ist verständiges Experimentieren in der Schule nicht möglich (vgl. Darian, 2003, Emden 2021).

2. Die Untergliederung des Experiments in Teilprozesse bildet eine notwendige didaktische Reduktion des Experiments (Osborne, 2011), worauf später ein elaborierteres Experimentierverständnis unter Beachtung der Aspekte der *Nature of Science* (etwa Berland et al., 2016) aufgebaut werden kann.

3. Die Teilprozesse dienen den Lernenden als Strukturierungs- und Orientierungshilfe und können so, als Teil von *Scaffolding*-Maßnahmen, zu einer Reduzierung des *Cognitive Load* (Entlastung des Arbeitsgedächtnisses) führen (vgl. Engl, 2017). Unter *Scaffolds* versteht man im Allgemeinen und direkter Übersetzung Gerüste zur Lernunterstützung, um eine Überforderung der Lernenden bei komplexen Aufgaben, wie etwa dem Experimentieren, zu reduzieren (vgl. Arnold, Kremer & Mayer, 2017).

P2: Experimentieren als expliziter Unterrichtsgegenstand

Die Lernenden empfinden praktische Arbeit als relevant und angenehm (Murray & Reiss, 2005), doch führt praktische Arbeit bzw. das Durchführen von Experimenten – also ein implizites Erfahren naturwissenschaftlichen Arbeitens – allein nicht unbedingt zu einem vertieften Verständnis des Experiments (Bell, Blair, Crawford & Lederman, 2003; Gott & Duggan, 2007; Lederman, Lederman, Sug Kim & Kyung Ko, 2012; Osborne, 2014a). Lange Zeit wurde *Doing Science,* also das händische, praktische Erfahren von naturwissenschaftlichen Erkenntnismethoden, mit *Learning Science,* dem Erlernen und Verstehen von Erkenntnismethoden gleichgesetzt. Dadurch hat sich in der Argumentation für den Einsatz von *Scientific Inquiry* die Vorstellung, dass naturwissenschaftliche Arbeitsweisen implizit gelehrt werden können, lange erhalten (Chinn & Malhotra, 2002). Obwohl sowohl explizite als auch implizite Ansätze zum Aufbau von Experimentierkompetenz bei Lernenden beitragen können (Dean & Kuhn, 2007), kommen empirische Studien mehrheitlich zu dem Ergebnis, dass explizite Ansätze deutlich effektiver sind (Vorholzer, 2015).

Studien zeigen, dass Lernende – aber auch Lehrende (Watson, Swain & McRobbie, 2007) – oft annehmen, der Zweck für die durchgeführte Untersuchung sei es, eine Folge von Anweisungen abzuarbeiten (‚Kochrezept') oder das ‚richtige' Ergebnis zu erzielen (vgl. Hofstein & Lunetta, 2004). Bei bisherigen naturwissenschaftlichen Untersuchungen wurde der Fokus häufig auf handwerkliche Tätigkeiten gelegt (Windschitl et al., 2008), auf Kosten eines tieferen Verständnisses der Methode und des

epistemischen Rahmens. Lernende führen also naturwissenschaftliche Untersuchungen durch, ohne ein klares Verständnis der Prozedur und des Zwecks der Untersuchung zu haben (Fisher, Harrison, Henderson & Hofstein, 1998; Roth & Garnier, 2007). Es ist also notwendig, das Experiment in Form eines expliziten Unterrichtsgegenstands mit den Lernenden explizit zu besprechen (Clough, 2006). Bei Studierenden, also im Hochschulkontext, scheint eine explizite Thematisierung allenfalls leicht positive Effekte zu zeigen (Kalthoff, Theyssen & Schreiber, 2018). Erklärungsmächtig für diese Ergebnisse, welche im scheinbaren Widerspruch zu den Resultaten von Vorholzer (2015) stehen, können hier – knapp angerissen – möglicherweise zwei Aspekte werden:

1. Die Studierenden haben bereits, etwa durch eine in der Schule bereits ausgeprägte Affinität zu den Naturwissenschaften, ein stabiles Verständnis des Experimentierens, ein explizites Thematisieren gibt den Lernenden keine zusätzlichen Informationen.
2. Die Studierenden haben eine erhöhte Problemlösekompetenz und kognitive Flexibilität, womit wiederum fehlende Informationen selbst erlangt werden können oder deren Erwerb umgangen werden kann.

Das explizite Thematisieren kann nachgelagert in Reflexionsphasen, aber auch durchaus *während* des Experimentierens geschehen (Henke, 2016). So zeigen Studien die Wirksamkeit der expliziten Thematisierung von Teilaspekten oder Teilprozessen des Experiments im Rahmen einer konkreten Untersuchung zur Entwicklung eines Verständnisses über *Nature of Science* (Abrahams & Millar, 2008; Ryder & Leach, 1999). Es muss also sichergestellt werden, dass Lernende auch die inhaltlichen und methodischen Hintergründe beim Experimentieren verstehen und nicht nur schlimmstenfalls die einzelnen Prozessschritte in einem ‚*hands on*, *minds off*'-Modus abarbeiten. Dazu müssen das Experiment und seine Teilprozesse explizit vermittelt werden (Hammann & Mayer, 2012).

P3: Experimentieren üben

Über die Notwendigkeit von Übungen zum nachhaltigen Lernen besteht lernpsychologisch (Bachmann, 2003) und allgemeindidaktisch (Brinkmann, 2009; Heymann, 2012) Konsens, dennoch wird das Üben zum Erlernen naturwissenschaftlicher Arbeitsweisen, wie etwa dem Experimentieren, kaum explizit eingefordert (Bewersdorff et al., 2020). Gräber (2002) postuliert, dass selbstgesteuertes Lernen in den Naturwissenschaften von den Lehrenden nicht unvorbereitet gefordert werden darf, sondern gezielt eingeführt werden muss sowie ein intensives Einüben der notwendigen Kompetenzen voraussetzt. Diese notwendigen Kompetenzen können im Anfangsunterricht, als Fundament für das später angestrebte selbstständige Experimentieren, teils

noch über „kochrezeptartige Versuchsvorschriften“ eingeübt werden (Emden, 2021; Metzger & Sommer, 2010). Nathan (2012) argumentiert noch strenger und fordert in seinem „*Formalisms First*“-Ansatz, dass Lernende zu Beginn reduzierte Problemlösealgorithmen einüben, bevor diese in praxisnahen Problemstellungen angewendet werden. Zentral ist hierbei, dass dieser Kompetenzaufbau nicht auf dem Niveau der mechanischen Problemlösung verbleibt und das Experiment – oder naturwissenschaftliche Erkenntnisgewinnung allgemein – von den Lernenden letztlich als eine starre Reihenfolge von Teilprozessen („*the Scientific Method*“: Alexakos, 2010; Lederman, 2007; Windschitl et al., 2008) missverstanden wird. Denn mit dieser übervereinfachten algorithmischen Form naturwissenschaftlichen Denkens ist es wahrscheinlich, dass die Lernenden wissenschaftliche Argumente als wenig relevant für realweltliche Entscheidungen ansehen (Chinn & Malhotra, 2002). Bei diesem Befund ist jedoch kritisch zu hinterfragen, ob die algorithmische Vorstellung von Erkenntnisgewinnung Ursache oder doch nur ein weiteres Symptom eines allgemein reduzierten Naturwissenschaftsverständnisses ist. Die Variablenkontrollstrategie als ein für die Lernenden kognitiv besonders anspruchsvoller Teilaspekt des Experimentierens muss durch Übung und Anleitung entwickelt werden (Schwichow, Croker, Zimmerman, Höffler & Härtig, 2016). Nur Übung und explizite Anwendung werden also den Lernenden helfen, ein angemessenes Verständnis des Experimentierens zu entfalten (Schauble, Klopfer & Raghavan, 1991).

P4: Experimentieren zum Problemlösen einsetzen

Naturwissenschaftliche Erkenntnisprozesse wie das Experimentieren können als Problemlöseprozess beschrieben werden (Hammann, Ho & Bayrhuber, 2007; Mayer, 2007). Ein Problem ist nach Dörner (1976) gekennzeichnet durch einen unerwünschten Anfangszustand, einen erwünschten Zielzustand sowie eine Barriere, welche die einfache Transformation, d. h. auf routiniertem oder vorgegebenem Weg, vom Anfangs- zum Zielzustand verhindert. Der Zielzustand kann sich unterschiedlich manifestieren, z. B. in Form eines Erkenntnisinteresses oder einer formulierten Forschungsfrage. Erst der Einsatz des Experiments im Rahmen eines Problemlöseprozesses zeigt die Einsatzmöglichkeiten des Experiments wie auch die Bezüge zwischen Problem und Experiment (vgl. Hammann et al., 2007). Aus konstruktivistischer Sicht führt die Auseinandersetzung mit authentischen Problemsituationen dazu, dass Wissensstrukturen vom Individuum generiert werden, indem dem Wissen durch Kontextualisierung Bedeutung verliehen wird (Gerstenmaier & Mandl, 1995). Das problembasierte Lernen (*Problem-Based Learning*) wird dementsprechend vom konstruktivistischen Standpunkt als Konzept effektiven Unterrichts propagiert (vgl. Savery & Duffy, 1998; in Deutschland Ansätze bereits bei Schmidkunz & Lindemann, 1976/1999). Nur über die Aufnahme von eigenen Daten (also im Rahmen einer Problemstellung) kann anschließend auch eine realitätsnahe Auswertung und Diskussion

der Ergebnisse erfolgen, um damit der Forderung nach Authentizität im Naturwissenschaftsunterricht gerecht zu werden (Höttecke, 2013).

Dass Experimentieren im Unterricht zur Lösung authentischer Probleme eingesetzt werden soll, scheint, auch im Lichte der zuvor vorgebrachten Argumente, unstrittig (vgl. Bewersdorff et al., 2020). Empirische Evidenz kann die Forderung nach der Einbettung des Experimentierens in problemorientierte Kontexte untermauern: Die Studie von Allen, Duch & Groh (1996) untersucht unterschiedliche Lehr-Lern-Szenarien in naturwissenschaftlichen Einführungskursen. Lernende, welche durch problembasiertes Lernen naturwissenschaftliche Erkenntnisse gewinnen, werden als „*active, participating and questioning throughout class*“ beschrieben (Allen et al., 1996, S. 50). In einem Review von Studien zur Wirkung von problembasiertem Lernen auf den Lernerfolg kommen Yew & Goh (2016) zu dem Schluss, dass durch problembasiertes Lernen langfristig verfügbares Wissen entsteht und die Anwendung des Wissens besser gelingt. Merritt, Lee, Rillero & Kinach (2017) finden im Rahmen eines Reviews von Studien zum problembasierten Lernen von der Grundschule bis zur 8. Klasse (sog. K–8 Schulen; Alter 3–14 Jahre) in Mathematik und den Naturwissenschaften positive Effekte auf die Lernleistung, das Behaltensvermögen von erworbenem Wissen und konzeptuellem Verständnis sowie die Werthaltung.

P5: Prozessreflektierend experimentieren

Ein prozessreflektierendes Experimentieren soll dazu beitragen, dass die Lernenden nicht mechanistisch die einzelnen Prozessschritte durchlaufen, sondern diese auch aufeinander beziehen und hinterfragen (Warum benötige ich für meine Planung eine Hypothese?; Wie hängen Erkenntnis und Hypothese miteinander zusammen?), um damit das Verständnis der Zusammenhänge der Teilprozesse anzuleiten. Eine Reflexion des *Lernprozesses*, welcher im Rahmen des Experimentierens parallel zum *Erkenntnisprozess* verläuft, wie auch das Besprechen und Diskutieren alternativer Erkenntniswege scheinen allein schon aus konstruktivistischer Sicht sinnvoll (Savery & Duffy, 1998) können aber schnell zur kognitiven Überforderung der Lernenden führen. Dennoch fordern Lehrende selten die Lernenden dazu auf, ihre Beobachtungen zu reflektieren (Abrahams & Millar, 2008), im Allgemeinen wird nur wenig Unterrichtszeit auf die Diskussion und Reflexion der Erkenntnisse verwendet (Akuma & Callaghan, 2018; Minner et al., 2010, Tesch & Duit, 2004). Eine Fallstudie von Donnelly, O’Reilly & McGarr (2013) kommt zu ähnlichen Erkenntnissen: So sahen drei von den vier befragten Lehrkräften das Analysieren, Besprechen und Reflektieren von gewonnenen Erkenntnissen nicht als integralen Teil des Experimentierens, son-

dern nur als ‚Add-On' zur händischen Durchführung des Experiments an. Ein Verständnis des Experimentierens wird jedoch effektiv durch explizit prozessreflektierende Ansätze gefördert, bei welchen sich die Lehrkräfte ausreichend Zeit nehmen, mit den Schülern die Zusammenhänge der einzelnen Teilprozesse zu beleuchten (Clough, 2006; Flick, 2006; Lederman, Lederman & Antink, 2013).

P6: Grenzen von Experimenten ansprechen

Über die Nature of Science, im Besonderen über Grenzen und Unsicherheiten gewonnener Erkenntnisse, gibt es eine Vielzahl von Mythen (McComas, 2002, 2020), welche sich teils sehr hartnäckig bei den Lernenden – und auch Lehrenden – halten (Clough, 2006). Zu diesen Mythen (bzw. Fehlkonzepten) zählen unter anderem, dass

- eine Sammlung von gewissenhaft erhobenen Daten zu sicherem Wissen führt,
- Naturwissenschaften und deren Methoden absolute Beweise liefern,
- die Anerkennung neuer naturwissenschaftlicher Erkenntnisse einfach und unproblematisch erfolgt,
- Gesetze und Prinzipien der Naturwissenschaften absolut und unumstößlich sind.

Zu Ansichten welche in der Wissenschaft als Konsens gelten können kann in dem Kontext wissenschaftlicher Grenzen folgende Charakteristika gezählt werden (Priemer, 2006 nach McComas, Clough & Ylmazroa, 2002):

- Wissen in den Naturwissenschaften ist, obwohl es zuverlässig ist, nicht unveränderlich.
- Neue Erkenntnisse müssen klar und offen dargestellt werden.

Experimente sind, entgegen der Auffassung vieler Lernender, nicht in der Lage, unabhängig von der Anzahl der Wiederholungen, eine *Theorie* zu ‚beweisen'. Die Übereinstimmung der Ergebnisse von Experimenten mit den theoretischen Vorhersagen dient lediglich als eine Bestärkung dieser Theorie (Carrier, 2001). Oft variieren Lernende zu viele Variablen zugleich und können dann keine wissenschaftlich valide Schlussfolgerung ziehen (Glaser, Schauble, Raghavan & Zeitz, 1992): Im Besonderen bei der Variation mehrerer Teilaspekte (Variablen) des Experiments ist kein *Modus tollens* (Wenn gilt „aus Theorie folgt Hypothese" und „Hypothese ist falsch", dann gilt auch „Theorie ist falsch (Falsifikation)") mehr möglich (Abd-El-Khalick, 2006b).

Im Unterrichtsalltag können diese Fehlkonzepte dazu führen, dass Lernende Erkenntnisse aus einmaliger Durchführung des Experiments bezüglich ihrer Aussagekraft überschätzen. Diese Präkonzepte müssen mit den Lernenden im Rahmen einer Diskussion sowohl über die Grenzen und Unsicherheiten von durchgeführten Experimen-

ten als auch der Naturwissenschaft allgemein angesprochen werden. Über den planvollen Umgang mit unerwarteten Daten (Chinn & Brewer, 1993) oder die Besprechung von Untersuchungen, welche nicht erwartungskonform verliefen (Nott & Wellington, 1996), kann eine Diskussion angestoßen werden. Als expliziten Diskussionseinstieg mit den Lernenden bieten sich auch direkte Vergleiche der unterrichtlichen Praxis mit der Wissenschaft an (vgl. zur Modellbildung: Tries & Upmeier zu Belzen, 2009).

Eine Besprechung der Grenzen eines Experiments und der Vorläufigkeit des gewonnenen Wissens ist im Lichte der identifizierten Fehlkonzepte für ein verständiges Experimentieren notwendig. Dies belegt auch eine Studie von Müller und Reiners (2020), welche auch noch bei Studierenden Schwierigkeiten bezüglich der Begründung der Vorläufigkeit von Erkenntnissen identifizieren konnten.

P7: Wissenschaftliche Strenge walten lassen

Ein zentrales Merkmal des Experiments ist die wissenschaftliche Strenge (z. B. Faraday, 1828). Nur durch wissenschaftliche Strenge können zentrale Charakteristika der Epistemologie – Nachvollziehbarkeit und Reproduzierbarkeit (McComas et al., 2002) – erreicht werden. Roberts und Johnson (2015) identifizierten ein Verständnis für evidenzbasiertes Arbeiten (*Concepts of Evidence*) als notwendige Grundlage für die Aufnahme und Auswertung von gewonnenen Daten. Zur naturwissenschaftlichen Datengewinnung und -erfassung wurden von der OECD (2017) im Rahmen des PISA 2015 Science Frameworks unter prozeduralem Wisse folgende Teilaspekte aufgeschlüsselt:

- das Konzept von Variablen (abhängige Variablen, unabhängige Variablen und Kontrollvariablen)
- Konzepte der Messwerterfassung (qualitativ, quantitativ, Umgang mit Skalen)
- Wege, um Messunsicherheit zu minimieren (Messwiederholung, Mittelwertbildung)
- Mechanismen, um die Replizierbarkeit und Genauigkeit zu gewährleisten
- die Variablenkontrollstrategie
- angemessene Abstraktion und Darstellung von Daten
- zielführende Wahl des Untersuchungsdesigns

Der wissenschaftlichen Strenge wird von den Lernenden oft, zugunsten eines Modus des unsystematischen Explorierens, nicht die notwendige Beachtung geschenkt (Meier & Mayer, 2012; Wahser & Sumfleth, 2008). Insgesamt haben die Lernenden Schwierigkeiten bei der adäquaten Berücksichtigung von Störvariablen, Messdauer sowie Messwiederholungen (Arnold, Kremer & Mayer, 2012). Diese Defizite beim Experimentieren fügen sich in das Bild der zuvor bereits dargestellten Befunde: Der Erkenntnismethode selbst wird von Lehrkräften oft nur wenig Beachtung geschenkt,

ein Besprechen der Charakteristika des Experimentierens findet kaum statt. Somit haben die Lernenden auch wenig Gelegenheit, diese Methode der Erkenntnisgewinnung zu erlernen – und fallen dann zurück in ein ineffektives, unsystematisches Explorieren. Es scheint zur Förderung wissenschaftlicher Strenge beim Experimentieren notwendig, die fachspezifischen Regeln und Prozeduren sowohl implizit beim Experimentieren als auch explizit zu thematisieren (Schwichow et al., 2016; Vorholzer, 2017).

P8: Selbstständigkeit beim Experimentieren

Im reformpädagogischen Ansatz von Bruner (1961) wird die Selbstständigkeit der Lernenden beim Erkenntnisgewinn mit erhöhter geistiger Leistungsfähigkeit (*Intellectual Potency*), der Verlagerung von extrinsischer hin zu intrinsischer Motivation, dem Erlernen der Heuristiken des Erkenntnisgewinns sowie einer Unterstützung für die Verarbeitung des Gelernten (*Aid to Memory Processing*) in Verbindung gebracht. Den Ansatz der selbstständigen Entdeckung ergänzt Bruner jedoch mit den Worten „*Discovery […] favors the well prepared mind*“ (Bruner, 1961, S. 1) und kritisiert damit ein planloses Explorieren. Aebli (1993) schreibt, „nicht Stoff, sondern Tätigkeit ist attraktiv“ (Aebli, 1993, S. 20), Lernen ist dabei nur ein Nebenprodukt der Tätigkeit. Aebli stellt jedoch zwei Bedingungen für das Gelingen der Lernendenselbstständigkeit: Die Aufgabe muss angemessen sein und die Tätigkeit muss gelingen. Auch konstruktivistische Lerntheorien fordern eine aktive, lernendenselbstständige Problemlösung (Gerstenmaier & Mandl, 1995; Savery & Duffy, 1998; Widodo & Duit, 2004), differenzieren aber zunehmend aus: So wird anstatt eines rein ‚aktiven Tuns‘ eine ‚aktive (kognitive) Informationsverarbeitung‘ oder im besten Falle eine ‚aktive, *fokussierte* Informationsverarbeitung‘ als erklärungsmächtigste Perspektive für effektiven Wissenserwerb angenommen (Renkl, 2015). Empirische Befunde scheinen diese Perspektive auch für den Erwerb von experimenteller Kompetenz zu stützen. In ihrer Synthese von Forschungsergebnissen der Jahre 1984 bis 2002 zu *Inquiry-based Science Instruction* kommen Minner et al. (2010) zu dem Schluss, dass allein Aktivität der Lernenden in Form von „*Hands-On*“-Aktivitäten wenig effektiv ist. Dennoch konnte anhand der 138 herangezogenen Studien ein Nachweis für den positiven Einfluss von *Inquiry*-basierten Unterrichtskonzepten auf das konzeptuelle Verständnis sowie das Inhaltswissen der Lernenden erbracht werden – besonders in Formaten, welche das aktive Denken (*Active Thinking*) der Lernenden betonen (Minner et al., 2010). Zu ähnlichen Erkenntnissen kommt eine Meta-Analyse von 164 Studien zum Discovery Learning (Alfieri, Brooks, Aldrich & Tenenbaum, 2011). Im Rahmen von Unterrichtskonzepten, welche eine Lernendenselbstständigkeit ohne eine Art von Lernunterstützung propagieren, lernten Schülerinnen und Schüler mehr als in traditionellen, instruktionsbasierten Unterrichtskonzepten. Die Ergebnisse deuten in weiterer Analyse aber darauf hin, dass eine Lernendenselbstständigkeit ohne eine Art von

Lernunterstützung zu niedrigeren Lernzuwächsen im Vergleich zu einer in Teilen geführten Selbstständigkeit führt, da diese zu einer kognitiven Überforderung der Lernenden führen kann (vgl. Kirschner et al., 2006; Sweller et al., 2007). In diese Befunde lassen sich die Ergebnisse der Analysen von PISA-Daten aus dem Jahr 2006 (Jiang & McComas, 2015) und 2015 (Aditomo & Klieme, 2020) sowie einer weiteren Meta-Analyse (Lazonder & Harmsen, 2016) einreihen. Jiang und McComas (2015) fanden in ihren Analysen der PISA-Daten 2006 heraus, dass die höchsten Effekte bezüglich eines Lernzuwachses bei einer moderaten Schülerselbstständigkeit im Erkenntnisprozess erreicht wurden. Bei dieser moderaten Schülerselbstständigkeit werden Fragen und Untersuchungsdesign von der Lehrkraft vorgegeben, die Lernenden führen die Untersuchung selbstständig durch, werten die aufgenommenen Daten aus und ziehen Schlussfolgerungen. Aditomo und Klieme (2020) formulieren aufgrund ihrer Ergebnisse ebenfalls die Forderung nach angemessenen Unterstützungsmaßnahmen und moderater Lernendenselbstständigkeit. Lazonder und Harmsen (2016) finden wie Alfieri et al. (2011) Hinweise, dass auch *Inquiry*-basierte Unterrichtskonzepte darstellenden Unterrichtskonzepten bezüglich des Lernerfolgs überlegen scheinen, doch dass dieser Lernerfolg wiederum stark von angemessenen Unterstützungsmaßnahmen für die Lernenden abhängt. Führen die Unterstützungsmaßnahmen jedoch zu einer starken Lehrerzentrierung und damit einer systematischen Führung der Lernenden durch den Erkenntnisprozess, limitiert dies die Entwicklung kritischen und wissenschaftlichen Denkens (Sadeh & Zion, 2012). Die Lehrkraft ist also gefordert, zwischen einer engen Führung, welche die kognitive Last reduziert, und einem vertieften geistigen Durchdringen des Experiments abzuwägen (Schmidt, Stiller & Wilde, 2019).

Eine *umfassende* Selbstständigkeit beim Experimentieren ist also vornehmlich als Zielvorstellung zu sehen, da zu stark geöffnete Experimentiersettings oft wenig effektiv sind. Darum wird im Rahmen dieses Unterrichtsprinzips eine *weitgehende* Selbstständigkeit beim Experimentieren eingefordert. Hier ist die Lehrkraft aufgefordert, unter abwägendem Blick auf die Lernenden, den Inhalt und sich selbst (didaktisches Dreieck: Glöckel, 2003) den Grad der Selbstständigkeit und angemessene Unterstützungsmaßnahmen für die Lernenden festzulegen.

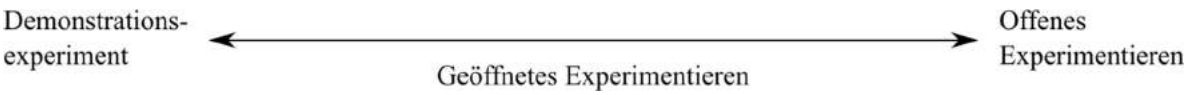


Abbildung 2: Kontinuum der Selbstständigkeit beim Experimentieren (adaptiert nach Furtak (2006))

Lehrkräfte haben jedoch oft Schwierigkeiten, zu entscheiden, wann und in welchem Maß Schülerselbstständigkeit im Erkenntnisprozess befördert und wann diese zurückgenommen werden soll (Crawford, 2007; Furtak, 2006). Um diesen Schwierigkeiten zu begegnen wurden Planungsraster für eine schrittweise Öffnung entwickelt. Diese sahen etwa eine schrittweise Öffnung in zwei (Banchi & Bell, 2008; Bell et al., 2005)

oder drei (Mayer & Ziemek, 2006) Öffnungsgraden einzelner Teilprozesse beginnend jeweils bei der Schlussfolgerung bis hin zur Fragestellung vor. Auf dieser Basis wurden die Öffnungsgrade in einem weiteren Modell ausdifferenziert (Baur & Emden, 2020; Baur et al., 2020; vgl. Tabelle 1) und um die Möglichkeit einer nichtlinearen Öffnung erweitert: Die entwickelte Matrix erzwingt keine Formalismen sondern ermöglicht explizit für einzelne Teilprozesse unterschiedliche Öffnungsgrade. Auf dem Weg zu einer schrittweisen Öffnung des gesamten Experimentierprozesses können Öffnungen, wo nötig, auch zurückgenommen werden.

Tabelle 1: Zweidimensionale Matrix zur Planung und Klassifizierung eines Experiments (Baur, Ehrenfeld, Emden, Hummel & Krieg, 2017)

	Öffnungsgrad			
	geschlossen	**leicht geöffnet**	**geöffnet**	**offen**
Fragestellung	Lernende beschäftigen sich mit einer vorgegebenen Fragestellung	Lernende wählen aus verschiedenen vorgegebenen Fragestellungen aus	Lernende erstellen mit Hilfestellung (medial, verbal) eine Fragestellung	Lernende entwickeln eine Fragestellung
Hypothese	Lernende beschäftigen sich mit einer vorgegebenen Hypothese	Lernende wählen aus verschiedenen vorgegebenen Hypothesen aus	Lernende stellen mit Hilfestellung (medial, verbal) eine eigene Hypothese auf	Lernende stellen eine Hypothese auf
Planung – Experiment	Lernende arbeiten nach einer vorgegebenen Anleitung	Lernende wählen ihre Vorgehensweise aus verschiedenen vorgegebenen Möglichkeiten aus	Lernende planen mit Hilfestellung (medial, verbal) ihr Vorgehen	Schüler planen ihr Vorgehen
Auswertung	Lernende werten die Daten nach Anleitung aus	Bei manchen Experimenten möglich: Lernende wählen aus verschiedenen vorgegebenen Darstellungsmöglichkeiten (Tabelle, Balken, Säulen, Liniendiagramm) aus	Lernende werten mit Hilfestellung (medial, verbal) die ermittelten Daten aus	Lernende suchen sich selbstständig ein Verfahren zur Datenauswertung aus
Schlussfolgerung	Lernende setzen sich mit vorgegebener Schlussfolgerung auseinander	Lernende wählen aus verschiedenen vorgegebenen Schlussfolgerungen aus	Lernende stellen mit Hilfestellung (medial, verbal) eine Schlussfolgerung auf	Lernende ziehen Schlussfolgerung

Eine reflektierte Auswahl des jeweiligen Öffnungsgrads über die verschiedenen Teilprozesse unter dem Credo „So offen wie möglich, so geschlossen wie nötig“ kann als ein Merkmal des Prinzips der Selbstständigkeit beim Experimentieren verstanden werden.

Zusammenfassend kann Unterrichtshandeln zum Experimentieren als effektiv gelten, wenn sowohl die Struktur als auch die innere Logik sowie Grenzen des Experiments vermittelt werden. Ultimatives Ziel ist dabei ein zunehmend selbstständiges und verständiges Experimentieren unter der Wahrung wissenschaftlicher Strenge.

In diesem Kapitel wurden Unterrichtsprinzipien zum effektiven Unterrichtshandeln zum Experimentieren abgeleitet und empirisch fundiert. Die fundierten Unterrichtsprinzipien können bei der Planung und Durchführung von Unterricht sowie in der Fachdidaktik (Forschungs- und Konzeptionsarbeit) Berücksichtigung finden (vgl. Baur et al., 2019). In dieser Forschungsarbeit dienen die Unterrichtsprinzipien zur Identifikation und Validierung geeigneter Kodiermanuale.

2.4 Wirkungen von Lehrkräftefortbildungen auf das Unterrichtshandeln

In diesem Kapitel werden Wirkungen von Lehrkräftefortbildungen auf das Unterrichtshandeln diskutiert. Dazu findet zuerst eine Einführung in Transferbegriffe, -theorien und -formen statt. Anschließend folgt eine Zusammenschau von Studien zu Lehrkräftefortbildungen zum Experimentieren mit dem Blick auf die Fortbildungscharakteristika, Erhebungsmethoden und Ergebnissen.

Lehrkräftefortbildungen sollen das Wissen von Lehrkräften erweitern und ihr Unterrichtshandeln verbessern (Desimone, Porter, Garet, Yoon & Birman, 2002; Garet, Porter, Desimone, Birman & Suk Yoon, 2001). Die Forschungslage weist jedoch darauf hin, dass der Transfer von Fortbildungsinhalten in die Unterrichtspraxis zumindest eine komplexe Angelegenheit ist, so der Transfer denn überhaupt gelingt (McDonald, 2014). So konstatierte Gris (2008) beruflichen Fortbildungen im Allgemeinen eine verschwindend geringe Effektivität. Wenn die Kritik auch zu pauschal gewesen sein mag, wurde doch die unklare Befundlage bezüglich des tatsächlichen Transfers deutlich (Hense & Mandl, 2011).

In diesem Kapitel wird zu Beginn der Begriff *Transfer* geklärt. Anschließend werden Transfertheorien und -modelle vorgestellt und mit empirischen Befunden der Transferforschung in Bezug gesetzt. Studien zu den Wirkungen von Lehrkräftefortbildungen auf das Unterrichtshandeln – speziell auch im Bereich naturwissenschaftlicher Erkenntnisgewinnung – werden beleuchtet. Implikationen für das Untersuchungsdesign der Studie werden abgeleitet.

2.4.1 Transferbegriff, -formen und -theorien

Der Begriff des Transfers wird je nach disziplinärem und kontextuellem Zusammenhang unterschiedlich gebraucht (Gräsel, 2010). Prenzel (2010) spricht von Transfer, wenn in einer Situation (A, Lernkontext) eine Problemlösung entwickelt oder etwas gelernt wird und das Gelernte bzw. die Problemlösung in einer anderen Situation (B, Anwendungskontext) angewendet wird.

Hense & Mandl (2011) unterscheiden drei verschiedene Bedeutungs- und Verwendungsebenen des Transferbegriffs:

1. Mikro-Ebene: Hier wird der Transfer im Rahmen der kognitiven Lern- und Wissenspsychologie untersucht. Der Transfer auf Mikro-Ebene findet im Rahmen ähnlich strukturierter Aufgaben statt. Mit dem Fokus der Lern- und Wissenspsychologie werden als Einflussfaktoren auf den Transfer – neben der Transferaufgabe selbst – vornehmlich kognitive Strukturen und Prozesse betrachtet.
2. Meso-Ebene: Auf dieser Ebene wird der Transfer von Gelerntem aus dem Lernkontext in eine *reale* Anwendung beschrieben. Transfergegenstand sind nun im Vergleich zur Mikro-Ebene breitere Kompetenzen und Wissensinhalte. Als Einflussfaktoren des Transfers stehen auf dieser Ebene die Realbedingungen im Fokus.
3. Makro-Ebene: Diese Ebene untersucht die Gelingensbedingungen für einen erfolgreichen Transfer von Erkenntnissen aus Forschungsprogrammen o. Ä. auf weitere Anwendungskontexte. Einflussfaktoren für gelingenden Wissenstransfer sind auf dieser Ebene Merkmale der Innovation selbst sowie Charakteristika der beteiligten Organisationen und Personen (vgl. Prenzel, 2010).

Der Transfer, welcher im Rahmen einer Lehrkräftefortbildung angeregt werden soll, kann folglich nach Hense & Mandl (2011) als Transfer auf der Meso-Ebene verstanden werden: Das Gelernte aus dem Lernkontext, also im Rahmen der Fortbildung, etwa durch Vorlesung, Hospitation oder Arbeitsaufgaben, angeeignetes Wissen und Kompetenzen, soll unter Realbedingungen, sprich in der Unterrichtspraxis, angewendet werden.

Der im Rahmen einer Lehrkräftefortbildung angestoßene Transferprozess in die unterrichtspraktische Anwendung bzw. Übertragung lässt sich nach Lipowsky (2014) in drei Ebenen konkretisieren (vgl. Abbildung 3:). Auf Ebene 1 findet eine Veränderung des Wissens und der Überzeugungen sowie der affektiv-motivationalen Voraussetzungen statt. Auf Ebene 2 kommt es zur Änderung des Unterrichtshandelns. Ebene 3 beschreibt die Wirkungen der Fortbildung auf die Leistung und Motivation der Lernenden.

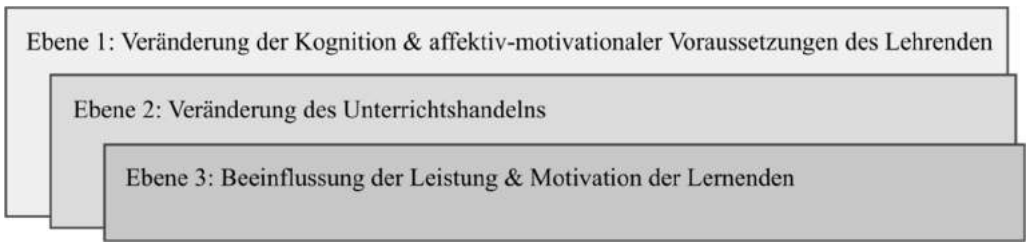


Abbildung 3: Ebenen des Transferprozesses nach Lipowsky (2014), eigene Darstellung

Dieses Ebenen-Modell zu den Wirkungen von Lehrkräftefortbildungen ist selbstverständlich idealisierend. Lipowsky (2014) merkt darüber hinaus an, dass Fortbildungen auf die Lehrkräfte nicht in gleicher Weise, sondern vielmehr differenziell wirken.

Transfer kann sich in unterschiedlichen Wirkungen äußern. Hasselhorn und Hager (2008) benennen eine Reihe von Transferwirkungen. Diese Transferwirkungen stellen sich oft als gegensätzliche Pole dar, sind dabei aber nicht als absolut zu verstehen, vielmehr bilden sie die Pole einer kontinuierlichen Achse. Der *positive Transfer* tritt auf, wenn das im Lernkontext Gelernte in produktivem Sinn im Anwendungskontext eingesetzt wird. *Negativer Transfer* tritt hingegen auf, wenn das im Lernkontext Gelernte zu Irritationen bereits gelernter Routinen und Prozeduren führt und damit die Anwendung schlechter gelingt. Das Gelernte wird offensichtlich – etwa im Sinne einer Übergeneralisierung – falsch angewendet. Mandl, Prenzel und Gräsel (1992) beschreiben in diesem Zusammenhang noch den *Nulltransfer*, es gibt also keinerlei Transfer vom Lernkontext in den Anwendungskontext. Der Positive Transfer kann weiter in *lateralen* (*horizontalen*) *Transfer* und *vertikalen Transfer* unterschieden werden (vgl. Mandl et al., 1992). Beim *lateralen Transfer* werden die im Lernkontext vermittelten Kompetenzen im Anwendungskontext eingesetzt. Beim *vertikalen Transfer* findet auf Basis der in dem Lernkontext vermittelten Kompetenzen im Anwendungskontext ein weiterer Kompetenzerwerb statt. Die im Lernkontext vermittelten Kompetenzen dienen also (auch) dem selbstständigen Erwerb übergeordneter Kompetenzen im Anwendungskontext (Hasselhorn & Hager, 2008).

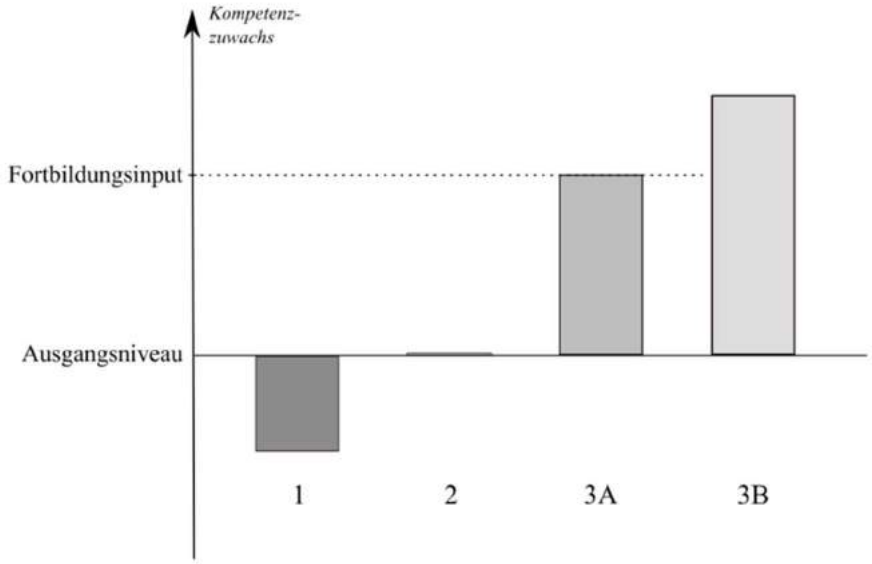


Abbildung 4: Die vier Transferwirkungen 1. Negativer Transfer, 2. Nulltransfer, 3A. und 3B. Positiver Transfer: A. Lateraler Transfer, B. Vertikaler Transfer, eigene Darstellung

Bezüglich der Adaptivität im Anwendungskontext kann zwischen *literalem Transfer* und *figuralem Transfer* unterschieden werden (Hasselhorn & Hager, 2008). *Literaler*

Transfer liegt vor, wenn erworbene Kompetenzen des Lernkontextes in Situationen *gleichen* Typs im Anwendungskontext übertragen werden. *Figuraler Transfer* hingegen bezeichnet die Anwendung von erworbenen Kompetenzen im Rahmen zum Lernkontext *analoger* Problemstellungen.

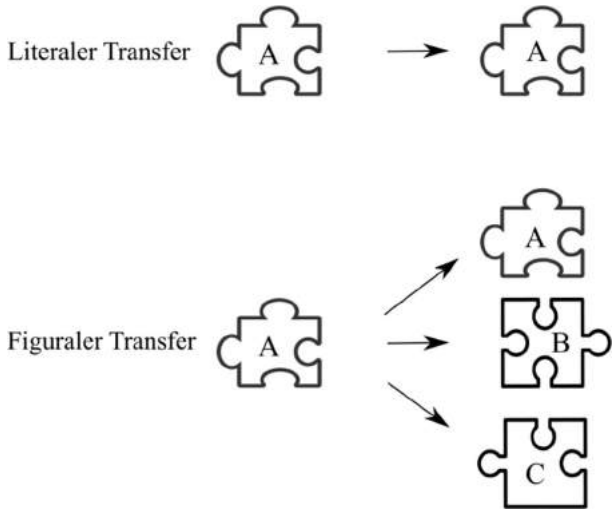


Abbildung 5: Literaler Transfer und Figuraler Transfer, eigene Darstellung

Der *spezifische Transfer* bezeichnet den Transfer eines klar abgegrenzten Lernkontextes auf einen Anwendungskontext. *Unspezifischer Transfer* liegt vor, wenn im Lernkontext allgemeine Problemlösekompetenzen oder Prinzipien vermittelt werden, welche erst in dem konkreten Anwendungskontext inhaltsgebunden eingesetzt werden.

Der *laterale*, der *literale* und der *spezifische Transfer* werden bezüglich ihrer Transferdistanz, also der Ähnlichkeit von Lernkontext und Anwendungskontext, als naher (proximaler) Transfer bezeichnet. Der *vertikale*, der *figurale* sowie der *unspezifische Transfer* werden hingegen dem *weiten* (*distalen*) *Transfer* zugeordnet.

Taxonomien

Haskell (2001) entwickelte eine Taxonomie zur Einteilung des Grades des Transfers in sechs Stufen: 1. unspezifischer Transfer (nonspecific transfer) 2. Anwendungstransfer (application transfer) 3. Kontexttransfer (context transfer) 4. naher Transfer 5. weiter Transfer und 6. kreativer Transfer (creative transfer) (vgl. Calais, 2006), wobei nur die Stufen 3–6 die Definition von Transfer nach (Prenzel, 2010) erfüllen, die Stufen 1 und 2 beschreiben hingegen vornehmlich Lernprozesse.

Bezüglich des Transfers speziell von Fortbildungsinhalten in die Unterrichtspraxis der Lehrkräfte identifizierten Joyce und Showers (2002) anhand von Fallstudien fünf Stufen des Transfers (vgl. Kreis & Unterköfler-Klatzer, 2017):

Stufe 1 – Imitation (imitative use): Die Lehrkräfte geben die in der Fortbildung gelernten Inhalte in gleicher Weise im Unterricht wieder bzw. reproduzieren gelernte Strategien in der Praxis.

Stufe 2 – mechanical use: In der Fortbildung gelernte Inhalte und Strategien werden von den Lehrkräften an das jeweilige Fach oder die Jahrgangsstufe angepasst.

Stufe 3 – routine use: Die Lehrkräfte ordnen Inhalte der Fortbildung Unterrichtssituationen zu und wenden diese an.

Stufe 4 – integrated use: Inhalte und Strategien der Fortbildung werden von den Lehrkräften in einer Vielzahl von unterschiedlichen Unterrichtssituationen angemessen angewendet.

Stufe 5 – executive use: Die Lehrkräfte haben ein Verständnis von den der Fortbildung zugrundeliegenden Theorien. Die Anwendung der gelernten Inhalte und Strategien erfolgt versiert themen-, fächer- und situationsübergreifend.

Bei dieser Taxonomie ist unter Berücksichtigung der Möglichkeit eines negativen Transfers sowie eines Nulltransfers (Hasselhorn & Hager, 2008) kritisch das Fehlen von einer Stufe 0 und einer Stufe -1 anzumerken. Eine mögliche Ergänzung um folgende Stufen wäre:

Stufe 0 – non use: Die in der Fortbildung gelernten Inhalte und Strategien werden von der Lehrkraft im Unterricht nicht angewendet.

Stufe -1 – problematic use: Die Lehrkraft wendet Inhalte und Strategien der Fortbildung an, doch kommt es zu Problemen bezüglich der Passung in konkreten Unterrichtssituationen.

Die in den Stufen 1-5 ausdifferenzierte Taxonomie scheint darüber hinaus bezüglich ihrer Charakteristiken und der Trennschärfe der Stufen untereinander diskutabel. Bezüglich der ‚angemessenen Anwendung' (Stufe 4) und der ‚versierten themen-, fächer- und situationsübergreifenden Anwendung' (Stufe 5) wird diese Unschärfe deutlich: Ist eine angemessene Anwendung nicht stets – wenn es der Unterricht erfordert – auch themen-, fächer- und situationsübergreifend?

Diese Transfertaxonomie von Joyce und Showers (2002) kann daher nur unter Vorbehalten als Rahmen für die Wertung des Transfererfolgs von in der Fortbildung vermittelten Inhalten und Strategien in das Unterrichtshandeln der Lehrkräfte herangezogen werden.

Es gibt noch keine einheitliche Theorie, wie Transfer zustande kommt (Hasselhorn & Hager, 2008), einige Einflussfaktoren konnten jedoch bereits identifiziert werden. Im Folgenden werden Einflussfaktoren benannt, deren Wirkung (positiv wie negativ) bereits als empirisch gesichert gelten kann.

Neben *potenziell* wirksamen Einflussfaktoren auf den Transfer findet sich bei Zehetmeier (2008) eine Kategorisierung der Einflussfaktoren. Die Kategorien sind wie folgt benannt: Lehrkräfte, Fortbildner, Fortbildungsprogramm, innerschulischer Kontext sowie außerschulischer Kontext. Eine abstraktere – und im Folgenden zur Einordnung empirischer Befunde verwendete – Kategorisierung von Einflussfaktoren findet sich bei Ingvarson, Meiers und Beavis (2005) im Rahmen des Frameworks zu den Beziehungen und Einflussfaktoren von Lehrkräftefortbildungen auf Lehrende sowie Lernende (vgl. Abbildung 6). Als Kategorien wurden „Background variables“, „Structural features“ sowie „Opportunity to learn“ benannt. Als mediierender Faktor wird die „Professional community” identifiziert.

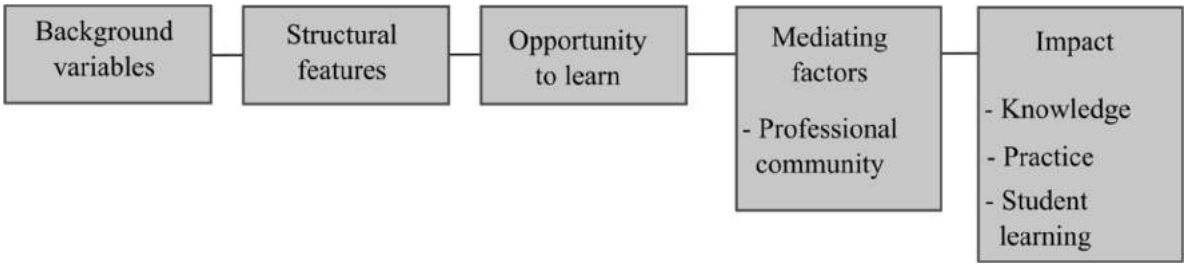


Abbildung 6: Einflussfaktoren von Lehrkräftefortbildungen auf Lehrende sowie Lernende nach Ingvarson et al. (2005)

Im Folgenden werden im Rahmen aller drei Kategorien „Background variables“, „Structural features“ sowie „Opportunity to learn“ positive Einflussfaktoren auf den Transfererfolg herausgestellt.

Ingvarson et al. (2005) berichten in ihrer Analyse von Befragungen von 3250 australischen Lehrkräften in der Kategorie „Background variables“ den Einflussfaktor der schulinternen Unterstützung (school support) als wichtige Grundvoraussetzung für erfolgreichen Transfer. Gräsel (2010) betont ebenfalls die Merkmale der jeweiligen Schule als bedeutungsvollen Einflussfaktor für gelingenden Transfer. Auch Blume, Ford, Baldwin und Huang (2010) berichten im Rahmen ihrer Metaanalyse von 89 Studien zum Transfer von beruflichen Fortbildungsmaßnahmen die Unterstützung

durch Kolleginnen und Kollegen und vor allem der Führungskräfte als wichtigen Einflussfaktor von Transfererfolg. Als weitere signifikante Einflussfaktoren, welche sich der Kategorie „Background variables“ zuordnen lassen, nennen sie individuelle kognitive Fähigkeiten, Gewissenhaftigkeit und Motivation der Teilnehmenden (Blume et al., 2010).

In der Kategorie „*Structural features“* konnten die Kontaktzeiten im Rahmen der Fortbildung und die kollegiale Teilnahme als positive Einflussfaktoren auf den Transfer von Fortbildungsinhalten und -strategien in die Unterrichtspraxis identifiziert werden. Für den Einflussfaktor des zeitlichen Umfangs von Lehrkräftefortbildungen als ‚*Structural feature*‘ konnte ebenfalls ein positiver Zusammenhang zum Transfererfolg nachgewiesen werden (Desimone et al., 2002; Zucker, Shields & Adelmann, 1998). Ingvarson et al. (2005) berichten für den Einflussfaktor des zeitlichen Umfangs von Lehrkräftefortbildungen substanzielle, jedoch vornehmlich indirekte Effekte auf den Transfererfolg.

Bezüglich der Kategorie „*Opportunity to learn“* wird vor allen Dingen „*Active Learnin*g“ der teilnehmenden Lehrkräfte als positiver Einflussfaktor für den Transfererfolg berichtet (Desimone et al., 2002; Garet et al., 2001; Ingvarson et al., 2005). Außerdem berichten sowohl Ingvarson et al. (2005) als auch Desimone et al. (2002) starke Effekte auf den Transfererfolg, wenn die Fortbildung inhaltlich fokussiert ist.

Die den Kategorien „*Structural features*“ sowie „*Opportunity to learn*“ einbeschriebenen Einflussfaktoren haben nach Ingvarson et al. (2005) die stärksten Effekte auf den Transfererfolg. Die „Background variables“ der teilnehmenden Lehrkräfte (demografische Variablen wie Geschlecht und Alter) haben kaum Einfluss auf den Transfererfolg (Blume et al., 2010), vielmehr wirken auf der Ebene der Einzelpersonen kognitiv-motivationale Faktoren auf den Transfererfolg. Die „Background variables“ der teilnehmenden Institution (etwa Unterstützung der Leitungsebene), aber im Besonderen die kollegiale Unterstützung („*professional community*“; Blume et al., 2010; Ingvarson et al., 2005), zeigen hingegen Einfluss auf den Transfererfolg.

Erfassung von Transfer

Um ein umfassendes Bild des Transfers zu erzielen, ist nach Hasselhorn und Hager (2008) eine Überprüfung des Transfers idealerweise auf drei sich gegenseitig ergänzende Arten vorzunehmen:

1. Situationstransfer: Die Rahmenbedingungen des Anwendungskontexts werden im Vergleich zum Lernkontext konstant gehalten. Die Situation des Anwendungskontexts unterscheidet sich jedoch in einem (oder wenigen) Merkmalen

von der Situation des Lernkontexts. Das Situationsmerkmal kann in der Ausprägung der Transferdistanz zwischen scheinbar marginal (Variation des Testleiters) und moderat (Variation der Sozialform) aber auch bedeutend (Transfer des Lernkontexts in eine Alltagssituation) variieren.

2. Zeitlicher Transfer: Der Transfer des Lernkontexts in den Anwendungskontext wird mit zeitlichem Abstand zum Lernkontext gemessen. Als Richtschnur für einen angemessenen zeitlichen Abstand nennen Hasselhorn und Hager (2008) drei Monate.
3. Anforderungstransfer: Im Anwendungskontext werden Anforderungen definiert, welche von den im Lernkontext fokussierten und darauf entwickelten Problemlösestrategien hinreichend verschieden sind. Hiermit soll die Fähigkeit zur Generalisierbarkeit im Anwendungskontext ermittelt werden.

Die praktische Transferforschung im Bildungsbereich, besonders die Untersuchung des Transfers vom Lernkontext „Fortbildung“ in den Anwendungskontext „Unterrichtshandeln“ erweist sich als methodisch anspruchsvoll (Lipowsky, 2010).

Zur Messung von Transferwirkungen vom Lern- in den Anwendungskontext lassen sich im Wesentlichen fünf Ansätze identifizieren (vgl. Hense & Mandl, 2011; Lipowsky, 2010). Der bevorzugte, da unmittelbare Ansatz, ist die objektive Messung des Transfers im Anwendungskontext über eine Operationalisierung der erwarteten Verhaltensänderungen (etwa Änderungen des Unterrichtshandelns). Problematisch bei der objektiven Messung von Transfer ist die steigende Schwierigkeit der Operationalisierung, je komplexer die zu erfassenden Verhaltensweisen (bzw. Änderungen dieser) sind. Probates Mittel im Kontext des Transfers auf den Anwendungskontext „Unterrichtshandeln“ ist bei diesem Ansatz die Unterrichtsvideografie (Lipowsky, 2010; Schmaltz, 2019).

Ein weiterer – und aus forschungsökonomischen Gründen oft angewendeter Ansatz (Ingvarson et al., 2005) – ist die Befragung der teilnehmenden Lehrkräfte zur Einschätzung des Transfererfolgs (Hense & Mandl, 2011). Der Nachteil dieses Ansatzes ist, dass durch die Einschätzung der Lehrkräfte nur ein indirekter Nachweis von Transferwirkungen möglich ist. Gerade im Wissen um eine verzerrte Selbstwahrnehmung sind diese Auskünfte kritisch zu prüfen.

Oft wird die Erweiterung der Lehrerkognition, wozu sowohl Überzeugungen und Werthaltungen (Beliefs) als auch das fachliche, fachdidaktische und pädagogische Wissen gezählt werden können, gemessen (Lipowsky, 2010). Dieser Ansatz wird in zahlreichen Studien (ergänzend) eingesetzt, die Aussage der Ergebnisse bezüglich tatsächlicher Verhaltensänderungen muss jedoch kritisch gesehen werden: Auch wenn

neues Wissen erlernt wurde, ist noch keinesfalls gesichert, dass sich dieses auch in angepassten Verhaltensweisen (etwa dem Unterrichtshandeln) niederschlägt.

Zur Analyse des Transfererfolgs kann außerdem die Einschätzung von dritten, mittelbar betroffenen Personen herangezogen werden (Hense & Mandl, 2011; Lipowsky, 2010). Im allgemeinen Berufsumfeld können für diese Einschätzung Vorgesetzte oder Kollegen, im Schulkontext Schülerinnen und Schüler befragt werden. Auch hier sind jedoch wieder mögliche Verzerrungen über den indirekten Nachweis der Transferwirkungen zu berücksichtigen.

Im Schulkontext bietet sich auch die Erhebung von Leistungsdaten von Lernenden zur Messung von Transferwirkungen an (Lipowsky, 2010). Auch dieser Ansatz erfasst Transferwirkungen vom Lern- in den Anwendungskontext nur indirekt. Die Leistungsdaten werden bei diesem Ansatz rückwirkend bezüglich der Transferwirkungen herangezogen. Auch dieser indirekte Nachweis von Transfer birgt Nachteile der Verzerrung durch unkontrollierte Störvariablen, die teils bei anderen Erhebungsmethoden stärker kontrolliert werden können.

Anhand der Wirkungskette von Lehrerbildung (u. a. bei Galluzzo und Craig (1990), um die Stufe der Lernerfahrung der Lernenden reduziert bei Terhart (2012)) sollen zur Übersicht die fünf Ansätze zur Erfassung von Transferwirkung in Abbildung 7 an ihrem jeweiligen Ausgangspunkt verortet werden.

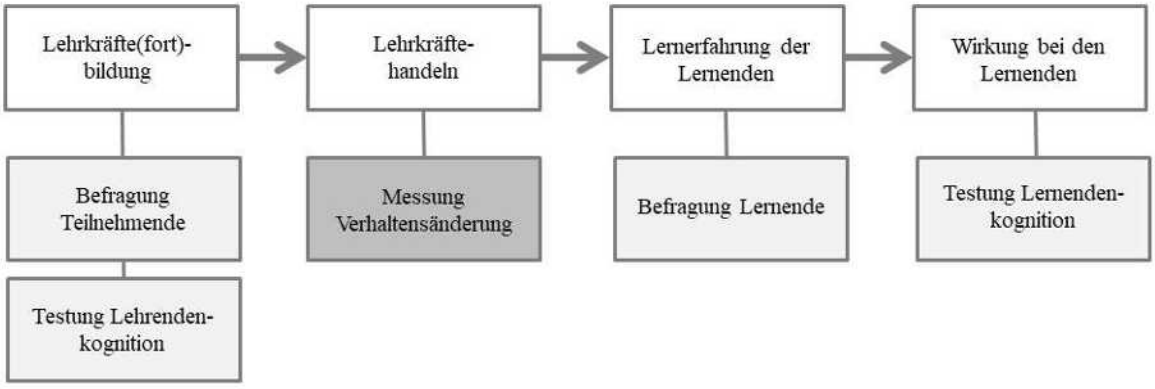


Abbildung 7: Ansätze zur Testung von Transferwirkungen in der Wirkungskette der Lehrerbildung nach Galluzzo und Craig (1990), eigene Darstellung. Die Messung der Verhaltensänderung (grau hervorgehoben) ist der unmittelbare Ansatz zur Messung von Transferwirkungen (Lipowsky, 2010)

2.4.2 Empirische Befunde zur Wirkung von Lehrkräftefortbildungen auf das Unterrichtshandeln – eine Zusammenschau

Im folgenden Abschnitt sollen aktuelle Studien zur Wirkung von Lehrkräftefortbildungen auf das Unterrichtshandeln speziell im Bereich des (*Scientific*) *Inquiry* vorgestellt und ihr Design und ihre Ergebnisse diskutiert und zu Erkenntnissen fachspezifischer Metaanalysen und Reviews in Bezug gesetzt werden. Für die Zusammenschau

wurden die Forschungsdatenbanken ERIC und für den deutschsprachigen Raum FIS Bildung mit den Begriffen "scientific inquiry" UND "professional development" respektive "experimentieren" UND "Fortbildung" abgefragt. Weitere Studien konnten mithilfe des Schneeballverfahrens (Döring & Bortz, 2016) identifiziert werden. Aufgrund dieses Verfahrens ist die Zusammenschau exemplarischen Charakters und erhebt keinen Anspruch auf Vollständigkeit. Aufgrund der reduzierten methodischen Strenge bei der Auswahl der Studien und damit einhergehend des vornehmlich illustrativen Charakters dieser Zusammenschau wird im Folgenden explizit nicht von einem Review gesprochen.

Kriterien für die Auswahl der Studien waren:

- der Fortbildungsinhalt (*Inquiry* im weiten Sinn),
- der Untersuchungsgegenstand (Unterrichtshandeln),
- sowie das Publikationsjahr (nach 2009).

Die Auswahl der Studien anhand des Fortbildungsinhalts (*Inquiry*) sowie ihr Untersuchungsgegenstand (Unterrichtshandeln) wurden aufgrund der Konzeption der eigenen Studie getroffen. Das Publikationsjahr wurde im Hinblick auf das Review von Capps et al. (2012) festgelegt. Die jüngste Studie des Reviews von Capps et al. (2012) wurde im Jahr 2009 publiziert. Dieses Reviews und die darauf abgeleiteten Implikationen, etwa die Forderung nach der Durchführung von Studien, welche die Wirkungen von Fortbildungen auf allen Ebenen bis zu den Lernenden untersuchen, war mit ursächlich für die Durchführung des Forschungsprojekts, in dessen Rahmen diese Arbeit entstanden ist. Darum wurden nur Publikationen herangezogen, welche *nicht* in dem Review von Capps et al. (2012) berücksichtigt wurden.

Da weitere Teilprojekte den Einfluss der Lehrkräftefortbildung auf die Beliefs (Emden, Baur & Bewersdorff, 2020) und das PCK der Lehrkräfte sowie die Schülerkognition untersuchen, werden diese Aspekte bei der Zusammenschau ebenfalls berichtet, soweit sie in den Studien neben dem Unterrichtshandeln ebenfalls untersucht wurden.

In Tabelle 2 sind die elf identifizierten Studien sowie wesentliche Merkmale der untersuchten Fortbildung (inhaltlicher Schwerpunkt, Land, pädagogische Ausrichtung und Aufbau, Dauer sowie Schulstufe) als auch Merkmale der begleitenden Studie (Teilnehmerzahl, Erhebungsmethoden sowie Befunde auf Ebene der Beliefs/PCK, des Unterrichtshandelns und der Kognition der Lernenden) dargestellt.

Tabelle 2: Zusammenschau von Studien zur Wirkung von Lehrkräftefortbildungen zu *Inquiry* (im weiten Sinne) u. a. auf das Unterrichtshandeln. # TN = Anzahl der Teilnehmenden; UH = Unterrichtshandeln; PCK = pedagogical content knowledge; LK = Lehrkraft/Lehrkräfte; LFB = Lehrkräftefortbildung

Studie	Schwerpunkt	Land	Ausrichtung/ Aufbau der LFB	Dauer	Schul-stufe	# TN	Methode	Befunde UH	Befunde PCK	Befunde Beliefs	Befunde Lernende
Antoniou & Kyriakides, 2013	Inquiry in Mathematik	Zypern	Dynamischer integrierter Ansatz (Dynamic Integrated Approach)	1 Schuljahr	Primar	130	classroom observations at the beginning, at the end, one year after the end of the intervention Lernende: Tests	Positive Entwicklung, konstant bei follow-up	-	-	Zeigten Kompetenzzuwachs
Brand & Moore, 2011	Scientific Inquiry	USA	Kokonstruktiv, soziokulturell	3 Jahre, 2 Phasen: workshop, Implementation und Reflexion	Primar	30	interviews and semantic maps	LK adaptierten gelernte Strategien in Richtung einer Lernenden-zentrierung	-	Entwickelten ein neues Verständnis	-
Chin, Lin & Tuan, 2016	Inquiry in Mathematik	Taiwan	Lernpartner in 12 Fortbildungstreffen: 1.Erfahren, 2.Implementieren, 3.Verstehen	1 Schuljahr	Sek I (Kl. 7–9)	4	interviews and concept maps, classroom observations	Alle vier LK entwickelten ihr UH weiter, tiefgehende Änderungen aber selten	Alle vier LK machten Fortschritte	-	-
Grigg, Kelly, Gamoran & Borman, 2013	scientific inquiry	USA	Workshoptreffen (kompakt-immersiv *oder* getrennt nach Themenbereichen)	1-5 Tage	Primar (Kl. 4–5)	96	Classroom observations	Die Ergebnisse zeigen, dass die Implementation auf einzelne Aspekte des scientific inquiry limitiert blieb	-	-	-
Harlow, 2013	Inquiry based science teaching	USA	Inquiry Learning am Beispiel von Magnetismus	6 Wochen	Primar	5	Interviews und Unterrichtsvideografie	3 LK zeigten Transfer, jedoch wurden sehr unterschiedliche, ihren individuellen Bedürfnissen und Vorstellungen entsprechende Aspekte übernommen	-	-	-

Lotter, Rushton & Singer, 2013	Scienfic inquiry	USA	Zwei Wochen Seminar, anschließend weitere Betreuung (Follow-up Workshops)	1 Schuljahr	Sek. I	36	Interviews and Final Written Reflection, Unterrichts-videografie	LK übernehmen unterschiedliche Aspekte, 4 Gruppen: Integrated (12), Emerging (7), Laboratory-based (12), und Activity-focused (5)	-	Beliefs erhöhten sich, doch teils rudimentäres Verständnis von scientific inquiry als ‚hands-on'-Aktivität	-
Lotter et al., 2016	Scienfic inquiry	USA	Zwei Wochen Seminar, anschließend weitere Betreuung (Follow-up Workshops)	1 Schuljahr	Sek. I	58	Fragebogen, Videografie	Signifikante Erhöhung, Wandel von lehrerzentriert zu schülerzentriert	-	Beliefs erhöhten sich signifikant	-
Marshall & Smart, 2013	Scientific Inquiry	USA	Zwei Wochen Seminar, follow up support inkl. Hospitationen	1 Schuljahr	Sek. I	3	classroom observations; field notes; teacher interviews; transcripts of classroom recordings	Alle LK zeigten Transfer der Inhalte in ihr Unterrichthandeln, teils aber nur oberflächliche Umsetzung	Alle LK verbesserten ihr Verständnis	Erhöht oder auf hohem Niveau gehalten	-
McGee & Nutakki, 2017	Scientific Inquiry	USA	Workshops	0,25 Jahre, 2–7 Workshops	Primar und Sek. I	91	Fragebogen Lehrkräfte und SuS bzgl. der berichteten Umsetzung von Inquiry	Significant impact on teaching practice	-	-	-
Rushton, Lotter & Singer, 2017	Scientific Inquiry	USA	Zwei Wochen Seminar, anschließend weitere Betreuung	1 Schuljahr	Sek. I	7	Beliefs: Interviewbogen, Reflexionen, Videografie	Besonders einzelne Aspekte der LFB wurden umgesetzt (Phenomena first, Wissenschaftsorientierung, Reflexion)	-	Entwicklung eines vollumfänglicheren Verständnis von Inquiry	-
Schmitt, 2016	Erkenntnisgewinnung	Deutschland	Vortrag, Planung und Erprobung von Experimenten und Unterrichtsmaterialien	1 Tag (4,5h)	Sek. I	54 (8 Video)	Fragebögen/ Tests Videografie/ Interviews	Lehrer waren nach der Fortbildung in der Lage, die Inhalte im Unterricht umzusetzen	Konnte unmittelbar (post) und langfristig (follow-up) gesteigert werden	Konnte unmittelbar (post) und langfristig (follow-up) gesteigert werden	Schülergruppen waren, in der Lage, Experimente gemäß dem Erkenntnisweg durchzuführen

Im Folgenden sollen nun Aspekte des Fortbildungsdesigns, des Studiendesigns und die jeweiligen Ergebnisse der in Tabelle 2 dargestellten Studien diskutiert werden.

Fortbildungsdesign

Leitlinien für die Beleuchtung der Fortbildungsdesigns sollen analog zu dem Review von Capps et al. (2012) die von Desimone et al. (2002) identifizierten und in der Fortbildungsforschung stark rezipierten Charakteristika effektiver Lehrkräftefortbildungen (*Content Focus*, *Active Learning*, *Coherence*, *Duration*, *Collective Partizipation*) sein (vgl. auch Garet et al., 2001, weitere Ausschärfungen bei Desimone & Garet, 2015). Gräsel, Fußangel & Parchmann (2006) identifizierten als Charakteristika effektiver Lehrkräftefortbildung ebenfalls Langfristigkeit, Kohärenz mit den Gegebenheiten und Bedürfnissen vor Ort sowie die kollegiale, diskursive Teilnahme. Auch Borko (2004) hebt die aktive Teilnahme der Lehrkräfte im Rahmen der Fortbildung als effektiv hervor.

Da bezüglich der Koheränz (*coherence*) und des Inhaltsfokus (*content focus*) der Lehrkräftefortbildung eine objektive, kriteriale Bewertung der zugrunde liegenden Veröffentlichungen schwer darstellbar ist, werden diese Charakteristika in der Diskussion nicht berücksichtigt: Nur manche der dargestellten Studien berichten explizit, dass Koheränz und/oder ein inhaltlicher Fokus bei der Konzeption berücksichtigt wurde, bei den anderen Studien kann die Berücksichtigung dieser beiden Charakteristika nicht per se ausgeschlossen werden. In welchem Umfang diese beiden Charakteristika, so sie berichtet werden, umgesetzt wurden, bleibt ebenfalls oft unklar.

Die Gelegenheit für eigenaktives Lernen (*Active Learning*) manifestiert sich etwa über die Implementation und Reflexion von Unterrichtskonzepten, welche im Rahmen der Fortbildung vermittelt wurden. Antoniou und Kyriakides (2013) ermöglichen das eigenaktive Lernen im Rahmen von Reflexion und Entwicklung an regelmäßigen Arbeitstreffen sowie der wiederholten Implementation zwischen den Arbeitstreffen. Brand und Moore (2011), Chin et al. (2016) und Grigg et al. (2013) gaben den Lehrkräften im Rahmen der Fortbildung ebenfalls durch Implementation und Reflexion die Möglichkeit des eigenaktiven Lernens. Auch bei Lotter et al. (2013), Lotter et al. (2016), Marshall und Smart (2013), McGee & Nutakki (2017) und Rushton et al. (2017) finden sich durch Phasen der Implementation und Reflexion Gelegenheiten für ein eigenaktives Lernen. In den Studien von Harlow (2013) und Schmitt (2016) ist das eigenaktive Lernen hingegen nicht explizit im Fortbildungskonzept angelegt.

Die kollegiale Teilnahme an der Fortbildung, und damit die Möglichkeit, von einem schulinternen Diskurs zu profitieren, findet sich nur in der Studie von Brand

und Moore, (2011). McGee und Nutakki (2017) betonen zwar in ihrer theoretischen Fundierung des Fortbildungsdesigns die Relevanz einer kollegialen Teilnahme. Ob diese auch in der Studie eingefordert wurde, wird jedoch nicht klar. Eine erklärungsmächtige Ursache für das fast vollständige Ausbleiben kollegialer Teilnahme in den von den Studien untersuchten Fortbildungen mag die in den jeweiligen Ländern institutionalisierte Fortbildungspraxis sein: Meist werden Fortbildungen an einer zentralen Stelle angeboten, zu welcher Lehrkräfte aus verschiedenen Schulen (USA: meist einem Schulbezirk) zusammenkommen.

Bei der Dauer der Fortbildungen muss man zum einen zwischen *Langfristigkeit*, also der Zeitspanne von Beginn bis zum Ende der Fortbildungsmaßnahme, sowie zum anderen der *Präsenzzeit*, also der Summe der Präsenzstunden je Lehrkraft unterscheiden. In den meisten Studien dieser Zusammenschau wird diese Präsenzzeit jedoch nicht explizit berichtet, sodass dieses Maß bezüglich der Dauer einer Fortbildung nicht herangezogen kann. Alternativ wird sich im Folgenden auf die Langfristigkeit der jeweiligen Fortbildungsmaßnahmen bezogen. Desimone und Garet (2015) definieren als Mindestanforderung zur Erfüllung des Charakteristikums der Dauer (duration) bezüglich der *Langfristigkeit* eine Spanne von mindestens einem Schuljahr (als weiteres Kriterium empfehlen sie eine Präsenzzeit von mindestens 20 Stunden, die aber aufgrund der dazu lückenhaften Angaben in den untersuchten Studien nicht als Maß angelegt werden kann). Bezüglich der Langfristigkeit zeichnet die Zusammenschau der Fortbildungskonzepte ein heterogenes Bild: Die kürzeste Fortbildung in der Zusammenschau dauert einen Tag (Schmitt, 2016), die längste Fortbildung erstreckt sich über drei Schuljahre (Brand & Moore, 2011). Sieben Studien der Zusammenschau erfüllen das Charakteristikum der Langfristigkeit nach Desimone und Garet (2015). Die jeweils genaue Fortbildungsdauer ist in Tabelle 2: Zusammenschau dargestellt.

Mehr als die Hälfte der Studien (Antoniou & Kyriakides, 2013; Brand & Moore, 2011; Grigg et al., 2013; Lotter et al., 2016; Marshall & Smart, 2013; McGee & Nutakki, 2017) beruft sich in der Konzeption (teilweise) auf die Arbeiten von Garet und Desimone.

Tabelle 3: Die Charakteristika ‚eigenaktives Lernen', ‚kollegiale Teilnahme' und ‚Dauer' nach Desimone und Garet (2015)

Studie	Eigenaktives Lernen	Kollegiale Teilnahme	Dauer (Langfristigkeit)
Antoniou & Kyriakides, 2013	X	-	X
Brand & Moore, 2011	X	X	X
Chin et al., 2016	X	-	X
Grigg et al., 2013	X	-	-
Harlow, 2013	-	-	-
Lotter et al., 2013	X	-	X
Lotter et al., 2016	X	-	X
Marshall & Smart, 2013	X	-	X
McGee & Nutakki, 2017	X	(X)	-
Rushton et al., 2017	X	-	X
Schmitt, 2016	-	-	-

Nur die Studie von Brand und Moore (2011) erfüllt alle drei hier näher untersuchten Charakteristika, die Studien von Harlow (2013) und Schmitt (2016) hingegen erfüllen keines dieser Charakteristika.

Studiendesign und -ergebnisse

Zur Messung von Änderungen des Unterrichtshandelns wurde von den Studien eine Vielzahl von Untersuchungsansätzen gewählt und teilweise kombiniert. Diese Untersuchungsmethoden umfassen Unterrichtsbeobachtungen, Interviews, Concept Maps, Unterrichtsvideografie, Fragebögen und Beobachtungsnotizen (vgl. Tabelle 2). Bei Studien, welche sich vornehmlich auf Unterrichtsbeobachtungen oder Selbstauskünfte der Lehrkräfte stützen, muss bezüglich der berichteten Effektivität kritisch berücksichtigt werden, dass sich die Sichtweisen unterschiedlicher Personengruppen auf Unterricht teils signifikant unterscheiden (Clausen, 2002).

Bemerkenswert ist bei der Zusammenschau (vgl. Tabelle 2), dass alle Studien Transferwirkungen der Lehrkräftefortbildung auf das Unterrichtshandeln nachweisen konnten. Dies gilt im Besonderen für Schmitt (2016), die zeigen konnte, dass sich Transferwirkungen bereits nach einer eintägigen Fortbildung nachweisen lassen (sogar weiter bis auf die Ebene der Lernendenkognition). Beim Blick auf das Unterrichtshandeln zeigt sich bei genauerer Betrachtung, dass die über die Fortbildung intendierten Änderungen des Unterrichtshandelns selten vollumfänglich erreicht werden (Antoniou und Kyriakides (2013), Lotter et al., (2016)). McGee und Nutakki (2017) und Schmitt (2016) berichten einen ganzheitlichen Transfer der Lerninhalte auf das Unterrichtshandeln. Antoniou und Kyriakides (2013) sowie McGee und Nutakki (2017) stützten sich bei ihren Erhebungen allein

auf Unterrichtsbeobachtungen beziehungsweise Fragebögen. Die Studien, welche eine teilweise Umsetzung der Fortbildungsinhalte in der Unterrichtspraxis berichten, identifizierten teils Aspekte des jeweiligen Transfers. Es zeigt sich, dass der Transfer der Fortbildungsinhalte auf das Unterrichtshandeln teils

- abhängig von der einzelnen Lehrkraft ist (Harlow, 2013; Lotter et al., 2013),
- nur auf einzelne Aspekte beschränkt ist (Grigg et al., 2013; Rushton et al., 2017),
- nur oberflächlich umgesetzt wird (Chin et al., 2016; Marshall & Smart, 2013).

Änderungen der Beliefs, des PCK und der Schülerkognition werden, soweit ebenfalls erhoben, durchweg als (signifikante) Steigerungen beschrieben.

Diese aus der Zusammenschau extrahierten Erkenntnisse lassen sich in Teilen mit den Erkenntnissen aus der Arbeit von Gräsel et al. (2006) und den Befunden von Reusser & Tremp (2008) zur Deckung bringen: Wird eine Fortbildung allein punktuell angeboten und es werden keine Transferstrategien vermittelt (Wie kann ich das Gelernte im eigenen Unterricht anwenden?), sind Effekte im Unterrichtshandeln wenig wahrscheinlich. Ein persönlicher Gewinn für die Lehrkraft ist schnell wieder abgefallen. Daraus schließen Reusser und Tremp (2008), das angebotene Lehrkräftefortbildungen langfristig angelegt werden sollen und über die rein akademische Einführung hinaus den Transfer in die Praxis begleiten müssen. Es ist außerdem davon auszugehen, dass kollegiale Teilnahme, etwa durch Kokonstruktion (Grosche, Fußangel & Gräsel, 2020), die Umsetzung (Transfer) von neuen Inhalten fördern (Gräsel, 2008).

2.5 Abgeleitete Forschungsfragen und Hypothesen

Ziel dieser Forschungsarbeit ist es, ein Instrument zur Erfassung von Änderungen des Unterrichtshandelns, bezogen auf das Experimentieren (speziell: das Entdeckende Experimentieren; Emden & Baur, 2017), zu entwickeln. Dieses Instrument ist notwendig, um Änderungen des Unterrichtshandelns im Rahmen von Fortbildungen zum Experimentieren zu erfassen, um damit den Forderungen von Capps & Crawford (2013) nach einer umfassenden Wirkanalyse von Fortbildungen zu *Scientific Inquiry* zu begegnen (Wirkungen auf Beliefs und PCK der Lehrkräfte sowie die Schülerkognition werden in assoziierten Teilprojekten untersucht). Durch zwei Teilstichproben soll die Wirkung des Grades der Kooperation der Lehrkräfte – ein mutmaßlich relevanter Bedingungsfaktor für den Transfererfolg in die Unterrichtspraxis (siehe Kapitel 2.4.2) – auf den Transfer in das Unterrichtshandeln erfasst werden.

Daraus wird folgende Forschungsfrage abgeleitet:

Forschungsfrage:

Wie wirken sich die beiden Fortbildungsformate (kokonstruktiv vs. individualisiert-konstruktiv) auf das Unterrichtshandeln der teilnehmenden Lehrkräfte aus?

Hypothese H_1:

> Die Lehrkräftefortbildung führt zu einer signifikanten Veränderung des Unterrichtshandelns.

Für die Hypothese H_1 gilt die Alternativhypothese:

> H_0: Es gibt durch die Lehrkräftefortbildungen keine signifikanten Veränderungen im Unterrichtshandeln der teilnehmenden Lehrkräfte.

Hypothese H_2:

> Das Fortbildungsformat mit kokonstruktiven Elementen zeigt im Vergleich zum Fortbildungsformat mit vornehmlich individuellen Elementen größere Veränderungen bezüglich des Unterrichtshandelns.

Für die Hypothese H_2 gilt die Alternativhypothese:

> H_0: Die Änderungen des Unterrichtshandelns der Lehrkräfte sind nicht abhängig vom Fortbildungsformat.

Die Hypothesen werden sowohl mit zweiseitigen Tests als auch einem einseitigen Test geprüft. Ist das Konstrukt unklar oder herrschen unauflösbare wechselseitige Beziehungen (etwa Aufteilung der Unterrichtszeit auf einzelne Teilprozesse des Experiments) wurden zweiseitige Tests verwendet. Beim Berichten der jeweiligen Ergebnisse wird die Wahl des jeweiligen Tests begründet. Diese Forschungsarbeit kann in doppelter Hinsicht einen Beitrag zur fachdidaktischen Forschung leisten: 1.) Durch die Entwicklung und Validierung eines Instruments zur Erfassung von Änderungen des Unterrichtshandelns bezogen auf das Experimentieren steht nun der Forschung ein Werkzeug zur Verfügung, um Änderungshypothesen bezüglich des Unterrichtshandelns beim Experimentieren prüfen zu können. 2.) Mit dieser Forschungsarbeit kann ein Beitrag zur Schließung eines lange bestehenden Desiderates in der Fortbildungs- und Transferforschung geleistet werden (Capps & Crawford, 2013; Grosche et al., 2020).

2.6 Existierende Rating- und Kodiermanuale: Potenziale und Limitationen

In folgendem Abschnitt sollen publizierte Rating- und Kodiermanuale vorgestellt und bezüglich ihrer Eignung zur Erfassung von Änderungen des Unterrichtshandelns beim Experimentieren (speziell bezüglich des Entdeckenden Experimentierens) bewertet werden. Kriterien für berücksichtigte Rating- und Kodiermanuale waren:

- Ein Fokus auf das Unterrichtshandeln der Lehrperson bzw. des Unterrichtsgeschehens (so wurden Manuale, welche die Lernenden in den Blick nehmen, nicht berücksichtigt).
- Ein Fokus auf das Experimentieren beziehungsweise naturwissenschaftliche Erkenntnisgewinnung.
- In Deutschland eine Publikation nach dem Jahr 2005. Diese ‚Baseline' ist auf die Zäsur durch die Einführung von Bildungsstandards durch die Kultusministerkonferenz in den Naturwissenschaften (KMK, 2005a, KMK, 2005b, KMK, 2005c) und damit auch eine Neuorientierung des Naturwissenschaftsunterrichts zu begründen.
- Die Studie wurde mit Lehrkräften der Sekundarstufe durchgeführt.

Im Besonderen für die internationale Forschung soll kein Anspruch auf Vollständigkeit erhoben werden. Internationale Arbeiten wurden ergänzt, wenn sie in diese Forschungsarbeit, etwa im Rahmen einer Studie zur Wirksamkeit von Fortbildungen (vgl. Kap. 2.3.2), Eingang gefunden haben oder über das Schneeballverfahren (Döring & Bortz, 2016) über bereits benannte Rating- und Kodiermanuale erschlossen wurden.

Tabelle 4: Übersicht über existierende Kodier- und Ratingmanuale

Autor	Themenkomplex	Schwerpunkte/Facetten	Zweck	Ebene	Kodierverfahren
Börlin (2012) (basierend auf Tesch, (2005))	Experimentelles Handeln	Phasen des Experiments Art des Experiments (qual/quant; Demo/Schüler) Organisationsform Materialien (Alltag/Phys. Geräte/Bildschirm)	Ländervergleich D/CH/FINN	Sichtstruktur	1-min-Intervalle
	Experimentelles Handeln	Kontextorientierung Reflexivität Theorieleitung	Ländervergleich D/CH/FINN	Tiefenstruktur	4-stufig Likertskaliert
Brandon, Taum, Young & Pottenger (2008)	Scientific Inquiry	Lehrer-Schüler-Interaktionsketten, beginnend aufgrund einer Frage (klärend, anregend, zusammenfassend) der Lehrkraft	Prüfung Implementation von Standards	Sichtstruktur	Interaktionsketten
Horizon Research (2005)	Scientific Inquiry	Stundenplanung; Stundenumsetzung; naturwissenschaftliche Inhalte; Klassenklima	Änderung	Tiefenstruktur	5-stufig Likertskaliert
Marshall, Smart & Horton, (2009)	Inquiry-based instruction	Art der Instruktion; Diskussions- und Fragekultur; Reflexion und Sicherung	Baseline	Tiefenstruktur	4-stufig Likertskaliert
Stiller, (2015)	Scientific Inquiry	Kultur- und schulstufenspezifische Handlungsmuster Teilprozesse: Schülerselbstständigkeit (explizites/implizites Schülerhandeln)	Ländervergleich D/SWE	Tiefenstruktur	Merkmalsausprägung über jeweils unterschiedliche Anzahl von Modifiern
Schulz (2010)	Experimentierspezifische Qualitätsmerkmale im Chemieunterricht	Unterrichtsdauer, Redeanteile, Organisationsformen, Phasen, Funktion des Experiments	Evaluation	Sichtstruktur	10-sec-Intervalle
	Experimentierspezifische Qualitätsmerkmale im Chemieunterricht	Instruktionseffizienz, Sicherheit, Offenheit, Klarheit & Strukturiertheit, Problemlösender Unterricht, Schülerorientierung, Gelingen des Experiments	Evaluation	Tiefenstruktur	Binärcodierung
Schmitt (2016)	Phasen des naturwiss. Erkenntniswegs	Teilprozesse	Post	Sichtstruktur	10-sec-Intervall
	Phasen des naturwiss. Erkenntniswegs und angemessene Strukturierungshilfen	Setzt die Lehrperson das Gelernte um?	Post	Sichtstruktur	4-stufig Likertskaliert
Turner, Keiffer & Salamo, (2018)	Scientific Inquiry	Arbeiten mit Hypothesen; Kommunikation; Praktische Untersuchung	Änderung	Tiefenstruktur	4-stufig Likertskaliert
Wee, Shepardson, Fast & Harbor, (2007)	Scientific Inquiry	Lehrer als Lotse; wissenschaftsortient. Fragen; Nachweis ist vorrangig; Entwicklung & Ausführung einer Untersuchung; Datenanalyse; Formulierung der Deutung; Kommunikation	Änderung	Tiefenstruktur	5-stufig Likertskaliert

Es zeigt sich mit Blick auf Tabelle 4, dass die Manuals sehr verschiedene Facetten des Unterrichtshandelns zum Experimentieren fokussieren. Auch der Zweck der eingesetzten Manuale unterscheidet sich erheblich. In Kapitel 3.4 soll die Eignung der vorgestellten Manuale für die Erfassung von Unterrichtshandeln zum (Entdeckenden) Experimentieren geprüft werden.

3 Design und Methodik

Ein Ziel dieser Forschungsarbeit ist die Entwicklung geeigneter Instrumente zur Analyse von Änderungen des Unterrichtshandelns im Rahmen von Lehrkräftefortbildungen zum Experimentieren. Im Folgenden sollen die Lehrkräftefortbildung (3.1), die Anlage und Ablauf der Studie (3.2), die Stichprobe (3.3) sowie die eingesetzten Messinstrumente (3.5) vorgestellt werden. Die Eignung bereits vorhandener Messinstrumente wird in Kapitel 3.4 geprüft.

3.1 Die Lehrkräftefortbildung

3.1.1 Grundlage des Fortbildungsdesigns

Dem in von Emden und Baur (2015) vorgestellten Fortbildungsdesign liegt das COACTIV-Mediationsmodell der professionellen Kompetenz (Kunter, Kleickmann & Richter, 2011) zugrunde. Dieses aus vorangegangenen Forschungsarbeiten abgeleitete Wirkungsmodell beschreibt für die Entwicklung professioneller Kompetenz von Lehrkräften sowohl individuelle Voraussetzungen der Lehrkraft als auch externe Wirkfaktoren aus dem Handlungsfeld der Lehrkraft. Das Modell legt dar, dass die professionelle Kompetenz der Lehrkraft, und damit mittelbar ihr Unterrichtshandeln sowie der Lernerfolg der Lernenden, nicht allein von Dispositionen der Lehrkraft selbst abhängig ist, sondern auch von externen Faktoren beeinflusst wird. „Professionelle Kompetenz“ steht bei COACTIV als Bezeichnung für ein umfassendes Bündel von Kompetenzen: Professionswissen (Professional Knowledge), Werthaltungen (Beliefs, Values, Goals), Motivation (Motivational Orientations) sowie Selbstregulation (Self-regulation) (Kunter et al., 2013). Das Professionswissen wiederum lässt sich mindestens in die drei Domänen Inhaltswissen (Content Knowledge, CK), pädagogisches Wissen (Pedagogical Knowledge, PK) sowie Pädagogisches Inhaltswissen (Pedagogical Content Knowledge, PCK) unterteilen (Fischer, Borowski & Tepner, 2012). Das COACTIV-Modell beschreibt neben dem CK, PK und PCK noch die Domänen Organisationswissen (Organizational knowledge, OK) und Beratungswissen (Counseling Knowledge) (Kunter et al., 2013).

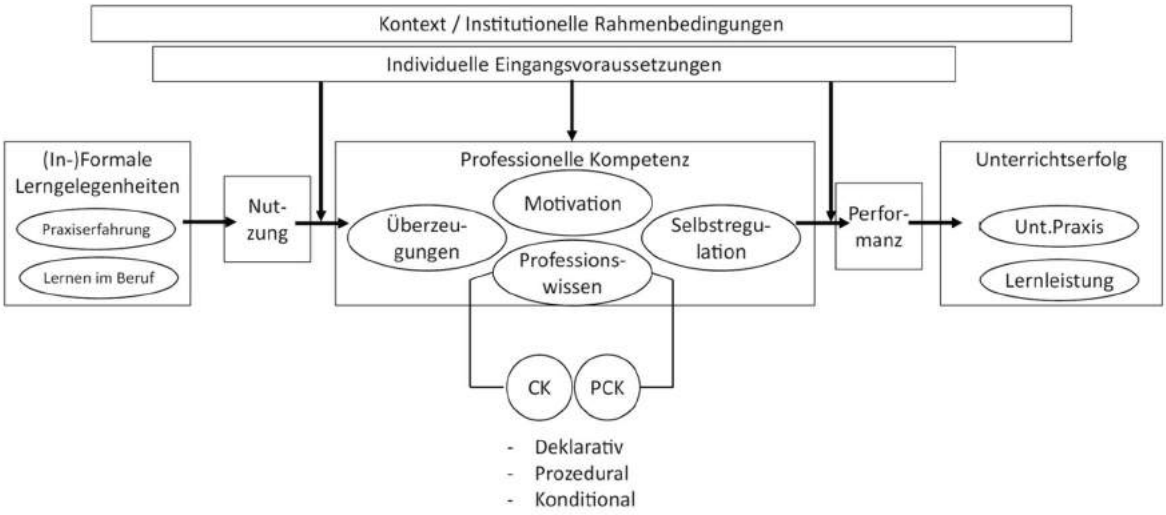


Abbildung 8: COACTIV-Mediationsmodell (Kunter et al., 2011) der professionellen Kompetenz (Darstellung nach Emden & Baur (2017))

Die einzelnen Einflussfaktoren des COACTIV-Mediationsmodells der professionellen Kompetenz wurden von Emden und Baur (2017) mit den Charakteristika effektiver Lehrkräftefortbildungen (Garet, Birman, Porter, Desimone & Herman, 1999; Garet et al., 2001), vgl. auch Kapitel 2.4.2, unterlegt, um somit ein integriertes Wirk- und Gestaltmodell effektiver Lehrkräftefortbildung zu gewinnen (Abbildung 9).

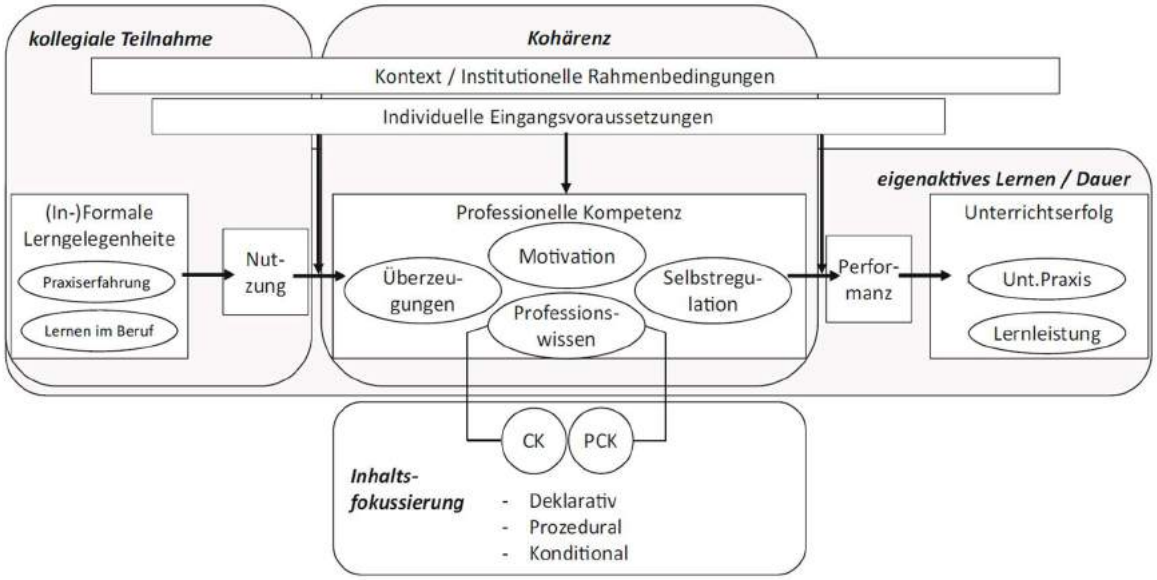


Abbildung 9: Integriertes Wirk- und Gestaltmodell effektiver Lehrkräftefortbildung (Emden & Baur, 2017)

Dieses heuristisch zusammengeführte Modell ermöglicht es nun, orientiert an den Wirkzusammenhängen professioneller Kompetenzentwicklung die Wirksamkeit einzelner Charakteristika der Lehrkräftefortbildung gezielt empirisch zu untersuchen.

Methodische Ausrichtung der Lehrkräftefortbildung

Methodisch sowie in der organisatorischen Anlage greift das Fortbildungskonzept die Charakteristika effektiver Lehrkräftefortbildung nach Garet et al. (2001) (Kohärenz, Inhaltsfokussierung, aktives Lernen, Dauer, kollegiale Teilnahme) konsequent auf (Emden & Baur, 2015; 2017):

- Kohärenz: Orientierung der Fortbildungsinhalte an den Bildungsstandards sowie Berücksichtigung schulspezifischer Gegebenheiten (Ausstattung, Stundentafel, Lernstand) und individueller Dispositionen der Lehrkraft (Ausbildungsstand, Lehrerfahrung, Werthaltungen zu Fortbildungsinhalten und -zielen).
- Inhaltsfokussierung: Eingrenzung des Fortbildungsinhalts auf das Unterrichtskonzept des „Entdeckenden Experimentierens".
- Aktives Lernen: Gemeinsame Diskussion von Fortbildungsinhalten, Entwicklungsaufgaben zwischen den Fortbildungstreffen.
- Dauer: Langfristige Anlage der Lehrkräftefortbildung in drei Phasen über drei Schulhalbjahre mit einer Gesamtkontaktzeit > 20 Stunden.
- Kollegiale Teilnahme: Bildung einer professionellen Lerngemeinschaft des jeweils gesamten naturwissenschaftlichen Kollegiums einer Schule und damit eine gemeinsame Weiterentwicklung des naturwissenschaftlichen Unterrichts im Fachkollegium.

3.1.2 Inhalt und Ziele der Fortbildung

Den Schwerpunkt der untersuchten Lehrkräftefortbildungen bildet das Unterrichtskonzept des ‚Entdeckenden Experimentierens' (Emden & Baur, 2017). Die Fortbildung selbst basiert dabei nicht auf konkreten Fachinhalten, vielmehr wird im Sinne der Kohärenz flexibel auf die Bedarfe der jeweiligen Schulen eingegangen. Vermittelt wird das Konzept in einer Phase fachdidaktisch geprägten Theorie-Inputs. Während dieser Phase finden drei halbtägliche Arbeitstreffen (AT) mit folgenden Schwerpunkten statt:

1. AT: Einführung in die Begriffe *Scientific Literacy* und *Scientific Inquiry* sowie Definition des Experiments und Klärung seiner prototypischen Teilprozesse sowie Möglichkeiten zu deren Öffnung.

2. AT: Fehler und Schwierigkeiten von Lernenden beim Experimentieren.

3. AT: Möglichkeiten und Herausforderungen beim Bewerten von Experimentierkompetenz.

Ziel der Fortbildung ist es, dass die teilnehmenden Lehrkräfte über den fachdidaktischen Input und die Erprobung in der eigenen Praxis ihre professionelle Kompe-

tenz bezüglich des Experimentierens weiterentwickeln und in den Unterricht einbringen, um damit schließlich die Kompetenzentwicklung der Lernenden zu stärken.

Die in dieser Studie untersuchte Lehrkräftefortbildung zielt somit auf einen weiten (distalen) Transfer ab: Die in der Lehrkräftefortbildung erlernten Kompetenzen im Lernkontext sollen auf analoge Problemstellungen übertragen werden (Figuraler Transfer). Auch soll die Lehrkräftefortbildung allgemeine Prinzipien zum geöffneten Experimentieren (etwa über die Matrix der Öffnungsgrade, vgl. Tabelle 1) fördern. Es liegt somit – im Falle positiven Transfers und trotz des klaren inhaltlichen Fokus – ein sogenannter unspezifischer Transfer vor.

3.1.3 Die Fortbildungsformate

Im Rahmen dieser Studie werden zwei nahezu identische Fortbildungsformate untersucht. Allein der Aspekt der kollegialen Teilnahme (vgl. 3.1.1 Grundlage des Fortbildungsdesigns) wird in einem Fortbildungsformat – dem sogenannten kokonstruktiven (vgl. Gräsel et al., 2006; Grosche et al., 2020) Format – vornehmlich durch Gruppenhospitationen und die kollegiale Bearbeitung der Arbeitsaufträge besonders gestärkt; diese Gruppe stellt im Studiendesign die Experimentalgruppe. In dem als individualisiert-konstruktiv beschriebenen Format finden die Hospitationen und Arbeitsaufträge hingegen individualisiert statt. Es kommt zu Einzelhospitationen und -reflexionen des Unterrichts und die Arbeitsaufträge aus den Arbeitstreffen müssen nicht kooperativ bearbeitet werden. Dieses Fortbildungsformat bildet die Kontrollgruppe (für eine detaillierte Gegenüberstellung der Formate siehe auch Abbildung 10 in Kapitel 3.2).

3.2 Anlage und Ablauf der Studie

Bei der vorliegenden Studie handelt es sich um ein quasiexperimentelles Untersuchungsdesign mit Pre-/Post-Erhebung.

Die Naturwissenschaftslehrkräfte einer Schule bilden jeweils eine Fortbildungsgruppe, welche anhand zuvor selbst eingeschätzter Kooperationstiefe innerhalb des Kollegiums von den Fortbildnern einem der beiden Fortbildungsformate zugeordnet wurde. Wurde die Kooperation als stark eingeschätzt, wurden diese Schulen eher dem kokonstruktiven Fortbildungsformat zugeordnet. Durch Zuordnung zu dem individualisiert-konstruktiven Format wäre eine bereits existierende Kooperation mutwillig gestört worden.

Die Fortbildung gliedert sich je Erhebungswelle in drei jeweils halbjährliche Phasen: Erarbeitung, Implementation und Nachbereitung. Vor Beginn der eigentlichen Fortbildung erfolgt die Pre-Erhebung des Unterrichtshandelns. In der Erarbeitungsphase finden Arbeitstreffen mit den Fortbildnern statt. In der Phase der Im-

plementation sollen die Lehrerinnen und Lehrer das in den Arbeitstreffen erworbene Wissen im eigenen Unterricht umsetzen. Dabei finden je nach Fortbildungsformat Gruppenhospitationen und kollegiale Rückmeldung (kokonstruktiv) oder Einzelhospitationen und Entwicklungsempfehlungen (individualisiert-konstruktiv) durch die Fortbildner statt. Die Hospitationen dienen als Grundlage einer Rückmeldung an die Fortbildungsteilnehmenden. In der sich anschließenden Post-Erhebung wird nochmals mittels Videografie das Unterrichtshandeln erfasst. In der Nachbereitungsphase wurden durch eine Onlineplattform (www.EntdeckendesExperimentieren.de) Entwicklungsmöglichkeiten aufgezeigt, weitere Unterrichtsprojekte angestoßen und Reflexionsimpulse gegeben.

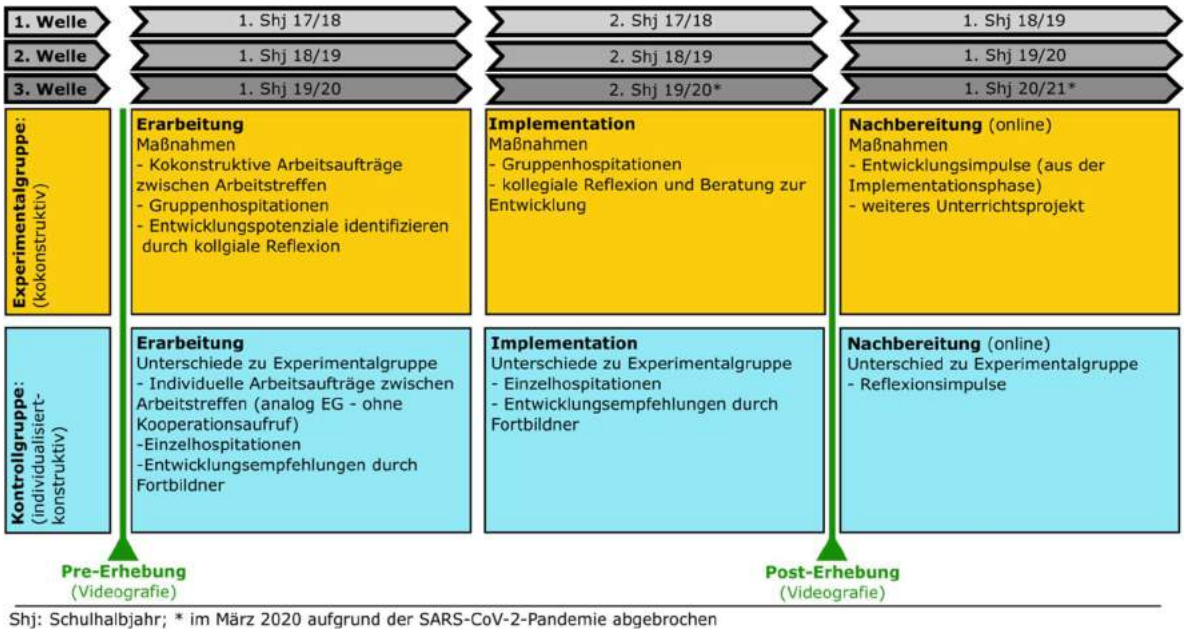


Abbildung 10: Die Phasen der Lehrkräftefortbildung sowie die jeweiligen Unterschiede der beiden Formate

3.3 Die Stichprobe

An der Lehrkräftefortbildung nahmen insgesamt $N = 42$ Lehrkräfte an sieben Schulen (Gemeinschafts- sowie Realschulen) aus Baden-Württemberg teil. Als ‚teilgenommen' wurden alle Lehrkräfte gewertet, welche die Arbeitstreffen (Phase 1) besuchten und mindestens eine der beiden vorgesehenen Hospitationen absolviert haben (Phase 2). Acht Lehrkräfte haben die Fortbildung begonnen, konnten diese jedoch aus unterschiedlichen Gründen nicht abschließen (siehe unten).

Die Fortbildungen wurden in drei zeitlich gestaffelten Wellen durchgeführt. So kann zum einen eine große Stichprobe erfasst werden, zum anderen auch eine qualitativ hochwertige Betreuung durch die beiden Fortbildner (gleichzeitig auch Antragsteller des Projekts) gewährleistet werden. In der ersten Welle nahmen drei Schulen (15 Lehrkräfte) an der Fortbildung teil, in der zweiten vier (21 Lehrkräfte).

Die dritte Welle wurde mit einer Schule (6 Lehrkräfte) begonnen, musste aber wegen der Schulschließungen aufgrund der Verordnung des Landes Baden-Württemberg zur Eindämmung der SARS-CoV-2-Pandemie (Regierung des Landes Baden-Württemberg, 2020) im Frühjahr 2020 nach dem zweiten Arbeitstreffen abgebrochen werden.

Die Teilnahme an der Videografie zur Analyse des Einflusses der Fortbildung auf das Unterrichtshandeln war den Lehrkräften, trotz entsprechender Bitte von Seiten der Fortbildner, freigestellt und kein Teil des Fortbildungsprogramms (etwa über eine Diskussion der videografierten Unterrichtsstunden). Von den 36 Lehrkräften, welche an der Fortbildung teilgenommen hatten, erklärten sich 20 bereit, die Forschungsarbeit mit ihrer Teilnahme an der Unterrichtsvideografie zu unterstützen (24 inklusive der abgebrochenen 3. Welle). Von den 20 Lehrkräften, welche sich bereit erklärt hatten, an der Unterrichtsvideografie teilzunehmen, liegen von 15 Lehrkräften beide benötigten Unterrichtsvideografien vor, bei 5 (9 inkl. der abgebrochenen 3. Welle) fehlt die zweite Unterrichtsvideografie (post).

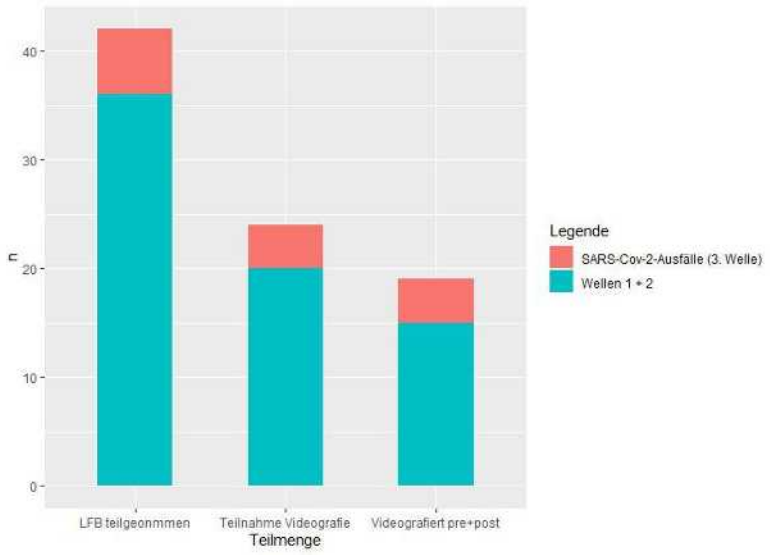


Abbildung 11: Reduktion der Stichprobe durch die SARS-Cov-2-Pandemie

3.4 Eignung der Manuale zur Erfassung der Änderung von Unterrichtshandeln zum (Entdeckenden) Experimentieren

Ziel des zu verwendenden Manuals oder der Manuale muss es sein, Änderungen im Unterrichtshandeln bezüglich des in der Fortbildung vermittelten Unterrichtskonzepts des Entdeckenden Experimentierens (siehe Kapitel 2.1.3) zu identifizieren.

Um das Unterrichtshandeln mit Fokus auf das Experimentieren abzubilden, ergeben sich bezüglich der in Kapitel 2.3 aus den Unterrichtszielen abgeleiteten Prinzipien folgende Bedarfe (Reihenfolge der Prinzipien nicht numerisch geordnet):

P1 Teilprozesse des Experimentierens berücksichtigen / P4 Experimentieren zum Problemlösen einsetzen

➔ Ein Instrument, welches den wissenschaftlichen Experimentierprozess in seinen einzelnen Teilprozessen abbildet und in den Unterrichtskontext einordnet.

P3 Experimentieren üben

➔ Ein Instrument, welches Übungsphasen zum Experimentieren oder einzelner Aspekte erfasst.

P2 Experimentieren als expliziter Unterrichtsgegenstand / P5 Prozessreflektierend experimentieren / P6 Grenzen von Experimenten ansprechen

➔ Ein Instrument, welches eine Reflexion der inneren Bezüge des Experiments sowie eine Reflexion der Charakteristika des Experiments abbildet.

P7 wissenschaftliche Strenge walten lassen

➔ Ein Instrument, welches ein wissenschaftsorientiertes Arbeiten und diesbezügliche Aussagen und Hinweise der Lehrperson abbildet.

P8 Selbstständigkeit der Lernenden beim Experimentieren

➔ Instrumente, welches die Lernendenselbstständigkeit (Öffnungsgrad) sowie entsprechende Angebote der Unterstützung abbildet.

Die Mehrzahl der in Kapitel 2.6 dargestellten Manuale (Brandon et al., 2008; Horizon Research, 2005; Marshall et al., 2009; Stiller, 2015; Turner et al., 2018; Wee et al., 2007) fokussieren auf *Scientific Inquiry* (oder Teilaspekten davon), also die Gesamtheit naturwissenschaftlicher Erkenntnisprozesse. Ist dieser Umstand etwa bezüglich des Kodierens von Lehrer-Schüler-Kommunikation weniger von Relevanz, ergeben sich zwangsläufig Lücken bei experimentierspezifischen Charakteristika wie etwa der Variablenkontrollstrategie und somit Herausforderungen bezüglich der Passung. Dies beschränkt ihre Eignung für den Einsatz in diesem Forschungsprojekt.

Im Folgenden soll geprüft werden, inwieweit die einzelnen Bedarfe durch bereits existierende Instrumente erfüllt werden können.

B1: Ein Instrument, welches den wissenschaftlichen Experimentierprozess in seinen einzelnen Teilprozessen abbildet und in den Unterrichtskontext einordnet.

Zur Identifikation der einzelnen Teilprozesse (Phasen) des Experiments existieren bereits eine Reihe von validierten Manualen, so etwa die Arbeiten von Schulz (2010), Börlin (2012), Stiller (2015) und Schmitt (2016; Phasen des naturwiss. Erkenntniswegs).

B2: Ein Instrument, welches Übungsphasen zum Experimentieren oder einzelner Aspekte abbildet.

Hierzu konnten keine Arbeiten gefunden werden.

B3: Ein Instrument, welches eine Reflexion der inneren Bezüge des Experiments sowie eine Reflexion der Charakteristika des Experiments abbildet.

Wiederum finden sich in zahlreichen Arbeiten Ansätze zur abschließenden Reflexion einer durchgeführten Untersuchung (Horizon Research, 2005; Marshall et al., 2009; Turner et al., 2018; Wee et al., 2007). Diese Manuale eignen sich aufgrund des jeweils allgemeineren Fokus auf Scientific Inquiry nur partiell zur Abbildung der Reflexion des Experiments. Besonders die inneren Bezüge *während* des Experimentierens (etwa Aspekte wie: „Warum benötigen wir eine Hypothese zur Planung unseres Experiments?“, „Passen unsere Hypothesen zur Forschungsfrage?“) sowie speziell der Reflexion der Charakteristika des Experiments werden nur rudimentär erfasst.

B4: Ein Instrument, welches ein wissenschaftsorientiertes Arbeiten und diesbezügliche Aussagen und Hinweise der Lehrperson abbildet.

Auch hier existieren bereits Arbeiten zum Unterrichtshandeln bezüglich des allgemeinen wissenschaftsorientierten Arbeitens (Turner et al., 2018; Wee et al., 2007), jedoch nicht speziell ausgerichtet auf das Experimentieren. Diese speziell auf das Experimentieren ausgerichteten Instrumente müssten beispielsweise Aspekte bezüglich der Umsetzung von angemessenen Versuchs- sowie entsprechenden Testansätzen beinhalten.

B5: Ein Instrument, welches Angebote zur Unterstützung der Lehrperson im Experimentierprozess abbildet.

Das Instrument von Brandon et al. (2008) erfasst durch Lehrer-Schüler-Interaktionsketten ausgehend von Inquiry-spezifischen Lehrerfragen die Kommunikationsstrategien von Lehrkräften. Diese Interaktionsketten bilden jedoch, wie Brandon et al. (2008) selbst herausstellen, nur einen kleinen Teil möglicher Unterstützung (verbale Impulse im Rahmen einer Art sokratischen Dialogs) ab.

B6: Ein Instrument welches die Lernendenselbstständigkeit (Öffnungsgrad) in den einzelnen Teilprozessen des Experiments abbildet.

Stiller (2015) identifizierte für Teilprozesse naturwissenschaftlicher Erkenntnisgewinnung Merkmale expliziten beziehungsweise impliziten Schülerhandelns, um damit Grade von Lernendenselbstständigkeit zu erfas sen. Eine Einteilung nach Öffnungsgraden findet sich in dem Instrument nicht.

Fazit

Bezüglich der Bedarfe B1, und mit Einschränkungen bezüglich B4, kann bei der Konzeption von Instrumenten auf bereits validierte Instrumente vorangegangener Arbeiten zurückgegriffen werden. Zur Befriedigung des Bedarfs nach einem Instrument, welches eine Reflexion der inneren Bezüge des Experiments sowie eine Reflexion der Charakteristika des Experiments und damit den Kern des in der Fortbildung vermittelten Konzepts des Entdeckenden Experimentierens abbildet (B3), kann nur in geringen Umfang auf Vorarbeiten zurückgegriffen werden. Dies ist auch darauf zurückzuführen, dass hier ein neues Fortbildungskonzept untersucht werden soll.

Um die beschriebenen Bedarfe voll zu erfüllen, sollen im Rahmen dieser Forschungsarbeit – teils unter Rückgriff auf bestehende Arbeiten – drei Instrumente angepasst beziehungsweise entwickelt werden:

I1: Eine Basiskodierung der Teilprozesse des Experiments (B1) (intervallcodiert). Die Basiscodierung dient der Einteilung des Geschehens auf der Ebene der „Sichtstruktur“ (siehe Kapitel 3.5.1, direkt beobachtbares Handeln) in Organisationsformen, Unterrichtsphasen und Aktivitäten (Marquardt, 2011). In dieser Studie dient die Basiskodierung der Einteilung des Experiments in seine Teilprozesse.

I2: Eine Kodierung der Tiefenstruktur (siehe Kapitel 3.5.1) des Unterrichtshandelns zum Experimentieren (B2, B3, B4) inklusive seiner Einbettung in die Unterrichtsstunde (B1) sowie Angebote zur Unterstützung der Lernenden im Experimentierprozess (B5).

I3: Eine Kodierung der Sichtstruktur zur Erfassung des Offenheitsgrades der einzelnen Teilprozesse (B6).

3.5 Messinstrumente

3.5.1 Theoretische Herleitung und Begründung der Methode

Das Ziel der vorliegenden Forschungsarbeit ist die Untersuchung des Einflusses zweier Fortbildungstypen auf das Unterrichtshandeln mittels einer Zusammenhangsanalyse (Mayring, 2007). Zur Erfassung von Unterrichtshandeln eignen sich prinzipiell verschiedene Ausprägungen der wissenschaftlichen Beobachtung

(Herrle & Breitenbach, 2016). Im Folgenden sollen die verschiedenen Dimensionen der Beobachtung vorgestellt und ihre Eigenschaften in Bezug auf die Anforderungen zum Einsatz in dieser Studie diskutiert werden.

3.5.1.1 Beobachtung

Helmke bezeichnet die Beobachtung als „Königsweg“ zur Beschreibung und Bewertung von Unterricht (Helmke, 2017, S. 292). Das Beobachten als Datenerhebungsmethode im sozialwissenschaftlichen Kontext kann als eine 1. absichtliche, 2. zielgerichtete, 3. bewusst selektive Form des Wahrnehmens verstanden werden (Graumann, 1966). Das Beobachten ist ein Begriff mit einer Vielzahl von Ausprägungen, welche sich anhand von Dimensionen charakterisieren lassen (Gniewosz, 2011; Helmke, 2017; Kochinka, 2010).

Eine Dimension der Beobachtung bildet der Grad der Anleitung mit den Polen der strukturierten (standardisierten) und der unstrukturierten (offenen) Beobachtung (Helmke, 2017; Kochinka, 2010). Zwischen diesen Polen existieren zahlreiche graduelle Abstufungen. Während eine unstrukturierte Beobachtung vor der Erhebung darauf verzichtet, den Gegenstand der Beobachtung explizit zu fixieren, fokussiert die strukturierte Beobachtung auf einzelne zuvor spezifizierte Merkmale, Verhaltensweisen oder Ähnliches. Die unterschiedlichen Pole dieser Dimension erfordern auch entsprechend ausgerichtete Erhebungsinstrumente. Je strukturierter die Beobachtung erfolgen soll, desto detaillierter müssen die Instrumente ausgearbeitet sein (vgl. Faßnacht, 1995). Mögliche Beobachtungsinstrumente zur strukturierten Beobachtung sind (Appel & Rauin, 2016; Böhm-Kasper & Weishaupt, 2004):

- Zeichensysteme: Es wird das Auftreten eines zuvor definierten Merkmals/Verhaltens erfasst.
- Kategoriensysteme: Zuvor definierte (Verhaltens-)Merkmale werden in einem System klassifiziert.
- Ratingsysteme (Schätzskalen): Die Ausprägung eines Merkmals oder Verhaltens wird anhand einer Skala geschätzt.

Die unstrukturierte Beobachtung eignet sich für explorative, hypothesengenerierende Forschungsdesigns, die strukturierte Beobachtung eignet sich demgegenüber zur Prüfung zuvor abgeleiteter Hypothesen.

Zur Beantwortung der in dieser Forschungsarbeit aufgeworfenen Forschungsfragen und daraus abgeleiteten Hypothesen ist somit eine strukturierte Beobachtung sinnvoll.

Eine weitere Dimension ist die Transparenz der Beobachtung (Bortz & Schuster, 2010) gegenüber den Beforschten: Wissen die Studienteilnehmenden, dass sie be-

obachtet werden, wird dies als offene Beobachtung bezeichnet. Bleibt die Beobachtung den Teilnehmenden verborgen, spricht man von verdeckter Beobachtung. Ein Vorteil der verdeckten Beobachtung ist eine dadurch ausbleibende mögliche Reaktion der Beobachtungssubjekte auf die Beobachtungssituation. Die verdeckte Beobachtung ist jedoch – besonders im geschützten Raum „Schule“ – aus ethischer Sicht nicht zu rechtfertigen (Ethik-Rat der DGfE, 2016).

Im Rahmen dieser Forschungsarbeit soll die Änderung des Unterrichtshandelns von Lehrkräften untersucht werden. Dies erfordert die Beobachtung von möglichst authentischem Unterricht. Scheint es aus der Forschungsperspektive verlockend, durch eine verdeckte Beobachtung die Interaktion mit dem Beobachtungsfeld zu minimieren, so verbietet sich forschungsethisch – aber auch rein rechtlich – eine verdeckte Beobachtung von Schulunterricht.

Beobachtungsformen können außerdem danach unterschieden werden, inwieweit der Beobachter in das Beobachtungsfeld involviert ist (Gniewosz, 2011). Diese Dimension beschreibt somit den Grad der Teilnahme anhand der beiden Pole teilnehmender Beobachtung und nicht teilnehmender Beobachtung. Eine teilnehmende Beobachtung kann passiv (der Beobachter ist anwesend, aber interagiert nicht aktiv mit dem Beobachtungsfeld) als auch aktiv (der Beobachter nimmt eine Doppelrolle ein und beteiligt sich, neben der Beobachtung, am sozialen Geschehen) durchgeführt werden. Eine teilnehmende Beobachtung kann aus pragmatischen Gründen sinnvoll sein (etwa eine Lehrkraft, welche gleichzeitig das Sozialverhalten der Lernenden beobachtet) oder notwendig, um das Beobachtungsfeld überhaupt erst zu erschließen (etwa Studien in Subkulturen).

Eine teilnehmende Beobachtung ist im Rahmen dieser Forschungsarbeit zur Erschließung des Beobachtungsfelds nicht notwendig. Eine nicht teilnehmende Beobachtung ist sinnvoll, um das Beobachtungsfeld möglichst gering zu beeinflussen.

Die Beobachtung kann in der Dimension der Mittelbarkeit (technisch) vermittelt oder unvermittelt erfolgen (Kochinka, 2010). Setzt der Beobachter allein seine Sinne im Beobachtungsfeld ein und protokolliert seine Beobachtungen während oder im Anschluss, ist dies eine unvermittelte Beobachtung. Eine vermittelte Beobachtung liegt vor, wenn das Beobachtungsfeld beziehungsweise Ausschnitte davon, aufgenommen werden (beispielsweise audiovisuell über Kamera oder rein auditiv über ein Tonbandgerät), um anschließend die Aufnahmen zu beobachten. Die Aufnahme selbst stellt noch keine Beobachtung dar. Die vermittelte Beobachtung mittels Videografie (videobasierte Analyse) hat sich zu einer eigenständigen Forschungsmethode in der erziehungswissenschaftlichen Forschung (Dinkelaker & Herrle, 2009; Janik & Seidel, 2009) und auch speziell der naturwissenschaftlichen Fachdidaktiken entwickelt (Brückmann & Duit, 2014). Im Folgenden sollen

die Vor- und Nachteile unvermittelter sowie vermittelter Beobachtung (am Beispiel der Videografie, da in dieser Studie das Unterrichtshandeln analysiert werden soll) diskutiert und für den Einsatz in dieser Studie abgewogen werden.

Die unvermittelte Beobachtung findet bei der Erfassung von Unterrichtshandeln bzw. Merkmalen des Unterrichts weiterhin großen Zuspruch (etwa bei Antoniou & Kyriakides, 2013; Grigg et al., 2013; Marshall & Smart, 2013, vgl. Zusammenschau von Studien zum Unterrichtshandeln bezogen auf *Inquiry*, Kapitel 2.4.2).

Für den Einsatz einer unvermittelten, direkten Beobachtung gegenüber der Videografie sprechen oft pragmatische Gründe. Die direkte Beobachtung ist mit deutlich reduziertem technischem, rechtlichem und zeitlichem Aufwand durchzuführen. Es wird weder eine Hardware (Kameras zur Aufnahme des Beobachtungsfelds) noch eine Software (zur Aufbereitung der Videos, etwa Videoschnitt und Bearbeiten der Tonspuren) benötigt. Die rechtlichen Hürden (Datenschutz) und damit administrativen Anforderungen sind bei einer videografischen Aufnahme von Unterricht ungleich höher als bei einer Unterrichtshospitation mit unmittelbarer Beobachtung.

Diesen pragmatischen Gewinnen steht eine Reihe von Nachteilen einer direkten Beobachtung im Vergleich zum Einsatz von Videografie entgegen:

- Erfolgt das Protokollieren bei der unvermittelten Beobachtung nicht augenblicklich, können Erinnerungsverzerrungen die Beobachtung verfälschen (Bortz & Schuster, 2010; Gniewosz, 2011).
- Beobachtung ist aufgrund der Fülle der potenziell zu verarbeitenden Informationen stets selektiv (Kochinka, 2010; Reh, 2012). Eine wiederholte Beobachtung oder eine Transkription zur umfassenden Sammlung aller relevanten Informationen ist nicht möglich.
- Wird ein Beobachtungssystem zu komplex, besteht die Gefahr, dass die Beobachter überfordert werden und es dadurch zu Informationsverlust kommt (Gniewosz, 2011). Wurde das Beobachtungsfeld videografiert, kann diese Überforderung des Beobachters durch Zurückspulen, wiederholtes Abspielen oder Abspielen in Zeitlupe aufgelöst werden.
- Die unvermittelte Beobachtung ist nicht reproduzierbar. Bei der Videografie können jederzeit die einzelnen Schritte von der standardisierten Aufnahme der Videos (gegebenenfalls über eine Transkription) zum Rating/Kodieren und der Auswertung durch externe Dritte nachvollzogen werden (Tuma, Schnettler & Knoblauch, 2013). Der Schluss von der unvermittelten Beobachtung zu den Beobachtungsaufzeichnungen (Protokoll, Rating- oder Kodiermanual) kann hingegen bei einem einzelnen Beobachter nicht mehr nachvollzogen werden. Damit ist die Beobachtung nicht mehr

wie für eine wissenschaftliche Beobachtung gefordert intersubjektiv überprüfbar (Döring & Bortz, 2016). Ist als einziger im Beobachtungsfeld anwesender Beobachter der publizierende Forscher präsent, existiert potenziell ein nicht zu kontrollierendes Manipulationsrisiko, die Objektivität der Forschungsarbeit ist nicht zu gewährleisten.

Um diesen Fehlerquellen zu begegnen, ist es notwendig, dass bei *jeder* Beobachtung mindestens zwei Beobachter unabhängig voneinander Daten erheben, welche anschließend zur Prüfung der wissenschaftlichen Gütekriterien herangezogen werden. Dies wiederum erhöht den zeitlichen Aufwand der Studie. Zur Kritik an der unmittelbaren Beobachtung sei angemerkt, dass auch die Videografie – wie jede Forschungsmethode – nicht frei von potenziellen Fehlerquellen ist, dass sie durch die ihr inhärente Eigenschaft zur wiederholten Beobachtung aber oben genannten Fehlerquellen der unmittelbaren Beobachtung, welche die Aussagekraft der gesamten Studie unterminieren können, zu begegnen vermag.

Zur Analyse von Unterrichtshandeln ist die wissenschaftliche Beobachtung des Unterrichts die einzig probate Forschungsmethode. Um mit dem Beobachtungsfeld möglichst wenig in Interaktion zu treten und damit die Objektivität der Erhebung zu erhöhen, ist eine nicht teilnehmende Beobachtung sinnvoll. Die Beobachtung muss außerdem aus forschungsethischen Gründen transparent erfolgen. Zur Prüfung der Änderungshypothese bezüglich des Unterrichtshandelns ist eine objektive und reliable Beobachtung über alle Messzeitpunkte hinweg zu gewährleisten. Dies ist über eine eng strukturierte Beobachtung zu erreichen. Eine mittelbare Beobachtung mittels Videografie ist bei einem komplexen Handlungsmuster wie dem Unterrichtshandeln und entsprechend umfangreichen Erhebungsinstrumenten der unmittelbaren Beobachtung vorzuziehen.

Roth (2009) führt als Gründe, warum in der TIMS-Studie (Hiebert et al., 2003) die Videografie der unmittelbaren Beobachtung (‚*first-hand classroom observations*') als Erhebungsmethode vorgezogen wurde unter anderem an, dass dadurch 1. eine präzisere und umfassendere Analyse von Unterrichtshandeln möglich ist, 2. eine höhere Interraterreliabilität und eine objektivere Beobachtung zu erzielen ist und 3. die Videos für weitere Analysen zur Verfügung stehen.

3.5.1.2 Videografie als mittelbare Beobachtung

Zur Untersuchung des Unterrichtshandelns (beziehungsweise der Änderungen im Unterrichtshandeln) von Lehrkräften hat sich die Videografie in den letzten Jahrzehnten zunehmend als fruchtbare Forschungsmethode erwiesen. Besonders zur Identifikation von Handlungsmustern und unterrichtsmethodischen Abläufen wird die Videografie auch in der naturwissenschaftsdidaktischen Forschung eingesetzt (Brückmann & Duit, 2014). Sie erlaubt durch eine „unvergleichliche Fülle von Wahrnehmungsaspekten" (Tuma et al., 2013, S. 31) nicht nur eine ganzheitliche

Erfassung des Unterrichtsgeschehens, sondern auch die Verbindung von qualitativen sowie quantitativen Analysemethoden (Helmke, 2017).

Auch in der Transfer- und Fortbildungsforschung ist die Videografie aufgrund ihrer Unmittelbarkeit die zu bevorzugende Erhebungsmethode (Lipowsky, 2010 und Kapitel 2.4.1).

Im Folgenden werden wesentliche Charakteristika der Videografie, speziell zur Erfassung von Unterrichtshandeln, vorgestellt und im Hinblick auf die Erfordernisse dieser Forschungsarbeit diskutiert.

Invasivität und Selektivität der Videografie

Obwohl die Videografie eine nicht teilnehmende Beobachtungsform ist, tritt sie doch mit dem Beobachtungsfeld in Resonanz, sie ist invasiv (Herrle & Breitenbach, 2016). Allein durch die Anwesenheit von Kameras (und möglicherweise Kameraleuten) verändern sich die Rahmenbedingungen des Beobachtungsfelds. Diese Resonanz lässt sich im Rahmen einer offenen Beobachtung nicht unterbinden. Die Stärke der Resonanz ist zum einen davon abhängig, wie stark („störend“) die Videografie von den Personen im Beobachtungsfeld wahrgenommen wird (eine Videokamera auf einem Stativ in der Ecke des Klassenraums auf das Beobachtungsfeld ist weniger invasiv als eine von einer/m Kamerafrau/mann dynamisch im Klassenraum geführte Videokamera) und zum anderen, wie die Videokameras rezipiert werden: Werden die Videokameras und ihre Bediener als Gäste des Unterrichts wahrgenommen oder entsteht durch die Anwesenheit eine (bedrückende) Prüfungsatmosphäre (Herrle & Breitenbach, 2016)?

Um die Invasivität der Videografie und damit eine ungewollte Änderung der Handlungsabläufe und des Verhaltens der Personen im Beobachtungsfeld möglichst zu minimieren, werden in dieser Forschungsarbeit statisch positionierte Videokameras eingesetzt. Außerdem werden vor der Videografie die Lernenden über den Zweck der Videoaufnahmen informiert. Anschließend haben diese auch die Möglichkeit, Fragen an den im Klassenraum anwesenden Mitarbeiter des Projekts zu stellen. Dadurch soll einer negativen Rezeption der Videografie bei den Lernenden begegnet werden. Es ist festzustellen, dass Irritationen, welche durch die Videokamera im Beobachtungsfeld erzeugt werden, in der Regel im Verlauf der Videografie schnell abnehmen (Herrle & Breitenbach, 2016) und somit Befürchtungen, dass sich Teilnehmende im Beobachtungsfeld aufgrund der Videoaufzeichnung unnatürlich verhalten, durch Forschung entkräftet werden konnte (Helmke, 2017).

Ebenen der Unterrichtsbeobachtung

Unterricht lässt sich im Wesentlichen in den vier voneinander getrennten Betrachtungsebenen ‚Organisationsformen', ‚Methoden des Unterrichts', ‚Sozialformen' sowie ‚Lehr-Lern-Prozesse' beschreiben (Kunter & Trautwein, 2013). Eine weitere Form der Klassifizierung von Unterricht ist die Unterscheidung von Sicht- und Tiefenstrukturen (Kunter & Trautwein, 2013; Seidel, 2003). Als Sichtstrukturen werden dabei jene Strukturen bezeichnet, welche für den Beobachter mit wenig eigener Interpretation erschlossen werden können. Die Betrachtungsebenen ‚Organisationsformen', ‚Methoden des Unterrichts' und ‚Sozialformen' lassen sich bereits gut über Sichtstrukturen erschließen. Es ist beispielsweise in der Betrachtungsebene ‚Sozialform' nur wenig Interpretation notwendig, um zu erfassen, ob im Plenum, in Kleingruppen, in Partner-, oder in Einzelarbeit gearbeitet wird. Tiefenstrukturen sind hingegen jene Strukturen, welche vom Beobachter nicht unmittelbar zu erfassen sind, wie etwa die Art und Qualität der Interaktion zwischen Lehrperson und Lernenden. Hier ist eine gewisse Interpretation vonseiten des Beobachters notwendig. Geben die über die Sichtstruktur zu erfassenden Merkmale den Rahmen der Unterrichtsgestaltung vor, so hat die Tiefenstruktur hingegen eine größere Erklärungsmacht im Hinblick auf die Unterrichtsqualität (Helmke, 2017).

Bei der Erfassung von Sichtstruktur und Tiefenstruktur gibt es jeweils eigene Klassen von Beobachtungsinstrumenten. Zur Erfassung der Sichtstruktur werden niedrig inferente Beobachtungsinstrumente, zur Erfassung der Tiefenstruktur hingegen hoch inferente Instrumente eingesetzt. Die Inferenz bezeichnet in diesem Zusammenhang den Grad der Interpretation durch den Beobachter, der bei der Beobachtung erforderlich ist (Lotz, Gabriel & Lipowsky, 2013, siehe Tabelle 5).

Tabelle 5: Niedrig und hoch inferente Verfahren (Lotz et al., 2013)

	Niedrig inferente Verfahren	Hoch inferente Verfahren
Bezeichnung	Kodierung/Kategoriensystem	Rating/Schätzverfahren
Art der Datengewinnung	Erfassen der Häufigkeit und Dauer leicht beobachtbarer Unterrichtsereignisse	Schätzverfahren zum Erfassen der Ausprägung eines Merkmals auf einer vorab definierten Skala
Ziel	Beschreibung der Unterrichtsgestaltung	Bewertung der Unterrichtsqualität
Analyseeinheit	in der Regel kurze Abschnitte, z. B. 10-Sekunden-Intervalle oder kurze Ereignisse	in der Regel längere Unterrichtssequenzen oder ganze Unterrichtsstunden
Grad der Interpretation	Verfahren orientieren sich fast ausschließlich an direkt beobachtbarem Verhalten; geringe Spielräume für die Beobachter	Verfahren orientieren sich nur teilweise an direkt beobachtbarem Verhalten; Schlussfolgerungen der Beobachter nötig
Beispiele	Kodierung der Sozialformen	Einschätzung des Unterrichtsklimas

Im Rahmen dieser Forschungsarbeit zur Analyse von Änderungen im Unterrichtshandeln soll sowohl auf niedrig inferente Verfahren als auch auf hoch inferente Verfahren zurückgegriffen werden, um ein möglichst ganzheitliches Bild von allen Betrachtungsebenen des Unterrichts zeichnen zu können.

Setting der Videografie

Unterricht ist ein „überkomplexes Geschehen“ (Herrle & Dinkelaker, 2016, S. 78), das den Forschenden dazu zwingt, Selektionsentscheidungen zu treffen und Relevantes von Irrelevantem zu trennen (Herrle & Breitenbach, 2016). Diese Selektivität der Videografie wird von der Forschungsfrage geleitet.

Bei der Videoaufnahme von Unterricht zur anschließenden Beobachtung und Analyse von Unterrichtshandeln ist eine möglichst „konstruktionslogische Vollständigkeit des Datums“, also die umfassende Konservierung von „multimodalen Ausdrucksressourcen“ (Stimmen, Mimik, Gestik, Bewegungen der Personen) sowie der „räumlich-visuellen Sachverhalte“ (Raumstruktur) erstrebenswert (vgl. Deppermann, 2014, S. 20). Dieses Ziel der umfassenden Konservierung muss gegen den damit steigenden technischen Aufwand (zusätzliche Videokameras, Mikrofone) – und somit auch mittelbar die Invasivität – sowie auch eine Steigerung des Auswertungsaufwands abgewogen werden (Dinkelaker & Herrle, 2009). Der Einsatz von zwei Videokameras inklusive Mikrofon im Klassenraum hat sich bei der Aufnahme von Unterrichtssettings bewährt (Brückmann & Duit, 2014; Helmke, 2017; Herrle & Breitenbach, 2016).

Die Position der Kameras im Beobachtungsfeld, in dieser Studie dem Klassenraum, wird im Wesentlichen von zwei Determinanten geleitet: Es ist aufgrund des Erkenntnisinteresses erstrebenswert, ein möglichst umfassendes Bild aller Lehr-Lern-Prozesse zu gewinnen. Da zur Erfassung des Unterrichtshandelns der Fokus nicht nur auf der Lehrperson selbst, sondern auch auf der Interaktion mit den Lernenden und der Lernenden untereinander (setzen diese etwa geforderte Arbeitsaufträge um?, erkennt die Lehrkraft Verständnisprobleme einzelner Lerngruppen?) liegt, ist es sinnvoll eine Videokamera auf die Lehrperson auszurichten und eine Videokamera möglichst vollständig auf die Lernenden auszurichten (Dinkelaker & Herrle, 2009). Technisch limitiert wird dieses Streben dadurch, dass die Bildqualität bei Gegenlicht stark gemindert wird (Helmke, 2017).

Aufgrund der oben erörterten Überlegungen wurden in der hier vorliegenden Forschungsarbeit beide Videokameras an die Fensterfront des Klassenraums gestellt mit dem Fokus gerichtet auf die Lehrperson respektive die Lernenden (siehe Abbildung 12). Zur weiteren Standardisierung und Vergleichbarkeit der Aufnahmen wurden stets Stative eingesetzt und der Fokus der Videokamera im Verlauf der Unterrichtsstunde nicht mehr verändert, das Kamerahandeln ist neutral (Tuma et al., 2013). Wenn möglich wurden die im Unterricht verwendeten Unterrichtsmaterialien (Arbeitsblätter, Experimentiermaterial, Präsentationsfolien etc.) sowie der Tafelanschrieb umfassend fotografisch dokumentiert (vgl. Herrle & Breitenbach, 2016).

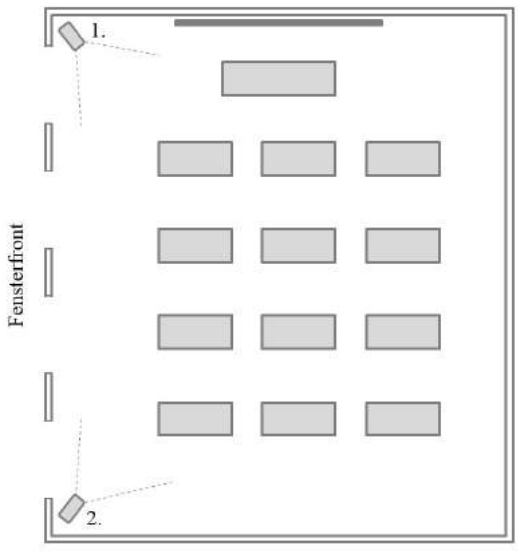


Abbildung 12: Skizze zur Positionierung der Videokameras (1. und 2.) im Klassenraum

Transkription der Videoaufnahmen

Während bei Forschungsarbeiten mit rein auditiven Aufzeichnungen (etwa Interviews) die Transkription dieser Aufzeichnungen – vor allem zum Zweck der Validierung – praktiziert wird und forschungsmethodischer Standard ist, so werden

Videografien aktuell weit weniger häufig vor der Auswertung transkribiert (Moritz, 2014). Die Argumente diesbezüglich sollen im Folgenden knapp vorgestellt und anschließend bezüglich der vorliegenden Forschungsarbeit abgewogen werden.

Für eine Transkription des Videomaterials spricht, dass damit Beobachtungen und Analysen ermöglicht werden, deren Erfassen respektive Erstellen allein aus der Videografie die Auffassungsgabe der Forschenden übersteigen würde (Moritz, 2014). Außerdem wird das Videomaterial seiner ‚Flüchtigkeit' enthoben (Moritz, 2014) und damit erst für bestimmte forschungsmethodische Zugänge (etwa einer induktiven Kategorienbildung im Zuge einer Inhaltsanalyse) produktiv nutzbar. Die Nachvollziehbarkeit einzelner Schritte der Erkenntnisgewinnung wird erhöht: Das zentrale Argument der Videotranskription, wie auch bei einer reinen auditiven Transkription, ist seine validierende Funktion (Moritz, 2014), also der durch die Transkription erzeugte empirische Beleg auch für nichtsprachliche Ereignisse (etwa Bewegungen und Interaktionen im Raum). Auf diese Belege kann sich im Verlauf der Forschungsarbeit sowohl bei der Entwicklung von Messinstrumenten als auch bei der Ergebnisdiskussion gestützt werden.

Ein zentrales Argument gegen die Anfertigung einer Transkription ist die durch die Überkomplexität der Videoaufnahmen bedingte Verdichtung, Reduktion und Fokussierung bei der Transformation von audiovisuellen Daten in eine Zeichensprache (meist Schriftsprache; Moritz, 2014) und der damit potenziell einhergehende Informationsverlust. Dieser damit einhergehende *Codewechsel* wird forschungsmethodisch teils kritisch rezipiert (Kurt, 2010).

Um diesem Argument zu begegnen, sei angemerkt, dass die Videotranskription nicht, wie bei Audiotranskriptionen üblich, als alleinige Substitution der originären Videografie dienen muss: Mittels Hypercoding (Irion, 2010), also der parallelen Ausbeutung der Videoaufnahme und des entsprechenden Transkripts (Hypercoding 1. Ordnung) oder sogar noch unter zusätzlicher Berücksichtigung weiterer, ergänzender Daten (Hypercoding 2. Ordnung) bei der Analyse lassen sich die Vorteile der Transkription ausschöpfen, ohne Gefahr zu laufen, für die Forschungsfrage wesentliche Inhalte im Transkriptionsprozess zu verlieren.

Im Rahmen dieser Forschungsarbeit wird aus oben erläuterten Gründen das Hypercoding 2. Ordnung bei der Analyse der Videografien eingesetzt, um die Vorteile der Transkription (gesteigerte Validität durch empirischen Beleg, Entzug der ‚Flüchtigkeit') zu nutzen, ohne damit potenziellem Informationsverlust durch die alleinige Verwendung des Transkripts zu unterliegen. Vielmehr wird durch Berücksichtigung weiterer, ergänzender Informationen (Arbeitsblätter, Vortragsfolien, Fotografien von Materialkisten und Tafelaufschrieben) der Blick auf das Beobachtungsfeld ergänzt.

Die Transkription selbst wird als Beobachtungsprotokoll (Dinkelaker & Herrle, 2009) beziehungsweise *Running Text* (Edwards & Lampert, 1993) angefertigt und lehnt sich dabei im Format Theatertexten an. Verbale sowie nonverbale Kommunikation werden in einem zeitlichen Kontinuum analog zu ihrem Auftreten im Video erfasst. Für weitere im Folgenden nicht erläuterte Typen der Transkription siehe Edwards & Lampert, (1993) und Moritz (2014). Auf phonetische Explikationen wird, außer durch die Verwendung von Frage- und Ausrufezeichen, zugunsten der Lesbarkeit und Ökonomie verzichtet. Der Text wird insgesamt geglättet und in normales Schriftdeutsch überführt (vgl. Kuckartz, 2007).

3.5.2 Entwicklung und Qualität der Kategoriensysteme

Im Folgenden wird die Entwicklung bzw. Ableitung der in der vorliegenden Arbeit verwendeten Kategoriensysteme beschrieben und bezüglich der Gütekriterien Objektivität, Reliabilität und Validität diskutiert. Die in der Studie verwendeten und im Folgenden diskutierten Kategoriensysteme finden sich in Tabelle 6:

Tabelle 6: Übersicht über die verwendeten Kategoriensysteme

Nr.	Bezeichnung	Kodierverfahren	Stichprobenplan	Interferenzen
1	Unterrichtshandeln von Lehrkräften beim Experimentieren	Dichotomes Kategoriensystem	Ereignisstichprobe	niedrig/ mittel
2	Phasen des Experimentierprozesses	Mehrstufiges Kategoriensystem	Zeitstichprobe (15s)	niedrig
3	Öffnungsgrade der Teilprozesse beim Experimentieren	Schätzverfahren (4-stufig Likert-skaliert)	Ereignisstichprobe	hoch
4	Definition des Experiments	Schätzverfahren (3-stufig Likert-skaliert)	Ereignisstichprobe	mittel

Entwicklung des Kategoriensystems „Unterrichtshandeln von Lehrkräften beim Experimentieren“ mittels qualitativer Inhaltsanalyse (Mayring, 2015) und Diskussion seiner Güte

Erst durch eine sich der Transkription anschließende Kodierung entstehen aus den (verschriftlichten) Beobachtungen Daten, welche nachfolgend einer Analyse zugänglich sind (Döring & Bortz, 2016). Dazu wird wo möglich auf bestehende oder zu entwickelnde Instrumente zurückgegriffen. Wie in Kapitel 3.4 gezeigt werden konnte, erfüllt keines der vorgestellten Kodier- bzw. Ratingmanuale die für diese Forschungsarbeit aufgestellten Kriterien vollständig. Aus diesem Grund wird, neben dem Rückgriff auf die Adaption zweier existierender Kodiermanuale (Stiller, 2015, Baur et al., 2020), ein Kategoriensystem induktiv neu mittels zusammenfassender qualitativer Inhaltsanalyse nach Mayring (2015) entwickelt. Dieses soll anschließend punktuell deduktiv ergänzt werden. Der Ansatz, zuerst induktiv das

Kategoriensystem zu entwickeln und dieses anschließend deduktiv zu ergänzen, wurde aus zwei Gründen gewählt:

1. Erhöhung der Passung und Flexibilität: Durch die Verwendung von Aufzeichnungen, welche im Rahmen des Forschungsprojekts entstanden sind, und eine primäre induktive Kategorienbildung wird die Passung zwischen Instrument (Kategoriensystem) und den Aufzeichnungen erhöht. Gerade da eine erstmalig durchgeführte Lehrkräftefortbildung analysiert werden soll, ist unklar, welche spezifischen Merkmale in welcher Ausprägung in den Aufzeichnungen auftreten werden. Es scheint sinnvoll, dieser Unsicherheit mit einer möglichst offenen – also induktiven – Entwicklung zu begegnen.
2. Über eine induktive Kategorienbildung können neue, noch nicht in der Theorie beschriebene Merkmale identifiziert werden – ein zuvor deduktives Vorgehen birgt die Gefahr in sich, für diese Merkmale ‚blind' zu werden.

Kategoriensysteme können im Wesentlichen auf einzelne auftretende (respektive unterbleibende) Ereignisse (Ereignisstichprobe, *event sample*) oder zeitbasiert, also auf die Dauer entsprechender Ereignisse (Zeitstichprobe, *time sample*), ausgerichtet werden (Brückmann & Duit, 2014; Döring & Bortz, 2016). Zur Identifikation von Unterrichtsprinzipien in den Aufzeichnungen spielt die Dauer einer Unterrichtshandlung eine untergeordnete Rolle: Wie viel Zeit etwa die Lehrperson auf eine Vergleichbarkeit von Kontroll- und Testansatz im Rahmen wissenschaftsorientierten Arbeitens (Prinzip P 7, Wahrung wissenschaftlicher Strenge) verwendet, ist unter anderem abhängig vom Niveau der Lernenden wie auch dem Ziel der konkreten Unterrichtsstunde. Über eine Zeitstichprobe würde also in diesem Fall kein Mehrwert an Erkenntnis generiert werden (anders beim Instrument zur Erfassung der Teilprozesse, sieht Abschnitt 3.5.2.2).

Zusammenfassende Inhaltsanalyse

Die zusammenfassende Inhaltsanalyse dient dazu, Material derart zu reduzierten, dass allein die wesentlichen Inhalte kondensiert werden, welche im Hinblick auf die Forschungsfrage ein Abbild des Urmaterials darstellen. Bei der zusammenfassenden Inhaltsanalyse ist ein standardisiertes und systematisches Vorgehen nach einem zuvor bestimmten Ablaufmodell notwendig (Mayring, 2015). Hierzu soll im Folgenden das entsprechende Ablaufmodell zusammenfassender Inhaltsanalyse nach Mayring (2015) angewendet werden. Zuerst wird das Verfahren theoretisch beschrieben und anschließend wird das konkrete Vorgehen in dieser Forschungsarbeit dargestellt.

In einem ersten Schritt ist das Material genau festzulegen und zu beschreiben. Nachdem anschließend die Analyseeinheiten bestimmt wurden, werden diese nach

Mayring (2015) in vier Schritten paraphrasiert, generalisiert und reduziert (hier gekürzt dargestellt):

1. Paraphrasierung: Herstellung einer einheitlichen Sprachebene.
2. Generalisierung: Generalisierung auf eine zuvor definierte Abstraktionsebene.
3. Erste Reduktion: Übernahme zentraler, inhaltstragender Paraphrasen.
4. Zweite Reduktion: Zusammenfassung von Paraphrasen mit ähnlichem Gegenstand.

Die genaue Ausgestaltung dieser Schritte kann je nach Erfordernissen der Analyse variiert werden.

Der Prozess der zusammenfassenden Inhaltsanalyse

Für die zusammenfassende Inhaltsanalyse wird als Materialbasis auf Transkripte der Videografien zurückgegriffen. Als Material für die zusammenfassende Inhaltsanalyse dienen 12 Transkriptionen von Unterrichtsvideografien. Der Anzahl von 12 Transkriptionen wurde sich dabei in drei Iterationen bis zur Sättigung (vgl. Akremi, 2014) des vorläufigen Kategoriensystems in den Schritten 8, 10 und 12 angenähert. Dabei wurden, möglichst zu gleichen Teilen, sowohl Transkripte von Videografien der Test- wie auch der Kontrollgruppe herangezogen. Es wurde außerdem bei der Auswahl der Transkripte darauf geachtet, den Testzeitpunkt (pre/post) wiederum möglichst zu gleichen Teilen abzubilden. Zusätzlich sollten sowohl Schulen der ersten wie auch zweiten Fortbildungswelle (siehe Kapitel 3.2) in ähnlichem Umfang Berücksichtigung finden. Bevorzugt wurden Transkripte von Unterrichtsvideografien herangezogen, welche nicht im Rahmen der Studie verwertet wurden. Dabei handelt es sich vornehmlich um Material von Lehrkräften, für die zu einem der beiden korrespondierenden Messzeitpunkte (pre oder post) keine Videografie vorliegt (etwa weil die Lehrkraft aus der Studie ausgestiegen ist) und somit eine solche nicht für die Studie selbst verwendet werden kann. Konkret wurden sieben Transkripte aus der ersten und fünf Transkripte aus der zweiten Welle berücksichtigt. Sieben Transkripte stammen von Lehrkräften, welche an der individualisiert-konstruktiven Fortbildung teilgenommen haben, fünf von Lehrkräften, welche dem kokonstruktiven Format zugewiesen wurden. Diese leichte Divergenz folgt aus der überproportional besetzten Gruppe der Teilnehmer der individualisiert-konstruktiven Fortbildung in der 1. und 2. Welle. Die Divergenz von 8 zu 4 zwischen pre und post hat hingegen forschungspraktische Ursachen: Zum Zeitpunkt der Arbeit am Kategoriensystem waren vornehmlich Videografien des ersten Messzeitpunkts transkribiert (Aufschlüsselung siehe Tabelle 7). Sechs der 12 Transkripte werden im Rahmen der Studie weiterverwertet.

Tabelle 7: Aufschlüsselung der für die zusammenfassende Inhaltsanalyse verwendeten Videos

			ind.-konstr.	kokonstr.
			7	5
pre	8	1.Welle	2	2
		2. Welle	3	1
post	4	1.Welle	1	2
		2. Welle	1	0

Das hier beschriebene Vorgehen lehnt sich an die von Mayring (2015) theoretisch beschriebene Schrittfolge (siehe oben) an, nimmt sich jedoch – ebenfalls im Sinne von Mayring (2015) - die Freiheit einer Anpassung des Vorgehens auf die studienspezifischen Erfordernisse. Hierzu gehört in dieser Studie die Gruppierung von Elementaraussagen zu Sinneinheiten.

Jedes der nun vorliegenden 12 Transkripte wurde im ersten Schritt um (im weitesten Sinne zur Beantwortung der Forschungsfragen) nicht inhaltstragende Bestandteile des Transkripts reduziert. Außerdem wurden die während der Transkription automatisch generierten Zeitstempel herausgenommen. Im zweiten Schritt wurden die nun verbliebenen Elementaraussagen zu Sinneinheiten zusammengefasst. Im dritten Schritt werden diese Sinneinheiten weiter generalisiert und auf eine Abstraktionsebene gebracht (siehe Abbildung 13).

Abbildung 13: Originalmaterial, Paraphrasierung, Gruppierung, Generalisierung. Hinweis: In diesem Beispiel wurde im Schritt der Paraphrasierung kaum gekürzt, da die meisten Aussagen inhaltstragend sind

In dem letzten Schritt wurden die generalisierten Aussagen aller Transkripte induktiv gruppiert und nochmals im Hinblick auf die angestrebte Funktion des späteren Kategoriensystems gefiltert. Dieses Vorgehen (Iteration bis zur Sättigung) lieferte 57 Kategorien für den Aufbau des Kategoriensystems.

Entwicklung des Kategoriensystems

Die Entwicklung des Kategoriensystems vollzog sich in mehreren Schritten und ist in Abbildung 14 im Überblick dargestellt.

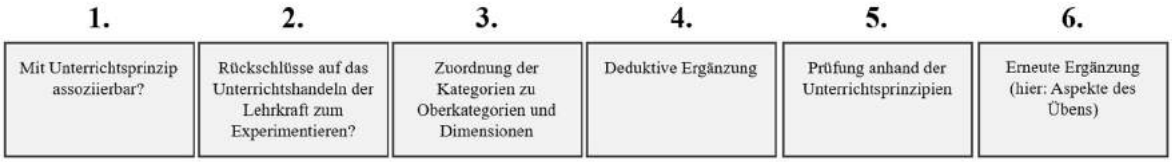


Abbildung 14: Überblick über die Prozessschritte bei der Entwicklung des Kategoriensystems

Die Basis bei der Entwicklung des Kategoriensystems bilden die 57 induktiv durch die zusammenfassende Inhaltsanalyse erhaltenen Kategorien. Diese Kategorien wurden im Zuge einer ersten Inhaltsvalidierung um 13 – vornehmlich dem Klassenmanagement zuzuordnende – Kategorien reduziert. Für den Verbleib der Kategorie im sich entwickelnden Kategoriensystem musste mindestens eines der beiden Kriterien erfüllt werden:

1. Die Kategorie lässt sich mit einem der postulierten Unterrichtsprinzipien assoziieren
2. Die Kategorie dient dazu, dass Rückschlüsse auf das Unterrichtshandeln der Lehrkraft zum Experimentieren gezogen werden können.

Die Kategorien „L beobachtet“, „L gibt Hinweise“, „Hinweise bzgl. Darstellung/Skizze“, „Organisatorisches“, „Arbeitsauftrag erteilen“, „Wechsel Sozialform/Gruppenbildung“, „Materialbeschaffung von L und S“, „Aufforderung zur Aufmerksamkeit“, „Zeitmanagement“, „Aufräumen“, „Unterrichtsstörung“, „Begrüßung“, „Verabschiedung“ erfüllen keines der beiden Kriterien und wurden aus dem Kategoriensystem entfernt.

Anschließend wurden die verbliebenen Kategorien in drei Dimensionen (Prozessstruktur, Prozessreflexion, Angemessenheit und Unterstützung) gegliedert. Hier wäre auch eine Ordnung nach den Unterrichtszielen (Kapitel 2.3) denkbar gewesen. Eine Ordnung anhand der Unterrichtsziele wurde jedoch verworfen, da den Unterrichtszielen die Trennschärfe untereinander fehlt: Einzelne Unterrichtsprinzipien wirken auf mehrere Unterrichtsziele (vgl. Abbildung 1), diese Unschärfe wäre auch für einzelne Kategorien des Kategoriensystems gegeben.

Die Eliminierung und Zuordnung der Kategorien zu den Dimensionen wurde im Prozess von Projektbeteiligten und auf einer Doktorierendentagung (ESERA Summer School 2019) diskursiv begleitet.

Als vermittelnde Ebene wurden Oberkategorien definiert, welche zwei bis fünf Kategorien beinhalten (vgl. Bewersdorff, Baur & Emden, 2020):

- Dimension Prozessstruktur: Die Dimension Prozessstruktur untergliedert sich in die Oberkategorien ‚Problem und Fragestellung‘ (Art und Auftreten), ‚Hypothesen aufstellen und besprechen‘ (Arten und Umgang durch die

Lehrperson), ‚Entwickeln einer Untersuchung', ‚wissenschaftsorientiertes Arbeiten anregen' (Vergleichbarkeit, Generalisierbarkeit, Umsetzen von Kontroll- und Testansatz), sowie ‚Fixieren im Erkenntnisprozess'.

- Dimension Prozessreflexion: Die Dimension Prozessreflexion ist zweigeteilt: Mesoreflexion bezeichnet eine Reflexion im Rahmen der konkret vorgenommenen Untersuchung und wird in drei Oberkategorien (‚Klärung und Zusammenfassung von Vermutungen' ‚Besprechung und Reflexion der Durchführung', ‚Besprechung und Reflexion von Erkenntnissen') erfasst. Metareflexion indessen umfasst Oberkategorien, welche das Unterrichtshandeln zur Reflexion über die Erkenntnismethode des Experimentierens an sich beschreiben: ‚Begriffsklärung ‚Vermutung', ‚Sichtbarmachung der Prozessstruktur', ‚wissenschaftsorientiertes Arbeiten' (Erklärungen zur Variablenkontrollstrategie).
- Dimension Angemessenheit und Unterstützung: Diese Dimension umfasst zwei Oberkategorien. ‚Unterstützte Offenheit' dient der Identifikation von Strukturierungs- und Hilfsangeboten durch die Lehrkraft sowie die den Lernenden im Experimentierprozess gegebene Freiheiten. Die Oberkategorie ‚Lehrkraft als Ansprechpartner, Trainer und Experte' klassifiziert direkte Interaktionen zwischen Lehrkraft und Lernenden.

Anschließend wurde das Kategoriensystem im Hinblick auf seine angestrebte Funktion – die Erfassung von Änderungen des Unterrichtshandelns im Hinblick auf die in Baur et al. (2019) abgeleiteten Unterrichtsprinzipien – punktuell deduktiv ergänzt. Grundlage war hierfür die Zusammenschau existierender Kategorien- und Ratingsysteme, welche in Kapitel 2.6 vorgestellt wurden. Traten in diesen vorgestellten Manualen Kategorien auf, welche bezüglich eines der abgeleiteten Unterrichtsprinzipien einen Beitrag zu leisten vermögen und nicht bereits durch induktiv generierte, inhaltsgleiche Kategorien im Kategoriensystem repräsentiert werden, wurden diese ergänzt.

In einem weiteren Schritt wurde durch inhaltliche Zuordnung der Kategorien zu den Unterrichtsprinzipien überprüft, inwieweit alle aufgestellten Unterrichtsprinzipien mit diesem Kategoriensystem erfasst werden können. Dabei wurde auffällig, dass weder über die induktive Erzeugung noch über die deduktive Ergänzung Aspekte des Übens experimenteller Denk- und Handlungsmuster (Unterrichtsprinzip P3, vgl. Abbildung 1) in dem Kategoriensystem Eingang gefunden haben: In der für die induktive Erzeugung gewählten Stichprobe treten Aspekte des Übens nicht auf. Auch in den für die deduktive Ergänzung herangezogenen Kategoriensystemen (vgl. Tabelle 4) wurden Aspekte des Übens nicht berücksichtigt. Die fehlende Berücksichtigung des Übens beim Vermitteln von Experimentierkompe-

tenz ist bereits in den Unterrichtskonzepten, welche sich für den Aufbau von Experimentierkompetenz empfehlen, auffällig (Bewersdorff et al., 2020). Im Kategoriensystem wurde die Kategorie „Das Experimentieren wird explizit eingeübt" ergänzt, um das Unterrichtsprinzip P3 („Üben") abzubilden.

3.5.2.1 Gütekriterien des Kategoriensystems

Die Qualität der vorliegenden Studie hängt maßgeblich von der Qualität der verwendeten Messinstrumente ab. Zur Überprüfung der Qualität des entwickelten Kategoriensystems wird dieses anhand der drei Gütekriterien Objektivität, Reliabilität und Validität (Döring & Bortz, 2016) geprüft. Im Folgenden werden für die Gütekriterien das jeweilige Vorgehen beschrieben und ggf. ermittelte Kennzahlen berichtet. Auf Basis dieser Befunde wird das Kategoriensystem gegebenenfalls optimiert bzw. angepasst. Die Anpassungen werden jeweils berichtet.

Objektivität

Die Objektivität bzw. Anwenderunabhängigkeit gibt an, ob ein Test (Messinstrument) bzw. dessen Ergebnis von der Person des Testanwenders unabhängig ist. Ein Test oder Messinstrument gilt als objektiv, wenn verschiedene Testanwender bei derselben Testperson zu denselben Testergebnissen kommen (Döring & Bortz, 2016). Bei der Objektivität kann dabei zwischen der Durchführungsobjektivität und der Auswertungsobjektivität unterschieden werden (Tachtsoglou & König, 2017).

Die Durchführungsobjektivität ist dann gewährleistet, wenn die Unabhängigkeit des Testergebnisses von der Person, welche den Test durchführt, gewährleistet ist (Döring & Bortz, 2016). Ist das Vorgehen – speziell bei Videografie – bei der Datenerhebung, -analyse und -auswertung standardisiert (etwa in Form einer präzisen Anleitung) und sind die zu beurteilenden Kategorien klar und eindeutig beschrieben, dann ist das Instrument nicht anfällig für Verzerrungen durch verschiedene Beobachter, sprich, das Instrument ist objektiv (Helmke, 2017). Diese Vorgaben sind vorab genau zu bestimmen und von den Personen, welche mit der Durchführung betraut sind, streng einzuhalten.

Um die Durchführungsobjektivität bei der Erhebung (Videografie) zu sichern, wurde eine detaillierte Anleitung erstellt (siehe Anhang 11.4) und vor der Videografie mit den damit betrauten Personen (im Allgemeinen studentische Hilfskräfte) besprochen, soweit die Aufnahmen nicht selbst durchgeführt wurden. Die Transkription erfolgte stets nach einem Transkriptionsleitfaden (Anhang 11.5).

Die Auswertungsobjektivität ist gewährleistet, wenn das Test- bzw. Messergebnis von der Person, welche den Test auswertet, unabhängig ist (Döring & Bortz, 2016).

Das Vorgehen der Auswertung muss dabei im Test- respektive Kodiermanual oder Kategoriensystem genau vorgegeben sein und von den Anwendern umgesetzt werden. Das Instrument ist standardisiert, also die Kategorien sind klar definiert und ihre Ausprägungen eindeutig beschrieben, sodass nach einem Beobachtertraining alle Beobachtenden das Instrument in gleicher Weise anwenden können (vgl. Anhang 11.1 und 11.2).

Nach Döring und Bortz (2016) kann somit die Objektivität als erfüllt betrachtet werden, wenn für die einzelnen Schritte der Durchführung und Auswertung des Tests stets präzise Angaben gemacht wurden und Hilfestellungen in Form von Anleitungen, Leitfäden und Interpretationsbeispielen gegeben wurden. Dies ist, wie gezeigt werden konnte, für die Aufnahme und Aufbereitung von Daten mit Hilfe des hier vorgestellten Instruments (vgl. Anhang 11.1 und 11.2) der Fall.

Reliabilität

Die Reliabilität (Zuverlässigkeit, Präzision oder Messgenauigkeit) gibt an, wie stark ein Test durch Messfehler verzerrt ist (Döring & Bortz, 2016). Die Reliabilität bei Videostudien hängt dabei von zwei Faktoren ab: dem zugrunde liegenden Messinstrument sowie der Urteilskompetenz der Beobachter (Rater) (Helmke, 2017). Je stärker der Ermessensspielraum bei der Formulierung einer Kategorie, desto stärker streuen auch die Urteile der Beobachter, was wiederum eine verminderte Reliabilität darstellt. Durch eine iterative Entwicklung des Kategoriensystems und Training der Beobachter kann die Reliabilität gesteigert werden (Helmke, 2017).

Die Reliabilität eines Beobachtungsinstruments wird mittels der Beobachterübereinstimmung (Interraterreliabilität oder Urteilerübereinstimmung) empirisch überprüft: Jede einzelne Kategorie im Kategoriensystem wird einer Reliabilitätsanalyse unterzogen. Die ermittelten Reliabilitätskoeffizienten schätzen die Messgenauigkeit der jeweiligen Kategorien (Döring & Bortz, 2016, S. 569). Im Folgenden orientiert sich die Durchführung streng nach dem in Döring & Bortz (2016) beschriebenen Vorgehen, wohl wissend, dass die Interraterreliabilität in Teilen der naturwissenschftdidaktischen Forschung auch als Maß für die Bestimmung der Objektivität herangezogen wird.

Zur Bestimmung der Interraterreliabilität wenden zwei geschulte Beobachter parallel und unabhängig voneinander das Kategoriensystem auf dieselbe Teilmenge der Datenbasis der Videografien an. Als angemessene Teilmenge nennen Brückmann und Duit (2014) 15 bis 20 % der Datenbasis. Mithilfe von Maßen zur Be-

stimmung der Übereinstimmung der Urteile beider Beobachter wird ein Reliabilitätskoeffizient quantifiziert. Die Wahl des Maßes ist wiederum vom Skalenniveau der Kategorien abhängig (Wirtz & Caspar, 2002).

Bei dem vorliegenden Instrument liegt ein nominalskaliertes Kategoriensystem vor. Bei mindestens nominalskalierten Kategoriensystemen werden typischerweise als statistische Reliabilitätsmaße Cohens Kappa und die prozentuale Übereinstimmung gewählt (Wirtz & Caspar, 2002). Die Datenbasis besteht aus 47 Videos, wobei nur 30 Videos für die Studie verwendet werden konnten (siehe Kapitel 3.3), für die Reliabilitätsprüfung wurden 11 Videos herangezogen. Dies entspricht 37 % der in der Studie verwendeten Videos. Der von Brückmann und Duit (2014) geforderte Anteil von 15 bis 20 % der Datenbasis wird auch aufgrund Datenverlusten durch die SARS-CoV-2-Pandemie deutlich übertroffen. Geplant wurde vor den Datenverlusten mit einer Datenbasis der Studie von $N = 48$ Videos, der Anteil der Videos für die Reliabilitätsprüfung wäre mit 11/48 (23%) über dem geforderten Anteil (15–20 %).

Bei den für die Beobachterübereinstimmung ausgewählten Videos wurde auf eine möglichst ausgeglichene Verteilung bzgl. des Messzeitpunktes (prä/post), der Experimentalgruppe (Kontrollgruppe/Testgruppe) geachtet. Die genaue Verteilung findet sich in Tabelle 8:

Tabelle 8: Verteilung der Stichprobe zur Bestimmung der Beobachterübereinstimmung

		prä	post
		6	5
Kontrollgruppe	6	4	2
Testgruppe	5	2	3

Mit diesen 11 von zwei unabhängigen Beobachtern parallel ausgewerteten Videos wurden für alle Kategorien die prozentuale Übereinstimmung und Cohens Kappa (Cohen, 1960) bestimmt. Die prozentuale Übereinstimmung berechnet sich nach Wirtz & Caspar (2002) als

$$p\ddot{U} = \frac{Anzahl\ der\ richtigen\ Entscheidungen\ * 100\ \%}{Anzahl\ der\ Entscheidungen}$$

Cohens Kappa κ ist über folgende Gleichung definiert (Cohen, 1960):

$$\kappa = \frac{p_0 - p_c}{1 - p_c}$$

Dabei ist p_0 der gemessene Übereinstimmungswert der beiden Beobachter und p_c die zufällig erwartete Übereinstimmung. p_0 und p_c können durch die Kreuztabelle beider Schätzer hergeleitet werden. Der theoretische Aufbau einer Kreuztabelle mit zwei Schätzern und zwei Ausprägungen (wie für die Berechnung von Cohens Kappa in dieser Studie verwendet) ist in Tabelle 9 dargestellt.

Tabelle 9: Schematischer Aufbau einer Kreuztabelle mit zwei Schätzern und dichotomer Merkmalsausprägung

		Schätzer 1		Randsumme	
		0	1		
Schätzer 2	0	A	B	A+B	o = (A+B)/N
	1	C	D	C+D	p = (C+D)/N
Randsumme		A+C	B+D	N = A+B+C+D	
		m = (A+C)/N	n = (B+D)/N		

p_0 berechnet sich aus der Summe der Diagonalwerte geteilt durch die Summe aller Entscheidungen:

$$p_0 = \frac{(A + D)}{N}$$

p_c berechnet sich aus den auf die Summe aller Entscheidungen normierten Randhäufigkeiten der Kreuztabelle:

$$p_c = m * n + o * p$$

Der berechnete κ-Wert kann Werte zwischen -1 (gegensätzliche Übereinstimmung) und 1 (perfekte Übereinstimmung) annehmen. Zur Einschätzung der Güte der Beurteilerübereinstimmung liefern sowohl Wirtz & Caspar (2002) als auch Landis und Koch (1977) Interpretationshilfen (siehe Tabelle 10).

Tabelle 10: Wertebereiche und Interpretation von Cohens Kappa nach Wirtz & Caspar (2002) bzw. Landis & Koch (1977)

Wirtz & Caspar (2002)		Landis & Koch (1977)	
Wertebereich	**Interpretation**	**Wertebereich**	**Interpretation**
> .75	gute bis sehr gute Übereinstimmung	> .80	Almost Perfect
.60 – .75	zufriedenstellende Übereinstimmung	.60 – .80	Substantial
.40 – .60	akzeptable Übereinstimmung	.40 – .60	Moderate
< .40	schlechte Übereinstimmung	.20 – .40	Fair
		< .20	Slight

Hierbei wird klar, dass diese Werte eine normative Setzung darstellen. So findet sich bei Altman (1991, ausgerichtet auf medizinische Forschung) wiederum eine leicht abweichende Interpretation: Werte $\kappa > .80$ gelten als „sehr gut“, $\kappa > .60$ noch als gut, zwischen $.60 > \kappa > .40$ als moderat und im Bereich zwischen $.40 > \kappa > .20$ als gering. Im Folgenden wird auf die von Wirtz und Caspar (2002) vorgeschlagenen Wertebereiche Bezug genommen.

Die prozentuale Übereinstimmung sowie Cohens Kappa (κ) wurden mithilfe von erstellten Skripten der freien und auf statistische Berechnungen ausgelegten Programmiersprache R (R Core Team, 2018) berechnet:

Tabelle 11: Beurteilerübereinstimmung für die einzelnen Kategorien, wurde der κ-Wert von 0,75 um mehr als 0,01 unterschritten, wurden die Kategorien hervorgehoben

Kategorie	Name	%Ü	κ
S1	Ein Problem (als Anstoß einer Untersuchung) wird aufgeworfen/identifiziert	.82	.62
S1A	Ein Problem wird aufgeworfen/identifiziert („Ingenieurmodus“)	1	1
S2	Eine Fragestellung wird präsentiert/hergeleitet	1	1
S2A	Eine naturwissenschaftliche Fragestellung wird präsentiert/hergeleitet	.91	.82
S3	Hypothesen, Vermutungen, Ideen werden (im Kontext eines Versuchs) aufgestellt	.91	.62
S3A	Hypothesen werden (im Kontext eines Versuchs) aufgestellt -> ausgerichtet auf Beobachtung	.73	.49
S3B	Hypothesen werden (im Kontext eines Versuchs) aufgestellt -> mit Begründung	.82	.65

S4	Problematischer Hypothesenbegriff von L: Aufforderung zum Raten; nicht wissenschaftlich; Zweck des Versuchs	.91	.74
S5	Begründungen für Hypothesen werden von L eingefordert	1	1
S6	Ideen zur Durchführung werden erfragt/ genannt/gesammelt	.91	.82
S7	Durchführung/Planung wird (vor-) besprochen	.91	.81
S8	Achtung einer Vergleichbarkeit	1	1
S9	Kontrollansatz: Aufforderung zur Umsetzung	1	1
S10	Aufforderung/Hinweis zu protokollieren.	.82	.63
S11	Einschränkungen und Generalisierbarkeit der Beobachtung/Ergebnisse (Vorläufigkeit d. Wissens, Grenzen) werden diskutiert	.91	.62
S12	Erfragen und besprechen von Beobachtungen	.91	0
S13	Festhalten von Beobachtungen/(Mess-)Ergebnissen	.91	.81
S14	Erkenntnisse werden notiert	1	1
S15A	Teilprozesse werden *im Prozess* schriftlich festgehalten	1	1
S15B	Teilprozesse werden *abschließend* schriftlich festgehalten	1	1
R1	Vermutungen/Ideen werden zusammengefasst/wiederholt	.82	.54
R2	Vermutungen werden verifiziert/falsifiziert	1	1
R3	Erfragen von Problemen bei/nach der Durchführung	.91	.79
R4	Besprechung von Verbesserungs-möglichkeiten nach der Durchführung	1	1
R5	Berichten und Besprechen nach Durchführung	.82	.56
R6	Kommentierung/Besprechung von Produkten und Objekten	.91	.62
R7	Erfragen und besprechen von Erkenntnissen	1	1
R8	Begriffsklärung/Problematisierung „Vermutung“	1	1
R9	Wiederholung Teilprozesse	1	1
R10	Bezug einzelner Teilprozesse im Prozess aufeinander durch L	1	1
R11	Klären von Funktion und Ausgestaltung eines Kontrollansatzes	1	n.v.
R12	Das Experimentieren wird explizit eingeübt	1	n.v.
K1	L gibt die Beobachtung vor	.91	1

K2	L greift händisch in die Durchführung ein	.91	.81
K3	Umgang mit/Funktionsweise von Labormaterial	1	.79
K4	Verweis auf Selbsttätigkeit	1	1
K5	Verweis auf Material	.91	n.v.
K6	Verweis auf AB	.91	.62
K7	Verweis auf Hilfen	1	0
K8	Verweis auf andere Gruppen	.82	1
K9	Freiheit bzgl. der Materialauswahl	.91	.62
K10	L gibt Versuchsanleitung, welche kochrezeptartig strukturiert ist	1	.74
K11	L weist auf die Möglichkeit der Inanspruchnahme von Hilfen hin	1	1
K12	L gibt S einen Protokollbogen aus	.91	1

Es wurde, soweit der κ-Wert berechnet werden konnte, für alle bis auf vier Kategorien eine „akzeptable Übereinstimmung“ ($\kappa > .60$) erzielt. Für 29 Kategorien (63 % aller berechenbaren κ-Werte) konnte eine „gute bis sehr gute Übereinstimmung“ ($\kappa > .75$) erzielt werden.

Im Folgenden sollen Kategorien mit κ-Werten < .74 und Kategorien ohne bestimmbaren κ-Wert (n. v.: nicht vorhanden) diskutiert werden. Bei Kategorien, welche in der Stichprobe selten (oder fast immer) auftreten, fällt der κ-Wert aufgrund ‚ungünstiger‘ Gewichtung der Randsummen ab, obwohl die prozentuale Übereinstimmung hoch ist (siehe Wirtz & Caspar, 2002). Tritt eine Kategorie nie (oder immer) auf, kann rein mathematisch (Teilen durch Null) kein κ-Wert bestimmt werden. Kategorien mit einem κ-Wert < .75 bzw. nicht zu bestimmendem κ-Wert, prozentualer Übereinstimmung > 90 % und solche, die nur selten (oder fast immer) kodiert wurden, sind in Tabelle 12 mit der relativen Häufigkeit ihres Kodierens in der Stichprobe zur Beurteilung der Interraterreliabilität aufgeführt.

Tabelle 12: Kategorien mit κ-Wert < 0,74, prozentualer Übereinstimmung > 90 % und rel. Häufigkeit < 15 % oder > 85 %

Kategorie	%Ü	κ	rel. Häufigkeit
S3	.91	.62	19/22 (86 %)
S11	.91	.62	3/22 (14 %)
S12	.91	0	21/22 (95 %)
R6	.91	.62	2/22 (9 %)
R11	1	n.v.	0/22 (0 %)
R12	1	n.v.	0/22 (0 %)
K5	1	n.v.	0/22 (0 %)
K6	.91	.62	3/22 (14 %)

Nach der ersten Bestimmung der Interraterreliabilität wurden zwischen den beiden Beobachtern alle abweichenden Kodierungen nochmals besprochen. Wesentliches Ziel dieses Vorgehens war es, den Einfluss von Unachtsamkeit und Lässlichkeit der Beobachter auf die κ-Werte möglichst zu minimieren. Besonders bei einem umfangreichen Kategoriensystem wie dem hier vorliegenden, welches zudem Kategorien beinhaltet, die sehr selten (oder sehr oft) auftreten, kann es zu schlichtem Übersehen einzelner relevanter Merkmale kommen. Im überkomplexen Geschehen des Schulunterrichts und unter Bezugnahme auf mehrere Quellen wie Videografie, Transkript und auch Fotografien von Unterrichtsmaterial (Hypercoding: Irion, 2010) ist eine Anfälligkeit für Unaufmerksamkeit und Flüchtigkeit beim Kodieren prinzipiell gegeben. Die prozentuale Übereinstimmung sowie κ-Werte für die Beurteilerübereinstimmung nach der Besprechung und ggf. Klärung von Differenzen finden sich in Tabelle 13.

Tabelle 13: Beurteilerübereinstimmung für die einzelnen Kategorien nach erneuter Besprechung abweichender Kodierungen

Kategorie	%Ü	κ
S1	.91	.81
S1A	1	1
S2	1	1
S2A	.91	.82
S3	1	1
S3A	.91	.82
S3B	1	1
S4	1	1
S5	1	1
S6	.91	.82
S7	1	1

S8	1	1
S9	1	1
S10	1	1
S11	1	1
S12	.91	0
S13	.91	.81
S14	1	1
S15A	1	1
S15B	1	1
R1	.91	.74
R2	1	1
R3	.91	.79
R4	1	1
R5	1	1
R6	.91	.62
R7	1	1
R8	1	1
R9	1	1
R10	1	1
R11	1	n.v.
K1	1	1
K2	.91	.81
K3	.91	.79
K4	1	1
K5	1	n.v.
K6	1	1
K7	1	1
K8	1	1
K9	1	1
K10	1	1
K11	1	1
K12	1	1

Alle Kategorien erreichen, soweit zu berechnen, κ-Werte über .60 („zufriedenstellende Übereinstimmung“). Für die große Mehrheit der Kategorien (44 von 47 Kategorien mit berechenbarem κ-Wert; 94 %) kann sogar eine „gute bis sehr gute Übereinstimmung“ ($\kappa > 0{,}75$) ausgewiesen werden. Die prozentuale Übereinstimmung fällt für alle Kategorien nie unter den Wert von 90 %. Der κ-Wert konnte durch diesen Schritt im Mittel über alle Kategorien von .79 auf .92 (prozentuale Übereinstimmung von 93 % auf 98 %) gesteigert werden.

Die hier vorgelegte Reliabilitätsprüfung offenbart zweierlei: Zum einen konnte gezeigt werden, dass durch die Entwicklung eines dichotomen Kategoriensystems (Merkmal tritt auf / Merkmal tritt nicht auf) eine sehr hohe Beobachterübereinstimmung erreicht werden kann. In der vorliegenden Arbeit wurde Hypercoding, also das Heranziehen verschiedener, teils auch widersprüchlicher Quellen (etwa, wenn im Video Handlungen zu sehen sind welche im Transkript nicht aufgezeichnet wurden), zur Kodierung überkomplexer Handlungsabläufe (hier: Unterrichtshandeln der Lehrperson) verwendet. Diese Praxis scheint zum anderen eine zuverlässige Identifikation von relevanten Merkmalen – und damit das Kodieren – zu erschweren.

Validität

Neben der Überprüfung der Reliabilität und der Objektivität ist auch die Überprüfung der Validität eines Beobachtungsinstruments von zentraler Bedeutung (Bortz & Schuster, 2010) und wird als das grundlegendste der drei Hauptgütekriterien wissenschaftlicher Tests angesehen (Schmiemann & Lücken, 2014). Die Validität gibt an, inwieweit die Beobachtungsinstrumente das messen, was gemessen werden soll bzw. was sie zu messen vorgeben. Unterrichtsbeobachtungen mit Videografien wird aufgrund der Natur des Datenmaterials eine hohe Validität zugeschrieben, diese nimmt jedoch umso mehr ab, je spezieller die zu beobachtenden Ereignisse sind (Brückmann & Duit, 2014).

Im Folgenden soll die Inhaltsvalidität des Kategoriensystems durch eine Anbindung der einzelnen Kategorien an die zuvor theoretisch hergeleiteten Unterrichtsprinzipien vorgenommen werden. Da die Kategorien induktiv hergeleitet und deduktiv ergänzt wurden, ist eine Zuordnung zu einem einzelnen Unterrichtsprinzip nicht immer eindeutig und möglicherweise auch diskutabel. Die Unterrichtsprinzipien selbst wurden in Baur et al. (2019) theoretisch hergeleitet und in Kapitel 2.3 weiter fundiert.

Anschließend erfolgt eine weitere Validierung durch ein Expertenrating.

Dimension „Prozessstruktur“

Die Kategorie S1 ‚Ein Problem (als Anstoß einer Untersuchung) wird aufgeworfen/identifiziert‘ ist bei dem Unterrichtsprinzip P4 ‚Problemlösen‘ zu verorten. Allein wenn ein Problem aufgeworfen wird, kann das Experimentieren auch tatsächlich zum Problemlösen eingesetzt werden.

Die Kategorie S1A ‚Ein Problem wird aufgeworfen/identifiziert (Ingenieurmodus)‘ ist ebenfalls dem Unterrichtsprinzip P4 zuzuordnen. Diese Kategorie dient

der Bestimmung, ob das Problem in einer Art Ingenieur-Modus (Suche nach dem besten Resultat) aufgeworfen bzw. identifiziert wurde.

Die Kategorien S2 bis S7 dienen dem Einblick in die Teilprozesse der Fragestellung und Hypothesenbildung und können somit dem Unterrichtsprinzip P1 ‚Teilprozesse' zugeordnet werden.

Die Kategorien S2 ‚Eine Fragestellung wird präsentiert/hergeleitet' und S2A ‚Eine naturwissenschaftliche Fragestellung wird präsentiert/hergeleitet' dienen der Erfassung des Teilprozesses ‚Fragestellung', wobei die Kategorie S2A den Zweck hat zu erfassen, ob die Fragestellung von der Lehrkraft derart gewählt wurde, dass sie überhaupt durch eine naturwissenschaftliche Untersuchung zu beantworten ist.

Die Kategorien S3 bis S5 dienen der Bestimmung, ob und in welcher Art und Weise von der Lehrkraft in der Unterrichtsstunde eine Phase der Hypothesenbildung betrieben wird. Die Kategorie S3 ‚Hypothesen, Vermutungen, Ideen werden (im Kontext eines Versuchs) aufgestellt' soll allgemein erfassen, ob Hypothesen aufgestellt werden. Die Kategorien S3A ‚Hypothesen werden (im Kontext eines Versuchs) aufgestellt -> ausgerichtet auf Beobachtung' und S3B ‚Hypothesen werden (im Kontext eines Versuchs) aufgestellt -> mit Begründung' dienen der genaueren Erfassung, des entsprechenden Hypothesentyps. Unter Hinzunahme der Kategorie S5 ‚Begründungen für Hypothesen werden von L eingefordert' lassen sich damit Rückschlüsse auf das Unterrichtshandeln der Lehrperson im Rahmen der Hypothesenbildung ziehen. Vermittelt die Lehrkraft im Unterricht einen problematischen Hypothesenbegriff, hat dies unmittelbar negative Auswirkungen auf das Unterrichtsprinzip P1 ‚Teilprozesse'. Die Kategorie S5 lässt sich zusätzlich auch aus dem Unterrichtsprinzip des Übens (P3) begründen, wobei das Üben hier nur implizit angesprochen wird.

Die Planung des Experiments ist Grundlage für die sich anschließende Durchführung des Experiments. Das Nennen, Erfragen und Sammeln von Ideen zur Durchführung (S6) sowie das Besprechen der Planung bzw. Vorbesprechen der Durchführung (S7) sind dabei wesentliche Aspekte des Lehrkräftehandelns in dieser Phase. Das Vorbesprechen dient dazu, methodische und inhaltliche Hintergründe mit den Lernenden zu klären, um ein rezeptartiges Abarbeiten zu verhindern (Hammann & Mayer, 2012).

Die Kategorien S8 ‚Achtung einer Vergleichbarkeit', S9 ‚Kontrollansatz: Aufforderung zur Umsetzung' und S10 ‚Aufforderung/Hinweis zu protokollieren' können dem Prinzip P3 ‚wissenschaftliche Strenge' zugeordnet werden. Alle drei Kategorien dienen der Beobachtung von Hinweisen und Aufforderungen der Lehrkraft zu wissenschaftlicher Strenge bei den Lernenden. Die Kategorien S8, S9 und

S10 lassen sich ferner auch noch weiteren Unterrichtsprinzipien zuordnen. Alle drei Kategorien explizieren einen Teilaspekt des Experiments (P2 ‚expliziter Unterrichtsgegenstand') und beinhalten durch den Hinweis der Lehrperson auch das (implizite) Einüben (P3 ‚Üben') der experimentellen Methode.

Die Kategorie S11 ‚Einschränkungen und Generalisierbarkeit der Beobachtung/Ergebnisse werden diskutiert' ist die einzige Kategorie, welche dem Unterrichtsprinzip P6 ‚Grenzen' zuzuordnen ist.

Das Erfragen und Besprechen von Beobachtungen (Kategorie S12) ist dem Unterrichtsprinzip P1 (‚Teilprozesse') zuzuordnen. Wird das Beobachten zwar nicht stets als eigener Teilprozess, sondern als Teil der Durchführung gesehen, ist das Beobachten dennoch ein wesentlicher Schritt im Erkenntnisweg. Das Besprechen von Hintergründen und Zusammenhängen ist in jedem Teilprozess des Experimentierens ein wichtiger Aspekt des Unterrichtshandelns (Hammann & Mayer, 2012).

Zum Prinzip der wissenschaftliche Strenge zähen Nachvollziehbarkeit, Dokumentation und Reproduzierbarkeit. Dies sind zentrale Charakteristika der Naturwissenschaften (McComas et al., 2002 und Kapitel 2.3). Relevant für das Unterrichtshandeln ist zum einen, ob Beobachtungen (S13) und Erkenntnisse (S14) festgehalten werden, zum anderen, *wann* diese schriftlichen Fixierungen (hier auch bezüglich anderer Teilprozesse) auftreten: Erfolgt das schriftliche Festhalten der einzelnen Prozessschritte fortlaufend im Experimentierprozess (S15A) oder als ein nachträgliches Fixieren von Fragestellung, Hypothese und den weiteren Teilprozessen (S15B)? Die Kategorien S15A und S15B können ferner noch den Prinzipien P1 ‚Teilprozesse' sowie P3 ‚Üben' zugeordnet werden: Die Teilprozesse dienen als Rahmen der Verschriftlichungen (P1 ‚Teilprozesse'), die Erkenntnismethode Experiment wird implizit eingeübt (P3 ‚Üben').

Dimension ‚Reflexion'

Ein prozessreflektierendes Experimentieren (Unterrichtsprinzip P5) soll dazu beitragen, dass die Lernenden nicht mechanistisch die einzelnen Prozessschritte durchlaufen, sondern diese auch aufeinander beziehen und hinterfragen (vgl. Kapitel 2.3). Die Kategorien R1 bis R7 können alle diesem Unterrichtsprinzip zugeordnet werden: Die Kategorie R1 ‚Vermutungen werden zusammengefasst/wiederholt' erfasst die Reflexion der Hypothesenbildung, die Kategorie R2 ‚Vermutungen werden verifiziert/falsifiziert' die Reflexion der Ergebnisse hinsichtlich der aufgestellten Hypothese. Überdies haben die beiden Kategorien R1 und R2 auch Bezüge zum Unterrichtsprinzip P1 ‚Teilprozesse'. Die Kategorie R1 erfasst die

Sammlungsphase des Teilprozesses Hypothesenbildung, die Kategorie R2 fokussiert auf den Rückbezug des Teilprozesses Schlussfolgerung auf die Hypothesenbildung.

Die Kategorien R3 ‚Erfragen von Problemen bei/nach der Durchführung', R4 ‚Besprechung von Verbesserungsmöglichkeiten nach der Durchführung', R5 ‚Berichten und Besprechen nach Durchführung' sowie R6 ‚Kommentierung/Besprechung von Produkten und Objekten' dienen der Erfassung von Reflexionen der Durchführungsphase. Hervorzuheben ist die Kategorie R6, welche als Indikator für ein möglicherweise übergeneralisierendes Experimentierverständnis und damit auch Unterrichtshandeln bei den Lehrkräften dient. Diese drei Kategorien dienen auch der Explikation (P2) und dem Üben (P3) der experimentellen Methode.

Durch das ‚Erfragen und Besprechen von Erkenntnissen' (R7) erfolgt eine Reflexion der Schlussfolgerung, somit dient diese Kategorie der Erfassung eines Reflexionsprozesses beim Experimentieren (P5, ‚Prozessreflektierend'). Da mittels dieser Kategorie die Reflexion eines Teilprozesses dokumentiert wird, ist auch eine Zuordnung zum Unterrichtsprinzip P1 (‚Teilprozesse') passend.

Die Kategorien R9 ‚Wiederholung Teilprozesse' und R10 ‚Bezug einzelner Teilprozesse im Prozess aufeinander durch L' sollen die explizite Sichtbarmachung der Prozessstruktur des Experiments durch die Lehrkraft erfassen und können somit dem Prinzip P2 ‚expliziter Unterrichtsgegenstand' zugeordnet werden. Auch die Kategorien R8 ‚Begriffsklärung, Vermutung' und R11 ‚Klären von Funktion und Ausgestaltung eines Kontrollansatzes' dienen der Identifikation von Unterrichtshandlungen der Lehrperson, bei denen diese Aspekte des Experiments explizit mit den Lernenden bespricht. Die Kategorien R8, R9, R10, R11 können ferner auch dem Unterrichtsprinzip P2 ‚expliziter Unterrichtsgegenstand' zugeordnet werden, da diese Aspekte des Experiments durch die Lehrkraft nur explizit erfolgen können.

Dimension ‚Angemessenheit und Unterstützung'

Gibt die Lehrperson die zu erwartende Beobachtung vor (K1) oder greift ohne Not händisch in den Versuchsaufbau ein (K2), sind dies potenziell Entmündigungen der Lernenden im Experimentierprozess und limitieren eine zunehmende Selbstständigkeit (P8) allein auf das korrekte Nachahmen.

Gibt die Lehrperson Hinweise zum eingesetzten Labormaterial (K3), ermöglicht sie damit ein selbstständiges Durchführen des Experiments durch die Lernenden (P8).

Zur Unterstützung einer zunehmenden Selbstständigkeit (P8) ist die Lehrperson gefordert, bei Unterstützungsanfragen vonseiten der Lernenden angemessen zu reagieren. Bei Hilfsersuchen der Lernenden sind also nach Art der erfragten Unterstützung und Niveau differenzierte Handlungen der Lehrperson geboten. Diese differenzierten Lehrerreaktionen sind über die Kategorie K4 ‚Verweis auf Selbsttätigkeit', K5 ‚Verweis auf Material', K6 ‚Verweis auf Arbeitsblatt', K7 ‚Verweis auf Hilfen' und K8 ‚Verweis auf andere Gruppen' zu erfassen.

Gibt die Lehrperson den Lernenden die Freiheit bezüglich ihrer Materialauswahl (K9) oder wird hingegen eine kochrezeptartige Experimentieranleitung vorgelegt (K10), sind dies Merkmale für den Grad der Öffnung des Experimentierprozesses. Wird auf die Inanspruchnahme von Hilfen hingewiesen (K11) oder gibt die Lehrperson einen Protokollbogen aus (K12), sind dies Merkmale einer unterstützten Offenheit.

Alle Kategorien dieser Dimension K1–K12 können somit dem Unterrichtsprinzip P8 ‚zunehmende Selbstständigkeit' zugeordnet werden.

Validierung durch ein Expertenrating

Nach der Inhaltsvalidierung durch die Bindung der Kategorien an die abgeleiteten Unterrichtsprinzipien wird noch eine weitere Inhaltsvalidierung mittels Expertenrating durchgeführt. Ein Expertenrating wird empfohlen, um durch Personen, welche nicht an der Instrumentenentwicklung beteiligt waren, eine externe Perspektive auf die Eignung der einzelnen Kategorien für den vorgegebenen Zweck zu prüfen (Schmiemann & Lücken, 2014). Für das Rating konnten fünf Experten mit Publikationstätigkeit im Themenfeld ‚Experimentieren' gewonnen werden: Als Experten konnten vier Professoren und ein Doktor aus der Naturwissenschaftsdidaktik mit einschlägigen Publikationen zum Experimentieren und Expertise in der Lehrerbildung gewonnen werden. Für jede Kategorie waren die Experten aufgefordert, auf einer 4-stufigen Likert-Skala zu werten, inwieweit diese Kategorie dazu geeignet ist, einen Beitrag zur Beschreibung von Unterrichtshandeln der Lehrperson zum Experimentieren zu leisten.

Die Stufen der Likert-Skala sind:

0 = diese Kategorie kann KEINEN Beitrag leisten,

1 = diese Kategorie kann einen geringen Beitrag leisten,

2 = diese Kategorie kann einen Beitrag leisten,

3 = diese Kategorie kann einen starken Beitrag leisten (vgl. Anhang 11.3).

Bei der Diskussion der Ergebnisse des Expertenratings werden nun nur die Kategorien, welche nach Ansicht von einem Experten keinen Beitrag (Stufe 0) zur Beschreibung von Unterrichtshandeln der Lehrperson beim Experimentieren leisten, betrachtet. Leisten die Kategorien nach Ansicht eines Experten allein einen geringen Beitrag (Stufe 1) zur Beschreibung, so ist die Aufnahme dieser Kategorien in das Kategoriensystem, welches das Unterrichtshandeln *umfassend* beschreiben soll, gerechtfertigt.

Ein vollständiger Konsens unter den Experten ist aufgrund der Subjektivität der Experteneinschätzungen nicht zu erwarten (Döring & Bortz, 2016). Diese teils gegebene Heterogenität spiegelt sich auch bei dem Blick auf den Median der von einem einzelnen Experten als ungeeignet gewerteten Kategorien wider.

Alle Kategorien im Kategoriensystem, welche von mindestens zwei der fünf Experten mit der Stufe 0 bewertet wurden, werden aus dem Kategoriensystem entfernt. Bei sechs Kategorien erkannte jeweils einer der Experten keinen Beitrag zum Unterrichtshandeln (Stufe 0). Es muss folglich keine Kategorie aus dem Kategoriensystem entfernt werden.

Alle Kategorien, bei welchen nach Ansicht *eines* der fünf Experten kein Beitrag zur Erfassung des Unterrichtshandelns der Lehrperson zum Experimentieren geleistet wird, sind in Tabelle 14 aufgeführt. Zusätzlich ist noch der jeweilige Median des Expertenratings angegeben.

Tabelle 14: Kategorien, welche aus Sicht eines der fünf Experten keinen Beitrag zum Erfassen von Unterrichtshandeln von Lehrkräften leisten können

Kategorie	Median des Expertenratings
S4: Problematischer Hypothesenbegriff von L: Aufforderung zum Raten; nicht wissenschaftlich; Zweck des Versuchs	1
R1: Vermutungen/Ideen zusammengefasst/wiederholt	2
K1: L gibt die Beobachtung vor	2
K2: L greift händisch in die Durchführung ein	1
K3: Umgang mit/Funktionsweise von Labormaterial	1
K8: Verweis auf andere Gruppen	2

3.5.2.2 Ableitung des Kategoriensystems „Phasen des Experimentierprozesses" und Diskussion seiner Güte

Ziel des Kategoriensystems „Phasen des Experimentierprozesses" ist es, durch die zeitbasierte Kodierung von Teilprozessen des Experimentierens Rückschlüsse auf Änderungen

- der im Unterricht auftretenden Teilprozesse,
- ihrer (relativen) Dauer,

im Rahmen der Lehrkräftefortbildung zu bekommen. Das hierfür benötigte Kategoriensystem wurde nicht vollständig neu entwickelt, vielmehr konnte auf bereits existierende Vorarbeiten zurückgegriffen werden. Basis der strukturellen Auslegung des Kategoriensystems sind die Arbeiten von Tesch (2005) sowie Seidel (2003). Inhaltlich orientiert sich das Kategoriensystem an den von Mayer & Ziemek (2006, S. 6) beschriebenen Teilprozessen des wissenschaftlichen Erkenntnisprozesses (Problem – Planung – Durchführung – Deutung) und hat damit Ähnlichkeit zu dem von Stiller (2015) beschriebenen Kategoriensystem. Auf die von Mayer & Ziemek (2006) beschriebenen Teilprozesse wurde auch in der Lehrkräftefortbildung Bezug genommen. Es wurde ein Kategoriensystem abgeleitet, dessen Kern die Teilprozesse des deduktiven Experiments bilden, angereichert um allgemeine, nicht speziell auf das Experimentieren bezogene Unterrichtsphasen (siehe Tabelle 15). Strukturelle Eigenschaften des Kategoriensystems sind Exklusivität und Exhaustivität der gewählten Kategorien. Zur Anwendung des Kategoriensystems wurden alle Videos in 15-Sekunden-Intervalle eingeteilt und anschließend das Kategoriensystem auf jedes Intervall angewendet. Das vollständige Kategoriensystem mit exakter Definition der einzelnen Kategorien sowie einer Anleitung zur Anwendung auf Unterrichtsvideografien findet sich in Anhang 11.2.

Tabelle 15: Die Kategorien des Kategoriensystems Nr. 2 „Phasen des Experimentierprozesses“

Kategorie	Code
Organisatorisches	110
Fach- und Methodenwissen	120
Einführung in das Problem (Problemstellung)	130
Fragestellung	141
Hypothesenfindung	142
Planung	143
Durchführung und Beobachtung	144
Auswertung	145
Schlussfolgerung	146
Präsentation und Reflexion von Erkenntnissen	147
Kein Unterricht mit naturwiss. Bezug	200
Restkategorie/nicht zuweisbar	300

Angewendet wurde das Kategoriensystem aufgrund der zugrundeliegenden Zeitstichprobe allein auf die Videografien. Andere Quellen (Transkript, Fotos von Unterrichtsmaterialien) wurden nicht für die Analyse hinzugezogen.

Objektivität

Bezüglich der Durchführungsobjektivität greift im Wesentlichen die identische Argumentation wie bei dem zuvor ausführlich vorgestellten Kategoriensystem Nr. 1 („Unterrichtshandeln von Lehrkräften beim Experimentieren“). Das Instrument ist standardisiert, also die Kategorien sind klar definiert und ihre Ausprägungen eindeutig beschrieben, sodass nach einem Beobachtertraining alle Beobachtenden das Instrument in gleicher Weise anwenden können (vgl. Anhang 11.2).

Nach Döring & Bortz (2016) kann somit auch bei diesem Instrument die Objektivität als erfüllt betrachtet werden, da die einzelnen Schritte der Durchführung und Anwendung stets durch präzise Angaben und Hilfestellungen (Anleitungen, Leitfäden und Interpretationsbeispielen) gestützt wurden.

Validität

Die Validität soll im Folgenden durch Diskussion der Inhaltsvalidität des Kategoriensystems geprüft werden.

Die Strukturierung des Experimentierprozesses ist ein zentrales Merkmal effektiven Unterrichts zum Experimentieren (vgl. Kapitel 2.3, Prinzip „Teilprozesse“). Die Kategorien des Manuals (Codes 130–147) orientieren sich lose an den von Mayer (2007) und Gut-Glanzmann und Mayer (2018) postulierten Teilprozessen. Andere Autoren schlagen ähnliche Strukturierungen des Experimentierprozesses

vor (vgl. Bewersdorff et al., 2020; Emden, 2011). Die Strukturierung nach Mayer (2007) ist anschlussfähig an die aktuelle fachdidaktische Forschung. Die in der Fortbildung verwendete Strukturierung des Experimentierprozesses stützte sich, ergänzt um den Teilprozess der Durchführung (Emden & Baur, 2017), ebenfalls auf die von Mayer vorgeschlagene Teilprozessfolge. Durch den Rückbezug auf die in der Fortbildung verwendeten und fachdidaktisch anschlussfähige Teilprozesse ist eine Inhaltsvalidität gegeben.

Reliabilität

Zur Bestimmung der Reliabilität wurde dieselbe Teilstichprobe von 12 Videos, welche für die Messung der Beurteilerübereinstimmung von Kategoriensystem Nr. 1 verwendet wurden, herangezogen (vgl. Kapitel 3.5.2.1). Erneut wurden die bei nominalskalierten Instrumenten typischerweise verwendeten statistischen Reliabilitätsmaße Cohens Kappa und die prozentuale Übereinstimmung gewählt (Wirtz & Caspar, 2002).

Folgende Tabelle (Tabelle 16) bildet die Kodierentscheidungen beider Beobachter von allen zwölf herangezogenen Videos ab:

Tabelle 16: Kreuztabelle der Kodierentscheidungen von Beobachter 1 und Beobachter 2

		Beobachter 1											
		110	120	130	141	142	143	144	145	146	147	200	300
Beobachter 2	110	319	14	2	1	7	9	11	19	1	4	9	0
	120	7	240	13	3	0	0	0	0	0	1	2	0
	130	8	12	28	8	0	0	0	0	0	0	0	0
	141	0	1	1	24	0	1	0	0	0	0	0	0
	142	4	20	0	1	146	0	0	0	0	0	0	0
	143	8	26	0	0	2	321	10	7	0	0	0	0
	144	8	0	0	0	0	13	464	6	0	0	0	0
	145	12	0	0	0	0	0	12	248	0	0	0	0
	146	9	0	0	0	0	0	0	5	64	2	0	0
	147	4	2	0	0	0	0	0	0	0	172	0	0
	200	0	0	1	0	0	0	0	0	0	0	62	0
	300	0	0	0	0	0	0	0	0	0	0	0	0

Auf den Zellen der Diagonalen finden sich die Summen der Kodierentscheidungen für jeden Teilprozess, bei welchem Beobachter 1 und Beobachter 2 zu gleichen Kodierentscheidungen gekommen sind. In den Kästen hervorgehoben sind die voneinander häufig abweichenden Kodierentscheidungen betreffend die Kategorien 120 (Fach- u. Methodenwissen) und 130 (Problemstellung). Diese Unschärfe wird später diskutiert werden.

Für die einzelnen Kategorien ergeben sich damit folgende Werte für die prozentuale Übereinstimmung sowie Cohens Kappa (κ):

Tabelle 17: Beurteilerübereinstimmung (%Ü, κ) für die Kategorien im Kategoriensystem Nr. 2 „Phasen des Experimentierprozesses“

Code	Kategorie	%Ü	κ
110	Organisatorisches	.94	.79
120	Fach- u. Methodenwissen	.96	.80
130	Problemstellung	.98	.54
141	Fragestellung	.99	.75
142	Hypothesenfindung	.99	.89
143	Planung	.97	.88
144	Durchführung und Beob.	.97	.92
145	Auswertung	.97	.88
146	Schlussfolgerung	.99	.88
147	Präs. und Refl. von Erkenntn.	.99	.96
200	Kein Unterricht	.99	.91
300	Restkategorie	1,00	n.v.

Alle Kategorien, für welche κ-Werte berechnet werden konnten, ausgenommen die Kategorie „Problemstellung“ (Code 130), zeigen nach Wirtz und Caspar (2002) „gute bis sehr gute“ Beobachterübereinstimmung an ($\kappa > .75$). Die Kategorie „Problemstellung“ mit einer prozentualen Beobachterübereinstimmung von 98 % fällt mit einem κ-Wert von 0,54 deutlich ab. In diesem Fall sei zum einen analog zu der Diskussion von Kategoriensystem Nr. 1 darauf verwiesen, dass bei Kategorien, welche in der Stichprobe selten (oder fast immer) auftreten, der κ-Wert aufgrund ‚ungünstiger‘ Gewichtung der Randsummen abfallen kann, obwohl die prozentuale Übereinstimmung hoch ist (siehe Wirtz & Caspar, 2002). Die Kategorie „Problemstellung“ wurde nur in etwa 3 % der Zeitintervalle von einem der Beobachter gewählt. Zum anderen zeigen die Beobachterurteile in Tabelle 16 (im Kasten hervorgehoben) tatsächlich, dass bezüglich der Kategorie „Problemstellung“ und der Kategorie „Fach- u. Methodenwissen“ oft divergierende Ratings auftreten. Dies kann möglicherweise in Teilen darauf zurückgeführt werden, dass in den Unterrichtsstunden oft im Rahmen einer Phase der Vermittlung von Fach- und Methodenwissen auch die Problemstellung für das darauffolgende Experiment bereits angebahnt wird. Diese Doppelfunktion von Wissensaneignung (bzw. -wiederholung) und Identifikation eines Problems kann auch aufgrund struktureller Merkmale des Kategoriensystems (Zeitstichprobe) nur ungenügend aufgelöst werden. Außerdem zeigte das Kategoriensystem bezüglich der beiden Kategorien auch

eine inhaltlich-definitorische Schwäche. Anhand der Gegenüberstellung von Ausschnitten der jeweiligen Kategoriedefinition soll die Problematik verdeutlich werden (Tabelle 18):

Tabelle 18: Gegenüberstellung der Definitionen der Kategorien „Fach- u. Methodenwissen“ sowie „Problemstellung“

Definition „Fach- u. Methodenwissen“	Definition „Problemstellung“
Jedes Unterrichtsgeschehen, welches einen fachwissenschaftlichen oder fachmethodischen Beitrag für das Experimentieren legt oder *den Themenkomplex einführt*. […]	Es wird in die Problemstellung für *das Experiment eingeführt*, das Experiment wird motiviert […]

Die hier offenkundig gewordene fehlende Trennschärfe zwischen der Einführung in den Themenkomplex sowie der Einführung in das konkrete Experiment wurde versucht durch die Überarbeitung der beiden Definitionen herzustellen (siehe Anhang 11.2). Anschließend wurden mit dem überarbeiteten Kategoriensystem Kodierungen mit diesen beiden Kategorien (über die Prüfung der Beobachterübereinstimmung hinaus) annulliert und erneut durch einen Beobachter kodiert.

Für die Kategorie „Restkategorie“ konnte kein κ-Wert bestimmt werden, da kein Zeitintervall von einem der Beobachter im Rahmen der Prüfung der Beobachterübereinstimmung kodiert wurde. Dementsprechend beträgt die prozentuale Übereinstimmung 100 %.

3.5.2.3 Ableitung der Kategoriensysteme „Öffnungsgrade der Teilprozesse beim Experimentieren“ sowie „Definition des Experiments“ und Diskussion der jeweiligen Güte

Ziel des Kategoriensystems 3 „Öffnungsgrade der Teilprozesse beim Experimentieren“ ist es, Rückschlüsse auf Änderungen des jeweiligen Öffnungsgrades im Rahmen der Lehrkräftefortbildung zu bekommen. Dazu wird der jeweilige Offenheitsgrad („geschlossen“, „leicht geöffnet“, „geöffnet“, „offen“) für jeden Teilprozess bestimmt.

Das Ziel des Kategoriensystems 4 „Definition des Experiments“ ist es, durch die Kodierung der einzelnen Aspekte des Experiments Rückschlüsse zu bekommen, ob im Rahmen der Fortbildung bei den Lehrkräften eine Sensibilisierung bezüglich des Experimentierbegriffs stattgefunden hat und ob diese sich im vermittelten Experimentierbegriff niederschlägt.

Die Kategorien sind – im Gegensatz zu den dichotomen Kategorien der Instrumente 1 und 2 – ordinalskaliert. Bei dem Kategoriensystem 3 „Öffnungsgrade der

Teilprozesse beim Experimentieren“ bildet jeder Teilprozess des Experimentierens je eine Kategorie. Diese sind stets 4-stufig Likert-skaliert („geschlossen“, „leicht geöffnet“, „geöffnet“, „offen“) und ergänzt um die Option „nicht anwendbar“. Kategoriensystem 4 „Definition des Experiments“ besteht ebenfalls aus fünf Kategorien, welche die Teilaspekte der Definition des Experiments abdecken. Es ist 3-stufig Likert-skaliert („voll erfüllt“, „zum Teil erfüllt“, „nicht erfüllt“), wiederum ergänzt um die Option „nicht anwendbar“.

Die jeweils verwendeten Kategoriensysteme (siehe Anhang 11.1) wurden weder neu entwickelt, noch adaptiert, sondern direkt aus Fortbildungsmaterialien abgeleitet. Publiziert wurde das Kategoriensystem „Öffnungsgrade der Teilprozesse beim Experimentieren“ als „Öffnungsmatrix“ bei Baur et al. (2017) und Baur und Emden (2020). Das Kategoriensystem „Definition des Experiments“ wurde als Taxonomie zur Nachbesprechung von Unterrichtshospitationen im Rahmen der Lehrkräftefortbildung eingesetzt. Die zugrunde liegende Definition des Experiments findet sich ebenfalls bei Baur et al. (2017).

Validität

Durch die Ableitung der beiden Kategoriensysteme direkt aus Fortbildungsmaterialien ist eine Inhaltsvalidität gegeben, da die einzelnen Kategorien(-systeme) einen unmittelbaren Ausschnitt aus dem Verhaltensbereich darstellen, welcher untersucht werden soll (Moosbrugger & Kelava, 2012). Über die Inhaltsvalidität hinaus wurde außerdem auch die Augenscheinvalidität geprüft und bestätigt. Die Augenscheinvalidität gibt an, inwieweit der Validitätsanspruch eines Instruments vom bloßen Augenschein her einem Laien gerechtfertigt erscheint (Moosbrugger & Kelava, 2012, S. 15).

Objektivität

Bezüglich der Objektivität der beiden Kategoriensysteme 3 und 4 wird dieselbe Argumentation angeführt wie bei den Kategoriensystemen 1 und 2: Die Instrumente sind standardisiert, also die Kategorien klar definiert und ihre Ausprägungen eindeutig beschrieben, sodass nach einem Beobachtertraining mit Feedbackschleifen alle Beobachtenden die Instrumente in gleicher Weise anwenden können.

Reliabilität

Zur Bestimmung der Reliabilität der beiden Kategoriensysteme wurden 12 Videos herangezogen. Zur Bestimmung der Beurteilerübereinstimmung wurde nach dem in Döring und Bortz (2016) ein gewichtetes Cohens Kappa (κ) gewählt und mittels R bestimmt. Die Gewichtung wurde linear gewählt: Kommen beide Beobachter zu

demselben Urteil, wird mit 1,0 gewichtet. Bei jeder Abweichung um einen Skalenwert wurde das Gewicht jeweils um 1/(Anzahl der Merkmalsausprägungen – 1) reduziert (Lineare Gewichtung, vgl. Tabelle 19). Ist die Absicht allein ein möglichst hoher κ-Wert, scheint eine andere, etwa quadratische Gewichtung oder die „Fleiss-Cohen"-Gewichtung ziel-, aber bezüglich der Aussagekraft des κ-Wertes irreführend.

Tabelle 19: Lineare Gewichte bei der Berechnung von Cohens Kappa (κ) bei einer 4-stufigen Likert-Skala

		Beobachter 1			
		geschlossen	leicht geöffnet	geöffnet	offen
Beobachter 2	geschlossen	1	.66	.33	.00
	leicht geöffnet	.66	1	.66	.33
	geöffnet	.33	.66	1	.66
	offen	.00	.33	.66	1

Für die einzelnen Kategorien der beiden Kategoriensysteme ergeben sich damit folgende Werte für das linear gewichtete Cohens Kappa (κ). Außerdem wird, da intuitiv zugänglich, der Wert der prozentualen Übereinstimmung (%Ü) berichtet. Dafür wurde das Verhältnis aus der Anzahl der Urteile, welche beide Beobachter exakt gleich getroffen haben, und der Gesamtzahl der Urteile abzüglich Wertungen von „nicht anwendbar" gebildet.

Tabelle 20: Beurteilerübereinstimmung (%Ü, κ) für die einzelnen Kategorien des Kategoriensystems 3

Kategoriensystem 3 „Öffnungsgrade der Teilprozesse beim Experimentieren"		
Kategorie	%Ü	κ (gewichtet)
Fragestellung	.80	.44
Hypothese	.55	.52
Planung-Experiment	.55	.57
Auswertung	.75	.63
Schlussfolgerung	.67	.38

Tabelle 21: Beurteilerübereinstimmung (%Ü, κ) für die einzelnen Kategorien des Kategoriensystems 4

Kategoriensystem 4 „Definition des Experiments"		
Kategorie	%Ü	κ (gewichtet)
Faktoren konstant	.83	.65
Kontrollansatz	.92	.75
Testansatz	1	n. v.
Wiederholbarkeit	1	n. v.
Gleiche Ergebnisse	1	n. v.

Die Beurteilerübereinstimmung des Kategoriensystems 3 „Öffnungsgrade der Teilprozesse beim Experimentieren" ist durchweg geringer im Vergleich zu den

anderen drei Kategoriensystemen, liegt aber bis auf die Kategorie „Schlussfolgerung“ noch im akzeptablen ($\kappa > 0{,}40$) oder zufriedenstellenden ($\kappa > 60$) Bereich (vgl. Wirtz & Caspar, 2002). Die im Vergleich geringe Beurteilerübereinstimmung kann mit dem Wechsel der Metrik von dichotomen Kategorien in den Kategoriensystemen 1 und 2 zu ordinalskalierten Kategorien in den Kategoriensystemen 3 und 4 zusammenhängen. Bei den Kategorien mit besonders geringen κ-Werten (Fragestellung, $\kappa = 0{,}44$; Schlussfolgerung, $\kappa = 0{,}38$) ist die Datenbasis auf 10 respektive 9 Videos reduziert, da diese Teilprozesse nicht in allen Videos aufgetreten sind: In zwei Videos wurde eine bereits zuvor hergeleitete Fragestellung aufgegriffen, in drei Videos wurde die Schlussfolgerung erst in der nachfolgenden Unterrichtsstunde gezogen. Die prozentuale Übereinstimmung liegt bei 80 % (Fragestellung) beziehungsweise bei 67 % (Schlussfolgerung).

Die Beurteilerübereinstimmung des Kategoriensystems 4 „Definition des Experiments“ liegt – soweit berechenbar – im zufriedenstellenden bzw. guten Bereich (Wirtz & Caspar, 2002). Die prozentuale Übereinstimmung ist stets > 80 %. Für einige Kategorien konnte kein κ-Wert bestimmt werden, hier beträgt die prozentuale Übereinstimmung jeweils 100 % (beide Beobachter kamen stets zu demselben Urteil).

3.6 Beschreibung des Vorgehens bei der Clusteranalyse

Die Kategorien des Kategoriensystems zur Erfassung von Unterrichtshandeln bei Lehrkräften beim Experimentieren sind stets dichotomer Ausprägung (Merkmal tritt auf / Merkmal tritt nicht auf). Somit ist das Gesamtrating (oder Ratings einzelner Teilaspekte des Kategoriensystems) einer Unterrichtsstunde als ein binärer Vektor darstellbar. Eine Clusteranalyse, sprich der Vergleich der Ähnlichkeit dieser binären Vektoren, also von Ratings der Unterrichtstunden, lässt sich mittels binärer Ähnlichkeitsmaße (Seung-Seok Choi, Sung-Hyuk Cha & Charles C. Tappert, 2010) herstellen. Dadurch wird ein vertiefter Einblick über deskriptive Statistik hinaus in die Struktur der Ratings – und somit mittelbar in die Struktur des Videomaterials und schließlich der videografierten Unterrichtsstunden – möglich. Im Folgenden wird aus Gründen der Lesbarkeit von Ähnlichkeiten der videografierten Unterrichtsstunden bzw. einzelner Teilaspekte gesprochen. Basis dieser Aussagen ist aber stets die Ähnlichkeit der Rating-Vektoren.

Zur Bestimmung der binären Ähnlichkeit eines Vektors gibt es eine Vielzahl unterschiedlicher Ähnlichkeitsmaße. Allen Ähnlichkeitsmaßen ist gemeinsam, dass sie auf der Kontingenztabelle der beiden zu vergleichenden binären Vektoren aufbauen.

Tabelle 22: Schematischer Aufbau der Kontigenztabelle für die binären Vektoren $\vec{a}$ und $\vec{b}$

		Vektor $\vec{b}$	
		1	0
Vektor $\vec{a}$	1	a	b
	0	c	d

Das Auftreten eines Merkmals soll im Vergleich zu dem Nichtauftreten eines Merkmals nicht höher gewichtet werden. Darum wurde bei der Auswahl des Ähnlichkeitsmaßes darauf geachtet, dass sowohl das gemeinsame Auftreten eines Merkmals in beiden zu vergleichenden Unterrichtsstunden (*a*) als auch das gemeinsame Nichtauftreten eines Merkmals in beiden Unterrichtsstunden (*d*) gleich gewichtet wird. Es bieten sich unter anderem folgende Ähnlichkeitsmaße an (Seung-Seok Choi et al., 2010):

- Sokal & Michener $s_1 = \frac{a+d}{a+b+c+d}$
- Rogers & Tanimoto $s_2 = \frac{a+d}{a+2(b+c)+d}$
- Hamann Koeffizient $s_3 = \frac{a-(b+c)+d}{a+b+c+d}$

Für die Clusteranalyse wird das Ähnlichkeitsmaß von Sokal & Michener, auch als *Simple Matching Coefficient* bezeichnet, verwendet. Das Ähnlichkeitsmaß von Rogers & Tanimoto sowie der Hamann Koeffizient liefern bei der Anwendung auf die Ratings ähnliche, teils sogar identische Ergebnisse. Das Ähnlichkeitsmaß nimmt stets Werte zwischen 0 (keine Ähnlichkeit) und 1 (Gleichheit der Vektoren) an.

Ist das Ähnlichkeitsmaß berechnet, so wird anhand eines Clustering-Algorithmus (Fusionierungsalgorithmus) die eigentliche Gruppierung der Daten vorgenommen. Hierzu wurde in einem ersten Schritt auf eine hierarchische Clusteranalyse mittels Single-Linkage-Kriterium zurückgegriffen (vgl. Leonhart, 2017). Traten ‚Ausreißer'-Unterrichtsstunden auf, also Unterrichtstunden, welche bei dieser ersten Clusterung keinem Cluster zugeordnet werden konnten (bzw. diese Unterrichtstunden bildeten jeweils als Cluster eine einelementige Menge), wurden diese Stunden von der weiteren Clusteranalyse ausgeschlossen.

In einem zweiten Schritt wurde die Anzahl der optimalen Cluster bestimmt. Das Identifizieren einer „optimalen" Anzahl von Clustern ist für die Stabilität und damit auch die Qualität der Ergebnisse grundlegend. Meist wird zur Bestimmung der „optimalen" Clusteranzahl auf das sogenannte „Elbow-Kriterium" zurückgegriffen (siehe Backhaus, Erichson, Plinke & Weiber, 2016). Es gibt aber noch eine Reihe weiterer probater Kriterien, etwa das Silhouette-Kriterium (Rousseeuw,

1987). Die optimale Anzahl an Clustern lässt sich jedoch nicht allein mathematisch bestimmen, sondern ist auch abhängig von inhaltlichen Überlegungen. Darum sind die ermittelten Clusteranzahlen mehr als Hinweisgeber denn als unumstößliche Festlegung zu interpretieren (Backhaus et al., 2016).
Führte das Elbow-Kriterium in dieser Analyse nicht zu einer inhaltlich klar zu interpretierenden Anzahl von Clustern oder zu keiner eindeutigen Clusteranzahl, wurde auf den Silhouette-Algorithmus (Rousseeuw, 1987) zurückgegriffen.
Im dritten Schritt wurde dann die Clusteranalyse erneut mit der festgelegten Anzahl der Cluster durchgeführt.

4 Ergebnisse

In diesem Kapitel werden, nach einer kurzen Darstellung der Datengrundlage, die Ergebnisse aus der Auswertung der Daten der vier Kodier- bzw. Ratingmanuale „Unterrichtshandeln von Lehrkräften beim Experimentieren“ (4.1), „Phasen des Experimentierprozesses“ (4.2), „Öffnungsgrade der Teilprozesse beim Experimentieren“ (4.3) und „Definition des Experiments“ (4.4) berichtet. Abschließend findet eine Zusammenfassung der Ergebnisse über die einzelnen Teilaspekte hinweg statt (4.5).

Folgende Tabelle gibt einen Überblick über das jeweilige Thema der Unterrichtsstunde, ihre Dauer sowie den Messzeitpunkt im Verlauf der Studie (pre/post). Die jeweils aufeinanderfolgenden Pre/Post-Paare, farblich alternierend hervorgehoben, stellen die beiden Videografien jeweils einer Lehrkraft dar. Tabelle 23 beinhaltet die 18 Videografien ($n = 9$) der individualisiert-konstruktiv betreuten Fortbildungsgruppen (Kontrollgruppe). In Tabelle 24 sind die 12 Videografien ($n = 6$) der kokonstruktiv betreuten Fortbildungsgruppen (Treatmentgruppe) aufgelistet. Die Unausgeglichenheit bezüglich der Mächtigkeit beider Gruppen ist auf den durch Corvid-19 bedingten Ausfall der 3. Fortbildungswelle zurückzuführen. Ohne diesen Ausfall wäre das Verhältnis der Videografien 18 (Kontrollgruppe) zu 20 (Experimentalgruppe). Das Wiederholen der abgebrochenen 3. Welle im darauffolgenden Schuljahr war aufgrund der anhaltenden Pandemie nicht mehr im Rahmen der Projektlaufzeit möglich.

Individualisiert-konstruktiv:

Tabelle 23: Videografierte Unterrichtsstunden im individualisiert-konstruktiven Format

Schule	Thema der Unterrichtsstunde	Dauer	Messzeit-punkt
1	Wir bringen Sand zum Schwimmen	00:48:24	post
1	Wann löst sich Zucker am schnellsten auf?	00:44:34	pre
1	Warum wächst die Pflanze schräg?	00:47:02	post
1	Was hält die Tiere im Winter warm?	00:46:00	pre
1	Wärmeleitfähigkeit	00:32:54	post
1	Ist Sauerstoff für die Verbrennung notwendig?	00:23:08	pre
2	Verbrennung von Nichtmetallen	00:57:06	post
2	Was alles braucht ein Feuer, damit es brennt?	00:55:33	pre
3	Wie können Vögel fliegen? (Thermik)	00:43:06	post
3	Warum werden Blätter im Herbst bunt? (Chro-matografie)	00:48:40	pre
3	Sauerstoffproduktion der Wasserpest in Abhän-gigkeit von Licht	00:45:43	post
3	Warum können Katzen in der Dunkelheit erfolg-reich jagen?	00:41:48	pre
3	Eigenschaften von CO_2	00:44:22	post
3	Herstellen von Speisesalz	01:23:56	pre
3	Eigenschaften des Lichts (Optik)	00:45:37	post
3	Phänomene des Lichts (Optik)	00:46:37	pre
4	Wohin verschwindet die Kerze? (Verbrennungs-produkte)	00:49:37	post
4	Glasbearbeitung mit dem Bunsenbrenner	00:42:58	pre
	Gesamtdauer	14:07:05	

Kokonstruktiv:

Tabelle 24: Videografierte Unterrichtsstunden im kokonstruktiven Format

Schule	Thema der Unterrichtsstunde	Dauer	Messzeit-punkt
5	Der elektrische Widerstand	01:03:48	post
5	Was leitet elektrischen Strom?	01:18:00	pre
5	Wie entsteht Nebel?	00:50:27	post
5	Wie wirkt sich Abholzung auf das Klima im Regenwald aus?	01:25:17	pre
5	Schallphänomene	00:38:07	post
5	Löslichkeit von Stoffen in unterschiedlichen Flüssigkeiten	01:07:33	pre
6	Ablenkungen beim Autofahren	00:37:55	post
6	Schallphänomene	01:19:45	pre
6	Phänomene aus ‚Star Wars'	00:41:00	post
6	Brennwert von Nüssen	01:19:27	pre
7	Ausdehnung von Luft in Abhängigkeit der Temperatur	01:16:43	post
7	Zeitverschiebung	00:44:23	pre
	Gesamtdauer	12:22:25	

Die Gesamtdauer aller Videografien der Kontrollgruppe beträgt 14:07:05 h, die durchschnittliche Länge der Videografien 00:47:04 h (Experimentalgruppe: Gesamtdauer: 12:22:25, Durchschnitt: 01:01:52).

4.1 Unterrichtshandeln von Lehrkräften beim Experimentieren

In diesem Unterkapitel werden die Daten, welche durch das Kategoriensystem „Unterrichtshandeln von Lehrkräften beim Experimentieren" gewonnen werden, mittels unterschiedlicher Verfahren ausgewertet. Durch eine Clusteranalyse (4.1.1) werden qualitative, intraindividuelle Entwicklungen nachgezeichnet. In Abschnitt 4.1.2 werden ebenfalls auf der Ebene der einzelnen Lehrkräfte Änderungen des Unterrichtshandelns qualitativ beschrieben. Eine quantitative Vergleichsanalyse der Teilstrichproben (interindividuell) bildet den dritten Blickwinkel auf diese Daten (4.1.3).

4.1.1 Ergebnisse der Clusteranalyse

Eine Clusteranalyse wurde sowohl für die jeweiligen Ratings des gesamten Kategoriensystems als auch für einzelne Dimensionen bzw. Oberkategorien des Kategoriensystems durchgeführt. Einzelne Kategorien wurden zuvor ‚umgepolt'. Im Folgenden sollen die Ergebnisse der einzelnen Clusteranalysen vornehmlich anhand von Abbildungen vorgestellt werden. Liefern die Clusteranalysen keine inhaltlich sinnvolle Interpretationsgrundlage, werden diese Analysen im Folgenden

nicht berichtet. Alle nicht berichteten Clusteranalysen finden sich (nicht aufbereitet) in Anhang 11.6.
Die Clusteranalysen wurden in Abbildungen überführt, welche stets noch mit weiteren Informationen angereichert wurden. Folgende Informationen finden sich in den Abbildungen:

1. Die identifizierten Cluster mit ihren jeweiligen Unterrichtsstunden inklusive einer Kurzbezeichung.
2. Gegebenenfalls zuvor aufgrund des Single-Linkage-Kriteriums ausgeschlossene ‚Ausreißer'-Unterrichtstunden. Hiermit sind Unterrichtsstunden gemeint, welche vom Algorithmus keinem Cluster zugeordnet werden können.
3. Eine Entwicklungsachse, stets von links nach rechts: Bei Clustern (oder einzelnen ‚Ausreißer'-Unterrichtsstunden) auf der linken Seite der Abbildung besteht im Allgemeinen noch Entwicklungsbedarf des Unterrichtshandelns hinsichtlich der untersuchten Dimension bzw. Oberkategorie. Je weiter rechts sich die Cluster bzw. die einzelnen ‚Ausreißer'-Unterrichtsstunden befinden, als desto fortgeschrittener kann das Unterrichtshandeln angesehen werden. Operationalisiert ist diese Achse durch den Anteil der erfüllten Kategorien bezogen auf alle in die jeweilige Clusteranalyse einbezogenen Kategorien. Innerhalb der Cluster wird nicht anhand der Entwicklungsachse differenziert. Die Entwicklungsachse gibt nur eine grobe Orientierung, da durch die Clusteranalyse teils verschiedene vergleichbare Entwicklungsstufen zusammengefasst werden. Aus Gründen der Transparenz sind bei Analysen mit fünf oder weniger Kategorien die Binärvektoren der Cluster sowie der ‚Ausreißer' angegeben.
4. Die pre-Unterrichtstunden sind hell eingefärbt, die post-Unterrichtstunden dunkel. Jeweils die pre- und die post-Unterrichtsstunden einer Lehrkraft werden in einer Zeile abgebildet. Findet zwischen der pre- und der post-Unterrichtsstunde einer Lehrkraft eine Entwicklung entlang der Entwicklungsachse statt, wird dies durch eng gepunktete Pfeile (positive Entwicklung) bzw. weit gepunktete Pfeile (nicht erwünschte Entwicklung) dargestellt.

5. Um Unterrichtstunden den beiden Experimentalbedingungen (individualisiert-konstruktiv vs. kokonstruktiv) zuordnen zu können, wurden die Unterrichtstunden folgendermaßen über verschiedene Indizes durchnummeriert:

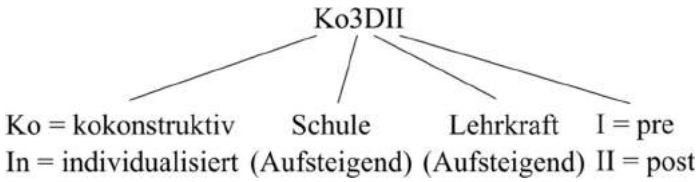


Abbildung 15: Kodierung der Lehrkräfte bzw. Unterrichtsstunden

Der Index Ko3DII steht somit für die post-Videografie (*II*) der Lehrkraft *D* an der Schule *3* aus der kokonstruktiven Teilstichprobe (*Ko*).

Clusteranalyse mit allen Kategorien (gesamtes Kategoriensystem)

Mit dem Single-Linkage-Kriterium wurden keine Ausreißer identifiziert, welche für die weitere Clusteranalyse ausgeschlossen werden sollten. Da das ‚Elbow'-Kriterium keine eindeutig interpretierbare Anzahl von Clustern liefert, wurde anschließend der Silhouette-Algorithmus berechnet. Dieser weist auf 2 oder 6 Cluster hin. Nur die Einteilung in sechs Cluster lässt sich auch inhaltlich sinnvoll interpretieren, weswegen diese Clusterung gewählt wurde. Trotz zuvor – erfolglos – durchgeführter Suche nach Ausreißern finden sich bei der nun durchgeführten Clusteranalyse zwei Cluster mit jeweils nur einem Eintrag. Das Resultat der Clusteranalyse ist, ergänzt um weitere Informationen, in Abbildung 16 dargestellt.

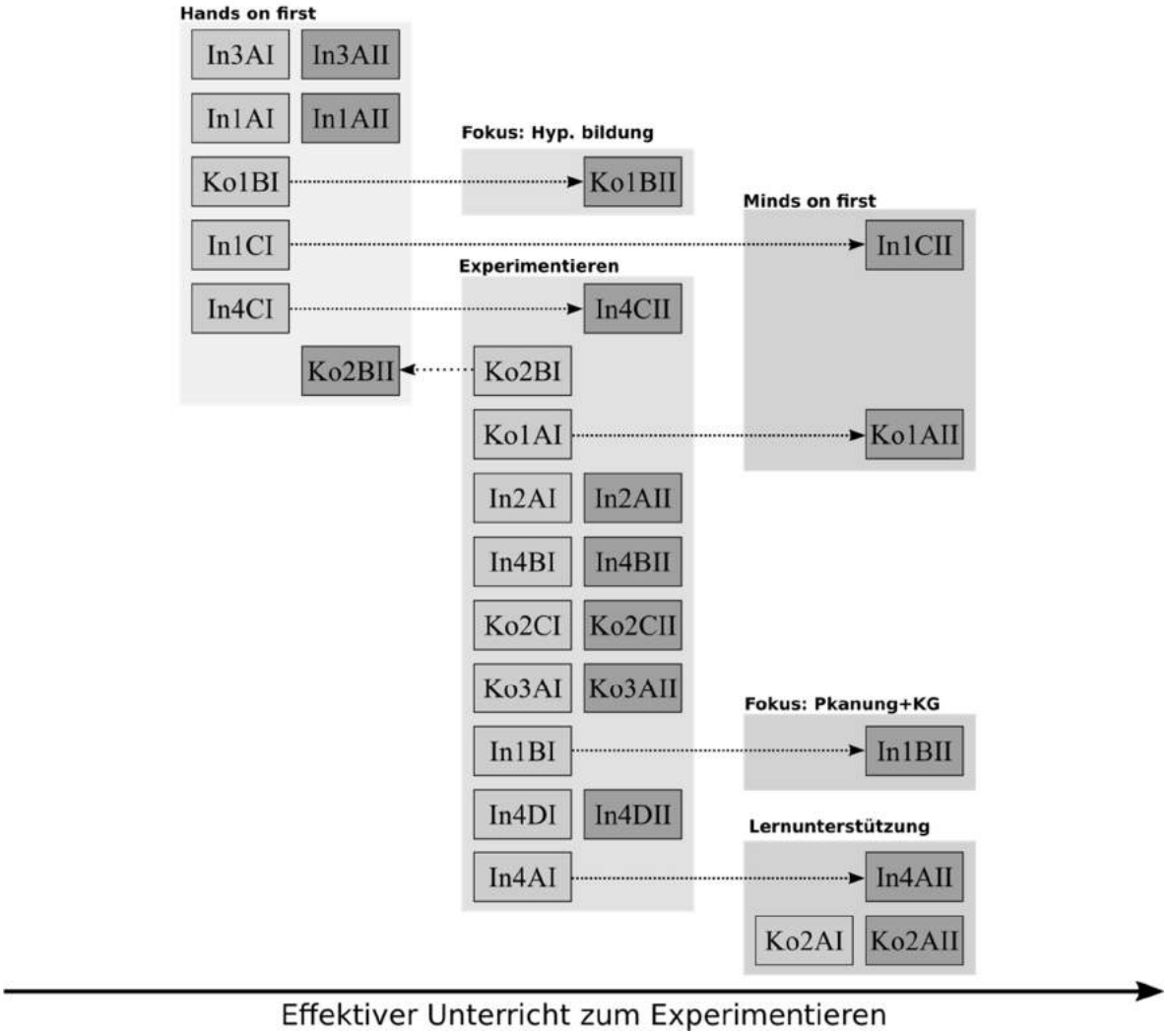


Abbildung 16: Clusteranalyse „Effektiver Unterricht zum Experimentieren“

Die global über alle Dimensionen des Experimentierens gefundenen Cluster können inhaltlich interpretiert werden:
Das Cluster „Hands on first“ beinhaltet Unterrichtsstunden, bei welchen die Durchführung von chemischen Versuchen (‚Brennwert von Nüssen‘, ‚Wohin verschwindet die Kerze? (Verbrennungsprodukte)‘, ‚Löslichkeit von Stoffen in unterschiedlichen Flüssigkeiten‘, ‚Lösen von Zucker‘) oder technischen Prüfungen und Verfahren im Fokus stehen (‚Wir bringen Sand zum Schwimmen‘, ‚Glasbearbeitung mit dem Bunsenbrenner‘, ‚Schallphänomene‘, ‚Herstellen von Speisesalz‘).
Experimente, welche sich an der ‚klassischen‘ Prozessfolge orientieren, finden sich im Cluster „Experimentieren“. Eine besondere Betonung der Epistemologie sowie Reflexion der Erkenntnismethode findet allenfalls rudimentär statt.
Im Cluster „Fokus: Hypothesenbildung“ findet sich eine Unterrichtstunde, welche sich auf die Hypothesenbildung und anschließende Planung – inklusive der epistemologischen Zusammenhänge und Hintergründe – fokussiert.

Das Cluster „Minds on first“ beinhaltet zwei Unterrichtstunden, in welchen die Lehrkraft den Lernenden explizit Zeit gewährt, sich mit der Planung der Durchführung zu beschäftigen. Auf die Implementation von sowohl Test- als auch Kontrollansatz wird hingewiesen, Teilaspekte des Experimentierens werden besprochen.
Ähnlich wie bei dem Cluster „Minds on first“ wird in dem Cluster „Fokus: Planung, KG“ von der Lehrkraft der Fokus auf die Planung des Experiments gelegt und die Funktion von Test- und Kontrollansatz beispielgebunden besprochen und reflektiert. Der Unterschied zum Cluster „Minds on first“ besteht darin, dass der Fokus der Unterrichtsstunde vornehmlich auf der Planung liegt, eine Durchführung, Auswertung und Schlussfolgerung findet keinen Raum.
Die Unterrichtsstunden des Clusters „Lernunterstützung“ zeichnen sich durch einen ausgedehnten Einsatz von Hilfen zur Steuerung (etwa Protokollbogen) und Entlastung der Lernenden beim Experimentieren aus. Im Vergleich zu den anderen ausgewerteten Unterrichtstunden wird in diesen Unterrichtsstunden ein besonderer Wert auf umfassende Lernunterstützung gelegt.
Wie Abbildung 16 zeigt, bleibt global über alle Aspekte des Experimentierens das Unterrichtshandeln etwa bei der Hälfte der Lehrkräfte – auf globaler Ebene – über die Fortbildung hinweg weitgehend konstant: Zwei Lehrkräfte verharren im Cluster „Hands on first“, bei fünf Lehrkräften wird sowohl vor als auch nach der Fortbildung grundständig experimentiert. Sechs der 15 Lehrkräfte akzentuieren ihr Unterrichtshandeln: Sie setzen in ihren Unterrichtsstunden bewusst Foki zur Hypothesenbildung bzw. Planung und besprechen (teils reflektieren) dabei epistemologische Aspekte des Experiments. Eine Lehrkraft wechselt aus dem Cluster „Experimentieren“ in das Cluster „Hands on first“.
Da bei dieser ersten Clusteranalyse alle Kategorien (der ‚Globalvektor‘) und somit unterschiedliche Aspekte berücksichtigt wurden ist die inhaltliche Interpretation der Cluster besonders unsicher.
Zur Illustration der Änderungen im Unterrichtshandeln sollen die pre- sowie post-Unterrichtstunden von zwei Lehrpersonen (In1C und In1A) näher betrachtet werden.
Die Lehrkraft In1C lässt in der pre-Unterrichtsstunde (Code In1CI) die Lernenden nachweisen, dass Sauerstoff für die Verbrennung notwendig ist. Anhand einer von der Lehrperson beschriebenen Beobachtung (Feuer lodert auf, wenn man die Kamintür öffnet) werden die Lernenden nach kurzer Diskussion möglicher Einflussfaktoren aufgefordert, zu zeigen, dass Sauerstoff für eine Verbrennung notwendig ist. Die Lernenden bekommen zügig (nach Minute 03:56) Materialien und beginnen sogleich mit Versuchen. Die Lehrperson betreut während dieser Phase die Gruppen. Nach der Durchführung werden einzelne Gruppen aufgefordert, ihr Vorgehen zu beschreiben (L: „Erklärt einfach, was ihr gemacht habt“, Z. 251/252).

Anschließend werden die Materialien aufgeräumt und die Unterrichtstunde endet. Diese Unterrichtsstunde ist dem Cluster „Hands on first“ zugeordnet.
In der post-Unterrichtstunde (In1CII) wird die Wärmeleitfähigkeit von unterschiedlichen Materialien untersucht: Die Lehrperson präsentiert das Bild einer Holzsitzbank und einer Sitzbank aus Metall im Sommer, zahlreiche Hypothesen auf die Frage, auf welcher Bank man angenehmer sitzt, werden genannt und an der Tafel aufgeschrieben (z. B. S19: „Metall nimmt Wärme schneller an. Also, es wird schneller warm“, Z. 98). Anschließend fordert die Lehrkraft die Lernenden auf, ihre Hypothesen mit bereitgestellten Materialien zu prüfen. Vor der Durchführung planen die Lernenden ihre Versuchsansätze und bauen sie auf, die Lehrperson gibt individuell die Durchführung frei. Die Notwendigkeit von Kontroll- und Testansatz sowie die Vergleichbarkeit der beiden Ansätze wird besprochen (L: „Auf was muss man denn achten […] damit man es auch vergleichen kann?“, Z. 893/894). Bei der Schlussfolgerung wird von der Lehrkraft nochmals Bezug auf die zuvor aufgestellten Hypothesen genommen. Diese Unterrichtsstunde ist dem Cluster „Minds on first“ zugeordnet.
Es kann somit eine Entwicklung des Unterrichtshandelns beobachtet werden: Steht in der pre-Unterrichtstunde der händische Aspekt des Experimentierens im Vordergrund – und wird das Experiment im didaktischen Kontext als ‚Versuche durchführen‘ verstanden – so wird in der post-Erhebung ein differenziertes Verständnis des Experiments vermittelt: Die Hypothesenbildung und die Planung werden gestärkt sowie einzelne Teilaspekte des Experiments (hier: Test- und Kontrollansatz sowie die Vergleichbarkeit der Ansätze) explizit vermittelt.
In der pre-Unterrichtsstunde (In1AI) der Lehrkraft In1A wird anhand einer detaillierten Anleitung geprüft, welche Stoffe sich in Wasser lösen (L: „Da gibt es vier Dinge, die ihr nachher macht und ausprobiert, wann löst es sich schneller auf? Die Anleitung dazu lest ihr mal bitte auf Seite 3“, Z. 133/134). Eine Hypothesenbildung oder anderweitige Orientierung an den Teilprozessen bzw. Teilaspekten des Experiments findet nicht statt. Nach der Durchführung werden die Beobachtungen miteinander verglichen, bevor die Unterrichtsstunde endet.
In der post-Erhebung (In1AII) lässt die Lehrperson die Lernenden aus Materialien wie Alufolie und Milchkartons Boote bauen, welche Sand tragen können sollen („S5 liest das gemeinsame Ziel, das auf der Tafel steht, laut vor. S5: Wir bringen Sand zum Schwimmen“, Z. 38–40). Nach dem Bau der Boote werden diese gemeinsam auf ihre Schwimmfähigkeit geprüft und nach Kriterien (Aussehen, Tragfähigkeit) beurteilt.
Die Lehrkraft zeigt also über die Fortbildung hinweg keine wesentlichen Fortschritte des Unterrichtshandelns in Bezug auf das Experimentieren: In beiden Unterrichtsstunden wird Experimentieren als ein händisches Ausprobieren und Explorieren verstanden. Eine Orientierung an den Teilprozessen des Experiments fin-

det nicht statt. Charakteristika des Experimentierens (Variablenkontrolle, wissenschaftliche Strenge, Grenzen) werden nicht angesprochen. Eine Reflexion der Methode findet jeweils nicht statt.
Zusammenfassung der Clusteranalyse „Effektiver Unterricht zum Experimentieren“: Unterrichtshandeln ist robust. Mehr als die Hälfte der Lehrkräfte bleiben über die Fortbildung hinweg demselben Cluster zugeordnet. Bei sechs Lehrkräften findet eine Akzentuierung des Unterrichtshandelns hin zu einem potenziell effektiveren Unterricht zum Experimentieren statt.

Clusteranalyse „Hypothesenbildung“

Für die Clusteranalyse wurden die Kategorien S3, S3A, S3B, S4 sowie S5 des Kategoriensystems herangezogen. Die Kategorien S3A sowie S4 wurden ‚umgepolt‘, sprich das Nicht-Auftreten dieser Kategorien wird durch eine ‚1‘ repräsentiert, das Auftreten mit einer ‚0‘. Damit wird eine sinnvolle Gruppierung entlang der Entwicklungsachse ermöglicht.
Mit dem Single-Linkage-Kriterium wurden fünf ‚Ausreißer‘-Unterrichtsstunden identifiziert. Diese wurden in der nun folgenden erneuten Clusteranalyse nicht berücksichtigt, jedoch in der Abbildung aufgenommen (vgl. Abbildung 17). Über das ‚Elbow‘-Kriterium wurde die optimale Anzahl an Clustern zu 4 bestimmt. Das Resultat der Clusteranalyse ist, ergänzt um weitere Informationen, in Abbildung 17 dargestellt.

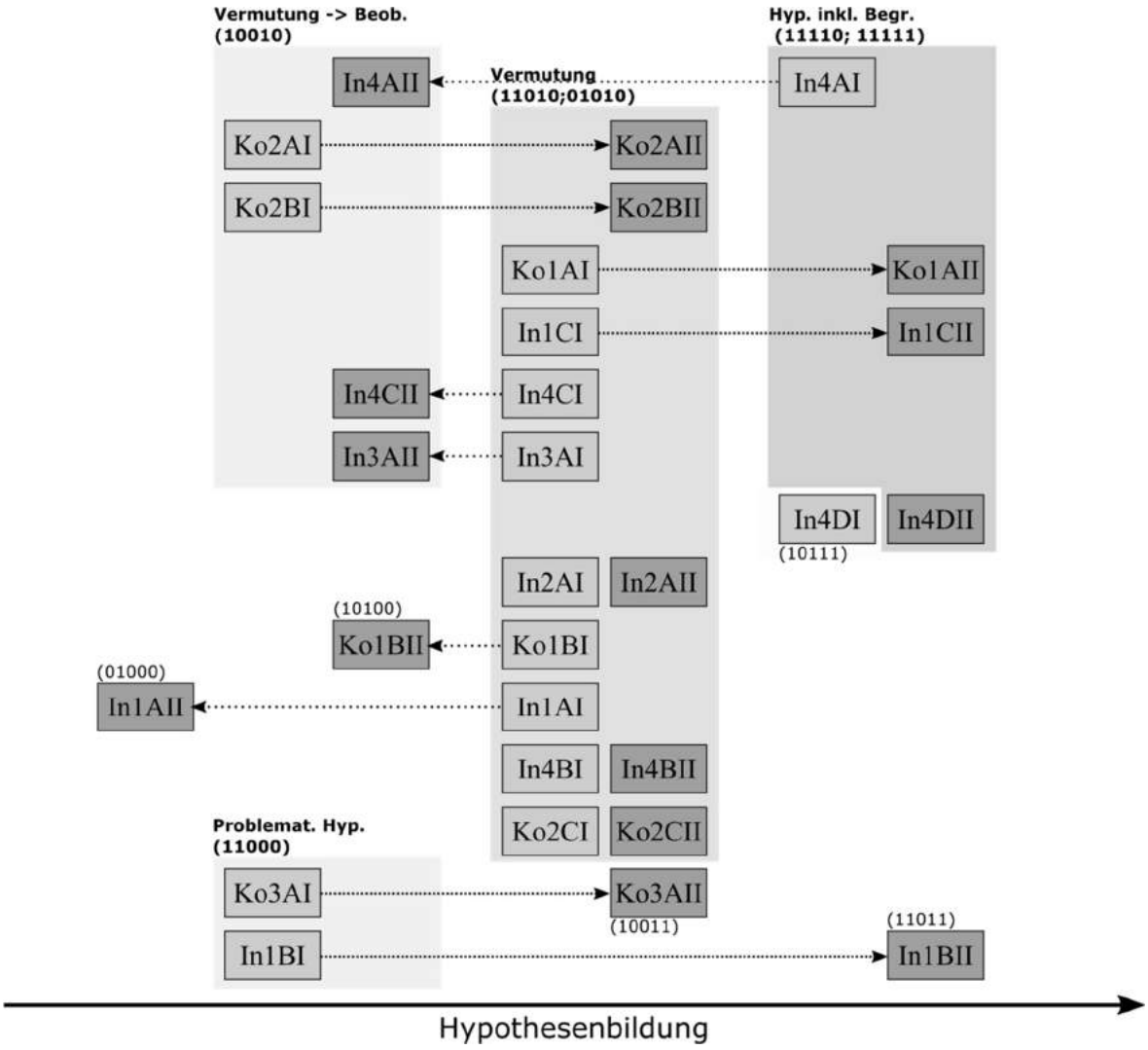


Abbildung 17: Clusteranalyse „Hypothesenbildung“

In fast allen Unterrichtstunden findet eine Hypothesenbildung statt. Oft sind die Hypothesen (hier meist treffender: Vermutungen) dabei allein auf die zu erwartende Beobachtung ausgerichtet („Was glaubt ihr, was werdet ihr beobachten?“; Cluster „Vermutung -> Beobachtung“: 2 pre / 3 post) oder es werden sehr allgemein Vermutungen ohne Begründung und nicht notwendigerweise mit Bezug auf die Fragestellung im Plenum gesammelt (Cluster „Vermutung“: 9 pre / 5 post). In wenigen Unterrichtstunden wurde ein problematischer Hypothesenbegriff bei der Lehrkraft augenfällig (etwa Cluster „Problematischer Hypothesenbegriff“: 2 pre / 0 post). Ein problematischer Hypothesenbegriff liegt etwa vor, wenn die Lehrkraft als Hypothesenbildung die Lernenden zum Erraten des Zwecks eines Versuchs/Experiments auffordert oder unwissenschaftliche Hypothesen („Superkräfte“ etc.) als Hypothesen von der Lehrperson unkritisch akzeptiert werden. Vereinzelt wurden Hypothesen (mit der Aufforderung zur Begründung, sprich theoretischer Fundierung; Cluster „Hypothese inkl. Begründung“: 1 pre / 3 post) aufgestellt sowie teils durch die explizite Einforderung von Begründungen durch die Lehrperson eine umfassende Hypothesenbildung erzielt.

Bezüglich der Entwicklung des Unterrichtshandelns in dieser Oberkategorie kann kein eindeutiger Trend identifiziert werden: Zwar zeigen sechs der 15 Lehrkräfte bezüglich dieser Oberkategorie eine (teils deutliche) Weiterentwicklung, doch gibt es bei 5 Lehrkräften auch Rückschritte zu verzeichnen.
Die Entwicklung des Unterrichtshandelns bezüglich der Hypothesenbildung soll exemplarisch anhand der Lehrkraft Ko3A nachvollzogen werden.
In der pre-Unterrichtsstunde (Ko3AI) von Lehrkraft Ko3A zum Thema ‚Zeitverschiebung' lässt die Lehrkraft zwar Vermutungen aufstellen, hierbei handelt es sich aber eher um ‚Mutmaßungen': Die Frage „Warum kann ich nicht um 12 Uhr heute Mittag mit ihm telefonieren?" (Lehrkraft, Z. 49) ist keine wissenschaftliche Fragestellung und lässt damit auch keine Hypothesenbildung im wissenschaftlichen Sinn zu. Die Lernenden stellen Mutmaßungen an (z. B. „weil sie gar keine Uhrzeit haben", Z. 380; „Oder weil ich bei der Arbeit bin oder er", Z. 382/383). Der vermittelte Hypothesenbegriff der Lehrperson ist hier problematisch zu sehen.
In der post-Unterrichtsstunde (Ko3AII) lässt die Lehrkraft zur Frage, was mit einem Schwimmreifen passiert, wenn man ihn in der Sonne liegen lässt, wieder Hypothesen aufstellen. Die Lernenden formulieren nun naturwissenschaftlich prüfbare Hypothesen („S11: Da geht Luft raus", Z. 40), die Lehrkraft fordert darüber hinaus auf, die jeweils aufgestellte Hypothese zu begründen („Aber warum geht da Luft raus?", Z. 42).
Die Lehrkraft verändert zwischen beiden Untersuchungszeiträumen den vermittelten Hypothesenbegriff von einem ‚Mutmaßen' hin zu einer Hypothesenbildung, welche Ausgangspunkt für die Planung einer wissenschaftlichen Untersuchung sein kann. Möglicherweise ist jedoch das ‚Mutmaßen' in der pre-Erhebung auch inhaltlichen Zwängen geschuldet.
Zusammenfassung der Clusteranalyse „Hypothesenbildung": In den meisten Stunden findet eine – wenn auch nur rudimentäre – Hypothesenbildung statt. Über den Inverventionszeitraum hinweg konnte der Vermittlung eines problematischen Hypothesenverständnisses in Teilen begegnet werden. Bezüglich der Entwicklung des Unterrichtshandelns in dieser Oberkategorie gibt es keinen eindeutigen Trend.

Clusteranalyse „Aufforderung zu wissenschaftsorientiertem Arbeiten"

Für die Clusteranalyse wurden die Kategorien S8–S11 des Kategoriensystems herangezogen. Mit dem Single-Linkage-Kriterium wurden fünf ‚Ausreißer'-Unterrichtsstunden identifiziert. Diese wurden in der nun folgenden erneuten Clusteranalyse nicht berücksichtigt, jedoch in die Abbildung aufgenommen (vgl. Abbildung 18). Über das ‚Elbow'-Kriterium wurde die optimale Anzahl an Clustern mit 3 bestimmt. Das Resultat der Clusteranalyse ist, wiederum ergänzt um weitere Informationen, in Abbildung 18 dargestellt.

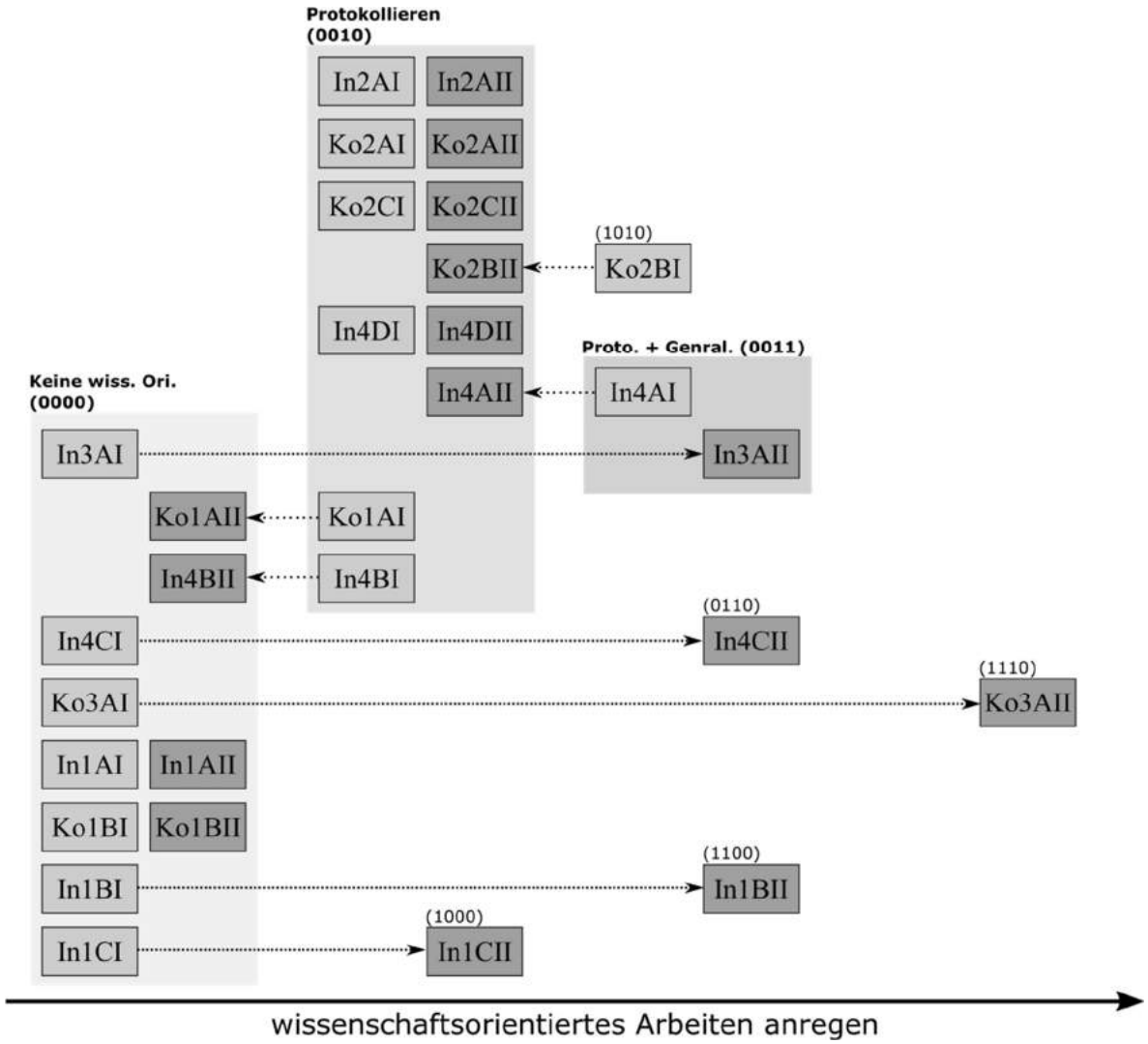


Abbildung 18: Clusteranalyse „Aufforderung zu wissenschaftsorientiertem Arbeiten“

Bei einem großen Teil der Unterrichtsstunden wurde wissenschaftsorientiertes Arbeiten nicht eingefordert (7 pre / 4 post) oder beschränkte sich allein darauf, dass die Lehrkraft die Lernenden beim Experimentieren darauf hinwies, zu protokollieren (6 pre / 6 post). In zwei Unterrichtstunden wurde das Führen eines Protokolls eingefordert und die Generalisierbarkeit der gewonnenen Erkenntnisse diskutiert. Fünf der sieben Lehrkräfte, welche vor der Fortbildung keinerlei wissenschaftsorientiertes Arbeiten anregten setzten nach der Fortbildung in verschiedenem Ausmaß Aspekte wissenschaftsorientierten Arbeitens um. Vier Lehrkräfte machten einen Rückschritt.

Exemplarisch soll wieder eine Lehrkraft (K03A) detaillierter betrachtet werden. Während die Lehrkraft Ko3A in der pre-Erhebung (Thema ‚Zeitverschiebung‘) die Lernenden nicht explizit zum wissenschaftsorientierten Arbeiten angeregt hat, konnten in der post-Unterrichtsstunde drei verschiedene Aspekte zum Anregen wissenschaftsorientierten Arbeitens identifiziert werden: 1. Die Lehrkraft fordert die Achtung der Vergleichbarkeit der einzelnen Ansätze (L: „Bei einem Experiment war wichtig, dass wir immer die gleichen Dinge verwenden und erst einmal

immer nur eine Sache abändern“, Z. 683–685). 2. Sie fordert die Lernenden auf, sowohl Test- als auch Kontrollansatz zu implementieren (L: „[…] einfach wie es jetzt im aktuellen Zustand ist. Das ist sozusagen unsere Kontrolle, die wir haben“, Z. 830/831) und 3. Die Lernenden werden aufgefordert, während des Experimentierens Protokoll zu führen.
Die Lehrkraft Ko3A regt vor der Intervention wissenschaftliches Arbeiten nicht explizit an, implementiert aber nach der Intervention zahlreiche Gelegenheiten zur Vermittlung wissenschaftlicher Arbeitsweisen. Eine fehlende wissenschaftliche Orientierung in der Pre-Erhebung könnte jedoch auch mit dem inhaltlichen Kontext (‚Zeitverschiebung‘) zu begründen sein.
Zusammenfassung der Clusteranalyse „Aufforderung zu wissenschaftsorientiertem Arbeiten“: Die Orientierung an wissenschaftlichem Arbeiten ist schwach ausgeprägt. Bei mehr als zwei Dritteln der Unterrichtsstunden beschränkt es sich auf einen Verweis der Lehrkraft, zu protokollieren, oder fehlt ganz. Fünf Lehrkräfte, welche zuvor keine Wissenschaftsorientierung von ihren Lernenden eingefordert haben, entwickelten – individualisiert – Wege zur Steigerung der Wissenschaftsorientierung bei den Lernenden.

Clusteranalyse „Metareflexion“

Für die Clusteranalyse wurden die vier Kategorien R8 bis R11 genutzt. Über das Single-Linkage-Kriterium wurden fünf ‚Ausreißer‘ identifiziert und aus der weiteren Clusteranalyse herausgenommen. Das ‚Elbow‘-Kriterium deutet auf zwei bis drei Cluster, der Silhouette-Algorithmus auf drei Cluster hin. Da auch inhaltlich drei Cluster sinnvoll unterschieden werden können, wurde die Clusteranzahl mit 3 festgesetzt. Das Ergebnis der Clusteranalyse ist in Abbildung 19, aufbereitet und ergänzt, dargestellt.

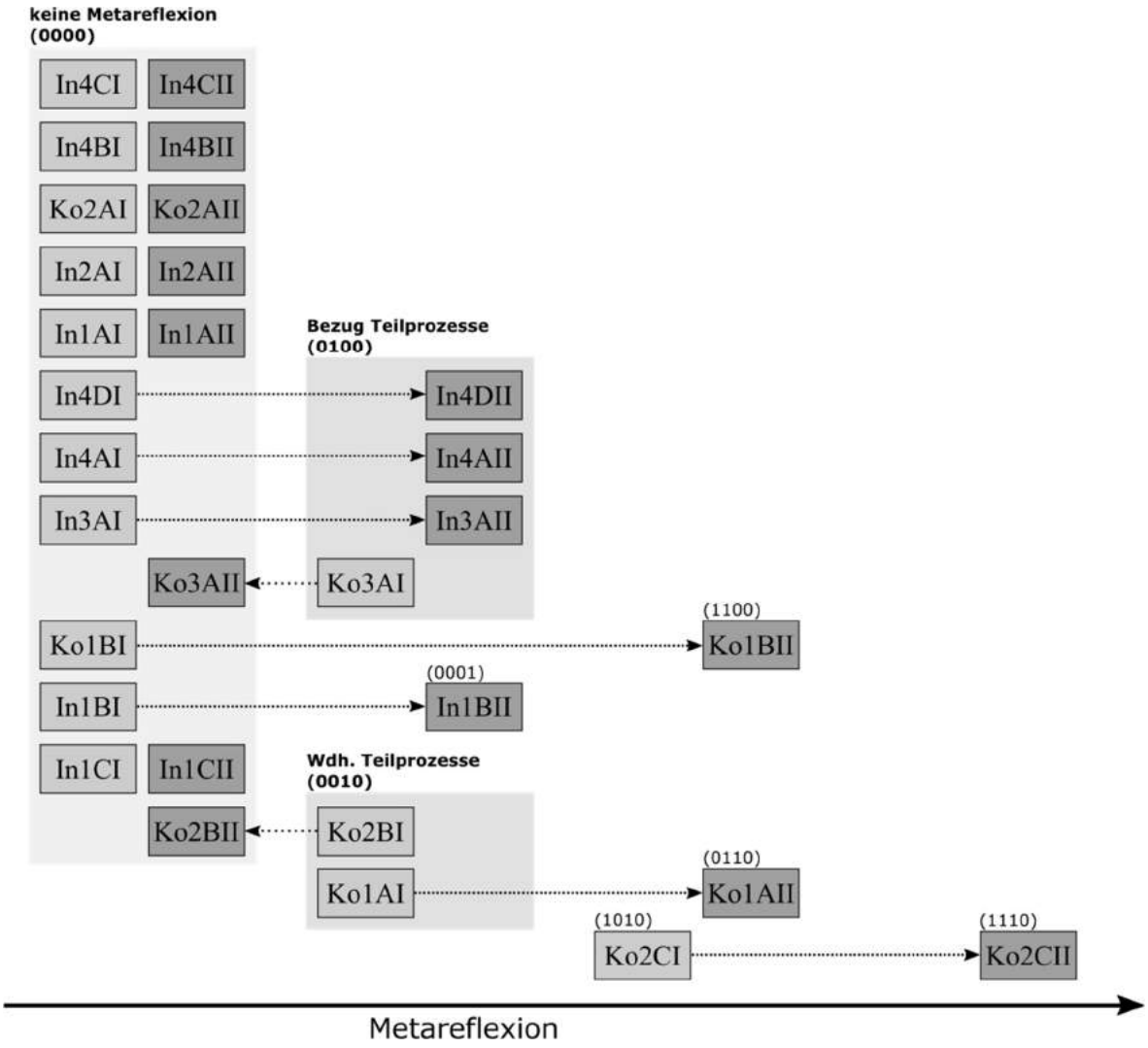


Abbildung 19: Clusteranalyse „Metareflexion“

In fast zwei Drittel der Unterrichtsstunden konnte keine Reflexion der Erkenntnismethode ‚Experiment‘ beobachtet werden (Cluster „keine Metareflexion“; 11 pre/8 post). Vier Lehrkräfte diskutierten mit den Lernenden den Bezug einzelner Teilprozesse zueinander (Cluster „Bezug Teilprozesse“; 1 pre/3 post). Zwei Lehrkräfte wiederholten begrifflich die einzelnen Teilprozesse mit den Lernenden (Cluster „Wiederholung Teilprozesse“; 2 pre/0 post).
Bei etwa der Hälfte der Lehrkräfte, die vor der Fortbildung anhand der Unterrichtsvideografie keine explizite Reflexion der Erkenntnismethode Experiment erkennen ließen, führt die Intervention zu keiner Veränderung des diesbezüglichen Unterrichtshandelns. Fünf von elf Lehrkräften aus dem Cluster „keine Metareflexion“ implementieren nach der Fortbildung zumindest Teilaspekte einer Metareflexion in ihr Unterrichtshandeln. Insgesamt zeigen 7 der 15 Lehrkräfte nach der Intervention ein elaborierteres Unterrichtshandeln bezüglich dieser Oberkategorie – zwei Lehrkräfte machen jedoch, möglicherweise bedingt durch äußere Umstände bei der zweiten Videografie wie etwa dem Lerngegenstand, Rückschritte. Neben Zu-

wächsen über die Fortbildung hinweg lässt sich außerdem bei einem Teil der Lehrkräfte eine Akzentuierung des Unterrichtshandelns, erkennbar an den ‚Ausreißer'-Unterrichtstunden, welche sich einer Clusterung verwehren, bezüglich dieser Oberkategorie feststellen.
Zur Veranschaulichung wird im Folgenden die Entwicklung des Unterrichtshandelns einer Lehrkraft (Ko1B) bezüglich der ‚Metareflexion' genauer vorgestellt. Während Lehrkraft Ko1B in der pre-Erhebung (Ko1BI) keine Aspekte einer Reflexion der Erkenntnismethode Experiment in ihre Unterrichtsstunde implementiert, werden in der Unterrichtstunde nach der Fortbildung (Ko1BII) die Teilprozesse des Experiments wiederholt („Lehrperson zeigt auf Plakat mit den Teilprozessen beim Experimentieren", Z. 395) und der Hypothesenbegriff – jedoch mit einem teils problematischen Hypothesenverständnis der Lehrkraft – diskutiert („L: Ich glaube wir müssen uns über den Begriff [Hypothese] unterhalten", Z. 345; im Folgenden wird der Hypothesenbegriff im Plenum diskutiert). Es kann also bei der Lehrkraft eine Zunahme von Lerngelegenheiten über die Erkenntnismethode Experiment im Sinne einer Metareflexion der Methode beobachtet werden.
Eine gegensätzliche Beobachtung kann bei der Lehrkraft Ko2B gemacht werden: Findet in der Unterrichtsstunde vor der Intervention (Ko2BI) noch eine Wiederholung bzw. Klärung der Teilprozesse statt („L: […] ihr könnt schon mal kurz überlegen, welche Infos wir in ein Protokoll schreiben", Z. 90/91; im Folgenden Verlauf des Unterrichts werden die Teilprozesse wiederholt), verzichtet die Lehrkraft in der post-Erhebung (Ko2BII) trotz ausführlicher Reflexion der Durchführung ganz auf Elemente der Metareflexion.
Zusammenfassung der Clusteranalyse „Metareflexion": Eine Mehrheit der Lehrkräfte bindet keine Reflexion des Experiments auf der Metaebene in ihr Unterrichtshandeln ein, ein weiterer Teil allenfalls rudimentär. Nach der Intervention implementieren einige Lehrkräfte – stark individualisiert und teils noch auf niedrigem Niveau – Teilaspekte einer Metareflexion in ihren Unterricht.

Clusteranalyse „Unterstützte Offenheit"

Für die Clusteranalyse wurden die Kategorien K9–K12 des Kategoriensystems verwendet. Die Kategorie K10 wurde dabei zuvor ‚umgepolt'. Mithilfe des Single-Linkage-Kriteriums wurden sechs ‚Ausreißer'-Unterrichtsstunden identifiziert und für die zweite Clusteranalyse ausgeschlossen. Mit dem ‚Elbow'-Kriterium wurde die Anzahl der Cluster mit 3 bestimmt. Diese Clusterung ist in Abbildung 20 dargestellt. Es wurden wieder die zuvor ausgeschlossenen Unterrichtstunden sowie Entwicklungen über den Interventionszeitraum ergänzt.

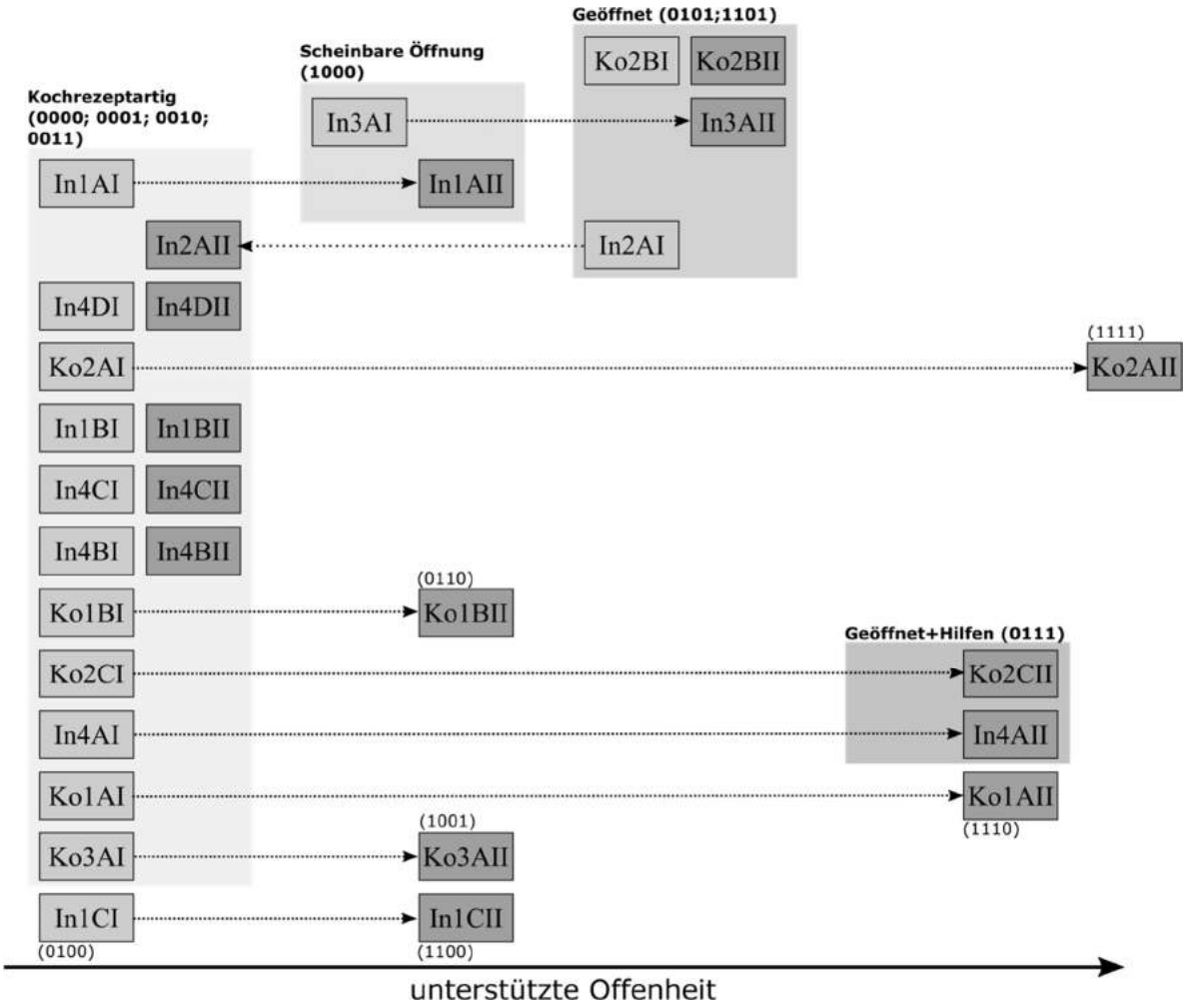


Abbildung 20: Clusteranalyse „Unterstützte Offenheit“

In mehr als der Hälfte der untersuchten Unterrichtsstunden (11 pre/5 post) wurden experimentelle Arbeitsphasen durch die Lehrkraft, meist über eine detaillierte Experimentieranleitung, kochrezeptartig strukturiert (Cluster „Kochrezeptartig“). Zwei Unterrichtstunden bieten den Lernenden zwar eine vorgeblich freie Materialauswahl aus einem Materialpool, geben das Experiment dennoch im Stil eines Kochrezepts (sprich: als detaillierte Anleitung) vor (Cluster „Scheinbare Öffnung“). In weiteren Stunden ist eine Öffnung zu beobachten (Cluster „Geöffnet“), teilweise weist die Lehrkraft dabei auch – im Sinne einer Akzentuierung – auf die Inanspruchnahme von Hilfen hin. Die Verwendung eines Protokollbogens (K12) tritt sowohl bei sehr eng geführten Experimentierphasen als auch offenen Settings auf.

Über den Verlauf der Interventionsmaßnahme zeigen sich zwei Entwicklungsbewegungen: Erstens ist eine allgemeine und individuell teils deutliche Verschiebung weg von kochrezeptartigen Experimentierphasen hin zu einer stärkeren Öffnung, teils ergänzt um individualisierende Hilfsangebote, zu verzeichnen. Neun der 15 Lehrkräfte zeigen eine positive Entwicklung hinsichtlich dieser Oberkategorie, bei nur einer Lehrkraft sind diesbezüglich leicht negative Entwicklungen zu

beobachten. Möglicherweise spielen bezüglich dieser beobachteten Schließung jedoch Sicherheitsaspekte eine Rolle, da in der Post-Unterrichtstunde (wie in der Pre-Unterrichtstunde) die Lernenden mit Bunsenbrennern arbeiten. Zweitens findet eine Individualisierung des Unterrichtshandelns bezüglich der Oberkategorie „unterstützte Offenheit“ statt: Alle 4 ‚Ausreißer‘-Unterrichtsstunden können der post-Erhebung zugeordnet werden.

Exemplarisch sollen wieder zwei Lehrkräfte (Ko2C und In2A) detaillierter betrachtet werden. Zwar weist die Lehrperson Ko2C in der Unterrichtsstunde vor der Intervention (Ko2CI) auf Hilfen hin, doch ist das Vorgehen bei dem Experiment bereits von der Lehrkraft und mittels eines Arbeitsblattes so detailliert erklärt („L: So, ihr bekommt von mir ein AB, lest dieses AB bitte durch, dort steht genau, was ihr braucht. Überlest nichts, wenn dort steht, dass ihr etwas langsam einfüllen sollt, dann sollt ihr es auch langsam einfüllen. […]“, Z. 645–647), dass die Lernenden das Experiment meist einfach ‚nachbauen‘ und die Hilfen auf dem Arbeitsblatt kaum in Anspruch genommen werden.

Die Untersuchung der post-Unterrichtsstunde (Ko2CII) ist hingegen nicht mehr ‚kochrezeptartig‘ vorgegeben. Als Unterstützung der gewonnenen Offenheit werden den Lernenden Hilfen („Lehrperson zeigt zur Tafel und weist auf die Merkblätter zum Wasserkreislauf hin“, Z. 894) und als Strukturierung des Experiments ein Protokollbogen ausgegeben.

Bei der Lehrkraft Ko2C ist nicht nur ein Abrücken von der eng geführten ‚Kochrezeptanleitung‘, sondern auch deren Begleitung der Öffnung durch Hilfen und Strukturierungen zu beobachten.

Lehrkraft In2A fällt als einzige Lehrkraft zwischen der pre- und post-Erhebung bezüglich der ‚unterstützten Offenheit‘ etwas zurück: Ist zu Beginn das Experiment noch nicht kochrezeptartig vorgegeben, aber mittels eines Protokollbogens strukturiert, wird von der Lehrkraft in der post-Erhebung auf eine detaillierte Anleitung zurückgegriffen. Angemerkt sei hierbei, dass es sich um chemische Experimente handelt: In der pre-Unterrichtsstunde ist das Thema „Was alles braucht ein Feuer, damit es brennt?“, in der post-Unterrichtstunde wird die Verbrennung von Nichtmetallen durchgeführt. Eine enge Führung anhand einer detaillierten Anleitung scheint aus Gründen der Sicherheit in beiden Fällen von der Lehrkraft gewollt – auch wenn diese enge Führung aufgrund von Sicherheitsaspekten allein für die Durchführung zwingend notwendig scheint.

Bei der Lehrkraft In2A ist im pre-post-Vergleich eine zunehmend engere Führung zu beobachten. Diese zunehmende Führung ist möglicherweise vor dem Hintergrund sicherheitsbezogener Überlegungen (oder allein des Sicherheitsempfindens) der Lehrkraft einzuordnen.

Zusammenfassung der Clusteranalyse „Unterstützte Offenheit“: Besonders in Unterrichtsstunden der pre-Erhebung sind Experimentierphasen kochrezeptartig

strukturiert. Über die Fortbildung hinweg findet eine breite und teils deutliche Entwicklung hin zu geöffneten und mit Hilfsangeboten begleiteten Experimentierphasen statt. Eine Individualisierung des Unterrichtshandelns im Verlauf der Fortbildung zeichnet sich ab.

4.1.2 Änderungen des Unterrichtshandeln mit Fokus auf die einzelnen Lehrkräfte

Im Folgenden werden die Ergebnisse für die einzelnen Lehrkräfte auf den drei Dimensionen des Kategoriensystems 1 ‚Unterrichtshandeln von Lehrkräften beim Experimentieren' vorgestellt. Die zugrundeliegenden Daten sind hierbei mit den Daten, welche für die Clusteranalyse verwendet wurden, identisch.

In den Tabellen sind jeweils die Mittelwerte pre und post sowie der *p*-Wert, berechnet mittels eines einseitigen, exakten Binomial-Tests, angegeben. Die Binomial-Tests werden nach Bortz & Schuster (2010) mit den Parametern $\pi = 1/2$, $N = B+C$ und $x = min(B,D)$ durchgeführt. π ist hierbei der Populationsparameter, die Parameter *B*, *C* und *D* beziehen sich auf die jeweiligen Zellen in Tabelle 22. Da das dichotome Kategoriensystem gerichtet ist (basierend auf den Unterrichtszielen- und prinzipien), wird auch stets nur ein einseitiger Test durchgeführt. Außerdem finden sich in den Tabellen jeweils die Anzahl der Kategorien mit positiver (+) und negativer (-) Entwicklung. Hierzu wurde auf die Einträge c und b der jeweiligen Kreuztabellen zurückgegriffen (vgl. Tabelle 22). Lehrkräfte, welche an dem kokonstruktiven Fortbildungsformat teilgenommen haben, sind grau unterlegt. Die Codierung der Lehrkräfte orientiert sich an der in Abbildung 15 eingeführten Kodierung. In der Spalte „Entwicklung" findet sich die jeweilige Anzahl der Kategorien, in welche die Lehrkraft zwischen pre und post zu einem effektiveren Unterrichtshandeln wechselt (+) oder vormals gezeigtes, effektives Unterrichtshandeln einstellt (-). Findet in einer Kategorie keine Änderung des Unterrichtshandelns statt, ist diese Kategorie auch nicht als Strich (|) in der Spalte „Entwicklung" vertreten. Viele Striche (|) sind somit ein Indikator für eine starke Änderung des Unterrichtshandelns. Liegen diese Striche vornehmlich im positiven Sektor (+), ist dies ein Zeichen für eine Änderung hin zu einem effektiveren Unterrichtshandeln bezüglich des Experimentierens.

Tabelle 25: Dimension ‚Prozessstruktur' (nach Lehrkraft aufgeschlüsselt)

Mittelwerte (MW), Entwicklung (+; -), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig)					
	pre	**post**	**Entwicklung**		
Lehrkraft	**MW**	**MW**	**-**	**+**	***p*-Wert (W-Wert)**
Ko2C	.45	.75		IIIII I	.016* (0)
Ko3A	.55	.85	I	IIIII II	.035* (1)
In1B	.579	.84		IIIII	.031* (0)
In3A	.3	.55	II	IIIII II	.09 (2)
In1C	.4	.50	II	IIII	.344 (2)
In4C	.4	.50	II	IIII	.344 (2)
Ko1A	.65	.65	IIII	IIII	.637 (4)
Ko1B	.421	.421	IIII	IIII	.637 (4)
In4D	.7	.65	III	II	.5 (2)
In1A	.4	.3	III	I	.313 (1)
Ko2A	.55	.45	IIII	II	.344 (2)
In4A	.684	.58	IIIII	III	.363 (3)
In2A	.65	.50	III		.125 (0)
Ko2B	.6	.45	IIII	I	.188 (1)
In4B	.65	.40	IIIII		.031* (0)

Bei vier Lehrkräften zeigen sich signifikante Änderungen des Unterrichtshandelns bezüglich der Dimension ‚Prozessstruktur' (20 Kategorien). Bei drei dieser Lehrkräfte zeigt sich ein positiver Effekt, eine Lehrkraft zeigt eine signifikant negative Entwicklung.

Tabelle 26: Dimension ‚Reflexion' (nach Lehrkraft aufgeschlüsselt)

Mittelwerte (MW), Entwicklung (+; -), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig)					
	pre	**post**	**Entwicklung**		
Lehrkraft	**MW**	**MW**	**-**	**+**	***p*-Wert (W-Wert)**
Ko2C	.167	.583		IIIII	.031* (1)
Ko1A	0	.333		IIII	.5 (0)
Ko3A	.143	.571		III	.125 (2)
In4A	0	.286		II	-
In3A	.25	.375		I	.5 (0)
In1C	.167	.250		I	.5 (0)
In4C	.25	.333		I	-
Ko1B	.083	.167		I	.5 (0)
In4D	.25	.250	I	I	.75 (1)
In1A	.5	.500	I	I	.75 (1)
Ko2A	.333	.333	II	II	.688 (0)
Ko2B	.333	.333	I	I	.75 (1)
In4B	.333	.333	I	I	.75 (0)
In2A	.25	.167	I		.5 (0)
In1B	.25	.083	III	I	.313 (0)

Die Dimension ‚Reflexion' beinhaltet 12 Kategorien. Einzig eine Lehrkraft zeigt in ihrem Unterrichtshandeln bezüglich dieser Kategorie eine signifikant (p = .031) positive Entwicklung.

Tabelle 27: Dimension ‚Angemessenheit und Unterstützung' (nach Lehrkraft aufgeschlüsselt)

Mittelwerte (MW), Entwicklung (+; -), Signifikanz (p-Wert jeweils über exakten Binomial-Test, einseitig)

Lehrkraft	pre MW	post MW	Entwicklung -	Entwicklung +	p-Wert (W-Wert)
In1A	.083	.583		IIIII I	.016* (0)
Ko2B	.167	.500		IIII	.063 (0)
Ko2C	.333	.500	II	IIII	.344 (2)
Ko3A	.083	.250		II	.25 (0)
In4C	.333	.417	II	III	.5 (2)
Ko1B	.5	.583	I	II	.5 (1)
In4D	.5	.583	I	II	.5 (1)
Ko2A	.25	.333	II	III	.5 (2)
Ko1A	.333	.333	II	II	.688 (2)
In4A	.25	.250	II	II	.688 (2)
In3A	.25	.250	I	I	.75 (1)
In1C	.333	.333	I	I	.75 (1)
In2A	.333	.333	I	I	.75 (1)
In4B	.333	.250	I		.5 (0)
In1B	.5	.250	III		.125 (0)

Es zeigt nur eine Lehrkraft in dieser Dimension ‚Angemessenheit und Unterstützung' (12 Kategorien) signifikante (p = .016) Änderung des Unterrichtshandelns. In dieser Dimension zeigen alle Lehrkräfte des kokonstruktiven Fortbildungsformats neutrale oder positive Entwicklungen.

Tabelle 28: Effektiver Unterricht zum Experimentieren (gesamtes Kategoriensystem)

(MW), Entwicklung (+; -), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig)

Lehrkraft	pre MW	post MW	Entwicklung -	Entwicklung +	*p*-Wert (W-Wert)
Ko2C	.364	.636	I	IIIII IIIII III	.001* (1)
In3A	.227	.455	IIII	IIIII IIIII IIII	.015* (2)
Ko3A	.409	.568	IIIII	IIIII IIIII II	.072 (4)
Ko1A	.432	.591	I IIIII	IIIII IIIII III	.084 (1)
In1B	.41	.564	II	IIIII III	.055 (1)
In4A	.395	.5	IIIII	IIIII IIII	.212 (0)
Ko1B	.342	.421	I IIIII	IIIII IIII	.304 (1)
In1C	.295	.364	III	IIIII I	.254 (1)
In4C	.364	.409	IIII	IIIII I	.377 (2)
Ko2A	.409	.409	IIIII	IIIII	.623 (1)
In1A	.273	.227	I IIIII	IIII	.377 (3)
In2A	.455	.409	IIII	II	.344 (1)
Ko2B	.477	.432	I IIIII	IIII	.377 (2)
In4D	.545	.455	II IIIII	III	.172 (1)
In4B	.455	.295	II IIIII		.008* (0)

Bezüglich der Entwicklung bei den einzelnen Lehrkräften über alle drei Dimensionen (44 Kategorien) zeigen drei Lehrkräfte (positiv: 2; negativ: 1) signifikante Änderungen des Unterrichtshandelns. Unter den beiden sich signifikant positiv entwickelnden Lehrkräften findet sich auch die Lehrkraft In3A, welche in der pre-Erhebung das niedrigste Niveau belegte (pre MW: .227).

Bezüglich aller drei Dimensionen und auch in der Gesamtbetrachtung wird sichtbar, dass sich die Lehrkräfte in ihrem Unterrichtshandeln individuell unterschiedlich stark weiterentwickelt haben. Die Überlegenheit des kokonstruktiven Fortbildungsformats zeigt sich nicht eindeutig.

4.1.3 Ergebnisse der Vergleichsanalyse

Im Folgenden werden die Ergebnisse für die Test- sowie Kontrollgruppe als auch die Gesamtstichprobe auf verschiedenen Ebenen des Kategoriensystems 1 ‚Unterrichtshandeln von Lehrkräften beim Experimentieren' vorgestellt. Die zugrundeliegenden Daten sind hierbei mit den Daten, welche für die Clusteranalyse verwendet wurden, identisch. Zuerst werden die Ergebnisse auf der untersten Ebene – der einzelnen Kategorien – berichtet.

Ergebnisse auf der Ebene der Kategorien

Trotz der geringen Stichproben ($n_{Testgruppe} = 6$; $n_{Kontrollgruppe} = 9$; $N_{gesamt} = 15$) sollen auch auf der Ebene der Kategorien, aus Gründen der wissenschaftlichen Transparenz, die Ergebnisse berichtet werden, obgleich die Aussagekraft, etwa aufgrund fehlender Signifikanzwerte, limitiert ist. Wegen der geringen Stichprobengrößen

schlägt eine Bestimmung der Signifikanz durch den Binomialtest (oder sogar dem McNemar-Test) oft fehl oder liefert für die Test- und Kontrollgruppe meist keine interpretierbaren Ergebnisse. Darum wird eine durch den Binomialtest berechnete Signifikanz – auf der Ebene der Kategorien – nur für die Gesamtstichprobe angegeben, wenngleich auch dort eine Berechnung aus mathematischen Gründen nicht immer möglich ist. Die Auftrittshäufigkeiten werden pre und post in Prozent angegeben. Die Auftrittshäufigkeit wird in Prozent und nicht absolut angegeben, da nicht alle Kategorien auch auf alle Unterrichtsvideografien angewendet können. Bei der dichotomen 0/1-Codierung sind die Auftrittshäufigkeiten mathematisch mit dem arithmetischen Mittel für die Kategorie und dem jeweiligen Messzeitpunkt äquivalent. Als alternative Zusammenhangsmaße werden sowohl die prozentuale Veränderung als auch die Veränderung in Prozentpunkten berichtet. Die Angabe der Änderung der Prozentpunkte dient bei jenen Kategorien als Orientierung, bei welchen zum ersten Messzeitpunkt die Kategorie gar nicht auftritt und so die prozentuale Änderung mathematisch nicht berechnet werden kann (Division durch null). Wie bereits in Kapitel 4.1.1 erläutert, sind Kategorien, welche ein nicht erstrebenswertes Handeln (z. B. K11: ‚L gibt Versuchsanleitung, welche kochrezeptartig strukturiert ist') beschreiben, ‚umgepolt': Je höher hier der Mittelwert, desto seltener ist die Kategorie aufgetreten. Diese Kategorien sind durch eine Tilde (~) gekennzeichnet. Entspricht die Stichprobengröße aufgrund fehlender Werte (Wertung ‚nicht anwendbar') nicht dem in der Tabellenbeschreibung berichteten Umfang, so ist hinter der Kategorie die angepasste Stichprobengröße aufgeführt.

Die Ergebnisse für die Kontrollgruppe (Individualisiert-konstruktiv; $N = 9$) sind in Tabelle 29 dargestellt.

Tabelle 29: Auftrittshäufigkeiten der Kategorien in Prozent pre und post sowie die Änderung in Prozentpunkten und die prozentuale Änderung für die Kontrollgruppe

Kategorie	pre MW (SD)	post MW (SD)	prozentuale Änderung	Änderung Proz.punkte
S1 (Problem)	.33 (.50)	.56 (.53)	.67	.22
S1A ~ (Ingenieur)	.78 (.44)	.67 (.50)	-.14	-.11
S2 (Frage)	.56 (.53)	.33 (.50)	-.40	-.22
S2A (nw. Frage)	.56 (.53)	.22 (.44)	-.60	-.33
S3 (Hyp. aufstellen)	.56 (.53)	.67 (.50)	.20	.11
S3A ~ (Hyp. -> Beob.)	.89 (.33)	.67 (.50)	-.25	-.22
S3B (Hyp. mit Begr.)	.22 (.44)	.22 (.44)	.00	.00
S4 ~ (probl. Hyp.)	.89 (.33)	.89 (.33)	.00	.00
S5 (Begründung)	.11 (.33)	.22 (.44)	1.00	.11
S6 (Ideen)	.44 (.53)	.56 (.53)	.25	.11
S7 (Planung bespr.)	.78 (.44)	.78 (.44)	.00	.00
S8 (Vergleichbarkeit)	.00 (.00)	.22 (.44)	-	.22
S9 (Kontrollansatz)	.00 (.00)	.22 (.44)	-	.22

S10 (Protokollieren)	.44 (.53)	.56 (.53)	.25	.11
S11 (Grenzen)	.11 (.33)	.11 (.33)	.00	.00
S12 (Erfragen Beob.)	1.00 (.00)	1.00 (.00)	.00	.00
S13 (Festhalten Beob.)	.78 (.44)	.67 (.50)	-.14	-.11
S14 (*n* = 7) (Erk. notieren)	.57 (.54)	.57 (.54)	.00	.00
S15A (Teilprozess)	.67 (.50)	.67 (.50)	.00	.00
S15B ~ (Teilprozess)	.89 (.33)	.89 (.33)	.00	.00
R1 (Whd. Vermutung)	.11 (.33)	.22 (.44)	1.00	.11
R2 (*n* = 7) (Veri./Falsi.)	.14 (.38)	.29 (.49)	1.00	.14
R3 (Erfragen Prob.)	.33 (.50)	.22 (.44)	-.33	-.11
R4 (*n* = 7) (Verbess.)	.00 (.00)	.14 (.38)	-	.14
R5 (*n* = 7) (Bespr. DF)	.71 (.49)	.57 (.54)	-.20	-.14
R6 ~ (*n* = 8) (Objekte)	.88 (.35)	.88 (.35)	.00	.00
R7 (*n* = 7) (Bespr.Erk.)	.57 (.54)	.71 (.49)	.25	.14
R8 (Begriff Vermutung)	.00 (.00)	.00 (.00)	-	.00
R9 (Wdh. TP)	.00 (.00)	.00 (.00)	-	.00
R10 (Bezug TP)	.00 (.00)	.33 (.50)	-	.33
R11 (Funkt. Kontrollan.)	.00 (.00)	.11 (.33)	-	.11
R12 (Üben)	.00 (.00)	.00 (.00)	-	.00
K1 ~ (Gibt Beob.)	.89 (.33)	1.00 (.00)	.12	.11
K2 ~ (Händ. Eingriff)	.67 (.50)	.56 (.53)	-.17	-.11
K3 (Labormaterial)	.44 (.53)	.22 (.44)	-.50	-.22
K4 (Verweis: selbst)	.11 (.33)	.00 (.00)	-1,00	-.11
K5 (Verweis: Material)	.00 (.00)	.11 (.33)	-	.11
K6 (Verweis: AB)	.22 (.44)	.44 (.53)	1,00	.22
K7 (Verweis: Hilfen)	.11 (.33)	.00 (.00)	-1,00	-.11
K8 (Verweis: Gruppen)	.00 (.00)	.00 (.00)	-	.00
K9 (Freiheit Material)	.11 (.33)	.22 (.44)	1,00	.11
K10 ~ (Kochrezept)	.22 (.44)	.33 (.50)	.50	.11
K11 (Hinweis Hilfe)	.11 (.33)	.33 (.50)	2,00	.22
K12 (Protokoll)	.56 (.53)	.56 (.53)	.00	.00
		Summe Änderung Prozentpunkte		1.16

Aus der Tabelle 29 geht hervor, dass einige Kategorien sowohl in der pre-Erhebung als auch in der post-Erhebung überhaupt nicht auftreten: Es werden zu beiden Messzeitpunkten in keiner der Unterrichtstunden der Kontrollgruppe der Begriff ‚Vermutung' erklärt (R8), die Teilprozesse des Experiments wiederholt (R9), das Experimentieren explizit eingeübt (R12) sowie beim Experimentieren auf andere Gruppen verwiesen (K8). Die Kategorien R10 (‚Bezug einzelner Teilprozesse im Prozess aufeinander durch L'), R11 (‚Klären von Funktion und Ausgestaltung eines Kontrollansatzes') sowie K5 (‚Verweis auf Material') treten nach der Intervention erstmalig auf. Die letzte Zeile berichtet die Summe der Änderungen der Prozentpunkte über alle Kategorien und hat hier rein illustrativen Charakter: Es ist ein Hinweis, dass sich Gewinne der einzelnen Kategorien mit den Verlusten nicht vollkommen aufwiegen.

In Tabelle 30 sind die Ergebnisse für die Testgruppe (kokonstruktiv; $N = 6$) abgebildet.

Tabelle 30: Auftrittshäufigkeiten der Kategorien in Prozent pre und post sowie die Änderung in Prozentpunkten und die prozentuale Änderung für die Testgruppe

Kategorie	pre MW (SD)	post MW (SD)	prozentuale Änderung	Änderung Proz.punkte
S1 (Problem)	.67 (.52)	.83 (.41)	.25	.17
S1A ~ (Ingenieur)	.50 (.55)	.67 (.52)	.33	.17
S2 (Frage)	.83 (.41)	.50 (.55)	-.40	-.33
S2A (nw. Frage)	.33 (.52)	.50 (.55)	.50	.17
S3 (Hyp. aufstellen)	.83 (.41)	.67 (.52)	-.20	-.17
S3A ~ (Hyp. -> Beob.)	.67 (.52)	.67 (.52)	.00	.00
S3B (Hyp. mit Begr.)	.00 (.00)	.33 (.52)	-	.33
S4 ~ (probl. Hyp.)	.83 (.41)	.83 (.41)	.00	.00
S5 (Begründung)	.00 (.00)	.33 (.52)	-	.33
S6 (Ideen)	.50 (.55)	.67 (.52)	.33	.17
S7 (Planung bespr.)	.50 (.55)	.67 (.52)	.33	.17
S8 (Vergleichbarkeit)	.17 (.41)	.17 (.41)	.00	.00
S9 (Kontrollansatz)	.00 (.00)	.17 (.41)	-	.17
S10 (Protokollieren)	.67 (.52)	.67 (.52)	.00	.00
S11 (Grenzen)	.00 (.00)	.00 (.00)	-	.00
S12 (Erfragen Beob.)	1 (.00)	1 (.00)	.00	.00
S13 (Festhalten Beob.)	1 (.00)	.83 (.41)	-.17	-.17
S14 ($n = 7$) (Erk. notieren)	.45 (.52)	.80 (.45)	.79	.35
S15A (Teilprozess)	.67 (.52)	.83 (.41)	.25	.17
S15B ~ (Teilprozess)	.83 (.41)	.83 (.41)	.00	.00
R1 (Whd. Vermutung)	.17 (.41)	.33 (.52)	.99	.17
R2 ($n = 7$) (Veri./Falsi.)	.45 (.20)	.40 (.55)	-.11	-.05
R3 (Erfragen Prob.)	.45 (.20)	.20 (.45)	-.55	-.25
R4 ($n = 7$) (Verbess.)	.00 (.00)	.00 (.00)	-	.00
R5 ($n = 7$) (Bespr. DF)	.00 (.00)	.60 (.55)	-	.60
R6 ~ ($n = 8$) (Objekte)	1 (.00)	1 (.00)	.00	.00
R7 ($n = 7$) (Bespr.Erk.)	.55 (.60)	1 (.00)	.82	.45
R8 (Begriff Vermutung)	.17 (.41)	.33 (.52)	.99	.17
R9 (Wdh. TP)	.33 (.52)	.50 (.55)	.50	.17
R10 (Bezug TP)	.33 (.52)	.33 (.52)	.00	.00
R11 (Funkt. Kontrollan.)	.00 (.00)	.00 (.00)	-	.00
R12 (Üben)	.00 (.00)	.00 (.00)	-	.00
K1 ~ (Gibt Beob.)	.67 (.52)	.83 (.41)	.25	.17
K2 ~ (Händ. Eingriff)	.67 (.52)	.67 (.52)	.00	.00
K3 (Labormaterial)	.17 (.41)	.17 (.41)	.00	.00
K4 (Verweis: selbst)	.17 (.41)	.33 (.52)	.99	.17
K5 (Verweis: Material)	.00 (.00)	.17 (.41)	-	.17
K6 (Verweis: AB)	.67 (.52)	.33 (.52)	-.50	-.33
K7 (Verweis: Hilfen)	.17 (.41)	.50 (.55)	1.99	.33
K8 (Verweis: Gruppen)	.17 (.41)	.00 (.00)	-1.00	-.17

K9 (Freiheit Material)	.17 (.41)	.67 (.52)	2.99	.50
K10 ~ (Kochrezept)	.17 (.41)	.83 (.41)	3,99	.67
K11 (Hinweis Hilfe)	.17 (.41)	.67 (.52)	2,99	.50
K12 (Protokoll)	.83 (.41)	.67 (.52)	-.20	-.17
		Summe Änderung Prozentpunkte		4.63

Wie anhand der Tabelle 30 zu sehen ist, werden in der Testgruppe sowohl vor als auch nach der Fortbildung sowohl die Vorläufigkeit der gewonnenen Erkenntnisse (S11) als auch Verbesserungsmöglichkeiten der Durchführung (R4) nicht besprochen. Die Funktion des Kontrollansatzes wird in keiner Unterrichtsstunde besprochen (R12), das Experimentieren nicht explizit eingeübt (R12). Jedoch werden nach der Fortbildung erstmals Hypothesen mit Begründung eingefordert (S5) und aufgestellt (S3B). Auch die Aufforderung zur Umsetzung des Kontrollansatzes (S9) findet sich erstmalig in der post-Erhebung der Testgruppe. Auch hier berichtet die letzte Zeile die Summe der Änderungen der Prozentpunkte über alle Kategorien (rein illustrativer Charakter): Es ist ein Hinweis, dass Gewinne der einzelnen Kategorien die Verluste in anderen Kategorien in der Summe übertreffen.

Im Folgenden wird die Gesamtstichprobe auf der Ebene der Kategorien analysiert. Dazu können nun aufgrund der Stichprobengröße Signifikanzen berechnet werden. Zur Berechnung der Signifikanz der einzelnen Kategorien der Gesamtstichprobe wird ein exakter, einseitiger Binomial-Test verwendet. Für die beiden Teilstichproben wurde aufgrund der geringen Stichprobengröße ($n = 6$ bzw. $n = 9$) auf die Bestimmung der Signifikanz verzichtet. Die Verwendung des McNemar-Tests ist aufgrund häufig zu schwach besetzter Zellen B und C nicht möglich, weswegen auf Binomial-Tests ausgewichen wird. Aufgrund der geringen Stichprobengröße ($N = 15$) wird das Signifikanzniveau auf $p < .1$ angehoben. Dieses Vorgehen ist durch kleine Stichprobengrößen sowie den explorativen Charakter des erstmals zur Anwendung gebrachten Kategoriensystems zu begründen (Bortz & Schuster, 2010; Leonhart, 2017). Signifikanzen zwischen $.1 > p > .05$ bedürfen dennoch einer kritischen Diskussion und im besten Fall einer Überprüfung mit einer größeren Stichprobe.
Die Ergebnisse für die Gesamtstichprobe ($N = 15$) sind in Tabelle 31 dargestellt. Aus Platzgründen wird auf die Kurzbeschreibung der Kategorien verzichtet.

Tabelle 31: Auftrittshäufigkeiten der Kategorien in Prozent pre und post sowie die Änderung in Prozentpunkten und die prozentuale Änderung für die Gesamtstichprobe

Kategorie	pre MW (SD)	post MW (SD)	prozentuale Änderung	Änderung Proz.punkte	*p*-Wert (*W*)
S1	.47 (.52)	.67 (.49)	.43	.20	.188 (1)
S1A ~	.67 (.49)	.67 (.49)	.00	.00	.637 (4)
S2	.67 (.49)	.40 (.51)	-.40	-.27	.109 (1)
S2A	.47 (.52)	.33 (.49)	-.29	-.13	.313 (1)
S3	.67 (.49)	.67 (.49)	.00	.00	.656 (3)

S3A ~	.80 (.41)	.67 (.49)	-.17	-.13	.363 (3)
S3B	.13 (.35)	.27 (.46)	1,01	.13	.313 (1)
S4 ~	.87 (.35)	.87 (.35)	.00	.00	.688 (2)
S5	.07 (.26)	.27 (.46)	2,99	.20	.188 (1)
S6	.47 (.52)	.60 (.51)	.28	.13	.363 (3)
S7	.67 (.49)	.73 (.46)	.10	.07	.500 (2)
S8	.07 (.26)	.20 (.41)	1,99	.13	.313 (1)
S9	.00 (.00)	.20 (.41)	-	.20	-
S10	.53 (.52)	.60 (.51)	.13	.07	.500 (2)
S11	.07 (.26)	.07 (.26)	.00	.00	.750 (1)
S12	1.00 (.00)	1.00 (.00)	.00	.00	-
S13	.87 (.35)	.73 (.46)	-.15	-.13	.313 (1)
S14 ($N = 12$)	.67 (.49)	.67 (.49)	.00	.00	.500 (0)
S15A	.67 (.49)	.73 (.46)	.10	.07	.500 (2)
S15B ~	.87 (.35)	.87 (.35)	.00	.00	.750 (1)
R1	.13 (.35)	.27 (.46)	1,01	.13	.344 (2)
R2($N = 12$)	.17 (.39)	.33 (.49)	.99	.17	.500 (2)
R3($N = 14$)	.29 (.47)	.21 (.43)	-.25	-.07	.500 (2)
R4($N = 12$)	.00 (.00)	.08 (.29)	-	.08	-
R5($N = 12$)	.42 (.52)	.58 (.52)	.40	.17	.323 (1)
R6 ~ ($N = 14$)	.93 (.26)	.93 (.27)	.00	.00	.750 (1)
R7($N = 12$)	.58 (.52)	.83 (.39)	.43	.25	.250 (0)
R8	.07 (.26)	.13 (.35)	.99	.07	.500 (1)
R9	.13 (.35)	.20 (.41)	.50	.07	.500 (0)
R10	.13 (.35)	.33 (.49)	1,50	.20	.227 (2)
R11	.00 (.00)	.07 (.26)	-	.07	-
R12	.00 (.00)	.00 (.00)	-	.00	-
K1 ~	.80 (.41)	.93 (.26)	.17	.13	.250 (0)
K2 ~	.67 (.49)	.60 (.51)	-.10	-.07	.500 (2)
K3	.33 (.49)	.20 (.41)	-.40	-.13	.344 (2)
K4	.13 (.35)	.13 (.35)	.00	.00	.750 (2)
K5	.00 (.00)	.13 (.35)	-	.13	-
K6	.40 (.51)	.40 (.51)	.00	.00	.656 (3)
K7	.13 (.35)	.20 (.41)	.50	.07	.500 (2)
K8	.07 (.26)	.00 (.00)	-1,00	-.07	-
K9	.13 (.35)	.40 (.51)	2,01	.27	.145 (2)
K10 ~	.20 (.41)	.53 (.52)	1,67	.33	.063* (1)
K11	.13 (.35)	.47 (.52)	2,51	.33	.063* (1)
K12	.67 (.49)	.60 (.51)	-.10	-.07	.500 (2)
		Summe Änderung Prozentpunkte		2.6	

Für die Gesamtstichprobe zeigen die Kategorien K10 „L gibt Versuchsanleitung, welche kochrezeptartig strukturiert ist“ und K11 „L weist auf die Möglichkeit der Inanspruchnahme von Hilfen hin“ jeweils signifikante Unterschiede zwischen der pre- und post-Erhebung. Diese Signifikanzen stehen jedoch mit einem Wert von jeweils $p = .063$ unter Vorbehalt.

Als einzige Kategorie tritt R11 (‚Das Experimentieren wird explizit eingeübt') in der Gesamtstichprobe weder vor noch nach der Intervention auf. Erstmalig nach der Fortbildung treten die Kategorien S9 (‚Kontrollansatz: Aufforderung zur Umsetzung') und R11 (‚Klären von Funktion und Ausgestaltung eines Kontrollansatzes') auf. Die letzte Zeile zeigt erneut die Summe der Änderungen der Prozentpunkte über alle Kategorien. Diese Zeile gibt einen Hinweis, dass die Gewinne der einzelnen Kategorien die Verluste in anderen Kategorien in der Summe mindestens aufheben. Die Summation der Prozentpunkte hat hierbei rein illustrativen Charakter und ist keine gängige statistische Aussage.

Oberkategorien und Dimensionen

Zur Berechnung der Signifikanz der einzelnen Oberkategorien und Dimensionen wird erneut der exakte Binomial-Test mit den Parametern $\pi = 1/2$, $N = B+C$ und $x = min(B,D)$ durchgeführt. Die Verwendung des McNemar-Tests ist aufgrund häufig zu schwach besetzter Zellen B und C (vgl. Tabelle 9) wieder nicht durchgehend möglich.
Der Vergleich der *p*-Werte des McNemar-Tests und Binomial-Tests bei der Anwendung auf der gleichen Stichprobe zeigt, dass beide Tests zu vergleichbaren Resultaten führen, eine Abweichung des *p*-Werts tritt meist erst ab der dritten Nachkommastelle auf. Aus diesem Grund wird im Folgenden auf den Binomial-Test ausgewichen, aus Gründen der Vergleichbarkeit auch dort, wo ein McNemar-Test gerechnet werden könnte.
Aufgrund der Stichprobengrößen ($N = 15$, $n_{min} = 6$) wird das Signifikanzniveau erneut auf $p < .1$ angehoben. Diese Entscheidung wird analog zu der Argumentation bezüglich des Signifikanzniveaus für die einzelnen Kategorien auf Seite 112 begründet.
Für die einzelnen Oberkategorien (die Zusammenfassung einzelner Kategorien) und Dimensionen (die Zusammenfassung von Oberkategorien) werden im Folgenden die Ergebnisse dargestellt. Die Ergebnisse beinhalten den Mittelwert (arithmetisches Mittel) und die Standardabweichung für die pre- und post-Erhebung, den mittels exaktem Binomial-Test errechneten *p*-Wert sowie als Maß der Effektstärke Cohens *d*. Die Effektstärke d wird aufgrund der für die Berechnung der Signifikanz oft geringen Gruppengröße im Folgenden stets ebenfalls berichtet. Folgt man der Interpretation nach Cohen (1988), ergeben sich folgende Effektstärken:

Tabelle 32: Interpretation von *d* nach Cohen (1988)

Kleiner Effekt	$\|d\| \geq 0.2$
Mittlerer Effekt	$\|d\| \geq 0.5$
Großer Effekt	$\|d\| \geq 0.8$

Die hier berichteten Effektstärken dienen jedoch nur zur Orientierung und sind nicht als absolut zu interpretieren (Cohen, 1988).

Die jeweils zu den Dimensionen zugehörigen Oberkategorien können in Kapitel 3.5.2 nachvollzogen werden.

Tabelle 33: Oberkategorie ‚Problem und Fragestellung'

Mittelwerte (MW), Standardabweichungen (SD), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig) und Effektstärke (Cohens *d*)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*W*)**	**Cohens *d***
KG (n = 9)	.556 (.504)	.444 (.504)	.212 (5)	-.222
TG (n = 6)	.583 (.504)	.625 (.495)	.500 (4)	.084
Gesamt (N = 15)	.567 (.5)	.517 (.504)	.339 (10)	-.100

Die Oberkategorie ‚Problem und Fragestellung' beinhaltet die vier Kategorien S1–S2A (ein Problem (als Anstoß einer Untersuchung) wird aufgeworfen/identifiziert; ein Problem wird aufgeworfen/identifiziert („Ingenieurmodus"); eine Fragestellung wird präsentiert/hergeleitet; eine naturwissenschaftliche Fragestellung wird präsentiert/hergeleitet). Weder in den Teilstichproben noch in der Gesamtstichprobe ändert sich das Unterrichthandeln signifikant.

Tabelle 34: Oberkategorie ‚Hypothesen aufstellen und besprechen'

Mittelwerte (MW), Standardabweichungen (SD), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig) und Effektstärke (Cohens *d*)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*W*)**	**Cohens *d***
KG (n = 9)	.533 (.505)	.533 (.505)	.605 (7)	.000
TG (n = 6)	.467 (.507)	.567 (.504)	.291 (5)	.198
Gesamt (N = 15)	.507 (.503)	.547 (.501)	.351 (12)	.080

Die Kategorien S3 bis S5 (Hypothesen, Vermutungen, Ideen) werden (im Kontext eines Versuchs) aufgestellt; Hypothesen werden (im Kontext eines Versuchs) aufgestellt -> ausgerichtet auf Beobachtung; Hypothesen werden (im Kontext eines Versuchs) aufgestellt -> mit Begründung; problematischer Hypothesenbegriff von L: Aufforderung zum Raten; nicht wissenschaftlich; Zweck des Versuchs; Begründungen für Hypothesen (werden von L eingefordert) bilden die Oberkategorie ‚Hypothesen aufstellen und besprechen'. In dieser Oberkategorie zeigen sich sowohl in den Teil- als auch der Gesamtstichprobe keine signifikanten Veränderungen des Unterrichtshandelns.

Tabelle 35: Oberkategorie ‚Entwickeln einer Untersuchung'

Mittelwerte (MW), Standardabweichungen (SD), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig) und Effektstärke (Cohens *d*)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*W*)**	**Cohens *d***
KG (n = 9)	.611 (.502)	.667 (.485)	.500 (4)	.113
TG (n = 6)	.5 (.522)	.667 (.492)	.344 (2)	.329
Gesamt (N = 15)	.567 (.504)	.667 (.479)	.304 (6)	.203

In der Oberkategorie ‚Metareflexion' sind die Kategorien ‚Entwickeln einer Untersuchung' und die beiden Kategorien S6 und S7 (Ideen zur Durchführung werden erfragt/genannt/gesammelt; Durchführung/Planung wird (vor-)besprochen) zusammengefasst. Bezüglich dieser Oberkategorie gibt es weder in den beiden Teilstichproben noch in der Gesamtstichprobe signifikante Effekte im Unterrichtshandeln.

Tabelle 36: Oberkategorie ‚Wissenschaftsorientiertes Arbeiten anregen'

Mittelwerte (MW), Standardabweichungen (SD), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig) und Effektstärke (Cohens *d*)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*W*)**	**Cohens *d***
KG (n = 9)	.139 (.351)	.278 (.454)	.090* (2)	.343
TG (n = 6)	.208 (.415)	.25 (.442)	.500 (2)	.098
Gesamt (N = 15)	.167 (.376)	.267 (.446)	.009** (4)	.242

Die Oberkategorie ‚Wissenschaftsorientiertes Arbeiten anregen' beinhaltet die Kategorien S8 bis S11 (Achtung einer Vergleichbarkeit; Kontrollansatz: Aufforderung zur Umsetzung; Aufforderung/Hinweis zu protokollieren; Einschränkungen und Generalisierbarkeit der Beobachtung/Ergebnisse (Vorläufigkeit d. Wissens, Grenzen) werden diskutiert). Bei der Kontrollgruppe zeigt sich ein kleiner bis mittlerer ($d = .343$) signifikanter ($p = .090$) Effekt im Unterrichtshandeln. Die Gesamtgruppe ändert ihr Unterrichtshandeln somit hoch signifikant ($p = 0.09$) mit einem kleinen Effekt ($d = .242$).

Die Lehrkräfte – vornehmlich aus der Kontrollgruppe – regen nach der Fortbildung die Lernenden vermehrt zu einem wissenschaftsorientierten Arbeiten an.

Tabelle 37: Oberkategorie ‚Fixierung im Erkenntnisprozess'

Mittelwerte (MW), Standardabweichungen (SD), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig) und Effektstärke (Cohens *d*)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*W*)**	**Cohens *d***
KG (n = 9)	.791 (.412)	.767 (.427)	.500 (2)	-.060
TG (n = 6)	.862 (.379)	.862 (.351)	.500 (4)	.000
Gesamt (N = 15)	.819 (.387)	.806 (.399)	.387 (5)	-.041

Die Oberkategorie ‚Fixierung im Erkenntnisprozess' beinhaltet die beiden Kategorien S12 –S15B (Erfragen und besprechen von Beobachtungen; Festhalten von Beobachtungen/ (Mess-)Ergebnissen; Erkenntnisse werden notiert; Teilprozesse werden im Prozess schriftlich festgehalten; Teilprozesse werden abschließend schriftlich festgehalten). Es zeigen sich in dieser Oberkategorie in allen drei Gruppen keine signifikanten Änderungen des Unterrichtshandelns. Das ist möglicherweise mit einem Deckeneffekt zu erklären: Bereits in der pre-Erhebung werden Mittelwerte um .800 erzielt.

Tabelle 38: Dimension ‚Prozessstruktur'

Mittelwerte (MW), Standardabweichungen (SD), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig) und Effektstärke (Cohens *d*)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*W*)**	**Cohens *d***
KG (n = 9)	.528 (.501)	.534 (.500)	.500 (25)	.012
TG (n = 6)	.533 (.501)	.597 (.493)	.174 (17)	.129
Gesamt (N = 15)	.530 (.500)	.559 (.497)	.233 (42)	.058

In der Dimension ‚Prozessstruktur' sind alle zuvor beschriebenen Oberkategorien zusammengefasst, sie beinhaltet somit die Kategorien S1–S15B. Weder in einer der beiden Teilstichproben noch in der Gesamtstichprobe zeigen sich signifikante Änderungen.

Tabelle 39: Oberkategorie ‚Mesoreflexion'

Mittelwerte (MW), Standardabweichungen (SD), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig) und Effektstärke (Cohens *d*)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*W*)**	**Cohens *d***
KG (n = 9)	.389 (.492)	.426 (.499)	.500 (6)	.070
TG (n = 6)	.334 (.475)	.514 (.507)	.055* (2)	.366
Gesamt (N = 15)	.363 (.483)	.462 (.501)	.105 (17)	.201

Die Oberkategorie ‚Mesoreflexion' beinhaltet die Kategorien R1–R7 (Vermutungen/Ideen werden zusammengefasst/wiederholt; Vermutungen werden verifiziert/falsifiziert; Erfragen von Problemen bei/nach der Durchführung; Besprechung von Verbesserungsmöglichkeiten nach der Durchführung; Berichten und Besprechen nach Durchführung; Kommentierung/Besprechung von Produkten und Objekten; Erfragen und besprechen von Erkenntnissen). In der Testgruppe zeigt sich ein kleiner bis mittlerer (d = .366), signifikanter (p = .055) Effekt zwischen beiden Erhebungen.
Die Lehrkräfte erweitern also ihr Unterrichtshandeln – vornehmlich in der Testgruppe – um Aspekte der Mesoreflexion.

Tabelle 40: Oberkategorie ‚Metareflexion'

Mittelwerte (MW), Standardabweichungen (SD), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig) und Effektstärke (Cohens *d*)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*W*)**	**Cohens *d***
KG (n = 9)	0 (0)	.089 (.288)	.063* (0)	.440
TG (n = 6)	.167 (.379)	.233 (.43)	.344 (2)	.163
Gesamt (N = 15)	.067 (.251)	.147 (.356)	.055* (2)	.260

In der Oberkategorie ‚Metareflexion' sind die Kategorien R8–R12 zusammengefasst (Begriffsklärung/Problematisierung „Vermutung"; Wiederholung Teilprozesse; Bezug einzelner Teilprozesse im Prozess aufeinander durch L; Klären von Funktion und Ausgestaltung eines Kontrollansatzes; Das Experimentieren wird explizit eingeübt). In der Kontrollgruppe gibt es einen signifikanten (p = .063), mittleren (d = .440) Effekt auf das Unterrichtshandeln, in der Gesamtstichprobe einen kleinen (*d* = .260), signifikanten (*p* = .055) Effekt.

Es zeigt sich, dass die Lehrkräfte in der Gesamtbetrachtung der beiden Teilstichproben ihr Unterrichtshandeln um Elemente der Metareflexion ergänzen. Die auftretenden Signifikanzen, besonders in der Kontrollgruppe, erklären sich auch durch das niedrige Ausgansniveau (KG: MW = 0).

Tabelle 41: Dimension ‚Reflexion'

Mittelwerte (MW), Standardabweichungen (SD), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig) und Effektstärke (Cohens *d*)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*W*)**	**Cohens *d***
KG (n = 9)	.212 (.411)	.273 (.448)	.304 (6)	.140
TG (n = 6)	.254 (.438)	.388 (.491)	.090* (4)	.288
Gesamt (N = 12)	.229 (.421)	.319 (.468)	.068* (12)	.202

Die Dimension ‚Reflexion' beinhaltet die beiden Oberkategorien ‚Mesoreflexion' und ‚Metareflexion'. Sowohl in der Test- als auch in der Gesamtstichprobe gibt es signifikante (TG: *p* = .090; Gesamt: *p* = .068) Änderungen des Unterrichtshandelns. Fällt der Effekt bei der Gesamtstichprobe (*d* = .202) klein aus, ist der Effekt bei der Testgruppe (*d* = .288) – auf ebenfalls kleinem Niveau – etwas ausgeprägter. Die beobachteten Signifikanzen sind in beiden Fällen jedoch schwach und können allenfalls als Hinweis für Änderungen im Unterrichtshandeln interpretiert werden.

Tabelle 42: Oberkategorie ‚Lehrkraft als Ansprechpartner, Trainer und Experte'

Mittelwerte (MW), Standardabweichungen (SD), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig) und Effektstärke (Cohens *d*)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*W*)**	**Cohens *d***
KG (n = 9)	.306 (.464)	.292 (.458)	.500 (11)	-.030
TG (n = 6)	.333 (.476)	.375 (.489)	.395 (8)	.087
Gesamt (N = 15)	.317 (.467)	.325 (.47)	.500 (19)	.017

Die Oberkategorie ‚Lehrkraft als Ansprechpartner, Trainer und Experte' beinhaltet die Kategorien K1–K8 (L gibt die Beobachtung vor; L greift händisch in die Durchführung ein; Umgang mit/Funktionsweise von Labormaterial; Verweis auf Selbsttätigkeit; Verweis auf Material; Verweis auf AB; Verweis auf Hilfen; Verweis auf andere Gruppen). Sowohl in den Teilstichproben als auch der Gesamtstichprobe gibt es keine statistisch signifikante Änderung des Unterrichtshandelns.

Tabelle 43: Oberkategorie ‚Unterstützte Offenheit'

Mittelwerte (MW), Standardabweichungen (SD), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig) und Effektstärke (Cohens *d*)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*W*)**	**Cohens *d***
KG (n = 9)	.25 (.439)	.361 (.487)	.194 (4)	.239
TG (n = 6)	.333 (.482)	.708 (.464)	.011* (2)	.793
Gesamt (N = 15)	.283 (.454)	.5 (.504)	.007** (6)	.452

Die Oberkategorie ‚Unterstützte Offenheit' beinhaltet die vier Kategorien K9–K12 (Freiheit bzgl. der Materialauswahl; L gibt Versuchsanleitung, welche kochrezeptartig strukturiert ist, aus; L weist auf die Möglichkeit der Inanspruchnahme von Hilfen hin; L gibt S einen Protokollbogen aus). In dieser Oberkategorie verändert sich das Unterrichtshandeln in der Testgruppe signifikant (p = .011) positiv, die Änderung kann, so man nach Cohen (1988) die Signifikanzniveaus tatsächlich als Faustregel und nicht streng mathematisch interpretiert, als großer Effekt (*d* = .793) bezeichnet werden. In der Gesamtstichprobe zeigt sich ein sehr signifikanter (*p* = .007) kleiner bis mittlerer Effekt (*d* = .452).

Die Lehrkräfte ändern also ihr Unterrichtshandeln, vornehmlich in der Testgruppe, merklich hin zu einer ‚Unterstützten Offenheit'.

Tabelle 44: Dimension ‚Angemessenheit und Unterstützung'

Mittelwerte (MW), Standardabweichungen (SD), Signifikanz (*p*-Wert jeweils über exakten Binomial-Test, einseitig) und Effektstärke (Cohens *d*)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*W*)**	**Cohens *d***
KG (n = 9)	.287 (.454)	.315 (.467)	.345 (6)	.061
TG (n = 6)	.333 (.475)	.486 (.503)	.026* (6)	.313
Gesamt (N = 15)	.306 (.462)	.383 (.488)	.035* (13)	.162

In der Dimension ‚Angemessenheit und Unterstützung' sind die Oberkategorien ‚Lehrkraft als Ansprechpartner, Trainer und Experte' und ‚Unterstützte Offenheit' zusammengefasst. Bei der Testgruppe gibt es einen kleinen bis mittleren (d = .313), signifikanten (p = .026) Effekt auf das Unterrichtshandeln. Auch in der Gesamtstichprobe zeigt sich ein signifikanter (p = .035), aber allenfalls kleiner (d = .162) Effekt.

4.2 Phasen des Experimentierprozesses

Im Folgenden sollen die Entwicklungsbewegungen bezüglich der aufgewendeten Unterrichtszeit für die einzelnen Phasen des Experimentierprozesses (Teilprozesse) dargestellt werden. Hierzu werden Daten ausgewertet, welche mithilfe des Kategoriensystems 2 „Phasen des Experimentierprozesses" aufgenommen wurden. Im ersten Abschnitt (4.2.1) wird qualitativ und intraindividuell berichtet, welche Teilprozesse in den Unterrichtsstunden beobachtet wurden. Anschließend (4.2.2) werden die Änderungen der aufgewendeten Unterrichtszeit für die einzelnen Teilprozesse auf der Ebene der Teilstichproben quantitativ berichtet.

4.2.1 Welche Teilprozesse treten auf?

In Tabelle 45 ist für jede untersuchte Lehrkraft dargestellt, welche Teilprozesse in der Unterrichtsstunde überhaupt durchlaufen wurden. Nur wenn der jeweilige Teilprozess in keinem einzigen 15-Sekunden-Intervall auftritt, wurde der Teilprozess als nicht aufgetreten gewertet. Die Reihenfolge der Teilprozesse zeigt einen prototypischen Gang eines Experiments, die Teilprozesse müssen nicht in dieser Reihenfolge durchlaufen worden sein.

Tabelle 45: Auftreten der einzelnen Teilprozesse in den untersuchten Unterrichtsvideografien (pre = weiß, post = grau)

		Problemstellung	Fragestellung	Hypothesenfindung	Planung	Durchführung und Beobachtung	Auswertung	Schlussfolgerung	Präs. und Refl. von Erk.
KG	In1AI	0	1	0	1	1	1	0	1
	In1BI	1	1	1	1	1	1	0	0
	In1CI	1	0	1	0	1	0	0	1
	In2AI	1	1	0	1	1	0	0	1
	In3AI	0	1	0	0	1	1	0	0
	In4AI	1	1	1	1	1	1	1	1
	In4BI	1	0	1	1	1	1	1	1
	In4CI	1	1	0	1	1	1	1	1
	In4DI	1	1	1	1	1	1	1	1
TG	Ko1AI	0	1	1	1	1	1	0	0
	Ko1BI	0	1	1	1	1	1	1	1
	Ko2AI	0	1	0	1	1	1	0	0
	Ko2BI	1	1	0	1	1	1	1	1
	Ko2CI	1	1	1	1	1	1	1	1
	Ko3AI	1	0	1	0	1	1	1	1
KG	In1AII	1	0	0	1	1	0	0	1
	In1BII	1	1	1	1	1	1	0	0
	In1CII	1	1	1	0	1	1	1	0
	In2AII	0	0	0	1	1	1	1	1
	In3AII	1	0	1	1	1	0	1	1
	In4AII	1	1	1	1	1	1	1	1
	In4BII	1	1	1	1	1	1	0	0
	In4CII	0	0	1	1	1	1	1	1
	In4DII	1	0	1	1	1	1	1	1
TG	Ko1AII	1	0	1	1	1	1	0	0
	Ko1BII	1	1	1	1	1	1	1	1
	Ko2AII	1	1	0	1	1	1	0	0
	Ko2BII	0	1	1	1	1	1	1	0
	Ko2CII	0	0	0	0	1	1	1	0
	Ko3AII	1	1	1	1	1	1	1	1

Bei einigen Unterrichtsstunden traten einzelne Teilprozesse, obwohl die Unterrichtseinheiten bei allen Videografien stets neu begonnen wurden, nicht auf. Das Nichtauftreten von Teilprozessen am Anfang des Experiments (Fragestellung, Hypothesenbildung) kann also nicht durch unterrichtsorganisatorische Zwänge erklärt werden. Bezüglich der Teilprozesse ‚Auswertung' und ‚Schlussfolgerung'

kann nicht immer zweifelsfrei geklärt werden, ob die Lehrperson aufgrund mangelnden Verständnisses oder Zeitknappheit die letzten Teilprozesse oder den letzten Teilprozess mit den Lernenden nicht durchläuft. Fehlende Teilprozesse sind also möglicherweise sowohl Indikatoren für ein unvollständiges Verständnis des Experiments als auch unter Umständen unterrichtsorganisatorisch bedingt. Änderungen bezüglich des Auftretens der Teilprozesse im Rahmen der Intervention werden nicht auffällig, wie auch in Tabelle 46 nachzuvollziehen ist:

Tabelle 46: Auftreten der einzelnen Teilprozesse in den Teilstichproben pre (weiß) und post (grau)

	Problemstellung	Fragestellung	Hypothesen-findung	Planung	Durchführung und Beobachtung	Auswertung	Schlussfolgerung	Präs. und Refl. von Erk.
KG	7	6	5	7	9	7	4	7
TG	3	5	4	5	6	6	4	4
gesamt	**10**	**11**	**9**	**12**	**15**	**13**	**8**	**11**
KG	7 (±)	4 (-2)	7 (+2)	8 (+1)	9 (±0)	7 (±0)	6 (+2)	6 (-1)
TG	4 (+1)	4 (-1)	4 (±0)	5 (±0)	6 (±0)	6 (±0)	4 (±0)	2 (-2)
gesamt	**11(+1)**	**8 (-3)**	**11 (+2)**	**13 (+1)**	**15 (±0)**	**13 (±0)**	**10 (+2)**	**8 (-3)**

4.2.2 Analyse der aufgewendeten Unterrichtszeit

Da die Daten metrisch skaliert sind, wurde die Normalverteilung der Daten mittels des Shapiro-Tests geprüft. Eine Normalverteilung der Daten ist für keinen der Teilprozesse gegeben, weswegen immer ein zweiseitiger, exakter Wilcoxon-Test für verbundene Stichproben durchgeführt wird. Ein zweiseitiger Test wurde verwendet, da ‚gewonnene' Unterrichtszeit in einem Teilprozess durch ‚verlorene' Unterrichtszeit aus anderen Teilprozessen kompensiert werden muss. Bei der Anlage der Studie und dem Aufstellen der Hypothesen waren die zu erwartenden Verschiebungen in der Unterrichtzeit unklar, weswegen ungerichtet zweiseitig getestet wird. Ist $n < 15$ wurde bei einer angezeigten Signifikanz noch der T-Wert mit den kritischen Werten aus der Tabelle (Leonhart, 2017, S. 782) geprüft. Dieses Vorgehen wurde stets bei dem Einsatz des exakten Wilcoxon-Tests im Rahmen dieser Arbeit gewählt.

Einführung in das Problem, Fragestellung, Hypothesenfindung und Planung

Bei den Teilprozessen ‚Einführung in das Problem', ‚Fragestellung', ‚Hypothesenfindung' und ‚Planung' findet – verstärkt bei der Testgruppe – sowohl eine Erhöhung als auch eine Individualisierung bzw. Akzentuierung (Erhöhung der Spannbreite) der jeweils aufgewendeten Unterrichtszeit statt. Bezüglich des Teilprozesses der Hypothesenfindung zeigt sich bei der Kontrollgruppe ein signifikanter (p = .031), mittlerer bis starker (d = .668) Effekt. Dieser Effekt ist mit ähnlicher Stärke (d = .710) auch in der Gesamtstichprobe signifikant (p = .021).

Tabelle 47: Unterrichtszeit Teilprozess ‚Einführung in das Problem' (Problemstellung)

Mittelwerte (MW), Standardabweichungen (SD), Effektstärke (Cohens *d*) und Signifikanz (*p*-Wert über exakten, zweiseitigen Wilcoxon-Test)

	pre MW (SD)	**post MW (SD)**	***p*-Wert (*T*)**	**Cohens *d***
KG (n = 9)	.025 (.025)	.015 (.011)	.301 (32)	-.525
TG (n = 6)	.008 (.011)	.040 (.055)	.313 (5)	.800
Gesamt (N = 15)	.018 (.022)	.025 (.036)	.978 (61)	.224

Tabelle 48: Unterrichtszeit Teilprozess ‚Fragestellung'

Mittelwerte (MW), Standardabweichungen (SD), Effektstärke (Cohens *d*) und Signifikanz (*p*-Wert über exakten, zweiseitigen Wilcoxon-Test)

	pre MW (SD)	**post MW (SD)**	***p*-Wert (*T*)**	**Cohens *d***
KG (n = 9)	.010 (.013)	.017 (.026)	.641 (14)	.306
TG (n = 6)	.003 (.002)	.011 (.011)	.156 (3)	1.049
Gesamt (N = 15)	.008 (.010)	.015 (.021)	.233 (33)	.426

Tabelle 49: Unterrichtszeit Teilprozess ‚Hypothesenfindung'

Mittelwerte (MW), Standardabweichungen (SD), Effektstärke (Cohens *d*) und Signifikanz (*p*-Wert über exakten, zweiseitigen Wilcoxon-Test)

	pre MW (SD)	**post MW (SD)**	***p*-Wert (*T*)**	**Cohens *d***
KG (n = 9)	.031 (.039)	.067 (.065)	.031* (1)	.668
TG (n = 6)	.019 (.028)	.109 (.160)	.438 (4)	.788
Gesamt (N = 15)	.026 (.034)	.084 (.109)	.021* (9)	.710

Tabelle 50: Unterrichtszeit Teilprozess ‚Planung'

Mittelwerte (MW), Standardabweichungen (SD), Effektstärke (Cohens *d*) und Signifikanz (*p*-Wert über exakten, zweiseitigen Wilcoxon-Test)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*T*)**	**Cohens *d***
KG (*n* = 9)	.069 (.075)	.111 (.097)	.742 (15)	.485
TG (*n* = 6)	.075 (.084)	.136 (.128)	.844 (9)	.564
Gesamt (*N* = 15)	.071 (.076)	.121 (.107)	.451 (40)	.536

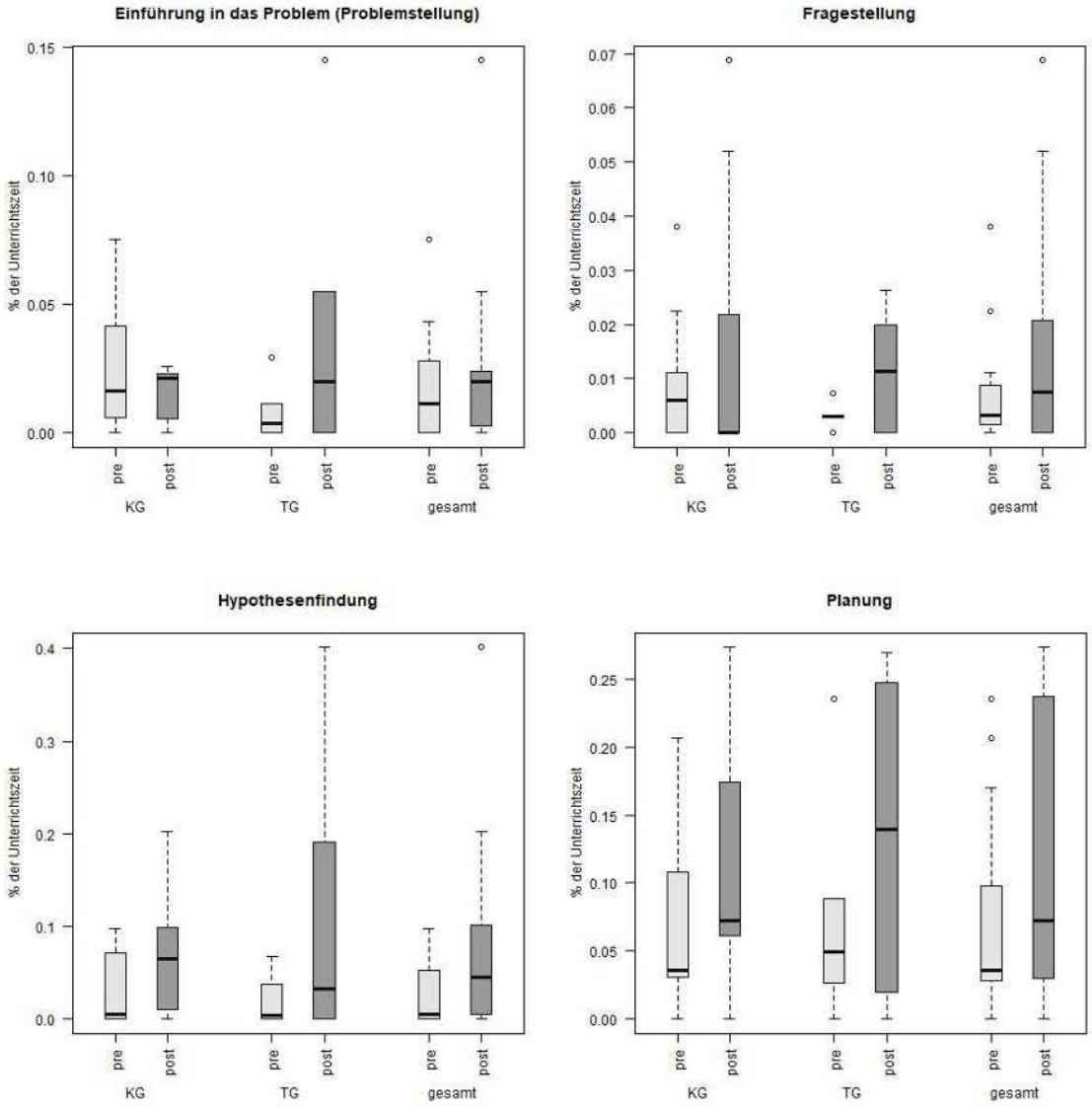


Abbildung 21: Die Teilprozesse ‚Problemstellung', ‚Fragestellung', ‚Hypothesenfindung' und ‚Planung'; KG = Kontrollgruppe, TG = Treatmentgruppe, Gesamt = Gesamtstichprobe. Hinweis: Die Skalierung der Ordinate ist nicht über die Diagramme hinweg fixiert, 1 entspricht 100%.

Durchführung und Beobachtung

Der Teilprozess ‚Durchführung und Beobachtung' gewinnt bei der Unterrichtszeit sowohl in der Testgruppe als auch in der Gesamtstichprobe an Spannbreite, die Lehrkräfte akzentuieren die für diesen Teilprozess aufgewendete Unterrichtszeit

(gewichten Sie nach ihren Bedarfen). Die durchschnittlich verwendete Unterrichtszeit sinkt in allen Gruppen – in der Kontrollgruppe signifikant (p = .098) mit starkem (d = -.779) Effekt.

Tabelle 51: Unterrichtszeit Teilprozess ‚Durchführung und Beobachtung'

Mittelwerte (MW), Standardabweichungen (SD), Effektstärke (Cohens *d*) und Signifikanz (*p*-Wert über exakten, zweiseitigen Wilcoxon-Test)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*T*)**	**Cohens *d***
KG (*n* = 9)	.342 (.133)	.240 (.123)	.098* (37)	-.799
TG (*n* = 6)	.334 (.104)	.268 (.223)	.563 (14)	-.380
Gesamt (*N* = 15)	.339 (.118)	.251 (.163)	.107 (89)	-.616

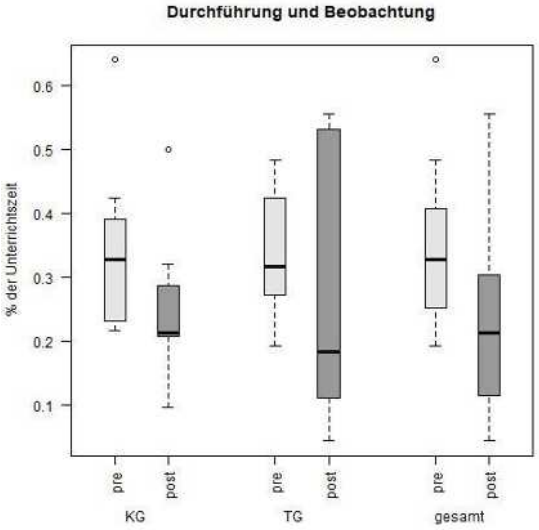


Abbildung 22: Der Teilprozess ‚Durchführung und Beobachtung'; KG = Kontrollgruppe, TG = Treatmentgruppe, Gesamt = Gesamtstichprobe

Auswertung, Schlussfolgerung und Präsentation und Reflexion von Erkenntnissen

Bei den drei Teilprozessen ‚Auswertung', ‚Schlussfolgerung' und ‚Präsentation und Reflexion von Erkenntnissen' gibt es in den Teilstichproben und der Gesamtstichprobe bezüglich der verwendeten Unterrichtszeit keine signifikanten Änderungen. Auch die Spannbreiten ändern sich nicht wesentlich.

Tabelle 52: Unterrichtszeit Teilprozess ‚Auswertung'

Mittelwerte (MW), Standardabweichungen (SD), Effektstärke (Cohens *d*) und Signifikanz (*p*-Wert über exakten, zweiseitigen Wilcoxon-Test)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*T*)**	**Cohens *d***
KG (*n* = 9)	.057 (.053)	.095 (.112)	.426 (15)	.444
TG (*n* = 6)	.141 (.125)	.062 (.048)	.156 (18)	-.839
Gesamt (*N* = 15)	.09 (.095)	.082 (.091)	.804 (65)	-.091

Tabelle 53: Unterrichtszeit Teilprozess ‚Schlussfolgerung'

Mittelwerte (MW), Standardabweichungen (SD), Effektstärke (Cohens *d*) und Signifikanz (*p*-Wert über exakten, zweiseitigen Wilcoxon-Test)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*T*)**	**Cohens *d***
KG (*n* = 9)	.038 (.052)	.028 (.028)	.688 (17)	-.255
TG (*n* = 6)	.021 (.020)	.057 (.089)	.375 (2)	.546
Gesamt (*N* = 15)	.032 (.042)	.039 (.059)	.689 (28)	.150

Tabelle 54: Unterrichtszeit Teilprozess ‚Präsentation und Reflexion von Erkenntnissen'

Mittelwerte (MW), Standardabweichungen (SD), Effektstärke (Cohens *d*) und Signifikanz (*p*-Wert über exakten, zweiseitigen Wilcoxon-Test)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*T*)**	**Cohens *d***
KG (*n* = 9)	.077 (.074)	.085 (.087)	1.000 (18)	.109
TG (*n* = 6)	.032 (.025)	.013 (.02)	.250 (9)	-.833
Gesamt (*N* = 15)	.059 (.062)	.056 (.076)	.610 (46)	-.033

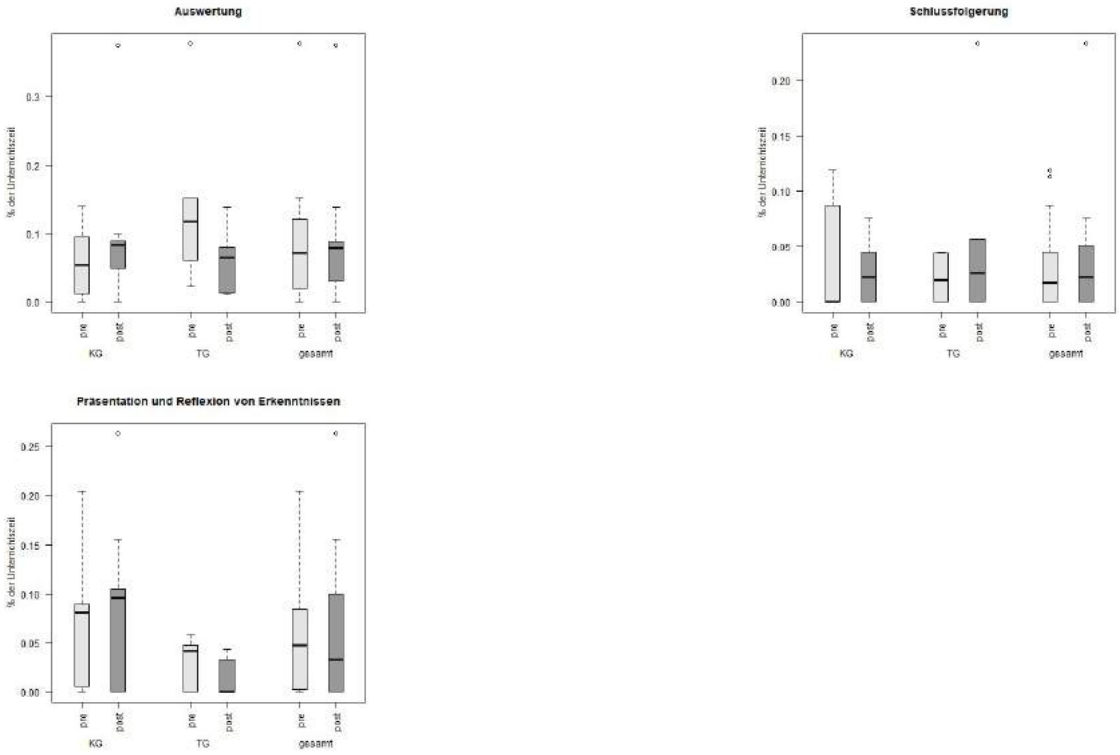


Abbildung 23: Die Teilprozesse ‚Auswertung', ‚Schlussfolgerung' und ‚Präsentation und Reflexion von Erkenntnissen'; KG = Kontrollgruppe, TG = Treatmentgruppe, Gesamt = Gesamtstichprobe

Weitere Kategorien

Die folgenden vier Kategorien stehen nicht direkt mit dem Experimentierprozess in Verbindung. Um jedoch der Exhaustivität des Kategoriensystems (jedes Unter-

richtsgeschehen kann einer Kategorie zugeordnet werden) gerecht zu werden, wurden diese Kategorien in das Kategoriensystem aufgenommen. Es treten bezüglich dieser vier Kategorien keine Signifikanzen auf.

Tabelle 55: Unterrichtszeit Teilprozess ‚Fach- und Methodenwissen'

Mittelwerte (MW), Standardabweichungen (SD), Effektstärke (Cohens *d*) und Signifikanz (*p*-Wert über exakten, zweiseitigen Wilcoxon-Test)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*T*)**	**Cohens *d***
KG (*n* = 9)	.104 (.083)	.121 (.125)	.945 (17)	.166
TG (*n* = 6)	.192 (.114)	.116 (.084)	.156 (11)	-.755
Gesamt (*N* = 15)	.139 (.103)	.119 (.107)	.315 (69)	-.188

Organisatorisches

Tabelle 56: Unterrichtszeit Teilprozess ‚Organisatorisches'

Mittelwerte (MW), Standardabweichungen (SD), Effektstärke (Cohens *d*) und Signifikanz (*p*-Wert über exakten, zweiseitigen Wilcoxon-Test)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*T*)**	**Cohens *d***
KG (*n* = 9)	.202 (.055)	.189 (.064)	.652 (27)	-.214
TG (*n* = 6)	.160 (.079)	.146 (.065)	1.000 (11)	-.205
Gesamt (*N* = 15)	.185 (.066)	.172 (.066)	.639 (69)	-.206

Tabelle 57: Unterrichtszeit Teilprozess ‚Kein Unterricht mit naturwissenschaftlichem Bezug'

Mittelwerte (MW), Standardabweichungen (SD), Effektstärke (Cohens *d*) und Signifikanz (*p*-Wert über exakten, zweiseitigen Wilcoxon-Test)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*T*)**	**Cohens *d***
KG (*n* = 9)	.017 (.013)	.031 (.040)	.461 (12)	.476
TG (*n* = 6)	.014 (.017)	.041 (.063)	.313 (3)	.574
Gesamt (*N* = 15)	.016 (.014)	.035 (.049)	.108 (22)	.534

Tabelle 58: Unterrichtszeit Teilprozess ‚Restkategorie' (nicht zuweisbar)

Mittelwerte (MW), Standardabweichungen (SD), Effektstärke (Cohens *d*) und Signifikanz (*p*-Wert über exakten, zweiseitigen Wilcoxon-Test)				
	pre MW (SD)	**post MW (SD)**	***p*-Wert (*T*)**	**Cohens *d***
KG (*n* = 9)	.028 (.068)	0 (0)	.500 (3)	-.577
TG (*n* = 6)	0 (0)	.001 (.002)	1.000 (0)	.577
Gesamt (*N* = 15)	.017 (.053)	0 (.001)	.423 (5)	-.432

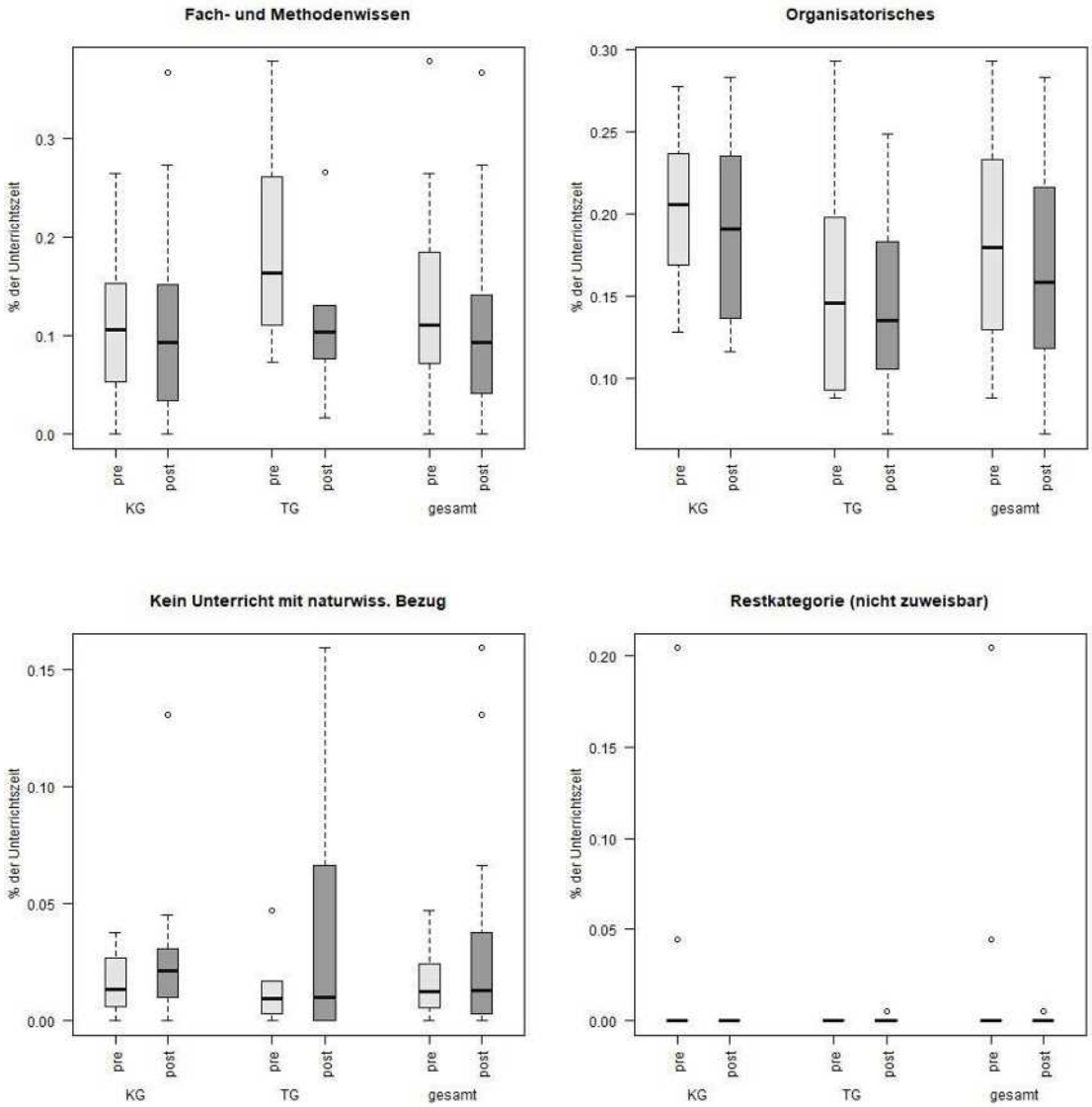


Abbildung 24: Die Kategorien ‚Fach- und Methodenwissen', ‚Organisatorisches', ‚kein Unterricht mit naturwissenschaftlichem Bezug' und ‚Restkategorie' als Boxplot dargestellt. KG = Kontrollgruppe, TG = Treatmentgruppe, Gesamt = Gesamtstichprobe

4.3 Öffnungsgrade der Teilprozesse beim Experimentieren

Im Folgenden sollen die Entwicklungsbewegungen bezüglich der Öffnung der einzelnen Teilprozesse des Experimentierprozesses dargestellt werden. Hierzu werden Daten ausgewertet, welche mithilfe des Kategoriensystems 3 „Öffnungsgrade der Teilprozesse beim Experimentieren" aufgenommen wurden.

Die Daten sind ordinal skaliert, weswegen stets ein zweiseitiger, exakter Wilcoxon-Test für verbundene Stichproben durchgeführt wird. Bei der Anlage der Studie und dem Aufstellen der Hypothesen waren die zu erwartenden Verschiebungen in den Öffnungsgraden unklar, weswegen ungerichtet zweiseitig getestet wird. Eine Entwicklung hin zu einer Öffnung des gesamten Experimentierprozesses wurde zwar von den Fortbildnern als Ziel ausgegeben (ein Argument für eine

einseitige Testung), stets wurde aber betont, dass auch das begründete Schließen einzelner Teilprozesse legitim ist.

Bei einigen Unterrichtsstunden traten einzelne Teilprozesse, obwohl der Experimentierprozess bei allen Videografien stets neu angestoßen wurde, nicht auf (vgl. Tabelle 45):

In der pre-Erhebung wurde der Teilprozess ‚Fragestellung' in vier Fällen (KG: 2; TG: 2), in der post-Erhebung sogar in sieben Fällen (KG: 5; TG: 2) nicht berücksichtigt. Eine Hypothesenbildung konnte in der pre-Erhebung in fünf Fällen (KG: 4; TG: 1) nicht beobachtet werden. In der post-Erhebung tritt dieser Teilprozess viermal nicht auf (KG: 2; TG: 2). Der Teilprozess ‚Planung-Experiment' tritt in allen untersuchten Unterrichtstunden auf.

Bezüglich der Teilprozesse ‚Auswertung' und ‚Schlussfolgerung' kann nicht immer zweifelsfrei geklärt werden, ob die Lehrperson aufgrund mangelnden Verständnisses oder Zeitknappheit die letzten beiden Teilprozesse (‚Auswertung' und ‚Schlussfolgerung') oder den letzten Teilprozess (‚Schlussfolgerung') mit den Lernenden nicht durchläuft. Wird der Teilprozess ‚Auswertung' (pre: KG: 2; TG: 0; post: KG: 0; TG: 1) ausgelassen, fehlt aber auch stets der Teilprozess Schlussfolgerung. Eine Schlussfolgerung wird in der prä-Erhebung in drei Fällen (KG: 3; TG: 0) und in der post-Erhebung in vier Fällen (KG: 3; TG: 1) nicht gezogen.

Fehlende Teilprozesse sind also möglicherweise Indikatoren für ein unvollständiges Verständnis des Experiments oder unter Umständen unterrichtsorganisatorisch bedingt.

Diese nicht wertbaren Öffnungsgrade wurden, zusammen mit der korrespondierenden Unterrichtsstunde der Lehrkraft, aus dem Datensatz des jeweiligen Teilprozesses entfernt. Dies führt teils zu stark reduzierten Stichproben und damit einhergehend zu einer limitierten Aussagekraft der Ergebnisse. Die Ergebnisse sind tabellarisch in Tabelle 59 und grafisch in Abbildung 25 dargestellt.

Der Öffnungsgrad ist ordinalskaliert, doch wurde durch die Bindung des Öffnungsgrades an eine numerische Skala (0 - geschlossen, 1 - leicht geöffnet , 2 - geöffnet, 3 - offen) eine Gleichabständigkeit, wie für metrische Daten gefordert, von Baur & Emden (2020) implizit berücksichtigt. Damit ist, wenn auch aus dem Blickwinkel der Statistik kritisch zu betrachten, eine Anwendung metrischer Statistiken (Mittelwert, Standardabweichung) prinzipiell möglich und führt auch zu interpretierbaren Resultaten. Der Gewinn an Informationsgehalt aus den metrischen Statistiken ist hier pragmatisch höher einzuschätzen als ein puristisches Festhalten an Skalenniveaus da aus den metrischen Statistiken die Messwertverteilung in der Population sichtbar wird. Aus diesem Grund werden bei der Auswertung der

Öffnungsgrade der Teilprozesse auch der Mittelwert, die Standardabweichung und die Effektstärke berichtet.

Unter ‚Total' wurden alle die Öffnungsgrade aller fünf Teilprozesse der jeweiligen Stichprobe aufsummiert und durch das Dividieren durch die Anzahl der Teilprozesse normiert. Damit wird die durchschnittliche Öffnung über den Gesamten Experimentierprozess für die einzelnen Stichproben dargestellt. In der Grafik ‚Gruppenvergleich' werden diesbezüglich die einzelnen Stichproben nochmals vergleichend dargestellt.

Tabelle 59: Entwicklung des Öffnungsgrades der Teilprozesse beim Experimentieren

Mittelwerte (MW), Standardabweichungen (SD), Effektstärke (Cohens *d*) und Signifikanz (*p*-Wert über exakten, einseitigen Wilcoxon-Rangsummen-Test)

Fragest.		**pre MW (SD)**	**post MW (SD)**	***p*-Wert (*T*)**	**Cohens *d***
	KG (*n* = 3)	0 (0)	1 (1)	.250 (0)	1.414
	TG (*n* = 4)	.5 (1)	.25 (.5)	1.000 (1)	-.316
	Ges. (*N* = 7)	.25 (.71)	.57 (.79)	.500 (2.5)	.430
Hypothese					
	KG (*n* = 5)	2.2 (.84)	2.6 (.55)	.375 (1.5)	.566
	TG (*n* = 3)	1.4 (1.34)	2.5 (.58)	.500 (1)	1.065
	Ges. (*N* = 8)	1.8 (1.14)	2.56 (.53)	.250 (4)	.854
Planung-					
Experiment	KG (*n* = 9)	1 (1.23)	1.11 (1.05)	.500 (9)	.097
	TG (*n* = 6)	.67 (1.21)	1.83 (.98)	.094* (1.5)	1.058
	Ges. (*N* = 15)	.87 (1,19)	1.4 (1.06)	.110 (18)	.475
Auswert.					
	KG (*n* = 7)	1.25 (1.5)	1.67 (1.37)	.750 (8)	.290
	TG (*n* = 5)	1.33 (1.03)	2.4 (.55)	.125 (0)	1.290
	Ges.(*N* = 12)	1.3 (1.16)	2 (1.10)	.250 (10.5)	.621
Schlussfolg.					
	KG (*n* = 3)	1.67 (1.53)	1.67 (.58)	.500 (1)	.000
	TG (*n* = 5)	1.2 (.45)	1.6 (.55)	.250 (0)	.800
	Ges. (*N* = 8)	1.375 (.916)	1.625 (.52)	.250 (2.5)	.336
Total					
	KG (*n* = 27)	1.2 (1.26)	1.577 (1.10)	.162 (63.5)	.319
	TG (*n* = 23)	1.04 (1.04)	1.75 (.99)	.022 (18)	.702
	Ges. (*N* = 50)	1.12 (1.14)	1.66 (1.04)	.014* (147)	.496

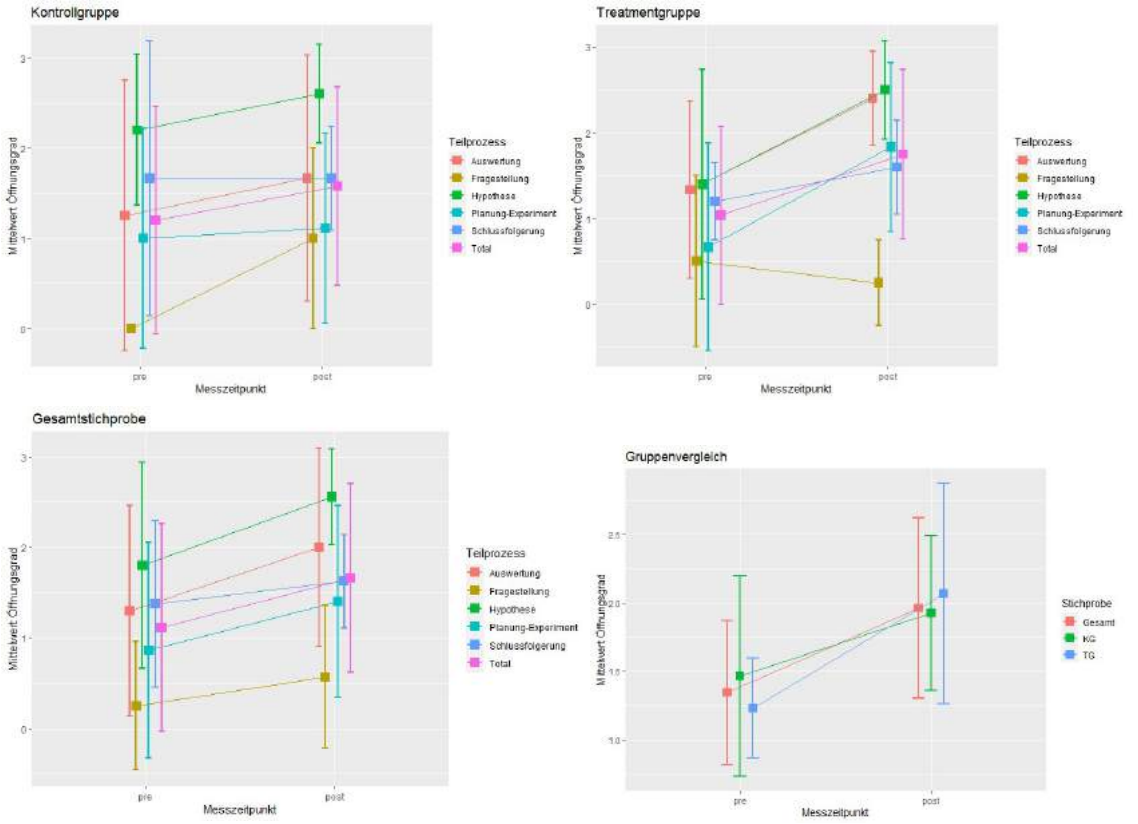


Abbildung 25: Die Entwicklung der Öffnungsgrade für die Teilstichproben sowie die Gesamtstichprobe. Der jeweilige Stichprobenumfang ist in Tabelle 60 abgebildet

In allen Teilprozessen aller Stichproben – mit Ausnahme des Teilprozesses ‚Schlussfolgerung' in der Kontrollgruppe und des Teilprozesses ‚Fragestellung' in der Treatmentgruppe – gibt es zwischen den beiden Messzeitpunkten eine Zunahme des Öffnungsgrades. Bezüglich des Teilprozesses ‚Planung-Experiment' zeigt sich ein signifikanter (p = .094), großer (d = 1.058) Effekt bezüglich der Öffnung. Auch die Gesamtstichprobe zeigt, über den gesamten Experimentierprozess gesehen, eine signifikante (p = .014), mittlere (d = .496) Zunahme des Öffnungsgrades.

4.4 Definition des Experiments

Im Folgenden werden die Entwicklungsbewegungen bezüglich der Passung des durchgeführten Experiments mit Aspekten der Definition eines Experiments geprüft. Hierzu werden Daten ausgewertet, welche mithilfe des Kategoriensystems 4 „Definition des Experiments" aufgenommen wurden. Die Daten sind ordinal skaliert, ein zweiseitiger, exakter Wilcoxon-Test für verbundene Stichproben wird durchgeführt.

Bei einer Unterrichtsstunde in der Kontrollgruppe konnten keine Aspekte des Experiments gewertet werden (Kategorie ‚Nicht anwendbar'). Die entsprechende Videografie zusammen mit der korrespondierenden Unterrichtsstunde der Lehrkraft wurde aus dem Datensatz entfernt.

Wieder liegt, genau wie bei dem Kategoriensystem „Öffnungsgrade der Teilprozesse beim Experimentieren“ eine Ordinalskalierung vor, wieder wurde durch die Bindung des Öffnungsgrades an eine numerische Skala (1 – voll erfüllt, 2 – zum Teil erfüllt, 3 – nicht erfüllt) eine Gleichabständigkeit, wie für metrische Daten gefordert, impliziert. Damit ist, wenn auch aus dem Blickwinkel der Statistik kritisch zu betrachten, eine Anwendung metrischer Statistiken (Mittelwert, Standardabweichung) prinzipiell möglich und führt auch zu interpretierbaren Resultaten.

Der Gewinn an Informationsgehalt aus den metrischen Statistiken ist hier pragmatisch höher einzuschätzen als ein puristisches Festhalten an Skalenniveaus da aus den metrischen Statistiken die Messwertverteilung in der Population sichtbar wird. Aus diesem Grund werden bei der Auswertung der Aspekte des Experiments auch der Mittelwert, die Standardabweichung und die Signifikanz berichtet.

Die Ergebnisse sind in Tabelle 60 und Abbildung 26 dargestellt.

Tabelle 60: Auswertung des Kategoriensystems ‚Definition des Experiments'

Mittelwerte (MW), Standardabweichungen (SD) und Signifikanz (*p*-Wert über exakten, einseitigen Wilcoxon-Rangsummen-Test)				
kontrolliertes Experiment		**pre Median (Modus)**	**post Median (Modus)**	***p*-Wert**
	KG (*n* = 8)	1.5 (1)	2 (1)	.383 (17.5)
	TG (*n* = 6)	2 (2)	1 (1)	.125 (13)
	Ges. (*N* = 14)	2 (2)	1 (1)	.096* (13)
Kontrollansatz				
	KG (*n* = 8)	3 (3)	3 (3)	.750 (4)
	TG (*n* = 6)	2.5 (3)	2 (2)	.500 (3.5)
	Ges. (*N* = 14)	3 (3)	2.5 (3)	.594 (3.5)
Testansatz				
	KG (*n* = 8)	1 (1)	1 (1)	.313 (7.5)
	TG (*n* = 6)	1 (1)	1 (1)	1.000 (0)
	Ges. (*N* = 14)	1 (1)	1 (1)	.500 (0)
Wiederholbar				
	KG (*n* = 8)	1 (1)	1 (1)	1.000 (0)
	TG (*n* = 6)	1 (1)	1 (1)	1.000 (0)
	Ges.(*N* = 14)	1 (1)	1 (1)	1.000 (0)
Gleiche Ergebnisse				
	KG (*n* = 8)	1 (1)	1 (1)	1.000 (0)
	TG (*n* = 6)	1 (1)	2 (1)	1.000 (0)
	Ges. (*N* = 14)	1 (1)	1 (1)	1.000 (0)
GESAMT				
	KG (*n* = 27)	1 (1)	1 (1)	.417 (75)
	TG (*n* = 23)	1 (1)	1 (1)	.186 (31.5)
	Ges. (*N* = 50)	1 (1)	1 (1)	1.000 (31.5)

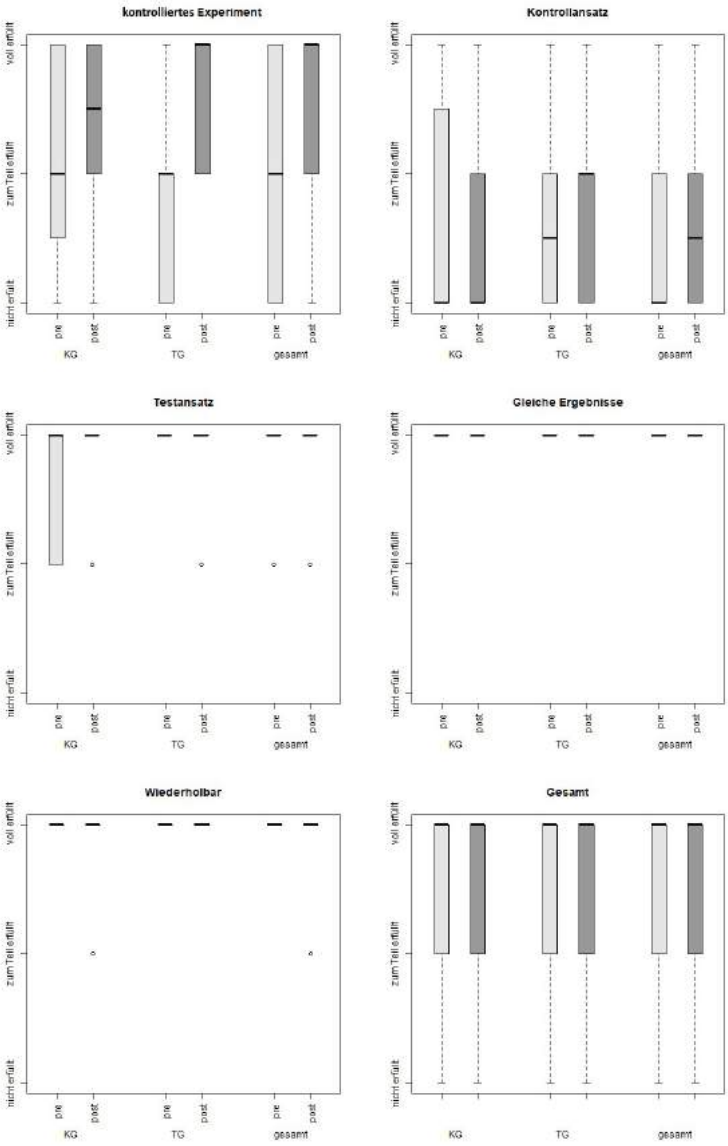


Abbildung 26: Die Boxplots zeigen die Entwicklung der Aspekte des Experiments für die Teilstichproben sowie die Gesamtstichprobe. Unter „Gesamt“ ist die jeweilige Entwicklung aller Aspekte der Definition des Experiments der einzelnen Stichproben nochmals separat dargestellt

Bezüglich der Aspekte der Definition gibt es mit Ausnahme des signifikanten (p = .096) Effekts auf die Kontrolle aller Faktoren eines Experiments (Aspekt ‚kontrolliertes Experiment‘) in der Gesamtstichprobe keine signifikanten Effekte. Median und Modus bleiben in den Kategorien „Testansatz“, „Gleiche Ergebnisse“ und „Wiederholbar“ über den Verlauf der Studie konstant (Ausnahme: TG in der Kategorie „Gleiche Ergebnisse“). Auch anhand der Abbildung 26 wird deutlich, dass speziell die Kategorien „Testansatz“, „Gleiche Ergebnisse“ und „Wiederholbar“ auf dem höchsten der drei Niveaus bis auf wenige Ausreißer über die Intervention hinweg konstant bleiben. Dies deutet darauf hin, dass diese Kategorien des Kategoriensystems möglicherweise zu basal sind, um formale Aspekte aus der Definition des Experiments differenziert zu erfassen (siehe Kapitel 5.4.2 Limitationen der eingesetzten Instrumente).

4.5 Zusammenfassung der Ergebnisse

Im Folgenden werden die umfangreichen Ergebnisse nochmals kondensiert anhand von vier zentralen Ergebnissen der Studie zusammengefasst.

Weniger Kochrezept – mehr Öffnung und Unterstützung

Bezogen auf die Gesamtstichprobe werden signifikant (p = .063) weniger Versuchsanleitungen im Stil von Kochrezepten ausgegeben, dafür wird signifikant (p = .063) häufiger von der Lehrperson im Experimentierprozess auf Hilfen verwiesen. Erweitert um die beiden Kategorien ‚Freiheit bzgl. der Materialauswahl' und ‚L gibt S einen Protokollbogen aus' und somit zu der Oberkategorie ‚Unterstützte Offenheit' zusammengefasst, zeigen sich in dieser Oberkategorie signifikante Änderungen sowohl in der Treatmentgruppe (p = .011; d = .793) als auch der Gesamtstichprobe (p = .007; d = .452).

Dieser Befund einer Abkehr von kochrezeptartigen Anleitungen hin zu offeneren Settings mit Lernunterstützungen wird gestützt durch Ergebnisse des Kategoriensystems ‚Öffnungsgrade der Teilprozesse beim Experimentieren': In der Treatmentgruppe und der Gesamtstichprobe zeigt sich ein signifikanter (TG: p = .022; Gesamt: p = .014) Effekt hin zu einem stärker geöffneten Experimentieren.

Die Öffnung des Experimentierprozesses führt nicht dazu, dass weniger wissenschaftsorientiert experimentiert wird. Das Gegenteil scheint der Fall, trotz Öffnung werden in der Gesamtstichprobe signifikant häufiger (p = .096, Signifikanz unter Vorbehalt da $.1 > p > 0.05$) sowohl Kontroll- als auch Testansatz implementiert. Auch die Ergebnisse des folgenden Abschnitts sprechen gegen eine Öffnung auf Kosten der Wissenschaftsorientierung.

‚Minds on' statt ‚Hands on' – Lehrkräfte regen stärker zu wissenschaftsorientiertem Arbeiten an

In den Oberkategorien ‚Wissenschaftsorientiertes Arbeiten anregen', ‚Mesoreflexion' und Metareflexion' zeigen sich sowohl bezüglich Teilstichproben als auch bezüglich der Gesamtstichprobe signifikante Änderungen im Unterrichtshandeln.

Wird vor der Fortbildung häufig gar nicht zu einem wissenschaftsorientierten Arbeiten aufgefordert, ist dies nach der Intervention vermehrt zu beobachten – es treten dann auch, wenn auch noch mit niedriger Frequenz (3 von 15 Lehrkräfte) Aspekte auf, welche zuvor gar nicht zu beobachten waren: Lehrpersonen forderten etwa nach der Fortbildung überhaupt erst die Lernenden zur Implementation des Kontrollansatzes auf.

Auch die Bezüge und Reflexion innerhalb des Experiments (Mesoreflexion), etwa das Diskutieren von Verbesserungsmöglichkeiten oder der Bezug von Ergebnissen

auf zuvor aufgestellte Hypothesen, nimmt zu: in der Treatmentgruppe zeigt sich ein signifikanter (p = .055), niedriger bis mittlerer (d = .366) Effekt.

Eine Metareflexion, also das Diskutieren über die Erkenntnismethode Experiment und ihre Aspekte, findet, bezogen auf die Gesamtstichprobe und ausgehend von einem sehr niedrigen Niveau, nach der Intervention signifikant (p = .055) häufiger statt (d = .260).

Die verwendete Unterrichtszeit für den Teilprozess ‚Hypothesenbildung' verdoppelt (KG) bzw. verfünffacht (TG) sich, ausgehend von jeweils niedrigem Niveau (KG: 3,1 % auf 6,7 % TG: 1,9 % auf 10,9 % der Unterrichtzeit). In der Kontrollgruppe ebenso wie in der Gesamtstichprobe ist der Effekt jeweils mittel bis hoch (KG: d = .668; Gesamt: d = .710) und signifikant (KG: p = .031; Gesamt: p = .021). Dem Teilprozess ‚Hypothesenbildung' wird also von den Lehrkräften oft erst nach der Fortbildung wesentliche Bedeutung in ihrem Unterrichtshandeln beigemessen.

Individualisierung und differenzielle Entwicklung des Unterrichtshandelns

Über die Clusteranalysen ‚Aufforderung zu wissenschaftsorientiertem Arbeiten', ‚Metareflexion' und ‚Unterstützte Offenheit' zeigt sich jeweils eine Individualisierung des Unterrichtshandelns: Während sich die Unterrichtsstunden der pre-Erhebung in den genannten Oberkategorien stark ähneln (clustern) ist das Unterrichtshandeln in der post-Erhebung stärker individualisiert (akzentuiert). Dieser Befund einer zunehmenden Individualisierung des Unterrichtshandelns zeigt sich auch bei der jeweils auf die einzelnen Teilprozesse aufgewendeten Unterrichtszeit. Bei den Teilprozessen ‚Fragestellung', ‚Hypothesenfindung', ‚Planung' und ‚Durchführung und Beobachtung' streut die Unterrichtszeit deutlich stärker, die Standardabweichung steigt (vgl. entsprechende Boxplots).

Die Lehrkräfte implementieren Inhalte der Lehrkräftefortbildung – augenscheinlich unabhängig von dem Fortbildungstyp – unterschiedlich (vgl. Kapitel 4.1.2). Wenige Lehrkräfte zeigen signifikante positive Entwicklungen ihres Unterrichtshandelns. Eine Lehrkraft (KG) hat sich signifikant verschlechtert.

Das Unterrichtshandeln ist robust gegenüber Interventionen

In vielen Aspekten des Unterrichtshandelns zeigen sich keine signifikanten Änderungen des Unterrichtshandelns. So konnten etwa bezüglich der Hypothesenfindung (‚Hypothesen aufstellen und besprechen') wie auch der Kommunikation (‚Lehrkraft als Ansprechpartner, Trainer und Experte') keine signifikanten Änderungen gemessen werden.

Die meisten Lehrkräfte zeigen bezogen auf die einzelnen Dimensionen nur leicht positive, jedoch nicht signifikante bis gar keine Änderungen ihres Unterrichtshandelns. Mitursächlich für fehlende Signifikanzen können hier auch die geringen

Stichprobengrößen sein. Die positiven signifikanten Veränderungen des Unterrichtshandelns in den einzelnen Teilen der Kategoriensysteme gehen auf wenige Lehrkräfte zurück, welche ihr Unterrichtshandeln signifikant verändern (vgl. auch die Bewegungen in den Clusteranalysen).

5 Diskussion

5.1 Klärung der Hypothesen

Die Hypothese H_1

> „Die Lehrkräftefortbildung führt zu einer signifikanten Veränderung des Unterrichtshandelns“

konnte bestätigt werden, wenngleich die über die Fortbildung angestoßenen Veränderungen nicht in allen untersuchten Teilbereichen zu signifikanten Änderungen geführt haben. Die Alternativhypothese H_0 „Es gibt durch die Lehrkräftefortbildungen keine signifikanten Veränderungen im Unterrichtshandeln der teilnehmenden Lehrkräfte“ kann zurückgewiesen werden.

Eine detaillierte Darstellung der beobachteten Veränderungen findet sich in Kapitel 4.

Die Hypothese H_2

> „Das Fortbildungsformat mit kokonstruktiven Elementen zeigt im Vergleich zum Fortbildungsformat mit vornehmlich individuellen Elementen größere Veränderungen bezüglich des Unterrichtshandelns“

konnte nicht bestätigt werden. Vielmehr deuten die Befunde darauf hin, dass die Hypothese H_0 „Die Änderungen des Unterrichtshandelns der Lehrkräfte ist nicht abhängig von dem Fortbildungsformat“ angenommen werden muss. Hinweise für diese Annahme liefern signifikante Änderungen in jeweils nur *einer* der beiden Teilstichproben sowohl in der Treatmentgruppe (Oberkategorien ‚Mesoreflexion‘, ‚Unterstützte Offenheit‘; Öffnung des Teilprozesses ‚Planung und Experiment‘), als auch in der Kontrollgruppe (Oberkategorien ‚wissenschaftsorientiertes Arbeiten‘ und ‚Metareflexion‘ sowie Teilprozesse ‚Hypothesenfindung‘ und ‚Durchführung und Beobachtung‘). In beiden Teilstichproben sind also signifikante Änderungen – meist sogar bei den einzelnen untersuchten Aspekten – zu beobachten.

Eine weitere Diskussion soll im Folgenden in Bezug auf Aspekte effektiven Unterrichts zum Experimentieren (Kapitel 5.2) sowie hinsichtlich des Lerntransfers (Kapitel 5.3) vorgenommen werden.

5.2 Einordnung und Interpretation der Befunde bzgl. der Aspekte effektiven Unterrichts

In Kapitel 2.3 wurden acht aus vier Unterrichtszielen abgeleitete Unterrichtsprinzipien vorgestellt (vgl. Abbildung 1 und Baur et al., 2019) und anhand empirischer

Befunde weiter legitimiert. Die in der Studie identifizierten Änderungen des Unterrichtshandelns sollen im Folgenden den einzelnen Unterrichtsprinzipien zugeordnet werden. Die Passung der Unterrichtsprinzipien mit Oberkategorien oder Teilaspekten der Kategoriensysteme ist aufgrund der unabhängigen Entwicklungshistorie nicht immer umfänglich gegeben. Aussagen zu den Änderungen des Unterrichtshandelns bezüglich des Erfüllens der Unterrichtsprinzipien können aber dennoch getroffen werden: Tatsächlich ermöglicht eine Betrachtung und Diskussion der Ergebnisse aus der Perspektive der Unterrichtsprinzipien erweiterte und geordnete Einblicke in durch die Fortbildung angeregte Änderungsprozesse des Unterrichtshandelns der Lehrkräfte.

P1: Teilprozesse des Experimentierens berücksichtigen und P8: Selbstständigkeit beim Experimentieren

Sowohl vor als auch nach der Fortbildung berücksichtigen die Lehrkräfte die Teilprozesse im Wesentlichen in gleichen Maßen (vgl. Tabellen 45 und 46). Bezüglich der zeitlichen Gewichtung der einzelnen Teilprozesse tritt aber eine Individualisierung bzw. Akzentuierung ein. Diese Individualisierung betrifft vor allen Dingen die Teilprozesse ‚Fragestellung', ‚Hypothesenfindung' und den Teilprozess ‚Planung' (vgl. steigende Standardabweichung in Tabellen 48–50). Die für diese Teilprozesse aufgewendete Unterrichtszeit steigt über die Intervention in den Teilstichproben, bezüglich des Teilprozesses ‚Hypothesenfindung' in der Kontrollgruppe und der Gesamtstichprobe sogar signifikant (vgl. Einführung in das Problem, Fragestellung, Hypothesenfindung und Planung). Eine Hypothesenbildung findet nach der Fortbildung etwas häufiger statt (pre: 9x, post 11x), vor allen Dingen ist die Phase der Hypothesenbildung in beiden Teilstichproben deutlich länger. Die Hypothesen werden nicht einfach nur von der Lehrkraft vorgegeben oder randständig entwickelt, sondern werden vermehrt dezidiert hergeleitet. Ein Blick auf die Clusteranalyse der Oberkategorie „Hypothesenbildung" zeigt aber auch, dass nur ein Teil der Lehrkräfte diese Entwicklung vollzieht. Einige Lehrkräfte scheinen in Ihrem Muster zu verharren, wogegen bei manchen Lehrkräften auch ein leichter Rückschritt erkennbar wird. Ursächlich dafür können unterschiedliche Transfertypen (vgl. Kapitel 2.4.1 und 5.3) oder aber ein bewusst verschobener Fokus im Unterrichtshandeln sein.

Die Befunde bezüglich der Individualisierung (Teilprozesse werden von der Lehrkraft unterschiedlich akzentuiert) und mehr prozentualer Unterrichtszeit (die Summe des Zeitbudgets einzelner Teilprozesse steigt in der gesamten Stichprobe) lassen sich möglicherweise durch zwei weitere Beobachtungen erklären: Die Lehrkräfte setzten signifikant seltener Experimentieranleitungen im Stil von Kochrezepten ein (vgl. Kapitel 4.1.3), womit eine zunehmende Individualisierung der Länge der einzelnen Teilprozesse erklärt werden kann. Auch nahm die Selbstständigkeit der Lernenden (vgl. Kapitel 4.3) im Experimentierprozess, vornehmlich in

der Treatmentgruppe, aber auch signifikant in der Gesamtstichprobe zu (P8). Die erhöhte Schülerselbstständigkeit (signifikant in der Treatmentgruppe bezüglich des Teilprozesses ‚Planung-Experiment' sowie signifikant in der Treatmentgruppe und der Gesamtstichprobe bezüglich des Gesamtprozesses) erklärt dadurch möglicherweise die Zunahme der prozentualen Unterrichtszeit in den Teilprozessen ‚Fragestellung', ‚Hypothesenfindung' (signifikant in der Kontrollgruppe und der Gesamtstichprobe), ‚Planung' sowie ‚Durchführung und Beobachtung'. Die Verlagerung der Verantwortung einzelner Teilprozesse hin zu den Lernenden erfordert für diese Teilprozesse auch mehr Unterrichtszeit: Bei einem eng vorgegebenen Experimentiersetting im Stil eines Kochrezepts lassen sich die einzelnen Teilprozesse, besonders bis zur eigentlichen Durchführung, zeitlich effizienter durchlaufen – jedoch auf Kosten der Lernendenselbstständigkeit.

P2: Experimentieren als expliziter Unterrichtsgegenstand und P5: Prozessreflektierend experimentieren

Nach der Fortbildung wird die Erkenntnismethode Experiment in der Kontrollgruppe und der Gesamtstichprobe signifikant häufiger reflektiert und es werden innere Bezüge herausgestellt (vgl. Tabelle 40). Viele Lehrkräfte zeigen in der pre-Erhebung überhaupt keine Aspekte einer Reflexion der Erkenntnismethode an sich (vgl. Clusteranalyse, Abbildung 19). Die Besprechung der Funktion des Kontrollansatzes (R11) etwa tritt überhaupt erst in der post-Erhebung auf. Die Teilprozesse ‚Fragestellung', ‚Hypothesenfindung' und ‚Planung' gewinnen an Unterrichtszeit (‚Hypothesenfindung' in der Treatmentgruppe und Gesamtstichprobe signifikant). Einzelne Lehrkräfte scheinen Teilprozesse dabei bewusst hervorzuheben. So verwendet in der post-Erhebung eine Lehrkraft 7 % der Unterrichtszeit auf das Herleiten einer Fragestellung (Max. pre: 4 %), eine andere Lehrkraft verwendet in der post-Erhebung 40 % der Unterrichtszeit auf die Hypothesenbildung inklusive ausführlicher Begriffsklärung (max. pre: 10 %).

Bei einem Teil der Lehrkräfte scheint das Experimentieren also zunehmend nicht mehr nur allein der Rahmen für eine händische Durchführung von Untersuchungen zu sein, sondern die Erkenntnismethode selbst scheint teils zum Unterrichtsgegenstand zu werden.

P3: Experimentieren üben

Dieses Unterrichtsprinzip wurde weder in einer pre- noch post-Erhebung von einer Lehrkraft explizit berücksichtigt. Hierbei muss jedoch beachtet werden, dass der Begriff ‚Üben' vergleichsweise streng aufgefasst wurde. Außer Frage steht, dass auch das wiederholte Experimentieren eine – teils ungerichtete – Art des Übens ist (Wiederholung als Aspekt des Übens, Brinkmann, 2009). Explizite Übungsphasen, etwa das Ordnen von Teilprozessen oder Identifizieren probater Hypothesen

zu naturwissenschaftlichen Fragestellungen auf einem Arbeitsblatt haben jedoch nicht stattgefunden.

Es ist damit das einzige Unterrichtsprinzip, welches bei den Lehrkräften keine Beachtung findet. Das Unterrichtsprinzip war jedoch auch kein expliziter Teil der Lehrkräftefortbildung, sondern wurde erst später bei Bewersdorff et al. (2020) ergänzt. Es ist jedoch keinesfalls ausgesagt, dass eine explizite Übung des Experimentierens bei den teilnehmenden Lehrkräften gar nicht stattfindet. Möglicherweise haben sich die Lehrkräfte dazu entschieden, im Rahmen der Unterrichtsstunden, welche videografiert werden, auf Aspekte des Übens zu verzichten und stattdessen eine Unterrichtstunde vorgezogen, in welcher das Experimentieren zur Anwendung kommt.

P4: Experimentieren zum Problemlösen einsetzen

Die beiden problembezogenen Kategorien S1 und S1A sowie die zugehörigen Oberkategorie ‚Problem und Fragestellung' des Kategoriensystems ‚Unterrichtshandeln von Lehrkräften beim Experimentieren' zeigen in den Teilstichproben und der Gesamtstichprobe keinerlei signifikante Änderungen. Dennoch sind nach der Intervention in allen Stichproben häufiger identifizierte Probleme der Ausgangspunkt des Experiments (vgl. Kategorie S1, Tabellen 29-31). Die für den Teilprozess ‚Einführung in das Problem' verwendete Unterrichtszeit steigt in der Gesamtstichprobe nur geringfügig, in der Kontrollgruppe sinkt sie sogar. Einzig in der Treatmentgruppe ist ein starker, wenn auch nicht signifikanter Effekt zu beobachten (vgl. Tabelle 47 und Abbildung 21).

P6: Grenzen von Experimenten ansprechen

Hinsichtlich der Diskussion von Einschränkungen bezüglich der gewonnenen Ergebnisse sowie der Generalisierbarkeit von Erkenntnissen gibt es im pre-post-Vergleich keine Änderungen. Zu beiden Messzeitpunkten wurden diese Limitationen jeweils nur von einer Lehrkraft in der Kontrollgruppe angesprochen. Diese Stagnation auf niedrigem Niveau zeigt zum einen, dass Entwicklungsbedarf bezüglich der Beachtung dieses Unterrichtsprinzips besteht und zum andern, dass die Fortbildung diesbezüglich keine Sensibilisierung erreichen konnte.

P7: Wissenschaftliche Strenge walten lassen

Wissenschaftliche Strenge im Sinne eines zunehmend wissenschaftsorientierten Arbeitens wird nach der Fortbildung von den Lehrkräften in der Kontrollgruppe sowie auch der Gesamtstichprobe von der Lehrkraft signifikant häufiger angeregt. Die Lehrperson fordert etwa beim Experimentieren mit mehreren Ansätzen nach der Intervention öfter zur Achtung der Vergleichbarkeit (S8) und zum Protokollieren (S10) auf. Es wird in der Treatmentgruppe signifikant öfter auf die Kontrolle

aller Faktoren eines Experiments (Kategoriensystem Definition des Experiments) geachtet.

Nach der Fortbildung wird also der Experimentierbegriff teilweise von Lehrpersonen enger interpretiert. Lehrpersonen scheinen außerdem zunehmend gegenüber den Lernenden wissenschaftliche Strenge einzufordern.

P8: Selbstständigkeit beim Experimentieren

Die Steigerung der Schülerselbstständigkeit beim Experimentieren findet, vor allem bezüglich der Treatmentgruppe, nach der Fortbildung signifikant mehr Beachtung (vgl. die Oberkategorie ‚Unterstützte Offenheit' (Tabelle 43) und das Kategoriensystem ‚Öffnungsgrade der Teilprozesse beim Experimentieren' (Tabelle 59, Abbildung 25)).

5.3 Einordnung und Interpretation der Befunde hinsichtlich des Lerntransfers

Im Folgenden sollen die in der Studie beobachteten Transferformen nach Hasselhorn und Hager (2008) (vgl. Kapitel 2.4.1) benannt werden. Eine Einordnung in die Taxonomie von Joyce & Showers (2002), um den Grad des Transfers genauer zu bestimmen, kann anhand der vorliegenden Ergebnisse (nur eine post-Videografie pro Lehrkraft) nicht gesichert vorgenommen werden (siehe Kapitel 5.4).

Zur Bestimmung der Transferform werden die Ergebnisse aus der Auswertung des Kategoriensystems ‚Unterrichtshandeln von Lehrkräften beim Experimentieren' herangezogen, da nur bei diesem Kategoriensystem die Ergebnisse auch auf der Ebene der Lehrkräfte berichtet werden. Die Prüfung der Ergebnisse der Kategoriensysteme II–IV hinsichtlich der zu beobachtenden Transfergrade wäre Aufgabe einer sich noch anschließenden Analyse.

Der Transfer von Lehrkräftefortbildungen, etwa zu Sicherheitsbestimmungen beim Experimentieren, lässt sich als literaler, spezifischer Transfer charakterisieren (vgl. Kapitel 2.4.1). Erworbene Kompetenzen des Lernkontextes (der Fortbildung) sollen auf Situationen gleichen Typs im Anwendungskontext (der Unterrichtsplanung bzw. -durchführung) übertragen werden. Dieser Transfer wird auch als naher Transfer bezeichnet.

Die in dieser Studie untersuchte Lehrkräftefortbildung zielt hingegen auf einen weiten (distalen) Transfer ab, es liegt– im Falle positiven Transfers und trotz des klaren inhaltlichen Fokus – ein sogenannter unspezifischer Transfer vor.

Mit Blick auf das gesamte Kategoriensystem „Effektiver Unterricht zum Experimentieren" (Tabelle 28) lassen sich alle drei grundlegenden Transferformen „Negativer Transfer", „Nulltransfer" und „Positiver Transfer" ablesen. Die Grenzen

zwischen den einzelnen Transferformen sind fließend, weswegen auf eine eindeutige Einteilung mittels normativer Zuordnungsvorschrift verzichtet wird. Bei mindestens einer Lehrkraft ist anhand der Tabelle 28 bezüglich der Fortbildung ein hoch signifikanter (p = .008) negativer Transfer in die Unterrichtspraxis zu erkennen. Die betroffene Lehrperson unterrichtete zum Zeitpunkt der Studie fachfremd in den Naturwissenschaften und stützte sich deshalb nach eigener Aussage auf bereits existierende Unterrichtsmaterialien. In beiden videografierten Unterrichtsstunden wurden im Wesentlichen in Gruppenarbeit Phänomene entdeckt und beschrieben.

Prozesse der Veränderung des Lehrens und Lernens sind nur mühsam und mit viel Unterstützung zu erreichen und sind störanfällig (vgl. Gräsel & Parchmann, 2004). Dazu passt die Beobachtung, dass viele Lehrkräfte keinen Transfer von Fortbildungsinhalten (‚Nulltransfer') zeigen.

Die dritte Form von Transferwirkungen, Positiver Transfer, konnte ebenfalls in den Ergebnissen beider Teilstichproben beobachtet werden (vgl. Tabellen 25–27): zwei Lehrkräfte zeigen eine signifikante (p = .016 und p = .035) Zunahme von erwünschten Aspekten des Unterrichtshandelns.

Die Ergebnisse der Studie lassen somit den Schluss zu, dass auf der Ebene der Lehrkräfte bezüglich der einzelnen untersuchten Oberkategorien, Dimensionen des Kategoriensystems sowie bezüglich des gesamten Kategoriensystems verschiedene Transferformen aufgetreten sind, welche auch zu unterschiedlichem Transfererfolg geführt haben. Die Heterogenität der Transferwirkungen deckt sich mit dem Befund von Lipowsky (2014), dass Fortbildungen auf die Lehrkräfte nicht in gleicher Weise, sondern vielmehr differenziell wirken. Dieser Befund deckt sich auch mit den Ergebnissen der Zusammenschau von Studien zu den Wirkungen von Lehrkräftefortbildungen zum Themenfeld ‚*inquiry*' (vgl. Kapitel 2.4.2).

Warum zeigen viele Lehrkräfte nach einer Fortbildungsmaßnahme, nicht nur in dieser Studie, keine oder allenfalls marginale Änderungen in ihrem Unterrichtshandeln? Im Folgenden soll ein möglicher Erklärungsansatz aus der Verhaltensökonomie diskutiert werden.

Sind (Lehrkräfte-)Fortbildungen Gemischte Lotterien?

Kahneman, Knetsch und Thaler (1991) konnten zeigen, dass sich bei sogenannten ‚Gemischten Lotterien', in denen sowohl ein Gewinn als auch ein Verlust realisiert werden kann, Menschen stark risikoavers verhalten (Verlustaversion). Mögliche Verluste werden bei einer Entscheidung höher gewichtet als potenzielle Gewinne. Beispielhaft untersuchten Hoorens, Remmers und van de Riet (1999) dieses Phänomen in Bezug auf Arbeitszeit und Einkommen: Die Studienteilnehmenden for-

derten für Überstunden einen deutlich über ihrer Bezahlung liegenden Stundenlohn. Für die Möglichkeit von weniger Arbeitsstunden waren dieselben Studienteilnehmenden jedoch nicht bereit, einen ähnlich starken Einschnitt ihres Gehalts (oder auch nur äquivalent zu ihrem regulären Stundenlohn) zu akzeptieren: Sie bevorzugten jeweils – weitgehend unabhängig vom aktuellen Arbeitspensum und Einkommensniveau – den Status quo, obwohl sie objektiv ihr Verhältnis von Einkommen zu Arbeitszeit jeweils erhöhen konnten. Die Arbeit von Tversky und Kahneman (1992) deutet darauf hin, dass ein Verlust doppelt so stark hemmt wie ein gleichwertiger Gewinn eine Person zum Handeln aktiviert. Auch für andere Bereiche konnten Hinweise auf die Gültigkeit der Theorie gefunden werden.

Lehrkräftefortbildungen zum Unterrichtshandeln sowie ähnlich gelagerte Fortbildungstypen können möglicherweise als eine Art der ‚Gemischten Lotterie' angesehen werden, bei der ebenfalls sowohl ein Gewinn als auch ein Verlust erzielt werden kann. Der Gewinn ist potenziell effektiveres Unterrichtshandeln, welches sich wiederum auf den Lernerfolg der Lernenden auswirken kann (vgl. Wirkungskette der Lehrkräftebildung: Lehrkräftebildung -> Lehrkräftehandeln -> Lernerfahrungen der Lernenden, -> Wirkungen bei den Lernenden; Galluzzo & Craig, 1990). Der Verlust bedeutet den Verzicht auf bereits elaborierte und in den Augen der Lehrkraft bewährte Handlungsskripte. Handlungsskripte sind Routinen, welche Lehrkräfte entwickelt haben, um ihr Unterrichtshandeln im Vorfeld zu strukturieren (Shavelson & Stern, 1981). Die Implikation, dass Lehrkräftefortbildungen ähnlich wie Gemischte Lotterien betrachtet werden können, kann möglicherweise für oft unbefriedigende Effekte von Fortbildungen erklärungsmächtig werden.

Keine Änderung des Unterrichtshandelns aufgrund einer Verlustaversion?

Da die Verlustaversion die Aussicht auf den Gewinn, also ein effektiveres Unterrichtshandeln, weit übersteigt (je nach Setting um das 1,5–2 Fache; Tversky & Kahneman, 1992) neigen Lehrkräfte möglicherweise in Teilen dazu, an ihren tradierten Handlungsskripten festzuhalten, anstatt Fortbildungsinhalte in ihr Unterrichtshandeln zu integrieren.

Falls eine Integration von Fortbildungsinhalten in das Unterrichtshandeln stattfindet, scheinen diese oft oberflächlich (Chin et al., 2016; Marshall & Smart, 2013 und Kapitel 2.4.2) oder werden nur bezüglich einzelner Facetten wirksam (Grigg et al., 2013; Rushton et al., 2017 und Kapitel 2.4.2). Es werden also oft Aspekte des Unterrichtshandels verändert, welche

1. nicht den Kern des Unterrichtshandelns – und damit auch der Lehrpersönlichkeit – berühren,
2. schnell auch wieder rückgängig gemacht werden können.

Je höher das Alter, desto risikoaversiver – also bewahrender – handeln Personen (Gächter, Johnson & Herrmann, 2010). Dieser Befund klärt damit eventuell zu einem gewissen Teil auf, warum die Transferleistung der Fortbildungsteilnehmer heterogen ausfällt. Intuitiv scheint der Befund sinnvoll: Je länger eine Lehrkraft bereits in ihren tradierten Handlungsskripten unterrichtet, desto höher die Verlustaversion bezüglich Änderungen des Unterrichtshandelns. Dazu müsste in einer Reanalyse das Alter der Fortbildungsteilnehmer mit einbezogen werden. Dieser demografische Parameter wurde in der vorliegenden Studie bisher nicht erfasst. Die Metastudie von Blume et al., (2010) wie auch die Studie von Florian (2008) konnten jedoch jeweils keinen Einfluss einer „Background Variable“ (vgl. Kapitel 2.4.1) auf den Transfererfolg identifizieren.

5.4 Limitationen und kritische Punkte der Studie

Die Studienergebnisse unterliegen einer Reihe von Limitationen. Diese sollen im Folgenden berichtet und kritisch diskutiert werden.

5.4.1 Allgemeine Limitationen

Die vorliegende Studie wurde im laufenden Schulbetrieb durchgeführt, die Fortbildung und darüber hinaus noch die freiwillige Videografie stellten für die Lehrkräfte eine nicht geringe, zusätzliche organisatorische Belastung dar. Aufgrund von Krankheit oder terminlichen Überschneidungen konnten nicht alle Lehrkräfte an allen drei Fortbildungstreffen und/oder beiden Hospitationen teilnehmen. Eine Lehrkraft hat an einer der beiden Hospitationen nicht teilgenommen. Die Teilnahme an den Fortbildungstreffen kann nicht mehr für alle Lehrkräfte exakt nachvollzogen werden, von vereinzelter Abwesenheit ist jedoch auszugehen.

Um möglichst zeitliche Überschneidungen der fortbildungsbezogenen Veranstaltungen (Fortbildungstreffen, Hospitationen, Videografien) mit anderen Terminen zu vermeiden, variieren die zeitlichen Abstände zwischen den fortbildungsbezogenen Veranstaltungen für die einzelnen Lehrkräfte in gewissem Maße. Kam es vor der Fortbildung nicht bereits zu einem Erstkontakt, wurde die pre-Erhebung (Videografie) aus ethischen Gründen (Pflicht der ausführlichen Information zum Forschungsvorhaben vor Studienbeginn (Ethik-Rat der DGfE, 2016)), erst nach dem ersten Fortbildungstreffen durchgeführt. Vor der Videografie wurden die Lehrkräfte stets über das Forschungsprojekt und das Ziel der Datenerhebung, -speicherung und -verarbeitung informiert, um anschließend ihre Entscheidung zur freiwilligen Teilnahme treffen zu können. Auch organisatorische Gründe (etwa das Einholen von Einverständniserklärungen bei den Eltern der teilnehmenden Lernenden) führte teils dazu, dass die pre-Erhebung teilweise erst nach dem ersten Fortbildungstreffen durchgeführt wurde. In jedem Fall wurde die pre-Erhebung vor den zwei weiteren Fortbildungstreffen und den beiden Hospitationen durchgeführt. Somit könnten Lehrkräfte Aspekte des Experimentierens, welche zum ersten

Fortbildungstreffen vermittelt wurden, bereits in der pre-Erhebung in ihr Unterrichtshandeln eingebunden haben. Dies führt gegebenenfalls dazu, dass Kompetenzzugewinne bezüglich des Experimentierens, und damit durch die Fortbildung angestoßenen Änderungen des Unterrichtshandelns, nicht vollumfänglich erfasst werden konnten: Möglicherweise sind die durch die Fortbildung angestoßenen Änderungen im Unterrichtshandeln also stärker als berichtet.

Bedingt durch Lehrkräfte, welche nur an einer Videografie teilgenommen haben, sowie die SARS-Cov2-Pandemie, welche zu einem Abbruch der dritten Fortbildungswelle geführt hat, konnte nicht der geplante Studienumfang erreicht werden. Der Stichprobenumfang von $N = 15$ Lehrkräften, welche an beiden Videografien teilgenommen haben, lässt keine allgemeinen Schlüsse, sondern allenfalls Hinweise und Tendenzen auf durch die Fortbildung ausgelöste Änderungen des Unterrichtshandelns zu.

Der Abbruch der dritten Fortbildungswelle hat außerdem zu einer Ungleichverteilung der beiden Teilstichproben geführt. Statt dem mit der Durchführung der 3. Fortbildungswelle angepeilten Verhältnis von 9 (Kontrollgruppe) zu 10 (Treatmentgruppe) Lehrkräften wurde ein Verhältnis von 9 (KG) zu 6 (TG) Lehrkräften erzielt. Diese Schieflage ist besonders beim Vergleich von Signifikanzen relevant, da gilt: Je größer die Stichprobe, desto höher ihre statistische Signifikanz (Metzler & Krause, 1997). Aufgrund der unterschiedlichen Stichprobengrößen der Teilstichproben lassen sich keine eindeutigen vergleichenden Aussagen bezüglich der Signifikanz einzelner Effekte der beiden Teilstichproben treffen.

Lehrkräfte sind vom Dienstherren dazu verpflichtet, sich fortlaufend zu professionalisieren (Konferenz der Kultusminister der Länder, 2020). Dies kann informell durch die eigene Schulpraxis, im Selbststudium, in schulinternen Formaten oder durch externe Fortbildner erfolgen. Effekte dieser parallel zu der Intervention stattfindenden Weiterbildungsmaßnahmen auch auf das Unterrichtshandeln beim Experimentieren sind nicht auszuschließen.

Es ist möglich, dass einzelne Lehrkräfte bei den Videografien nicht ihr eigentliches Unterrichtsverhalten zum Experimentieren zeigten. Sie handelten möglicherweise im Sinne sozialer Erwünschtheit vielmehr so, wie sie denken, dass die Fortbildner/der Forscher das Unterrichtshandeln gerne sehen würden. Die Fortbildner haben versucht, diesen Einfluss durch ihre Bitte, auch während der Hospitationen und Videografien ‚alltäglichen' Unterricht zu halten, zu begrenzen. Ein Effekt der sozialen Erwünschtheit kann dennoch nicht ausgeschlossen werden.

Der Fachinhalt und daraus möglicherweise resultierende Beschränkungen (etwa aufgrund von Sicherheitsaspekten) sowie die Lage der videografierten Unterrichts-

stunde in der Lehr-Lern-Einheit wirken sich auf die jeweilige videografierte Unterrichtstunde aus. Um diesen Einflüssen zu begegnen wurde den Lehrkräften durch ein jeweils mehrwöchiges Zeitfenster für die Videografie hohe Flexibilität eingeräumt. Damit sollten den Lehrkräften jeweils die Möglichkeit gegeben werden, geeignete Unterrichtsstunden zum Experimentieren zu wählen.

5.4.2 Limitationen der eingesetzten Instrumente

Das Unterrichtsprinzip P3 ‚Üben der experimentellen Methode' wird von den eingesetzten Kategoriensystemen nur rudimentär erfasst. Dies hat zwei Ursachen: Im Rahmen der induktiven Entwicklung des Kategoriensystems traten Aspekte des Übens nicht auf oder wurden übersehen. Eine deduktive Ergänzung durch die Auswertung bereits existierender Kodiermanuale (vgl. Kapitel 2.6) ergab, dass Aspekte des Übens in diesen Kodiermanualen nicht berücksichtigt sind, ebenfalls keine dieses Unterrichtsprinzip beschreibende Kategorien. So wurde allein die Kategorie R12 ‚Das Experimentieren wird explizit eingeübt' dem Kategoriensystem I hinzugefügt. Das Nichtauftreten dieser Kategorie in der gesamten Stichprobe lässt sich in zwei Weisen deuten:

1. Aspekte des Übens sind in den Kategoriensystemen nicht sensibel genug erfasst.

2. Das Einüben von Aspekten des Experimentierens findet tatsächlich nicht explizit statt.

Um diese beiden Punkte aufzuklären, wäre weitere Forschung, etwa über die Entwicklung eines weiteren Kategoriensystems speziell zum Üben (oder die Weiterentwicklung der bestehenden Instrumentenbatterie bzw. einzelner Teile) und die Re-Analyse des Materials, nötig.

Die getroffene Ordnung der Kategorien zu Oberkategorien und Dimensionen in Kategoriensystem I ist diskutabel, auch andere Ordnungen sind denkbar und auch für eine Re-Analyse offen. Die Ordnung wurde aufgrund inhaltlicher Passung nach der induktiven Herleitung des Kategoriensystems getroffen und nach der deduktiven Ergänzung nochmals geprüft. Andere Ordnungen können zu anderen Blickwinkeln auf das Unterrichtshandeln führen.

Die Ergebnisse zu den Kategorien „Wiederholbarkeit" und „Gleiche Ergebnisse" des Kategoriensystems „Definition des Experiments" zeigen – sowohl in der pre-Erhebung als auch in der post-Erhebung in beiden Teilstichproben – an, dass die Experimente in den videografierten Unterrichtsstunden fast ausschließlich wiederholbar sind und zu gleichen Ergebnissen führen. Dies ist bei genauerer Betrachtung wenig überraschend. Diese Teilaspekte der Definition eines Experiments, welche allein bereits aufgrund unseres kausalen Weltverständnisses Gültigkeit beanspruchen, verlieren nur schwer beim Experimentieren und eigenständigen Explorieren

und Versuchen im schulischen/alltäglichen Rahmen an Gültigkeit. Bei diesem Kategoriensystem wäre eine Präzision (Setzt die Kategorie „Wiederholbarkeit“ eine wissenschaftliche Strenge voraus oder ist eine theoretische Wiederholbarkeit ausreichend?) und möglicherweise Erweiterung um weitere Teilaspekte des Experiments wie etwa „Es existiert eine naturwissenschaftliche Hypothese“, „Die Schlussfolgerung bezieht sich auf die Hypothese“, sinnvoll.

5.4.3 Limitationen der Auswertung

Die Qualität der Ton- und Bildaufnahmen variiert zwischen den einzelnen Videografien teils deutlich. Wurde zu Beginn der Datenaufnahme noch mit technisch vergleichsweise beschränkten Videokameras gearbeitet, konnte bereits in der ersten Welle nach der ersten Videografie auf eine professionelle Kameraausrüstung zurückgegriffen werden. Aufgrund der technischen Limitationen zu Beginn sind manche Verbalaussagen, vor allem von den Lernenden, in den ersten Videografien teils schwer verständlich oder es können Handlungen nicht genau erkannt werden. Dies schlägt sich auf die Qualität der daraus angefertigten Transkriptionen nieder und wirkt sich damit auch mittelbar auf Kodierentscheidungen aus.

Aufgrund der geringen Stichprobengröße ($N = 15$ bzw. $n = 9$ (KG) respektive $n = 6$ (TG)) wurde das Signifikanzniveau von $p < .05$ auf $p < .1$ angehoben. Dieses Vorgehen ist durch kleine Stichprobengrößen sowie den explorativen Charakter des erstmals zur Anwendung gebrachten Kategoriensystems zu begründen (Bortz & Schuster, 2010; Leonhart, 2017). Signifikanzen zwischen $.1 > p > .05$ bedürfen dennoch, wie geschehen, einer besonders vorsichtigen Interpretation, Diskussion und im besten Fall einer Überprüfung mit einer größeren Stichprobe.

5.4.4 Limitationen bezüglich der Interpretation

Im Rahmen der post-Erhebung steht jeweils nur eine videografierte Unterrichtsstunde pro Lehrkraft zur Verfügung. Diese Datengrundlage ist zu dünn, um über eine vertiefte Analyse des Transfers mittels der Transfertaxonomie von Joyce und Showers (2002) zu validen Interpretationen zu kommen, fordern doch Joyce und Showers (2002) eine Interpretation der Passung des angewendeten Fortbildungsinhalts mit *unterschiedlichen* Unterrichtssituationen und Unterrichtsgegenständen. Diese Passung kann anhand einer Unterrichtsstunde nicht valide beurteilt werden. Aus diesem Grund wurde allein eine Bestimmung der vorliegenden Transferformen vorgenommen.

6 Zusammenfassung, Fazit und Ausblick

Ausgangspunkt dieser Arbeit sind noch fehlende eindeutige Befunde zur Wirkweise von einzelnen Aspekten von Lehrkräftefortbildungen (Desimone et al., 2002; Gräsel et al., 2006). Es wurden zwar Aspekte von Lehrkräftefortbildungen identifiziert, welche als effektiv gelten können, doch existieren hier allenfalls Hinweise auf Effektivität. Besonders die Wirkung des Aspekts der Kokonstruktion stellt nach wie vor ein Desiderat in der Fortbildungsforschung dar (Grosche et al., 2020). Speziell gerichtet auf Fortbildungsforschung zum Experimentieren fordern Capps et al. (2012) breitere Evidenz für die Wirkung von Lehrkräftefortbildung auf das Unterrichtshandeln, das PCK sowie die Beliefs der Lehrkräfte wie auch die Schülerkognition.

Mit dieser Studie zur Untersuchung zweier Formate (individualisiert-konstruktiv und kokonstruktiv) einer Lehrkräftefortbildung zum Experimentieren wird hinsichtlich der Untersuchung der Änderung des Unterrichtshandelns ein Beitrag zur Schließung der Desiderate geleistet.

Zur Untersuchung des Unterrichtshandelns wurden zuerst Aspekte effektiven Unterrichtshandelns zum Experimentieren – anhand zuvor abgeleiteter Unterrichtsprinzipien – beschrieben und empirisch fundiert. Anschließend wurden geeignete Rating- und Kodiermanuale zur Erfassung des Unterrichtshandelns zusammengetragen. Hier zeigte sich, dass die Manuale sehr verschiedene Facetten des Unterrichtshandelns zum Experimentieren fokussieren. Auch der Zweck der eingesetzten Manuale unterscheidet sich erheblich.

Aus dem Befund, welche Aspekte von Unterrichtshandeln als effektiv gelten können, der Zusammenschau existierender Erhebungsinstrumente und den Erkenntnissen der Transferforschung wurde ein Abgleich durchgeführt, um zu erfassen, welche Instrumente für die Erfassung des Unterrichtshandelns in dieser Studie geeignet sind und wo Neuentwicklungen oder Adaptionen notwendig werden. Es wurde deutlich, dass bezüglich der Kodierung von Änderungen von effektivem Unterrichtshandeln kein probates Kodiermanual existiert. Darum wurde ein geeignetes Kodiermanual grundlegend neu entwickelt. Ergänzend wurde auf existierende Instrumente (adaptiert) zurückgegriffen.

Eine Zusammenschau bisheriger Befunde zu Lehrkräftefortbildungen zum Experimentieren diente als Grundlage der später folgenden Einordnung der in dieser Studie gewonnenen Erkenntnisse in das Forschungsfeld.

Aus dieser Zusammenschau konnte im Wesentlichen kondensiert werden, dass der Transfer der Fortbildungsinhalte auf das Unterrichtshandeln oft

- abhängig von der einzelnen Lehrkraft ist (Harlow, 2013; Lotter et al., 2013),
- nur auf einzelne Aspekte beschränkt ist (Grigg et al., 2013; Rushton et al., 2017),
- nur oberflächlich umgesetzt wird (Chin et al., 2016; Marshall & Smart, 2013).

Die in dieser Arbeit berichteten Ergebnisse stützen sich auf N = 15 Teilnehmende (TG: n = 6, KG: n = 9) einer Lehrkräftefortbildung. Bei allen teilnehmenden Lehrkräften wurden vor und nach der Präsenzfortbildung Videografien des Unterrichtshandelns angefertigt und ausgewertet.

Es zeigten sich mit Blick auf das Unterrichtshandeln beim Experimentieren zwei wesentliche Änderungen: dass die Lehrkräfte nach der Fortbildung weniger kochrezeptartige Experimentieranleitungen einsetzen und Experimentierphasen – begleitet von Unterstützungsangeboten – öffnen. Außerdem regen Lehrkräfte nach der Fortbildung Lernende stärker zu wissenschaftsorientiertem Arbeiten an (‚*Minds on*' statt ‚*Hands on*').

Mit Blick auf den Transfer von Fortbildungsinhalten in das Unterrichtshandeln ist eine starke Individualisierung und differenzielle Entwicklung zu beobachten. Daneben scheint das Unterrichtshandeln auch gegenüber (langfristigen) Interventionen robust: In der Mehrzahl der Aspekte des Unterrichtshandelns zeigen sich keine signifikanten Änderungen.

Diese Befunde decken sich sowohl mit den Befunden der Transferforschung allgemein (vgl. Lipowsky, 2014) als auch der Zusammenschau von Studien zu Wirkungen von Fortbildungen auf das Unterrichtshandeln beim Experimentieren (Kapitel 2.4.2).

Nicht bestätigt werden konnte die Hypothese, dass ein kokonstruktives Fortbildungsformat gegenüber einem individualisiert-konstruktiven Format zu stärkeren Effekten im Unterrichtshandeln führt. Ursächlich dafür mag, neben weiteren Faktoren, die Stichprobenverteilung sein (Kapitel 5.4.3). Anhand dieses Befundes muss es als diskutabel gelten, ob das organisatorisch bedeutend anspruchsvollere kokonstruktive Fortbildungsformat (gleichzeitige Hospitationen aller Lehrkräfte im laufenden Schulbetrieb) angesichts der Ergebnisse gerechtfertigt ist.

Der Ertrag dieser Studie soll in drei Ausblicken deutlich gemacht werden:

1. Im Rahmen dieser Studie wurde ein Instrument zur Erfassung von Änderungen des Unterrichtshandelns von Lehrkräften beim Experimentieren grundlegend entwickelt. Dieses Instrument hat sich in dieser Studie, auch

gemessen an statistischen Kenngrößen wie der Beobachterübereinstimmung (Interraterreliabilität), bewährt und steht der Unterrichtsforschung für weitere Studien zur Verfügung.

2. Es konnte gezeigt werden, dass das untersuchte Fortbildungskonzept prinzipiell zu Änderungen im Unterrichtshandeln führt. Die Überlegenheit des kokonstruktiven Fortbildungsformats auf das Unterrichtshandeln gegenüber dem individualisierten Fortbildungsformat konnte jedoch in dieser Studie nicht bestätigt werden. Weitere Aspekte effektiver Lehrkräftefortbildung (Garet et al., 2001) bieten sich für eine Überprüfung in ähnlichem Studiendesign an.
3. In der Diskussion der Befunde wird als Erklärungsansatz für Hemmnisse effektiver (Lehrkräfte-)Fortbildungen die Theorie der Verlustaversion aus dem Gebiet der Verhaltensökonomie (Kahneman et al., 1991) vorgestellt. Damit wird mit den oft heterogenen und in der Summe meist nicht befriedigenden Transfererfolgen eine Theorie verbunden, welche in der Transfer- und Fortbildungsforschung bisher kaum Beachtung findet. Ob dieser Ansatz tatsächlich erklärungsmächtig werden kann, wäre über eine umfangreiche theoretische Fundierung und abschließende empirische Belege zu prüfen.

7 Literaturverzeichnis

Abd-El-Khalick, F. (2006a). Over and over again: College student's views of Nature of Science. In L. B. Flick & N. G. Lederman (Hrsg.), *Scientific Inquiry and Nature of Science. Implications for Teaching, Learning, and Teacher Education* (Science & Technology Education Library, S. 389–424). Dordrecht: Springer.

Abd-El-Khalick, F. (2006b). Socioscientific Issues in Pre-College Science Classrooms. In D. L. Zeidler (Ed.), *The Role of Moral Reasoning on Socioscientific Issues and Discourse in Science Education* (Science & Technology Education Library). Dordrecht: Kluwer Academic Publishers.

Abrahams, I. & Millar, R. (2008). Does Practical Work Really Work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education, 30* (14), 1945–1969.

Aditomo, A. & Klieme, E. (2020). Forms of inquiry-based science instruction and their relations with learning outcomes. Evidence from high and low-performing education systems. *International Journal of Science Education, 117* (1), 1–22.

Aebli, H. (1993). *Grundlagen des Lehrens. Eine allgemeine Didaktik auf psychologischer Grundlage* (2. Aufl.). Stuttgart: Klett-Cotta.

Akremi, L. (2014). Stichprobenziehung in der qualitativen Sozialforschung. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 265–279). Wiesbaden: Springer VS.

Akuma, F. V. & Callaghan, R. (2018). A systematic review characterizing and clarifying intrinsic teaching challenges linked to inquiry-based practical work. *Journal of Research in Science Teaching, 30* (10), 1405.

Alexakos, K. (2010). Teaching the practice of science, unteaching the "scientific method". *Science Scope, 33* (9), 74–79.

Alfieri, L., Brooks, P. J., Aldrich, N. J. & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103* (1), 1–18.

Allen, D. E., Duch, B. J. & Groh, S. E. (1996). The power of problem-based learning in teaching introductory science courses. *New Directions for Teaching and Learning, 6* (68), 43–52.

Altman, D. G. (1991). *Practical statistics for medical research*. Boca Raton, Fla: Taylor & Francis.

Anderson, L. W. (Ed.). (2009). *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives* [Nachdr.]. New York: Longman.

Antoniou, P. & Kyriakides, L. (2013). A Dynamic Integrated Approach to teacher professional development. Impact and sustainability of the effects on improving teacher behaviour and student outcomes. *Teaching and Teacher Education, 29,* 1–12.

Appel, J. & Rauin, U. (2016). Quantitative Analyseverfahren in der videobasierten Unterrichtsforschung. In U. Rauin, M. Herrle & T. Engartner (Hrsg.), *Videoanalysen in der Unterrichtsforschung. Methodische Vorgehensweisen und Anwendungsbeispiele* (Grundlagentexte Methoden, S. 130–153). Weinheim: Beltz Juventa.

Arnold, J., Kremer, K. & Mayer, J. (2012). Wissenschaftliches Denken beim Experimentieren – Kompetenzdiagnose in der Sekundarstufe II. In D. Krüger, A. Upmeier zu Belzen, P. Schmiemann, A. Möller & D. Elster (Hrsg.), *Erkenntnisweg Biologiedidaktik 11* (S. 7–20).

Arnold, J., Kremer, K. & Mayer, J. (2017). Scaffolding beim Forschenden Lernen. *Zeitschrift für Didaktik der Naturwissenschaften, 23* (1), 21–37.

Australian Curriculum, Assessment And Reporting Authority (2016). *The Australian Curriculum. Science* (Australian Curriculum, Assessment And Reporting Authority, Hrsg.).

Bachmann, H. (2003). *Auch Lernen will gelernt sein. Von der Theorie zur Praxis*. Oberentfelden: Sauerländer.

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2016). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (14., überarbeitete und aktualisierte Auflage). Berlin: Springer Gabler.

Banchi, H. & Bell, R. L. (2008). The Many Levels of Inquiry. Inquiry comes in various forms. *Science and Children,* 26–29.

Barron, B. & Darling-Hammond, L. (2008). Teaching for Meaningful Learning. A Review of Research on Inquiry-Based and Cooperative Learning. Book Excerpt. *George Lucas Educational Foundation*.

Baumert, J. (Hrsg.). (2012). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Baumert, J., Bos, W. & Lehmann, R. (2000). *TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie -- Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 1 Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Baur, A., Ehrenfeld, U., Emden, M., Hummel, E. & Krieg, A. (2017). *Naturwissenschaften zum Leben erwecken: Chemie. Unterrichtsideen, Materialien und didaktische Grundlagen zum offenen Experimentieren*. Hamburg: Persen.

Baur, A. & Emden, M. (2020). How to open inquiry teaching? An alternative teaching scaffold to foster students' inquiry skills. *Chemistry Teacher International, 2* (1), 1–12.

Baur, A., Emden, M. & Bewersdorff, A. (2019). Welche Unterrichtsprinzipien sollten für den Aufbau von Kompetenzen zum Experimentieren Beachtung finden? Eine Ableitung auf Basis multiperspektivisch begründeter Unterrichtsziele. *Zeitschrift für Didaktik der Biologie, 23* (1), 10–24.

Baur, A., Hummel, E., Emden, M. & Schröter, E. (2020). Wie offen sollte offenes Experimentieren sein? Ein Plädoyer für das geöffnete Experimentieren. *MNU Journal, 73* (2), 125–128.

Bell, R. L., Blair, L. M., Crawford, B. A. & Lederman, N. G. (2003). Just do it? Impact of a science apprenticeship program on high school students' understandings of the nature of science and scientific inquiry. *Journal of Research in Science Teaching, 40* (5), 487–509.

Bell, R. L., Smetana, L. & Binns, I. (2005). Simplifying Inquiry Instruction. Assessing the inquiry level of classroom activities. *The Science Teacher, 7* (72), 30–33.

Bell, T. (2012). Entdeckendes und forschendes Lernen. In S. Mikelskis-Seifert & T. Rabe (Hrsg.), *Physik-Methodik. Handbuch für die Sekundarstufe I und II* (3. Auflage, S. 70–81). Berlin: Cornelsen.

Bell, T., Urhahne, D., Schanze, S. & Ploetzner, R. (2010). Collaborative Inquiry Learning. Models, tools, and challenges. *International Journal of Science Education, 32* (3), 349–377.

Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S. & Reiser, B. J. (2016). Epistemologies in practice. Making scientific practices meaningful for students. *Journal of Research in Science Teaching, 53* (7), 1082–1112.

Bewersdorff, A., Baur, A. & Emden, M. (2020). Untersuchung der Wirksamkeit einer Lehrkräftefortbildung zum Experimentieren. In: S. Habig (Hrsg.), *Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen.* Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Wien 2019, S. 860-863.

Bewersdorff, A., Baur, A. & Emden, M. (2020). Analyse von Unterrichtskonzepten zum Experimentieren hinsichtlich theoretisch begründeter Unterrichtsprinzipien. Bestandsaufnahme und kriteriale Gegenüberstellung. *Zeitschrift für Didaktik der Biologie, 24* (1), 108–130.

Blume, B. D., Ford, J. K., Baldwin, T. T. & Huang, J. L. (2010). Transfer of Training. A Meta-Analytic Review. *Journal of Management, 36* (4), 1065–1105.

Böhm-Kasper, O. & Weishaupt, H. (2004). Quantitative Ansätze und Methoden in der Schulforschung. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (1. Auflage, S. 93–126). Wiesbaden: VS Verlag für Sozialwissenschaften.

Borko, H. (2004). Professional Development and Teacher Learning: Mapping the Terrain. *Educational Researcher, 33* (8), 3–15.

Börlin, J. (2012). *Das Experiment als Lerngelegenheit. Vom interkulturellen Vergleich des Physikunterrichts zu Merkmalen seiner Qualität*. Universität Basel, Basel. Zugriff am 15.06.2018.

Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Berlin: Springer.

Brand, B. R. & Moore, S. J. (2011). Enhancing Teachers' Application of Inquiry-Based Strategies Using a Constructivist Sociocultural Professional Development Model. *International Journal of Science Education, 33* (7), 889–913.

Brandon, P. R., Taum, A. K. H., Young, D. B. & Pottenger, F. M. (2008). The development and validation of The Inquiry Science Observation Coding Sheet. *Evaluation and program planning, 31* (3), 247–258.

Brinkmann, M. (2009). Üben. Wissen — Können — Wiederholen. *Vierteljahrsschrift für Wissenschaftliche Pädagogik, 85* (4), 413–434.

Brückmann, M. & Duit, R. (2014). Videobasierte Analyse unterrichtlicher Sachstrukturen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 189–201). Heidelberg: Springer.

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31* (1), 21–32.

Bybee, R. W. (2002). Scientific Literacy - Mythos oder Realität? In P. Nentwig & T. Koballa (Hrsg.), *Scientific Literacy. Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung* (S. 21–42). Heidelberg: Springer.

Cairns, D. & Areepattamannil, S. (2017). Exploring the Relations of Inquiry-Based Teaching to Science Achievement and Dispositions in 54 Countries. *Research in Science Education, 49* (1), 1–23.

Calais, G. J. (2006). Haskell's Taxonomies Of Transfer Of Learning: Implications For Classroom Instruction. *National Forum Of Applied Educational Research Journal, 20* (3), 1–8.

Capps, D. K. & Crawford, B. A. (2012). Inquiry-Based Instruction and Teaching About Nature of Science. Are They Happening? *Journal of Science Teacher Education, 24* (3), 497–526.

Capps, D. K. & Crawford, B. A. (2013). Inquiry-Based Professional Development. What does it take to support teachers in learning about inquiry and nature of science? *International Journal of Science Education, 35* (12), 1947–1978.

Capps, D. K., Crawford, B. A. & Constas, M. A. (2012). A Review of Empirical Literature on Inquiry Professional Development. Alignment with Best Practices and a Critique of the Findings. *Journal of Science Teacher Education, 23* (3), 291–318.

Capps, D. K., Shemwell, J. T. & Young, A. M. (2016). Over reported and misunderstood? A study of teachers' reported enactment and knowledge of inquiry-based science teaching. *International Journal of Science Education, 38* (6), 934–959.

Carrier, R. (2001). *Test Your Scientific Literacy!,* The Secular Web. Verfügbar unter https://infidels.org/library/modern/richard_carrier/SciLit.html

Chin, E.-T., Lin, Y.-C. & Tuan, H.-l. (2016). Analyzing Changes in Four Teachers' Knowledge and Practice of Inquiry-Based Mathematics Teaching. *The Asia-Pacific Education Researcher, 25* (5-6), 845–862.

Chinn, C. A. & Brewer, W. F. (1993). The Role of Anomalous Data in Knowledge Acquisition. A Theoretical Framework and Implications for Science Instruction. *Review of Educational Research, 63* (1), 1.

Chinn, C. A. & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools. A theoretical framework for evaluating inquiry tasks. *Science Education, 86* (2), 175–218.

Clausen, M. (2002). *Unterrichtsqualität. Eine Frage der Perspektive? : empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Münster: Waxmann.

Clough, M. P. (2006). Learners' Responses to the Demands of Conceptual Change. Considerations for Effective Nature of Science Instruction. *Science & Education, 15* (5), 463–494.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement, 20* (1), 37–46.

Cohen, S. (1988). Perceived stress in a probability sample of the United States. In S. Spacapan & S. Shirlynn (Hrsg.), *The Claremont Symposium on Applied Social Psychology. The social psychology of health*. Newbury Park: Sage.

Crawford, B. A. (2007). Learning to teach science as inquiry in the rough and tumble of practice. *Journal of Research in Science Teaching, 44* (4), 613–642.

Crawford, B. A. (2014). From Inquiry to Scientific Practices in the Science Classroom. In N. G. Lederman & S. K. Abell (Hrsg.), *Handbook of research on science education* (S. 515–541). New York, NY: Routledge.

Darian, S. G. (2003). *Understanding the language of science* (1st ed.). Austin: University of Texas Press.

Dean, D., JR. & Kuhn, D. (2007). Direct instruction vs. discovery. The long view. *Science Education, 91* (3), 384–397.

Department for Education (2014). The national curriculum in England - Key stages 3 and 4 framework document.

Deppermann, A. (2014). Konversationsanalyse: Elementare Interaktionsstrukturen am Beispiel der Bundespressekonferenz. In S. Staffeldt & J. Hagemann (Hrsg.), *Pragmatiktheorien* (Stauffenburg Einführungen, Band 27, S. 19–47). Tübingen: Stauffenburg Verlag.

Desimone, L. M. (2009). Improving Impact Studies of Teachers' Professional Development. Toward Better Conceptualizations and Measures. *Educational Researcher, 38* (3), 181–199.

Desimone, L. M. & Garet, M. S. (2015). Best Practices in Teachers' Professional Development in the United States. *Psychology, Society, & Education, 7* (3), 252–263.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S. & Birman, B. F. (2002). Effects of Professional Development on Teachers' Instruction. Results from a Three-year Longitudinal Study. *Educational Evaluation and Policy Analysis, 24* (2), 81–112.

Dewey, J. (1910). Science as Subject-Matter and as Method. *Science, 31* (787), 121–127.

Dewey, J. (1938). *Logic: The Theory of Inquiry*. New York: Henry Holt.

Dinkelaker, J. & Herrle, M. (2009). *Erziehungswissenschaftliche Videographie*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Donnelly, D., O'Reilly, J. & McGarr, O. (2013). Enhancing the Student Experiment Experience. Visible Scientific Inquiry Through a Virtual Chemistry Laboratory. *Research in Science Education, 43* (4), 1571–1592.

Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin, Heidelberg: Springer.

Dörner, D. (1976). *Problemlösen als Informationsverarbeitung*. Stuttgart: Kohlhammer.

Edwards, J. A. & Lampert, M. D. (Hrsg.). (1993). *Talking Data. Transcription and Coding in Discourse Research*. Hoboken: Taylor and Francis.

Emden, M. (2011). *Prozessorientierte Leistungsmessung des naturwissenschaftlich-experimentellen Arbeitens. Eine vergleichende Studie zu Diagnoseinstrumenten zu Beginn der Sekundarstufe I*. Berlin: Logos.

Emden, M. (2021): Reintroducing "the" Scientific Method to Introduce Scientific Inquiry in Schools? In: *Science & Education*.

Emden, M.; Baur, A. (2015): "Entdeckendes Experimentieren" - Wirksamkeit einer Lehrerinnen- und Lehrerfortbildung im Format der Schulwerkstatt: DFG-Förderantrag.

Emden, M. & Baur, A. (2017). Effektive Lehrkräftebildung zum Experimentieren – Entwurf eines integrierten Wirkungs- und Gestaltungsmodells. *Zeitschrift für Didaktik der Naturwissenschaften, 23* (1), 1–19.

Emden, M., Baur, A. & Bewersdorff, A. (2020). Professional Development in the School Insitute "Discovery Experimentation" - Framework and first Results. *Progress in Science Education, 3* (2), 1–9.

Emden, M., Bewersdorff, A. & Baur, A. (2019). Kann Experimentieren in der Schule bilden? Ein Beitrag zur bildungstheoretischen Legitimation eines selbstverständlichen Gegenstandes des Naturwissenschaftsunterrichts. *Zeitschrift für Pädagogik, 65* (5), 710–729.

Engeln, K., Euler, M. & Maass, K. (2013). Inquiry-based learning in mathematics and science. A comparative baseline study of teachers' beliefs and practices across 12 European countries. *ZDM, 45* (6), 823–836.

Engeln, K., Mikelskis-Seifert, S. & Euler, M. (2014). Inquiry-Based Mathematics and Science Education Across Europe: A Synopsis of Various Approaches and Their Potentials. In C. Bruguière, A. Tiberghien & P. Clément (Hrsg.), *Topics and Trends in Current Science Education. 9th ESERA Conference Selected Contributions*. Dordrecht: Springer Netherlands.

Engl, L. (2017). *Bedeutung des Protokollierens für den naturwissenschaftlichen Erkenntnisprozess*. Dissertation. Universität Koblenz-Landau, Landau.

Ethik-Rat der DGfE (2016). *Ethik-Kodex der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE)* (Deutsche Gesellschaft für Erziehungswissenschaft, Hrsg.). Berlin. Verfügbar unter https://www.dgfe.de/fileadmin/Ordner-Redakteure/Satzung_etc/Ethikkodex_2016.pdf

Faraday, M. (1828). *Chemische Manipulation. Oder das eigentlich Practische der sichern Ausführung chemischer Arbeiten und Experimente*. Weimar: Verlag des Großherzoglich Sächsisch privilegierten Landes-Industrie-Comptoirs.

Faßnacht, G. (1995). *Systematische Verhaltensbeobachtung. Eine Einführung in die Methodologie und Praxis*. München: Reinhardt.

Fischer, H. E., Borowski, A. & Tepner, O. (2012). Professional Knowledge of Science Teachers. In B. J. Fraser, K. Tobin & C. J. McRobbie (Hrsg.), *Second International Handbook of Science Education*. Dordrecht: Springer Netherlands.

Fisher, D., Harrison, A., Henderson, D. & Hofstein, A. (1998). Laboratory learning environments and practical tasks in senior secondary science classes. *Research in Science Education, 28* (3), 353–363.

Flick, L. B. (2006). Developing Understanding Of Scientific Inquiry In Secondary Students. In L. B. Flick & N. G. Lederman (Hrsg.), *Scientific Inquiry and Nature of Science. Implications for Teaching, Learning, and Teacher Education*. Dordrecht: Springer.

Florian, A. (2008). *Blended Learning in der Lehrerfortbildung. Evaluation eines onlinegestützten, teambasierten und arbeitsbegleitenden Lehrerfortbildungsangebots im deutschsprachigen Raum*. Dissertation. Universität Augsburg, Augsburg.

Forbes, C. T., Biggers, M. & Zangori, L. (2013). Investigating Essential Characteristics of Scientific Practices in Elementary Science Learning Environments. The Practices of Science Observation Protocol (P-SOP). *School Science and Mathematics, 113* (4), 180–190.

Furtak, E. M. (2006). The problem with answers. An exploration of guided scientific inquiry teaching. *Science Education, 90* (3), 453–467.

Furtak, E. M., Seidel, T., Iverson, H. & Briggs, D. C. (2012). Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching. *Review of Educational Research, 82* (3), 300–329.

Gächter, S., Johnson, E. J. & Herrmann, A. (2010). *Individual-level loss aversion in riskless and risky choices*. CeDEx Discussion Paper Series. University of Nottingham, Nottingham.

Galluzzo, G. & Craig, J. (1990). Evaluation of Preservice Teacher Education Programs. In R. W. Houston (Hrsg.), *Handbook of Research on Teacher Education* (S. 599–616). New York.

Garet, M. S., Birman, B. F., Porter, A. C., Desimone, L. M. & Herman, R. (1999). *Designing Effective Professional Development: Lessons from the Eisenhower ProgramAndTechnical Appendices* (American Institutes for Reseach, Hrsg.). Washington, D.C.

Garet, M. S., Porter, A. C., Desimone, L. M., Birman, B. F. & Suk Yoon, K. (2001). What Makes Professional Development Effective? Results from a National Sample of Teachers. *American Educational Research Journal, 38* (4), 915–945.

Gerstenmaier, J. & Mandl, H. (1995). Wissenserwerb unter konstruktivistischer Perspektive. *Zeitschrift für Pädagogik, 41* (6), 867–888.

Glaser, R., Schauble, L., Raghavan, K. & Zeitz, C. (1992). Scientific Reasoning Across Different Domains. In E. de Corte, M. C. Linn, H. Mandl & L. Verschaffel (Hrsg.), *Computer-Based Learning Environments and Problem Solving.* Berlin, Heidelberg: Springer.

Glöckel, H. (2003). *Vom Unterricht. Lehrbuch der allgemeinen Didaktik* (4., durchges. u. erg. Aufl.). Bad Heilbrunn: Klinkhardt.

Gniewosz, B. (2011). Beobachtung. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Strukturen und Methoden* (1. Aufl., S. 99–119). Wiesbaden: VS Verlag für Sozialwissenschaften.

Gott, R. & Duggan, S. (2007). A framework for practical work in science and scientific literacy through argumentation. *Research in Science & Technological Education, 25* (3), 271–291.

Gräber, W. (2002). „Scientific Literacy" - Naturwissenschaftliche Bildung in der Diskussion. In P. Döbrich (Hrsg.), *Qualitätsentwicklung im naturwissenschaftlichen Unterricht. Fachtagung am 15. Dezember 1999.* Frankfurt am Main.

Gräsel, C. (2008). Die Anregung zur Kooperation im Rahmen von Fortbildungen: unterschiedliche Wege. *Beiträge zur Lehrerinnen- und Lehrerbildung,* 64–71.

Gräsel, C. (2010). Stichwort. Transfer und Transferforschung im Bildungsbereich. *Zeitschrift für Erziehungswissenschaft, 13* (1), 7–20.

Gräsel, C., Fußangel, K. & Parchmann, I. (2006). Lerngemeinschaften in der Lehrerfortbildung. Kooperationserfahrungen und -überzeugungen von Lehrkräften. *Zeitschrift für Erziehungswissenschaft* (9), 545–561.

Gräsel, C. & Parchmann, I. (2004). Implementationsforschung - oder: der steinige Weg, Unterricht zu verändern. *Unterrichtswissenschaft, 32* (3), 196–214.

Graumann, C. F. (1966). Grundzüge der Verhaltensbeobachtung. In E. Meyer (Hrsg.), *Fernsehen in der Lehrerbildung. Neue Forschungsansätze in Pädagogik, Didaktik und Psychologie* (S. 86–107). München: Manz Verlag.

Grigg, J., Kelly, K. A., Gamoran, A. & Borman, G. D. (2013). Effects of Two Scientific Inquiry Professional Development Interventions on Teaching Practice. *Educational Evaluation and Policy Analysis, 35* (1), 38–56.

Gris, R. (2008). *Die Weiterbildungslüge. Warum Seminare und Trainings Kapital vernichten und Karrieren knicken*. Frankfurt: Campus.

Gropengießer, H. (2018): Experimentieren. In: Harald Gropengießer, Ute Harms und Ulrich Kattmann (Hg.): Fachdidaktik Biologie. Die Biologiedidaktik. Seelze: Aulis Verlag, S. 284–293.

Grosche, M., Fußangel, K. & Gräsel, C. (2020). Kokonstruktive Kooperation zwischen Lehrkräften. Aktualisierung und Erweiterung der Kokonstruktionstheorie sowie deren Anwendung am Beispiel schulischer Inklusion. *Zeitschrift für Pädagogik, 66* (4), 461–479.

Gut-Glanzmann, C. & Mayer, J. (2018). Experimentelle Kompetenz. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 121–140). Berlin, Heidelberg: Springer.

Hammann, M., Ho, T. T. & Bayrhuber, H. (2007). Experimentieren als Problemlösen: Lässt sich das SDDS- Modell nutzen, um unterschiedliche Dimensionen beim Experimentieren zu messen? *Zeitschrift für Erziehungswissenschaft, 10* (8), 33–49.

Hammann, M. & Mayer, J. (2012). Was lernen Schülerinnen und Schüler beim Experimentieren? *Biologie in unserer Zeit, 42* (5), 284–285.

Harlen, W. (1999): Effective Teaching of Science. A Review of Research: The Scottish Council for Research in Education.

Harlow, D. B. (2013). An Investigation of How a Physics Professional Development Course Influenced the Teaching Practices of Five Elementary School Teachers. *Journal of Science Teacher Education, 25* (1), 119–139.

Härtig, H., Neumann, K. & Erb, R. (2017). Experimentieren als Interaktion von Situation und Person. *Zeitschrift für Didaktik der Naturwissenschaften, 23* (1), 71–80.

Haskell, R. E. (2001). *Transfer of learning. Cognition, instruction, and reasoning*. San Diego, Calif: Academic Press.

Hasselhorn, M. & Hager, W. (2008). Transferwirkungen kognitiver Trainings. In W. Schneider, M. Hasselhorn & J. Bengel (Hrsg.), *Handbuch der pädagogischen Psychologie*. Göttingen: Hogrefe.

Hazelkorn, E., Ryan, C., Beernaert, Y., Constantinou, C. P., Deca, L., Grangeat, M. et al. (2015). *Science Education for Responsible Citizenship. Report to the European Commission of the expert group on science education* (European Commission, Hrsg.). Directorate-General for Research and Innovation.

Helmke, A. (2017). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze-Velber: Klett/Kallmeyer.

Henke, A. (2016). Lernen über die Natur der Naturwissenschaften – Forschender und historisch orientierter Physikunterricht im Vergleich. *Zeitschrift für Didaktik der Naturwissenschaften, 22* (1), 123–145.

Hense, J. & Mandl, H. (2011). Transfer in der beruflichen Weiterbildung. In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen Empirischer Bildungsforschung* (S. 249–263). Wiesbaden: VS Verlag für Sozialwissenschaften.

Herrle, M. & Breitenbach, S. (2016). Planung, Durchführung und Nachbereitung videogestützter Beobachtungen im Unterricht. In U. Rauin, M. Herrle & T. Engartner (Hrsg.), *Videoanalysen in der Unterrichtsforschung. Methodische Vorgehensweisen und Anwendungsbeispiele*. Weinheim: Beltz Juventa.

Herrle, M. & Dinkelaker, J. (2016). Qualitative Analyseverfahren in der videobasierten Unterrichtsforschung. In U. Rauin, M. Herrle & T. Engartner (Hrsg.), *Videoanalysen in der Unterrichtsforschung. Methodische Vorgehensweisen und Anwendungsbeispiele*. Weinheim: Beltz Juventa.

Heymann, H. W. (2012). Schüler beim Aufbau von Kompetenzen unterstützen. Üben, Anwenden, Vertiefen - Gelingensbedingungen für nachhaltiges Lernen. *Pädagogik, 64* (12), 6–11.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J. et al. (2003). *Teaching Mathematicsin Seven Countries. Results From the TIMSS 1999 Video Study* (U.S. Department of Education, Hrsg.). Washington, DC: National Center for Education Statistics.

Hodson, D. (2014). Learning Science, Learning about Science, Doing Science. Different goals demand different learning methods. *International Journal of Science Education, 36* (15), 2534–2553.

Hofstein, A. & Lunetta, V. N. (2004). The laboratory in science education. Foundations for the twenty-first century. *Science Education, 88* (1), 28–54.

Holzberger, D., Philipp, A. & Kunter, M. (2016). Ein Blick in die Black-Box. Wie der Zusammenhang von Unterrichtsenthusiasmus und Unterrichtshandeln bei angehenden Lehrkräften erklärt werden kann. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 48* (2), 90–105.

Hood Cattaneo, K. (2017). Telling Active Learning Pedagogies Apart. From theory to practice. *Journal of New Approaches in Educational Research, 6* (2), 144–152.

Hoorens, V., Remmers, N. & van de Riet, K. (1999). Time is an amazingly variable amount of money. Endowment and ownership effects in the subjective value of working time. *Journal of Economic Psychology, 20* (4), 383–405.

Horizon Research, I. (2005). *Local Systemic Change Classroom Observation Protocol.*

Höttecke, D. (2013). Forschend-entdeckenden Unterricht authentisch gestalten - ein Problemaufriss. In S. Bernholt (Hrsg.), *Inquiry-based Learning - Forschendes Lernen. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Hannover 2012*. Kiel: IPN.

Höttecke, D. & Rieß, F. (2007). Rekonstruktion der Vorstellungen von Physikstudierenden über die Natur der Naturwissenschaften - eine explorative Studie. *Physik und Didaktik in Schule und Hochschule, 6* (1), 1–14.

Ingvarson, L., Meiers, M. & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes & efficacy. *education policy analysis archives, 13,* 10.

Irion, T. (2010). Hypercoding in der empirischen Lehr-Lern- Forschung. Möglichkeiten der synchronen Analyse multicodaler Datensegmente zur Rekonstruktion subjektiver Perspektiven in Videostudien. In M. Corsten, M. Krug & C. Moritz (Hrsg.), *Videographie praktizieren. Herangehensweisen, Möglichkeiten und Grenzen*. Wiesbaden: VS Verlag.

Janik, T.; Seidel, T. (Hrsg.) (2009): The power of video studies in investigating teaching and learning in the classroom. Münster: Waxmann.

Jiang, F. & McComas, W. F. (2015). The Effects of Inquiry Teaching on Student Science Achievement and Attitudes. Evidence from Propensity Score Analysis of PISA Data. *International Journal of Science Education, 37* (3), 554–576.

Joyce, B. & Showers, B. (2002). *Student Achievement Through Staff Development, 3rd Edition* (3rd ed.). Alexandria: ASCD.

Kahneman, D., Knetsch, J. L. & Thaler, R. H. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectiv, 5* (1), 193–206.

Kalthoff, B., Theyssen, H. & Schreiber, N. (2018). Explicit promotion of experimental skills. And what about the content-related skills? *International Journal of Science Education, 40* (11), 1305–1326.

Kipnis, M. & Hofstein, A. (2008). The Inquiry Laboratory as a Source for Development of Metacognitive Skills. *International Journal of Science and Mathematics Education, 6* (3), 601–627.

Kirschner, P. A., Sweller, J. & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Pyschologist, 41* (2), 75–86.

Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12* (1), 1–48.

Kochinka, A. (2010). Beobachtung. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 449–461). Wiesbaden: VS Verlag.

Köller, O. (2014). Naturwissenschaftliche Leistungen, demographische Veränderungen und Lehrerbildung. *Zeitschrift für Didaktik der Naturwissenschaften, 20* (1), 3–9.

Konferenz der Kultusminister der Länder (2005a). *Beschlüsse der Kultusministerkonferenz. Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss* (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland., Hrsg.). Berlin.

Konferenz der Kultusminister der Länder (2005b). *Beschlüsse der Kultusministerkonferenz. Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss* (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland., Hrsg.). Berlin.

Konferenz der Kultusminister der Länder (2005c). *Beschlüsse der Kultusministerkonferenz. Bildungsstandards im Fach Physik für den Mittleren Schulabschluss* (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland., Hrsg.). Berlin.

Konferenz der Kultusminister der Länder (2020). *Ländergemeinsame Eckpunkte zur Fortbildung von Lehrkräften als ein Bestandteil ihrer Professionalisierung in der dritten Phase der Lehrerbildung* (Sekretariat der Ständigen Konferenz der Kultusminister der Länder, Hrsg.). Berlin.

Kreis, I. & Unterköfler-Klatzer, D. (2017). *Fortbildung Kompakt. Wissenschaftstheoretische und praktische Modelle zur wirksamen Lehrer/innenfortbildung*. Innsbruck: Studienverlag.

Krumbacher, C. (2016). *Die Relevanz lernprozessorientierter Sequenzierung im physikbezogenen Sachunterricht. Eine Videostudie zur Berücksichtigung von Tiefenstrukturen beim Experimentieren*. Dissertation. Universität Duisburg-Essen, Duisburg-Essen.

Kuckartz, U. (2007). *Einführung in die computergestützte Analyse qualitativer Daten* (3. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project* (Mathematics teacher education, S. 25–48). New York: Springer.

Kunter, M., Kleickmann, T. & Richter, D. (2011). Die Entwicklung professioneller Kompetenz von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 55–68). Münster: Waxmann.

Kunter, M. & Trautwein, U. (2013). *Psychologie des Unterrichts* (StandardWissen Lehramt, Bd. 3895). Paderborn: Schöningh.

Kurt, R. (2010). Diener zweier Damen. Videobasierte Sozialforschung zwischen Datendokumentation und Filmproduktion. In M. Corsten, M. Krug & C. Moritz (Hrsg.), *Videographie praktizieren. Herangehensweisen, Möglichkeiten und Grenzen*. Wiesbaden: VS Verlag.

Labudde, P. & Möller, K. (2012). Stichwort: Naturwissenschaftlicher Unterricht. *Zeitschrift für Erziehungswissenschaft, 15* (1), 11–36.

Landis, J. R. & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33* (1), 159–174.

Lazonder, A. W. & Harmsen, R. (2016). Meta-Analysis of Inquiry-Based Learning. Effects of Guidance. *Review of Educational Research, 86* (3), 681–718.

Lederman, J. S., Lederman, N. G., Sug Kim, B. & Kyung Ko, E. (2012). Teaching and Learning of Nature of Science and Scientific Inquiry: Building Capacity Through Systematic Research-Based Professional Development. In M. S. Khine (Hrsg.), *Advances in Nature of Science Research. Concepts and Methodologies* (S. 125–152). Heidelberg: Springer Dordrecht.

Lederman, N. (2007). Nature of science. Past, present, and future. In S. K. Abell & N. G. Lederman (Hrsg.), *Handbook of research on science education*. New York, NY: Routledge.

Lederman, N. G. (2006). Syntax Of Nature Of Science Within Inquiry And Science Instruction. In L. B. Flick & N. G. Lederman (Hrsg.), *Scientific Inquiry and Nature of Science. Implications for Teaching, Learning, and Teacher Education*. Dordrecht: Springer. Z

Lederman, N. G., Lederman, J. S. & Antink, A. (2013). Nature of Science and Scientific Inquiry as Contexts for the Learning of Science and Achievement of Scientific Literacy. *International Journal of Education in Mathematics, Science and Technology, 3* (1), 138–147.

Leonhart, R. (2017). *Lehrbuch Statistik. Einstieg und Vertiefung* (4., überarbeitete und erweiterte Auflage). Bern: Hogrefe Verlag.

Li, Y. & Oliveira, H. (2015). Research on Classroom Practice. In Sung Je Cho (Hrsg.), *The Proceedings of the 12th International Congress on Mathematical Education. Intellectual and Attitudinal Challenges* (S. 489–496). Heidelberg: Springer.

Lipowsky, F. (2010). Lernen im Beruf – Empirische Befunde zur Wirksamkeit von Lehrerfortbildung. In F. H. Müller (Hrsg.), *Lehrerinnen und Lehrer lernen. Konzepte und Befunde zur Lehrerfortbildung* (S. 51–72). Münster: Waxmann.

Lipowsky, F. (2014). Theoretische Perspektiven und empirische Befunde zur Wirksamkeit von Lehrerfort- und -weiterbildung. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 398–417). Münster: Waxmann.

Lotter, C., Rushton, G. T. & Singer, J. (2013). Teacher Enactment Patterns. How Can We Help Move All Teachers to Reform-Based Inquiry Practice Through Professional Development? *Journal of Science Teacher Education, 24* (8), 1263–1291. https://doi.org/10.1007/s10972-013-9361-0

Lotter, C. R., Thompson, S., Dickenson, T. S., Smiley, W. F., Blue, G. & Rea, M. (2016). The Impact of a Practice-Teaching Professional Development Model on Teachers' Inquiry Instruction and Inquiry Efficacy Beliefs. *International Journal of Science and Mathematics Education, 16* (2), 255–273.

Lotz, M., Gabriel, K. & Lipowsky, F. (2013). Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung. Analysen zu deren gegenseitiger Validierung. *Zeitschrift für Pädagogik, 59* (3), 357–380.

Maaß, K. & Euler, M. *Promoting inquiry-based learning in mathematics and science education across Europe.* University of Education Freiburg.

Mandl, H., Prenzel, M. & Gräsel, C. (1992). Das Problem des Lerntransfers in der betrieblichen Weiterbildung. *Unterrichtswissenschaft, 20* (2), 126–143.

Marquardt, C. (2011). *Videoanalysen in der Unterrichtsforschung: Momentaufnahme oder typischer Schulalltag? Eine Betrachtung der DESI-Videostudie*. Berlin: Diplomica Verlag.

Marshall, J. C. & Smart, J. B. (2013). Teachers' Transformation to Inquiry-Based Instructional Practice. *Creative Education, 04* (02), 132–142.

Marshall, J. C., Smart, J. & Horton, R. M. (2009). The design and validation of EQUIP: An instrument to assess inquiry-based instruction. *International Journal of Science and Mathematics Education* (8), 299–321.

Maurer, C. (2016). *Strukturierung von Lehr-Lernsequenzen*. Universität Regensburg, Regensburg.

Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung. Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 177–186). Berlin Heidelberg: Springer-Verlag.

Mayer, J. & Ziemek, H.-P. (2006). Offenes Experimentieren. Forschendes Lernen im Biologieunterricht. *Unterricht Biologie* (317), 4-13.

Mayring, P. (2007): Designs in qualitativ orientierter Forschung. In: *Journal für Psychologie* 15 (2), S. 1–10.

Mayring, P. (2010). Design. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 225–237). Wiesbaden: VS Verlag.

Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (12. Aufl.). Weinheim u.a.: Beltz.

McComas, W. F. (2002). The Principal Elements of the Nature of Science: Dispelling the Myths. In W. F. McComas (Hrsg.), *The nature of science in science education* (S. 53–72). New York, Boston, Dordrecht, London, Moscow: Kluwer Academic Publishers.

McComas, W. F. (2020). Principal Elements of Nature of Science: Informing Science Teaching while Dispelling the Myths. In W. McComas (Hrsg.), *Nature of Science in Science Instruction. Rationales and Strategies* (S. 35–65).

McComas, W. F., Clough, M. P. & Nouri, N. (2020). Nature of Science and Classroom Practice: A Review of the Literature with Implications for Effective NOS Instruction. In W. McComas (Hrsg.), *Nature of Science in Science Instruction. Rationales and Strategies* (S. 67–111).

McComas, W. F., Clough, M. P. & Ylmazroa, H. (2002). The Role And Character of The Nature of Science in Science Education. Rationales and strategies. In W. F. McComas (Hrsg.), *The nature of science in science education* (S. 3–40). New York, Boston, Dordrecht, London, Moscow: Kluwer Academic Publishers.

McDonald, L. (2014). Transfer of the learning. Teacher professional development. *European Journal of Social and Behavioural Sciences, 11* (4), 1569–1584.

McGee, S. & Nutakki, N. (2017). The Impact of Adapting a General Professional Development Framework to the Constraints of In-Service Professional Development on the Next Generation Science Standards in Urban Settings. *Journal of Urban Learning, Teaching, and Research, 13* (73), 73–89.

Meier, M. & Mayer, J. (2012). Experimentierkompetenz praktisch erfassen. Entwicklung und Validierung eines anwendungsbezogenen Aufgabendesigns. In U. Harms & F. X. Bogner (Hrsg.), *"Didaktik der Biologie - Standortbestimmung und Perspektiven". Internationale Tagung der Fachsektion Didaktik der Biologie im VBIO, Bayreuth 2011* (S. 81–98). Innsbruck: Studienverlag.

Merritt, J., Lee, M. Y., Rillero, P. & Kinach, B. M. (2017). Problem-Based Learning in K–8 Mathematics and Science Education. A Literature Review. *Interdisciplinary Journal of Problem-Based Learning, 11* (2).

Metzger, S. & Sommer, K. (2010). "Kochrezept" oder experimentelle Methode? Eine Standortbestimmung von Schülerexperimenten unter dem Gesichtspunkt der Erkenntnisgewinnung. *Der mathematische und naturwissenschaftliche Unterricht, 63* (1), 4–11.

Metzler, P. & Krause, B. (1997). Methodischer Standard bei Studien zur Therapieevaluation. *Methods of Psychological Research, 2* (1), 55–67.

Meyer, H. (1987). *Unterrichtsmethoden*. Frankfurt am Main: Scriptor.

Ministerium für Kultus, Jugend und Sport Baden-Württemberg (Hrsg.) (2016). *Gemeinsamer Bildungsplan der Sekundarstufe I. Physik*. Stuttgart.

Ministry of Education (2007). *The Ontario Curriculum Grades 1-8: Science and Technology* (The Ministry of Education, Hrsg.).

Minner, D. D., Levy, A. J. & Century, J. (2010). Inquiry-based science instruction-what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching, 47* (4), 474–496.

Miranda, R. J. & Damico, J. B. (2015). Changes in Teachers' Beliefs and Classroom Practices Concerning Inquiry-Based Instruction Following a Year-Long RET-PLC Program. *Science Educator, 24* (1), 23–35.

Moosbrugger, H. & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion*.

Moritz, C. (2014). Vor, hinter, für und mit der Kamera: Viergliedriger Video-Analyserahmen in der Qualitativen Sozialforschung. In C. Moritz (Hrsg.), *Transkription von Video- und Filmdaten in der Qualitativen Sozialforschung* (S. 17–54). Wiesbaden: Springer Fachmedien Wiesbaden.

Müller, S. & Reiners, C. S. (2020). Verständnisfördernde Lernarrangements zu ausgewählten NOS-Aspekten. In S. Habig (Hrsg.), *Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen. Jahrestagung in Wien 2019* (S. 150–153).

Murray, I. & Reiss, M. (2005). The student review of the science curriculum. *School Science Review, 87* (318), 1–11.

Nathan, M. J. (2012). Rethinking Formalisms in Formal Education. *Educational Psychologist, 47* (2), 125–148.

Neubert, S., Reich, K. & Voß, R. (2001). Lernen als konstruktiver Prozess. In T. Hug (Hrsg.), *Einführung in das wissenschaftliche Arbeiten* (S. 253–265). Baltmannsweiler: Schneider-Verlag.

Nott, M. & Wellington, J. (1996). When the black box springs open. Practical work in schools and the nature of science. *International Journal of Science Education, 18* (7), 807–818.

NRC. (2012). *A Framework for K-12 Science Education. Practices, Crosscutting Concepts, and Core Ideas*. Washington, D.C.: National Academies Press. https://doi.org/10.17226/13165

OECD (2017). *PISA 2015 Assessment and Analytical Framework* (OECD Publishing, Hrsg.). Paris.

Osborne, J. (2011): Science teaching methods: a rationale for practices. In: *The School science review* (93), S. 93–103.

Osborne, J. (2014a). Scientific Practices and Inquiry in the Science Classroom. In N. G. Lederman & S. K. Abell (Hrsg.), *Handbook of research on science education* (S. 579–599). New York, NY: Routledge.

Osborne, J. (2014b). Teaching Scientific Practices. Meeting the Challenge of Change. *Journal of Science Teacher Education, 25* (2), 177–196.

Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (2013). *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann Verlag.

Prenzel, M. (2010). Geheimnisvoller Transfer? *Zeitschrift für Erziehungswissenschaft, 13* (1), 21–37.

Priemer, B. (2006). Deutschsprachige Verfahren der Erfassung von epistemologischen Überzeugungen. *Zeitschrift für Didaktik der Naturwissenschaften, 12,* 159–175.

Prince, M. & Felder, R. (2007). The Many Faces of Inductive Teaching and Learning. *Journal of College Science Teaching, 36* (5), 14–20.

R Core Team. (2018). R: A Language and Environment for Statistical Computing [Computer software]. Wien: R Foundation for Statistical Computing.

Regierung des Landes Baden-Württemberg. (2020). Verordnung der Landesregierung über infektionsschützende Maßnahmen gegen die Ausbreitung des Virus SARS-Cov-2. CoronaVO, 1–11.

Reh, S. (2012). Beobachtungen aufschreiben. Zwischen Beobachtungen, Notizen und „Re-writing“. In H. d. Boer & S. Reh (Hrsg.), *Beobachtung in der Schule - Beobachten lernen* (S. 115–129). Wiesbaden: VS Verlag für Sozialwissenschaften.

Renkl, A. (2015). Wissenserwerb. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 3–24). Berlin, Heidelberg: Springer Berlin Heidelberg.

Reusser, K. & Tremp, P. (2008). Diskussionsfeld «Berufliche Weiterbildung von Lehrpersonen». *Beiträge zur Lehrerinnen- und Lehrerbildung, 26* (1), 5–10.

Roberts, R. & Johnson, P. (2015). Understanding the quality of data. A concept map for ‘the thinking behind the doing’ in scientific practice. *The Curriculum Journal, 26* (3), 345–369.

Roth, K. (2009). Using video studies to compare and understand science teaching: Results from the TIMSS video study of 8th grade science teaching. In T. Janik & T. Seidel (Hrsg.), *The power of video studies in investigating teaching and learning in the classroom* (S. 23–37). Münster: Waxmann.

Roth, K. & Garnier, H. (2007). What Science Teaching Looks Like: An International Perspective. *Educational Leadership, 64* (4), 16–23.

Rousseeuw, P. J. (1987). Silhouettes. A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20,* 53–65.

Rushton, G. T., Lotter, C. & Singer, J. (2017). Chemistry Teachers' Emerging Expertise in Inquiry Teaching. The Effect of a Professional Development Model on Beliefs and Practice. *Journal of Science Teacher Education, 22* (1), 23–52.

Ryder, J. & Leach, J. (1999). University science students' experiences of investigative project work and their images of science. *International Journal of Science Education, 21* (9), 945–956.

Sadeh, I. & Zion, M. (2012). Which Type of Inquiry Project Do High School Biology Students Prefer. Open or Guided? *Research in Science Education, 42* (5), 831–848.

Savery, J. R. & Duffy, T. M. (1998). Problem Based Learning: An Instructional Model and Its Constructivist Framework. In B. G. Wilson (Ed.), *Constructivist learning environments. Case studies in instructional design* (S. 135–148). Englewood Cliffs, N.J: Educational Technology Publications.

Schauble, L., Klopfer, L. E. & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching, 28* (9), 859–882.

Schmaltz, C. (2019). *Heterogenität als Herausforderung für die Professionalisierung von Lehrkräften. Entwicklung der Unterrichtsplanungskompetenz im Rahmen einer Fortbildung*. Wiesbaden: Springer.

Schmidkunz, H. & Lindemann, H. (1999). *Das forschend-entwickelnde Unterrichtsverfahren. Problemlösen im naturwissenschaftlichen Unterricht*. Magdeburg: Westarp-Wiss (Original erschienen 1976).

Schmidt, S., Stiller, C. & Wilde, M. (2019). Hilfen beim Experimentieren Auswirkungen unterschiedlicher Arten der Unterstützung auf den extraneous Cognitive Load. *Erkenntnisweg Biologiedidaktik,* 9–23.

Schmiemann, P. & Lücken, M. (2014). Validität – Misst mein Test, was er soll? In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 107–118). Heidelberg: Springer.

Schmitt, A.-K. (2016). *Entwicklung und Evaluation einer Chemielehrerfortbildung zum Kompetenzbereich Erkenntnisgewinnung*. Dissertation. Logos.

Schulz, A. (2010). *Experimentierspezifische Qualitätsmerkmale im Chemieunterricht - Eine Videostudie* (Studien zum Physik- und Chemielernen, Bd. 113). Berlin: Logos.

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.

Seidel, T. (2003). *Lehr-Lernskripts im Unterricht. Freiräume und Einschränkungen für kognitive und motivationale Lernprozesse ; eine Videostudie im Physikunterricht*. Zugl.: Kiel, Univ., Diss., 2002. Münster: Waxmann.

Seidel, T.; Reiss, K. (2014): Lerngelegenheiten im Unterricht. In: Tina Seidel und Andreas Krapp (Hrsg.): Pädagogische Psychologie. Weinheim: Julius Beltz S. 253–295.

Seung-Seok Choi, Sung-Hyuk Cha & Charles C. Tappert (2010). A Survey of Binary Similarity and Distance Measures, *8* (1), 43–48.

Shavelson, R. J. & Stern, P. (1981). Research on Teachers' Pedagogical Thoughts, Judgments, Decisions, and Behavior. *Review of Educational Research, 51* (4), 455–498.

Stiller, J. (2015). *Scientific Inquiry im Chemieunterricht. Eine Videoanalyse zur Umsetzung von Erkenntnisgewinnungsprozessen im internationalen und schulstufenübergreifenden Vergleich*. Humboldt-Universität, Berlin.

Sweller, J., Kirschner, P. A. & Clark, R. E. (2007). Why Minimally Guided Teaching Techniques Do Not Work. A Reply to Commentaries. *Educational Psychologist, 42* (2), 115–121.

Tachtsoglou, S. & König, J. (2017). *Statistik für Erziehungswissenschaftlerinnen und Erziehungswissenschaftler. Konzepte, Beispiele und Anwendungen in SPSS und R* (Lehrbuch). Wiesbaden: Springer Fachmedien Wiesbaden. h

Teig, N., Scherer, R. & Nilsen, T. (2018). More isn't always better. The curvilinear relationship between inquiry-based teaching and student achievement in science. *Learning and Instruction, 56* (1), 20–29.

Terhart, E. (2012). Wie wirkt Lehrerbildung? Forschungsprobleme und Gestaltungsfragen. *Zeitschrift für Bildungsforschung, 2* (1), 3–21.

Tesch, M. (2005). *Das Experiment im Physikunterricht. Didaktische Konzepte und Ergebnisse einer Videostudie*. Zugl.: Kiel, Univ., Diss., 2005. Berlin: Logos.

Tesch, M. & Duit, R. (2004). Experimentieren im Physikunterricht - Ergebnisse einer Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften, 10,* 51–70.

Tries, U. & Upmeier zu Belzen, A. (2009). „Wissenschaftler nutzen Modelle, um etwas Neues zu entdecken, und in der Schule lernt man einfach nur, dass es so ist.“. *Erkenntnisweg Biologiedidaktik* (9), 23–38.

Tuma, R., Schnettler, B. & Knoblauch, H. (2013). *Videographie. Einführung in die interpretatiove Videoanalyse sozialer Situationen*. Wiesbaden: Springer.

Turner, R. C., Keiffer, E. A. & Salamo, G. J. (2018). Observing Inquiry-Based Learning Environments Using the Scholastic Inquiry Observation Instrument. *International Journal of Science and Mathematics Education, 16* (8), 1455–1478.

Tversky, A. & Kahneman, D. (1992). Advances in prospect theory. Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5* (4), 297–323.

Vorholzer, A. (2017). Sprechen Sie über die Regeln! Zur Relevanz der expliziten Thematisierung von Regeln zum naturwissenschaftlichen Denken und Arbeiten. *Naturwissenschaften im Unterricht. Chemie, 28* (158), 34–38.

Vorholzer, A. & Aufschnaiter, C. von (2019). Guidance in inquiry-based instruction – an attempt to disentangle a manifold construct. *International Journal of Science Education, 41* (11), 1562–1577.

Vorholzer, A., Aufschnaiter, C. von & 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.

Vorholzer, A. S. (2015). *Wie lassen sich Kompetenzen des experimentellen Denkens und Arbeitens fördern?* Dissertation.

Wagner, R. (2016): Unterricht aus Sicht der Lehrerinnen und Lehrer. Subjektive Theorien zur Unterrichtsgestaltung und ihre Veränderung durch ein Training zu neuen Unterrichtsmethoden. Bad Heilbrunn: Verlag Julius Klinkhardt.

Wahser, I. (2007): Training von naturwissenschaftlichen Arbeitsweisen zur Unterstützung experimenteller Kleingruppenarbeit im Fach Chemie. Eine experimentelle Laborstudie. Universität Duisburg-Essen.

Wahser, I. & Sumfleth, E. (2008). Training experimenteller Arbeitsweisen zur Unterstützung kooperativer Kleingruppenarbeit im Fach Chemie. *Zeitschrift für Didaktik der Naturwissenschaften, 14.*

Walpuski, M. (2006). *Optimierung von experimenteller Kleingruppenarbeit durch Strukturierungshilfen und Feedback*. Berlin: Logos. Zugriff am 20.11.2018.

Watson, R. J., Swain, J. R. L. & McRobbie, C. (2007). Students' discussions in practical scientific

Wee, B., Shepardson, D., Fast, J. & Harbor, J. (2007). Teaching and Learning About Inquiry. Insights and Challenges in Professional Development. *Journal of Science Teacher Education, 18* (1), 63–89.

Widodo, A. & Duit, R. (2004). Konstruktivistische Sichtweisen vom Lehren und Lernen und die Parxis des Physikunterrichts. *Zeitschrift für Didaktik der Naturwissenschaften, 10,* 233–255.

Windschitl, M., Thompson, J. & Braaten, M. (2008). How Novice Science Teachers Appropriate Epistemic Discourses Around Model-Based Inquiry for Use in Classrooms. *Cognition and Instruction, 26* (3), 310–378.

Wirtz, M. A. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.

Yew, E. H.J. & Goh, K. (2016). Problem-Based Learning. An Overview of its Process and Impact on Learning. *Health Professions Education, 2* (2), 75–79.

Zehetmeier, S. (2008). *Zur Nachhaltigkeit von Lehrer/innenfortbildung*. Dissertation. Alpen-Adria-Universität, Klagenfurt.

Zion, M. & Mendelovici, R. (2012). Moving from structured to open inquiry: Challenges and limits. *Science Education International, 23* (4), 383–399.

Zucker, A. A., Shields, P. M. & Adelmann, N. E. (1998). *Evaluation of NSF's Statewide Systemic Initiatives (SSI) Program: The SSIs' impacts on classroom practice.* Menlo Park: National Science Foundation.

8 Danksagung

Von manchen Menschen bekommt man im Leben mehr gegeben als man wohl je zurückgeben kann. Dazu zählen zweifellos die eigenen Eltern. Hier aber auch meine beiden Betreuer Prof. Dr. Armin Baur und Prof. Dr. Markus Emden. Durch ihr gelebtes Vorbild, die herausragende fachliche Betreuung aber auch durch Diskussionen und den täglichen Umgang auf persönlicher Ebene bin ich – ganz einfach formuliert aber zutreffend – zu einem besseren Menschen geworden. Für ihre Betreuung während meiner Promotion werde ich immer dankbar sein.

Bedanken möchte ich mich außerdem bei der Deutschen Forschungsgemeinschaft (DFG) welche durch ihre Förderung (Projektnummer 375137339) das Projekt in diesem Rahmen erst möglich gemacht hat.

Prof. Dr. Manuela Welzel-Breuer sowie Prof. Dr. Lutz Kasper danke ich für die stete Unterstützung meiner akademischen Laufbahn. Sei es durch gemeinsame Publikationstätigkeit, Tipps für Förderungen oder das informelle Gespräch auf dem Flur oder später auf der Tagung.

Ich danke den Lehrkräften, trotz teils starker Mehrbelastung, für ihre freiwillige Teilnahme an dieser Studie und die Unterstützung meiner Forschung.

Meinem Kommilitonen David Weiler danke ich für die Zusammenarbeit und den Austausch in den tollen Projekten, welche wir seit unserem gemeinsamen Studium immer wieder starten.

Meinen Eltern Ulrike und Klaus Bewersdorff danke ich für die Entwicklungschancen welche sie mir gegeben haben. Mir ist bewusst welches Privileg es ist, in einem Haus aufgewachsen zu sein, wo Bildung und Bücher einen hohen Stellenwert haben. Ich danke auch für die vielfältige Unterstützung in den letzten Jahren.

Ich danke meinen guten Freunden Matthias und Bastian, Anna und Hannes, Georg und Clara dafür, dass sie meine guten Freunde sind.

Der große Dank geht an meinen Bruder Hannes Bewersdorff von dem ich so viel lernen durfte.

9 Abbildungsverzeichnis

Abbildung 1: Übersicht über die abgeleiteten Unterrichtsziele und die zugehörigen Unterrichtsprinzipien 21
Abbildung 2: Kontinuum der Selbstständigkeit beim Experimentieren 30
Abbildung 3: Ebenen des Transferprozesses nach Lipowsky (2014) 33
Abbildung 4: Die vier Transferwirkungen 34
Abbildung 5: Literaler Transfer und Figuraler Transfer 35
Abbildung 6: : Einflussfaktoren von Lehrkräftefortbildungen auf Lehrende und Lernende 37
Abbildung 7: Ansätze zur Testung von Transferwirkungen in der Wirkungskette der Lehrerbildung nach Galluzzo und Craig (1990) 40
Abbildung 8: COACTIV-Mediationsmodell der professionellen Kompetenz 52
Abbildung 9: Integriertes Wirk- und Gestaltmodell effektiver Lehrkräftefortbildung 52
Abbildung 10: Die Phasen der Lehrkräftefortbildung sowie die jeweiligen Unterschiede der beiden Formate 55
Abbildung 11: Reduktion der Stichprobe durch die SARS-Cov-2-Pandemie 56
Abbildung 12: Skizze zur Positionierung der Videokameras im Klassenraum ... 67
Abbildung 13: Originalmaterial, Paraphrasierung, Gruppierung, Generalisierung. 72
Abbildung 14: Überblick über die Prozessschritte bei der Entwicklung des Kategoriensystems 73
Abbildung 15: Kodierung der Lehrkräfte bzw. Unterrichtsstunden 105
Abbildung 16: Clusteranalyse „Effektiver Unterricht zum Experimentieren“ .. 106
Abbildung 17: Clusteranalyse „Hypothesenbildung“ 110
Abbildung 18: Clusteranalyse „Aufforderung zu wissenschaftsorientiertem Arbeiten“ 112
Abbildung 19: Clusteranalyse „Metareflexion“ 114
Abbildung 20: Clusteranalyse „Unterstützte Offenheit“ 116
Abbildung 21: Die Teilprozesse ‚Problemstellung‘, ‚Fragestellung‘, ‚Hypothesenfindung‘ und ‚Planung‘‘ 137
Abbildung 22: Der Teilprozess ‚Durchführung und Beobachtung‘‘ 138
Abbildung 23: Die Teilprozesse ‚Auswertung‘, ‚Schlussfolgerung‘ und ‚Präsentation und Reflexion von Erkenntnissen‘ 139
Abbildung 24: Die Kategorien ‚Fach- und Methodenwissen‘, ‚Organisatorisches‘, ‚kein Unterricht mit naturwissenschaftlichem Bezug‘ und ‚Restkategorie‘ als Boxplot 141
Abbildung 25: Die Entwicklung der Öffnungsgrade für die Teilstichproben sowie die Gesamtstichprobe 144

Abbildung 26: Entwicklung der Aspekte des Experiments für die Teilstichproben sowie die Gesamtstichprobe. 147

10 Tabellenverzeichnis

Tabelle 1: Zweidimensionale Matrix zur Planung und Klassifizierung eines Experiments 31
Tabelle 2: Zusammenschau von Studien zur Wirkung von Lehrkräftefortbildungen zu *Inquiry* 42
Tabelle 3: Die Charakteristika ‚eigenaktives Lernen', ‚kollegiale Teilnahme' und ‚Dauer' nach Desimone und Garet (2015) 46
Tabelle 4: Übersicht über existierende Kodier- und Ratingmanuale 50
Tabelle 5: Niedrig und hoch inferente Verfahren 66
Tabelle 6: Übersicht über die verwendeten Kategoriensysteme 69
Tabelle 7: Aufschlüsselung der für die zusammenfassende Inhaltsanalyse verwendeten Videos 72
Tabelle 8: Verteilung der Stichprobe zur Bestimmung der Beobachterübereinstimmung 77
Tabelle 9: Schematischer Aufbau einer Kreuztabelle mit zwei Schätzern und dichotomer Merkmalsausprägung 78
Tabelle 10: Wertebereiche und Interpretation von Cohens Kappa 79
Tabelle 11: Beurteilerübereinstimmung für die einzelnen Kategorien 79
Tabelle 12: Kategorien mit κ-Wert < 0,74, prozentualer Übereinstimmung > 90 % und rel. Häufigkeit < 15 % oder > 85 % 82
Tabelle 13: Beurteilerübereinstimmung für die einzelnen Kategorien nach erneuter Besprechung abweichender Kodierungen 82
Tabelle 14: Kategorien, welche aus Sicht eines der fünf Experten keinen Beitrag zum Erfassen von Unterrichtshandeln von Lehrkräften leisten können 90
Tabelle 15: Die Kategorien des Kategoriensystems Nr. 2 „Phasen des Experimentierprozesses" 91
Tabelle 16: Kreuztabelle der Kodierentscheidungen von Beobachter 1 und Beobachter 2 93
Tabelle 17: Beurteilerübereinstimmung (%Ü, κ) für die Kategorien im Kategoriensystem Nr. 2 „Phasen des Experimentierprozesses" 94
Tabelle 18: Gegenüberstellung der Definitionen der Kategorien „Fach- u. Methodenwissen" sowie „Problemstellung" 95
Tabelle 19: Lineare Gewichte bei der Berechnung von Cohens Kappa (κ) bei einer 4-stufigen Likert-Skala 97
Tabelle 20: Beurteilerübereinstimmung (%Ü, κ) für die einzelnen Kategorien des Kategoriensystems 3 97
Tabelle 21: Beurteilerübereinstimmung (%Ü, κ) für die einzelnen Kategorien des Kategoriensystems 4 97
Tabelle 22: Schematischer Aufbau der Kontigenztabelle für die binären Vektoren $\vec{a}$ und $\vec{b}$ 99

Tabelle 23: Videografierte Unterrichtsstunden im individualisiert-konstruktiven Format 102
Tabelle 24: Videografierte Unterrichtsstunden im kokonstruktiven Format 103
Tabelle 25: Dimension ‚Prozessstruktur' (nach Lehrkraft aufgeschlüsselt) 119
Tabelle 26: Dimension ‚Reflexion' (nach Lehrkraft aufgeschlüsselt) 119
Tabelle 27: Dimension ‚Angemessenheit und Unterstützung' (nach Lehrkraft aufgeschlüsselt) 120
Tabelle 28: Effektiver Unterricht zum Experimentieren (gesamtes Kategoriensystem) 121
Tabelle 29: Auftrittshäufigkeiten der Kategorien in Prozent pre und post sowie die Änderung in Prozentpunkten und die prozentuale Änderung für die Kontrollgruppe 122
Tabelle 30: Auftrittshäufigkeiten der Kategorien in Prozent pre und post sowie die Änderung in Prozentpunkten und die prozentuale Änderung für die Testgruppe 124
Tabelle 31: Auftrittshäufigkeiten der Kategorien in Prozent pre und post sowie die Änderung in Prozentpunkten und die prozentuale Änderung für die Gesamtstichprobe 125
Tabelle 32: Interpretation von *d* nach Cohen (1988) 127
Tabelle 33: Oberkategorie ‚Problem und Fragestellung' 128
Tabelle 34: Oberkategorie ‚Hypothesen aufstellen und besprechen' 128
Tabelle 35: Oberkategorie ‚Entwickeln einer Untersuchung' 129
Tabelle 36: Oberkategorie ‚Wissenschaftsorientiertes Arbeiten anregen' 129
Tabelle 37: Oberkategorie ‚Fixierung im Erkenntnisprozess' 129
Tabelle 38: Dimension ‚Prozessstruktur' 130
Tabelle 39: Oberkategorie ‚Mesoreflexion' 130
Tabelle 40: Oberkategorie ‚Metareflexion' 131
Tabelle 41: Dimension ‚Reflexion' 131
Tabelle 42: Oberkategorie ‚Lehrkraft als Ansprechpartner, Trainer und Experte' 132
Tabelle 43: Oberkategorie ‚Unterstützte Offenheit' 132
Tabelle 44: Dimension ‚Angemessenheit und Unterstützung' 132
Tabelle 45: Auftreten der einzelnen Teilprozesse in den untersuchten Unterrichtsvideografien 134
Tabelle 46: Auftreten der einzelnen Teilprozesse in den Teilstichproben 135
Tabelle 47: Unterrichtszeit Teilprozess ‚Einführung in das Problem' (Problemstellung) 136
Tabelle 48: Unterrichtszeit Teilprozess ‚Fragestellung' 136
Tabelle 49: Unterrichtszeit Teilprozess ‚Hypothesenfindung' 136
Tabelle 50: Unterrichtszeit Teilprozess ‚Planung' 137
Tabelle 51: Unterrichtszeit Teilprozess ‚Durchführung und Beobachtung' 138

Tabelle 52: Unterrichtszeit Teilprozess ‚Auswertung' 138
Tabelle 53: Unterrichtszeit Teilprozess ‚Schlussfolgerung' 139
Tabelle 54: Unterrichtszeit Teilprozess ‚Präsentation und Reflexion von Erkenntnissen' 139
Tabelle 55: Unterrichtszeit Teilprozess ‚Fach- und Methodenwissen' 140
Tabelle 56: Unterrichtszeit Teilprozess ‚Organisatorisches' 140
Tabelle 57: Unterrichtszeit Teilprozess ‚Kein Unterricht mit naturwissenschaftlichem Bezug' 140
Tabelle 58: Unterrichtszeit Teilprozess ‚Restkategorie' (nicht zuweisbar) 140
Tabelle 59: Entwicklung des Öffnungsgrades der Teilprozesse beim Experimentieren 143
Tabelle 60: Auswertung des Kategoriensystems ‚Definition des Experiments' 146

11 Anhang

11.1 Kategorien- und Ratingsysteme ‚Unterrichtshandeln von Lehrkräften beim Experimentieren', ‚Öffnungsgrade der Teilprozesse beim Experimentieren' und ‚Definition des Experiments'

Arne Bewersdorff, PH Heidelberg

Kodiermanual

Kategoriensysteme zum Unterrichtshandeln von Lehrkräften beim geöffneten Experimentieren

Stand: 07. Dezember 2020

Die Kategoriensysteme ‚*Prozessstruktur*', ‚*Prozessreflexion*', ‚*Angemessenheit, Kommunikation und Steuerung*', sowie die Wertungsraster ‚*Öffnungsgrade*' und ‚*Definition Experiment*' finden jeweils auf der gesamten Transkription einer Unterrichtsstunde Anwendung.

Ziele zur Erfassung im Rahmen der einzelnen Dimensionen:

Prozessstruktur:

- Sichtbarkeit der Teilprozesse des Experimentierens und der ihr von der Lehrkraft einbeschriebenen Konzepte
- Orientierung am wissenschaftlichen Arbeiten

Prozessreflexion:

- Gelegenheiten zur Meso- und Metareflexion im Rahmen der Unterrichtsstunde
- Beziehungen von Teilprozessen im fortschreitenden Erkenntnisprozess

Angemessenheit und Unterstützung

- Strukturierung des Erkenntnisprozesses
- Unterstützung sowie Eingriffe in die Schülerselbstständigkeit
- Bereitstellung von Hilfsangeboten

≡ drei Optionen: Auftreten, nicht auftreten, nicht anwendbar (wenn der Teilprozess nicht Gegenstand der Unterrichtsstunde ist)

Dimension „Prozessstruktur“			
Oberkategorie	Kategorie	Definition	Ankerbeispiel
Problem und Fragestellung	**S1**: Ein Problem (als Anstoß einer Untersuchung) wird aufgeworfen/identifiziert	Beim Problemlösen muss etwas Neues geschaffen werden, Aufgaben erfordern nur reproduktives Denken: Ein Problem ist jeder Zweifel, jede Schwierigkeit oder Unsicherheit, die eine Art von Lösung erfordert. ≡	*Lehrperson: Was könnte denn, das Problem hier sein? S15: Das man halt während dem Autofahren nicht trinken darf, nicht rauchen darf und nicht telefonieren darf und das der das halt [der Autofahrer im Bild] alles auf einmal macht [...]* *Lehrperson: Warum ist das denn gefährlich? S14: Weil ein Unfall passieren kann [...] //* *Lehrperson: Wie können wir herausfinden, dass [...]*
	S1A: Ein Problem wird aufgeworfen/identifiziert („Ingenieurmodus“)	Es soll von S ein Nachweis erbracht, ein Effekt maximiert werden. Regel: nur kodieren wenn auch **S1** kodiert! ≡	*Lehrperson: Eure Aufgabe ist es, herauszufinden, in welchen Lebensmitteln [...] hier ist jetzt eher gesund, dann habe ich jetzt hier zum Beispiel Erdnüsse. //* *Lehrperson: Ihr seht [auf dem Whiteboard], so einen Wissenschaftler. Und der einfach nicht mehr weiß, in welches Glas er Mineralwasser, abgekochtes Wasser oder Leitungswasser rein hat.*
	S2: Eine Fragestellung wird präsentiert/hergeleitet	Eine Fragestellung wird von L angeboten oder im Plenum entwickelt ≡	*Lehrperson: Die zentrale Frage ist nun, bekomme ich da einen Stromschlag?*
	S2A: Eine naturwissenschaftliche Fragestellung wird präsentiert/hergeleitet	Eine naturwissenschaftliche Fragestellung* wird von L angeboten oder im Plenum entwickelt Regel: nur kodieren wenn auch **S2** kodiert! ≡ * Eine Fragestellung, welche sich auf naturwiss. Weg (z.B. Untersuchung, Beobachtung, Experiment) beantworten lässt (OECD, 2007)	*Lehrperson schreibt etwas an die Tafel und liest es vor.* *Lehrperson: "Reagieren wir langsamer, wenn wir abgelenkt werden?"*
Hypothesen aufstellen, besprechen	**S3**: Hypothesen, Vermutungen, Ideen werden (im Kontext eines Versuchs) aufgestellt	S oder L nennen Vermutungen/Ideen (vorläufige Annahmen aufgrund von	

		Theoriewissen oder bisherigen Erfahrungen), diese werden evtl. notiert	
	S3A: Hypothesen werden (im Kontext eines Versuchs) aufgestellt -> ausgerichtet auf Beobachtung	Die Vermutung ist eine Annahme, ‚was bei dem Versuch passiert'. Regel: nur kodieren wenn auch **S3** kodiert!	*S13: Das schmeckt vielleicht anders.*
	S3B: Hypothesen werden (im Kontext eines Versuchs) aufgestellt -> mit Begründung	Vermutungen werden inkl. Begründung genannt Regel: nur kodieren wenn auch **S3** kodiert!	*S 16: In den Kartoffeln ist Stärke drin, weil, es gibt ja auch Kartoffelstärke.*
	S4: Problematischer Hypothesenbegriff von L: Aufforderung zum Raten; nicht wissenschaftlich; Zweck des Versuchs	L lässt S Raten um was es geht, Nennung von Themen	*Lehrperson: Unterhaltet euch bitte mal mit eurem Nachbarn. Warum zeige ich euch diesen Film zu (unv.) Stunde? //*
	S5: Begründungen für Hypothesen werden von L eingefordert	L fordert S auf seine Vermutung/Idee zu Begründen	*L: Kannst du deine Idee auch begründen? Warum denkst du, dass die Knetmasse untergeht? //* *L: Und bei der Vermutung möchte ich nicht nur ein Ja oder Nein hören, ich möchte ein Ja, weil hören oder ein Nein, weil oder es hängt davon ab, dass..*
Entwickeln einer Untersuchung	**S6**: Ideen zur Durchführung werden erfragt/ genannt/ gesammelt	Skizzieren, erfragen, äußern und sammeln von gedanklichen Entwürfen zur praktischen Umsetzung	*Lehrperson: Okay, dann sagen nochmal die Gruppen was sie für eine Idee hatten. Wie sie ihre Vermutung überprüfen wollen. S2: wegen den Sonnenstrahlen. Lehrperson: Was meintet ihr mit 'wegen den Sonnenstrahlen'? Was habt ihr denn für einen Versucht überlegt? S2: Also wir stellen die Pflanze für ein paar Tage aufs Fensterbrett. Damit die Sonnenstrahlen (unv.). S1: Und dass sie nicht so schnell wächst, kann man sie auch (unv.)*

			stellen, auf ein Regal oder so. (unv.) Lehrperson: Damit ihr schauen könnt, ob es dann gerade wächst. Gute Idee.
	S7: Durchführung / Planung wird (vor)besprochen	Gespräch über die konkrete praktische Umsetzung	*Lehrperson: Welche Gruppe möchte denn ihre Lösungsmöglichkeit vorstellen? S17: […] Dann würden wir jetzt die Stoppuhr auf drei Minuten stellen und diese Pflanze mit, gleichzeitig reinmachen [in das Becherglas] wie das Licht anmachen und dann zählen wir, wie viele Bläschen dabei entstehen […].*
Wissenschafts-orientiertes Arbeiten anregen	**S8**: Achtung einer Vergleichbarkeit	Vergleichbarkeit (der Ansätze; bei Messwiederholung) wird gewahrt / Aufforderung zur Vergleichbarkeit	*Lehrperson: Wenn wir den Versuch so durchführen, er dann vergleichbar? Können wir unsere Werte vergleichen ohne Probleme? S4: Nein, weil manche einen größeren oder kleinere Trichter haben oder manche mehr oder weniger Wasser haben. Lehrperson: Was könnte man trotzdem das nicht alle Filter gleich sind, wo könnte man trotzdem vergleichen oder wer könnte sich vergleichen? S16: Dann muss man überall gleich viel Erde rein machen.*
	S9: Kontrollansatz: Aufforderung zur Umsetzung	L weist S auf (die Notwendigkeit) eines Kontrollansatzes hin Sprechen über Zweck des Kontrollansatzes R11 kodieren!	*Lehrperson: Okay, jetzt haben wir noch eine Pflanze rumstehen (unv.). Der Kontrollversuch, da könnt ihr euch schon mal was überlegen. // Lehrperson: Denkt ihr auch an den Kontollansatz? Wie könnte der aussehen?*
	S10: Aufforderung/ Hinweis zu protokollieren.	L ist es wichtig, dass die S ein Protokoll führen und weist darum (mehrmals, im Plenum,…) darauf hin, dass ein Versuchsprotokoll zu führen ist.	*Lehrperson: Okay. Dann fangt ihr mit eurem Protokoll an. Da macht ihr eure Skizze die malt ihr bitte ab und dann könnt ihr unten eure Ergebnisse eintragen.*

		Einmalige Aufforderung an ein S etwas zu notieren nicht ausreichend!	
	S11: Einschränkungen und Generalisierbarkeit der Beobachtung/Ergebnisse (Vorläufigkeit d. Wissens, Grenzen) werden diskutiert	Möglichkeiten der Verallgemeinerung/ der Beschränkung bzgl. einer allgemeinen Gültigkeit werden diskutiert	*Lehrperson: Katzen riechen gut?* *SuS. Ja.* *Lehrperson: Das Wissen wir jetzt im Moment noch nicht, das haben wir noch nicht überprüft.* (Baur, Emden, Bewersdorff 2019)
Fixierung im Erkenntnisprozess	**S12**: Erfragen und besprechen von Beobachtungen	L erfragt Beobachtungen, L spricht mit S über Beobachtungen ≡	*Lehrperson: Okay, was habt ihr aufgeschrieben bei 'Beobachtung vorher'? [...] S10: Ich habe bei dem normalen Luftballon, der ist normal geblieben...*
	S13: Festhalten von Beobachtungen / (Mess)Ergebnissen	L fordert auf Beobachtungen aufzuschreiben; Beobachtungen /(Mess)Ergebnisse werden notiert ≡	*Lehrperson: Vorne die Flaschen in eine der roten Tasche hinein und dann aufschreiben [die Beobachtung]. // L erfragt den Messwert einer Gruppe und notiert ihn an der Tafel*
	S14: Erkenntnisse werden notiert	L fordert S auf Erkenntnisse zu notieren; S notieren Erkenntnisse ≡	*Lehrperson: So fünf Minuten, dann wollen wir zum Schluss kommen. Eure Schlussfolgerung sollte notiert sein oder gemalt sein oder was weiß ich.*
	S15A: Teilprozesse werden *im Prozess* schriftlich festgehalten	S fixieren im Erkenntnisprozess wesentliche Überlegungen auf einem Protokoll/Arbeitsblatt/Heft (>25%, Intention von L)	*Lehrperson: Genau diese Vermutung sollt ihr auf euer Versuchsprotokoll übernehmen. [später..] Lehrperson: Wenn ihr besprochen habt, dürft ihr auch gerne schon das Material notieren, auf eurem Protokoll.*
	S15B: Teilprozesse werden *abschließend* schriftlich festgehalten	S fixieren im Anschluss an das Experiment wesentliche Überlegungen auf einem Protokoll/Arbeitsblatt/Heft (>25%, Intention von L) ≡	

Dimension Prozessreflexion (meso und meta)				
Oberkategorie		Kategorie	Definition	Ankerbeispiel
Mesoreflexion	**Vermutungen: Zusammenfassung, Klärung**	**R1**: Vermutungen/Ideen werden zusammengefasst/ wiederholt	Die aufgestellten Vermutungen/Ideen werden vorgelesen/an der Tafel gesammelt	*Lehrperson: Okay, dann fassen wir nochmals zusammen: S5, lese doch mal vor, was eure Vermutungen sind. S 5: 'Warum werden im Herbst die Blätter bunt? Wegen der Photosynthese; Sonne und Sauerstoff spielen eine Rolle; Färbung hat mit Wärme und Kälte zu tun; Macht die Natur selbst; Reste der Nährstoffe werden an Blätter weiter gegeben.'*
		R2: Vermutungen werden verifiziert/falsifiziert	Rückbezug auf zuvor aufgestellte Vermutungen: Verifizierung/ Falsifizierung ≡	*Lehrperson: Und ‚macht die Natur es selbst?' S4: Ja, eigentlich schon. Lehrperson: Eigentlich schon, also passt auch. Lehrperson setzt ein Haken hinter die Hypothese am Whiteboard.*
	Besprechung und Reflexion der Durchführung	**R3**: Erfragen von Problemen bei/nach der Durchführung	Probleme (praktisch/wissenschaftlich) bei der Durchführung werden angesprochen	*Lehrperson: Was ist da das Problem, was müsste man eigentlich schauen? S15: Den Trichter hoch, aber wir haben keinen passenden gefunden.*
		R4: Besprechung von Verbesserungs-möglichkeiten nach der Durchführung	Verbesserungsmöglichkeiten/Möglichkeiten der Optimierung bei der Durchführung werden von L erfragt, diskutiert ≡	*Lehrperson: Die[Gruppe von S11] können nicht richtig ablesen. Was brauchen wir dann nächstes Mal auch? Damit wir das können, weil das ist keine Einheit mehr. S5: einen kleinen Messbecher.*
		R5: Berichten und Besprechen nach Durchführung	Die Durchführung wird nach dem Experiment vorgestellt/besprochen ≡	

		R6: Kommentierung/ Besprechung von Produkten und Objekten	Hergestellte Produkte (Boote, Glasarbeiten) werden kommentiert, besprochen, getestet ≡	*Lehrperson: Erstmal jetzt kurz Aussehen bitte und Idee S4: Es hat Augen. S5: Nen Mund und einen Namen. S2: Unser Rettungsboot. Lehrperson: Okay, okay. Von der Idee her 1-10?*
	Besprechen und reflektieren von Erkenntnissen	**R7**: Erfragen und besprechen von Erkenntnissen	Erkenntnisse (Schlussfolgerungen) werden von L erfragt bzw. aus den gewonnenen Daten/ Beobachtungen abgeleitet, diskutiert und beurteilt ≡	*Lehrperson: Also was merken wir uns heute? S2: Die Erde dreht sich und damit ist auf unterschiedlichen Seiten Tag und Nacht.*
Metareflexion	**Vermutungen: Begriffsklärung**	**R8**: Begriffsklärung/ Problematisierung „Vermutung“	Der Begriff wird im Verständnis der SuS ausgeschärft (Passung Vermutung-Fragestellung, Begriffserklärung, Wissenschaftlichkeit…)	
	Sichtbarmachung Prozessstruktur	**R9**: Wiederholung Teilprozesse	Teilprozesse werden genannt; Frage, welcher Teilprozess folgt.	*Lehrperson zeigt [wiederholt] auf Plakat mit den Teilprozessen beim Experimentieren. // Lehrperson: Aber vielleicht erinnert sich der eine oder andere noch, welche Bausteine, welche Teile solch ein Versuchsprotokoll hat. S12: Als erstes kommt die Frage oder die Vermu…ja als erstes die Frage. Fragestellung […]*
		R10: Bezug einzelner Teilprozesse im Prozess aufeinander durch L	Teilprozesse werden aufeinander Bezogen Nicht Kodieren wenn R2 zutrifft	*Lehrperson: So, wir schauen ganz am Ende auf eure Vermutungen und schauen ob da sogar die richtige Lösung dabeistand. // Lehrperson : Ist es dann sinnvoll die Zeit zu stoppen, bis das erste Bläschen da ist? Jetzt überlegt einmal, dass es auch zu der Vermutung auch passt.*
	Kontrollansatz: Begriffsklärung	**R11**: Klären von Funktion und Ausgestaltung eines Kontrollansatzes	Die Funktion und Ausgestaltung eines Kontrollansatzes wird besprochen/wiederholt	*Lehrperson: Brauchen immer einen Kontrollansatz, um das zu Vergleichen. Wir verändern hier unsere Bedingungen sodass wir glauben, jetzt, so glaube ich, dass die Pflanze schräg wächst. Brauchen wir natürlich eine Pflanze im Vergleich […]*

	Üben	**R12**: Das Experimentieren wird explizit eingeübt	Teilaspekte des Experimentierens sind Gegenstand von Übungen	-
Dimension Angemessenheit und Unterstützung				
Oberkategorie		Kategorie	Definition	Ankerbeispiel
L als Trainer, Ansprechpartner und Experte	**trivial**	**K1**: L gibt die Beobachtung vor	L sagt S was diese gefunden haben (würden), offenbart Beobachtung des Experiments	*Lehrperson: Dann haben S sich überlegt, das könnte der beste Platz sein, dass von allen Richtungen, überall, Licht kommt und die sich nicht krumm zur Seite biegt. // Lehrperson: Wunderbar. Super. Da steigt es [die mit dem Amperemeter gemessene Spannung] an. (Horizon Research, 2005)*
	Unter-stützen/ Eingriff	**K2**: L greift händisch in die Durchführung ein	L bearbeitet/manipuliert händisch den Versuchsaufbau von S Regel: nicht kodieren wenn zur Gefahrenabwehr	*Lehrperson wendet sich S11, S12 und S13 zu und spricht mit ihnen und manipuliert ihren Versuchsaufbau.*
	Hinweise	**K3**: Umgang mit/Funktionsweise von Labormaterial	Hinweise zur Verwendung von Labor- und Verbrauchsmaterial. Erklärung der Funktionsweise von Labormaterial	*Lehrperson: Schraubt es auf und leert es [das dest. Wasser] hinein, das geht schneller.*
	Verweis (jeweils nur nach Inter-aktionen der SuS kodieren!)	**K4**: Verweis auf Selbsttätigkeit	Lehrperson bedeutet S das Problem/die Frage eigenständig zu lösen	*Lehrperson: Stimmt. Beobachte einfach mal und dann schau, ob du selber drauf kommst. (unv.).*
		K5: Verweis auf Material	Lehrperson verweist als Tipp auf zur Verfügung gestellte (Labor)Materialien Regel: Hilfen (Kärtchen) sind keine Materialien	*Lehrperson: Schaut doch mal auf das Material hier. Was könnte man damit machen?*
		K6: Verweis auf AB	Lehrperson weist auf Arbeitsblatt, Protokoll, Anleitung hin	*Lehrperson: Ihr braucht diese Anleitung. Das sieht dann nachher praktisch so aus.*
		K7: Verweis auf Hilfen	Lehrperson weist auf bereitgestellte Hilfen hin	*S8: Was ist das für ein Stoff? Lehrperson: Herausfinden. Da drüben wo die Aushänge hängen. Zum Informieren. //*

				Lehrperson: So also diejenigen, die gar nicht weiter kommen, ich habe hier vorne ein paar Bücher, die durchaus helfen könnten.
		K8: Verweis auf andere Gruppen	Lehrperson verweist darauf andere Gruppen zu fragen	*Lehrperson: Nein, ich würde sagen experimentiert mal. Und wenn man gar nicht weiter kommt, schaut man vielleicht bei der Nachbargruppe etwas ab oder holt sich ein Tipp bei mir.*
Unterstützte Offenheit	**Material-beschaffung**	**K9:** Freiheit bzgl. der Materialauswahl	L gibt S Freiheit(en) bei der Planung bzgl. der Materialauswahl, S haben mehr Material (Handlungsoptionen bzgl. des Materials) zur Verfügung als für die Durchführung notwendig	*Schülerinnen und Schüler gehen nach vorne an das Pult und holen sich ihr Material. Lehrperson: Ihr nehmt einfach das, was ihr denkt, was ihr braucht. Ihr müsst auch nicht alles nehmen. // S2: Das dürfen wir aber nicht zerschneiden oder? S2 zeigt auf ein Tetrapack. Lehrperson: Doch, ihr dürft alles machen, was ihr wollt damit.*
	Kochrezept	**K10:** L gibt Versuchsanleitung welche kochrezeptartig strukturiert ist	L gibt detaillierte Anweisungen (verbal, schriftlich) sodass S die von L gewünschten Ergebnisse zu erzielen	*Lehrperson: Wichtig ist, dass ihr euch natürlich das Material durchlest bevor ihr startet. Das es auch so funktioniert, wie es funktionieren soll.* *(Alexakos, 2010; Metzger & Sommer, 2010)*
	Hilfen und Strukturierungs-angebote	**K11:** L weist auf die Möglichkeit der Inanspruchnahme von Hilfen hin	L gibt den Hinweis, dass Hilfen bereitliegen, auf die zurückgegriffen werden darf	*Lehrperson: […] Danach stellt eine Gruppe ihr Ergebnis vor und wenn ihr nicht weiterkommt, dürft ihr vorne eine Tippkarte holen. (Lazonder & Harmsen, 2016).*
		K12: L gibt S ein Protokollbogen aus	S haben ein Protokollbogen zur Verfügung / S ist eine Strukturierung der Teilprozesse gegeben	*Lehrperson: An diesem Protokoll könnt ihr euch beim Experimentieren orientieren (Lazonder & Harmsen, 2016).*

Kategoriensystem Definition Experiment (Baur, Ehrenfeld, Emden, Hummel & Krieg, 2017)					
Aspekt	Voll erfüllt (1)	Zum Teil erfüllt (2)	Nicht erfüllt (3)	N/A (4)	Kommentar
E1: Naturwissenschaftliches Experiment = eine **Untersuchung** eines Naturphänomens, **bei der alle Faktoren bis auf einen** (oder wenige) **konstant gehalten werden.**					
E2: Ein Experiment besteht aus einem **Kontrollansatz** (alle Faktoren konstant)					
E3: und einem oder mehreren **Testansätzen** (ein oder wenige Faktoren werden gezielt verändert).					
E4: Das Experiment ist wiederholbar					
E5: und liefert bei einer Wiederholung (nahezu) gleiche Ergebnisse.					

Kategoriensystem Öffnungsgrade (Baur et al., 2017)				
	0 – geschlossen	**1 – leicht geöffnet**	**2 – geöffnet**	**3 - offen**
G1: Fragestellung	Schüler/innen beschäftigen sich mit einer vorgegebenen Fragestellung.	Schüler/innen wählen aus verschiedenen vorgegebenen Fragestellungen aus.	Schüler/innen erstellen mit Hilfestellung (medial, verbal) eine Fragestellung.	Schüler/innen entwickeln eine Fragestellung.
G2: Hypothese	Schüler/innen beschäftigen sich mit einer vorgegebenen Hypothese.	Schüler/innen wählen aus verschiedenen vorgegebenen Hypothesen aus.	Schüler/innen stellen mit Hilfestellung (medial, verbal) eine eigene Hypothese auf.	Schüler/innen stellen eine Hypothese auf.
G3: Planung-Experiment	Schüler/innen arbeiten nach einer vorgegebenen Anleitung.	Schüler/innen wählen ihre Vorgehensweise aus verschiedenen vorgegebenen Möglichkeiten aus.	Schüler/innen planen mit Hilfestellung (medial, verbal) ein Vorgehen.	Schüler/innen planen ihr Vorgehen.
G4: Auswertung	Schüler/innen werten die Daten nach Anleitung aus.	tw. möglich: Schüler/innen wählen aus verschiedenen vorgegebenen Darstellungsmöglichkeiten (Tabelle, Balken-, Säulen-, Liniendiagramm) aus.	Schüler/innen werten mit Hilfestellung (medial, verbal) die ermittelten Daten aus.	Schüler/innen suchen sich selbstständig ein Verfahren zur Datenauswertung aus.
G5: Schlussfolgerung	Schüler/innen setzen sich mit vorgegebener Schlussfolgerung auseinander.	Schüler/innen wählen aus verschiedenen vorgegebenen Schlussfolgerungen aus.	Schüler/innen stellen mit Hilfestellung (medial, verbal) eine Schlussfolgerung auf.	Schüler/innen ziehen Schlussfolgerung.

Literaturverzeichnis

Alexakos, K. (2010). Teaching the practice of science, unteaching the" scientific method. Science Scope, 33 (9), 74–79. Zugriff am 02.09.2019.

Baur, A., Ehrenfeld, U., Emden, M., Hummel, E. & Krieg, A. (2017). Naturwissenschaften zum Leben erwecken: Chemie. Unterrichtsideen, Materialien und didaktische Grundlagen zum offenen Experimentieren (Bergedorfer Unterrichtsideen, 1. Auflage). Hamburg: Persen.

Horizon Research, I. (2005). Local Systemic Change Classroom Observation Protocol. Zugriff am 23.03.2018.

Lazonder, A. W. & Harmsen, R. (2016). Meta-Analysis of Inquiry-Based Learning. Effects of Guidance. Review of Educational Research, 86 (3), 681–718. https://doi.org/10.3102/0034654315627366

Metzger, S. & Sommer, K. (2010). "Kochrezept" oder experimentelle Methode? Eine Standortbestimmung von Schülerexperimenten unter dem Gesichtspunkt der Erkenntnisgewinnung. Der mathematische und naturwissenschaftliche Unterricht, 63 (1), 4–11. Zugriff am 17.09.2019.

11.2 Kategoriensystem ‚Phasen des Experimentierprozesses'

Pädagogische Hochschule HEIDELBERG University of Education

Kategoriensystem zur Analyse von naturwissenschaftlichem Unterricht mit der Absicht des Entdeckenden Experimentierens

In offenem Unterricht arbeiten die Schülerinnen und Schüler selbstständig oder in Gruppen und befinden sich, je nach ihrer Arbeitsweise, möglicherweise in unterschiedlichen Phasen des Experimentierprozesses. Damit ist eine Einteilung des gesamten Klassengeschehens in eine Kategorie des Kategoriensystems nicht mehr möglich. In diesen Unterrichtsphasen soll eine Schülergruppe repräsentativ bei der Kodierung fokussiert werden. In der Regel kann hierfür die am besten erkennbare/beobachtbare Gruppe herangezogen werden. Dies erfordert jedoch eine kurze Absprache mit den anderen Ratern.

Die Allgemeine Handlung (die große Mehrheit der Schülerinnen und Schüler) steht immer über der Gruppenhandlung. Beispiel: Wird bereits aufgeräumt, obwohl die fokussierte Gruppe noch experimentiert, ist „Aufräumen“ zu kodieren.

Es wird immer diejenige Kategorie gewählt, auf welcher der Fokus der Aktivität liegt. Beispiel: Beschafft ein Gruppenteilnehmer vor der Experimentierphase Experimentiermaterial, die anderen Gruppenteilnehmer warten ab, so liegt der Fokus der Gruppe auf der Materialbeschaffung. Räumt ein Schüler der Gruppe bereits Materialien auf, während die anderen Gruppenteilnehmer die Daten auswerten liegt der Fokus auf der Datenauswertung.

Ist eine Phase, etwa durch eine externe Unterrichtsstörung, unterbrochen, wird diese Phase weiter codiert, da sich der geplante Fokus des Unterrichtsgeschehens nicht verändert hat.

Tritt in der Mitte eines 15 Sekunden-Slots eine neue Kategorie ein und löst die bisherige Kategorie ab, ist folgendermaßen zu verfahren: Tritt die neue Kategorie bis zur Sekunde 00:07 nicht auf, erfolgt eine Kodierung der bisherigen Kategorie. Tritt die neue Kategorie vor Sekunde 00:07 auf, wird die neue Kategorie für diesen zehn Sekunden-Slot gewählt.

1 Unterricht mit der Absicht des Entdeckenden Experimentierens

Der Unterricht hat die Absicht im weiten Sinne Entdeckendes Experimentieren oder deren Teilaspekte vorzubereiten, durchzuführen oder zu besprechen.

1 Organisatorisches

Jedes Unterrichtsgeschehen, welches dazu dient, eine soziale, lokale oder temporäre Lerngelegenheit zu strukturieren, vorzubereiten und einzurichten oder aufzulösen.

2 Fach- und Methodenwissen

Jedes Unterrichtsgeschehen welches einen fachwissenschaftlichen oder fachmethodischen Beitrag zum Experimentieren leistet. Dies kann auch eine (mündliche) Wiederholung bereits gelernten Wissens (deklaratives (Fach)wissen, Prozedurales Wissen) sein.

3 Einführung in das Problem (Problemstellung)

Es wird in die Problemstellung für das Experiment eingeführt, das Experiment wird motiviert OHNE eine Fragestellung aufzuwerfen (nur Motivieren) oder Fach- oder Methodenwissen zu vermitteln. Es wird ein Problem, eine besondere Situation etc. dargestellt.

4 Experimentelle Methode

1 Fragestellung

Identifizieren einer oder mehrere Fragestellungen in Bezug auf ein beschriebenes Phänomen.

2 Hypothesenfindung

Identifizieren einer oder mehrerer Hypothesen (Vermutungen, zunächst unbewiesener Aussagen), meist in Bezug auf die Fragestellung(en).

3 Planen

Planen eines Experiments. In der Planungsphase kann das für das Experiment notwenige Material beschafft werden (falls nicht in einer von der Lehrperson explizit ausgewiesenen Materialbeschaffungsphase geschehen) und Skizzen erstellt werden.

Die Anleitung des Vorgehens während des Experiments ist eine Handlungsanweisung.

4 Durchführen und Beobachten

Zentrales Unterrichtsgeschehen ist die **händische** Durchführung eines Experiments oder mehrerer Experimente. Der Beginn der Durchführungsphase kann den Aufbau des Experiments einschließen. Nachträgliche Materialbeschaffungen (es wurde etwas vergessen etc.) werden nicht neu als Materialbeschaffung codiert, da dies nicht mehr der Fokus des Handelns der Schülerinnen und Schüler ist.

5 Auswerten

Mögliche Arten der Auswertung werden besprochen und gewonnene Ergebnisse werden aufgeschrieben und ausgewertet: Daten werden notiert (z.B. in einer Tabelle), gesammelt, zueinander in Beziehung gesetzt, in Schaubildern visualisiert und aufbereitet. Beobachtungen werden verschriftlicht und miteinander verglichen.

Werden während der Durchführung bereits Messwerte aufgenommen, ist dies keine Auswertung sondern zählt weiter zur Durchführung. Messfehler und -abweichungen werden diskutiert.

6 Schlussfolgern

Die aus der Auswertung gewonnenen Daten werden mit Bezug auf die Hypothese interpretiert. Eine Schlussfolgerung ohne Bezug auf die Hypothese ist ebenfalls möglich, in diesem Falle ist

eine Schlussfolgerung in Bezug auf den Alltag, die Umwelt oder andere Lebensbereiche denkbar. Eine Schlussfolgerung ohne den Bezug zu gewonnenen Daten ist auch möglich, sofern es sich um ein Gedankenexperiment o.Ä. handelt.

7 Präsentation und Reflexion von Erkenntnissen

Gewonnene Erkenntnisse werden zusammengefasst, gesammelt, präsentiert, abgefragt und ggf. reflektiert. Wie sind wir zu dieser Erkenntnis gekommen? Hierbei können neue Fragestellungen oder Hypothesen auftreten, dies ist mit *Fragestellung* oder *Hypothese* zu kennzeichnen (Notiz: So werden mehrere Zyklen von Experimenten identifiziert -> Hinweis auf anregenden Unterricht und epistemologische Reife)

2 Kein Unterricht im Sinne des entdeckenden Experimentierens

Der Unterricht hat nicht die Absicht im weiten Sinne Entdeckendes Experimentieren oder deren Teilaspekte vorzubereiten, durchzuführen oder zu besprechen.

Indikatoren/Beispiele: Lehrperson bespricht Themen des nächsten Gelingensnachweises, Organisatorisches zum Schulalltag, den kommenden Ausflug, stellt die Anwesenheit fest, sanktioniert Schülerinnen oder Schüler oder Begrüßt die Klasse.

1 Unterricht hat noch nicht begonnen

Die Lehrperson hat den Unterricht noch nicht für die gesamte Klasse formal begonnen (z.B. Schülerinnen und Schüler betreten das Klassenzimmer)

2 Unterricht ist beendet

Der Unterricht wurde durch die Lehrperson beendet.

3 Organisatorisches

Beispiele: Lehrperson bespricht Themen des nächsten Gelingensnachweises, Organisatorisches zum Schulalltag, den kommenden Ausflug, stellt die Anwesenheit fest, sanktioniert Schülerinnen oder Schüler, weiteres Vorgehen in den folgenden Stunden

3 Restkategorie: Nicht zuweisbar

Unterrichtsgeschehen, welches nicht in eine der beschriebenen Kategorien zugeordnet werden kann.

11.3 Validierungsbogen Kategoriensystem ‚Unterrichtshandeln von Lehrkräften beim Experimentieren'

Pädagogische Hochschule HEIDELBERG University of Education

Validierung des Kategoriensystems

Frage: In welchem Umfang ist die jeweilige Kategorie dazu geeignet, einen Beitrag zur Beschreibung von Unterrichtshandeln der Lehrperson zum Experimentieren (Experimentieren im weiten Sinne) zu leisten?

0 = diese Kategorie kann KEINEN Beitrag leisten
1 = diese Kategorie kann einen geringen Beitrag leisten
2 = diese Kategorie kann einen Beitrag leisten
3 = diese Kategorie kann einen starken Beitrag leisten

Kategorie	Wertung (0 - 3)	ggf. Kommentar
S1		
S1A		
S2		
S2A		
S3		
S3A		
S3B		
S4		
S5		
S6		
S7		
S8		
S9		
S10		
S11		
S12		
S13		
S14		
S15A		
S15B		
R1		
R2		
R3		
R4		
R5		
R6		
R7		
R8		
R9		
R10		
R11		
R12		
K1		

K2		
K3		
K4		
K5		
K6		
K7		
K8		
K9		
K10		
K11		
K12		

11.4 Handlungsschritte zur Durchführung von Unterrichtsvideografien

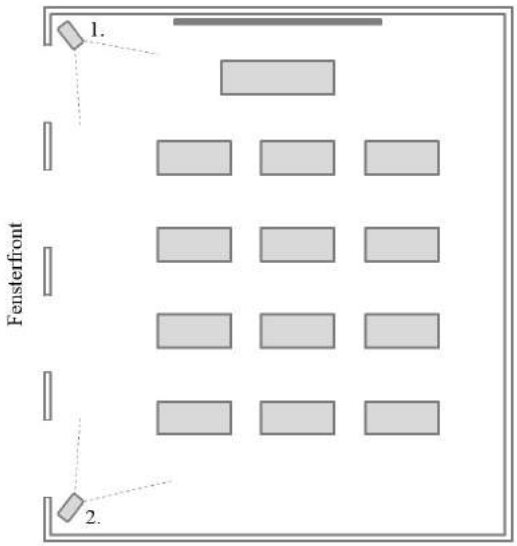
Pädagogische Hochschule
HEIDELBERG
University of Education

Handlungsschritte zur Durchführung von Unterrichtsvideographien

1. Vor der Abfahrt prüfe die Ausstattung anhand der Checkliste

    - Stative?
    - Kameras?
    - SD Karten (leer) in den Kameras?
    - Ladekabel inkl. Verlängerungskabel
    - Akkus geladen
    - Tickets gekauft

2. Achte auf eine rechtzeitige Anreise
3. Mache dich vor der Videografie mit dem Raum vertraut. Wo liegen Steckdosen, wo Fenster und Türen? Wo sind die Laufwege
4. Baue die Kameras auf. Prüfe Gefahren durch Kabel etc.



Kamerapositionen auf 1. und 2. Die Kameras werden horizontal aufgestellt.

5. Führe einen Testlauf durch.
6. Vor der Videogaphie die Einverständniserklärungen prüfen und der Klasse den Sinn und Zweck der Aufnahme kurz erklären.
7. Materialien wie Arbeitsblätter, Tafelanschriebe etc. bitte fotografisch dokumentieren.

11.5 Transkriptionsleitfaden

Transkriptionsregeln

Die Transkriptionsregeln, die bei der Verschriftlichung der Videos zum Einsatz kommen, orientieren sich stets an der Forschungsfrage. In unserem Fall stehen keine sprachwissenschaftlichen Aspekte im Vordergrund, sondern der Inhalt der Videos soll, in Bezug auf das Unterrichtshandeln der Lehrkraft, wiedergegeben werden.

Die Transkription soll einfach und ökonomisch sein.

1. Treten eine Handlung und eine Aussage gleichzeitig auf, wird in der Regel zuerst die Aussage transkribiert anschließend die Handlung beschrieben.
2. Auf jede Aussage oder Handlungsbeschreibung folgt ein Absatz.
3. Handlungen werden knapp und präzise beschrieben. Die Handlungen sollen das Unterrichtshandeln der Lehrperson im Fokus haben. Schülerhandeln wird zum allgemeinen Verständnis wo sinnvoll beschrieben.
4. In Gruppenarbeitsphasen wird das Verhalten der Lehrperson beschrieben.
5. Die Entwicklung des Tafelbilds wird in der Beschreibung dargestellt.
6. Sprachpausen in Aussagen werden nicht transkribiert.
7. Para- und außersprachliche Handlungen (lachen, husten) und Verzögerungssignale (äh, ahm) werden nicht mit transkribiert.
8. Werden Aussagen mit Nachdruck formuliert („Schönen guten Morgen!“, Denkt an das Aufräumen!“) wird die Aussage mit einem Ausrufezeichen (!) abgeschlossen.
9. Einheiten (Meter, Milliampere) in Aussagen werden ausgeschrieben.
10. Es werden keine Abkürzungen bei Aussagen verwendet (bzw., z.B. ...).
11. Werden Aussagen nicht abgeschlossen wird ein … angefügt, etwa wenn in das Wort gefallen wird. Werden Aussagen von einer anderen Person zu Ende geführt, wird ein … angeführt.
12. Dialekt wird in Hochdeutsch überführt: „Des weisch doch“ -> „Das weist du doch“ „kriege“ -> „bekomme“, „ranmachen“ -> „anschließen“, „gucken“ -> „schauen“, „tust du da rüber“ -> „räumst du da rüber“
13. Kann der Teil einer unverständlichen Aussage nicht rekonstruiert werden wird dies mit (unv.) gekennzeichnet.
14. Einfache Anführungszeichen werden zur Kennzeichnung von Zitaten verwendet. Wird ein Text vorgelesen, schreibt man diesen in ‚einfachen Anführungszeichen‘.
15. Alle Angaben, die den Rückschluss auf eine Person (Lehrperson, SchülerIn) erlauben, werden anonymisiert: aus „Hr. Winter“ -> „Lehrperson“, aus „Luise“ -> „S1“ etc.
16. Hinweise zum Verständnis einzelner Aussagen werden in [] gesetzt: „Wenn Sie das Blaue [Isolation] anfassen, dann schon, wenn Sie das oben [blankes Metall] anfassen, dann nicht.“
17. Abkürzungen:
    AB = Arbeitsblatt

SuS = Schülerinnen und Schüler
L = Lehrperson: S4: Herr L, kann ich mir die Batterie nehmen?

18. Es ist sinnvoll die SuS vor dem Transkribieren im Geiste durchzunummerieren (etwa vorne links S1 bis hinten rechts S25). Sprechen S, welche nicht im Bild sind, muss eine beliebige Zuweisung erfolgen.

Literatur:

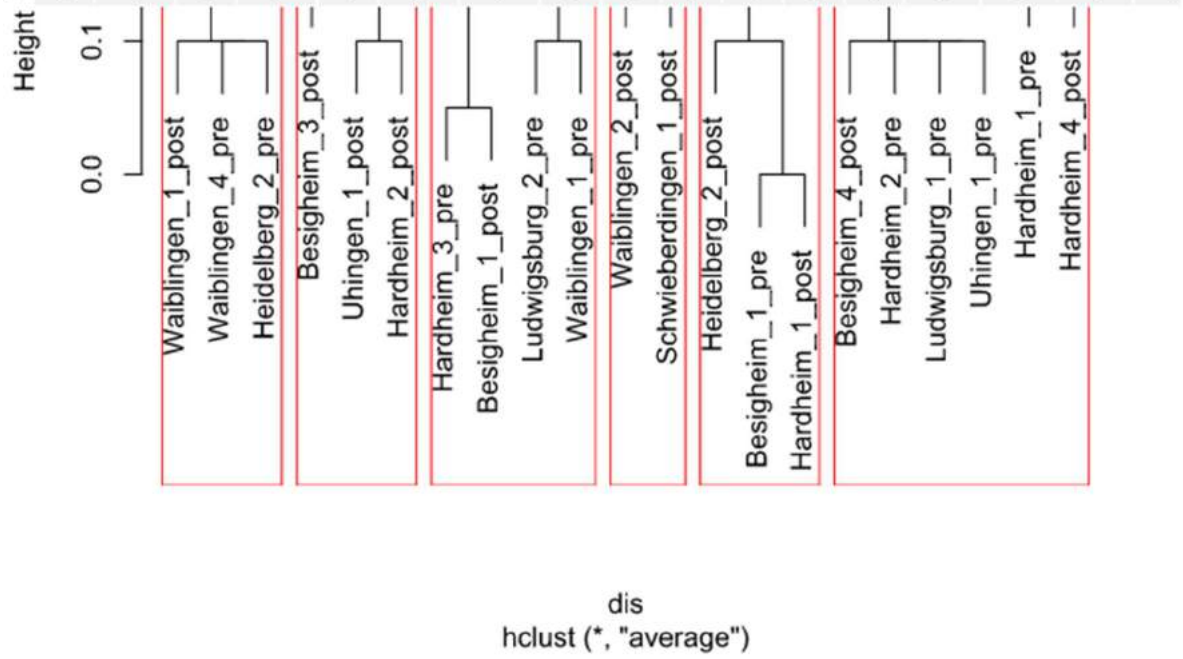
Kuckartz, U./ Dresing, Th./ Rädiker, S./ Stefer, C.(2007): Qualitative Evaluation – Der Einstieg in die Praxis; Wiesbaden: VS Verlag für Sozialwissenschaften

11.6 Clusteranalysen (Rohform)

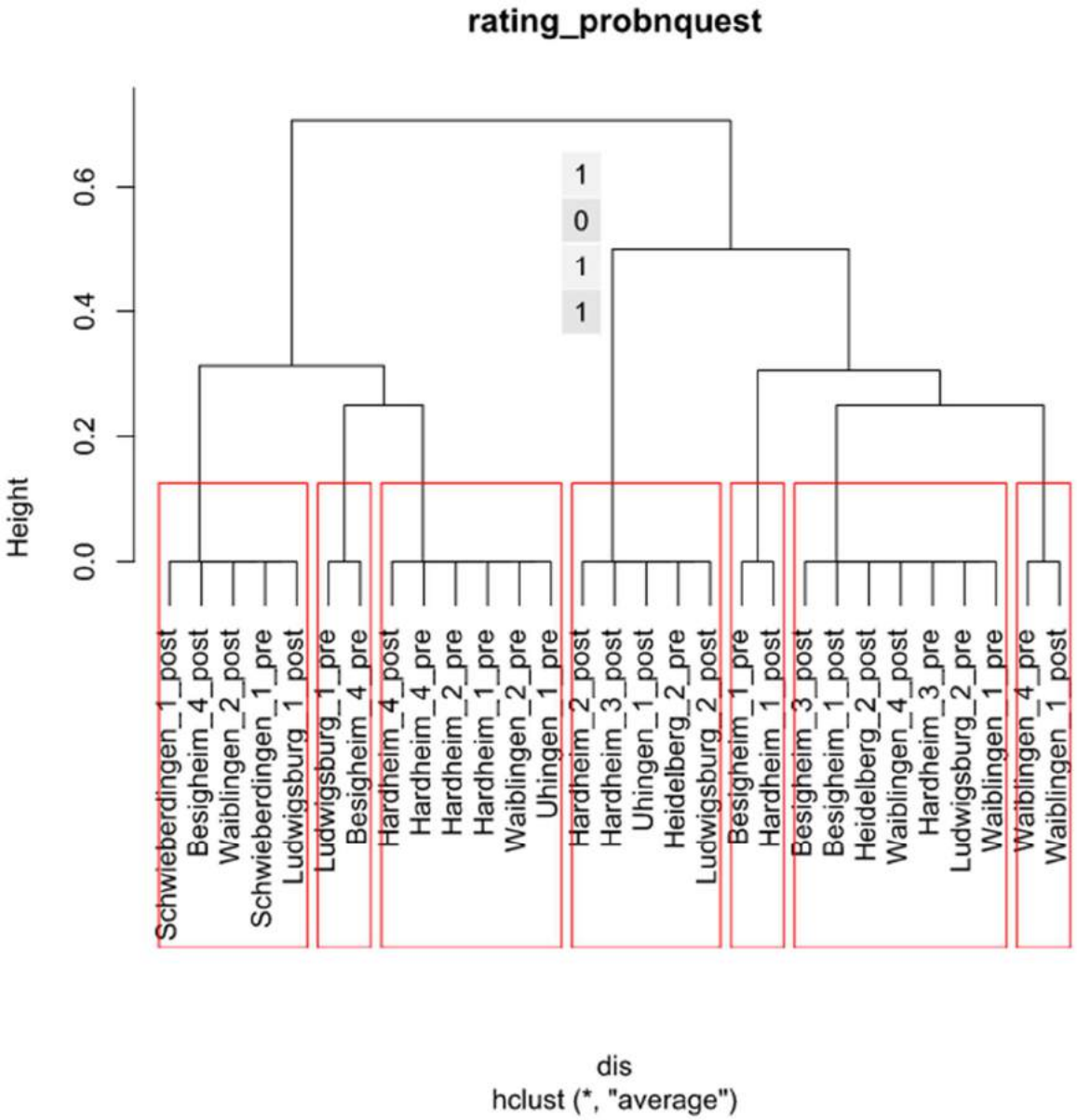
Im Folgenden sind alle Clusteranalysen in ‚Rohform' dargestellt. Können einzelne Videografien keinem Cluster zugeordnet werden, so sind diese als Tabellenzeile aufgeführt.

11.6.1 Dimension ‚Prozessstruktur'

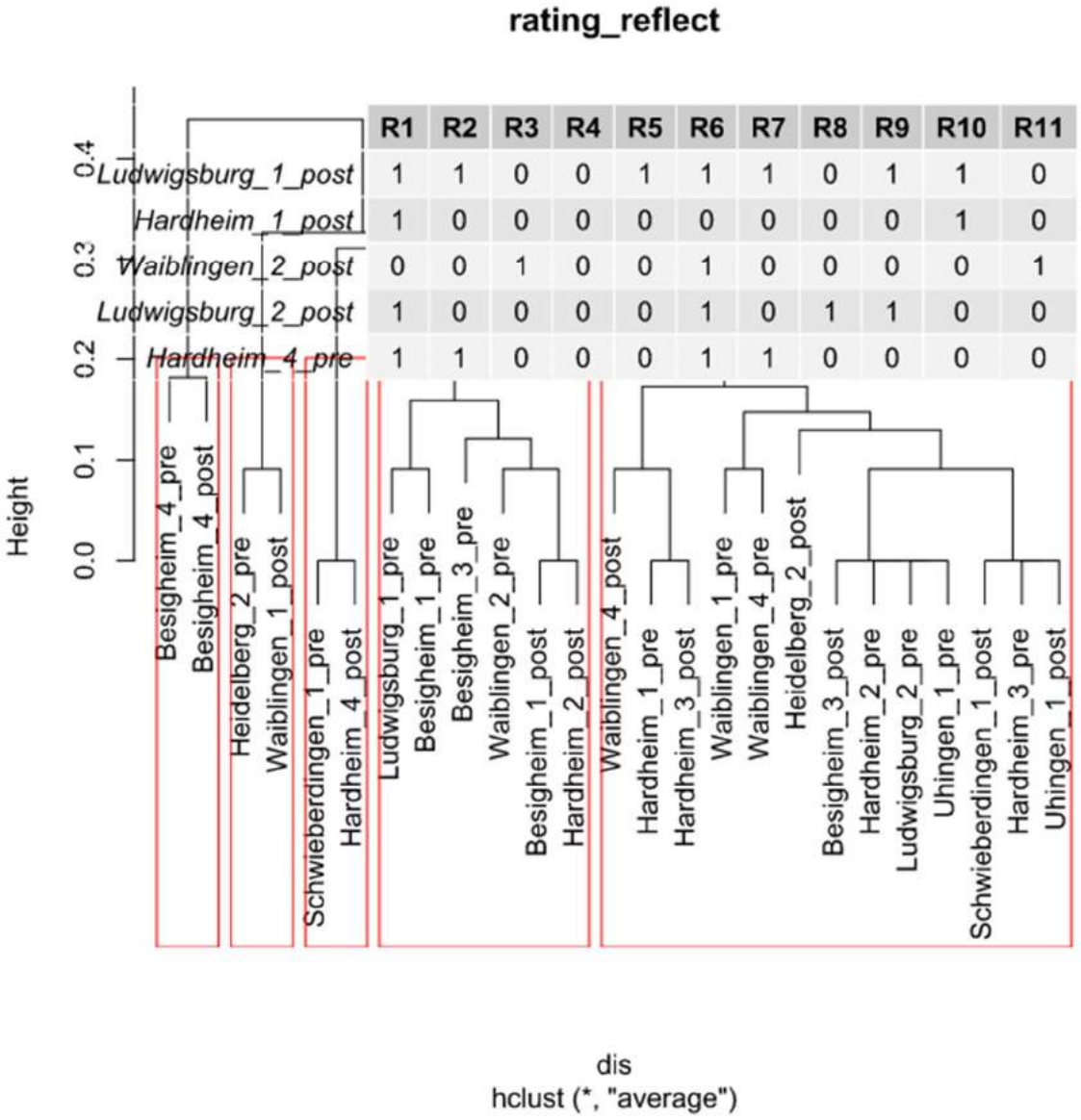
	S1	S1A	S2	S2A	S3	S3A	S3B	S4	S5	S6	S7	S8	S9	S10	S11	S12	S1
st	0	1	0	0	1	0	0	1	0	1	1	0	1	1	0	1	1
st	1	0	0	0	1	1	1	1	1	1	0	1	0	0	0	1	0
st	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	0
st	0	1	0	0	1	0	1	0	0	0	1	0	0	0	0	1	1
e	1	0	1	1	1	0	0	1	0	0	0	1	0	1	0	1	1
e	0	1	1	1	1	0	1	1	1	1	0	0	0	1	0	1	1
e	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1
e	0	1	1	0	1	1	0	1	0	0	0	0	0	1	0	1	1
e	0	1	1	1	1	1	0	0	0	1	1	0	0	0	0	1	1



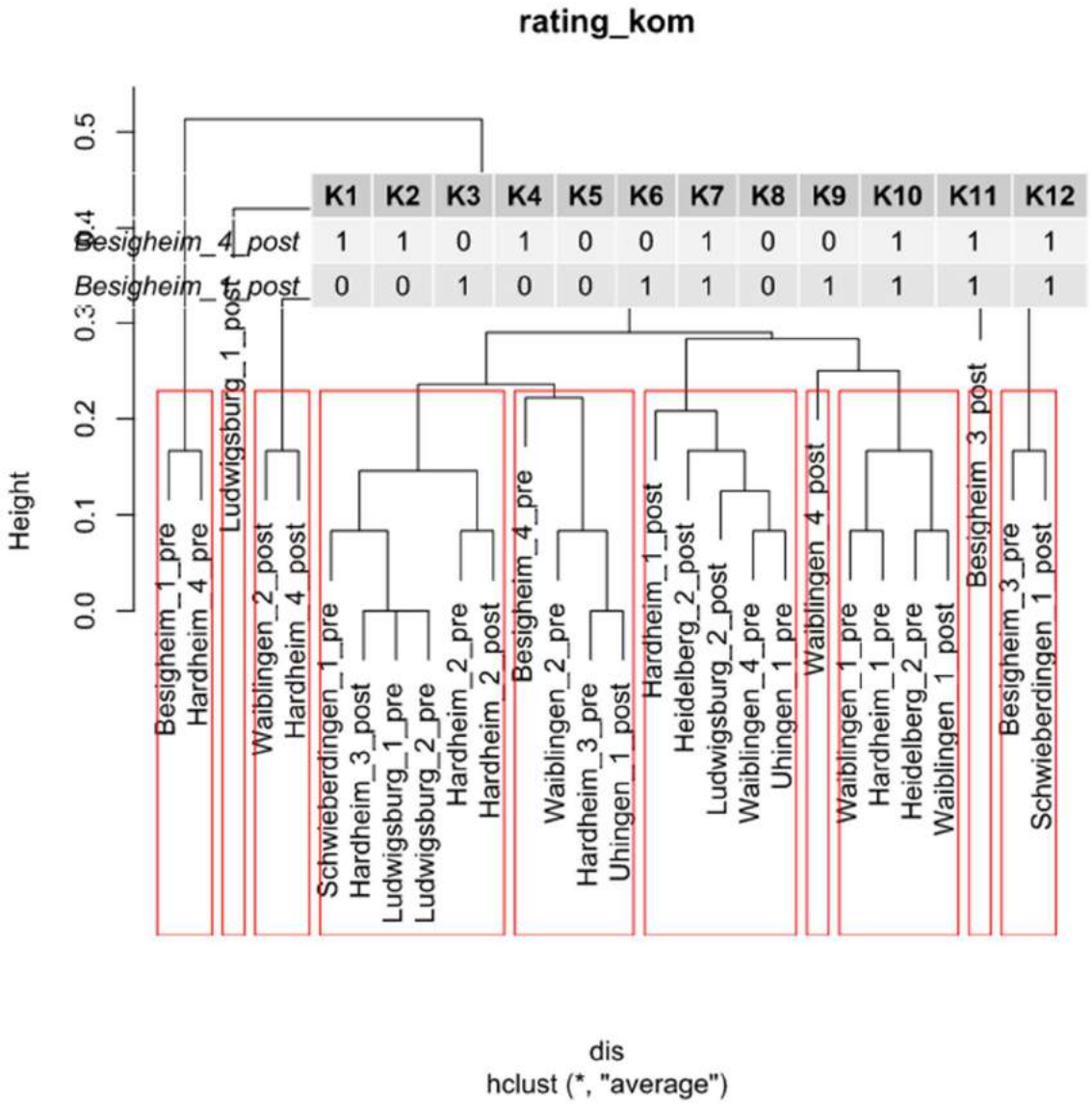
11.6.2 Oberkategorie ‚Problem und Fragestellung'



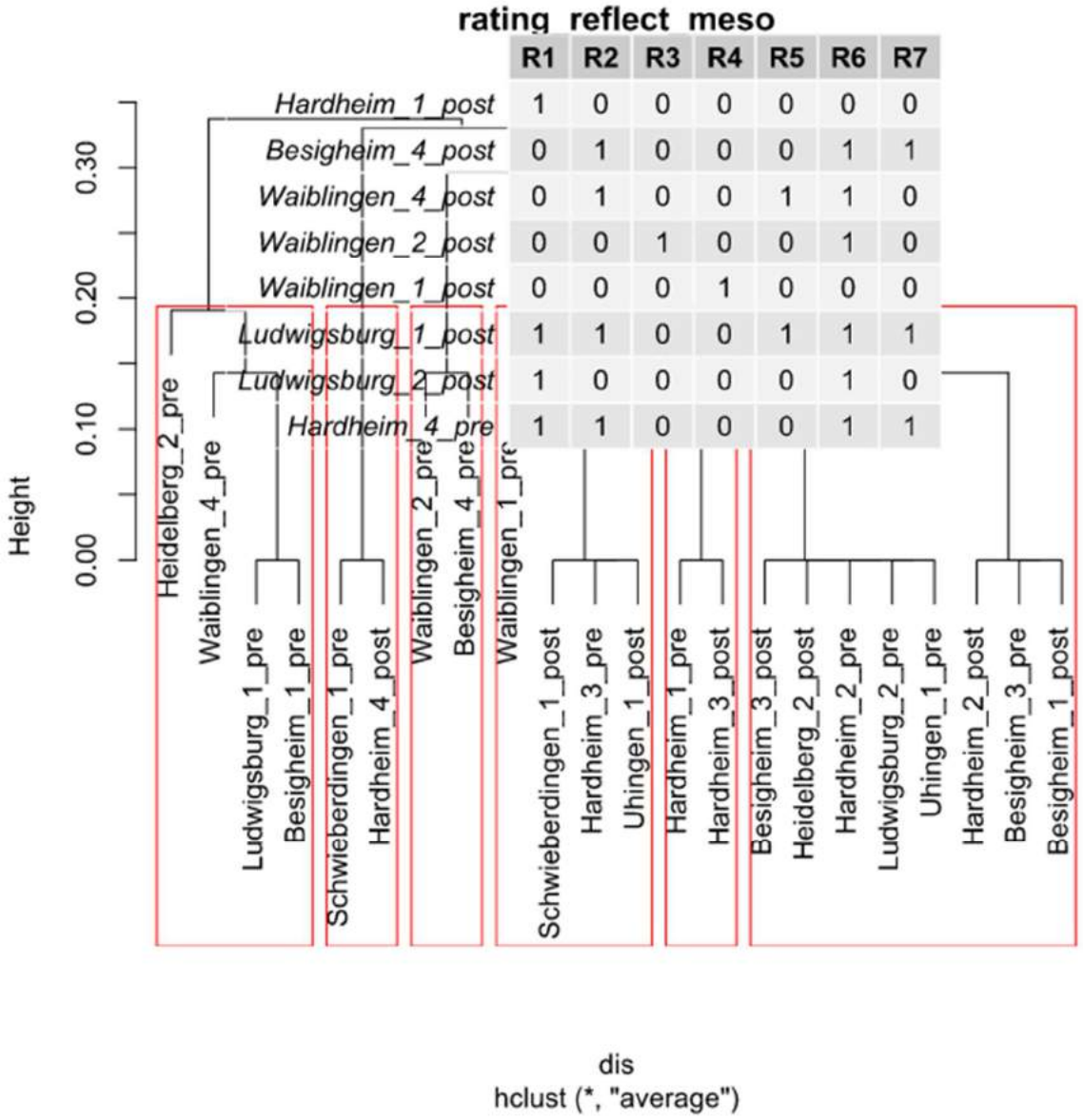
11.6.3 Dimension , Prozessreflexion (meso und meta)‘



11.6.4 Dimension ‚Angemessenheit und Unterstützung'

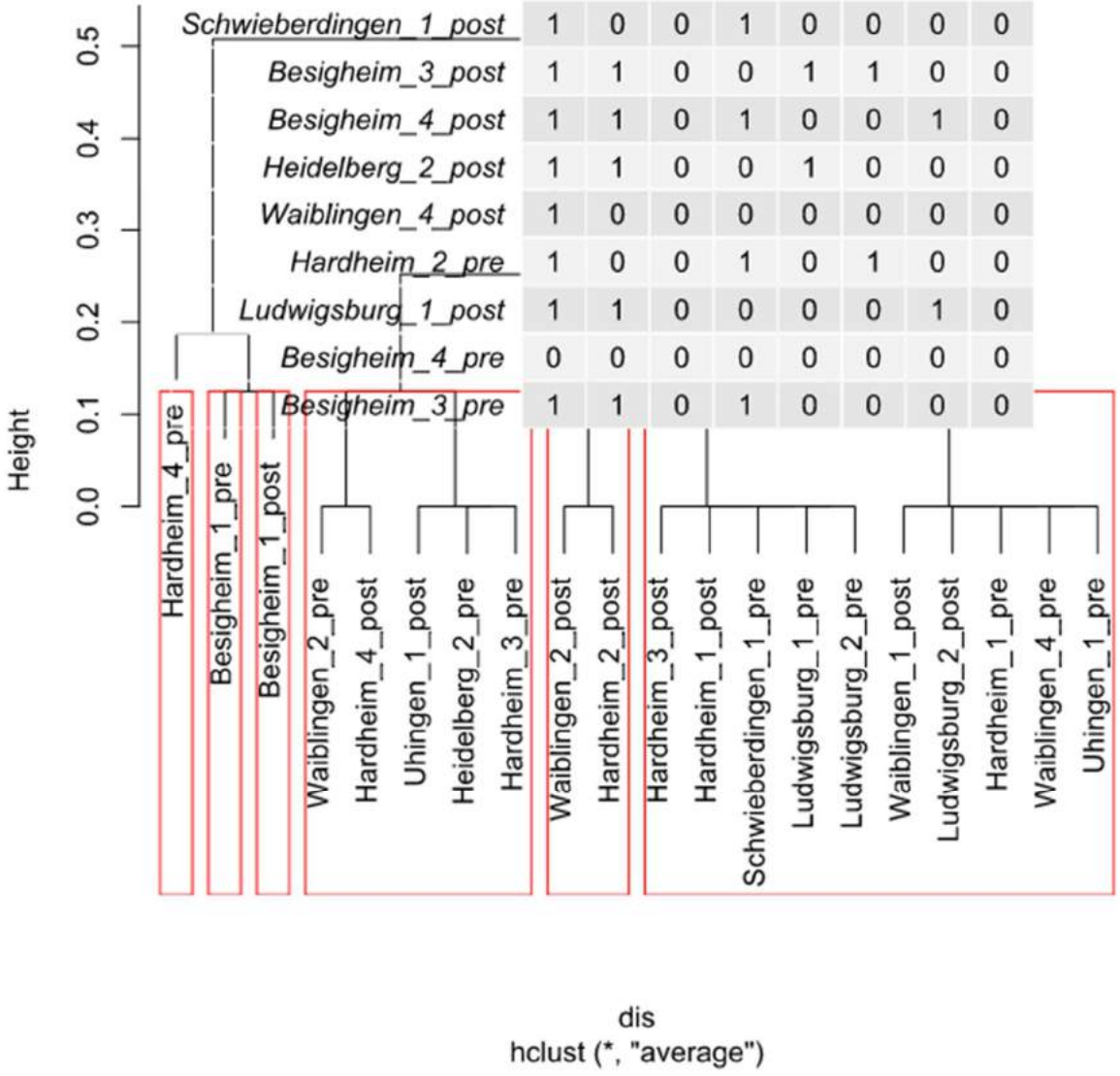


11.6.5 Oberkategorie ‚Mesoreflexion'



11.6.6 Oberkategorie ‚L als Trainer, Ansprechpartner und Experte'

ra	K1	K2	K3	K4	K5	K6	K7	K8
Waiblingen_1_pre	0	1	0	0	0	0	0	0
Schwieberdingen_1_post	1	0	0	1	0	0	0	0
Besigheim_3_post	1	1	0	0	1	1	0	0
Besigheim_4_post	1	1	0	1	0	0	1	0
Heidelberg_2_post	1	1	0	0	1	0	0	0
Waiblingen_4_post	1	0	0	0	0	0	0	0
Hardheim_2_pre	1	0	0	1	0	1	0	0
Ludwigsburg_1_post	1	1	0	0	0	0	1	0
Besigheim_4_pre	0	0	0	0	0	0	0	0
Besigheim_3_pre	1	1	0	1	0	0	0	0



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

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

1 Helmut Fischler, Jochen Peuckert (Hrsg.): Concept Mapping in fachdidaktischen Forschungsprojekten der Physik und Chemie
ISBN 978-3-89722-256-4 40.50 EUR

2 Anja Schoster: Bedeutungsentwicklungsprozesse beim Lösen algorithmischer Physikaufgaben. *Eine Fallstudie zu Lernprozessen von Schülern im Physiknachhilfeunterricht während der Bearbeitung algorithmischer Physikaufgaben*
ISBN 978-3-89722-045-4 40.50 EUR

3 Claudia von Aufschnaiter: Bedeutungsentwicklungen, Interaktionen und situatives Erleben beim Bearbeiten physikalischer Aufgaben
ISBN 978-3-89722-143-7 40.50 EUR

4 Susanne Haeberlen: Lernprozesse im Unterricht mit Wasserstromkreisen. *Eine Fallstudie in der Sekundarstufe I*
ISBN 978-3-89722-172-7 40.50 EUR

5 Kerstin Haller: Über den Zusammenhang von Handlungen und Zielen. *Eine empirische Untersuchung zu Lernprozessen im physikalischen Praktikum*
ISBN 978-3-89722-242-7 40.50 EUR

6 Michaela Horstendahl: Motivationale Orientierungen im Physikunterricht
ISBN 978-3-89722-227-4 50.00 EUR

7 Stefan Deylitz: Lernergebnisse in der Quanten-Atomphysik. *Evaluation des Bremer Unterrichtskonzepts*
ISBN 978-3-89722-291-5 40.50 EUR

8 Lorenz Hucke: Handlungsregulation und Wissenserwerb in traditionellen und computergestützten Experimenten des physikalischen Praktikums
ISBN 978-3-89722-316-5 50.00 EUR

9 Heike Theyßen: Ein Physikpraktikum für Studierende der Medizin. *Darstellung der Entwicklung und Evaluation eines adressatenspezifischen Praktikums nach dem Modell der Didaktischen Rekonstruktion*
ISBN 978-3-89722-334-9 40.50 EUR

10 Annette Schick: Der Einfluß von Interesse und anderen selbstbezogenen Kognitionen auf Handlungen im Physikunterricht. *Fallstudien zu Interessenhandlungen im Physikunterricht*
ISBN 978-3-89722-380-6 40.50 EUR

11 Roland Berger: Moderne bildgebende Verfahren der medizinischen Diagnostik. *Ein Weg zu interessanterem Physikunterricht*
ISBN 978-3-89722-445-2 40.50 EUR

12 Johannes Werner: Vom Licht zum Atom. *Ein Unterrichtskonzept zur Quantenphysik unter Nutzung des Zeigermodells*
ISBN 978-3-89722-471-1 40.50 EUR

13 Florian Sander: Verbindung von Theorie und Experiment im physikalischen Praktikum. *Eine empirische Untersuchung zum handlungsbezogenen Vorverständnis und dem Einsatz grafikorientierter Modellbildung im Praktikum*
ISBN 978-3-89722-482-7 40.50 EUR

14 Jörn Gerdes: Der Begriff der physikalischen Kompetenz. *Zur Validierung eines Konstruktes*
ISBN 978-3-89722-510-7 40.50 EUR

15 Malte Meyer-Arndt: Interaktionen im Physikpraktikum zwischen Studierenden und Betreuern. *Feldstudie zu Bedeutungsentwicklungsprozessen im physikalischen Praktikum*
ISBN 978-3-89722-541-1 40.50 EUR

16 Dietmar Höttecke: Die Natur der Naturwissenschaften historisch verstehen. *Fachdidaktische und wissenschaftshistorische Untersuchungen*
ISBN 978-3-89722-607-4 40.50 EUR

17 Gil Gabriel Mavanga: Entwicklung und Evaluation eines experimentell- und phänomenorientierten Optikcurriculums. *Untersuchung zu Schülervorstellungen in der Sekundarstufe I in Mosambik und Deutschland*
ISBN 978-3-89722-721-7 40.50 EUR

18 Meike Ute Zastrow: Interaktive Experimentieranleitungen. *Entwicklung und Evaluation eines Konzeptes zur Vorbereitung auf das Experimentieren mit Messgeräten im Physikalischen Praktikum*
ISBN 978-3-89722-802-3 40.50 EUR

19 Gunnar Friege: Wissen und Problemlösen. *Eine empirische Untersuchung des wissenszentrierten Problemlösens im Gebiet der Elektrizitätslehre auf der Grundlage des Experten-Novizen-Vergleichs*
ISBN 978-3-89722-809-2 40.50 EUR

20 Erich Starauschek: Physikunterricht nach dem Karlsruher Physikkurs. *Ergebnisse einer Evaluationsstudie*
ISBN 978-3-89722-823-8 40.50 EUR

21 Roland Paatz: Charakteristika analogiebasierten Denkens. *Vergleich von Lernprozessen in Basis- und Zielbereich*
ISBN 978-3-89722-944-0 40.50 EUR

22 Silke Mikelskis-Seifert: Die Entwicklung von Metakonzepten zur Teilchenvorstellung bei Schülern. *Untersuchung eines Unterrichts über Modelle mithilfe eines Systems multipler Repräsentationsebenen*
ISBN 978-3-8325-0013-9 40.50 EUR

23 Brunhild Landwehr: Distanzen von Lehrkräften und Studierenden des Sachunterrichts zur Physik. *Eine qualitativ-empirische Studie zu den Ursachen*
ISBN 978-3-8325-0044-3 40.50 EUR

24 Lydia Murmann: Physiklernen zu Licht, Schatten und Sehen. *Eine phänomenografische Untersuchung in der Primarstufe*
ISBN 978-3-8325-0060-3 40.50 EUR

25 Thorsten Bell: Strukturprinzipien der Selbstregulation. *Komplexe Systeme, Elementarisierungen und Lernprozessstudien für den Unterricht der Sekundarstufe II*
ISBN 978-3-8325-0134-1 40.50 EUR

26 Rainer Müller: Quantenphysik in der Schule
ISBN 978-3-8325-0186-0 40.50 EUR

27 Jutta Roth: Bedeutungsentwicklungsprozesse von Physikerinnen und Physikern in den Dimensionen Komplexität, Zeit und Inhalt
ISBN 978-3-8325-0183-9 40.50 EUR

28 Andreas Saniter: Spezifika der Verhaltensmuster fortgeschrittener Studierender der Physik
ISBN 978-3-8325-0292-8 40.50 EUR

29 Thomas Weber: Kumulatives Lernen im Physikunterricht. *Eine vergleichende Untersuchung in Unterrichtsgängen zur geometrischen Optik*
ISBN 978-3-8325-0316-1 40.50 EUR

30 Markus Rehm: Über die Chancen und Grenzen moralischer Erziehung im naturwissenschaftlichen Unterricht
ISBN 978-3-8325-0368-0 40.50 EUR

31 Marion Budde: Lernwirkungen in der Quanten-Atom-Physik. *Fallstudien über Resonanzen zwischen Lernangeboten und SchülerInnen-Vorstellungen*
ISBN 978-3-8325-0483-0 40.50 EUR

32 Thomas Reyer: Oberflächenmerkmale und Tiefenstrukturen im Unterricht. *Exemplarische Analysen im Physikunterricht der gymnasialen Sekundarstufe*
ISBN 978-3-8325-0488-5 40.50 EUR

33 Christoph Thomas Müller: Subjektive Theorien und handlungsleitende Kognitionen von Lehrern als Determinanten schulischer Lehr-Lern-Prozesse im Physikunterricht
ISBN 978-3-8325-0543-1 40.50 EUR

34 Gabriela Jonas-Ahrend: Physiklehrervorstellungen zum Experiment im Physikunterricht
ISBN 978-3-8325-0576-9 40.50 EUR

35 Dimitrios Stavrou: Das Zusammenspiel von Zufall und Gesetzmäßigkeiten in der nichtlinearen Dynamik. *Didaktische Analyse und Lernprozesse*
ISBN 978-3-8325-0609-4 40.50 EUR

36 Katrin Engeln: Schülerlabors: authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken
ISBN 978-3-8325-0689-6 40.50 EUR

37 Susann Hartmann: Erklärungsvielfalt
ISBN 978-3-8325-0730-5 40.50 EUR

38 Knut Neumann: Didaktische Rekonstruktion eines physikalischen Praktikums für Physiker
ISBN 978-3-8325-0762-6 40.50 EUR

39 Michael Späth: Kontextbedingungen für Physikunterricht an der Hauptschule. *Möglichkeiten und Ansatzpunkte für einen fachübergreifenden, handlungsorientierten und berufsorientierten Unterricht*
ISBN 978-3-8325-0827-2 40.50 EUR

40 Jörg Hirsch: Interesse, Handlungen und situatives Erleben von Schülerinnen und Schülern beim Bearbeiten physikalischer Aufgaben
ISBN 978-3-8325-0875-3 40.50 EUR

41 Monika Hüther: Evaluation einer hypermedialen Lernumgebung zum Thema Gasgesetze. *Eine Studie im Rahmen des Physikpraktikums für Studierende der Medizin*
ISBN 978-3-8325-0911-8 40.50 EUR

42 Maike Tesch: Das Experiment im Physikunterricht. *Didaktische Konzepte und Ergebnisse einer Videostudie*
ISBN 978-3-8325-0975-0 40.50 EUR

43 Nina Nicolai: Skriptgeleitete Eltern-Kind-Interaktion bei Chemiehausaufgaben. *Eine Evaluationsstudie im Themenbereich Säure-Base*
ISBN 978-3-8325-1013-8 40.50 EUR

44 Antje Leisner: Entwicklung von Modellkompetenz im Physikunterricht
ISBN 978-3-8325-1020-6 40.50 EUR

45 Stefan Rumann: Evaluation einer Interventionsstudie zur Säure-Base-Thematik
ISBN 978-3-8325-1027-5 40.50 EUR

46 Thomas Wilhelm: Konzeption und Evaluation eines Kinematik/Dynamik-Lehrgangs zur Veränderung von Schülervorstellungen mit Hilfe dynamisch ikonischer Repräsentationen und graphischer Modellbildung – mit CD-ROM
ISBN 978-3-8325-1046-6 45.50 EUR

47 Andrea Maier-Richter: Computerunterstütztes Lernen mit Lösungsbeispielen in der Chemie. *Eine Evaluationsstudie im Themenbereich Löslichkeit*
ISBN 978-3-8325-1046-6 40.50 EUR

48 Jochen Peuckert: Stabilität und Ausprägung kognitiver Strukturen zum Atombegriff
ISBN 978-3-8325-1104-3 40.50 EUR

49 Maik Walpuski: Optimierung von experimenteller Kleingruppenarbeit durch Strukturierungshilfen und Feedback
ISBN 978-3-8325-1184-5 40.50 EUR

50 Helmut Fischler, Christiane S. Reiners (Hrsg.): Die Teilchenstruktur der Materie im Physik- und Chemieunterricht
ISBN 978-3-8325-1225-5 34.90 EUR

51 Claudia Eysel: Interdisziplinäres Lehren und Lernen in der Lehrerbildung. *Eine empirische Studie zum Kompetenzerwerb in einer komplexen Lernumgebung*
ISBN 978-3-8325-1238-5 40.50 EUR

52 Johannes Günther: Lehrerfortbildung über die Natur der Naturwissenschaften. *Studien über das Wissenschaftsverständnis von Grundschullehrkräften*
ISBN 978-3-8325-1287-3 40.50 EUR

53 Christoph Neugebauer: Lernen mit Simulationen und der Einfluss auf das Problemlösen in der Physik
ISBN 978-3-8325-1300-9 40.50 EUR

54 Andreas Schnirch: Gendergerechte Interessen- und Motivationsförderung im Kontext naturwissenschaftlicher Grundbildung. *Konzeption, Entwicklung und Evaluation einer multimedial unterstützten Lernumgebung*
ISBN 978-3-8325-1334-4 40.50 EUR

55 Hilde Köster: Freies Explorieren und Experimentieren. *Eine Untersuchung zur selbstbestimmten Gewinnung von Erfahrungen mit physikalischen Phänomenen im Sachunterricht*
ISBN 978-3-8325-1348-1 40.50 EUR

56 Eva Heran-Dörr: Entwicklung und Evaluation einer Lehrerfortbildung zur Förderung der physikdidaktischen Kompetenz von Sachunterrichtslehrkräften
ISBN 978-3-8325-1377-1 40.50 EUR

57 Agnes Szabone Varnai: Unterstützung des Problemlösens in Physik durch den Einsatz von Simulationen und die Vorgabe eines strukturierten Kooperationsformats
ISBN 978-3-8325-1403-7 40.50 EUR

58 Johannes Rethfeld: Aufgabenbasierte Lernprozesse in selbstorganisationsoffenem Unterricht der Sekundarstufe I zum Themengebiet ELEKTROSTATIK. *Eine Feldstudie in vier 10. Klassen zu einer kartenbasierten Lernumgebung mit Aufgaben aus der Elektrostatik*
ISBN 978-3-8325-1416-7 40.50 EUR

59 Christian Henke: Experimentell-naturwissenschaftliche Arbeitsweisen in der Oberstufe. *Untersuchung am Beispiel des HIGHSEA-Projekts in Bremerhaven*
ISBN 978-3-8325-1515-7 40.50 EUR

60 Lutz Kasper: Diskursiv-narrative Elemente für den Physikunterricht. *Entwicklung und Evaluation einer multimedialen Lernumgebung zum Erdmagnetismus*
ISBN 978-3-8325-1537-9 40.50 EUR

61 Thorid Rabe: Textgestaltung und Aufforderung zu Selbsterklärungen beim Physiklernen mit Multimedia
ISBN 978-3-8325-1539-3 40.50 EUR

62 Ina Glemnitz: Vertikale Vernetzung im Chemieunterricht. *Ein Vergleich von traditionellem Unterricht mit Unterricht nach Chemie im Kontext*
ISBN 978-3-8325-1628-4 40.50 EUR

63 Erik Einhaus: Schülerkompetenzen im Bereich Wärmelehre. *Entwicklung eines Testinstruments zur Überprüfung und Weiterentwicklung eines normativen Modells fachbezogener Kompetenzen*
ISBN 978-3-8325-1630-7 40.50 EUR

64 Jasmin Neuroth: Concept Mapping als Lernstrategie. *Eine Interventionsstudie zum Chemielernen aus Texten*
ISBN 978-3-8325-1659-8 40.50 EUR

65 Hans Gerd Hegeler-Burkhart: Zur Kommunikation von Hauptschülerinnen und Hauptschülern in einem handlungsorientierten und fächerübergreifenden Unterricht mit physikalischen und technischen Inhalten
ISBN 978-3-8325-1667-3 40.50 EUR

66 Karsten Rincke: Sprachentwicklung und Fachlernen im Mechanikunterricht. *Sprache und Kommunikation bei der Einführung in den Kraftbegriff*
ISBN 978-3-8325-1699-4 40.50 EUR

67 Nina Strehle: Das Ion im Chemieunterricht. *Alternative Schülervorstellungen und curriculare Konsequenzen*
ISBN 978-3-8325-1710-6 40.50 EUR

68 Martin Hopf: Problemorientierte Schülerexperimente
ISBN 978-3-8325-1711-3 40.50 EUR

69 Anne Beerenwinkel: Fostering conceptual change in chemistry classes using expository texts
ISBN 978-3-8325-1721-2 40.50 EUR

70 Roland Berger: Das Gruppenpuzzle im Physikunterricht der Sekundarstufe II. *Eine empirische Untersuchung auf der Grundlage der Selbstbestimmungstheorie der Motivation*
ISBN 978-3-8325-1732-8 40.50 EUR

71 Giuseppe Colicchia: Physikunterricht im Kontext von Medizin und Biologie. *Entwicklung und Erprobung von Unterrichtseinheiten*
ISBN 978-3-8325-1746-5 40.50 EUR

72 Sandra Winheller: Geschlechtsspezifische Auswirkungen der Lehrer-Schüler-Interaktion im Chemieanfangsunterricht
ISBN 978-3-8325-1757-1 40.50 EUR

73 Isabel Wahser: Training von naturwissenschaftlichen Arbeitsweisen zur Unterstützung experimenteller Kleingruppenarbeit im Fach Chemie
ISBN 978-3-8325-1815-8 40.50 EUR

74 Claus Brell: Lernmedien und Lernerfolg - reale und virtuelle Materialien im Physikunterricht. *Empirische Untersuchungen in achten Klassen an Gymnasien (Laborstudie) zum Computereinsatz mit Simulation und IBE*
ISBN 978-3-8325-1829-5 40.50 EUR

75 Rainer Wackermann: Überprüfung der Wirksamkeit eines Basismodell-Trainings für Physiklehrer
ISBN 978-3-8325-1882-0 40.50 EUR

76 Oliver Tepner: Effektivität von Aufgaben im Chemieunterricht der Sekundarstufe I
ISBN 978-3-8325-1919-3 40.50 EUR

77 Claudia Geyer: Museums- und Science-Center-Besuche im naturwissenschaftlichen Unterricht aus einer motivationalen Perspektive. *Die Sicht von Lehrkräften und Schülerinnen und Schülern*
ISBN 978-3-8325-1922-3 40.50 EUR

78 Tobias Leonhard: Professionalisierung in der Lehrerbildung. *Eine explorative Studie zur Entwicklung professioneller Kompetenzen in der Lehrererstausbildung*
ISBN 978-3-8325-1924-7 40.50 EUR

79 Alexander Kauertz: Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben
ISBN 978-3-8325-1925-4 40.50 EUR

80 Regina Hübinger: Schüler auf Weltreise. *Entwicklung und Evaluation von Lehr-/ Lernmaterialien zur Förderung experimentell-naturwissenschaftlicher Kompetenzen für die Jahrgangsstufen 5 und 6*
ISBN 978-3-8325-1932-2 40.50 EUR

81 Christine Waltner: Physik lernen im Deutschen Museum
ISBN 978-3-8325-1933-9 40.50 EUR

82 Torsten Fischer: Handlungsmuster von Physiklehrkräften beim Einsatz neuer Medien. *Fallstudien zur Unterrichtspraxis*
ISBN 978-3-8325-1948-3 42.00 EUR

83 Corinna Kieren: Chemiehausaufgaben in der Sekundarstufe I des Gymnasiums. *Fragebogenerhebung zur gegenwärtigen Praxis und Entwicklung eines optimierten Hausaufgabendesigns im Themenbereich Säure-Base*
978-3-8325-1975-9 37.00 EUR

84 Marco Thiele: Modelle der Thermohalinen Zirkulation im Unterricht. *Eine empirische Studie zur Förderung des Modellverständnisses*
ISBN 978-3-8325-1982-7 40.50 EUR

85 Bernd Zinn: Physik lernen, um Physik zu lehren. *Eine Möglichkeit für interessanteren Physikunterricht*
ISBN 978-3-8325-1995-7 39.50 EUR

86 Esther Klaes: Außerschulische Lernorte im naturwissenschaftlichen Unterricht. *Die Perspektive der Lehrkraft*
ISBN 978-3-8325-2006-9 43.00 EUR

87 Marita Schmidt: Kompetenzmodellierung und -diagnostik im Themengebiet Energie der Sekundarstufe I. *Entwicklung und Erprobung eines Testinventars*
ISBN 978-3-8325-2024-3 37.00 EUR

88 Gudrun Franke-Braun: Aufgaben mit gestuften Lernhilfen. *Ein Aufgabenformat zur Förderung der sachbezogenen Kommunikation und Lernleistung für den naturwissenschaftlichen Unterricht*
ISBN 978-3-8325-2026-7 38.00 EUR

89 Silke Klos: Kompetenzförderung im naturwissenschaftlichen Anfangsunterricht. *Der Einfluss eines integrierten Unterrichtskonzepts*
ISBN 978-3-8325-2133-2 37.00 EUR

90 Ulrike Elisabeth Burkard: Quantenphysik in der Schule. *Bestandsaufnahme, Perspektiven und Weiterentwicklungsmöglichkeiten durch die Implementation eines Medienservers*
ISBN 978-3-8325-2215-5 43.00 EUR

91 Ulrike Gromadecki: Argumente in physikalischen Kontexten. *Welche Geltungsgründe halten Physikanfänger für überzeugend?*
ISBN 978-3-8325-2250-6 41.50 EUR

92 Jürgen Bruns: Auf dem Weg zur Förderung naturwissenschaftsspezifischer Vorstellungen von zukünftigen Chemie-Lehrenden
ISBN 978-3-8325-2257-5 43.50 EUR

93 Cornelius Marsch: Räumliche Atomvorstellung. *Entwicklung und Erprobung eines Unterrichtskonzeptes mit Hilfe des Computers*
ISBN 978-3-8325-2293-3 82.50 EUR

94 Maja Brückmann: Sachstrukturen im Physikunterricht. *Ergebnisse einer Videostudie*
ISBN 978-3-8325-2272-8 39.50 EUR

95 Sabine Fechner: Effects of Context-oriented Learning on Student Interest and Achievement in Chemistry Education
ISBN 978-3-8325-2343-5 36.50 EUR

96 Clemens Nagel: eLearning im Physikalischen Anfängerpraktikum
ISBN 978-3-8325-2355-8 39.50 EUR

97 Josef Riese: Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften
ISBN 978-3-8325-2376-3 39.00 EUR

98 Sascha Bernholt: Kompetenzmodellierung in der Chemie. *Theoretische und empirische Reflexion am Beispiel des Modells hierarchischer Komplexität*
ISBN 978-3-8325-2447-0 40.00 EUR

99 Holger Christoph Stawitz: Auswirkung unterschiedlicher Aufgabenprofile auf die Schülerleistung. *Vergleich von Naturwissenschafts- und Problemlöseaufgaben der PISA 2003-Studie*
ISBN 978-3-8325-2451-7 37.50 EUR

100 Hans Ernst Fischer, Elke Sumfleth (Hrsg.): nwu-essen – 10 Jahre Essener Forschung zum naturwissenschaftlichen Unterricht
ISBN 978-3-8325-3331-1 40.00 EUR

101 Hendrik Härtig: Sachstrukturen von Physikschulbüchern als Grundlage zur Bestimmung der Inhaltsvalidität eines Tests
ISBN 978-3-8325-2512-5 34.00 EUR

102 Thomas Grüß-Niehaus: Zum Verständnis des Löslichkeitskonzeptes im Chemieunterricht. *Der Effekt von Methoden progressiver und kollaborativer Reflexion*
ISBN 978-3-8325-2537-8 40.50 EUR

103 Patrick Bronner: Quantenoptische Experimente als Grundlage eines Curriculums zur Quantenphysik des Photons
ISBN 978-3-8325-2540-8 36.00 EUR

104 Adrian Voßkühler: Blickbewegungsmessung an Versuchsaufbauten. *Studien zur Wahrnehmung, Verarbeitung und Usability von physikbezogenen Experimenten am Bildschirm und in der Realität*
ISBN 978-3-8325-2548-4 47.50 EUR

105 Verena Tobias: Newton'sche Mechanik im Anfangsunterricht. *Die Wirksamkeit einer Einführung über die zweidimensionale Dynamik auf das Lehren und Lernen*
ISBN 978-3-8325-2558-3 54.00 EUR

106 Christian Rogge: Entwicklung physikalischer Konzepte in aufgabenbasierten Lernumgebungen
ISBN 978-3-8325-2574-3 45.00 EUR

107 Mathias Ropohl: Modellierung von Schülerkompetenzen im Basiskonzept Chemische Reaktion. *Entwicklung und Analyse von Testaufgaben*
ISBN 978-3-8325-2609-2 36.50 EUR

108 Christoph Kulgemeyer: Physikalische Kommunikationskompetenz. *Modellierung und Diagnostik*
ISBN 978-3-8325-2674-0 44.50 EUR

109 Jennifer Olszewski: The Impact of Physics Teachers' Pedagogical Content Knowledge on Teacher Actions and Student Outcomes
ISBN 978-3-8325-2680-1 33.50 EUR

110 Annika Ohle: Primary School Teachers' Content Knowledge in Physics and its Impact on Teaching and Students' Achievement
ISBN 978-3-8325-2684-9 36.50 EUR

111 Susanne Mannel: Assessing scientific inquiry. *Development and evaluation of a test for the low-performing stage*
ISBN 978-3-8325-2761-7 40.00 EUR

112 Michael Plomer: Physik physiologisch passend praktiziert. *Eine Studie zur Lernwirksamkeit von traditionellen und adressatenspezifischen Physikpraktika für die Physiologie*
ISBN 978-3-8325-2804-1 34.50 EUR

113 Alexandra Schulz: Experimentierspezifische Qualitätsmerkmale im Chemieunterricht. *Eine Videostudie*
ISBN 978-3-8325-2817-1 40.00 EUR

114 Franz Boczianowski: Eine empirische Untersuchung zu Vektoren im Physikunterricht der Mittelstufe
ISBN 978-3-8325-2843-0 39.50 EUR

115 Maria Ploog: Internetbasiertes Lernen durch Textproduktion im Fach Physik
ISBN 978-3-8325-2853-9 39.50 EUR

116 Anja Dhein: Lernen in Explorier- und Experimentiersituationen. *Eine explorative Studie zu Bedeutungsentwicklungsprozessen bei Kindern im Alter zwischen 4 und 6 Jahren*
ISBN 978-3-8325-2859-1 45.50 EUR

117 Irene Neumann: Beyond Physics Content Knowledge. *Modeling Competence Regarding Nature of Scientific Inquiry and Nature of Scientific Knowledge*
ISBN 978-3-8325-2880-5 37.00 EUR

118 Markus Emden: Prozessorientierte Leistungsmessung des naturwissenschaftlich-experimentellen Arbeitens. *Eine vergleichende Studie zu Diagnoseinstrumenten zu Beginn der Sekundarstufe I*
ISBN 978-3-8325-2867-6 38.00 EUR

119 Birgit Hofmann: Analyse von Blickbewegungen von Schülern beim Lesen von physikbezogenen Texten mit Bildern. *Eye Tracking als Methodenwerkzeug in der physikdidaktischen Forschung*
ISBN 978-3-8325-2925-3 59.00 EUR

120 Rebecca Knobloch: Analyse der fachinhaltlichen Qualität von Schüleräußerungen und deren Einfluss auf den Lernerfolg. *Eine Videostudie zu kooperativer Kleingruppenarbeit*
ISBN 978-3-8325-3006-8 36.50 EUR

121 Julia Hostenbach: Entwicklung und Prüfung eines Modells zur Beschreibung der Bewertungskompetenz im Chemieunterricht
ISBN 978-3-8325-3013-6 38.00 EUR

122 Anna Windt: Naturwissenschaftliches Experimentieren im Elementarbereich. *Evaluation verschiedener Lernsituationen*
ISBN 978-3-8325-3020-4 43.50 EUR

123 Eva Kölbach: Kontexteinflüsse beim Lernen mit Lösungsbeispielen
ISBN 978-3-8325-3025-9 38.50 EUR

124 Anna Lau: Passung und vertikale Vernetzung im Chemie- und Physikunterricht
ISBN 978-3-8325-3021-1 36.00 EUR

125 Jan Lamprecht: Ausbildungswege und Komponenten professioneller Handlungskompetenz. *Vergleich von Quereinsteigern mit Lehramtsabsolventen für Gymnasien im Fach Physik*
ISBN 978-3-8325-3035-8 38.50 EUR

126 Ulrike Böhm: Förderung von Verstehensprozessen unter Einsatz von Modellen
ISBN 978-3-8325-3042-6 41.00 EUR

127 Sabrina Dollny: Entwicklung und Evaluation eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Chemielehrkräften
ISBN 978-3-8325-3046-4 37.00 EUR

128 Monika Zimmermann: Naturwissenschaftliche Bildung im Kindergarten. *Eine integrative Längsschnittstudie zur Kompetenzentwicklung von Erzieherinnen*
ISBN 978-3-8325-3053-2 54.00 EUR

129 Ulf Saballus: Über das Schlussfolgern von Schülerinnen und Schülern zu öffentlichen Kontroversen mit naturwissenschaftlichem Hintergrund. *Eine Fallstudie*
ISBN 978-3-8325-3086-0 39.50 EUR

130 Olaf Krey: Zur Rolle der Mathematik in der Physik. *Wissenschaftstheoretische Aspekte und Vorstellungen Physiklernender*
ISBN 978-3-8325-3101-0 46.00 EUR

131 Angelika Wolf: Zusammenhänge zwischen der Eigenständigkeit im Physikunterricht, der Motivation, den Grundbedürfnissen und dem Lernerfolg von Schülern
ISBN 978-3-8325-3161-4 45.00 EUR

132 Johannes Börlin: Das Experiment als Lerngelegenheit. *Vom interkulturellen Vergleich des Physikunterrichts zu Merkmalen seiner Qualität*
ISBN 978-3-8325-3170-6 45.00 EUR

133 Olaf Uhden: Mathematisches Denken im Physikunterricht. *Theorieentwicklung und Problemanalyse*
ISBN 978-3-8325-3170-6 45.00 EUR

134 Christoph Gut: Modellierung und Messung experimenteller Kompetenz. *Analyse eines large-scale Experimentiertests*
ISBN 978-3-8325-3213-0 40.00 EUR

135 Antonio Rueda: Lernen mit ExploMultimedial in kolumbianischen Schulen. *Analyse von kurzzeitigen Lernprozessen und der Motivation beim länderübergreifenden Einsatz einer deutschen computergestützten multimedialen Lernumgebung für den naturwissenschaftlichen Unterricht*
ISBN 978-3-8325-3218-5 45.50 EUR

136 Krisztina Berger: Bilder, Animationen und Notizen. *Empirische Untersuchung zur Wirkung einfacher visueller Repräsentationen und Notizen auf den Wissenserwerb in der Optik*
ISBN 978-3-8325-3238-3 41.50 EUR

137 Antony Crossley: Untersuchung des Einflusses unterschiedlicher physikalischer Konzepte auf den Wissenserwerb in der Thermodynamik der Sekundarstufe I
ISBN 978-3-8325-3275-8 40.00 EUR

138 Tobias Viering: Entwicklung physikalischer Kompetenz in der Sekundarstufe I. *Validierung eines Kompetenzentwicklungsmodells für das Energiekonzept im Bereich Fachwissen*
ISBN 978-3-8325-3277-2 37.00 EUR

139 Nico Schreiber: Diagnostik experimenteller Kompetenz. *Validierung technologiegestützter Testverfahren im Rahmen eines Kompetenzstrukturmodells*
ISBN 978-3-8325-3284-0 39.00 EUR

140 Sarah Hundertmark: Einblicke in kollaborative Lernprozesse. *Eine Fallstudie zur reflektierenden Zusammenarbeit unterstützt durch die Methoden Concept Mapping und Lernbegleitbogen*
ISBN 978-3-8325-3251-2 43.00 EUR

141 Ronny Scherer: Analyse der Struktur, Messinvarianz und Ausprägung komplexer Problemlösekompetenz im Fach Chemie. *Eine Querschnittstudie in der Sekundarstufe I und am Übergang zur Sekundarstufe II*
ISBN 978-3-8325-3312-0 43.00 EUR

142 Patricia Heitmann: Bewertungskompetenz im Rahmen naturwissenschaftlicher Problemlöseprozesse. *Modellierung und Diagnose der Kompetenzen Bewertung und analytisches Problemlösen für das Fach Chemie*
ISBN 978-3-8325-3314-4 37.00 EUR

143 Jan Fleischhauer: Wissenschaftliches Argumentieren und Entwicklung von Konzepten beim Lernen von Physik
ISBN 978-3-8325-3325-0 35.00 EUR

144 Nermin Özcan: Zum Einfluss der Fachsprache auf die Leistung im Fach Chemie. *Eine Förderstudie zur Fachsprache im Chemieunterricht*
ISBN 978-3-8325-3328-1 36.50 EUR

145 Helena van Vorst: Kontextmerkmale und ihr Einfluss auf das Schülerinteresse im Fach Chemie
ISBN 978-3-8325-3321-2 38.50 EUR

146 Janine Cappell: Fachspezifische Diagnosekompetenz angehender Physiklehrkräfte in der ersten Ausbildungsphase
ISBN 978-3-8325-3356-4 38.50 EUR

147 Susanne Bley: Förderung von Transferprozessen im Chemieunterricht
ISBN 978-3-8325-3407-3 40.50 EUR

148 Cathrin Blaes: Die übungsgestützte Lehrerpräsentation im Chemieunterricht der Sekundarstufe I. *Evaluation der Effektivität*
ISBN 978-3-8325-3409-7 43.50 EUR

149 Julia Suckut: Die Wirksamkeit von piko-OWL als Lehrerfortbildung. Eine Evaluation zum Projekt *Physik im Kontext* in Fallstudien
ISBN 978-3-8325-3440-0 45.00 EUR

150 Alexandra Dorschu: Die Wirkung von Kontexten in Physikkompetenztestaufgaben
ISBN 978-3-8325-3446-2 37.00 EUR

151 Jochen Scheid: Multiple Repräsentationen, Verständnis physikalischer Experimente und kognitive Aktivierung: *Ein Beitrag zur Entwicklung der Aufgabenkultur*
ISBN 978-3-8325-3449-3 49.00 EUR

152 Tim Plasa: Die Wahrnehmung von Schülerlaboren und Schülerforschungszentren
ISBN 978-3-8325-3483-7 35.50 EUR

153 Felix Schoppmeier: Physikkompetenz in der gymnasialen Oberstufe.*Entwicklung und Validierung eines Kompetenzstrukturmodells für den Kompetenzbereich Umgang mit Fachwissen*
ISBN 978-3-8325-3502-5 36.00 EUR

154 Katharina Groß: Experimente alternativ dokumentieren. *Eine qualitative Studie zur Förderung der Diagnose- und Differenzierungskompetenz in der Chemielehrerbildung*
ISBN 978-3-8325-3508-7 43.50 EUR

155 Barbara Hank: Konzeptwandelprozesse im Anfangsunterricht Chemie. *Eine quasiexperimentelle Längsschnittstudie*
ISBN 978-3-8325-3519-3 38.50 EUR

156 Katja Freyer: Zum Einfluss von Studieneingangsvoraussetzungen auf den Studienerfolg Erstsemesterstudierender im Fach Chemie
ISBN 978-3-8325-3544-5 38.00 EUR

157 Alexander Rachel: Auswirkungen instruktionaler Hilfen bei der Einführung des (Ferro-)Magnetismus. *Eine Vergleichsstudie in der Primar- und Sekundarstufe*
ISBN 978-3-8325-3548-3 43.50 EUR

158 Sebastian Ritter: Einfluss des Lerninhalts Nanogrößeneffekte auf Teilchen- und Teilchenmodellvorstellungen von Schülerinnen und Schülern
ISBN 978-3-8325-3558-2 36.00 EUR

159 Andrea Harbach: Problemorientierung und Vernetzung in kontextbasierten Lernaufgaben
ISBN 978-3-8325-3564-3 39.00 EUR

160 David Obst: Interaktive Tafeln im Physikunterricht. *Entwicklung und Evaluation einer Lehrerfortbildung*
ISBN 978-3-8325-3582-7 40.50 EUR

161 Sophie Kirschner: Modellierung und Analyse des Professionswissens von Physiklehrkräften
ISBN 978-3-8325-3601-5 35.00 EUR

162 Katja Stief: Selbstregulationsprozesse und Hausaufgabenmotivation im Chemieunterricht
ISBN 978-3-8325-3631-2 34.00 EUR

163 Nicola Meschede: Professionelle Wahrnehmung der inhaltlichen Strukturierung im naturwissenschaftlichen Grundschulunterricht. *Theoretische Beschreibung und empirische Erfassung*
ISBN 978-3-8325-3668-8 37.00 EUR

164 Johannes Maximilian Barth: Experimentieren im Physikunterricht der gymnasialen Oberstufe. *Eine Rekonstruktion übergeordneter Einbettungsstrategien*
ISBN 978-3-8325-3681-7 39.00 EUR

165 Sandra Lein: Das Betriebspraktikum in der Lehrerbildung. *Eine Untersuchung zur Förderung der Wissenschafts- und Technikbildung im allgemeinbildenden Unterricht*
ISBN 978-3-8325-3698-5 40.00 EUR

166 Veranika Maiseyenka: Modellbasiertes Experimentieren im Unterricht. *Praxistauglichkeit und Lernwirkungen*
ISBN 978-3-8325-3708-1 38.00 EUR

167 Christoph Stolzenberger: Der Einfluss der didaktischen Lernumgebung auf das Erreichen geforderter Bildungsziele am Beispiel der W- und P-Seminare im Fach Physik
ISBN 978-3-8325-3708-1 38.00 EUR

168 Pia Altenburger: Mehrebenenregressionsanalysen zum Physiklernen im Sachunterricht der Primarstufe. *Ergebnisse einer Evaluationsstudie.*
ISBN 978-3-8325-3717-3 37.50 EUR

169 Nora Ferber: Entwicklung und Validierung eines Testinstruments zur Erfassung von Kompetenzentwicklung im Fach Chemie in der Sekundarstufe I
ISBN 978-3-8325-3727-2 39.50 EUR

170 Anita Stender: Unterrichtsplanung: Vom Wissen zum Handeln.
Theoretische Entwicklung und empirische Überprüfung des Transformationsmodells der Unterrichtsplanung
ISBN 978-3-8325-3750-0 41.50 EUR

171 Jenna Koenen: Entwicklung und Evaluation von experimentunterstützten Lösungsbeispielen zur Förderung naturwissenschaftlich-experimenteller Arbeitsweisen
ISBN 978-3-8325-3785-2 43.00 EUR

172 Teresa Henning: Empirische Untersuchung kontextorientierter Lernumgebungen in der Hochschuldidaktik. *Entwicklung und Evaluation kontextorientierter Aufgaben in der Studieneingangsphase für Fach- und Nebenfachstudierende der Physik*
ISBN 978-3-8325-3801-9 43.00 EUR

173 Alexander Pusch: Fachspezifische Instrumente zur Diagnose und individuellen Förderung von Lehramtsstudierenden der Physik
ISBN 978-3-8325-3829-3 38.00 EUR

174 Christoph Vogelsang: Validierung eines Instruments zur Erfassung der professionellen Handlungskompetenz von (angehenden) Physiklehrkräften. *Zusammenhangsanalysen zwischen Lehrerkompetenz und Lehrerperformanz*
ISBN 978-3-8325-3846-0 50.50 EUR

175 Ingo Brebeck: Selbstreguliertes Lernen in der Studieneingangsphase im Fach Chemie
ISBN 978-3-8325-3859-0 37.00 EUR

176 Axel Eghtessad: Merkmale und Strukturen von Professionalisierungsprozessen in der ersten und zweiten Phase der Chemielehrerbildung. *Eine empirisch-qualitative Studie mit niedersächsischen Fachleiter_innen der Sekundarstufenlehrämter*
ISBN 978-3-8325-3861-3 45.00 EUR

177 Andreas Nehring: Wissenschaftliche Denk- und Arbeitsweisen im Fach Chemie. Eine kompetenzorientierte Modell- und Testentwicklung für den Bereich der Erkenntnisgewinnung
ISBN 978-3-8325-3872-9 39.50 EUR

178 Maike Schmidt: Professionswissen von Sachunterrichtslehrkräften. Zusammenhangsanalyse zur Wirkung von Ausbildungshintergrund und Unterrichtserfahrung auf das fachspezifische Professionswissen im Unterrichtsinhalt „Verbrennung“
ISBN 978-3-8325-3907-8 38.50 EUR

179 Jan Winkelmann: Auswirkungen auf den Fachwissenszuwachs und auf affektive Schülermerkmale durch Schüler- und Demonstrationsexperimente im Physikunterricht
ISBN 978-3-8325-3915-3 41.00 EUR

180 Iwen Kobow: Entwicklung und Validierung eines Testinstrumentes zur Erfassung der Kommunikationskompetenz im Fach Chemie
ISBN 978-3-8325-3927-6 34.50 EUR

181 Yvonne Gramzow: Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik. Modellierung und Testkonstruktion
ISBN 978-3-8325-3931-3 42.50 EUR

182 Evelin Schröter: Entwicklung der Kompetenzerwartung durch Lösen physikalischer Aufgaben einer multimedialen Lernumgebung
ISBN 978-3-8325-3975-7 54.50 EUR

183 Inga Kallweit: Effektivität des Einsatzes von Selbsteinschätzungsbögen im Chemieunterricht der Sekundarstufe I. *Individuelle Förderung durch selbstreguliertes Lernen*
ISBN 978-3-8325-3965-8 44.00 EUR

184 Andrea Schumacher: Paving the way towards authentic chemistry teaching. *A contribution to teachers' professional development*
ISBN 978-3-8325-3976-4 48.50 EUR

185 David Woitkowski: Fachliches Wissen Physik in der Hochschulausbildung. *Konzeptualisierung, Messung, Niveaubildung*
ISBN 978-3-8325-3988-7 53.00 EUR

186 Marianne Korner: Cross-Age Peer Tutoring in Physik. *Evaluation einer Unterrichtsmethode*
ISBN 978-3-8325-3979-5 38.50 EUR

187 Simone Nakoinz: Untersuchung zur Verknüpfung submikroskopischer und makroskopischer Konzepte im Fach Chemie
ISBN 978-3-8325-4057-9 38.50 EUR

188 Sandra Anus: Evaluation individueller Förderung im Chemieunterricht.*Adaptivität von Lerninhalten an das Vorwissen von Lernenden am Beispiel des Basiskonzeptes Chemische Reaktion*
ISBN 978-3-8325-4059-3 43.50 EUR

189 Thomas Roßbegalle: Fachdidaktische Entwicklungsforschung zum besseren Verständnis atmosphärischer Phänomene. *Treibhauseffekt, saurer Regen und stratosphärischer Ozonabbau als Kontexte zur Vermittlung von Basiskonzepten der Chemie*
ISBN 978-3-8325-4059-3 45.50 EUR

190 Kathrin Steckenmesser-Sander: Gemeinsamkeiten und Unterschiede physikbezogener Handlungs-, Denk- und Lernprozesse von Mädchen und Jungen
ISBN 978-3-8325-4066-1 38.50 EUR

191 Cornelia Geller: Lernprozessorientierte Sequenzierung des Physikunterrichts im Zusammenhang mit Fachwissenserwerb. *Eine Videostudie in Finnland, Deutschland und der Schweiz*
ISBN 978-3-8325-4082-1 35.50 EUR

192 Jan Hofmann: Untersuchung des Kompetenzaufbaus von Physiklehrkräften während einer Fortbildungsmaßnahme
ISBN 978-3-8325-4104-0 38.50 EUR

193 Andreas Dickhäuser: Chemiespezifischer Humor. *Theoriebildung, Materialentwicklung, Evaluation*
ISBN 978-3-8325-4108-8 37.00 EUR

194 Stefan Korte: Die Grenzen der Naturwissenschaft als Thema des Physikunterrichts
ISBN 978-3-8325-4112-5 57.50 EUR

195 Carolin Hülsmann: Kurswahlmotive im Fach Chemie. Eine Studie zum Wahlverhalten und Erfolg von Schülerinnen und Schülern in der gymnasialen Oberstufe
ISBN 978-3-8325-4144-6 49.00 EUR

196 Caroline Körbs: Mindeststandards im Fach Chemie am Ende der Pflichtschulzeit
ISBN 978-3-8325-4148-4 34.00 EUR

197 Andreas Vorholzer: Wie lassen sich Kompetenzen des experimentellen Denkens und Arbeitens fördern? *Eine empirische Untersuchung der Wirkung eines expliziten und eines impliziten Instruktionsansatzes*
ISBN 978-3-8325-4194-1 37.50 EUR

198 Anna Katharina Schmitt: Entwicklung und Evaluation einer Chemielehrerfortbildung zum Kompetenzbereich Erkenntnisgewinnung
ISBN 978-3-8325-4228-3 39.50 EUR

199 Christian Maurer: Strukturierung von Lehr-Lern-Sequenzen
ISBN 978-3-8325-4247-4 36.50 EUR

200 Helmut Fischler, Elke Sumfleth (Hrsg.): Professionelle Kompetenz von Lehrkräften der Chemie und Physik
ISBN 978-3-8325-4523-9 34.00 EUR

201 Simon Zander: Lehrerfortbildung zu Basismodellen und Zusammenhänge zum Fachwissen
ISBN 978-3-8325-4248-1 35.00 EUR

202 Kerstin Arndt: Experimentierkompetenz erfassen.
Analyse von Prozessen und Mustern am Beispiel von Lehramtsstudierenden der Chemie
ISBN 978-3-8325-4266-5 45.00 EUR

203 Christian Lang: Kompetenzorientierung im Rahmen experimentalchemischer Praktika
ISBN 978-3-8325-4268-9 42.50 EUR

204 Eva Cauet: Testen wir relevantes Wissen? *Zusammenhang zwischen dem Professionswissen von Physiklehrkräften und gutem und erfolgreichem Unterrichten*
ISBN 978-3-8325-4276-4 39.50 EUR

205 Patrick Löffler: Modellanwendung in Problemlöseaufgaben. *Wie wirkt Kontext?*
ISBN 978-3-8325-4303-7 35.00 EUR

206 Carina Gehlen: Kompetenzstruktur naturwissenschaftlicher Erkenntnisgewinnung im Fach Chemie
ISBN 978-3-8325-4318-1 43.00 EUR

207 Lars Oettinghaus: Lehrerüberzeugungen und physikbezogenes Professionswissen. *Vergleich von Absolventinnen und Absolventen verschiedener Ausbildungswege im Physikreferendariat*
ISBN 978-3-8325-4319-8 38.50 EUR

208 Jennifer Petersen: Zum Einfluss des Merkmals Humor auf die Gesundheitsförderung im Chemieunterricht der Sekundarstufe I.
Eine Interventionsstudie zum Thema Sonnenschutz
ISBN 978-3-8325-4348-8 40.00 EUR

209 Philipp Straube: Modellierung und Erfassung von Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung bei (Lehramts-) Studierenden im Fach Physik
ISBN 978-3-8325-4351-8 35.50 EUR

210 Martin Dickmann: Messung von Experimentierfähigkeiten. *Validierungsstudien zur Qualität eines computerbasierten Testverfahrens*
ISBN 978-3-8325-4356-3 41.00 EUR

211 Markus Bohlmann: Science Education. Empirie, Kulturen und Mechanismen der Didaktik der Naturwissenschaften
ISBN 978-3-8325-4377-8 44.00 EUR

212 Martin Draude: Die Kompetenz von Physiklehrkräften, Schwierigkeiten von Schülerinnen und Schülern beim eigenständigen Experimentieren zu diagnostizieren
ISBN 978-3-8325-4382-2 37.50 EUR

213 Henning Rode: Prototypen evidenzbasierten Physikunterrichts. *Zwei empirische Studien zum Einsatz von Feedback und Blackboxes in der Sekundarstufe*
ISBN 978-3-8325-4389-1 42.00 EUR

214 Jan-Henrik Kechel: Schülerschwierigkeiten beim eigenständigen Experimentieren. *Eine qualitative Studie am Beispiel einer Experimentieraufgabe zum Hooke'schen Gesetz*
ISBN 978-3-8325-4392-1 55.00 EUR

215 Katharina Fricke: Classroom Management and its Impact on Lesson Outcomes in Physics. *A multi-perspective comparison of teaching practices in primary and secondary schools*
ISBN 978-3-8325-4394-5 40.00 EUR

216 Hannes Sander: Orientierungen von Jugendlichen beim Urteilen und Entscheiden in Kontexten nachhaltiger Entwicklung. *Eine rekonstruktive Perspektive auf Bewertungskompetenz in der Didaktik der Naturwissenschaft*
ISBN 978-3-8325-4434-8 46.00 EUR

217 Inka Haak: Maßnahmen zur Unterstützung kognitiver und metakognitiver Prozesse in der Studieneingangsphase. *Eine Design-Based-Research-Studie zum universitären Lernzentrum Physiktreff*
ISBN 978-3-8325-4437-9 46.50 EUR

218 Martina Brandenburger: Was beeinflusst den Erfolg beim Problemlösen in der Physik? *Eine Untersuchung mit Studierenden*
ISBN 978-3-8325-4409-6 42.50 EUR

219 Corinna Helms: Entwicklung und Evaluation eines Trainings zur Verbesserung der Erklärqualität von Schülerinnen und Schülern im Gruppenpuzzle
ISBN 978-3-8325-4454-6 42.50 EUR

220 Viktoria Rath: Diagnostische Kompetenz von angehenden Physiklehrkräften. *Modellierung, Testinstrumentenentwicklung und Erhebung der Performanz bei der Diagnose von Schülervorstellungen in der Mechanik*
ISBN 978-3-8325-4456-0 42.50 EUR

221 Janne Krüger: Schülerperspektiven auf die zeitliche Entwicklung der Naturwissenschaften
ISBN 978-3-8325-4457-7 45.50 EUR

222 Stefan Mutke: Das Professionswissen von Chemiereferendarinnen und -referendaren in Nordrhein-Westfalen. *Eine Längsschnittstudie*
ISBN 978-3-8325-4458-4 37.50 EUR

223 Sebastian Habig: Systematisch variierte Kontextaufgaben und ihr Einfluss auf kognitive und affektive Schülerfaktoren
ISBN 978-3-8325-4467-6 40.50 EUR

224 Sven Liepertz: Zusammenhang zwischen dem Professionswissen von Physiklehrkräften, dem sachstrukturellen Angebot des Unterrichts und der Schülerleistung
ISBN 978-3-8325-4480-5 34.00 EUR

225 Elina Platova: Optimierung eines Laborpraktikums durch kognitive Aktivierung
ISBN 978-3-8325-4481-2 39.00 EUR

226 Tim Reschke: Lesegeschichten im Chemieunterricht der Sekundarstufe I zur Unterstützung von situationalem Interesse und Lernerfolg
ISBN 978-3-8325-4487-4 41.00 EUR

227 Lena Mareike Walper: Entwicklung der physikbezogenen Interessen und selbstbezogenen Kognitionen von Schülerinnen und Schülern in der Übergangsphase von der Primar- in die Sekundarstufe. *Eine Längsschnittanalyse vom vierten bis zum siebten Schuljahr*
ISBN 978-3-8325-4495-9 43.00 EUR

228 Stefan Anthofer: Förderung des fachspezifischen Professionswissens von Chemielehramtsstudierenden
ISBN 978-3-8325-4498-0 39.50 EUR

229 Marcel Bullinger: Handlungsorientiertes Physiklernen mit instruierten Selbsterklärungen in der Primarstufe. *Eine experimentelle Laborstudie*
ISBN 978-3-8325-4504-8 44.00 EUR

230 Thomas Amenda: Bedeutung fachlicher Elementarisierungen für das Verständnis der Kinematik
ISBN 978-3-8325-4531-4 43.50 EUR

231 Sabrina Milke: Beeinflusst *Priming* das Physiklernen?
Eine empirische Studie zum Dritten Newtonschen Axiom
ISBN 978-3-8325-4549-4 42.00 EUR

232 Corinna Erfmann: Ein anschaulicher Weg zum Verständnis der elektromagnetischen Induktion. *Evaluation eines Unterrichtsvorschlags und Validierung eines Leistungsdiagnoseinstruments*
ISBN 978-3-8325-4550-5 49.50 EUR

233 Hanne Rautenstrauch: Erhebung des (Fach-)Sprachstandes bei Lehramtsstudierenden im Kontext des Faches Chemie
ISBN 978-3-8325-4556-7 40.50 EUR

234 Tobias Klug: Wirkung kontextorientierter physikalischer Praktikumsversuche auf Lernprozesse von Studierenden der Medizin
ISBN 978-3-8325-4558-1 37.00 EUR

235 Mareike Bohrmann: Zur Förderung des Verständnisses der Variablenkontrolle im naturwissenschaftlichen Sachunterricht
ISBN 978-3-8325-4559-8 52.00 EUR

236 Anja Schödl: FALKO-Physik – Fachspezifische Lehrerkompetenzen im Fach Physik. *Entwicklung und Validierung eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Physiklehrkräften*
ISBN 978-3-8325-4553-6 40.50 EUR

237 Hilda Scheuermann: Entwicklung und Evaluation von Unterstützungsmaßnahmen zur Förderung der Variablenkontrollstrategie beim Planen von Experimenten
ISBN 978-3-8325-4568-0 39.00 EUR

238 Christian G. Strippel: Naturwissenschaftliche Erkenntnisgewinnung an chemischen Inhalten vermitteln. *Konzeption und empirische Untersuchung einer Ausstellung mit Experimentierstation*
ISBN 978-3-8325-4577-2 41.50 EUR

239 Sarah Rau: Durchführung von Sachunterricht im Vorbereitungsdienst. *Eine längsschnittliche, videobasierte Unterrichtsanalyse*
ISBN 978-3-8325-4579-6 46.00 EUR

240 Thomas Plotz: Lernprozesse zu nicht-sichtbarer Strahlung. *Empirische Untersuchungen in der Sekundarstufe 2*
ISBN 978-3-8325-4624-3 39.50 EUR

241 Wolfgang Aschauer: Elektrische und magnetische Felder. *Eine empirische Studie zu Lernprozessen in der Sekundarstufe II*
ISBN 978-3-8325-4625-0 50.00 EUR

242 Anna Donhauser: Didaktisch rekonstruierte Materialwissenschaft. *Aufbau und Konzeption eines Schülerlabors für den Exzellenzcluster Engineering of Advanced Materials*
ISBN 978-3-8325-4636-6 39.00 EUR

243 Katrin Schüßler: Lernen mit Lösungsbeispielen im Chemieunterricht. *Einflüsse auf Lernerfolg, kognitive Belastung und Motivation*
ISBN 978-3-8325-4640-3 42.50 EUR

244 Timo Fleischer: Untersuchung der chemischen Fachsprache unter besonderer Berücksichtigung chemischer Repräsentationen
ISBN 978-3-8325-4642-7 46.50 EUR

245 Rosina Steininger: Concept Cartoons als Stimuli für Kleingruppendiskussionen im Chemieunterricht. *Beschreibung und Analyse einer komplexen Lerngelegenheit*
ISBN 978-3-8325-4647-2 39.00 EUR

246 Daniel Rehfeldt: Erfassung der Lehrqualität naturwissenschaftlicher Experimentalpraktika
ISBN 978-3-8325-4590-1 40.00 EUR

247 Sandra Puddu: Implementing Inquiry-based Learning in a Diverse Classroom: Investigating Strategies of Scaffolding and Students' Views of Scientific Inquiry
ISBN 978-3-8325-4591-8 35.50 EUR

248 Markus Bliersbach: Kreativität in der Chemie. *Erhebung und Förderung der Vorstellungen von Chemielehramtsstudierenden*
ISBN 978-3-8325-4593-2 44.00 EUR

249 Lennart Kimpel: Aufgaben in der Allgemeinen Chemie. *Zum Zusammenspiel von chemischem Verständnis und Rechenfähigkeit*
ISBN 978-3-8325-4618-2 36.00 EUR

250 Louise Bindel: Effects of integrated learning: explicating a mathematical concept in inquiry-based science camps
ISBN 978-3-8325-4655-7 37.50 EUR

251 Michael Wenzel: Computereinsatz in Schule und Schülerlabor. *Einstellung von Physiklehrkräften zu Neuen Medien*
ISBN 978-3-8325-4659-5 38.50 EUR

252 Laura Muth: Einfluss der Auswertephase von Experimenten im Physikunterricht. *Ergebnisse einer Interventionsstudie zum Zuwachs von Fachwissen und experimenteller Kompetenz von Schülerinnen und Schülern*
ISBN 978-3-8325-4675-5 36.50 EUR

253 Annika Fricke: Interaktive Skripte im Physikalischen Praktikum. *Entwicklung und Evaluation von Hypermedien für die Nebenfachausbildung*
ISBN 978-3-8325-4676-2 41.00 EUR

254 Julia Haase: Selbstbestimmtes Lernen im naturwissenschaftlichen Sachunterricht. *Eine empirische Interventionsstudie mit Fokus auf Feedback und Kompetenzerleben*
ISBN 978-3-8325-4685-4 38.50 EUR

255 Antje J. Heine: Was ist Theoretische Physik? *Eine wissenschaftstheoretische Betrachtung und Rekonstruktion von Vorstellungen von Studierenden und Dozenten über das Wesen der Theoretischen Physik*
ISBN 978-3-8325-4691-5 46.50 EUR

256 Claudia Meinhardt: Entwicklung und Validierung eines Testinstruments zu Selbstwirksamkeitserwartungen von (angehenden) Physiklehrkräften in physikdidaktischen Handlungsfeldern
ISBN 978-3-8325-4712-7 47.00 EUR

257 Ann-Kathrin Schlüter: Professionalisierung angehender Chemielehrkräfte für einen Gemeinsamen Unterricht
ISBN 978-3-8325-4713-4 53.50 EUR

258 Stefan Richtberg: Elektronenbahnen in Feldern. Konzeption und Evaluation einer webbasierten Lernumgebung
ISBN 978-3-8325-4723-3 49.00 EUR

259 Jan-Philipp Burde: Konzeption und Evaluation eines Unterrichtskonzepts zu einfachen Stromkreisen auf Basis des Elektronengasmodells
ISBN 978-3-8325-4726-4 57.50 EUR

260 Frank Finkenberg: Flipped Classroom im Physikunterricht
ISBN 978-3-8325-4737-4 42.50 EUR

261 Florian Treisch: Die Entwicklung der Professionellen Unterrichtswahrnehmung im Lehr-Lern-Labor Seminar
ISBN 978-3-8325-4741-4 41.50 EUR

262 Desiree Mayr: Strukturiertheit des experimentellen naturwissenschaftlichen Problemlöseprozesses
ISBN 978-3-8325-4757-8 37.00 EUR

263 Katrin Weber: Entwicklung und Validierung einer Learning Progression für das Konzept der chemischen Reaktion in der Sekundarstufe I
ISBN 978-3-8325-4762-2 48.50 EUR

264 Hauke Bartels: Entwicklung und Bewertung eines performanznahen Videovignettentests zur Messung der Erklärfähigkeit von Physiklehrkräften
ISBN 978-3-8325-4804-9 37.00 EUR

265 Karl Marniok: Zum Wesen von Theorien und Gesetzen in der Chemie. *Begriffsanalyse und Förderung der Vorstellungen von Lehramtsstudierenden*
ISBN 978-3-8325-4805-6 42.00 EUR

266 Marisa Holzapfel: Fachspezifischer Humor als Methode in der Gesundheitsbildung im Übergang von der Primarstufe zur Sekundarstufe I
ISBN 978-3-8325-4808-7 50.00 EUR

267 Anna Stolz: Die Auswirkungen von Experimentiersituationen mit unterschiedlichem Öffnungsgrad auf Leistung und Motivation der Schülerinnen und Schüler
ISBN 978-3-8325-4781-3 38.00 EUR

268 Nina Ulrich: Interaktive Lernaufgaben in dem digitalen Schulbuch eChemBook. *Einfluss des Interaktivitätsgrads der Lernaufgaben und des Vorwissens der Lernenden auf den Lernerfolg*
ISBN 978-3-8325-4814-8 43.50 EUR

269 Kim-Alessandro Weber: Quantenoptik in der Lehrerfortbildung. *Ein bedarfsgeprägtes Fortbildungskonzept zum Quantenobjekt „Photon“ mit Realexperimenten*
ISBN 978-3-8325-4792-9 55.00 EUR

270 Nina Skorsetz: Empathisierer und Systematisierer im Vorschulalter. *Eine Fragebogen- und Videostudie zur Motivation, sich mit Naturphänomenen zu beschäftigen*
ISBN 978-3-8325-4825-4 43.50 EUR

271 Franziska Kehne: Analyse des Transfers von kontextualisiert erworbenem Wissen im Fach Chemie
ISBN 978-3-8325-4846-9 45.00 EUR

272 Markus Elsholz: Das akademische Selbstkonzept angehender Physiklehrkräfte als Teil ihrer professionellen Identität. *Dimensionalität und Veränderung während einer zentralen Praxisphase*
ISBN 978-3-8325-4857-5 37.50 EUR

273 Joachim Müller: Studienerfolg in der Physik. *Zusammenhang zwischen Modellierungskompetenz und Studienerfolg*
ISBN 978-3-8325-4859-9 35.00 EUR

274 Jennifer Dörschelln: Organische Leuchtdioden. *Implementation eines innovativen Themas in den Chemieunterricht*
ISBN 978-3-8325-4865-0 59.00 EUR

275 Stephanie Strelow: Beliefs von Studienanfängern des Kombi-Bachelors Physik über die Natur der Naturwissenschaften
ISBN 978-3-8325-4881-0 40.50 EUR

276 Dennis Jaeger: Kognitive Belastung und aufgabenspezifische sowie personenspezifische Einflussfaktoren beim Lösen von Physikaufgaben
ISBN 978-3-8325-4928-2 50.50 EUR

277 Vanessa Fischer: Der Einfluss von Interesse und Motivation auf die Messung von Fach- und Bewertungskompetenz im Fach Chemie
ISBN 978-3-8325-4933-6 39.00 EUR

278 René Dohrmann: Professionsbezogene Wirkungen einer Lehr-Lern-Labor-Veranstaltung. *Eine multimethodische Studie zu den professionsbezogenen Wirkungen einer Lehr-Lern-Labor-Blockveranstaltung auf Studierende der Bachelorstudiengänge Lehramt Physik und Grundschulpädagogik (Sachunterricht)*
ISBN 978-3-8325-4958-9 40.00 EUR

279 Meike Bergs: Can We Make Them Use These Strategies? *Fostering Inquiry-Based Science Learning Skills with Physical and Virtual Experimentation Environments*
ISBN 978-3-8325-4962-6 39.50 EUR

280 Marie-Therese Hauerstein: Untersuchung zur Effektivität von Strukturierung und Binnendifferenzierung im Chemieunterricht der Sekundarstufe I. *Evaluation der Strukturierungshilfe Lernleiter*
ISBN 978-3-8325-4982-4 42.50 EUR

281 Verena Zucker: Erkennen und Beschreiben von formativem Assessment im naturwissenschaftlichen Grundschulunterricht. *Entwicklung eines Instruments zur Erfassung von Teilfähigkeiten der professionellen Wahrnehmung von Lehramtsstudierenden*
ISBN 978-3-8325-4991-6 38.00 EUR

282 Victoria Telser: Erfassung und Förderung experimenteller Kompetenz von Lehrkräften im Fach Chemie
ISBN 978-3-8325-4996-1 50.50 EUR

283 Kristine Tschirschky: Entwicklung und Evaluation eines gedächtnisorientierten Aufgabendesigns für Physikaufgaben
ISBN 978-3-8325-5002-8 42.50 EUR

284 Thomas Elert: Course Success in the Undergraduate General Chemistry Lab
ISBN 978-3-8325-5004-2 41.50 EUR

285 Britta Kalthoff: Explizit oder implizit? *Untersuchung der Lernwirksamkeit verschiedener fachmethodischer Instruktionen im Hinblick auf fachmethodische und fachinhaltliche Fähigkeiten von Sachunterrichtsstudierenden*
ISBN 978-3-8325-5013-4 37.50 EUR

286 Thomas Dickmann: Visuelles Modellverständnis und Studienerfolg in der Chemie. *Zwei Seiten einer Medaille*
ISBN 978-3-8325-5016-5 44.00 EUR

287 Markus Sebastian Feser: Physiklehrkräfte korrigieren Schülertexte. *Eine Explorationsstudie zur fachlich-konzeptuellen und sprachlichen Leistungsfeststellung und -beurteilung im Physikunterricht*
ISBN 978-3-8325-5020-2 49.00 EUR

288 Matylda Dudzinska: Lernen mit Beispielaufgaben und Feedback im Physikunterricht der Sekundarstufe 1. *Energieerhaltung zur Lösung von Aufgaben nutzen*
ISBN 978-3-8325-5025-7 47.00 EUR

289 Ines Sonnenschein: Naturwissenschaftliche Denk- und Arbeitsprozesse Studierender im Labor
ISBN 978-3-8325-5033-2 52.00 EUR

290 Florian Simon: Der Einfluss von Betreuung und Betreuenden auf die Wirksamkeit von Schülerlaborbesuchen. *Eine Zusammenhangsanalyse von Betreuungsqualität, Betreuermerkmalen und Schülerlaborzielen sowie Replikationsstudie zur Wirksamkeit von Schülerlaborbesuchen*
ISBN 978-3-8325-5036-3 49.50 EUR

291 Marie-Annette Geyer: Physikalisch-mathematische Darstellungswechsel funktionaler Zusammenhänge. *Das Vorgehen von SchülerInnen der Sekundarstufe 1 und ihre Schwierigkeiten*
ISBN 978-3-8325-5047-9 46.50 EUR

292 Susanne Digel: Messung von Modellierungskompetenz in Physik. *Theoretische Herleitung und empirische Prüfung eines Kompetenzmodells physikspezifischer Modellierungskompetenz*
ISBN 978-3-8325-5055-4 41.00 EUR

293 Sönke Janssen: Angebots-Nutzungs-Prozesse eines Schülerlabors analysieren und gestalten. *Ein design-based research Projekt*
ISBN 978-3-8325-5065-3 57.50 EUR

294 Knut Wille: Der Productive Failure Ansatz als Beitrag zur Weiterentwicklung der Aufgabenkultur
ISBN 978-3-8325-5074-5 49.00 EUR

295 Lisanne Kraeva: Problemlösestrategien von Schülerinnen und Schülern diagnostizieren
ISBN 978-3-8325-5110-0 59.50 EUR

296 Jenny Lorentzen: Entwicklung und Evaluation eines Lernangebots im Lehramtsstudium Chemie zur Förderung von Vernetzungen innerhalb des fachbezogenen Professionswissens
ISBN 978-3-8325-5120-9 39.50 EUR

297 Micha Winkelmann: Lernprozesse in einem Schülerlabor unter Berücksichtigung individueller naturwissenschaftlicher Interessenstrukturen
ISBN 978-3-8325-5147-6 48.50 EUR

298 Carina Wöhlke: Entwicklung und Validierung eines Instruments zur Erfassung der professionellen Unterrichtswahrnehmung angehender Physiklehrkräfte
ISBN 978-3-8325-5149-0 43.00 EUR

299 Thomas Schubatzky: Das Amalgam Anfangs-Elektrizitätslehreunterricht. *Eine multiperspektivische Betrachtung in Deutschland und Österreich*
ISBN 978-3-8325-5159-9 50.50 EUR

300 Amany Annaggar: A Design Framework for Video Game-Based Gamification Elements to Assess Problem-solving Competence in Chemistry Education
ISBN 978-3-8325-5150-6 52.00 EUR

301 Alexander Engl: Chemie Pur – Unterrichten in der Natur: *Entwicklung und Evaluation eines kontextorientierten Unterrichtskonzepts im Bereich Outdoor Education zur Änderung der Einstellung zu „Chemie und Natur“*
ISBN 978-3-8325-5174-2 59.00 EUR

302 Christin Marie Sajons: Kognitive und motivationale Dynamik in Schülerlaboren. *Kontextualisierung, Problemorientierung und Autonomieunterstützung der didaktischen Struktur analysieren und weiterentwickeln*
ISBN 978-3-8325-5155-1 56.00 EUR

303 Philipp Bitzenbauer: Quantenoptik an Schulen. *Studie im Mixed-Methods Design zur Evaluation des Erlanger Unterrichtskonzepts zur Quantenoptik*
ISBN 978-3-8325-5123-0 59.00 EUR

304 Malte S. Ubben: Typisierung des Verständnisses mentaler Modelle mittels empirischer Datenerhebung am Beispiel der Quantenphysik
ISBN 978-3-8325-5181-0 43.50 EUR

305 Wiebke Kuske-Janßen: Sprachlicher Umgang mit Formeln von LehrerInnen im Physikunterricht am Beispiel des elektrischen Widerstandes in Klassenstufe 8
ISBN 978-3-8325-5183-4 47.50 EUR

306 Kai Bliesmer: Physik der Küste für außerschulische Lernorte. *Eine Didaktische Rekonstruktion*
ISBN 978-3-8325-5190-2 58.00 EUR

307 Nikola Schild: Eignung von domänenspezifischen Studieneingangsvariablen als Prädiktoren für Studienerfolg im Fach und Lehramt Physik
ISBN 978-3-8325-5226-8 42.00 EUR

308 Daniel Averbeck: Zum Studienerfolg in der Studieneingangsphase des Chemiestudiums. *Der Einfluss kognitiver und affektiv-motivationaler Variablen*
ISBN 978-3-8325-5227-5 51.00 EUR

309 Martina Strübe: Modelle und Experimente im Chemieunterricht. *Eine Videostudie zum fachspezifischen Lehrerwissen und -handeln*
ISBN 978-3-8325-5245-9 45.50 EUR

310 Wolfgang Becker: Auswirkungen unterschiedlicher experimenteller Repräsentationen auf den Kenntnisstand bei Grundschulkindern
ISBN 978-3-8325-5255-8 50.00 EUR

311 Marvin Rost: Modelle als Mittel der Erkenntnisgewinnung im Chemieunterricht der Sekundarstufe I. *Entwicklung und quantitative Dimensionalitätsanalyse eines Testinstruments aus epistemologischer Perspektive*
ISBN 978-3-8325-5256-5 44.00 EUR

312 Christina Kobl: Förderung und Erfassung der Reflexionskompetenz im Fach Chemie
ISBN 978-3-8325-5259-6 41.00 EUR

313 Ann-Kathrin Beretz: Diagnostische Prozesse von Studierenden des Lehramts – *eine Videostudie in den Fächern Physik und Mathematik*
ISBN 978-3-8325-5288-6 45.00 EUR

314 Judith Breuer: Implementierung fachdidaktischer Innovationen durch das Angebot materialgestützter Unterrichtskonzeptionen. *Fallanalysen zum Nutzungsverhalten von Lehrkräften am Beispiel des Münchener Lehrgangs zur Quantenmechanik*
ISBN 978-3-8325-5293-0 50.50 EUR

315 Michaela Oettle: Modellierung des Fachwissens von Lehrkräften in der Teilchenphysik. *Eine Delphi-Studie*
ISBN 978-3-8325-5305-0 57.50 EUR

316 Volker Brüggemann: Entwicklung und Pilotierung eines adaptiven Multistage-Tests zur Kompetenzerfassung im Bereich naturwissenschaftlichen Denkens
ISBN 978-3-8325-5331-9 40.00 EUR

317 Stefan Müller: Die Vorläufigkeit und soziokulturelle Eingebundenheit naturwissenschaftlicher Erkenntnisse. *Kritische Reflexion, empirische Befunde und fachdidaktische Konsequenzen für die Chemielehrer*innenbildung*
ISBN 978-3-8325-5343-2 63.00 EUR

318 Laurence Müller: Alltagsentscheidungen für den Chemieunterricht erkennen und Entscheidungsprozesse explorativ begleiten
ISBN 978-3-8325-5379-1 59.00 EUR

319 Lars Ehlert: Entwicklung und Evaluation einer Lehrkräftefortbildung zur Planung von selbstgesteuerten Experimenten
ISBN 978-3-8325-5393-71 41.50 EUR

320 Florian Seiler: Entwicklung und Evaluation eines Seminarkonzepts zur Förderung der experimentellen Planungskompetenz von Lehramtsstudierenden im Fach Chemie
ISBN 978-3-8325-5397-5 47.50 EUR

321 Nadine Boele: 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: Entwicklung und Evaluation digitalisierungsbezogener Kompetenzen von angehenden Chemielehrkräften
ISBN 978-3-8325-5410-1 49.50 EUR

323 Lars-Frederik Weiß: 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: Kumulatives Lehren und Lernen im Lehramtsstudium Physik. *Theorie und Evaluation eines Lehrkonzepts*
ISBN 978-3-8325-5421-7 51.00 EUR

325 Kübra Nur Celik: 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: 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: 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: Schülerlabore als interessefördernde authentische Lernorte für den naturwissenschaftlichen Unterricht nutzen
ISBN 978-3-8325-5454-5 50.50 EUR

329 Christopher Kurth: Die Kompetenz von Studierenden, Schülerschwierigkeiten beim eigenständigen Experimentieren zu diagnostizieren
ISBN 978-3-8325-5457-6 58.50 EUR

330 Dagmar Michna: 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: 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: 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

333 Erik Heine: 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

334 Simon Goertz: 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: 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

336 Alina Behrendt: 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

337 Manuel Daiber: Entwicklung eines Lehrkonzepts für eine elementare Quantenmechanik *Formuliert mit In-Out Symbolen*
ISBN 978-3-8325-5507-8 48.50 EUR

338 Felix Pawlak: Das Gemeinsame Experimentieren (an-)leiten *Eine qualitative Studie zum chemiespezifischen Classroom-Management*
ISBN 978-3-8325-5508-5 46.50 EUR

339 Liza Dopatka: Konzeption und Evaluation eines kontextstrukturierten Unterrichtskonzeptes für den Anfangs-Elektrizitätslehreunterricht
ISBN 978-3-8325-5514-6 69.50 EUR

340 Arne Bewersdorff: Untersuchung der Effektivität zweier Fortbildungsformate zum Experimentieren mit dem Fokus auf das Unterrichtshandeln
ISBN 978-3-8325-5522-1 39.00 EUR

341 Thomas Christoph Münster: Wie diagnostizieren Studierende des Lehramtes physikbezogene Lernprozesse von Schüler*innen? Eine Videostudie zur Mechanik
ISBN 978-3-8325-5534-4 44.50 EUR

342 Ines Komor: Förderung des symbolisch-mathematischen Modellverständnisses in der Physikalischen Chemie
ISBN 978-3-8325-5546-7 46.50 EUR

343 Verena Petermann: 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

344 Jana Heinze: Einfluss der sprachlichen Konzeption auf die Einschätzung der Qualität instruktionaler Unterrichtserklärungen im Fach Physik
ISBN 978-3-8325-5545-0 47.00 EUR

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

Studien zum Physik- und Chemielernen

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

Die Reihe umfasst inzwischen eine große Zahl von wissenschaftlichen Arbeiten aus vielen Arbeitsgruppen der Physik- und Chemiedidaktik und zeichnet damit ein gültiges Bild der empirischen physik- und chemiedidaktischen Forschung im deutschsprachigen Raum.
Die Herausgeber laden daher Interessenten zu neuen Beiträgen ein und bitten sie, sich im Bedarfsfall an den Logos-Verlag oder an ein Mitglied des Herausgeberteams zu wenden.

Kontaktadressen:

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

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

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

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

In dieser Studie werden zwei längerfristig konzipierte Lehrkräftefortbildungsformate (individualisiert-konstruktiv vs. kokonstruktiv) zum geöffneten Experimentieren hinsichtlich ihrer Wirkung auf das Unterrichtshandeln untersucht. Es wird damit ein Beitrag zur Aufklärung der Wirksamkeit kokonstruktiver Elemente in Fortbildungen geleistet.

Induktiv und deduktiv wurden Kategorien des Unterrichtshandelns zum effektiven Experimentieren zusammengestellt, die bisher in keinem bestehenden Rating- oder Kodiermanual zusammengeführt wurden. Das Manual kam in einer quasi-experimentellen Videostudie zum Einsatz. Die Lehrkräfte ($N = 15$) wurden bzgl. ihres Unterrichtshandelns beim Experimentieren vor sowie nach der Fortbildung videografiert.

Die Studie zeigt mit Blick auf das Unterrichtshandeln beim Experimentieren zwei wesentliche Änderungen: einige Lehrkräfte setzen nach der Fortbildung weniger kochrezeptartige Experimentieranleitungen ein und Experimentierphasen – begleitet von Unterstützungsangeboten – werden geöffnet. Außerdem regen einzelne Lehrkräfte nach der Fortbildung Lernende stärker zu wissenschaftsorientiertem Arbeiten an (‘Minds on’ statt ‘Hands on’). Der Transfer von Fortbildungsinhalten in das Unterrichtshandeln kann nur bei einigen Lehrkräften, differenziell, beobachtet werden. Das Unterrichtshandeln scheint auch gegenüber längerfristigen Interventionen robust: In der Mehrzahl der Aspekte des Unterrichtshandelns zeigen sich keine signifikanten Änderungen.

Logos Verlag Berlin

ISBN 978-3-8325-5522-1