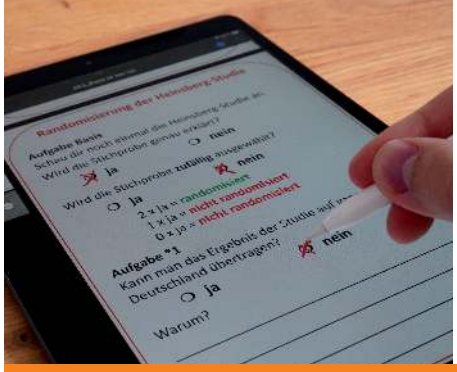


Band 10

Lernen in Naturwissenschaften

verstehen und entwickeln

A. Marohn (Hrsg.)

Christopher Kralisch

Wissenschaftskriterien verstehen und anwenden

Weiterentwicklung der Unterrichtskonzeption choice^{2}reflect zur Bewertung gesellschaftlicher Kontroversen

λογος

Lernen in
Naturwissenschaften
Band 10

Lernen in Naturwissenschaften

Band 10

Herausgegeben von

Prof. Dr. Annette Marohn

Westfälische Wilhelms-Universität Münster
Institut für Didaktik der Chemie

Christopher Kralisch

Wissenschaftskriterien verstehen und anwenden

Weiterentwicklung der Unterrichtskonzeption choice²reflect zur Bewertung gesellschaftlicher Kontroversen

Logos Verlag Berlin

λογος

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 978-3-8325-5517-7
ISSN 2566-493X

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

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

WWU
MÜNSTER

Inaugural-Dissertation

zur Erlangung des Doktorgrades in Erziehungswissenschaften

Wissenschaftskriterien verstehen und anwenden

Weiterentwicklung der Unterrichtskonzeption choice^{2}reflect zur Bewertung gesellschaftlicher Kontroversen

---

verfasst von:

Christopher Kralisch

wohnhaft in Münster

Erstgutachterin:

Prof. Dr. Annette Marohn

Westfälische Wilhelms-Universität Münster – Institut für Didaktik der Chemie

Zweitgutachterin:

Prof. Dr. Katrin Sommer

Ruhr-Universität Bochum – Lehrstuhl für Didaktik der Chemie

Münster, 2022

Chemie
Didaktik
Münster

Zusammenfassung

In der vorliegenden Arbeit wird im Rahmen des Design-Based Research-Ansatzes das Unterrichtskonzept choice^{2}reflect weiterentwickelt und erprobt.

In aktuellen gesellschaftlichen Debatten ist immer häufiger zu beobachten, dass Argumente von Falschinformationen beeinflusst sind und wichtige Entscheidungen nicht auf der Basis wissenschaftlich gesicherter Fakten getroffen werden (Cook et al. 2018). Letztere werden jedoch zum Treffen rationaler Entscheidungen benötigt. Kapitel 1 dieser Arbeit zeigt die Grundlagen der Entscheidungstheorie auf und beschreibt, dass Individuen häufig nur begrenzt rational auf der Basis von Heuristiken und beeinflusst durch Emotionen entscheiden (Pfister et al. 2017; Kahneman et al. 1982; Kahneman 2011). Im Kapitel 2 werden wissenschaftstheoretische Kriterien festgelegt, nach denen die Wissenschaftlichkeit von Informationen bewertet werden kann. Die Bewertungskompetenz (Kapitel 3) lässt sich nicht nur als das „Erkennen und Bewerten naturwissenschaftlicher Sachverhalte in verschiedenen Kontexten“ (z. B. KMK 2005a, S. 7) definieren. Viele weitere Einflüsse, Dimensionen und Kompetenzstrukturmodelle formen die Bewertungskompetenz zu einem umfangreichen Kompetenzbereich (Bögeholz et al. 2004; Reitschert et al. 2007; Eggert & Bögeholz 2006). Dessen Vermittlung hat im naturwissenschaftlichen Unterricht eine große Bedeutung für die Ausbildung einer Scientific Literacy (Prenzel et al. 2007).

Das Unterrichtskonzept choice^{2}reflect wurde entwickelt, um Schülerinnen und Schülern eine rationale Bewertung von gesellschaftlichen Kontroversen auf der Basis wissenschaftlicher Prüfkriterien zu ermöglichen (Jungkamp 2021). Es verläuft in fünf Phasen und ist sowohl fächer- als auch kontextübergreifend einsetzbar. In der Konzeption positionieren sich Schülerinnen und Schüler zu einem kontroversen Thema wie *Homöopathie* oder *Nahrungsergänzungsmittel*. Anschließend erarbeiten sie Wissenschaftskriterien, anhand derer sie Aussagen, Experimente und empirische Studien auf Wissenschaftlichkeit prüfen. Die Kriterien wenden sie auf die Kontroverse an,

um abschließend zu einer reflektierten Entscheidung zu gelangen.

Diese Arbeit verläuft nach dem Design-Based Research-Ansatz (Rohrbach-Lochner 2019; Collins et al. 2004; Collins 1992) und gliedert sich in die Elemente *Framing*, *Design-Experiment* und *Re-Framing*. Das Ziel der Arbeit wird im *Framing* (Teil IV) festgelegt. Es besteht in der Weiterentwicklung der Konzeption durch Ergänzen weiterer Lernmaterialien und Erproben dieser im realen Bildungskontext.
Das *Design-Experiment* (Teil V) verläuft iterativ und besteht aus drei Mesozyklen, in denen Lernmaterial zunächst entwickelt, dann erprobt und nach umfangreicher Analyse optimiert wird.

Der erste Mesozyklus (Kapitel 9) widmet sich der Materialentwicklung und -erprobung für die wissenschaftlichen Prüfkriterien *Falsifizierbarkeit*, *Randomisierung*, *Korrelation und Kausalität* sowie *Reproduzierbarkeit* im Rahmen der Phase III des Unterrichtskonzepts.
Es wird untersucht, inwiefern Lernende in der Lage sind, die Prüfkriterien zu erarbeiten und anzuwenden. Dabei werden Reaktionen auf *Lehrimpulse* untersucht, *Verständnisschwierigkeiten* der Schülerinnen und Schüler aufgedeckt und das *Off-Task-Verhalten* analysiert. Außerdem liegt ein Schwerpunkt auf der Untersuchung von *Aha-Erlebnissen*, die die Lernenden bei der Bearbeitung der Materialien erfahren.

Im zweiten Mesozyklus (Kapitel 10) wird das Unterrichtskonzept, das bei Jungkamp (2021) zur Bewertung der Kontroverse *Homöopathie* verwendet wurde, auf einen neuen Kontext übertragen: Anhand einer Fragebogenstudie wird die Thematik der *Nahrungsergänzungsmittel* als geeigneter Kontext identifiziert.
Aufbauend auf die Studie werden Lernmaterialien zu Nahrungsergänzungsmitteln konzipiert und für die Verwendung in den Phasen I, II, IV und V von choice^{2}reflect aufbereitet. Mittels Erprobungen des Lernmaterials mit Schülerinnen und Schülern der Sekundarstufe I wird die Eignung der entwickelten Lerneinheiten überprüft. Weiterhin wird erhoben, inwiefern Lernende nach Abschluss der fünf Phasen rationale Entscheidungen bezüglich der Kontroverse treffen.

Der dritte Mesozyklus (Kapitel 11) behandelt die Frage, wie Lernende dabei unterstützt werden können, das methodische Wissen aus Phase III auf gesellschaftliche Kontroversen anzuwenden. Dabei werden Beziehungen der Prüfkriterien zueinander geklärt und in einem Prüfschema als methodisches Instrument verortet.

Erste Erprobungen mit Studierenden liefern Erkenntnisse über die Herangehensweise, die Lernende bei der Bewertung einer Kontroverse verfolgen: Viele der teilnehmenden Studierenden können auch ohne Instruktion in die Funktionsweise des Prüfschemas eine rationale Entscheidung zum Thema *Veganismus* treffen und nutzen das Schema in einer zielführenden Weise.

Das *Re-Framing* (Teil VI) zeigt den praktischen Output der Arbeit auf und grenzt diesen vom theoretischen Output ab.
Auf der praktischen Seite (Kapitel 12-14) stehen neu entwickelte und optimierte Lernmaterialien für das Unterrichtskonzept choice^{2}reflect . Vier Prüfkriterien wurden ausgearbeitet und im Unterrichtseinsatz erprobt. Dabei hat sich gezeigt, dass ein großer Teil der Lernziele in den Erprobungen adressiert wurde und oft auch erreicht werden konnte. An vielen Stellen erkennbare *Aha-Erlebnisse* legen das Erreichen der Lernziele nahe.
Aufgetretene Verständnisschwierigkeiten und Off-Task-Interaktionen der Lernenden wurden für ein Re-Design der Materialien verwendet.

Weiterhin wurde das Thema *Nahrungsergänzungsmittel* als kontroverser und für Lernende relevanter Kontext herausgestellt. Das heterogene Meinungsbild in Lerngruppen bei geringer Informationslage der Schülerinnen und Schüler kann als Ausgangspunkt für die Erarbeitung von Wissenschaftskriterien verwendet werden.

Das Prüfschema hat sich als intuitiv verständliches methodisches Instrument zur Anwendung der Kriterien erwiesen.

In der Theoriebildung (Kapitel 15) werden Gestaltungsmerkmale für alle entwickelten Materialien herausgearbeitet, die für den Lernerfolg verantwortlich sein könnten. Für die Erarbeitung von Wissenschaftskriterien hat es sich als zielführend erwiesen, Lernende auf intuitive Weise an die Prüfkriterien heranzuführen und sowohl Experimente als auch empirische Studien zu thematisieren.
Zentral für die intuitive Verständlichkeit des Prüfschemas könnten die verwendete Symbolsprache oder auch die kurzen Fragesätze sein.
Das Zusammenspiel aller Elemente der Unterrichtskonzeption choice^{2}reflect kann bei Lernenden ein Bewusstsein für Wissenschaftlichkeit schaffen und einen Beitrag dazu leisten, eine kritische Grundhaltung im Alltag zu etablieren und Bewertungskompetenz zu stärken.

Inhaltsverzeichnis

I Einleitung **1**

II Theoretischer Hintergrund **7**

1 Entscheidungstheorie **9**

1.1 Rationales Verhalten: Die normative Entscheidungstheorie 9

1.1.1 Entscheiden bei Sicherheit 10

1.1.2 Entscheiden bei Unsicherheit 11

1.2 Selektion von Entscheidungsregeln 11

1.3 Beschränkt rationales Verhalten: Die deskriptive Entscheidungstheorie 12

1.3.1 Abgrenzung zur normativen Entscheidungstheorie 12

1.3.2 Prospect-Theorie und Framing 13

1.3.3 Entscheiden nach Heuristiken 13

1.3.4 Die Rolle von Emotionen 14

1.3.5 Der Ansatz der Zwei Systeme 15

1.4 Fazit Entscheidungstheorie 16

1.5 Urteilen vs. Entscheiden 17

1.5.1 Urteilstypen 18

1.6 Entscheidungsprozessmodell 18

1.6.1 präselektionale Phase 19

1.6.2 postselektionale Phase 19

2 Wissenschaftstheorie **21**

2.1 wahrheitsfähige Aussagen 22

2.1.1 logische Aussagen 22

2.1.2 empirische Aussagen 22

2.2 nicht-wahrheitsfähige Aussagen 23
2.2.1 normative Aussagen 23
2.2.2 meta-physische Aussagen 24
2.3 Gewinnen wissenschaftlicher Erkenntnis anhand von Aussagen 24
2.3.1 Verallgemeinerungen: Der Induktivismus 25
2.3.2 Falsifikationismus 26
2.3.3 wissenschaftstheoretische Grundbegriffe: Von der These zur Theorie 28
2.3.4 Operationalisierung 30
2.4 Gütekriterien empirischer Untersuchungen 31
2.4.1 Validität, Reliabilität, Objektivität 32
2.4.2 Kontrolle von Störfaktoren 33
2.4.3 Interpretation und Dokumentation 35
2.4.4 Zusammenfassung der Gütekriterien 36

3 Bewertungskompetenz im naturwissenschaftlichen Unterricht 37
3.1 Definition von Bewertungskompetenz 37
3.2 Kompetenzstrukturmodelle 40
3.2.1 Modell der ethischen Urteilskompetenz 40
3.2.2 Göttinger Modell der Bewertungskompetenz 41
3.2.3 Modell für die Evaluation der Bildungsstandards: Das ESNaS-Modell 43
3.2.4 Modell der epistemischen Kompetenz 44
3.3 Einflussfaktoren auf Bewertungskompetenz 46
3.4 Bedeutung für den naturwissenschaftlichen Unterricht 48
3.4.1 Allgemeine Bedeutung 48
3.4.2 Bewertungskompetenz in den schulischen Lehrplänen 49
3.5 Bewerten als Ausgangspunkt für *conceptual change* 55
3.6 Bewerten unter Kontroversität: Socio-Scientific Issues 56
3.7 Empirische Befunde zum Einsatz von SSI 59
3.8 Fazit zur Bewertungskompetenz 60

4 Unterrichtskonzeption choice^{2}reflect 63
4.1 Idee 63

4.2 Die fünf Phasen der Konzeption 64
4.3 Wissenschaftliche Prüfkriterien 66
4.4 Prüfkarten 70

III Projektverlauf gemäß Design-Based Research 71

5 Der Design-Based Research-Ansatz 73
5.1 Entwicklung des DBR 74
5.2 Charakteristika des Design-Based Research-Ansatzes 75
5.3 Kritik am DBR 76
5.4 Ablauf design-basierter Forschungsprojekte 77
5.4.1 Framing 77
5.4.2 Design-Experiment 79
5.4.3 Re-Framing 81
5.5 Einordnung des eigenen Projektverlaufs 81

IV Framing 83

6 Forschungsergebnisse des Vorgänger-Projekts 87
6.1 Mesozyklus 1 87
6.2 Mesozyklus 2 88
6.3 Mesozyklus 3 88
6.4 Mesozyklus 4 88

7 Zielformulierung und Forschungsfragen 91
7.1 Forschungsfragen des Mesozyklus 1 98
7.2 Forschungsfragen des Mesozyklus 2 99
7.3 Forschungsfragen des Mesozyklus 3 100

V Design-Experiment 103

8 Übersicht der Erprobungen und Forschungsmethodik 107

8.1 Datenerhebung . . . 107
8.1.1 Erprobung des Prüfkriteriums Falsifizierbarkeit . . . 107
8.1.2 Fragebogenerhebung zu Nahrungsergänzungsmitteln . . . 110
8.1.3 Erprobung von Phase I und II . . . 111
8.1.4 Erprobung des Prüfkriteriums Reproduzierbarkeit . . . 111
8.1.5 Gesamterprobung . . . 112
8.1.6 Erprobung des Prüfschemas . . . 114
8.2 Datenauswertung . . . 115
8.2.1 Qualitative Inhaltsanalyse nach Kuckartz . . . 115
8.2.2 weitere Analysemethoden . . . 117

9 Mesozyklus 1: Erarbeitung und Anwendung von Prüfkriterien 119

9.1 MZ 1.1: Falsifizierbarkeit . . . 120
9.1.1 Entwicklung des Lernmaterials . . . 121
9.1.2 Erprobung . . . 126
9.1.3 Fazit und Ausblick . . . 131
9.2 MZ 1.2: Korrelation vs. Kausalität . . . 132
9.2.1 Theoretischer Hintergrund zu Korrelationen und Kausalitäten 133
9.2.2 Entwicklung des Lernmaterials . . . 135
9.2.3 Erprobung . . . 142
9.2.4 Auswertung . . . 157
9.2.5 Re-Design . . . 161
9.2.6 Fazit und Ausblick . . . 165
9.3 MZ 1.3: Randomisierung . . . 167
9.3.1 Theoretischer Hintergrund zu Randomisierung . . . 167
9.3.2 Entwicklung des Lernmaterials . . . 168
9.3.3 Erprobung . . . 177
9.3.4 Auswertung . . . 189
9.3.5 Re-Design . . . 194
9.3.6 Fazit und Ausblick . . . 198

9.4 MZ 1.4: Reproduzierbarkeit 199
9.4.1 Theoretischer Hintergrund zur Reproduzierbarkeit 199
9.4.2 Entwicklung des Lernmaterials 200
9.4.3 Erprobung 206
9.4.4 Diskussion und Re-Design 213

10 Mesozyklus 2: *Nahrungsergänzungsmittel* als Kontroverse 215
10.1 Bewertung nach Kriterien eines guten Kontextes 215
10.2 Fragebogenerhebung zu Nahrungsergänzungsmitteln 219
10.2.1 Entwicklung des Fragebogens 219
10.2.2 Ergebnisse 223
10.2.3 Diskussion 230
10.3 Fachspezifischer Problemaufriss und Positionierung 233
10.3.1 Entwicklung des Lernmaterials 233
10.3.2 Erste Erprobung des Lernmaterials 235
10.3.3 Diskussion und Re-Design 240
10.3.4 Zweite Erprobung des Unterrichtsmaterials 244
10.4 Anwendung und Reflexion 247
10.4.1 Ziele der Anwendungs- und Reflexionsphase 247
10.4.2 Entwicklung des Lernmaterials 248
10.4.3 Erprobung 252
10.4.4 Diskussion und Re-Design 256

11 Mesozyklus 3: Prüfschema zur Unterstützung der Prüfkriterien 259
11.1 Stellung der Prüfkriterien 259
11.1.1 Abgrenzung der Goldstandard-Kriterien von übergeordneten Prinzipien 260
11.1.2 Goldstandard-Kriterien 261
11.2 Explizieren der Verortung von Prüfkriterien im Unterricht: Ein Prüfschema 265
11.3 Entscheidungsfindung nach Absolvieren des Prüfschemas 269
11.4 Erprobung des Prüfschemas 271
11.4.1 Methodik und Aufgabenentwicklung 271
11.4.2 Entwicklung des Kategoriensystems 274

11.4.3 Ergebnisse 276
11.5 Analyse und Diskussion 288
11.5.1 Vorwissen zu den Prüfkriterien 288
11.5.2 Umgang mit dem Prüfschema 290
11.5.3 Beantworten der Forschungsfrage 291
11.6 Re-Design 291

VI Re-Framing und Reflexion 295

12 Praktischer Output des ersten Mesozyklus 299
12.1 Falsifizierbarkeit 299
12.2 Korrelation und Kausalität 300
12.3 Randomisierung 301
12.4 Reproduzierbarkeit 302

13 Praktischer Output des zweiten Mesozyklus 305
13.1 Kriterien eines guten Kontexts 305
13.2 Validierung des Themas *Nahrungsergänzungsmittel* als guter Kontext . 305
13.3 Entwicklung und Erprobung der Phasen I und II 306
13.4 Entwicklung und Erprobung der Phasen IV und V 308

14 Praktischer Output des dritten Mesozyklus 311

15 Theoriebildung 313
15.1 Ausgangstheorien des Vorgängerprojektes 314
15.2 Gestaltungsmerkmale für Prüfkriterien 314
15.3 Präsentation eines kontroversen Kontexts in choice^{2}reflect 317
15.4 Gestaltungsmerkmale eines Prüfschemas 319

16 Fazit und Abschlussreflexion 323

VII Literatur 327

VIII Anhang **345**

A Theoretischer Hintergrund **347**

B Design-Experiment **351**

B.1 Kategoriensysteme 351
B.2 Impulse in Plenumsphasen 372
B.3 Codelines 382
B.4 Einzelaussagen 387
B.5 Arbeitsblätter 389
B.5.1 Phase I und II 389
B.5.2 Phase III 392
B.5.3 Phase IV und V 410
B.6 Prüfkarten 414
B.7 Stundenverlaufspläne 415
B.8 Fragebogen 421

Abbildungsverzeichnis **421**

Tabellenverzeichnis **440**

Danksagung **443**

Teil I

Einleitung

Einleitung

Wissenschaftliche Allgemeinbildung ist ein Impfstoff gegen Desinformation.

Nguyen-Kim (2021, S. 338)

Aktuelle gesellschaftliche Debatten werden nicht ausschließlich auf der Basis wissenschaftlich gesicherter Fakten geführt. *Fake-News*, *Verschwörungstheorien* und *Desinformation* nehmen immer mehr Raum ein (Cook et al. 2018). Teilweise dominieren sie öffentliche Diskussionen über Themen wie den Klimawandel, die Wirksamkeit von Homöopathika oder auch die Verbreitung des Corona-Virus *SARS-CoV-2* (Naeem et al. 2021; Mäntele 2021).

Die von Nguyen-Kim (2021) angeführte Metapher des „Impfstoffes" verdeutlicht die Versuche der Wissenschaft, die Verbreitung von Desinformation mit allen Mitteln einzudämmen. Gerade in sozialen Medien scheint eben diese jedoch besonders *ansteckend* zu sein: Mit Berufung auf die „Meinungsfreiheit" werden alternative unwissenschaftliche Fakten gepostet, geteilt, weiterverbreitet.

Eine Schülerin äußerte im Laufe dieses Projekts während einer Unterrichtsstunde im Kontext *Nahrungsergänzungsmittel*: „Auf Instagram, da kann ja jeder schreiben, was er möchte. Er muss ja nicht irgendwie Wissenschaftler sein oder sich besonders gut damit auskennen".

Und selbst *wenn* sich eine Person gut mit einem Thema auskennt, besondere Expertise besitzt oder eine Wissenschaftlerin ist, garantiert dies nicht für eine hohe wissenschaftliche Qualität der Äußerungen dieser Person.

Wie können wir aus der uns täglich begegnenden Informationsflut herausfiltern, ob eine Information *wahr* ist oder nicht? Die kurze Antwort lautet: Wir können es nicht. Die endgültige Wahrheit einer Hypothese lässt sich anhand von induktiven Methoden, also wissenschaftlichen Beobachtungen und Verallgemeinerungen, nicht feststellen (Popper 1989; Hume 2019, original: 1748, Section IV; Baumann 2015, S. 257).

Wissenschaftlerinnen und Wissenschaftler „irren sich empor“ (Vollmer 1995; Nguyen-Kim 2021; Mittelstraß 1997), sie widerlegen bestehende Hypothesen und stellen neue auf, die weiteren Falsifizierungs-Versuchen standhalten sollen. Gelangen viele Forschergruppen zu den gleichen Erkenntnissen, bildet sich ein *wissenschaftlicher Konsens* heraus (Bortz & Döring 2006, S. 41).

So lautet die ausführliche - und weniger ernüchternde - Antwort auf die obenstehende Frage: Aus der großen Informationsflut lassen sich mithilfe von *wissenschaftlicher Methodik* Falschmeldungen („Fake-News“) aussortieren und gut belegte Informationen identifizieren.
Letztere können dann verwendet werden, um in einem kontrovers diskutierten Thema zu einer rationalen Entscheidung zu gelangen.
Das dazu notwendige Wissen über wissenschaftliche Methodik wird im Eingangszitat als *Wissenschaftliche Allgemeinbildung* bezeichnet und gefordert.

Eine naturwissenschaftliche Grundbildung wird auch in den schulischen Lehrplänen gefordert, zum Beispiel in den Bildungsstandards im Fach Chemie (KMK 2005a, S. 6):

> *Naturwissenschaftliche Bildung ermöglicht dem Individuum eine aktive Teilhabe an gesellschaftlicher Kommunikation und Meinungsbildung über technische Entwicklungen und naturwissenschaftliche Forschung und ist deshalb wesentlicher Bestandteil von Allgemeinbildung.*

Mit welchen Mitteln diese Grundbildung, häufig auch *scientific literacy* genannt, im Unterricht erreicht werden soll, wird weder in den Bildungsstandards noch in fachspezifischen Kernlehrplänen konkreter ausgestaltet. Zwar schlüsseln die Kernlehrpläne vier verschiedene Kompetenzbereiche auf und spezifizieren übergeordnete Kompetenzerwartungen. Es gibt jedoch kaum Anweisungen für die konkrete unterrichtliche Umsetzung.
Benötigt werden Unterrichtskonzepte, die Schülerinnen und Schüler beim Treffen rationaler Entscheidungen in gesellschaftlichen Kontroversen unterstützen.
Dieser Forderung begegnen Jungkamp und Marohn mit der Unterrichtskonzeption choice^{2}reflect (Marohn & Jungkamp 2016; Jungkamp 2021). In fünf Phasen positionieren sich Lernende zu einer Kontroverse, erarbeiten wissenschaftliche Prüfkriterien und wenden diese auf Aussagen und Studien zur Bewertung der Kontroverse an.

Dabei nehmen sie die Methoden der untersuchten Studien genau in den Blick und überprüfen beispielsweise, ob die Stichproben wissenschaftlich korrekt ausgewählt wurden oder ob unter kontrollierten und verblindeten Bedingungen gearbeitet wurde.
Jungkamp (2021) entwickelte und erprobte im Rahmen des Design-Based Research-Ansatzes das Konzept im praktischen Unterrichtseinsatz, wobei die Lernenden das Thema *Homöopathie* bewerteten. Er gelangte u. a. zu der Erkenntnis, dass die Erarbeitung von Wissenschaftskriterien zu einem schrittweisen Aufbau des Wissenschaftsverständnisses von Lernenden im Sinne eines *Cluster Approaches* beiträgt und dass die Konzeption zur Bewertung von Kontroversen geeignet ist.

Das in dieser Arbeit beschriebene Projekt verläuft ebenso gemäß dem Design-Based Research-Ansatz und knüpft an Jungkamps Erkenntnisse an. Forschungsdesiderate werden aufgegriffen und zu einer übergreifenden Forschungsfrage zusammengeführt (Teil IV). Diese gliedert sich in unterschiedliche Teilbereiche, die im Rahmen von drei Mesozyklen erforscht werden:
Im ersten Mesozyklus (Kapitel 9) werden Unterrichtsmaterialien zu den Wissenschaftskriterien (weiter)entwickelt und erprobt.
Der zweite Mesozyklus (Kapitel 10) befasst sich mit der Einführung von einer neuen Kontroverse, die als Kontext in choice^{2}reflect verwendet werden kann: Das Thema *Nahrungsergänzungsmittel* wird für den Einsatz in der Konzeption aufbereitet.
Im Rahmen des dritten Mesozyklus (Kapitel 11) wird ein Prüfschema entwickelt, das als methodisches Instrument zur Anwendung der Wissenschaftskriterien dient und direkt in der vierten und fünften Phase eingesetzt werden kann.

Alle Unterrichtsmaterialien werden für den direkten unterrichtlichen Einsatz konzipiert und auch in diesem erprobt: Choice^{2}reflect wird in realen Bildungskontexten eingesetzt. Die Einsätze werden forschungsmethodisch begleitet und gewonnene Daten auf qualitativ-inhaltsanalytischer Basis ausgewertet.
Die gewonnenen Erkenntnisse fließen in eine direkte (iterative) Weiterentwicklung des Lernmaterials ein. Sie bilden neben dem praktischen Output dieses Projekts auch Ansatzpunkte für eine übergreifende Theoriebildung (Teil VI).
Dieses Projekt bringt die Etablierung der Unterrichtskonzeption choice^{2}reflect als praxistaugliche Intervention zur Schulung von Bewertungskompetenz im naturwissenschaftlichen Unterricht weiter voran.

Teil II

Theoretischer Hintergrund

1 Entscheidungstheorie

Im Rahmen dieses Projekts wird ein Unterrichtskonzept weiterentwickelt, das Lernende im Bezug auf gesellschaftlich kontrovers diskutierte Themen zu einer rationalen und sachlich begründeten Bewertung befähigen soll.
Das Ziel der Einheit besteht also darin, dass Schülerinnen und Schüler auf der Basis wissenschaftlich gesicherter Fakten einen eigenen Standpunkt gegenüber einer Kontroverse entwickeln. Um das zu ermöglichen, gilt es zunächst zwei Fragen zu klären: Welche Anforderungen müssen Informationen erfüllen, um als *wissenschaftlich gesicherte Fakten* angesehen werden zu können? Und auf welche Art und Weise treffen Individuen Entscheidungen zur Bildung eines eigenen Standpunktes?
Beide Fragen werden in den folgenden Kapiteln behandelt. Zunächst wird in diesem Kapitel ein umfangreicher Überblick über das Entscheidungsverhalten von Individuen gegeben: Die Entscheidungstheorie. Anschließend (Kapitel 2) werden wissenschaftstheoretische Überlegungen zur Güte von Aussagen und empirischen Untersuchungen angestellt.

1.1 Rationales Verhalten: Die normative Entscheidungstheorie

Die normative (bzw. präskriptive) Entscheidungstheorie verfolgt eine deduktive Herangehensweise an die Theorie über das Treffen von Entscheidungen. Sie nimmt an, dass Entscheidungen nach rationalen Mustern gefällt werden und geht von axiomatischer Gültigkeit bestimmter Annahmen aus. Abhängig von der Art des Informationsangebots werden unterschiedliche Prinzipien der Entscheidungsfindung beschrieben. Dabei wird unterschieden, wie zuverlässig und sicher die Konsequenzen der einzelnen Handlungsoptionen vorhergesagt werden können bzw. ob diese überhaupt bekannt sind (Pfister et al. 2017).

1.1.1 Entscheiden bei Sicherheit

Bei klarer und detailreicher Informationslage kann eine „Entscheidung unter Sicherheit“ vorgenommen werden. Sind alle Handlungsoptionen und alle damit verbundenen Ausprägungen der Attribute (also Konsequenzen) bekannt, kann das Entscheidungsverhalten durch kompensatorische und non-kompensatorische Entscheidungsregeln beschrieben werden. Bei kompensatorischen Regeln kann eine „gute“ Ausprägung eines Attributs die „schlechte“ Ausprägung eines anderen Attributs ausgleichen, während dies nach non-kompensatorischen Entscheidungsregeln nicht möglich ist (Betsch et al. 2011, S. 97).

Eine bekannte kompensatorische Entscheidungsregel ist das additive multiattribute Nutzenmodell (MAU-Theorie) (Pfister et al. 2017, S. 100). Alle Attribute werden gewichtet und aufaddiert, sodass der Gesamtnutzen einer Option berechnet werden kann. Dafür müssen alle Attribute gleichwertig und präferentiell voneinander unabhängig sein. Kompensatorische Entscheidungen sind häufig mit Kompromissen verbunden und erfordern Abwägungen anhand evaluativer Urteile (vgl. 1.5.1), daher tendieren Entscheider dazu, derartige Entscheidungen zu vermeiden oder aufzuschieben (Anderson 2003).

Non-kompensatorische Regeln sind z. B. die lexikografische Regel (LEX-Regel, Attribute werden nach ihrer Wichtigkeit geordnet) oder die „Elimination by Aspects“-Regel (EBA-Regel, Aspekte werden nach ihrer Wahrscheinlichkeit selektiert und unpassende Optionen eliminiert; Tversky 1972; Baron 2000). Diese Regeln lassen sich gut anhand von Kaufentscheidungen veranschaulichen. Soll beispielsweise eine neue Gitarre angeschafft werden, lassen sich verschiedene Attribute gegeneinander abwägen. Ein Modell ist in einer unerwünschten Farbe lackiert, hat jedoch aufgrund der Stahlsaiten einen als schöner empfundenen Klang als ein vergleichbares Modell mit Saiten aus Nylon. Das Attribut Klang überwiegt und kann die unpassende Farbe kompensieren, somit kommt es zu einer kompensatorisch getroffenen Kaufentscheidung. Non-kompensatorisch wird jedoch entschieden, wenn der Entscheider als zwingendes Attribut bspw. das Vorhandensein eines integrierten Stimmgerätes voraussetzt. Besitzt ein Modell kein Stimmgerät, dann kann kein anderes Attribut den Mangel ausgleichen, sodass kein Kauf zustande kommt.

1.1.2 Entscheiden bei Unsicherheit

Entscheidungsregeln wie die MAU-Theorie oder die LEX-Regel finden nur Anwendung, wenn alle Handlungskonsequenzen mit Sicherheit vorhergesagt werden können bzw. wenn diese feststehen. Fehlt diese Sicherheit, ist es nicht möglich, den Gesamtnutzen jeder Handlungsoption zu berechnen. Die normative Entscheidungsforschung spricht in diesem Fall von Entscheiden unter Unsicherheit (Pfister et al. 2017, S. 171).
Anstelle der Berechnung des Gesamtwerts unter Berücksichtigung multipler Attribute (MAU-Theorie) wird der subjektiv erwartete Nutzen bestimmt (SEU-Theorie).
Die SEU-Theorie (Edwards 1954) geht von bekannten Wahrscheinlichkeiten für unterschiedliche Handlungskonsequenzen aus, integriert diese zu einem SEU-Wert für jede Handlungsoption und bestimmt einen Gesamtwert, den „Subjektiv Erwarteten Nutzen“. Dieses Prinzip wird häufig am Beispiel von Lotterien verdeutlicht. Entweder man gewinnt oder verliert, jeweils mit einer bestimmten Wahrscheinlichkeit. Gewählt wird die Handlungsoption, die den höchsten erwarteten Gewinn verspricht, also deren SEU-Wert am größten ist.

1.2 Selektion von Entscheidungsregeln

Welche der beschriebenen Regeln letztendlich zur Entscheidungsfindung verwendet wird (bewusst oder unbewusst), ist neben der Sicherheit der Handlungskonsequenzen von weiteren Bedingungen abhängig. Dazu gehören die Komplexität des Problems (Pfister et al. 2017, 237ff) und die Art des Informationsangebots (Konkretheit, Übersichtlichkeit, Vollständigkeit, Darbietungsformat). Den gegebenen Bedingungen wird eine passende Entscheidungsregel zugeordnet. Welche Regel geeignet ist, entscheidet sich anhand der Merkmale, die die einzelnen Regeln unterscheiden:

- kompensatorisch oder non-kompensatorisch
  Entscheidungsregeln können kompensatorisch oder non-kompensatorisch sein, wie im Kapitel 1.1.1 beschrieben
- Verarbeitung von vorliegenden Informationen
  Die zu einem Thema vorhandenen Informationen können vollständig oder nur partiell zur Entscheidungsfindung verwendet werden

- Integration von Informationen
  Es gibt Entscheidungsregeln, die für eine Option alle Informationen zu einem Gesamtwert integrieren, während andere Regeln ohne Gesamtwert zu einer Entscheidung führen
- quantitative oder qualitative Bewertungen

1.3 Beschränkt rationales Verhalten: Die deskriptive Entscheidungstheorie

1.3.1 Abgrenzung zur normativen Entscheidungstheorie

Die verschiedenen Axiome rationalen Entscheidens wurden durch empirische Entscheidungsforschung teilweise widerlegt bzw. wurde gezeigt, dass die Axiome nicht in allen Situationen gültig sind (Pfister et al. 2017, S. 180). Daher beschäftigt sich die deskriptive Entscheidungstheorie damit, wie Entscheidungen tatsächlich unter realen Bedingungen getroffen werden. Es wird u. a. untersucht, in welcher Weise Individuen Wahrscheinlichkeitsurteile über ungewisse Ereignisse bilden, wie Ziele gebildet werden und sich im Laufe des Entscheidungsprozesses verändern und wie Informationsbeschaffungs- und -verarbeitungsprozesse ablaufen (Laux et al. 2018, S. 17).
Wird die Entscheidung getroffen, zu einem Problem relevante Informationen zu beschaffen und in die Problemlösung mit einzubeziehen, folgt nicht zwangsläufig eine rationale Entscheidung im Sinne normativ-entscheidungstheoretischer Ansätze: Die zentrale Annahme der deskriptiven Entscheidungstheorie stellt die begrenzte Informationsverarbeitungskapazität von Individuen dar (Langer & Rogowski 2009). Das reale Entscheidungsverhalten weicht von der optimal-rationalen Vorgehensweise ab, sodass Individuen teils irrationale Entscheidungen treffen.
In der Literatur werden vier Arten begrenzt rationalen Entscheidens unterschieden (Bea & Göbel 2002; Langer & Rogowski 2009):

- Irrtum (subjektiv rational, aber objektiv falsch),
- Eigennutz (persönlich rational, aber organisatorisch falsch),

- Identifikation mit Teilbereichen (lokal rational, aber in größerem Zusammenhang falsch),
- begrenzte Voraussicht (kurzfristig rational, aber langfristig falsch).

1.3.2 Prospect-Theorie und Framing

Weitere theoretische Ansätze zur Erforschung von realem Entscheidungsverhalten stellen Kahneman & Tversky (1979) mit der Prospect-Theorie und dem Konzept des Framing (Tversky & Kahneman 1981) bereit. Laut der Prospect-Theorie entstehen durch Editiermechanismen vereinfachte mentale Repräsentationen von objektiv gegebenen Informationen. Ähnlich zur SEU-Theorie wird anschließend eine Wertfunktion (entspricht in etwa der Nutzenfunktion in der SEU-Theorie) gebildet, die jedoch Referenzpunkt-basiert ist und durch weitere Funktionen charakterisiert wird. Eine umfassende Darstellung findet sich u. a. bei Langer & Rogowski (2009) und auch bei Kahneman & Tversky (1979).
Framing-Effekte bezeichnen Präferenzänderungen durch unterschiedliche Beschreibungen desselben Sachverhaltes. Je nach Ausdrucksweise der Handlungskonsequenzen lässt sich das Entscheidungsverhalten von Menschen beeinflussen, wie an folgendem Beispiel verdeutlicht wird. Beim Online-Shopping auf einer Seite für Nahrungsergänzungsmittel findet sich ein Präparat, das sowohl geliefert werden kann, als auch bei einem lokalen Händler in der Nähe verfügbar ist. Die Beschreibung: „Sparen Sie 3 € bei Abholung im Markt“ hat eine andere Wirkung als: „Sparen Sie sich den Weg zu uns - Für nur 3 € liefern wir Ihnen die Ware nach Hause“, obwohl dieselbe Botschaft vermittelt wird.

1.3.3 Entscheiden nach Heuristiken

Bei vielen Entscheidungen spielen klassische Heuristiken eine Rolle. Übersteigt der zur Informationsbeschaffung und -verarbeitung nötige Aufwand die kognitive Kapazität des Entscheidenden, wird nach Entscheidungsheuristiken vorgegangen. Diese werden unter Verwendung von sehr wenigen Informationen schnell und einfach getroffen („Fast and Frugal“,Gigerenzer & Gaissmaier 2011). Die Ergebnisse dieser Entscheidungen sind häufig akzeptabel, allerdings entstehen leicht systematische Fehlurteile (Pfister et al. 2017; Kahneman et al. 1982). Bekannte Heuristiken sind die Re-

präsentativitätsheuristik, Verfügbarkeitsheuristik, Verankerungs- und Anpassungsheuristik. Für eine ausführliche Darstellung der „Heuristics and Biases“ siehe Kahneman et al. (1982)und Kahneman (2011).
Ein weiterer Aspekt, der die Entscheidungsfindung beeinflusst, ist die Empfindung von Emotionen.

1.3.4 Die Rolle von Emotionen

Entscheidungen werden immer auch anhand von Emotionen gefällt. Das ist nicht in jedem Fall störend oder schädlich für einen Entscheidungsprozess, denn in unterschiedlichen Entscheidungssituationen haben Emotionen auch unterschiedliche Funktionen.
Emotionen haben u.a. die Funktion, eine Person darüber zu informieren, wie eine Beziehung zu einer bestimmten Sache in der Welt zu deuten ist: Etwas wird als angenehm oder unangenehm empfunden. Evolutionär gesehen waren diese Empfindungen in der Vergangenheit von Vorteil, so lösen beispielsweise der Geruch und Geschmack verdorbener Lebensmittel eine starke Abneigung aus und verhindern somit Lebensmittelintoxikationen. In der modernen Gesellschaft ist Entscheidungsfindung jedoch weitaus vielschichtiger geworden, sodass richtige von falschen Entscheidungen nicht mehr anhand einfacher Emotionen voneinander abgegrenzt werden können.
Emotionen können über die Vielschichtigkeit komplexer Situationen hinweg täuschen und damit zu Fehleinschätzungen führen. Außerdem sind sie gegenwartsgebunden und beziehen sich meist nur auf kurzfristige Situationen, sodass emotional gefällte Entscheidungen langfristig nicht immer die beste Option sind. (Pfister et al. 2017)
Der Einfluss inzidenteller, also während der Entscheidung auftretender Emotionen ist gut untersucht und wird durch das Appraisal-Tendency-Framework (ATF) von Lerner & Keltner (2000) beschrieben: Jede Emotion lässt sich durch ein bestimmtes Appraisal-Profil mit unterschiedlichen Ausprägungen der Dimensionen Sicherheit, Kontrolle, Anstrengung und Verantwortlichkeit charakterisieren.
Dadurch wirken Emotionen, die zwar dieselbe Grundstimmung besitzen, jedoch durch unterschiedliche Appraisal-Profile gekennzeichnet sind, in unterschiedlicher Weise auf entscheidungsrelevante Bewertungen. Die Theorie lässt sich anhand der Emotionen ärgerlich und furchterregend verdeutlichen. Beide Emotionen haben eine negati-

ve Grundstimmung. Das Appraisal-Profil der Emotion „ärgerlich“ zeichnet sich durch eine hohe Sicherheit und Kontrollierbarkeit aus, sie ist mit hoher Anstrengung verbunden und wird von einer anderen Person ausgelöst.
Die konkrete Emotion „furchterregend“ ist hingegen unsicher und unkontrollierbar. Aufgrund dieser Unterschiede wirken sich die Emotionen unterschiedlich auf Bewertungsurteile aus.
Auch andere, nicht mit der inzidentellen Emotion in Zusammenhang stehende Ereignisse, werden von den Appraisal-Profilen beeinflusst - es entsteht ein „carry-over-Effekt“. Deshalb führen auftretende Emotionen jeder Art dazu, dass Entscheidungen von ihnen beeinflusst werden (Lerner et al. 2004).
Die Theorie des ATF ist in der Literatur gut bestätigt (Lerner & Tiedens 2006; Han et al. 2007; Cavanaugh et al. 2007; Lerner et al. 2015; Pfister et al. 2017). Zur Frage, welche Einflüsse Emotionen auf die Entscheidungsfindung nehmen, existieren weitere umfangreiche Forschungsansätze. Als grundsätzliche Kernaussage lässt sich jedoch formulieren, dass inzidentelle Emotionen stets Einfluss auf Entscheidungen und Bewertungsprozesse nehmen. Möglichkeiten, diese Einflüsse zu minimieren, bestehen u. a. in der Eliminierung des Auslösers einer Emotion sowie Bewusstmachen der Emotionen und Entscheidungsprozesse (Lerner & Keltner 2000).

1.3.5 Der Ansatz der Zwei Systeme

Einer der einflussreichsten Ansätze der Entscheidungspsychologie ist der System 1 und System 2-Ansatz nach Kahneman (2011). Er postuliert ein Kontinuum, auf dem kognitive und affektive Prozesse zwischen automatisch-unbewusst bis kontrolliert-bewusst einzuordnen sind.
System 1 bezeichnet kognitive und emotionale Prozesse, die unwillkürlich, ohne Anstrengung und automatisiert ablaufen. Dazu gehören beispielsweise Sinneswahrnehmung, Sprachverstehen der Muttersprache oder Wiederkennen von Gesichtern. System 2 stellt kontrollierte, bewusste Prozesse mit verbundener kognitiver Anstrengung dar, wie z. B. die Anwendung logischer Regeln oder multiattribute Präferenzen mit Kompromissentscheidungen (Pfister et al. 2017).
System 1 ist der Standard-Modus des kognitiven Systems, während System 2 in Ausnahmefällen bzw. bei Schwierigkeiten interveniert. Nach dem Prinzip des geringsten kognitiven Aufwands ist das zweite System jedoch „faul“ (Kahneman 2011; Pfister

et al. 2017) und interveniert nicht, wenn System 1 bereits eine zufriedenstellende Lösung erreicht hat.
Im Rahmen des Zwei-System-Ansatzes lassen sich Heuristiken (Kapitel 1.3) und Intuitionen verorten. Heuristiken im System 1 liefern zwar häufig brauchbare Ergebnisse, jedoch ist das erste System fehleranfällig, wie Kahneman (2011) zeigt: Es arbeitet nach den Prinzipien der a) Assoziativen Aktivierung, b) Kohärenz und c) Kausalität.

a) Wird ein bestimmter Stimulus wahrgenommen, so werden automatisch alle mit diesem Stimulus im Gedächtnis assoziierten Elemente aktiviert. Diese Ausbreitung der Aktivierung folgt dabei keinen logischen Regeln.

b) Wahrgenommene Stimuli und damit aktivierte Assoziationen werden als einzige und erschöpfende Informationen wahrgenommen, sodass kohärente Bilder ohne Alternativen entstehen: Trifft man bspw. einen Apotheker, der den Einsatz homöopathischer Arzneimittel glorifiziert, neigt das kognitive System dazu, alle Apotheker:innen als Befürworter der Homöopathie zu generalisieren.

c) Das System 1 unterstellt dynamischen Beziehungen grundsätzlich, sie seien kausal und intendiert.

System 2 könnte diesen Fehlern entgegenwirken, wird jedoch nur unter erhöhtem kognitiven Aufwand aktiviert.

1.4 Fazit Entscheidungstheorie

Die in diesem Kapitel dargestellten Theorien zeigen auf, dass Entscheidungssituationen bei Individuen nach verschiedenen Regeln und Bedingungen ablaufen. Die normative Entscheidungstheorie hat Theorien über Muster und Axiome rationaler Entscheidungen aufgestellt, die bewertungsrelevante Entscheidungen vorhersagen sollen. Sie differenziert zwischen Entscheidungen unter Sicherheit und Entscheidungen unter Unsicherheit, während letztere durch bestimmte Heuristiken charakterisiert sind.
Die deskriptive Entscheidungstheorie zeigt die begrenzte Rationalität realer Entscheidungen auf, indem sie anhand empirischer Forschungsergebnisse die Allgemeingültigkeit der normativen Axiome infrage stellt. Sie identifiziert die begrenzte Kapazität

bei der Informationsverarbeitung als zentrale Ursache dafür, dass Individuen oft irrational handeln und stellt die besondere Rolle der Emotionen im Entscheidungsprozess heraus.

Der Ansatz der Zwei Systeme nach Kahneman (2011) hat weitere Mechanismen aufgeklärt, die die Fehleranfälligkeit von Entscheidungsheuristiken und Intuitionen als dem System 1 zugehörige Prinzipien untermauern. Das System 2 ist fähig, diese fehlerhaften Bewertungstendenzen auszugleichen - dafür müssen Individuen jedoch einen hohen kognitiven Aufwand aufbringen.

Individuen im Umgang mit den Unterschieden zwischen System 1 und 2 zu schulen, ihnen die Bedeutung von Emotionen, Intuitionen und Entscheidungsheuristiken zu verdeutlichen und bewusst zu machen, wird durch das Konstrukt der Bewertungskompetenz aufgegriffen. Im Kapitel 3 wird Bewertungskompetenz charakterisiert, verschiedene Kompetenzstrukturmodelle werden dargelegt und die Bedeutung im naturwissenschaftlichen Unterricht aufgezeigt.

1.5 Urteilen vs. Entscheiden

Urteilen wird definiert als „Psychologischer Prozess, der zugrunde liegt, wenn Menschen einem Urteilsobjekt einen Wert auf einer Urteilsdimension zuordnen und das daraus resultierende Urteil explizit zum Ausdruck bringen“ (Betsch et al. 2011, S. 12).

Urteilsobjekte können u. a. Personen, Objekte, Aussagen, innere Zustände und Situationen sein, dabei kann die Skalierung von Urteilsdimensionen sehr unterschiedlich aussehen. Im Alltag werden nur selten konkrete Skalen verwendet (z. B. *Homöopathie kann schon ganz hilfreich sein*).

Auf Urteile folgen häufig Handlungskonsequenzen, diese sind jedoch weniger Teil der psychologischen *Urteils*forschung. Vielmehr werden Urteile als Endprodukte psychologischer Prozesse angesehen. Wie Individuen dann mit gefällten Urteilen umgehen und welche Handlungskonsequenzen sie aus ihnen ableiten, ist Teil der *Entscheidungs*forschung. Besteht eine Lücke in einem Handlungsplan, so wird diese durch Problemlösen geschlossen.

1.5.1 Urteilstypen

Betsch et al. (2011) unterscheiden vier Arten von Urteilen, dabei werden Urteile je nach Struktur der Urteilsaufgabe entweder deduktiv oder induktiv gebildet.

evaluative Urteile

Evaluative Urteile ordnen einem zu beurteilenden Objekt eine Valenz zu. Auf einer bewertenden Skala, z. B. positiv-negativ oder gut-schlecht drücken Individuen ihre Einstellungen zu einem Urteilsobjekt aus. In der Entscheidungsforschung werden anhand evaluativer Urteile Präferenzurteile gefällt.

prädiktive Urteile

Prädiktive Urteile dienen der Einschätzung der Wahrscheinlichkeit, mit der bestimmte Ereignisse eintreten. Sie sind bei Entscheidungen unter Unsicherheit von besonderer Relevanz (vgl. 1.1.2).

Wahrheitsurteile

Wahrheitsurteile treffen Aussagen über die Korrektheit von logischen Schlüssen. Sie werden auf der Grundlage klassischer Syllogismen gefällt und ergeben sich durch die Untersuchung, ob eine Konklusion unter bestimmten Prämissen gültig ist.

soziale Urteile

Bei sozialen Urteilen handelt es sich um Urteile über Personen (Eiser 1990; Betsch et al. 2011). In dieser Arbeit werden soziale Urteile nicht näher betrachtet.

1.6 Entscheidungsprozessmodell

Die bisher dargelegten Theorien zu Entscheidungen behandeln den Moment, in dem eine Entscheidung getroffen wird: Die *selektionale Phase*. In einem Prozessmodell von Betsch & Haberstroh (2014) & Betsch et al. (2011) wird geklärt, welche Phasen

für einen Entscheidungsprozess charakteristisch sind. Dabei werden auch die *präselektionale* und *postselektionale Phase* in den Blick genommen, die den Prozess vor und unmittelbar nach Entscheidungen beschreiben.

1.6.1 präselektionale Phase

Bevor eine Entscheidung getroffen wird, findet eine Reihe entscheidungsvorbereitender Schritte statt (Laux et al. 2018), die als präselektionale Phase zusammengefasst werden. Zunächst wird vom Entscheidungsträger eine unbefriedigende Situation erkannt und als Problem formuliert. Die anschließende Präzisierung des Zielsystems gibt der Erforschung von Handlungsalternativen eine Richtung, sie liefert den Beurteilungsmaßstab für die Entscheidung. Auf Grundlage der Spezifika der Problemsituation sowie des Zielsystems werden Handlungsalternativen erforscht und ausdifferenziert, außerdem werden mögliche Handlungskonsequenzen prognostiziert.
Unter Berücksichtigung der in der präselektionalen Phase gewonnenen Erkenntnisse wird in der selektionalen Phase die beste Handlungsalternative ausgewählt und die eigentliche Entscheidung wird getroffen.

1.6.2 postselektionale Phase

Mit der Wahl der besten Alternative ist die Entscheidung keineswegs abgeschlossen. Postselektional wird die gewählte Alternative realisiert. Aus den Rückmeldungen der Umwelt wird ersichtlich, ob die prognostizierten Konsequenzen eintreten oder nicht. Die Erfahrung mit Handlungskonsequenzen nimmt einen Einfluss auf zukünftige Entscheidungen, die bereits getroffenen Entscheidungen ggf. ähnlich sind (Betsch et al. 2011; Laux et al. 2018). Das bedeutet, dass Entscheidungen nicht rein beschreibungsbasiert sind, sondern auch auf bereits gesammelte Erfahrungen gestützt werden. Besonders Alltagsentscheidungen sind erfahrungsbasiert (Betsch et al. 2011, S. 110).
Im Anschluss an Entscheidungen treten häufig kognitive Dissonanzeffekte auf, die eine Unzufriedenheit mit der getroffenen Entscheidung hervorrufen können. Diese können z. B. durch das Korrigieren der Entscheidung oder nachträgliche Aufwertung der gewählten Option abgebaut werden (Betsch et al. 2011, S. 118). In der postselektionalen Phase findet auch die Ausbildung von Routinen statt, die direkten Einfluss auf zukünftige Entscheidungen und den Prozess der Informationsbeschaffung neh-

men können. Routinierte Entscheider:innen können sich bei der Informationssuche für zukünftige Entscheidungen einerseits besser auf die wichtigen Aspekte fokussieren (Ross et al. 2005), andererseits führt Routinisierung in manchen Fällen zu einseitiger Betrachtung von Informationen und zum Auftreten von Bestätigungsfehlern (Betsch et al. 2001; Betsch et al. 2011).

2 Wissenschaftstheorie

Menschen fällen Urteile und Entscheidungen nur begrenzt rational, wie im Kapitel 1 beschrieben wird. Entscheidungsheuristiken, Emotionen und Intuition ermöglichen schnelle und einfache Entscheidungen, sind jedoch bei komplexeren Problemsituationen fehleranfällig. Die Annäherung an rationale Urteils- und Entscheidungsfindung zeigt auf, dass eine umfassende Informationsbeschaffung und -verarbeitung in der präselektionalen Phase des Entscheidungsprozesses stattfinden muss. Nur durch Integration der wissenschaftlichen Datenlage zum Thema sind rationale Entscheidungen möglich.

Die wissenschaftliche Datenlage ist bei alltäglichen Entscheidungen häufig klar: Ein rohes Ei lagert man besser in der Eierpappe als auf dem leicht schrägen Regalboden, da Gravitationskräfte ansonsten dafür sorgen würden, dass das Ei zu Boden rollt. Komplexe, kontrovers diskutierte Probleme sind jedoch vielfältig und die Faktenlage ist nicht immer eindeutig. Verschiedene Standpunkte werden durch verschiedene Argumente gestützt und begründet, die häufig nicht im Einklang miteinander stehen und sich teilweise widersprechen.

Komplexe Themen erfordern daher, zusätzlich zu der Informationsbeschaffung und -verarbeitung, eine Bewertung der vorliegenden Informationen. Der kognitive Aufwand zum Treffen einer rationalen Entscheidung wird dadurch weiter erhöht. Bevor im späteren Kapitel dieser Arbeit (Kapitel 3) die Bedeutung von Bewertungskompetenz für den naturwissenschaftlichen Unterricht aus verschiedenen Perspektive beleuchtet wird, ist es zunächst wichtig, Kriterien festzulegen, nach denen vorliegende Informationen bewertet werden können. Diese Kriterien sollen eine Einordnung ermöglichen, ob eine Information wissenschaftlich ist und damit zur Entscheidungsfindung verwendet werden kann oder nicht.

Informationen erreichen ein Individuum in der Form von Aussagen. Der Kanal, über den die Aussagen vermittelt werden, ist dabei zunächst nicht relevant. Entscheidend ist jedoch die Art der Aussage. Kornmeier (2007) listet die verschiedenen Aussagen-

typen auf und orientiert sich dabei an Alt (1994), Nienhüser (1989) und Raffée (1974), sowie an Nienhüser & Magnus (2003). Aussagen können logisch, empirisch, normativ oder meta-physisch ausgerichtet sein. Eine Handlungsrelevanz für praktische Probleme besitzen jedoch nur die logischen und empirischen Aussagen. Normative und meta-physische Aussagen sind nicht-wahrheitsfähig und damit für praktische Probleme nicht handlungsrelevant. Wahrheitsfähige empirische Aussagen lassen sich unterteilen in deskriptive, explikative und technologische Aussagen.

2.1 wahrheitsfähige Aussagen

2.1.1 logische Aussagen

Lässt sich der Wahrheitsgehalt einer Aussage anhand der Regeln der Logik auf logische Konsistenz überprüfen, handelt es sich um eine logische Aussage. Klassische Syllogismen führen zu einer eindeutigen Einordnung, ob die Aussage wahr ist.
Beispiel:

Prämisse 1:	Alle Halogene liegen zweiatomig vor.
Prämisse 2:	Brom ist ein Halogen.
Konklusion:	Brom liegt zweiatomig vor.

2.1.2 empirische Aussagen

Beschreiben Aussagen einen realen Sachverhalt, handelt es sich um empirische Aussagen. Sie lassen sich durch empirische Experimente oder Studien prüfen, also durch „Konfrontation mit der Realität“ (Kornmeier 2007, S. 46). Empirische Aussagen können sehr vielfältig sein und lassen sich ihrerseits in weitere Kategorien unterteilen.

Deskriptive Aussagen

Deskriptive Aussagen bilden die Basis empirischer Erkenntnis. Sie beschreiben durch konkreten Raum/Zeit-Bezug einzelne Sachverhalte und sind intersubjektiv überprüfbar, z. B. durch Beobachtung der Realität (Kornmeier 2007, S. 48; Raffée 1974, S. 29)
Beispiel:
Die Dichte des Gases Helium beträgt $178,5 \frac{g}{m^3}$.

Explikative Aussagen

Explikative Aussagen besitzen generelle Gültigkeit ohne Raum-Zeit-Bezug, sie erklären in beschreibender Weise einen umfassenderen Ausschnitt als deskriptive Aussagen. Sie stellen Ursache-Wirkungs-Beziehungen her und beantworten die Frage: „Warum ist das der Fall?“. Naturwissenschaftliche Gesetze, aber auch stochastische Aussagen und Tendenzen sind explikativer Natur (Kornmeier 2007, S. 49; Lingnau 1995, S. 125).
Beispiel:
Die Gravitationskraft bewirkt, dass Gegenstände zu Boden fallen.

Technologische Aussagen

Technologische Aussagen oder auch Ziel/Mittel-Aussagen sind konkreter als Gesetzesaussagen und beschreiben genau, wie mit einem bestimmten Mittel ein bestimmtes Ziel erreicht werden kann (Kornmeier 2007, S. 55).
Beispiel:
Wenn ein Gemisch aus rotem Phosphor und Kaliumchlorat einem starken Schlag ausgesetzt wird, dann explodiert es.

2.2 nicht-wahrheitsfähige Aussagen

2.2.1 normative Aussagen

Normative Aussagen sind wertsetzend und empfehlen eine bestimmte Handlung. Sie schreiben vor, wie etwas sein soll und basieren auf gesellschaftlichen Werturteilen. Diese Art von Aussagen kann nicht anhand des Wahrheitskriteriums geprüft werden, dennoch ist es in manchen Situationen ratsam, die Handlungskonsequenzen normativer Aussagen zu evaluieren (Kornmeier 2007, S. 47; Nienhüser & Magnus 2003, S. 11).

Beispiel: Während der Corona-Pandemie sollen die Menschen einen Mund-Nasen-Schutz tragen.

2.2.2 meta-physische Aussagen

Meta-physische Aussagen können empirisch nicht überprüft werden und entziehen sich damit auch jeglicher Möglichkeit, widerlegt zu werden. Sie können wahr sein oder nicht, allerdings ist es nicht möglich, ihre Wahrheit wissenschaftlich festzustellen. Dennoch können meta-physische Aussagen Denkanstöße liefern oder neue Sichtweisen für praktische Problemlösungen darstellen (Kornmeier 2007, S. 47).

Beispiel: Es könnte außerirdisches Leben geben.

2.3 Gewinnen wissenschaftlicher Erkenntnis anhand von Aussagen

Die Unterscheidung zwischen wahrheitsfähigen und nicht-wahrheitsfähigen Aussagen zeigt, dass nicht jede getroffene Aussage zur wissenschaftlichen Erkenntnisgewinnung herangezogen werden kann. Lediglich logische und empirische Aussagen sind derart belastbar, dass auf ihrer Grundlage neue Hypothesen, Theorien und Gesetze entstehen können. Um zu verstehen, in welcher Weise dies geschehen kann, ist ein tieferer Blick in die Wissenschaftstheorie notwendig.

Deskriptive Aussagen beschreiben, wie zuvor dargestellt, beobachtbare Aspekte der Realität und werden daher auch beobachtbare Tatsachen genannt (Chalmers et al. 2007). Unter der vereinfachten Annahme, dass Beobachtungsprozesse nicht fehlbar sind, gelten Beobachtungen als adäquates Mittel, um Tatsachen zu validieren (Chalmers et al. 2007, S. 5). Allerdings ist es auch durch sorgfältige Beobachtung nicht möglich, die Welt in ihrer Komplexität hinreichend zu erfassen. Zur Gewinnung von Tatsachen ist es daher zusätzlich notwendig, Experimente durchzuführen, um den zu untersuchenden Prozess zu isolieren und andere Effekte von der Untersuchung auszuschließen (Chalmers et al. 2007, S. 26).

Entscheidend ist die Sichtweise, mit der auf den weiteren Prozess der Datenverarbeitung geblickt wird. Wie werden aus Tatsachen (deskriptiven Aussagen) übergeordnete Theorien und Gesetze?

2.3.1 Verallgemeinerungen: Der Induktivismus

Der Induktivismus sieht Wissenschaft als induktives Schließen von beobachtbaren Tatsachen auf übergeordnete Theorien und Gesetze (Chalmers et al. 2007; Popper 1989, S. 4). Tatsachen bzw. deskriptive Aussagen werden verwendet, um ein theoretisches Gerüst zur Erklärung der Wirklichkeit zu entwerfen. Besondere Tatsachen, gewonnen durch Beobachtungen und Experimente, erfahren einen Verallgemeinerungs-Prozess. Dabei entziehen sie sich jedoch der klassischen Logik, da sie einen Schritt über das hinausgehen, was in Tatsachen enthalten ist (Chalmers et al. 2007, S. 38). Das folgende Beispiel zeigt einen typischen induktiven Schluss und steht im Gegensatz zum klassischen logisch-deduktiven Schluss einer logischen Aussage (Kapitel 2.1.1).

Beispiel:

Tatsache 1:	Fluor ist ein Halogen und liegt zweiatomig vor.
Tatsache 2:	Chlor ist ein Halogen und liegt zweiatomig vor.
Tatsache 3:	Brom ist ein Halogen und liegt zweiatomig vor.
Tatsache *n*:	Element *n* ist ein Halogen und liegt zweiatomig vor.
Konklusion:	Alle Halogene liegen zweiatomig vor.

Dieser Schluss ist logisch nicht gültig, da aus den Tatsachen keine eindeutige allgemeingültige Schlussfolgerung gezogen werden kann. Es wurden jedoch Gütekriterien für induktive Argumente aufgestellt, deren Beachtung dafür sorgen soll, dass induktive Schlüsse zulässig sind (Chalmers et al. 2007, S. 40):

> Wenn eine große Anzahl von A unter einer großen Vielfalt von Bedingungen beobachtet wird, und wenn alle diese beobachteten A ohne Ausnahme die Eigenschaft B besitzen, dann besitzen alle A die Eigenschaft B.

Ungeklärt bleibt, wie die „große Anzahl“ oder „große Vielfalt von Bedingungen“ definiert werden. Ab wann ist ein induktives Argument angemessen, um die Ableitung von Theorien und Gesetzen zu legitimieren? Diese Fragen lässt die induktivistische Sichtweise offen.

Auch Popper (1989) hat das Induktionsproblem erkannt und behandelt es in seiner Schrift „Logik der Forschung“ als ein Grundproblem der Erkenntnislogik. Popper postuliert, es sei nicht möglich, das Induktionsproblem zu überwinden, da es in

einen „unendlichen Regress“ führe: Bei der Beantwortung der Frage, wie das Induktionsprinzip wissenschaftlich gerechtfertigt werden kann, müssten wieder induktive Schlüsse angewendet werden (Popper 1989, S. 4). Das müsste nach folgendem Schema geschehen (Chalmers et al. 2007, S. 43):

Tatsache 1:	Das Induktionsprinzip hat im Fall x_1 funktioniert.
Tatsache 2:	Das Induktionsprinzip hat im Fall x_2 funktioniert.
Tatsache *n*:	Das Induktionsprinzip hat im Fall x_n funktioniert.
Konklusion:	Das Induktionsprinzip funktioniert immer.

Es wurde gezeigt, dass induktive Schlüsse nicht geeignet sind, um aus einem Satz deskriptiver Aussagen allgemeingültige Gesetzesaussagen abzuleiten. Den Ausweg aus diesem Problem sieht Popper im Prinzip des Falsifikationismus.

2.3.2 Falsifikationismus

Das Induktionsproblem sieht Popper (1989) als unüberwindbar an und betrachtet daher induktive Schlüsse als unzulässig für das endgültige Beweisen von Aussagen. Nur die deduktive Überprüfung von Hypothesen liefert zuverlässige Ergebnisse.

Das bedeutet, dass Theorien und Hypothesen von Wissenschaftlern (zunächst unbegründet) aufgestellt und als Basis für die experimentelle Überprüfung verwendet werden. Die deskriptiven Einzelaussagen können verwendet werden, um eine bereits aufgestellte Theorie zu bestätigen oder zu widerlegen; nicht jedoch, um neue allgemeingültige Theorien aufzustellen (Popper 1989, S. 7). Popper geht noch einen Schritt weiter und behauptet, Theorien seien niemals empirisch verifizierbar. Er bezieht sich dabei auf den bis dahin weit verbreiteten Gedanken, dass eine Aussage nur sinnvoll sei, wenn ihre Wahrheit endgültig bewiesen ist.

Bei Schlick oder Waismann ist die von Popper abgelehnte Haltung erkennbar: „… eine echte Aussage muss sich endgültig verifizieren lassen“ (Schlick 1931) bzw. „… der Sinn eines Satzes ist die Methode seiner Verifikation“ (Waismann 1930). Aufgrund des Induktionsproblems ist es genau diese Verifikation aber nicht möglich.

Dass sich eine Aussage durch Beobachtung/Experiment bestätigt bedeutet nicht, dass die Aussage auch wahr ist. Andererseits werden Aussagen nur als empirisch-wissenschaftlich betrachtet, wenn sie auch durch Erfahrungen überprüfbar sind. Anstelle der Wahrheit von „Systemen“ (also Theorien und Hypothesen) kann auch festgestellt

werden, dass sich Systeme bei Überprüfungen als falsch erweisen. Poppers Ausführungen setzen genau an dieser Stelle an (Popper 1989, S. 15):

> Wir fordern zwar nicht, daß das System auf empirisch-methodischem Wege endgültig positiv ausgezeichnet werden kann, aber wir fordern, daß es die logische Form des Systems ermöglicht, dieses auf dem Wege der methodischen Nachprüfung negativ auszuzeichnen: Ein empirisch-wissenschaftliches System muß an der Erfahrung scheitern können.

Mit dieser Sichtweise ist das Induktionsproblem ausgehebelt. Es wird nicht mehr gefordert, dass aufgrund von deskriptiven Einzelaussagen induktiv die Wahrheit eines Systems bewiesen wird. Stattdessen ist es möglich, anhand von deduktiv gewonnenen Einzelaussagen auf die Falschheit eines Systems zu schließen: Widerspricht eine einzige Beobachtung dem zu überprüfenden System, so ist dieses auf der Stelle widerlegt oder auch falsifiziert. Steht eine Beobachtung in Einklang mit dem System, so hat sich dieses *bewährt* - es wurde jedoch nicht *bewiesen*.
Die Falsifizierung einer Aussage bedeutet nicht, dass diese für den wissenschaftlichen Diskurs wertlos ist. Bei der Entscheidungsfindung sind Informationen in Form von falsifizierten Aussagen besonders relevant, denn sie treffen eindeutige Aussagen darüber, wie bestimmte Zusammenhänge *nicht* auftreten.
Poppers Sichtweise verdeutlicht den hohen Stellenwert der Hypothesenprüfung und der damit einhergehenden Versuche, Hypothesen zu falsifizieren. Eine Falsifikation kann jedoch nur gelingen, wenn die zu prüfende Aussage auch falsifizierbar formuliert ist: Die Falsifizierbarkeit wird als zwingendes Kriterium für Aussagen angesehen, die zur Entscheidungsfindung verwendet werden sollen. Welche weiteren Kriterien dabei erfüllt werden müssen, wird in einem späteren Kapitel dieser Arbeit dargelegt (Kapitel 2.4).
Dass dem Falsifikationismus keine uneingeschränkte Geltung zugesprochen werden sollte, zeigt eine Betrachtung von Extrem- bzw. Randpositionen. Sehr gewagte, neuartig-kühne Hypothesen zu falsifizieren, erwirkt keine großartigen Durchbrüche in der Wissenschaft: Die Hypothese, dass ab morgen alle Gegenstände nach oben steigen, statt zu Boden zu fallen, ist sehr gewagt und leicht falsifizierbar - die wissenschaftliche Tragweite ihrer Falsifizierung ist jedoch gering.
Vorsichtig formulierte Hypothesen zu widerlegen ist weitaus bedeutender, aber auch unwahrscheinlicher: „Das Einhalten eines Sicherheitsabstands von zwei Metern zwi-

schen Personen verringert das Ansteckungsrisiko mit dem Corona-Virus SARS-CoV-2“ ist eine vorsichtig formulierte Hypothese, deren Falsifizierung zwar sehr unwahrscheinlich ist, aber großes Aufsehen erregen würde.
Daher ist es wichtig, die Prinzipien des Falsifikationismus dahingehend zu erweitern, dass der Positivismus ebenfalls eine gewisse Bedeutung erhält: Bedeutsame wissenschaftliche Fortschritte entstehen durch die *Bewährung* von kühnen Hypothesen oder durch *Falsifikation* von behutsamen, vorsichtig-formulierten Aussagen (Chalmers et al. 2007, S. 67).

2.3.3 wissenschaftstheoretische Grundbegriffe: Von der These zur Theorie

Deskriptive Aussagen beschreiben erfahrbare Tatsachen und lassen sich, erkenntnistheoretisch betrachtet, mit einfachen Mitteln überprüfen. Der wissenschaftliche Fortschritt wird durch diese Aussagen jedoch nur bedingt vorangetrieben, da sie aufgrund ihres konkreten Raum-Zeit-Bezugs nicht allgemeingültig und daher nicht auf andere Sachverhalte anwendbar sind.
Explikative Aussagen, z. B. in Form von Gesetzesaussagen, sind von höherem wissenschaftlichen Wert, da sie einen größeren Geltungsbereich haben. Das Aufstellen explikativer Aussagen kann einfach und schnell durch jedes Individuum erfolgen, dabei muss zunächst keine Begründung vorliegen. Diese Folgerung geht logisch aus Poppers Umgehung des Induktionsproblems hervor. Die Aufgabe der empirischen Wissenschaften besteht in der Überprüfung von zuvor aufgestellten Aussagen.
In der Wissenschaftstheorie werden verschiedene Fachtermini verwendet, die den jeweiligen Status einer Aussage im Bezug auf dessen wissenschaftliche Güte definieren. Die Verwendung dieser Begriffe findet nicht immer trennscharf statt und einheitliche Definitionen existieren nur in Teilen. Daher wird im folgenden Abschnitt das in dieser Arbeit vertretene Verständnis der Grundbegriffe erläutert.

These

Eine These bezeichnet eine behauptend aufgestellte Aussage in Form eines Satzes oder Gedankenganges, deren empirischer Gehalt festgestellt werden muss (Duden: *These*). An eine These werden keine bestimmten Bedingungen geknüpft und sie ist

in jeder Umgebung vorzufinden: Thesen können in einem alltäglichen persönlichen Gespräch geäußert, in den sozialen Medien verbreitet, in Printmedien abgedruckt oder auch in den eigenen Gedanken entwickelt werden.
Die Bandbreite möglicher Thesen ist enorm groß und erstreckt sich über alle Themengebiete. Thesen haben keine richtungsweisende Relevanz in wissenschaftlichen Diskursen, solange sie nicht die Kriterien einer Hypothese erfüllen.

Hypothesen

Hypothesen drücken Annahmen über reale Sachverhalte in Form von Konditionalsätzen aus, d. h. sie beschreiben einen Zusammenhang zwischen mindestens zwei Variablen (Bortz & Döring 2006, S. 4; Schnell et al. 2005, 53ff). Sie sind frei von Widersprüchen, prinzipiell falsifizierbar und müssen sich empirischen Untersuchungen unterziehen, um entweder falsifiziert zu werden oder sich zu bewähren (Duden: *Hypothese*).

Gesetz

Gesetze sind zu verstehen als explikative Aussagen mit Allgemeingültigkeitsanspruch, deren Allgemeingültigkeit in mehrfachen empirischen Untersuchungen bisher nicht falsifiziert wurde. Sie beziehen sich auf Regelmäßigkeiten der Natur („Naturgesetze“) (Chalmers et al. 2007, S. 37).

Theorie

Die übergeordnete Begrifflichkeit der *Theorie* beschreibt eine Menge an Gesetzeshypothesen, die logisch widerspruchsfrei miteinander verbunden sind (Lingnau 1995, S. 125) und Zustände der Realität beschreiben (Franke 2002, S. 179). Eine Theorie kann Gesetze, Hypothesen und deskriptive Aussagen enthalten (Ben-Ari 2011, S. 24; American Museum of Natural History 2021).

Die Erläuterungen dieses Kapitels zeigen deutlich, dass die in der Entscheidungstheorie geforderte „Suche, Verarbeitung und Integration von Informationen“ keinen geradlinigen, eindimensionalen Prozess darstellt. Es ist von entscheidender Bedeutung, welche Art von Informationen vorliegt - im Optimalfall ist die wissenschaftliche Datenlage klar und wird durch eine empirisch gut bewährte Theorie beschrieben.

In komplexen Entscheidungssituationen sind die Handlungsoptionen und -konsequenzen jedoch häufig kontrovers, sodass verschiedene zueinander in Widerspruch stehende Theorien existieren.
Es wurde gezeigt, dass die Falsifizierbarkeit eine grundlegende Anforderung an Hypothesen darstellt. Deshalb ist es notwendig, verschiedene Theorien zu einem kontroversen Thema zunächst in ihre Bestandteile, die einzelnen Thesen, zu zerlegen und diese auf Falsifizierbarkeit zu überprüfen. Enthalten Theorien nicht-falsifizierbare Thesen, so sind sie unwissenschaftlich und in der Kontroverse irrelevant.
Nach der logischen Prüfung auf Falsifizierbarkeit bleiben nur Theorien übrig, die ausschließlich falsifizierbare Hypothesen enthalten. Das bedeutet jedoch nicht, dass diese Theorien unreflektiert für eigene Handlungsentscheidungen verwendet werden können. Die aussagenlogische Unbedenklichkeit ist nicht gleichzusetzen mit einer Aussage über die inhaltliche Qualität der Theorien: Die falsifizierbaren Theorien können dennoch auf inhaltlich falschen Annahmen beruhen oder wissenschaftlich fragwürdige Informationen enthalten. Daher müssen Theorien mittels empirischer Überprüfung einem Fakten-Check unterzogen werden.

2.3.4 Operationalisierung

Die Prüfung verschiedener Theorien auf ihre Falsifizierbarkeit stellt sicher, dass die einzelnen Bestandteile der zu überprüfenden Theorien logisch widerspruchsfrei und empirisch überprüfbar sind. Um Theorien als Systeme von Hypothesen und Gesetzen inhaltlich zu überprüfen, sind über die reine Logik hinausgehende Untersuchungen notwendig. Wie oben aufgezeigt, müssen Theorien empirischen Falsifizierungs-Versuchen standhalten und sich dadurch bewähren.
In den empirischen Wissenschaften wird versucht, den in einer Hypothese postulierten Zusammenhang von Variablen in der Realität zu beobachten. Dabei ist jedes Konstrukt operationalisierbar, also beschreibbar durch konkrete Beobachtungs- oder Mess-Anleitungen (Schnell et al. 2005, S. 75). Diese Anleitungen geben an, wie hypothetischen Konstrukten beobachtbare Sachverhalte zugeordnet werden können (Schnell et al. 2005, S. 130).
Die Operationalisierung von Hypothesen kann als Zwischenschritt auf dem Weg von der theoretisch-hypothetischen Ebene zur empirischen Ebene angesehen werden, sie dient der Überführung der theoretischen Konzepte in messbare und beobacht-

bare Variablen (Bortz & Döring 2006, S. 223). Die Erhebung der Messwerte, gefolgt von ihrer statistischen Analyse, erlaubt letztendlich Rückschlüsse auf die Gültigkeit der getesteten Hypothese.
Die konkrete Ausgestaltung der empirischen Überprüfung ist abhängig von der Fachrichtung und der Fragestellung. Quantitative Messungen werden häufig in den Naturwissenschaften durchgeführt, während psychologische und sozialwissenschaftliche Fragestellungen meist anhand von klinischen und epidemiologischen Studien und Befragungen untersucht werden.
Empirische Untersuchungen können also für einen Fakten-Check herangezogen werden. Entscheidend ist dabei jedoch die Qualität der Untersuchung. Für eine fundiert-rationale Entscheidungsfindung sollten nur Informationen hoher wissenschaftlicher Qualität verwendet werden. Nach welchen Kriterien festgestellt werden kann, wie gut eine empirische Untersuchung ist, wird im nächsten Abschnitt beleuchtet.

2.4 Gütekriterien empirischer Untersuchungen

In den Naturwissenschaften wird Erkenntnis durch Experimente gewonnen. Wie zuvor beschrieben, dienen die Experimente dabei nicht der induktiven Generierung von Hypothesen und Theorien, sondern der deduktiven Überprüfung bereits bestehender Hypothesen. Ein Experiment ist eine Versuchsanordnung zur empirischen Überprüfung von Hypothesen zu Ursache-Wirkungs-Zusammenhängen. Es wird unter kontrollierten, vorher festgelegten Umweltbedingungen durchgeführt und zeichnet sich dadurch aus, dass „eine oder mehrere unabhängige Variablen planmäßig variiert und deren Wirkung auf eine oder mehrere abhängige Variablen bestimmt werden [sic!]“ (Huber et al. 2014, S. 23; Meffert 1992, S. 207; Koch et al. 2016, S. 71).
Viele Fragestellungen können jedoch nicht mittels klassischer Laborexperimente ermittelt werden, da besonders bei psychologischen, sozialwissenschaftlichen und gesundheitsbezogenen Themen die unabhängige Variable nur schwer zu beeinflussen ist: Die Hypothese, dass Handystrahlung die Bildung von Krebserkrankungen bei Menschen hervorruft, lässt sich nicht einfach durch Variation der unabhängigen Variablen untersuchen - denn in dem Fall müssten Personen systematisch einer potentiellen Gefährdung ausgesetzt werden (Bestrahlen mit Handystrahlung).
Aus diesem Grund haben sich in verschiedenen Wissenschaftsdisziplinen weitere

empirische Untersuchungsmethoden herausgebildet. Dazu gehören u. a. Befragungen, Beobachtungen und Studien. Jede dieser Methoden besitzt eigene Besonderheiten und Ansprüche, einige grundlegende Kriterien lassen sich jedoch für alle empirischen Methoden formulieren.
In den folgenden Ausführungen wird mit dem Begriff der Untersuchung die Gesamtheit aller Experimente, Befragungen, Studien, etc. bezeichnet.
In realen Untersuchungssituationen gibt es neben der zu variierenden unabhängigen Variablen und den abhängigen Variablen häufig auch Störgrößen, die sich nicht beeinflussen lassen. Dazu gehören z. B. demographische Daten von Versuchspersonen oder bestimmte Einflüsse der äußeren Umwelt. Diese externen Störvariablen können das Ergebnis von Untersuchungen verzerren und für systematische Fehler sorgen (Huber et al. 2014, S. 25). Die empirische Forschung folgt bestimmten Gütekriterien, um den Einfluss der Störvariablen möglichst gering zu halten. Allgemein gefordert werden Validität, Reliabilität und Objektivität von Untersuchungen (Huber et al. 2014, S. 39; Koch et al. 2016, S. 231; Berekoven et al. 2009, 87ff).

2.4.1 Validität, Reliabilität, Objektivität

Die Validität als Genauigkeit bzw. Gültigkeit einer Messung ist notwendig, um Aussagen über Kausalitäten treffen zu können. Sie ist gegeben, wenn das Untersuchungsinstrument den Sachverhalt untersucht, der zu messen beabsichtigt war (Koch et al. 2016, S. 231). Eine Untersuchung ist intern valide, wenn die Variation der abhängigen Variablen ausschließlich auf die Manipulation der unabhängigen Variablen zurückgeführt werden kann. Dies ist nur der Fall, wenn die Störvariablen ausreichend beherrscht werden.
Externe Validität ist gegeben, wenn die Ergebnisse der Untersuchung verallgemeinerbar sind. Zwischen interner und externer Validität herrscht ein Spannungsverhältnis, denn mit steigender interner Validität entfernt sich eine Untersuchung von den Bedingungen in der realen Umwelt, sie nähert sich einem Laborexperiment an. Damit sinkt die externe Validität. Campbell & Stanley (1963) räumen der internen Validität dabei jedoch die höhere Priorität ein.
Eine Untersuchung ist *reliabel*, wenn wiederholte Messungen eines Objekts mit einem Messinstrument immer die gleichen Ergebnisse liefern (Huber et al. 2014, S. 41). Wird eine Untersuchung von verschiedenen voneinander unabhängigen Wissenschaft-

ler:innen durchgeführt, dabei aber stets das gleiche Ergebnis erhalten, dann ist die Untersuchung als *objektiv* anzusehen. Das Ergebnis einer Untersuchung muss also unabhängig sein von der Person, die diese Untersuchung durchführt (Bortz & Döring 2006, S. 442).
Die Gütekriterien stehen miteinander in Zusammenhang: Objektivität ist Voraussetzung für Reliabilität, diese ist wiederum die Grundlage für Validität (Berekoven et al. 2009, S. 80; Huber et al. 2014, S. 41).
In diesem Abschnitt wurden allgemeine Gütekriterien für empirische Untersuchungen hergeleitet und miteinander in Beziehung gesetzt. Unklar ist bisher jedoch, mit welchen Mitteln diese Kriterien sichergestellt werden können. Die übergeordneten Kriterien der Validität, Reliabilität und Objektivität erfordern, wie oben erklärt, die Beherrschung und Kontrolle von Störvariablen sowie die generelle Reproduzierbarkeit von Untersuchungen. Im folgenden Abschnitt wird die Kontrolle von Störfaktoren etwas detaillierter aufgeschlüsselt, um konkrete Handlungsanweisungen für das Design empirischer Untersuchungen generieren zu können.

2.4.2 Kontrolle von Störfaktoren

In einer idealen Untersuchungssituation werden alle (abhängigen) Variablen konstant gehalten, während ausschließlich die unabhängige Variable variiert wird. Die Ergebnisse der Untersuchung lassen dann eine eindeutige kausale Aussage zu, in welcher Weise die unabhängige Variable auf die abhängige(n) Variable(n) wirkt. Beispiel: Je größer die Temperatur von Wasser ist, desto geringer ist die Gaslöslichkeit. Eine systematische Erhöhung der Wassertemperatur unter Beobachtung der Gaslöslichkeit ist leicht zu kontrollieren.
Komplexere Untersuchungen, besonders jene, die Vorgänge im menschlichen Körper betreffen, sind jedoch schwer zu kontrollieren. Das konkrete Untersuchungsobjekt ist nicht „ein Mensch“, stattdessen wird eine große Gruppe an Personen untersucht. Die Variation der unabhängigen Variablen besteht dann darin, einen Teil der Gruppe der variierten Variable auszusetzen (Experimentalgruppe), während ein weiterer Teil nicht behandelt wird (Kontrollgruppe). Dieses Vorgehen wird als *Kontrollgruppendesign* bezeichnet und gilt als Qualitätskriterium für empirische Studien (Bortz & Döring 2006, S. 193; Brosius et al. 2008, S. 205).
Da sich alle Menschen in verschiedenen Merkmalen unterscheiden, ist es wichtig, die-

se Unterschiede in der Untersuchung zu berücksichtigen und Gruppen zu bilden, die exakt vergleichbar sind. Diese Anforderung kann - bei ausreichender Gruppengröße - durch zufällige Zuordnung der Untersuchungsobjekte zu den Gruppen sichergestellt werden (Scheffler 2000, S. 64). Studien, die diesem Prinzip der *Randomisierung* folgen, werden randomisierte Studien genannt (Bortz & Döring 2006, S. 193). Ein umfassender Überblick über Stichprobenauswahlverfahren und verschiedene Möglichkeiten, Randomisierung sicherzustellen, findet sich bei Schnell et al. (2005, S. 265ff), sowie bei Westermann (2000, S. 310ff) und Mossig (2012).

Ein weiterer Störfaktor wird durch Erwartungseffekte dargestellt. Sowohl Versuchsteilnehmende als auch -leitende sind durch persönliche Einstellungen gegenüber der Untersuchung gekennzeichnet und entwickeln daher bestimmte Erwartungen im Bezug auf die Untersuchungsergebnisse. Diese Effekte werden durch *Verblindung* minimiert. Die Details der Durchführung und Auswertung werden dabei den im Bezug zur Untersuchung stehenden Personen vorenthalten. Beispielsweise wird Versuchspersonen nicht mitgeteilt, ob sie der Experimental- oder Kontrollgruppe zugeordnet wurden (Bortz & Döring 2006, S. 197; Kabisch et al. 2011).

Einflüsse von Störfaktoren bei empirischen Untersuchungen können also durch Einhaltung des *Kontrollgruppendesigns* sowie der Prinzipien *Randomisierung* und *Verblindung* minimiert werden. Das steigert die interne Validität einer Untersuchung und somit auch die Möglichkeit, kausale Aussagen treffen zu können. Wie bereits erwähnt, sind Untersuchungen mit einer hohen internen Validität häufig durch eine geringe externe Validität gekennzeichnet und damit nur begrenzt generalisierbar. Um die Ergebnisse dieser Untersuchungen jedoch im Hinblick auf Realsituationen aussagekräftiger zu gestalten, können systematische *Replikationsstudien* durchgeführt werden (Bortz & Döring 2006, S. 198). Dabei wird die zuvor unter kontrollierten, randomisierten und verblindeten Bedingungen durchgeführte Studie erneut durchgeführt. Bei der Replikation werden jedoch im Hinblick auf die Zielpopulation etwas andere Bedingungen gewählt (z. B. werden an der Stelle von Jugendlichen nun Erwachsene fortgeschrittenen Alters untersucht). Durch diese Vorgehensweise wird der Geltungsbereich der Untersuchung, also die externe Valididät, erhöht.

2.4.3 Interpretation und Dokumentation

Die Ergebnisse empirischer Untersuchungen werden durch die Forscherteams veröffentlicht. Die Veröffentlichungen enthalten in der Regel neben einer Einleitung und einem Methodenteil auch einen Abschnitt, in dem die Ergebnisse dargestellt, interpretiert und diskutiert werden. Für die Forschenden ist es dabei wichtig, dass die Aussagekraft ihrer Ergebnisse zur Geltung kommt.

Werden die eigenen Erkenntnisse jedoch glorifiziert oder über die eigentliche, wissenschaftlich gültige Aussagekraft hinaus zu plakativ formuliert, können sie von den Rezipienten der Veröffentlichung überinterpretiert werden (Bortz & Döring 2006, S. 204). Daher ist es von besonderer Bedeutung, dass Forschende bei der Dokumentation ihrer Untersuchungen zwischen Korrelationen und Kausalitäten unterscheiden (Kornmeier 2007, 63ff).

Wird ein Zusammenhang zwischen der unabhängigen und einer untersuchten abhängigen Variablen festgestellt, so besteht eine *Korrelation* zwischen den Variablen (z. B. je höher die unabhängige Variable ausgeprägt ist, desto stärker tritt auch die abhängige Variable hervor).

Ob das Auftreten eines Faktors jedoch auch die Ursache des Auftretens des anderen Faktors ist, kann anhand einfacher Häufigkeitsbetrachtungen nicht festgestellt werden. Eine Aussage über die *Kausalität* zwischen zwei Variablen kann erst getroffen werden, wenn das Untersuchungsdesign ausreichend valide ist (siehe Kapitel 2.4.1), ansonsten handelt es sich um eine Scheinkorrelation (Kornmeier 2007, S. 63), wenngleich der Begriff einer Scheinkausalität passender erscheint.

Einige Untersuchungsarten sind allein von ihrer Forschungsmethodik so angelegt, dass eine Kausalitätsaussage kaum möglich ist. Das betrifft vor allem Studien, in denen personengebundene oder von der Umwelt abhängige Variablen untersucht werden: Es ist beispielsweise forschungspraktisch nicht realisierbar, personengebundene Variablen wie das Geschlecht, Links-/Rechtshändigkeit o. ä. zu variieren, daher lässt sich ein kausaler Einfluss derartiger Variablen nicht ermitteln (Bortz & Döring 2006, S. 203). Studien, die diese Faktoren untersuchen, sind in der Regel nicht-experimentell und können daher nur Korrelationsaussagen treffen.

Untersuchungen, die den Kriterien interner und externer Validität entsprechen, müssen zudem auch reliabel und objektiv gestaltet sein. Das bedeutet, dass die Untersuchung zu beliebigen Zeitpunkten von unterschiedlichen Personen unter Erhalten des-

selben Ergebnis wiederholbar sein muss. Dabei ist es wichtig, dass die Untersuchung nicht nur prinzipiell einem reproduzierbaren Forschungsdesign folgt, sondern auch nachvollziehbar dokumentiert wird.
Die Veröffentlichung von Untersuchungsergebnissen ist stets mit allen relevanten Details zu den Versuchsbedingungen zu versehen: Die Charakterisierung der Stichprobe oder des untersuchten Gegenstands ist dabei ebenso essentiell wie Angaben zur genauen Durchführung, zu untersuchten und nicht–betrachteten Variablen, Umweltbedingungen, Auswertungsverfahren, und so weiter (Huber et al. 2014). Diese Anforderung an wissenschaftliche Dokumentation wird zusammen mit den Kriterien der Reliabilität und Objektivität unter dem Begriff der *Reproduzierbarkeit* zusammengefasst.

2.4.4 Zusammenfassung der Gütekriterien

In diesem Kapitel wurden grundlegende Kriterien aufgezeigt, die empirische Untersuchungen erfüllen müssen, um wissenschaftlichen Ansprüchen zu genügen. Es wurde dargestellt, dass kontrollierte Untersuchungsbedingungen zusammen mit der Verblindung essentiell sind für die Validität empirischer Forschungsdesigns. Weiterhin muss die Reproduzierbarkeit sichergestellt werden, indem neben Reliabilität und Objektivität auch eine lückenlose und intersubjektiv nachvollziehbare Dokumentation aller Untersuchungsbedingungen beachtet wird. Bei der Auswertung, Interpretation und Darstellung der Ergebnisse müssen die Forschenden dem Unterschied zwischen statistischen Korrelationen und echten Kausalitäten besondere Aufmerksamkeit schenken, um Über- und Fehlinterpretationen zu vermeiden.

3 Bewertungskompetenz im naturwissenschaftlichen Unterricht

Das Kapitel 1 hat eine Einführung in die Entscheidungstheorie gegeben und beschreibt die Regeln und Bedingungen von Entscheidungssituationen. Es wurde herausgestellt, dass reale Entscheidungen aufgrund von begrenzter Kapazität bei der Informationsverarbeitung nur begrenzt rational getroffen werden. Nach dem Ansatz der Zwei Systeme (Kahneman 2011) können fehlerhafte Bewertungstendenzen durch das System 2 unter hohem kognitiven Aufwand ausgeglichen werden: Anstelle der Anwendung von Entscheidungsheuristiken werden wissenschaftliche Fakten aus verschiedenen Informationsquellen beschafft, verarbeitet und in den Entscheidungsprozess integriert.
Entscheidend dabei ist die wissenschaftliche Qualität der Informationen. Kapitel 2 gibt einen Überblick über den Weg der wissenschaftlichen Erkenntnisgewinnung und stellt verschiedene Gütekriterien empirischer Untersuchungen dar.
In diesem Abschnitt der Arbeit werden die zentralen Erkenntnisse aus Entscheidungs- und Wissenschaftstheorie zusammengeführt und in dem Konstrukt der Bewertungskompetenz zum Ausdruck gebracht.

3.1 Definition von Bewertungskompetenz

Bewertungskompetenz lässt sich aus verschiedenen Perspektiven beschreiben, basiert jedoch in den meisten Fällen auf dem Kompetenzbegriff nach Weinert (2001): Kompetenzen sind „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verant-

wortungsvoll nutzen zu können".

Die durch die Beschlüsse der Kultusministerkonferenz festgelegten Bildungsstandards schlüsseln den Kompetenzbegriff in die vier Kompetenzbereiche Fachwissen, Erkenntnisgewinnung, Kommunikation und Bewertung auf. Die Bildungsstandards der naturwissenschaftlichen Fächer definieren Bewertungskompetenz als Erkennen und Bewerten naturwissenschaftlicher Sachverhalte in verschiedenen Kontexten (KMK 2005a, S. 7; KMK 2005b, S. 7; KMK 2005c, S. 7). Etwas konkreter werden Rost et al. (2003, S. 13, zitiert nach BLK-Programm), sie definieren Bewertungskompetenz als „Fähigkeit und Bereitschaft, die Abhängigkeit menschlichen Handelns von subjektiven Wertmaßstäben zu erkennen und eigene Wertvorstellungen bei komplexen Entscheidungen zu berücksichtigen". Sie sei ebenso als Fähigkeit zu verstehen, „bei Entscheidungen unterschiedliche Werte zu erkennen, gegeneinander abzuwägen und in den Entscheidungsprozess einfließen zu lassen".

Weiterhin heißt es in den Bildungsstandards des Fachs Chemie (KMK 2005a, S. 10):

> Die Kenntnis und Reflexion der Beziehungen zwischen Naturwissenschaft, Technik, Individuum und Gesellschaft gehören zum Bereich Bewertung.
>
> Durch die Auswahl geeigneter Sachverhalte können die Schülerinnen und Schüler Vernetzungen der Chemie in Lebenswelt, Alltag, Umwelt und Wissenschaft erkennen. Darauf basierend sollen Schülerinnen und Schüler in der Lage sein, chemische Sachverhalte in ihrer Bedeutung und Anwendung aufzuzeigen.
>
> Diese gezielte Auswahl chemierelevanter Kontexte ermöglicht es den Schülerinnen und Schülern, Fachkenntnisse auf neue vergleichbare Fragestellungen zu übertragen, Probleme in realen Situationen zu erfassen, Interessenkonflikte auszumachen, mögliche Lösungen zu erwägen sowie deren Konsequenzen zu diskutieren.
>
> Bei der Betrachtung gesellschaftsrelevanter Themen aus unterschiedlichen Perspektiven erkennen die Lernenden, dass Problemlösungen von Wertentscheidungen abhängig sind. Sie sollen befähigt sein, Argumente auf ihren sachlichen und ideologischen Anteil zu prüfen und Entscheidungen sachgerecht, selbstbestimmt und verantwortungsbewusst zu treffen.

Der Kompetenzbereich Bewertung zielt demnach im Fach Chemie auf Interdisziplinarität, Perspektivenwechsel und eigenständige Problemlösung ab.
Die Bildungsstandards für das Fach Biologie behandeln Bewertungskompetenz ausführlicher. Die Wertschätzung einer intakten Natur, gesunder Lebensführung und ein Verständnis für nachhaltige Entwicklung stehen im Vordergrund. Es wird an ein Verantwortungsbewusstsein der Lernenden sowie an die Fähigkeit des Perspektivenwechsels appelliert. Das systematische Bewerten soll stets mit ethischen Werten in Beziehung gesetzt werden, wodurch Lernende sowohl eigene als auch fremde Urteile begründen können sollen und sich einen eigenen Standpunkt bilden. Hier findet eine starke Betonung der ethischen Bewertung statt (KMK 2005b, S. 12).
Für das Fach Physik halten die Bildungsstandards einen kurzen Abschnitt bereit. Die Lernenden sollen zwischen physikalischen, gesellschaftlichen und politischen Komponenten von Bewertungen unterscheiden können. Weiterhin erfordert der Kompetenzbereich Bewertung die „Fähigkeit zur Differenzierung nach physikalisch belegten, hypothetischen oder nicht naturwissenschaftlichen Aussagen in Texten und Darstellungen“ und die Kenntnis der Grenzen naturwissenschaftlicher Sichtweisen (KMK 2005c, S. 10).
Die angeführten Definitionen von Bewertungskompetenz verdeutlichen, dass bei diesem Themenbereich kognitive Fähigkeiten und Fertigkeiten beleuchtet werden. Während die Entscheidungstheorie wichtige Ansätze zur Erklärung von Entscheidungsverhalten liefert und in der Wissenschaftstheorie die Grundsätze wissenschaftlicher Erkenntnis definiert werden, geht es im Kompetenzbereich Bewertung darum, die Inhalte dieser Theorien zu explizieren.
Die Lernenden sollen sich bewusst machen, dass Entscheidungen und Urteilsprozesse häufig von Werten und persönlichen Einstellungen beeinflusst werden, dass Intuitionen eine wichtige Rolle spielen und komplexe Entscheidungen ein reflektiertes und rationales Abwägen der Handlungsoptionen unter Integration wissenschaftlicher Fakten erfordern. Lernende brauchen ein Verständnis für gesellschaftliche Problemstellungen mit naturwissenschaftlichen Bezügen, um an demokratischen Entscheidungsprozessen teilzunehmen (Bingle & Gaskell 1994; Heitmann & Tiemann 2011). Zu diesen Problemen existieren meist widersprüchliche Argumente, dennoch müssen Entscheidungen gefällt werden. Dazu befähigt Bewertungskompetenz als Bestandteil der naturwissenschaftlichen Grundbildung im Sinne der Bildungsstandards (Rost et al. 2003).

Der nächste Abschnitt stellt verschiedene Dimensionen und Strukturmodelle von Bewertungskompetenz dar.

3.2 Kompetenzstrukturmodelle

3.2.1 Modell der ethischen Urteilskompetenz

Eine Teildimension der Bewertungskompetenz bildet die ethische Urteilskompetenz. Sie ist vor allem in der biologiedidaktischen Forschung von großer Bedeutung. Bögeholz et al. (2004) geben in ihrem Beitrag einen Überblick über zehn Modelle zur Behandlung ethischer Fragen im Unterricht. Sie unterscheiden Modelle, die die ethische Reflexion in den Vordergrund stellen, von Modellen auf Grundlage von Kohlbergs Modell zur Entwicklung moralischer Urteilsfähigkeit. Außerdem führen sie Modelle mit den Schwerpunkten Bildung für nachhaltige Entwicklung oder Umweltbildung als eigene Gruppe auf.
Basierend auf dieser theoretischen Grundlage haben Reitschert et al. (2007) das Modell ethischer Urteilskompetenz entwickelt. Die Autoren beschreiben verschiedene Teildimensionen der ethischen Urteilskompetenz und konkretisieren Niveaustufen für die einzelnen Dimensionen. Dabei wurden die folgenden Teilkompetenzen identifiziert:

- Wahrnehmen und Bewusstmachen der eigenen Einstellung:
  Voreinstellungen und Neigungen zu einer ethischen Fragestellung spielen eine wichtige Rolle bei der Bewertung. Lernende benötigen ein Bewusstsein für die eigenen Einstellungen gegenüber einem Thema.
- Wahrnehmen und Bewusstmachen moralischer Relevanz:
  Nicht alle Aspekte einer Fragestellung sind moralisch relevant und basieren auf gesellschaftlichen Normen. Ein Bewusstsein für den moralischen Status eines Problems und die zugrundeliegenden ethischen Werte ist von großer Bedeutung.
- Folgenreflexion:
  Auf Entscheidungen folgen Konsequenzen - sowohl kurz- und mittelfristig für die Betroffenen als auch langfristig für die Gesellschaft. Mögliche Folgen einer

(ethisch relevanten) Entscheidung müssen daher aus verschiedenen Perspektiven reflektiert werden.

- Beurteilen:
  Die Teildimension des Beurteilens umfasst eine deskriptive Analyse des Problems und die Fähigkeit, Vor- und Nachteile geplanter Handlungen zu erörtern und eigene Beweggründe zu nennen. Mit steigender Beurteilungsfähigkeit können auch der eigenen Position entgegenstehende Gründe und Motive genannt, sowie Rückbezüge auf Werte vollzogen werden.
- Urteilen:
  Lernende erkennen, dass Meinungsäußerungen keine reflektierten Urteile darstellen - Für ein reflektiertes Urteil müssen alle anderen Dimensionen beachtet und für Schlussfolgerungen verwendet werden.
- Ethisches Basiswissen:
  Zentrale Begriffe der Ethik wie Moral, Wert oder Norm sind wichtige Bestandteile des (bio)ethischen Unterrichts. Eine grundlegende philosophische Basis sollte im Unterricht vermittelt werden.

Querliegend zu diesen Dimensionen werden die Fähigkeit zu Perspektivenwechsel und Argumentationsfähigkeit als über den gesamten Prozess der ethischen Urteilsbildung hinweg wichtige Kompetenzen beschrieben. Abbildung 3.1 zeigt die einzelnen Teildimensionen des Modells.
Zu den Teilkompetenzen *Wahrnehmen und Bewusstmachen moralischer Relevanz*, *Folgenreflexion*, *Perspektivenwechsel* und *Argumentieren* wurden je drei Niveaustufen aufgestellt (Mittelsten Scheid & Hößle 2008). Durch diese Einteilung in Niveaustufen liefert das Modell ethischer Urteilskompetenz ein Instrument zur Diagnose von Bewertungskompetenz bei Lernenden. Die Autor:innen betonen jedoch die bisher fehlenden empirischen Längsschnittstudien zur Passung des Instruments (Reitschert et al. 2007).

3.2.2 Göttinger Modell der Bewertungskompetenz

Das Göttinger Modell der Bewertungskompetenz nach Eggert & Bögeholz (2006) ist ein Modell für Gestaltungsaufgaben Nachhaltiger Entwicklung und dient zur Behandlung von komplexen naturwissenschaftlichen Fragestellungen mit persönlicher Re-

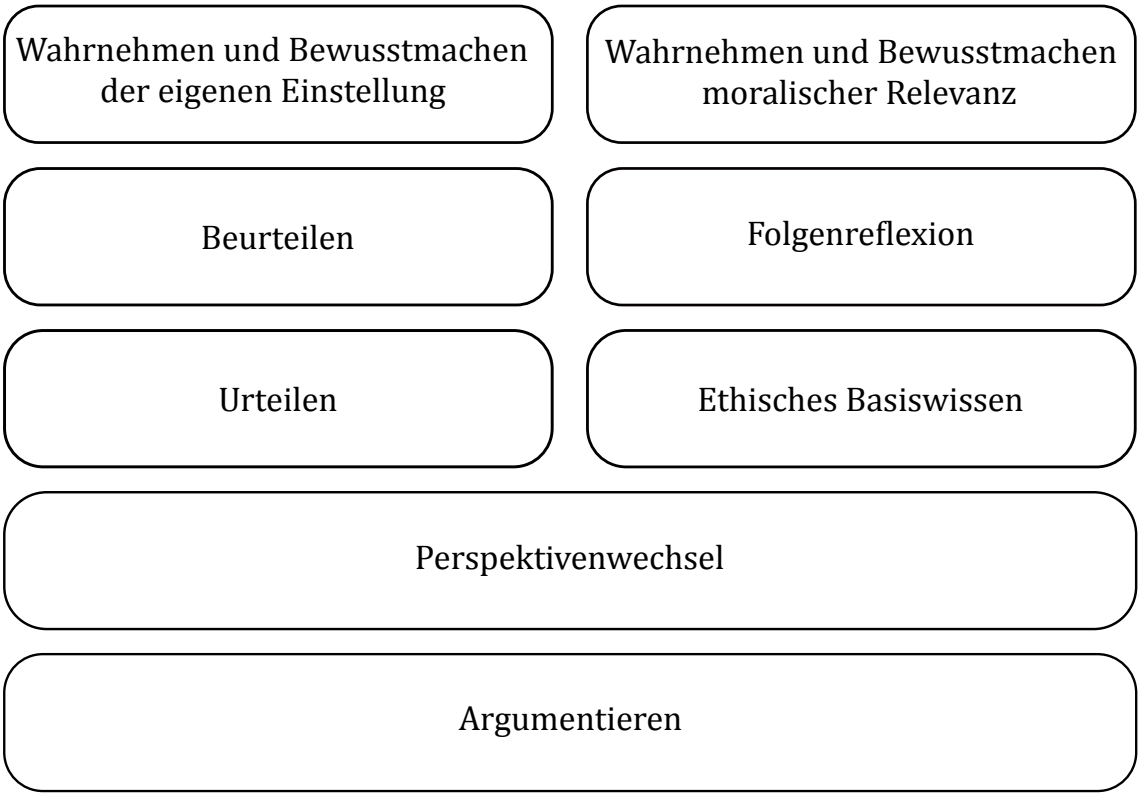


Abbildung 3.1: *Das Modell ethischer Urteilskompetenz nach Reitschert et al. (2007).*

levanz und gesellschaftlicher Tragweite (Bögeholz et al. 2018). Dabei wird Bewertungskompetenz durch einen transparenten, systematischen und reflektierten Umgang mit faktischer und ethischer Komplexität charakterisiert (Bögeholz & Barkmann 2005; Bögeholz et al. 2018). Es wurde aufbauend auf zentralen Erkenntnissen der Entscheidungstheorie (Betsch & Haberstroh 2014) und Prinzipien bestehender Kompetenzmodelle im Bezug auf *Scientific Literacy* (Bybee 1997; Hammann 2004; Eggert & Bögeholz 2006) entwickelt.

Eggert & Bögeholz (2006) unterteilen Bewertungskompetenz in die vier Teilkompetenzen *Kennen und Verstehen von Nachhaltiger Entwicklung*, *Kennen und Verstehen von Werten und Normen*, *Generieren und Reflektieren von Sachinformationen* und *Bewerten, Entscheiden und Reflektieren*.

Das Modell ist gezielt auf Gestaltungsaufgaben Nachhaltiger Entwicklung (GNE) ausgerichtet. Die Teilkompetenzen *Kennen und Verstehen von Nachhaltiger Entwicklung* und *Kennen und Verstehen von Werten und Normen* sind kontextspezifisch auf GNE bezogen und müssen bei der Anwendung des Modells in anderen Kontexten an die

jeweilige Domäne angepasst werden.
Die Teilkompetenz *Generieren und Reflektieren von Sachinformationen* bezieht sich auf die *präselektionale Phase* von Entscheidungsprozessen (siehe Kapitel 1.6.1). Dazu gehört die Fähigkeit, relevante Handlungsoptionen zu identifizieren und passende Untersuchungen zu planen, durchzuführen und deren Ergebnisse zu analysieren. Es wird explizit gefordert, dass Lernende auch Defizite von durchgeführten Datenerhebungen erkennen und reflektieren sollen, um auch bei defizitärer Informationslage handlungsfähig zu sein (Eggert & Bögeholz 2006).
Einen weiteren Teilbereich der Bewertungskompetenz stellt nach Eggert & Bögeholz (2006) das *Bewerten, Entscheiden und Reflektieren* dar. Bezugnehmend auf zentrale Erkenntnisse der Entscheidungstheorie werden hier Prozesse der *selektionalen Phase* (siehe Kapitel 1.6) adressiert. Die Lernenden wägen vorhandene Handlungsoptionen gegeneinander ab und vergleichen deren Vor- und Nachteile. Dabei gehen sie nicht nach Entscheidungsheuristiken oder Intuitionen vor, sondern wählen geeignete Entscheidungsstrategien aus. Zentral ist dabei auch die Fähigkeit, zwischen faktischen und ethischen Aussagen trennen zu können (Eggert & Bögeholz 2006).
Das Göttinger Modell wurde mit der Zeit weiterentwickelt zum *Modell zur Bewertungskompetenz für Gestaltungsaufgaben Nachhaltiger Entwicklung* (Bögeholz et al. 2014; Bögeholz et al. 2017; Bögeholz et al. 2018). Unterschieden werden die vier Teilkompetenzen 1. Beschreiben und Entwickeln von Handlungsoptionen, 2. Vollziehen von Perspektivenwechsel, 3. qualitatives Bewerten von Handlungsoptionen und 4. quantitatives Bewerten von Handlungsoptionen. Teilkompetenz 1 ist als analog zu *Generieren und Reflektieren von Sachinformationen* des Göttinger Modells anzusehen, während das Äquivalent zur Kompetenz 3 das *Bewerten, Entscheiden und Reflektieren* darstellt. Die einzelnen Teilkompetenzen wurden größtenteils bereits empirisch fundiert (Bögeholz et al. 2018).

3.2.3 Modell für die Evaluation der Bildungsstandards: Das ESNaS-Modell

Das ESNaS-Modell bedient ein breites Spektrum an Fragestellungen und ist sowohl für alltägliche Entscheidungen als auch für komplexe *socio-scientific issues* (siehe Kapitel 3.6) geeignet. Es dient als Grundlage für die Evaluation der Bildungsstandards und bietet ein Instrument zur Überprüfung des Erreichen der Bildungsstandards in

Deutschland (Hostenbach et al. 2011).
Die drei Dimensionen des Modells sind *Kognitive Prozesse*, *Komplexität* und *Kompetenzbereiche*. Auf letzterer Dimension wird auch die Modellkomponente zu *Bewertung* abgebildet.
Im ESNaS-Modell sind verschiedene Teilkompetenzen aus anderen Modellen verortet, nicht alle werden dabei jedoch dem Bereich *Bewertung* zugeschrieben. Die Teilkompetenz *Generieren und Reflektieren von Sachinformationen* des Göttinger Modells der Bewertungskompetenz wird hier dem Bereich der *Erkenntnisgewinnung* zugeordnet. Das *Argumentieren* aus dem Modell zur ethischen Urteilskompetenz findet seinen Platz im Bereich *Kommunikation* der Bildungsstandards (Hostenbach et al. 2011).
Innerhalb des Kompetenzbereichs *Bewertung* postuliert das ESNaS-Modell drei typische, aber nicht trennscharfe Kompetenzteilbereiche: *Bewertungskriterien*, *Handlungsoptionen* und *Reflexion*. Diese sind jeweils in fünf Komplexitätsniveaus unterteilt. Der Fokus kann in jedem Teilbereich auf persönliche, gesellschaftliche oder ethische Aspekte gelegt werden, um eine Vielfalt von Bewertungsperspektiven sicherzustellen (Hostenbach et al. 2011). Das Modell ist unabhängig von fachlichen Kontexten und beinhaltet ausschließlich quantitative Methoden.

3.2.4 Modell der epistemischen Kompetenz

Das Modell der ethischen Urteilskompetenz, das Göttinger Modell und das ESNaS-Modell beleuchten Bewertungskompetenz mit einem Fokus auf die Bewertung ethischer Aspekte. Dabei werden Werte und Normen integriert und persönliche Einstellungen gegenüber einem Thema bewertet.
Die fachdidaktische Forschung setzt sich mit dieser ethischen Perspektive auf Bewertungskompetenz umfangreich auseinander und es werden Unterrichtskonzepte zur Förderung der ethischen Kompetenz entwickelt (z. B. Bögeholz 2006). Der Bereich der Bewertungskompetenz beinhaltet jedoch noch eine weitere Dimension, die von Bildungsforscherinnen und -forschern deutlich weniger beleuchtet wird. Dabei handelt es sich um die *epistemische Kompetenz*, die die naturwissenschaftliche Wissensproduktion genauer unter die Lupe nimmt. Kötter & Hammann (2018) haben in Anlehnung an Leisen (2011) eine vorläufige Definition epistemischer Kompetenz aufgestellt:

> Epistemische Kompetenz ist der willentlich handelnde Umgang mit Wissen über naturwissenschaftliches Erkennen und naturwissenschaftliche Erkenntnisansprüche. Epistemisch kompetent wäre demnach eine Person, die in der Lage und willens ist, über Besonderheiten, Möglichkeiten und Grenzen naturwissenschaftlicher Wissensproduktion, sowie den gesellschaftlichen Status der Naturwissenschaften zu reflektieren, Stellung zu nehmen und naturwissenschaftliche Erkenntnisse, Methoden und Geltungsansprüche kritisch zu prüfen und zu würdigen.

Bei epistemischer Kompetenz geht es also nicht darum, verschiedene Standpunkte zu naturwissenschaftlichen Fragestellungen auf inhaltlicher Ebene ethisch zu bewerten. Stattdessen werden auf einer Meta-Ebene die Besonderheiten der naturwissenschaftlichen Erkenntnis vermittelt. Kötter & Hammann (2018) sprechen von dem Wissen über die *Natur der Naturwissenschaften*: „[...] not only *the* NOS [Nature of Science] but *about* NOS should be taught, too".
Die einzelnen Teilbereiche und Voraussetzungen für den Erwerb epistemischer Kompetenz haben Kötter & Hammann (2018) in einem Modell epistemischer Kompetenz (Abbildung 3.2) zusammengeführt und dieses in Analogie zum Modell ethischer Urteilskompetenz betrachtet.
Während die ethische Urteilskompetenz ein Bewusstsein der moralischen Relevanz sowie der eigenen Einstellung gegenüber einer ethischen Fragestellung voraussetzt, werden im Bereich der epistemischen Kompetenz ein *Bewusstsein für die lebensweltliche Bedeutung der Erkenntnistheorie* und ein *Bewusstsein eigener Einstellungen gegenüber Naturwissenschaften* gefordert.
Beide Bereiche erfordern eine Wissensbasis aus *Grundlagenwissen* über ethische bzw. epistemische Aspekte sowie über die *naturwissenschaftliche Praxis*.
Ethisch kompetente Personen sind in der Lage, ein eigenes begründetes Urteil zu fällen und verschiedene ethische Standpunkte zu beurteilen. Die entsprechende Dimension der epistemischen Kompetenz wird als *Fähigkeit und Bereitschaft, rationale Urteile zu fällen* dargestellt. Weiterhin wird beschrieben, dass epistemische Bewertungskompetenz ein hohes Maß an *Ambiguitätstoleranz und die Akzeptanz von Aporien* erfordert, während auf ethischer Seite eine Folgenreflexion stattzufinden hat. Querliegend zu allen Dimensionen steht im Modell der ethischen Urteilskompetenz die Fähigkeit zum Perspektivenwechsel und im epistemischen Bereich das kritische

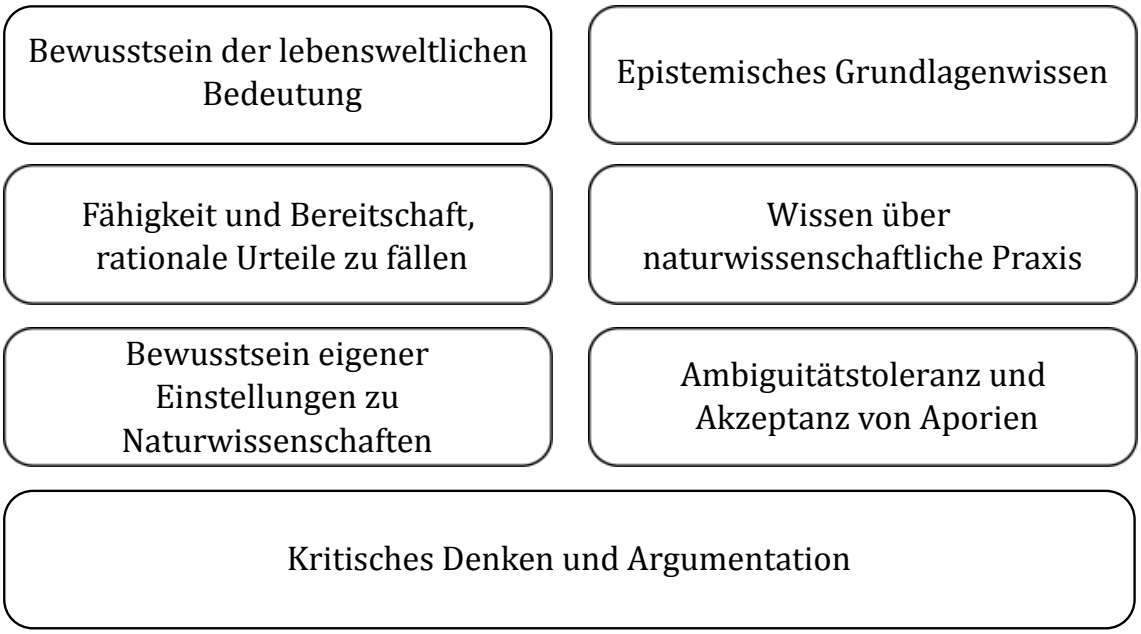


Abbildung 3.2: *Das Modell epistemischer Kompetenz nach Kötter & Hammann (2018).*

Denken. Argumentationsfähigkeiten sind über alle Bereiche der Bewertungskompetenz hinweg relevant. Diese Gegenüberstellung der beiden Modelle ist in der Abbildung 3.3 visualisiert.

3.3 Einflussfaktoren auf Bewertungskompetenz

Können bestimmte Faktoren als Prädiktoren für Bewertungskompetenz dienen? Diese Frage wird in der Kompetenzforschung untersucht. Für eine theoretische Herleitung einzelner Einflussfaktoren muss ein bestimmtes Kompetenzstrukturmodell zugrunde gelegt werden. Hostenbach & Walpuski (2013) verwenden für ihre Betrachtung das ESNaS-Modell (Kapitel 3.2.3).
Die Bewertungskompetenz wird von externen Faktoren beeinflusst, die das Wertesystem (Ajzen & Fishbein 1969) und die Kenntnisse (Sadler & Zeidler 2005) des Bewertenden betreffen. Die Faktoren werden in drei Gruppen unterteilt (fachliche Faktoren, überfachliche Faktoren, persönlich-individuelle Faktoren).
Fachliche Faktoren sind die Kenntnis von Fachwissen und der Umgang mit Fachwissen, überfachliche Faktoren bestehen in der Beurteilung der Datenqualität, Kenntnis von Bewertungsstrategien und Umgang mit Bewertungsstrategien. Persönlich-

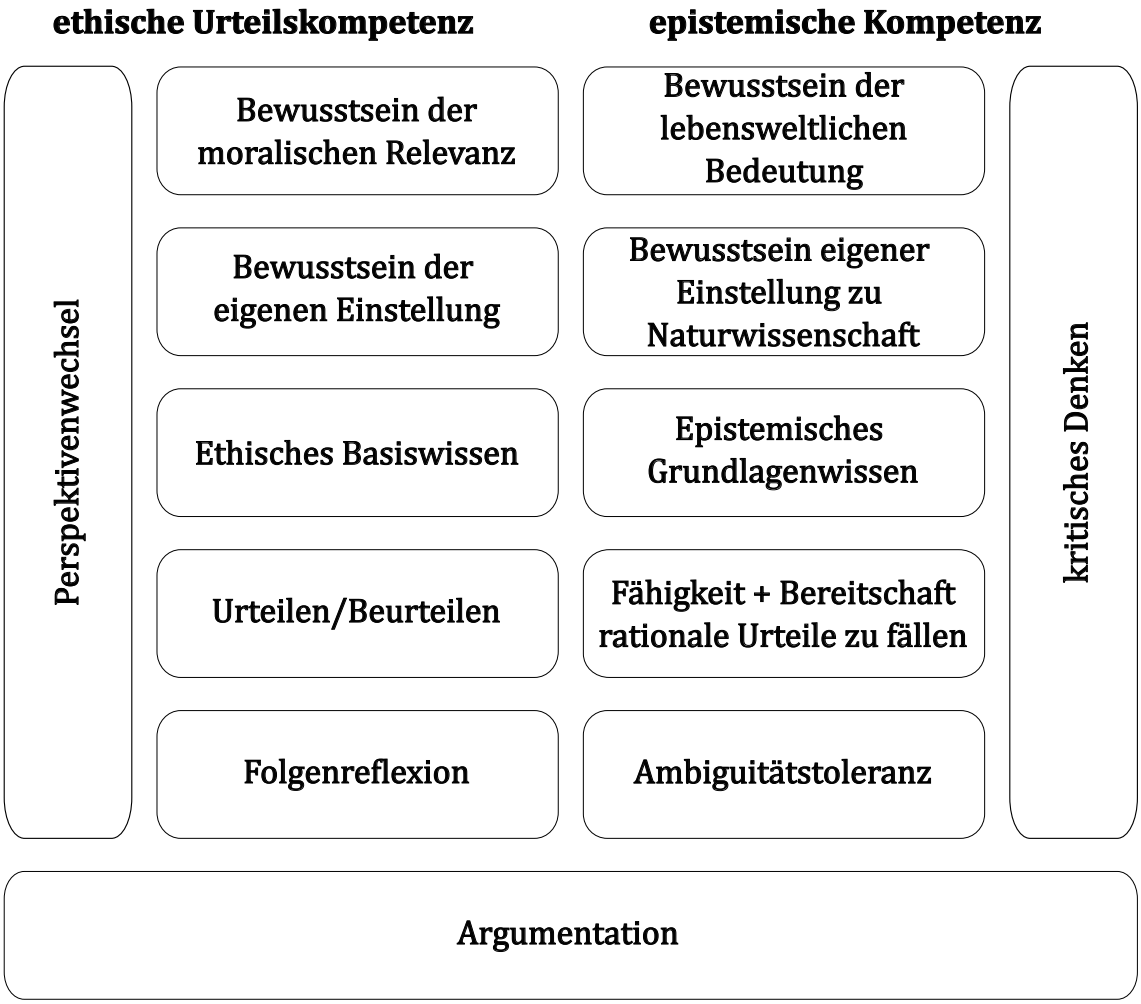


Abbildung 3.3: *Gegenüberstellung der Modelle ethischer Urteilskompetenz (Reitschert et al. 2007) und epistemischer Kompetenz (Kötter & Hammann 2018).*

individuelle Faktoren drücken sich in der eigenen Meinung und sozialer Erwünschtheit aus. Eine umfangreiche theoretische Herleitung der Faktoren ist bei Hostenbach & Walpuski (2013) zu finden.

Hostenbach & Walpuski (2013) zeigen, aufbauend auf der Herleitung der Einflussfaktoren, dass es sich bei Bewertungskompetenz in Abhängigkeit von verschiedenen Fachinhalten um ein mehrdimensionales Konstrukt handelt. Das bedeutet, dass die Bewertungskompetenz im Fach Biologie weder mit Bewertung im Fach Chemie noch mit Bewertung in außerfachlichen Kontexten gleichzusetzen ist. Schüler:innen, die im außerfachlichen Kontext über hoch ausgeprägte Bewertungskompetenz verfügen,

können in einem fachinhaltlichen Kontext (Biologie oder Chemie) dennoch geringe Bewertungskompetenzen aufweisen. Der Unterschied lasse sich darin begründen, dass Lernende im alltäglichen Leben häufig Entscheidungen treffen müssen, während der Fachunterricht nur selten explizite Bewertung erfordert (Hostenbach & Walpuski 2013).
In ihrer empirischen Untersuchung identifizieren die Autor:innen drei Hauptfaktoren, die sich als Prädiktoren für Bewertungskompetenz eignen. Die *außerfachliche Anwendung von Bewertungsstrategien* ist der stärkste Prädiktor. In Kompetenz-Tests erzielen Lernende in dieser Kategorie recht gute Ergebnisse, somit lässt sich schlussfolgern, dass Lernende zwar über allgemeine Bewertungsstrategien verfügen, diese jedoch für die Verwendung in fachlichen Kontexten spezifisch angepasst werden müssen (Eggert 2008; Hostenbach & Walpuski 2013).
Weitere Prädiktoren stellen die *Kenntnis von Fachwissen* und die *Einschätzung von Datenqualität* dar. Es ist anzumerken, dass die empirische Untersuchung nur eine begrenze Aussagekraft über das Verhalten in authentischen Entscheidungssituationen liefert, da im Rahmen der Studie lediglich fiktive Bewertungssituationen verwendet wurden, die von den Schüler:innen möglicherweise analytischer behandelt wurden. Für eine richtungsweisende Einschätzung lassen sich die Ergebnisse jedoch verwenden. Als Folgerung aus den Erkenntnissen der Studie formulieren die Autor:innen das Desiderat, im naturwissenschaftlichen Unterricht das Bewerten häufiger zu üben.

3.4 Bedeutung für den naturwissenschaftlichen Unterricht

3.4.1 Allgemeine Bedeutung

Der naturwissenschaftliche Unterricht hat seit langer Zeit eine besondere Verantwortung für die Reflexion wissenschaftlicher Forschung. Er ist kritisch-emanzipatorisch anzulegen und zielt auf Partizipation: Um an privaten und öffentlichen Diskursen teilnehmen zu können, benötigen Lernende eine ethisch-moralische Urteils- und Handlungsfähigkeit (Dittmer et al. 2016; Klieme et al. 2003, S. 65). Laut Klafki (1994) und Dittmer et al. (2016) geht es dabei um „Kritikbereitschaft und -fähigkeit, Argumentationsbereitschaft und -fähigkeit, Empathie und die Fähigkeit zu vernetztem Denken“.

Impliziert sind diese Forderungen auch in dem allgemeinen Bildungsziel der „Scientific Literacy“ (Prenzel et al. 2007), das eine allgemeine naturwissenschaftliche Grundbildung für alle Lernenden fordert. Der Beitrag des naturwissenschaftlichen Unterrichts zur *Scientific Literacy* besteht darin, „naturwissenschaftliche Basiskonzepte zur Interpretation von Natur, Mensch und einer durch Naturwissenschaften und Technik gestalteten Welt“ (BLK 1998) zu vermitteln.
Ein moralpsychologisch-soziologischer Blick auf die Bedeutung von Bewertungskompetenz ergibt ähnliche Sichtweisen. Lernende sollen Probleme rational begründet, argumentierend und reflektierend lösen (Luther-Kirner 2009). Das Zwei-Prozess-Modell des Bewertens (Smith & DeCoster 2000) unterscheidet zwischen einem Denkprozess und einem kognitiv aufwändigeren Verarbeitungsprozess und erinnert damit an die Theorie der Zwei Systeme nach Kahneman (2011), siehe auch Kapitel 1.3.5 dieser Arbeit. Denkprozesse (analog: System 1) verlaufen schnell und einfach, simultan zur Wahrnehmung. Sie haben heuristischen Charakter und sind intuitiv gesteuert. Die Wahl einer intuitiven Strategie, also die Verarbeitung eines Problems mittels Denkprozess, macht direkt handlungsfähig und liefert häufig sogar bessere Ergebnisse als die Lösung eines Problems durch kognitiv aufwändige Verarbeitungsprozesse (bzw. System 2) (Czerlinski et al. 1999; Dittmer et al. 2016).
Komplexe Problemlagen müssen dennoch rational abgewogen und reflektiert werden. Die moralpsychologische Perspektive auf Bewertungskompetenz stellt die Forderung dar, dass Lernende sich der eigenen Einstellungen, Intuitionen und Emotionen bewusst werden und dieses Bewusstsein in die Bewertungsprozesse mit einfließen lassen. Diese Betrachtung, wie auch die bildungstheoretische Sicht auf die Förderung von Bewertungskompetenz mündet in einer Darstellung der *conceptual change*-Theorie (Kapitel 3.5) zur Umstrukturierung von Schülervorstellungen.

3.4.2 Bewertungskompetenz in den schulischen Lehrplänen

Die schulischen Lehrpläne der naturwissenschaftlichen Fächer fordern explizit, dass im naturwissenschaftlichen Unterricht neben *Fachwissen*, *Kommunikation* und *Erkenntnisgewinnung* auch *Bewertungskompetenz* gefördert wird.
Wie oben zitiert, definieren die bundesweiten Bildungsstandards der Fächer Chemie, Biologie und Physik Bewertungskompetenz als „Erkennen und Bewerten naturwissenschaftlicher Sachverhalte in verschiedenen Kontexten“. Dazu gehören auch die

„Kenntnis und Reflexion der Beziehungen zwischen Naturwissenschaft, Technik, Individuum und Gesellschaft“ sowie das Übertragen von Fachkenntnissen auf neue Fragestellungen und Erfassen von Problemen in realen Situationen. Auch Perspektivenwechsel, Interdisziplinarität und eigenständige Problemlösung werden gefordert.
Die Kernlehrpläne der verschiedenen Schulformen und -fächer in Nordrhein-Westfalen halten ebenfalls eigene Definitionen der Bewertungskompetenz bereit. Ein umfänglicher Überblick wird an dieser Stelle nicht ausformuliert, befindet sich jedoch in tabellarischer Form im Anhang A.0.1. Dort wird das Konstrukt *Bewertungskompetenz* in den folgenden Kernlehrplänen miteinander verglichen: Gymnasium Sek I in den Fächern Biologie, Chemie, Physik; Gesamtschule im Fach Naturwissenschaften; Realschule in den Fächern Biologie, Chemie, Physik; Bildungsstandards für den mittleren Schulabschluss und die Bildungsstandards Abitur.
Der Vergleich beinhaltet folgende Gesichtspunkte:

- Die Definition der Bewertungskompetenz als die Fähigkeit, [...]
- Nennen von kriteriengeleitetem Bewerten
- Fordern von rationalen Entscheidungen
- Bezugnahme auf ethische Aspekte
- Berücksichtigung von Argumentationsmustern
- Formulierung, dass Bewertungskompetenz für gesellschaftliche Teilhabe wichtig ist
- Erwähnen der naturwissenschaftlichen Grundbildung
- Hinweise auf Fake-News, Verschwörungstheorien etc.
- Erklären von Unwissenschaftlichkeit

Das Ergebnis dieses Vergleichs ist eine Tabelle mit 17 Spalten und 9 Zeilen. Eine genauere Betrachtung zeigt, dass in den meisten Fällen zumindest die Fächer einer Schulform an einem Strang ziehen und eine ähnliche Ausdifferenzierung der Bewertungskompetenz wählen.
Allerdings gibt es Unterschiede zwischen Sekundarstufe I und II des Gymnasiums oder auch zwischen Gymnasium, Gesamtschule, Realschule und den Bildungsstandards. Ein derart umfangreicher „Flickenteppich“ an Definitionen erschwerte das Bil-

den einer passenden Arbeitsdefinition für diese Arbeit. In Kombination mit Ausführungen der einschlägigen Fachliteratur wurde die Bewertungskompetenz jedoch in den vorherigen Abschnitten eingegrenzt und für die Verwendung in dieser Arbeit abgesteckt.

In allen Lehrplänen wird jedoch das *kriteriengeleitete Bewerten* als zentral angesehen. Formulierungen wie: „[...] Handlungsmöglichkeiten zu entwickeln und auf der Grundlage von Kriterien gegeneinander abzuwägen [...]“ (KLP 2019, S. 14) sind häufiger zu entdecken.

In der Literatur wird für das *Treffen rationaler Entscheidungen* ein hohes Maß an Bewertungskompetenz vorausgesetzt (Rost et al. 2003), dieser Aspekt findet sich auch in den Lehrplänen, zum Beispiel im KLP der Gesamtschulen für das Fach Naturwissenschaften (KLP 2013, S. 19): „Auf dieser Grundlage ist es möglich, Entscheidungen zu finden und dafür zielführend zu argumentieren und Position zu beziehen“.

Die Trennung zwischen ethischen und epistemischen Aspekten, siehe dazu die Kompetenzstrukturmodelle im Kapitel 3.2, ist in den Lehrplänen nicht wiederzufinden. Häufig gibt es Formulierungen wie: „[Es ist] die Kenntnis und Berücksichtigung von normativen und ethischen Maßstäben bedeutsam, nach denen Interessen und Folgen naturwissenschaftlicher Forschung beurteilt werden können“ (KLP 2014, S. 18). Der Fokus verschiebt sich nur langsam in Richtung der epistemischen Bedeutung von Bewertungskompetenz, zum Beispiel im neueren Kernlehrplan für „G9“, also der Fassung aus dem Jahr 2019 (KLP 2019).

Die Konzeptualisierung von Bewertungskompetenz in dieser Fassung für das Fach Chemie wird nun etwas detaillierter beleuchtet.

Der Kernlehrplan für die Sekundarstufe I von Gymnasien für das Fach Chemie in NRW (KLP 2019) unterscheidet, wie auch andere Kernlehrpläne, *Kompetenzbereiche* von *Inhaltsfeldern* und *Kompetenzerwartungen*. Kompetenzbereiche strukturieren einzelne Teiloperationen, während Inhaltsfelder „die im Unterricht verbindlichen und unverzichtbaren Gegenstände“ spezifizieren und sich damit inhaltlichen Schwerpunkten widmen. Zusammengenommen beschreiben sie die *Komptenzerwartungen* als fachliche Anforderungen und intendierte Lernergebnisse.

Zu den Kompetenzbereichen zählen *Umgang mit Fachwissen*, *Kommunikation*, *Erkenntnisgewinnung* und *Bewertung*. Die dort aufgeführte Definition von Bewertungskompetenz wurde bereits oben erläutert und bezieht sich vor allem darauf, „sachlich fun-

diert und wertebasiert zu begründeten Entscheidungen zu kommen". Auffällig ist, dass diese aktualisierte Version stärker als die vorherige Version in den Fokus rückt, dass Bewerten *kriteriengeleitet* geschehen soll.
Die Inhaltsfelder konkretisieren die *Gegenstände*, an denen die Kompetenzen im Unterricht erworben werden sollen. Allerdings sind dort nicht viele Bezüge zur Bewertungskompetenz zu finden. Im Inhaltsfeld 9 „Saure und alkalische Lösungen" wird genannt, dass fundierte Kenntnisse in diesem Bereich die Beurteilung von Aussagen in Medien und Werbungen ermöglichen.
Das Inhaltsfeld 10 zur Organischen Chemie behandelt Kohlenwasserstoffverbindungen als Energieträger und bietet damit Anknüpfungspunkte für Diskussionen über die Effizienz chemischer Reaktionen oder die Herkunft der beteiligten Rohstoffe. Im Kernlehrplan wird beschrieben, dass dies die Urteilskraft in gesellschaftspolitisch relevanten Fragen stärke (KLP 2019, S. 16).

In den *übergeordneten Kompetenzerwartungen* finden sich neben den Erwartungen im Kompetenzbereich *Bewerten* auch in den Bereichen *Kommunikation* und *Erkenntnisgewinnung* Formulierungen, die auf rationale Entscheidungen und kritisches Hinterfragen abzielen. Folgende übergeordnete Kompetenzerwartungen werden als relevant für das Verständnis von Bewertungskompetenz in dieser Arbeit angesehen:

E7 Naturwissenschaftliches Denken und Arbeiten: In einfachen chem. Zusammenhängen Schritte der naturwissenschaftlichen Erkenntnisgewinnung nachvollziehen und Aussagen konstruktiv kritisch hinterfragen.

K4 Argumentation: Eigene Aussagen fachlich sinnvoll begründen, faktenbasierte Gründe von intuitiven Meinungen unterscheiden sowie bei Unklarheiten sachlich nachfragen.

B1 Fakten- und Situationsanalyse: In einer einfachen Bewertungssituation chemische Fakten nennen sowie die Interessen der Handelnden und Betroffenen beschreiben.

B2 Bewertungskriterien und Handlungsoptionen: Bewertungskriterien und Handlungsoptionen benennen.

B3 Abwägung und Entscheidung: Kriteriengeleitet eine Entscheidung für eine Handlungsoption treffen.

B4 Stellungnahme und Reflexion: Bewertungen und Entscheidungen begründen.

In diesen Kompetenzerwartungen wird klar gefordert, dass Schüler:innen faktenbasiert und auf der Grundlage von Kriterien Entscheidungen treffen und ihr Verhalten bzw. ihre Positionen reflektieren sollen. Wie diese Forderung im Chemieunterricht genau umgesetzt werden soll, könnte ein Blick auf die konkretisierten Kompetenzerwartungen verraten, denn diese beschreiben für jedes Inhaltsfeld jeweils für alle vier Kompetenzbereiche konkrete Inhalte, an denen die Kompetenzen erarbeitet werden können. In den Tabellen 3.1 und 3.2 sind alle konkretisierten Kompetenzerwartungen für den Bereich *Bewertung* in den Progressionsstufen 1 und 2 aufgeführt.

Tabelle 3.1: *Zusammenstellung aller im Kernlehrplan Chemie (KLP 2019) genannten konkretisierten Kompetenzerwartungen für den Kompetenzbereich Bewertung. IF = Inhaltsfeld. ÜK = übergeordnete Kompetenzerwartungen.*

IF	konkretisierte Kompetenzerwartung	ÜK
1	die Verwendung ausgewählter Stoffe im Alltag mithilfe ihrer Eigenschaften begründen	B1, K2
2	die Bedeutung chemischer Reaktionen in der Lebenswelt begründen	B1, K4
3	in vorgegebenen Situationen Handlungsmöglichkeiten zum Umgang mit brennbaren Stoffen zur Brandvorsorge sowie mit offenem Feuer zur Brandbekämpfung bewerten und sich begründet für eine Handlung entscheiden; Vor- und Nachteile einer ressourcenschonenden Energieversorgung auf Grundlage der Umkehrbarkeit chemischer Reaktionen am Beispiel von Wasser beschreiben	B2, B3, K4 B1
4	die Bedeutung des Metallrecyclings im Zusammenhang mit Ressourcenschonung und Energieeinsparung beschreiben und auf dieser Basis das eigene Konsum- und Entsorgungsverhalten bewerten; Maßnahmen zum Löschen von Metallbränden auf der Grundlage der Sauerstoff-Übertragungsreaktion begründet auswählen	B1, B4, K4 B3

Der Blick in die übergeordneten Kompetenzerwartungen zeigt, dass dem Bereich Bewertung eine ganze Reihe an Erwartungen zugeordnet sind. Allerdings steht hier das „fachliche Bewerten“ im Vordergrund, also z. B. das Abwägen von Risiken und Nutzen beim Umgang mit sauren und alkalischen Lösungen im Inhaltsfeld 9. Auch das *Begründen* der Relevanz chemischer Reaktionen oder ähnlicher Zusammenhänge wird unter diesen Erwartungen gefasst.

Nicht alle der Kompetenzerwartungen werden dem zuvor definierten Verständnis

Tabelle 3.2: *Zusammenstellung aller im Kernlehrplan Chemie (KLP 2019) genannten konkretisierten Kompetenzerwartungen für den Kompetenzbereich Bewertung der zweiten Progressionsstufe. IF = Inhaltsfeld. ÜK = übergeordnete Kompetenzerwartungen.*

IF	konkretisierte Kompetenzerwartung	ÜK
5	vor dem Hintergrund der begrenzten Verfügbarkeit eines chemischen Elements bzw. seiner Verbindungen Handlungsoptionen für ein ressourcenschonendes Konsumverhalten entwickeln	B3
6	unter Umwelt- und Gesundheitsaspekten die Verwendung von Salzen im Alltag reflektieren	B1
7	Kriterien für den Gebrauch unterschiedlicher elektrochemischer Energiequellen im Alltag reflektieren	B2, B3, K2
8	Informationen für ein technisches Verfahren zur Industrierohstoffgewinnung aus Gasen mithilfe digitaler Medien beschaffen und Bewertungskriterien auch unter Berücksichtigung der Energiespeicherung festlegen; unterschiedliche Darstellungen von Modellen kleiner Moleküle auch mithilfe einer Software vergleichend gegenüberstellen	B2, K2 B1, K1, K3
9	beim Umgang mit sauren und alkalischen Lösungen Risiken und Nutzen abwägen und angemessene Sicherheitsmaßnahmen begründet auswählen; Aussagen zu sauren, alkalischen und neutralen Lösungen in analogen und digitalen Medien kritisch hinterfragen	B3 B1, K2
10	Vor- und Nachteile der Nutzung von fossilen und regenerativen Energieträgern unter ökologischen, ökonomischen und ethischen Gesichtspunkten diskutieren; am Beispiel eines chemischen Produkts Kriterien hinsichtlich Verwendung, Ökonomie, Recyclingfähigkeit und Umweltverträglichkeit abwägen und im Hinblick auf die Verwendung einen eigenen sachlich fundierten Standpunkt beziehen	B4, K4 B3, B4, K4

der Bewertungskompetenz gerecht. In manchen Fällen werden zwar „ökologische, ökonomische und ethische Gesichtspunkte“ erwähnt oder es wird gefordert, unter „Umwelt- und Gesundheitsaspekten“ etwas zu bewerten. Kontroverse Kontexte oder konkrete Umsetzungsbeispiele befinden sich im Kernlehrplan jedoch nicht.

3.5 Bewerten als Ausgangspunkt für *conceptual change*

Aus bildungstheoretischer Sicht ist die Förderung von Bewertungskompetenz ein besonders geeignetes Mittel, um bei Lernenden eine „Krise“ als Anlass innerer Reflexion hervorzurufen (Dittmer et al. 2016).
Bildung wird als transformatorischer Prozess des Selbst- und Weltverhältnisses verstanden, durch den Individuen zu neuen Sichtweisen und Handlungsmöglichkeiten gelangen (Koller 1999, S. 17; Dittmer et al. 2016). Dieser transformatorische Prozess wird initiiert, wenn Lernende etwas erleben, das über ihre bisherigen Erfahrungen hinaus reicht oder wenn ein Ereignis anders eintritt als erwartet. In diesen Fällen geraten Lernende in eine „Krise“ (Combe & Gebhard 2012; Dittmer et al. 2016).
Der Begriff der Krise kann analog zur *conceptual change*-Theorie nach Strike & Posner (1982) verstanden werden. Nach dieser Theorie findet Lernen nicht als reine Akkumulation von Wissen statt, sondern als Transformation von bereits bestehenden Kenntnissen. Strike & Posner (1982) zeigen, dass jede Schülerin und jeder Schüler ihr bzw. sein eigenes *Konzept* im Umgang mit Aussagen und Hypothesen besitzt: „Individuals approach any inquiry with a prior set of concepts“. Dieses Konzept widerspricht häufig der fachlich anerkannten Sichtweise.
Die bei den Lernenden vorhandenen Konzepte, die auch als Schülervorstellungen (Gropengießer & Marohn 2018; Barke et al. 2015) bezeichnet werden, können durch einen *conceptual change* verändert werden. Ein Konzeptwechsel ist jedoch nur unter bestimmten Bedingungen möglich:
Zunächst ist es wichtig, dass eine *Unzufriedenheit mit dem bestehenden Konzept* hergestellt wird. Wenn ein Erklärungsmuster bei den Lernenden an seine Grenzen stößt, weil es gewisse Aspekte nicht mehr abdeckt oder mit ihnen in Konflikt steht, dann entsteht diese Unzufriedenheit (Strike & Posner 1982).
Häufig stehen Schülervorstellungen den wissenschaftlichen Vorstellungen konträr

gegenüber und es wird versucht, die Schülerinnen und Schüler mithilfe eines kognitiven Konflikts von der wissenschaftlichen Sicht zu überzeugen (Duit 1995). Der kognitive Konflikt kann laut Duit auf drei Arten hervorgerufen werden:

- Hypothesen der Lernenden stimmen nicht mit einem experimentellen Ergebnis überein,
- Es herrscht ein Konflikt zwischen Schülervorstellung und wissenschaftlicher Vorstellung,
- Verschiedene Lernende besitzen konträre Vorstellungen.

Neben der Unzufriedenheit mit eigenen Vorstellungen ist es wichtig, dass das neue Konzept *verständlich*, *plausibel* und *fruchtbar* für weitere Forschung erscheint (Strike & Posner 1982).
Dieses Verfahren zur Änderung von Konzepten ist auch heute noch anerkannt und wird in jüngerer Zeit zur Materialentwicklung verwendet (Macbeth 2000; Egbers & Marohn 2013).

3.6 Bewerten unter Kontroversität: Socio-Scientific Issues

Socio-Scientific Issues sind Problemstellungen, die eine hohe gesellschaftliche Relevanz besitzen, kontrovers diskutiert werden und umfangreiche wissenschaftliche Informationen zur Bekräftigung verschiedener Standpunkte erfordern (Sadler 2004). Sowohl gesellschaftliche als auch naturwissenschaftliche Aspekte spielen in diesen Problemstellungen eine Rolle.
In der internationalen Literatur wird bereits seit Längerem die Integration von socio-scientific issues (SSI) in den naturwissenschaftlichen Unterricht gefordert. In den 1980er Jahren entwickelte sich die STS-Bewegung („*Science, Technology and Society*“) mit dem Versuch, die Grenzen zwischen Naturwissenschaft, technologischer Entwicklung und der Gesellschaft im naturwissenschaftlichen Unterricht zu explizieren (Sadler 2004). Aufgrund des erhöhten Lebensweltbezugs von *socio-scientific issues* und dem spezifischeren Fokus auf naturwissenschaftliche Problemstellungen hat sich die SSI-Bewegung jedoch stärker durchgesetzt (Pedretti & Hodson 1995).

Kolstø et al. (2006) stellen im Zusammenhang mit SSI heraus, dass Lernende in der Lage sein sollen, zentrale naturwissenschaftliche Behauptungen und Argumente zu bewerten und zusammen mit gesellschaftlichen Überlegungen in eine durchdachte Entscheidungsfindung einfließen zu lassen. Diese Forderungen weisen Parallelen zu den Vorgaben zur Förderung der Bewertungskompetenz auf, die im Kapitel 3.1 dargestellt sind.

Sadler (2004) beschreibt vier Hauptbereiche der SSI-Forschung und bezieht sich dabei auf informale Logik: 1. gesellschaftlich-naturwissenschaftliches Argumentieren, 2. Beziehungen zwischen NOS-Konzeptualisierungen und SSI-bezogener Entscheidungsfindung, 3. Bewertung von SSI-betreffenden Informationen und 4. Einfluss des Konzeptverstehens auf logisches Denken bzw. informal reasoning. Für alle Teilbereiche hat Sadler (2004) Ergebnisse aus der Forschung zusammengetragen.

Es wurde herausgefunden, dass die Argumentationsfähigkeiten bezogen auf SSI bei den meisten Lernenden nur gering ausgeprägt sind (Sadler 2004).

Die Art und Weise, wie Lernende NOS (*Nature of Science*) wahrnehmen, hat nicht immer einen direkten Einfluss auf das Entscheidungsverhalten bezüglich SSI. Einige NOS-Aspekte scheinen jedoch Auswirkungen darauf zu haben. Es wird daher empfohlen, *Nature of Science*-Aspekte zur Förderung der gesellschaftlich-naturwissenschaftlichen Entscheidungsfindung in die Curricula des naturwissenschaftlichen Unterrichts zu integrieren (Bell & Lederman 2003; Zeidler et al. 2002; Abd-El-Khalick et al. 1998).

Empirische Studien haben herausgefunden, dass vielen Lernenden Kompetenzen im Bezug auf die Bewertung von wissenschaftlichen Informationen fehlen. Die Studienteilnehmer:innen erkannten die Notwendigkeit der fachgerechten Evaluation, wiesen jedoch nicht die dazu nötigen Fähigkeiten auf (Sadler 2004). S. Rundgren & C. Rundgren (2010) stellen heraus, dass die Thematisierung von SSI im Unterricht einen angemessenen Rahmen bietet, um diese Fähigkeiten zu fördern.

Weiterhin wird beschrieben, dass das konzeptuelle Verstehen der einzelnen Materialien zu SSI einen wichtigen Faktor zur SSI-bezogenen Entscheidungsfindung darstellt. Diese Ergebnisse seien jedoch vorsichtig zu betrachten, da erweiterte Forschung in dem Bereich noch fehlt (Tytler 2001; Hogan 2002).

Bisher bestehen einzelne naturwissenschaftliche Unterrichtskonzepte, die SSI berücksichtigen. Ein Beispiel ist das Konzept *Science Education for Public Unterstanding*,

kurz SEPUP (Wilson & Sloane 2000). Dort wird Entscheidungsfindung im Hinblick auf Biologie-bezogene SSI thematisiert. Lernende führen *trade-offs* durch, also das Abwägen von Vor- und Nachteilen verschiedener Lösungsalternativen. Außerdem planen sie naturwissenschaftliche Untersuchungen, führen sie durch und analysieren die Ergebnisse. Fünf SEPUP-Variablen werden auf verschiedenen Niveaustufen konkretisiert, sodass Schülerantworten kategorisiert und in Niveaus eingestuft werden können.

Im Rahmen von SEPUP ist beispielsweise das Programm „Issues, Evidence and You" entstanden, das 60 *activities* für die Sekundarstufe I zu je einer bis drei Unterrichtsstunden enthält und dabei verschiedene SSI im Fach Physik behandelt.

Im deutschsprachigen Raum entwickelten Marks et al. (2014) das *gesellschaftskritisch-problemorientierte Unterrichtsverfahren*. Sie diskutieren, „[...] wie man solche Themen [gemeint sind socio-scientific issues] auch im naturwissenschaftlichen Unterricht der Fächer Chemie und Physik aus einer gesellschaftlichen Perspektive behandeln kann [...]". In dem Konzept werden „authentische, relevante und kontroverse Diskussionen aus der Gesellschaft über naturwissenschaftlich-technische Fragestellungen in den Mittelpunkt des Unterrichts" gestellt (Marks & Eilks 2009; Marks et al. 2014). Es wird das Ziel verfolgt, Schüler:innen eine allgemeine Bildung, die auf gesellschaftliche Teilhabe vorbereitet, bereitzustellen.

Die Unterrichtseinheit verläuft in fünf Phasen: 1. Zugang/Analyse der Kontroverse, 2. Fachliche Klärung, 3. Wiederaufgreifen der kontroversen Problemlage, 4. Erarbeitung/Diskussion verschiedener Perspektiven, 5. Metareflexion.

Ein ähnliches fünf-phasiges Vorgehen unter Berücksichtigung kontroverser Themen verfolgt die von Jungkamp (2021) entwickelte Unterrichtskonzeption mit dem Namen choice^{2}reflect. Entscheidende Unterschiede liegen jedoch in den mittleren Teilen der Konzepte. Während Marks et al. (2014) eine *Fachliche Klärung* als zweite Phase einsetzen, liegt der Fokus bei Jungkamp (2021) auf der Erarbeitung von wissenschaftlichen Prüfkriterien. Das Ziel dieser Arbeit besteht in der Weiterentwicklung des Konzepts. Aus diesem Grund wird das Konzept in einem eigenen Kapitel (4) beleuchtet.

3.7 Empirische Befunde zum Einsatz von SSI

Die Entwicklung von Unterrichtskonzepten in naturwissenschaftlichen Fächern, die naturwissenschaftliche Kontroversen oder SSI berücksichtigen, schreitet voran. Die Unterschiede zwischen den Konzepten zeigen, dass nicht immer klar ist, wie mit konfligierenden Informationen verfahren werden soll.
Eine *fachliche Klärung* wie in Phase 2 des gesellschaftskritisch-problemorientierten Unterrichtsverfahrens erweist sich als schwierig, wenn zu einer Problemstellung verschiedene Standpunkte mit sich widersprechenden Informationen vorliegen. Die Schüler:innen müssten Informationen, mit denen sie konfrontiert werden, selbstständig prüfen und deren Güte einschätzen können.
Die *science*-Dimension von SSI enthält also häufig ungesicherte, vorläufige und umstrittene Informationen, die geprüft werden müssen. Damit Lernende sich eine Meinung bilden können, müssen sie einen ähnlichen Prüfungsvorgang durchführen wie auch Experten. Die Lernenden als wissenschaftliche Laien haben allerdings nur beschränktes epistemisches Wissen. Daher haben sich Kolstø et al. (2006) die Frage gestellt, welche Wissenschaftskriterien für Laien zumutbar sind.
In einer Studie mit 89 Lehramtsstudierenden in Norwegen wurde untersucht, in welcher Weise die Studierenden sich mit eigenständig recherchierten Informationen eines selbstgewählten SSI auseinandersetzen und welche Kriterien sie zur Bewertung der Informationen heranziehen.
Die Forschenden haben herausgefunden, dass die Studierenden sich auf empirische und theoretische Eignung der vorliegenden Informationen beziehen, sowie ihre Vollständigkeit überprüfen. Weiterhin beachten sie soziale und manipulative Strategien. Die untersuchten Studierenden nannten keine literaturbasierten Kriterien.
Unsicherheit der wissenschaftlichen Fakten begründeten die Studierenden häufig mit unzureichender Informationslage und erkannten damit nicht an, dass es zu einigen SSI keine eindeutige wissenschaftliche Datenlage gibt (Kolstø et al. 2006; Yang 2004).
Die Ergebnisse zeigen, dass Studierende versuchen, die Qualität wissenschaftlicher Informationen zu beurteilen und zu prüfen. Allerdings stützen sie sich dabei meist nur auf die gegebenen Informationen und führen kein „crosschecking“ durch. Es wird ein hohes Vertrauen in die Wissenschaftler:innen gelegt.
Die Autor:innen der Studie argumentieren, dass Laien bis zu einem bestimmten Grad

fähig sind, wissenschaftliche Aussagen auf empirische Güte zu überprüfen und fordern, dass im Unterricht Wissenschaftskriterien und Kontextfaktoren stärkere Beachtung finden.

3.8 Fazit zur Bewertungskompetenz

Bewertungskompetenz hat viele Facetten. Das hat sich in den Ausführungen der letzten Abschnitte gezeigt. Schon die Definition des Begriffs unter Einbeziehen von Bildungsstandards und Kernlehrplänen, aber auch unter Berücksichtigung relevanter Fachliteratur zum Beispiel von Rost et al. (2003) hat die verschiedenen Perspektiven der Bewertungskompetenz offengelegt.

Anhand von Strukturmodellen wurden Teildimensionen benannt und Anwendungsbereiche des Bewertens aufgezeigt. Neben dem Modell der ethischen Urteilskompetenz und dem Göttinger Modell der Bewertungskompetenz wurde auch ein Modell zur Evaluation der Bildungsstandards (das ESNaS-Modell) betrachtet.

Diese Modelle fokussieren stark auf *ethische* Aspekte des Bewertens; Werte und Normen sowie persönliche Einstellungen gegenüber einem Thema spielen dabei eine große Rolle. Es hat sich aber gezeigt, dass neben ethischem Bewerten auch epistemische Überlegungen mit stärkerem Bezug zu *Nature Of Science* einbezogen werden müssen.

Bewertungskompetenz wird von externen Faktoren beeinflusst, die vor allem von Hostenbach et al. (2011) charakterisiert wurden. Als Haupt-Prädiktor hat sich die *außerfachliche Anwendung von Bewertungsstrategien* herausgestellt. Das verdeutlicht die enorme Relevanz von *Bewertung* im Alltag: Lernende müssen befähigt werden, nicht nur innerhalb des schulischen Unterrichts, sondern auch in ihrem Alltag gute Entscheidungen auf Basis von Bewertungsstrategien zu treffen.

Der naturwissenschaftliche Unterricht kann und sollte einen Beitrag zur Ausbildung von Bewertungskompetenz bei Schüler:innen leisten. Das wird auch durch Bildungsstandards und Lehrpläne gefordert, wenn auch nur marginal konkretisiert.

In der Literatur werden *socio-scientific issues* als geeignetes Mittel angesehen, um Bewertungskompetenz im naturwissenschaftlichen Unterricht zu fördern. Empirische Befunde und bereits entwickelte Unterrichtskonzepte zu SSI bestätigen diese Haltung.

Die hier dargelegte Bedarfsanalyse im Zusammenhang mit der Bestandsaufnahme zur Vermittlung von Bewertungskompetenz hat gezeigt, dass vor allem im Bezug auf kontrovers diskutierte Themen, zu denen keine klare Position mit wissenschaftlich gesicherten Informationen besteht, weitere Unterrichtskonzepte nötig sind.
Jungkamp (2021) stellt mit choice^{2}reflect ein Unterrichtskonzept bereit, das zur Bewertung kontroverser Themen befähigen soll, indem Lernende anhand von Wissenschaftskriterien Aussagen und Studien bewerten. Im Rahmen dieses Projekts wird das Unterrichtskonzept auf dem Weg zu einer praxistauglichen Intervention weiterentwickelt.

4 Unterrichtskonzeption choice^{2}reflect

4.1 Idee

Die theoretischen Hintergründe zur Entscheidungstheorie haben gezeigt, dass reale Entscheidungen nur begrenzt rational getroffen werden. Individuen haben eine begrenzte Kapazität bei der Informationsverarbeitung und beziehen nicht alle vorhandenen Informationen in ihre Denkmuster ein. Weiterhin spielen Emotionen eine große Rolle beim Treffen von Entscheidungen.
Der Ansatz der Zwei Systeme nach Kahneman (2011) wurde vorgestellt und es wurde herausgearbeitet, dass Individuen im Umgang mit den Unterschieden zwischen System 1 und System 2 geschult werden müssen (siehe Kapitel 1).
Weiterführend spannte sich der Bogen über Inhalte der Wissenschaftstheorie bis hin zur Bewertungskompetenz. Die Unterschiede zwischen wahrheitsfähigen und nicht wahrheitsfähigen Aussagen und daraus folgende Implikationen für die Möglichkeit, aus bestimmten Aussagensätzen wissenschaftliche Erkenntnisse abzuleiten wurden für die Formulierung von Gütekriterien herangezogen.
Diese Gütekriterien werden in der obersten Gliederungsebene zunächst nach *Validität*, *Reliabilität* und *Objektivität* aufgeteilt und bilden auf einer konkreteren Ebene Anforderungen an empirische Untersuchungen.
Das Kapitel 3 kristallisiert den Begriff der Bewertungskompetenz aus der Fachliteratur heraus und spezifiziert Kompetenzstrukturmodelle, Einflussfaktoren und schließlich die Bedeutung der Bewertungskompetenz für den naturwissenschaftlichen Unterricht.
Verschiedene Ansätze für die Schulung von Bewertungskompetenz im Unterricht wurden vorgestellt, dabei wurde auch auf die von Jungkamp und Marohn entwickelte Un-

terrichtskonzeption choice^{2}reflect verwiesen (Marohn & Jungkamp 2016; Jungkamp 2021).

Choice^{2}reflect stellt eine unterrichtliche Intervention dar, die Schülerinnen und Schüler der Sekundarstufe I im Umgang mit naturwissenschaftlichen Kontroversen schult. Das Ziel ist es, Lernende zu befähigen, in gesellschaftlich kontrovers diskutierten Themen zu einer rationalen Entscheidung zu gelangen.
Um das zu erreichen, berücksichtigt das Konzept die Erkenntnisse aus Entscheidungs- und Wissenschaftstheorie. Das zentrale Element bildet eine Phase, in der die Lernenden *wissenschaftliche Prüfkriterien* erarbeiten, die an die zuvor herausgearbeiteten *Gütekriterien empirischer Untersuchungen* angelehnt sind.
Schüler:innen erlernen wissenschaftliche Denk- und Arbeitsweisen: Sie prüfen Aussagen auf ihre Falsifizierbarkeit, analysieren die Stichproben empirischer Studien im Hinblick auf die Auswahl- und Versuchsbedingungen und begeben sich auf die Suche nach Scheinkausalitäten in den Schlussfolgerungen der Studien.
Insgesamt identifizierte Jungkamp sieben Wissenschaftskriterien, die in der Konzeption verankert wurden.
Das Konzept ist sowohl unabhängig vom behandelten Kontext als auch fächerübergreifend angelegt: Als umspannende Kontexte können naturwissenschaftliche Kontroversen verwendet werden. Diese sind häufig nicht an eine der Naturwissenschaften gebunden, sondern beziehen sich stets auf mehrere Disziplinen. Jungkamp (2021) verwendet beispielsweise die Kontroverse *Homöopathie*, die sowohl aus chemischer Sicht betrachtet werden kann (Durchführen von Verdünnungen, Betrachtung von Stoffeigenschaften etc.) als auch mit ihren zahlreichen gesundheitsbezogenen Aspekten aus biologischer oder physikalischer Sicht.

4.2 Die fünf Phasen der Konzeption

Die Unterrichtskonzeption choice^{2}reflect gliedert sich in fünf Phasen, die in der Abbildung **??** dargestellt sind. Sie startet mit einem *Fachspezifischen Problemaufriss*. Dabei erfolgt eine neutrale Einführung in das kontrovers diskutierte Thema: Es werden Hintergrundinformationen zu einem ganz spezifischen Aspekt aus der Kontroverse gegeben, ohne bereits Vor- oder Nachteile zu nennen oder eine Position vorzugeben. Jungkamp entwickelte Lernmaterial zur Homöopathie. In der ersten Phase wurden

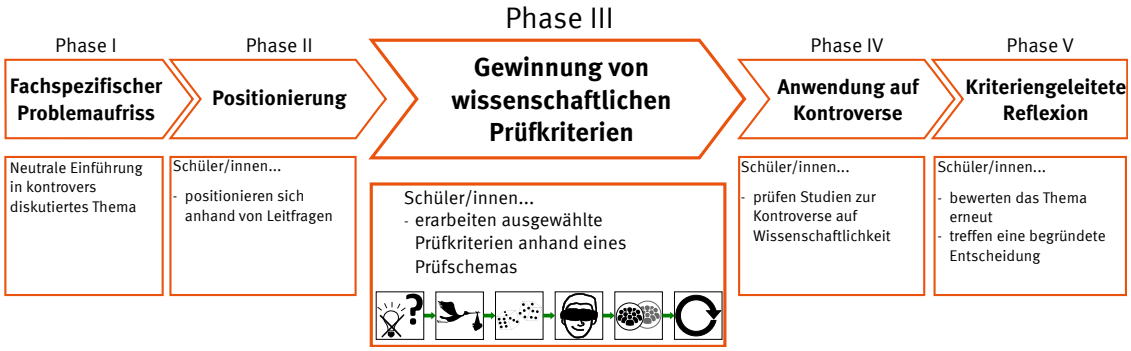


Abbildung 4.1: *Darstellung der fünf Phasen von choice² reflect (Jungkamp 2021; Kralisch & Marohn o.D.).*

auf einem Infoblatt die Regel „Ähnliches heilt Ähnliches“ und das Verfahren der Potenzierung kurz erläutert. Ein weiteres Arbeitsblatt wirft eine spezifischere Problemstellung auf. Die Lernenden werden mit der Frage konfrontiert, ob einem Mädchen, das von ADHS betroffen ist, mit homöopathischen Mitteln geholfen werden kann.

Damit wird die zweite Phase der Konzeption begonnen: Die *Positionierung*. In der Positionierungs-Phase legen die Lernenden ihren persönlichen Standpunkt zur Kontroverse dar, beispielsweise nehmen sie kurz Stellung zu ihrer Sichtweise auf Homöopathie. Eigene Vorerfahrungen mit der Kontroverse sowie kurze Bewertungsfragen können die Lernenden in der zweiten Phase beantworten.

Nachdem sich alle Schüler:innen positioniert haben, wird zum Kern der Konzeption übergeleitet: Zur Phase der *Gewinnung wissenschaftlicher Prüfkriterien*. Diese Phase steht losgelöst von der umspannenden Kontroverse. Sie beinhaltet die Erarbeitung der theoriebasierten Wissenschaftskriterien (siehe nächster Abschnitt).

Die Lernmaterialien sind dabei so gestaltet, dass jedes Kriterium an einem eigenen Kontext erarbeitet wird und eine abgeschlossene Unterrichtseinheit bildet. Somit können Lehrkräfte flexibel einzelne Prüfkriterien auswählen und im Rahmen einer Unterrichtsstunde erarbeiten. Es ist nicht obligatorisch, alle Prüfkriterien im Unterricht zu behandeln.

Das erlernte Wissen über Prüfkriterien wenden die Schülerinnen und Schüler in der vierten Phase auf die Kontroverse an (*Anwendung auf Kontroverse*). Dazu erhalten oder recherchieren sie eine echte Studie zu einem Teilaspekt der behandelten Kontroverse. Die Studie sollte zu der in Phase I generierten Problemstellung passen und

eine (vermeintliche) Antwort darauf liefern. Mithilfe der Wissenschaftskriterien prüfen die Lernenden diese Studie auf Wissenschaftlichkeit.

Mit den Erkenntnissen aus der Phase IV treffen die Lernenden eine Entscheidung in der letzten Phase, der *kriteriengeleiteten Reflexion*. Sie bewerten das Thema, zu dem sie sich in Phase II bereits positioniert haben, erneut; beziehen dabei jedoch nur wissenschaftlich gesicherte Informationen in ihre Entscheidung mit ein.
Hintergründe zu den fünf Phasen sind auch zu finden bei Marohn & Jungkamp (2016), Jungkamp (2021) und Kralisch & Marohn (o.D.).

4.3 Wissenschaftliche Prüfkriterien

In Phase III werden wissenschaftliche Prüfkriterien erarbeitet. Jungkamp (2021) identifizierte die sieben Prüfkriterien *Falsifizierbarkeit*, *Plausibilität*, *kontrollierte Bedingungen*, *Korrelation und Kausalität*, *Reproduzierbarkeit*, *Verblindung* und *Randomisierung*.
An dieser Stelle wird nur ein kurzer Überblick über die Hintergründe der Prüfkriterien gegeben. Einzelheiten zur theoretischen Fundierung, der genauen Stellung bei der Bewertung von Aussagen und Studien sowie den Beziehungen der Kriterien untereinander werden im *Design-Experiment* entwickelt. Die Grundlage dafür bietet der *Design-Based-Research*-Ansatz, der den forschungsmethodischen Rahmen dieser Arbeit bildet: Die Entwicklung und Erprobung von Lernmaterialien zu den Wissenschaftskriterien bilden den ersten Mesozyklus des Projekts (Kapitel 9).

Falsifizierbarkeit

Die Falsifizierbarkeit geht zurück auf die wissenschaftstheoretischen Überlegungen Karl Poppers, der beschrieb, dass das endgültige Beweisen von Aussagen anhand induktiver Schlüsse nicht möglich ist. Stattdessen müssen Hypothesen stets so gestaltet sein, dass sie mit deduktiven Methoden überprüfbar und widerlegbar sind (Popper 1989).
Im Kapitel 2.3.2 wird Falsifizierbarkeit als Gütekriterium hergeleitet. Die Entwicklung von Unterrichtsmaterial und dessen Erprobung fanden im Rahmen einer Masterarbeit in Zusammenarbeit mit Jungkamp statt und sind im Kapitel 9.1 beschrieben.

Plausibilität

Neben einer prinzipiellen, also aus logischer Sicht gegebenen Widerlegbarkeit von Hypothesen wird es auch als wichtig angesehen, dass Hypothesen *plausibel* sind. Auch wird im Alltag häufig von *interessanten*, *anregenden* oder *überzeugenden* Interpretationen gesprochen. Die letzten drei Kriterien scheinen beliebig auslegbar zu sein und beanspruchen keine richtungsübergreifende Akzeptanz (Winko 2015).
Die Auslegbarkeit des Begriffs *plausibel* lässt sich etwas zielführender eingrenzen. Nach Charles Peirce müssen plausible Hypothesen „evident, einfach und effektiv prüfbar" sein (Wirth 1995, S. 407). Allerdings wird Plausibilität auch als Alternative zum Falsifikationismus nach Popper angesehen, da sie sich nicht direkt auf Prüfbarkeit beziehe.
An anderer Stelle wird wiederum die *Effektive Prüfbarkeit* als Kriterium für gute plausible Hypothesen angesehen (Winko 2015).
Insgesamt lässt sich das Prädikat *plausibel* auf drei verschiedene Weisen verwenden. Es kann auf Gegenstände (Aussagen und Aussagenkomplexe) angewendet werden, Eigenschaften (z. B. „Überzeugtheitsgrad") oder auch im Bezug auf den Geltungsanspruch variieren (Winko 2015).
In der Wissenschaftstheorie wird *Plausibilität* jedoch in der Regel nicht mehr als Gültigkeitskriterium angeführt. Stattdessen geht es um die *Korrektheit des Schließens*: Ob eine Argumentation schlüssig ist, entscheidet zunächst nur die logische Struktur. Die *Plausiblität* einer Argumentation begründet sich über inhaltliche Faktoren.
Die inhaltliche Einschätzung, ob eine These plausibel ist oder nicht, wird letztendlich mit einer Portion „Bauchgefühl" getroffen. Weiterhin ist sie nur vorläufig. Denn sobald eine Hypothese auf ihre Wahrheit (bzw. Falschheit) hin beurteilt werden kann, löst das Ergebnis die vorläufig getroffene Einschätzung ab (Winko 2015).

Der theoretische Rahmen zur Plausibilität zeigt, dass dieses Prüfkriterium für Schüler:innen schwer handhabbar ist. Außerdem wird es obsolet, sobald die Prüfung auf Falsifizierbarkeit erfolgte. Auf das *Bauchgefühl* zur inhaltlichen Passung, Stimmigkeit oder eben Plausibilität sollten Lernende im Fall von choice^{2}reflect nicht setzen.
In dem Unterrichtskonzept geht es darum, auf der Grundlage von „harten Kriterien" eine Aussage als wissenschaftlich oder unwissenschaftlich einzustufen. Das kann das Prüfkriterium *Plausibilität* nicht leisten und wird daher im Rahmen dieses Projekts nicht weiter ausgearbeitet.

Kontrollierte Bedingungen

In experimentellen Settings darf stets nur die zu untersuchende Variable verändert werden, während alle anderen konstant gehalten werden. Diese Anforderung ist hergeleitet aus den wissenschaftstheoretischen Ausführungen des Kapitels 2.4.1. Nur unter Kontrolle aller relevanten Störfaktoren sind wissenschaftlich belastbare Schlussfolgerungen aus einem Experiment möglich.
Da in der Realität viele Faktoren nicht einfach variierbar sind, werden anstelle eines experimentellen Settings empirische Kohortenstudien durchgeführt, bei denen die Stichprobe eine Experimental- und eine Kontrollgruppe umfasst. Dieses Kontrollgruppendesign dient als Qualitätskriterium für empirische Studien.
Beide Aspekte zusammengenommen, also die adäquate Kontrolle der Variablen und die Umsetzung von Kontrollgruppendesigns, bilden den Rahmen für das Prüfkriterium *kontrollierte Bedingungen*.
Jungkamp (2021) entwickelte und erprobte eine Unterrichtseinheit zu diesem Prüfkriterium, in der die Lernenden einen möglichst stabilen Milchschaum herstellen. In der experimentellen Durchführung erlernen sie, dass stets nur eine Bedingung variiert werden darf, z. B. die Temperatur der Milch oder die Technik des Aufschäumens.

Korrelation und Kausalität

Die Kontrolle von Störfaktoren in experimentellen Settings wie auch in empirischen Studien ist elementar für die Wissenschaftlichkeit einer Untersuchung. Doch auch unter *kontrollierten Bedingungen* können häufig keine beliebig verallgemeinerbaren Gesetze aufgestellt werden. Das trifft vor allem auf gesundheitsbezogene empirische Studien zu.
Wird ein statistischer Zusammenhang zwischen zwei Variablen gefunden, so *korrelieren* beide Variablen miteinander. Allerdings lässt sich aus einer statistischen Korrelation nicht ableiten, inwiefern eine der Variablen die *Ursache* der anderen darstellt. Nur wenn alle Störvariablen berücksichtigt und konstant gehalten werden, lassen sich *Kausalaussagen* treffen, siehe dazu auch Kapitel 2.4.2.
Da dies bei empirischen Studien jedoch nur bedingt zu realisieren ist, muss der Unterschied zwischen *Korrelationen* und *Kausalitäten* stets berücksichtigt werden; sowohl beim Treffen von Schlussfolgerungen als auch bei der Interpretation bereits gegebener Schlussfolgerungen.

Unterrichtsmaterial zu diesem Prüfkriterium wird im Kapitel 9.2 entwickelt und erprobt.

Reproduzierbarkeit

Neben der Validität von Untersuchungen gehören auch *Reliabilität* und *Objektivität* zu den zu berücksichtigenden Gütekriterien. Im Kapitel 2.4.3 wurde hergeleitet, dass eine Untersuchung zu beliebigen Zeitpunkten von unterschiedlichen Personen unter Erhalten desselben Ergebnisses wiederholbar sein muss.

Auch die Versuchsdokumentation muss entsprechend reproduzierbar gestaltet sein. Diese Anforderungen bilden das Prüfkriterium *Reproduzierbarkeit*, das im Kapitel 9.4 ausgearbeitet wird.

Verblindung

In experimentellen Studiendesigns, in denen die Stichprobe auf eine Test- und eine Kontrollgruppe aufgeteilt wird, dürfen weder die Versuchspersonen noch die Durchführenden oder Auswertenden der Studie die genaue Aufteilung der Versuchsbedingungen kennen. Es darf nicht ersichtlich sein, ob eine Versuchsperson sich in der Testgruppe oder der Kontrollgruppe befindet.

Wird das Kriterium der *Verblindung* nicht eingehalten, können Einstellungen, Erwartungen und persönliche Überzeugungen zu Verzerrungseffekten führen (Bortz & Döring 2006, S. 197; Kabisch et al. 2011). Somit erlernen die Schüler:innen im Rahmen der Phase III von choice^{2}reflect die Bedeutung von Verblindung und Verfahrensweisen zum Erreichen von Verblindung kennen.

Sie führen eine Verkostung von eingefärbtem Naturjoghurt durch. Es handelt sich bei allen „Sorten“ um die gleiche Geschmacksrichtung, allerdings sind sie unterschiedlich eingefärbt und suggerieren dadurch jeweils eine bestimme Geschmacksrichtung.

Die Lernenden sollen daraufhin vermuten, welche Sorten in den Joghurtbechern vorliegen. Häufig „erschmecken“ sie dabei Geschmacksrichtungen, die aufgrund der Farbe zu vermuten sind (z. B. Erdbeergeschmack bei rot eingefärbtem Joghurt). An diesem Beispiel lernen die Schüler:innen, wie persönliche Erwartungen ein experimentelles Ergebnis beeinflussen können (Jungkamp 2021; Marohn & Jungkamp 2016).

Randomisierung

Eine weitere Anforderung zur Sicherung der Validität empirischer Studien ist das Zuteilen von Versuchspersonen zu der Test- oder Kontrollgruppe durch ein zufallsbasiertes Verfahren. Die Zuteilung darf keiner bewussten oder unbewussten Selektion unterliegen und wird idealerweise über ein „Münzwurf-Verfahren“ durchgeführt.
Nur durch Randomisierung kann sichergestellt werden, dass die Merkmale, in denen sich Versuchspersonen (natürlicherweise) unterscheiden, auf die verschiedenen Versuchsgruppen aufgeteilt werden und damit an Relevanz für die Auswertung verlieren (Scheffler 2000, S. 67).
Die Randomisierung betreffende Unterrichtsmaterialien werden im Kapitel 9.3 ausgearbeitet.

4.4 Prüfkarten

Die Erarbeitung der Prüfkriterien in Phase III wird ergänzt durch einen Pool an *Prüfkarten*. Zu jedem Prüfkriterium werden Prüfkarten entwickelt, die die wesentlichen Informationen zu den Kriterien enthalten.
Die Prüfkarten weisen ein einheitliches Design auf, sowohl von der visuellen als auch inhaltlichen Gestaltung. Somit werden pro Prüfkriterium sechs Prüfkarten entwickelt. Die erste Seite enthält als Deckblatt den Namen und das Symbol des Kriteriums. Es folgt eine allgemein informierende Seite, zum Beispiel: „Was ist Randomisierung?“ Die dritte Seite stellt dar, mit welchen Mitteln das Kriterium erreicht werden kann, zum Beispiel: „Wie erreicht man Randomisierung?“
Angefügt werden weitere Seiten mit Anwendungsmöglichkeiten, dabei hat sich die Aufteilung nach *Chemie*, *Forschung* und *Alltag* als hilfreich erwiesen (Jungkamp 2021). Die Prüfkarten enthalten auch Vertiefungsaufgaben, die im Anschluss an die Erarbeitung des Kriteriums mit den Schüler:innen durchgeführt werden können.
Die Entwicklung weiterer Prüfkarten ist angegliedert an den ersten Mesozyklus, in dem auch die Prüfkriterien weiterentwickelt werden (Kapitel 9).

Teil III

Projektverlauf gemäß Design-Based Research

5 Der Design-Based Research-Ansatz

> Ist und bleibt es alleinige Aufgabe der Lehr-Lernforschung [...], Innovationen in der Bildung zu evaluieren? Gibt es keine Möglichkeiten, Lehr-Lernforschung *zum Zwecke der Innovation* zu betreiben [...]?
>
> – Reinmann (2005)

Das Verständnis des Begriffs *Innovationen* hat sich im Laufe der Zeit verändert: Traditionell wurden unmittelbare, dramatische Effekte mit großen Schritten neuartiger Ideen erwartet; nach einem modernen Innovationsverständnis haben auch kleinschrittige undramatische Entwicklungen ihren Platz (Reinmann 2005).

In der Bildung sind es vor allem letztere kleinschrittige Entwicklungen, die sich als Innovationen erweisen können. Doch woher stammen Innovationen in der Bildung, wie lassen sich neue Konzepte in die Praxis integrieren?

Brown (1992) erkannte, dass in klassischer Experimentalforschung nachhaltige Innovationen ausbleiben bzw. dass die Erkenntnisse oft nicht ausreichend sind, um Konzepte und Instrumente für die Bildungspraxis zu liefern.

Die traditionelle Forschung befinde sich häufig im *belief mode*: Bestehende Ideen und Wissensstrukturen werden auf ihre Wahrheit überprüft, während gestaltungsorientierte Forschung nur selten stattfindet (Bereiter & Scardamalia 2003).

Demgegenüber steht der *design mode*, in dem es um die Gestaltung neuer Konzepte und deren Nützlichkeit in praktischen Anwendungen geht.

Eine stärker Design-orientierte Bildungsforschung wird immer häufiger gefordert (Reinmann 2005; Wilhelm & Hopf 2014; Collins et al. 2004) und wird durch den *Design-Based Research-Ansatz* umgesetzt.

Dieser Forschungsansatz wird im folgenden Kapitel charakterisiert.

5.1 Entwicklung des DBR

Design-Based Research geht zurück auf Collins (1992), der „design science“ vor dem Hintergrund entwickelte, technologische Innovationen in den naturwissenschaftlichen Unterricht einzubringen und zu evaluieren.

Brown (1992) verfolgte einen ähnlichen Ansatz und entwickelte den Begriff *design experiments*. Sie bezeichnet Laborstudien als ungeeignet, um Unterricht zu untersuchen, da die Separation einzelner Variablen dem realen Unterrichtsgeschehen nicht gerecht werde. Das Ziel besteht in der Entwicklung „laborferner“ Interventionen für die Praxis mit gleichzeitiger Theoriebildung zur Erklärung der Wirkungen entwickelter Interventionen (Rohrbach-Lochner 2019, 13f).

Der Forschungsrahmen ist auch bekannt als *Educational Design Research* (McKenney & Reeves 2012), beide Begriffe entwickelten sich parallel aus „design science“ (Brown 1992; Collins 1992; Cobb et al. 2003).

Im deutschsprachigen Raum wird Der Begriff *Design-Based Research* (DBR) häufiger verwendet (Reinmann 2005; Krüger 2010; Tobias 2010; Rohrbach-Lochner 2019).

Mit DBR wird das Spannungsfeld zwischen Grundlagenforschung und anwendungsbezogener Forschung adressiert; Stokes (1997) schlägt ein 2x2-Quadrat vor: Zwischen reiner Grundlagenforschung und reiner Anwendungsforschung liege die anwendungsorientierte Grundlagenforschung - darin sieht Stokes ein großes Potential und Fischer et al. (2005) sowie Reinmann (2005), Prediger et al. (2012) und Rohrbach-Lochner (2019) verorten dort Forschungsarbeiten zum DBR.

Die Lücke zwischen beiden Forschungsrichtungen kann DBR zwar nicht vollständig schließen (McKenney & Reeves 2012), allerdings ist der Ansatz wichtig für die Entwicklung neuer Theorien bei gleichzeitiger Anwendbarkeit der gewonnenen Ergebnisse in der Praxis.

Collins et al. (2004) beschreibt diese Zweiteilung als *dual goals*: „Design research should always have the dual goals of refining both theory and practice“.

5.2 Charakteristika des Design-Based Research-Ansatzes

Charakteristisch für Design-Based Research ist die kontinuierliche Weiterentwicklung und Verfeinerung von Interventionen. Die erste Version einer Intervention wird in einer realen Lernumgebung eingesetzt und die Situation genau beobachtet.

Aus den Ergebnissen der Beobachtung werden Rückschlüsse auf die Angemessenheit der Intervention gezogen: Es wird determiniert, welche Aspekte der Intervention in der realen Lernumgebung funktionieren („What works?“, siehe Collins et al. 2004; Rohrbach-Lochner 2019, S. 19).

Auf Grundlage der Schlussfolgerungen werden Interventionen weiterentwickelt und verbessert, um anschließend erneut in praxisnahen Bildungssituationen erprobt zu werden. Wichtig ist, dass dieser Prozess nicht nur auf die Verfeinerung von Praxiselementen abzielt, sondern auch konkrete theoretische Fragen und Probleme adressiert. Im Zuge der Weiterentwicklung einer Intervention werden somit sowohl der Praxiseinsatz als auch die dahinter liegenden Theorien gestärkt (Collins et al. 2004).

Design-Experimente unterscheiden sich von anderen klassischen Forschungsrichtungen in verschiedenen Merkmalen. Collins et al. (2004) ziehen einen Vergleich der Merkmale heran, um Laborstudien von Design-Based Research abzugrenzen. In der Tabelle 5.1 befindet sich eine zusammengefasste Übersicht über die wichtigsten Abgrenzungsmerkmale.

Es wird deutlich, dass Laborstudien unter anderen Bedingungen und mit anderer Zielsetzung stattfinden als Design-basierte Projekte. Während erstere unter Einsatz von standardisierten Methoden und Variablenkontrolle Effekte einzelner abhängiger Variablen zu messen beabsichtigen, ist das Ziel von DBR die Identifikation interessierender Variablen und das Erstellen eines quantitativen und qualitativen Profils zur Charakterisierung von realen Lernsituationen.

Rott & Marohn (2016) stellen den DBR zusammenfassend als *iterativ*, *prozessorientiert*, *nutzenorientiert*, *theorieorientiert*, *kooperativ* und *offen gegenüber allen Untersuchungs- und Auswertungsmethoden* dar.

DBR ist dabei allerdings eher als *Rahmenansatz* und nicht als Forschungsrichtung zu verstehen, da er in vielen Fachrichtungen verwendbar ist (Rohrbach-Lochner 2019, S. 31).

Tabelle 5.1: *Vergleich der Charakteristika von Laborstudien und Design-Based Research nach Collins et al. (2004).*

Merkmal	Laborstudien	Design-Based Research
Umgebung	Labor	„messy situations“, reale Lernsituationen
abhängige Variablen	eine	mehrere
Methode	Variablenkontrolle	Variablenidentifikation
Verfahrensweise	starres, geplantes Vorgehen	flexibles Vorgehen, erfolgsorientierte Revision
soziale Umgebung	Isolation	Interaktion
Ziel	Hypothesentestung	Profil zur Charakterisierung erstellen
Entscheidungsträger	Leiter:in des Experiments	mehrere Mitwirkende

5.3 Kritik am DBR

Der DBR-Rahmenansatz wurde und wird kontrovers diskutiert. Anwendungsbereich und Ausgangspunkt, von dem der DBR sich entwickelt hat, werden gleichzeitig auch als seine Schwäche angesehen: Der Einsatz von Interventionen in *realen Bildungssituationen* ist durch eine große Fülle an Variablen gekennzeichnet, deren Konstanthaltung durch Forschende nicht kontrollierbar ist (Collins et al. 2004).

Anstelle von systematischer Kontrolle der Variablen werden bei DBR-basierter Forschung die Begleitumstände möglichst detailgenau beschrieben und charakterisiert. Unterrichtssituationen werden häufig aufgezeichnet, um eine umfassende Erhebung möglichst vieler Variablen zu gewährleisten.

Dabei entstehen große Datenmengen, die nur unter einem hohen Aufwand analysiert werden können. In den meisten Projekten ist es nicht möglich, *alle* erhobenen Daten darzustellen und Schlüsse daraus zu ziehen, dies wird als Schwäche des Ansatzes angesehen (Collins et al. 2004; Rohrbach-Lochner 2019, S. 23).

Bei der Auswahl der „relevanten“ Daten können Forschende einem *bias* unterliegen, sodass sie unbewusst oder bewusst nur die Daten auswählen, die zu ihren Vermutungen passen. Diese Problematik sollte Forschenden stets präsent sein und bei der Forschungsdokumentation beachtet werden.

Dede (2004) sieht den DBR-Ansatz als „underconceptualized“ und „overmethodologized“ an (vgl. auch Rohrbach-Lochner 2019, S. 24). Weiterhin fehle dem Ansatz ein Konsens an Standards; und auch die Validität und Verallgemeinerbarkeit generierter Ergebnisse werden kritisch betrachtet (Hoadley 2004; Reinmann 2007).
Um dieser Kritik zu begegnen und einen hohen Grad an Nachvollziehbarkeit zu gewährleisten, folgen DBR-basierte Projekte festgelegten Ablaufschemata, für die das iterative Vorgehen in aufeinanderfolgenden Mesozyklen charakteristisch ist.

5.4 Ablauf design-basierter Forschungsprojekte

Rohrbach-Lochner (2019), erstmals veröffentlicht in Rohrbach-Lochner & Marohn (2018) hat ausgehend von Cobb et al. (2003), Akker et al. (2006) & McKenney & Reeves (2012) ein Strukturmodell des Design-Based Research entwickelt. Dieses ist in der Abbildung 5.1 dargestellt.
In dem Strukturmodell wird der Dreischritt von Design-basierten Projekten klar dargestellt. In der Vorbereitungsphase wird ein didaktisches Problem aufgeworfen, dessen Lösung im Design-Experiment iterativ entwickelt wird. Abschließend werden in einer retrospektiven Analyse praxisorientiert die entwickelte Intervention reflektiert sowie theorieorientiert Hypothesen zur Funktionsweise der Intervention dargestellt.

5.4.1 Framing

Die Vorbereitungsphase, auch Framing genannt (Rott & Marohn 2016), besteht in der Ausdifferenzierung der Problemstellung, der Zielformulierung und Literaturrecherche. Die Ziele des Projekts, eine praxistaugliche Intervention zu entwickeln und Theorien über dessen Einsatz in realen Lernsituationen zu bilden, werden bereits in dieser Phase ins Auge gefasst.
Ein Problem aus der Bildungspraxis wird aufgegriffen und in eine vorläufige Frage- oder Problemstellung überführt. Damit ist der Ausgangspunkt des Projekts geschaffen (Cobb et al. 2003; Rohrbach-Lochner 2019, S. 32). Es folgt die Formulierung von Zielen, die auf die konkrete Problemstellung bezogen sind. Dazu gehört z. B. die Entwicklung einer Lehr-Lern-Umgebung.

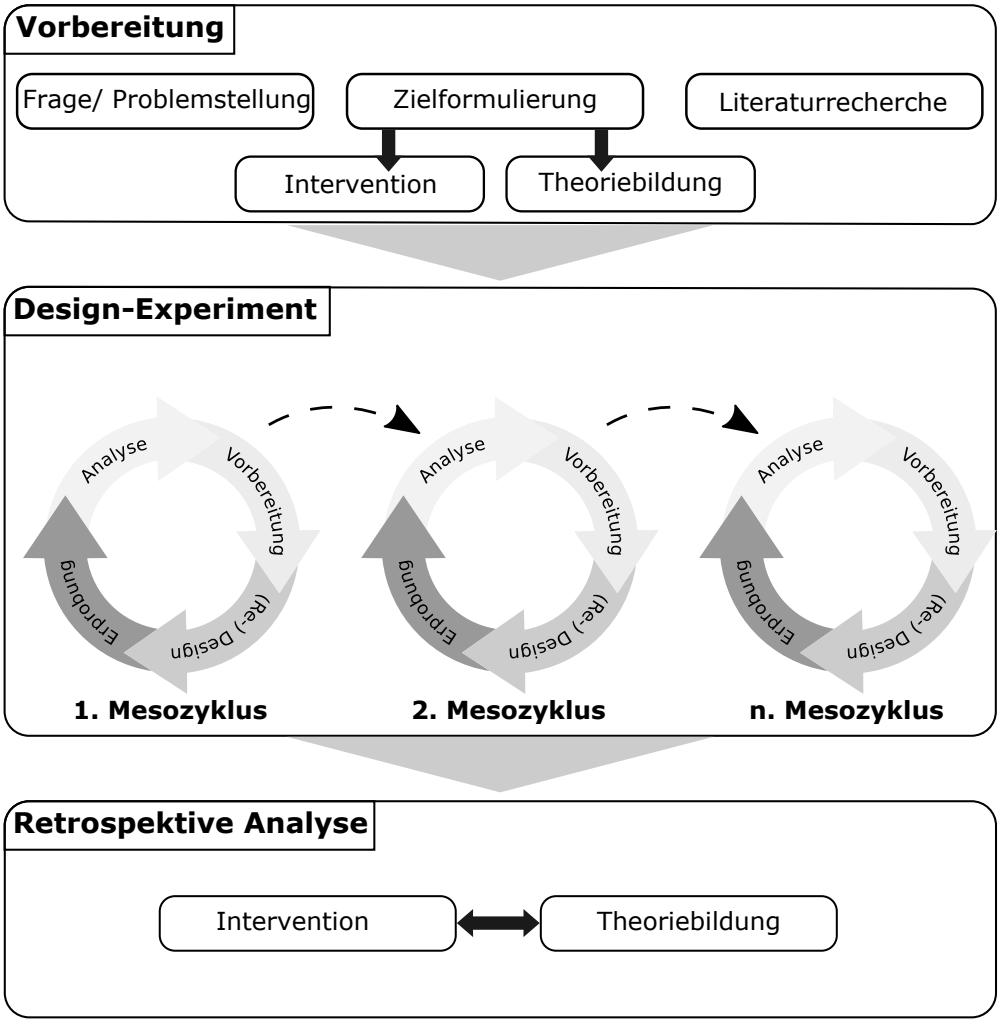


Abbildung 5.1: *Strukturmodell design-basierter Forschung, verändert nach Rohrbach-Lochner (2019, S. 31).*

Sowohl die übergeordneten Fragestellungen als auch Zielformulierungen sind tentativ. Sie können im Laufe des Forschungsprojekts im Einklang mit Ergebnissen aus den verschiedenen Mesozyklen verändert und angepasst werden (McKenney & Reeves 2012; Rohrbach-Lochner 2019).
Zum Framing gehört auch der Beginn einer intensiven Literaturrecherche. Die Forscherin oder der Forscher erarbeitet sich einen Einblick in den aktuellen Stand der Forschung zur jeweiligen Problemstellung. Die Recherche ist phasenübergreifend und kann zu verschiedenen Punkten des Projektverlaufs eine wichtige Rolle spielen. In der Framing-Phase erfolgt die Sichtung und Bearbeitung von Literatur jedoch meist am intensivsten (Rohrbach-Lochner 2019, S. 33).

5.4.2 Design-Experiment

Im Design-Experiment wird der Fokus auf die Lösung des Problems im Lehr-Lernkontext gelegt. Die übergeordnete Fragestellung wird in kleinere Sinneinheiten eingeteilt. Jede dieser Einheiten wird durch eigene Fragestellungen und Ziele angeführt und bildet daher einen Mesozyklus innerhalb des Design-Experiments.
Die Gliederung in Mesozyklen sorgt für den DBR-typischen iterativen Verlauf des Projekts. Jeder Mesozyklus besteht aus den Elementen *Synthese und Vorbereitung*, *Design*, *Erprobung* und *Analyse* (McKenney & Reeves 2012, 78f; Rott & Marohn 2016). Häufig werden Mesozyklen auch in *Analyse und Vorbereitung*, *Konstruktion und Durchführung* sowie *Evaluation und Reflexion* eingeteilt (Rohrbach-Lochner 2019). Die Bestandteile der Mesozyklen sind demnach nicht trennscharf voneinander abgrenzbar, definieren jedoch in ihrer Gesamtheit einen Durchlauf. Die Phasen (oder auch Mikrozyklen) eines Mesozyklus werden im Folgenden näher erläutert.

Synthese und Vorbereitung

Ein Mesozyklus beginnt mit der Synthese und Vorbereitung. Die übergeordnete Fragestellung wird, bezogen auf den zu untersuchenden Teilaspekt, konkretisiert und stärker auf den Einsatz im realen Lehr-Lernkontext fokussiert. Weiterhin werden die Rahmenbedingungen zur Erprobung und Analyse der Intervention geklärt: Lerngruppe und -ort werden charakterisiert und es werden die nötigen Vorbereitungen zur Erprobung getroffen (Rohrbach-Lochner 2019, S. 35; McKenney & Reeves 2012).

(Re-)Design

Die konkretisierte Fragestellung dient als Leitfaden für die Erstellung der Intervention. Die Forscherin oder der Forscher bezieht zum Einen theoretische und analytische Überlegungen in die Konstruktion der Intervention ein, weiterhin sind auch kreative Prozesse beim Interventions-Design von Bedeutung. Eine ausführliche Dokumentation der Entscheidungen, die zum Design der Materialien geführt haben, ist essentiell (McKenney & Reeves 2012).

Erprobung

Unter fortlaufender Dokumentation aller Arbeitsschritte und sorgfältiger Datenerhebung findet die Erprobung bzw. Durchführung der Intervention im realen Bildungskontext statt. Die verwendeten Erhebungsmethoden können vielfältig sein (Mixed Methods) und variieren je nach Fragestellung des Mesozyklus - häufig zeichnen sich design-basierte Projekte jedoch durch qualitativ orientierte Untersuchungsmethoden aus (Wilhelm & Hopf 2014; Rott & Marohn 2016).

Analyse

In dieser Phase findet eine ausführliche Auswertung und Reflexion der erhobenen Daten statt. Die oder der Forschende bedient sich sowohl qualitativer als auch quantitativer Methoden. Ergebnisse der verschiedenen Methoden werden triangulierend zusammengeführt und zur Beantwortung der eingangs formulierten Forschungsfrage herangezogen.
Durch das Forschungs-Setting entsteht eine große Menge an Daten, deren Auswertung jedoch in einem für den Forschenden zumutbaren Maß geschehen sollte. Der Umfang der Evaluation kann somit geringer ausfallen als in vergleichbaren Laborstudien, bei denen Einflüsse einzelner Variablen kontrolliert gemessen werden (Wilhelm & Hopf 2014; Rohrbach-Lochner 2019). Gütekriterien der einzelnen verwendeten Methoden sind jedoch weiterhin zu beachten.
Erkenntnisse aus dieser Phase fließen direkt in den nächsten Durchlauf des Mesozyklus ein und formen die Fragestellung für das weitere Vorgehen sowie die Materialentwicklung in der Phase des Re-Designs.

5.4.3 Re-Framing

Das Re-Framing, auch Retrospektive Analyse genannt, schließt den Makrozyklus des Forschungsprojekts ab. Unter Rückbezug auf die zu Beginn des Projekts aufgestellten Fragestellungen und Ziele wird ein Blick auf die im Design-Experiment entwickelte Intervention geworfen. Es wird reflektiert, inwiefern eine praxistaugliche und funktionierende Intervention entwickelt wurde. Weiterhin soll die Tragweite der Ergebnisse für andere Fachrichtungen in Form einer Theoriebildung abgesteckt werden. (Krüger 2010; Rohrbach-Lochner 2019)

5.5 Einordnung des eigenen Projektverlaufs

Das in dieser Arbeit beschriebene Forschungsprojekt orientiert sich am Forschungsrahmen des Design-Based Research. Es steht die Entwicklung einer praxistauglichen Intervention im Vordergrund, die sich für den Einsatz in Lehr-Lernumgebungen eignet.
Diese Arbeit knüpft an ein bereits bestehendes design-basiertes Forschungsprojekt, durchgeführt von Florian Jungkamp aus dem Arbeitskreis von Annette Marohn, an.
Das Design-Experiment in Jungkamps Projekt wird, dem Forschungsrahmen des DBR folgend, durch eine Framing-Phase initiiert und mittels Re-Framing in Richtung einer praxistauglichen Intervention geformt. Es zeichnet sich durch vier Mesozyklen mit unterschiedlichen Fragestellungen aus.
Das nun nachfolgende Projekt greift zentrale Erkenntnisse und identifizierte Gestaltungsmerkmale von Jungkamp (2021) auf mit dem Ziel, die Intervention weiterzuentwickeln. Damit ist nicht gemeint, dass die Mesozyklen des Design-Experiments unter ähnlicher Zielsetzung weitergeführt und in Jungkamps Projekt integriert werden. Vielmehr wird ein neuer Makrozyklus begonnen, der seinerseits von Framing- und Re-Framing-Elementen begleitet wird.
Jungkamp (2021) stellte Untersuchungen zum wissenschaftlichen Weltbild von Schüler:innen an, formulierte Prüfkriterien zum graduellen Aufbau des Wissenschaftsverständnisses Lernender, identifizierte Homöopathie als geeigneten unterrichtlichen Kontext und entwickelte die Unterrichtskonzeption *choice²reflect* als didaktische Rahmung zur Vermittlung der essentiellen Elemente wissenschaftlicher Denk- und Arbeitsweisen.

Diese Erkenntnisse liefern eine Basis für das nachfolgende Projekt und können direkt in der Framing-Phase als Ausgangspunkt verwendet werden (siehe Kapitel IV).

Teil IV

Framing

Im Abschnitt 5.4 wurde der allgemeine Ablauf design-basierter Projekte beschrieben und das von Rohrbach-Lochner (2019) entwickelte Strukturmodell präsentiert. Diese Arbeit beschreibt ein Forschungsprojekt im Rahmen des Design-Based Research-Ansatzes und lässt sich somit in die Phasen *Framing*, *Design-Experiment* und *Re-Framing* gliedern. Wie bereits beschrieben, knüpft die Arbeit direkt an ein Vorgängerprojekt an und bildet einen weiteren, durch Framing und Re-Framing umschlossenen Makrozyklus.
Grundlegende Literatur mit aktuellen Erkenntnissen aus Entscheidungstheorie, Wissenschaftstheorie und der Konzeptualisierung von Bewertungskompetenz wurde im Kapitel II dargestellt. Dieses bildet zusammen mit den Erkenntnissen des Vorgängerprojekts (Jungkamp 2021) die Basis für die initiierende Phase dieses Projekts - das Framing.

6 Forschungsergebnisse des Vorgänger-Projekts

Florian Jungkamp (2021) entwickelte zusammen mit Prof. Marohn die Unterrichtskonzeption *choice²reflect* zur Bewertung kontroverser Themen im naturwissenschaftlichen Unterricht. Die Entwicklung der Intervention erfolgte nach den Grundsätzen des Design-Based Research-Ansatzes, wobei innerhalb des Makrozyklus vier Mesozyklen durchlaufen wurden.
Unter Verwendung qualitativer und quantitativer Methoden erprobte Jungkamp verschiedene Materialien der Konzeption und analysierte die erhobenen Daten hinsichtlich unterschiedlicher Fragestellungen. Der nächste Abschnitt gibt einen Überblick über die zentralen Erkenntnisse von Jungkamp, um den Ausgangspunkt dieses Projekts klar abzustecken und nachfolgende Fragestellungen und Zielsetzungen einzurahmen („Framing“).

6.1 Mesozyklus 1

Der erste Mesozyklus des Projekts fokussierte in zwei Durchläufen auf Einstellungen Lernender bezüglich naturwissenschaftlicher Kontroversen sowie auf das Verständnis von Wissenschaft.
Jungkamp stellte fest, dass Lernende den Kontext Homöopathie eher auf der Grundlage persönlicher Erfahrungen und Emotionen als auf der Grundlage von Fachwissen bewerten und beurteilen. Weiterhin ermittelte er, dass Lernende kein einheitliches Verständnis von Wissenschaft besitzen: „Wissenschaft wird zwar weitestgehend positiv betrachtet, allerdings treten auch antiwissenschaftliche Ansichten auf. Dies könnte die Auseinandersetzung mit pseudowissenschaftlichen Themen komplizierter machen, da entsprechende Informationen nicht einfach voneinander abge-

grenzt werden können“ (Jungkamp 2021).
Es zeigte sich, dass bei Lernenden ein Bedarf an methodischen Werkzeugen zur rationalen Beurteilung emotionaler Themen besteht. Diesem Bedarf begegnet Jungkamp im Mesozyklus 2.

6.2 Mesozyklus 2

Der zweite Mesozyklus beschäftigt sich mit der Entwicklung und Erprobung von Prüfkriterien zur Förderung eines methodischen Verständnisses von Wissenschaft. Er wurde acht Mal durchlaufen, wobei nach dem ersten Durchlauf die Fragestellung konkretisiert und auf einzelne Prüfkriterien bezogen wurde. Jungkamp formuliert sieben Prüfkriterien als Elemente eines kohärenten naturwissenschaftlichen Weltbildes zur Förderung des Erwerbs von *Scientific Awareness* (vgl. auch Shamos 1995).
Konkrete Materialien zu einzelnen Prüfkriterien wurden entwickelt und erprobt. Die Ergebnisse zeigen, dass die Prüfkriterien von Lernenden verstanden und angewendet werden können und die Entwicklung eines einheitlichen wissenschaftlichen Weltbildes fördern.

6.3 Mesozyklus 3

Der dritte Mesozyklus in Jungkamps Projekt stellt unter Berücksichtigung der Kriterien geeigneter Kontexte heraus, dass Homöopathie ein geeigneter unterrichtlicher Kontext ist, um mithilfe von Prüfkriterien bewertet zu werden.

6.4 Mesozyklus 4

Im vierten Mesozyklus wurde unter Berücksichtigung der Erkenntnisse vorheriger Mesozyklen die Unterrichtskonzeption *choice^2 reflect* als didaktischer Rahmen zur Vermittlung und Anwendung von Prüfkriterien entwickelt und erprobt. Die fünf-phasige Konzeption wird im Kapitel 4 näher beschrieben.
Jungkamp (2021) fand heraus, dass choice^{2}reflect die nötigen Voraussetzungen liefert, um den Kontext Homöopathie mithilfe wissenschaftlicher Denk- und Arbeitsweisen adäquat zu bewerten. Insbesondere die Prüfkriterien *kontrollierte Bedingungen*

und *Verblindung* haben eine Schlüsselrolle bei der Bewertung von Homöopathie gespielt.
Die Konzeption unterstützt Lernende bei der Erkenntnis, dass die Prinzipien der Homöopathie unwissenschaftlich sind. Die Schüler:innen erachteten homöopathische Wirkprinzipien nach der Unterrichtseinheit als nicht sinnvoll und änderten somit ihre Einstellung gegenüber der Thematik, entschieden aber mehrheitlich, die Mittel trotzdem weiter zu verwenden (Jungkamp 2021).

7 Zielformulierung und Forschungsfragen

Die deskriptive Entscheidungstheorie hat gezeigt, dass Entscheidungen in realen Situationen nur begrenzt rational getroffen werden (Kapitel 1). Emotionen und Heuristiken nehmen Einfluss auf die Entscheidungsfindung. Diese Einflüsse können durch Bewusstmachen von Entscheidungstendenzen und -verzerrungen und Erlernen wissenschaftlicher Methodik minimiert werden. Diese Erkenntnisse stehen im Einklang mit den Forderungen der Bildungsstandards und Kernlehrpläne, in denen die Vermittlung einer naturwissenschaftlichen Grundbildung bzw. *scientific literacy*, sowie die gezielte Förderung von Bewertungskompetenz verlangt werden (siehe Kapitel 3.4.2).

Im Hinblick auf diese Anforderungen hat Jungkamp (2021) im Laufe von vier Mesozyklen untersucht, wie Lernende dazu befähigt werden können, kontroverse naturwissenschaftliche Kontexte kriteriengeleitet zu bewerten. Mit der Entwicklung der Unterrichtskonzeption choice^{2}reflect beantwortet Jungkamp die übergeordnete Fragestellung.

Jungkamp betont jedoch, dass choice^{2}reflect im Rahmen seiner Arbeit zu einem *Pilotprojekt* (Jungkamp 2021, S. 305) herangewachsen ist, dessen Einbindungsmöglichkeiten in den naturwissenschaftlichen Unterricht noch nicht vollständig empirisch untersucht werden konnten. Weitere Forschungen seien notwendig, um die Konzeption weiterzuentwickeln.

Damit ergibt sich für das vorliegende Projekt die folgende übergeordnete Fragestellung:

Wie kann die Unterrichtskonzeption choice^{2}reflect als praxistaugliche Intervention zur Bewertung kontroverser Kontexte weiterentwickelt werden?

Die Erforschung dieser Fragestellung kann durch die von Jungkamp formulierten Forschungsdesiderate weiter fokussiert werden. Er nennt u. a. die folgenden Forschungsdesiderate, die als Anknüpfungspunkte für weitere Forschung dienen sollten:

1. In welcher Weise erarbeiten und nutzen Lernende die wissenschaftlichen Prüfkriterien *Reproduzierbarkeit* und *Randomisierung*? Wie könnten geeignete Materialien designt werden?
2. Auf welche Weise könnten die wissenschaftlichen Denk- und Arbeitsweisen in Form von Prüfkriterien sinnvoll sortiert und strukturiert werden?
3. Welche weiteren Kontexte wären überhaupt geeignet, um die Vermittlung kriteriengeleiteter Bewertungskompetenz zu ermöglichen?

In Anknüpfung an diese Desiderate werden die Zielsetzungen des vorliegenden Projektes definiert. Sie gliedern sich in drei Mesozyklen und sind jeweils durch eigene Fragestellungen gekennzeichnet. In der Tabelle 7.1 befindet sich eine Übersicht über die in dieser Arbeit erforschten Mesozyklen und dessen Fragestellungen.

Tabelle 7.1: *Übersicht aller in diesem Projekt durchlaufenen Mesozyklen innerhalb des Makrozyklus (Design-Experiment) sowie der zugehörigen Forschungsfragen. MiZ = Mikrozyklus.*

Mesozyklus 1 (Kapitel 9) *In welcher Weise können die Wissenschaftskriterien Falsifizierbarkeit, Korrelation und Kausalität, Randomisierung und Reproduzierbarkeit im Unterricht expliziert werden?*		
MiZ	**Sub-Forschungsfrage**	**Kapitel**
1.1	Wie können die fachlichen Inhalte der Falsifizierbarkeit für den Einsatz im naturwissenschaftlichen Unterricht didaktisch aufbereitet werden? Inwieweit sind Lernende in der Lage, das Wissenschaftskriterium *Falsifizierbarkeit* zu erarbeiten und anzuwenden? a) Inwiefern werden die Lernziele 1-4 erreicht? b) Welche Aspekte führen Lernende bei der Einschätzung der Überprüfbarkeit von Aussagen an? c) Inwiefern ist bei den Lernenden eine Unzufriedenheit mit ihrem Konzept des Beweisens von Aussagen zu erkennen?	9.1
1.2	Wie können die Unterschiede zwischen Korrelationen und Kausalitäten für den Einsatz im naturwissenschaftlichen Unterricht didaktisch aufbereitet werden? Inwieweit sind Lernende in der Lage, das Wissenschaftskriterium *Korrelation und Kausalität* zu erarbeiten und anzuwenden? a) Inwiefern werden die Lernziele 1-4 erreicht? b) Inwiefern reagieren Lernende *passend* und *anschlussfähig* auf Impulse der Lehrkraft?	9.2

	c) Welche *Verständnisschwierigkeiten* erfahren Lernende bei der Bearbeitung der Lernmaterialien? d) Inwiefern erfahren Lernende *Aha-Erlebnisse* bei der Bearbeitung der Lernmaterialien? e) Zu welchen Zeitpunkten des Unterrichts führen Lernende *Off-Task-Interaktionen* durch?	
1.3	Wie kann *Randomisierung* als Anforderung an empirische Studien für den Einsatz im naturwissenschaftlichen Unterricht didaktisch aufbereitet werden? Inwieweit sind Lernende in der Lage, das Wissenschaftskriterium *Randomisierung* zu erarbeiten und anzuwenden? a) Inwiefern werden die Lernziele 1-4 erreicht? b) Inwiefern reagieren Lernende *passend* und *anschlussfähig* auf Impulse der Lehrkraft? c) Welche *Verständnisschwierigkeiten* erfahren Lernende bei der Bearbeitung der Lernmaterialien? d) Inwiefern erfahren Lernende *Aha-Erlebnisse* bei der Bearbeitung der Lernmaterialien? e) Zu welchen Zeitpunkten des Unterrichts führen Lernende *Off-Task-Interaktionen* durch?	9.3
1.4	Wie kann *Reproduzierbarkeit* als Anforderung an Experimente und empirische Studien für den Einsatz im naturwissenschaftlichen Unterricht aufbereitet werden? Inwieweit sind Lernende in der Lage, das Wissenschaftskriterium *Reproduzierbarkeit* zu erarbeiten und anzuwenden? a) Inwiefern werden die Lernziele 1-3 erreicht?	9.4

	b) Inwiefern erarbeiten die Schülerinnen und Schüler selbstständig spezifische Kriterien für Reproduzierbarkeit am Beispiel des Experiments „Die Zitronenbatterie“ und benennen sie diese Kriterien? c) Inwiefern bewerten die Schülerinnen und Schüler anhand von Kriterien für Reproduzierbarkeit das Versuchsprotokoll zum Experiment „Die Zitronenbatterie“ und wie begründen sie ihre Bewertung? d) Inwiefern bewerten die Schülerinnen und Schüler anhand von Kriterien für Reproduzierbarkeit die Studie zum Thema *E-Scooter* und wie begründen sie ihre Bewertung?	

Mesozyklus 2 (Kapitel 10) *In welcher Weise kann das Thema Nahrungsergänzungsmittel als gesellschaftlich kontroverser Kontext in choice^2 reflect verwendet werden?*		
2.1.	Inwiefern erfüllt das Thema *Nahrungsergänzungsmittel* die Kriterien eines guten Kontextes?	10.1
2.2	Mit welcher Intention werden welche NEM von Schülerinnen und Schülern der gymnasialen Sekundarstufe I (Jahrgang 8 bis 10) konsumiert? a) Wie sind Vorwissen und Alltagserfahrungen bezüglich NEM bei Lernenden ausgeprägt? b) Welche Einnahmeprävalenz zeigt sich unter den Lernenden? c) Welche Verwendungsgewohnheiten geben Lernende an? d) Aus welchen Quellen beziehen Lernende NEM? e) Welche Gründe nennen Lernende für die Einnahme von NEM? f) Wie gestaltet sich das Informationsverhalten der Lernenden?	10.2

	Wie bewerten SuS verschiedene Aspekte zu NEM im Zusammenhang mit der eigenen Ernährung? a) Wie bewerten Lernende das Risiko der Einnahme von NEM? b) Inwiefern stehen Lernende NEM kritisch gegenüber?	
2.3	In welcher Weise können die Erkenntnisse aus der Fragebogenstudie in Lernmaterial der Phasen I und II integriert werden? a) Inwieweit werden beide Lernziele erreicht? b) Wie positionieren sich Lernende zur zentralen Problemfrage: „Sollte Paula gegen Muskelkrämpfe Magnesium-Tabletten einnehmen oder jeden Tag eine Banane essen“? c) Inwiefern sind in den Antworten der Lernenden *Unsicherheiten* erkennbar?	10.3
2.4	In welcher Weise können die Wissenschaftskriterien auf das Thema *Nahrungsergänzungsmittel* angewendet werden und zu einer kriteriengeleiteten Reflexion anregen? a) Welche reale Studie zur in Phase I aufgeworfenen Problematik ist geeignet, um anhand von Prüfkriterien bewertet zu werden? b) Wie gelingt es Lernenden, die Wissenschaftskriterien *Randomisierung* und *Korrelation und Kausalität* auf die reale Studie anzuwenden? c) Inwiefern treffen Lernende zum Ende des Unterrichtskonzepts eine rationale Entscheidung?	10.4

Mesozyklus 3 (Kapitel 11) *Wie können Lernende dabei unterstützt werden, das methodische Wissen aus Phase III auf gesellschaftliche Kontroversen anzuwenden?*		
3.1	In welchem Verhältnis stehen die Prüfkriterien zueinander? a) Inwiefern trägt die Unterscheidung von *übergeordneten Prinzipien* und *Goldstandard-Kriterien* nach Jungkamp (2021) zur Verortung der Prüfkriterien bei? b) Inwiefern existieren Abhängigkeiten zwischen den Prüfkriterien? c) Inwiefern unterscheiden sich die Anwendungsbereiche der Prüfkriterien?	11.1
3.2	In welcher Weise kann die Verortung der Prüfkriterien als methodisches Instrument in das Unterrichtskonzept choice^{2}reflect integriert werden? a) Welche zentralen Gestaltungselemente benötigt die *linke Seite* des Prüfschemas? b) Welche zentralen Gestaltungselemente benötigt die *rechte Seite* des Prüfschemas? c) In welcher Weise kann der Übergang von der linken zur rechten Seite gestaltet werden? d) Inwiefern ist die Verwendung einer plakativen Symbolsprache im Prüfschema angemessen?	11.2

3.3	Welche Herangehensweise verfolgen Studierende bei der Bewertung eines kontroversen Themas unter Nutzung des entwickelten Prüfschemas? a) Welches Vorwissen haben die Studierenden bereits zu den einzelnen Prüfkriterien? b) Inwiefern gehen die Studierenden gemäß der europäischen Leserichtung den richtigen Weg des Schemas? c) Inwiefern treffen die Studierenden mithilfe des Prüfschemas eine rationale Entscheidung bei der Bewertung einer These oder Studie? d) Welche Reaktionen zeigen Studierende beim ersten Blick auf das Prüfschema?	11.4

7.1 Forschungsfragen des Mesozyklus 1

Der erste Mesozyklus dieses Projekts steht unter der Fragestellung:

In welcher Weise können die Wissenschaftskriterien Falsifizierbarkeit, Korrelation und Kausalität, Randomisierung und Reproduzierbarkeit im Unterricht expliziert werden?

In einem zweiten Schritt werden Lernmaterialien entwickelt und hinsichtlich folgender Fragestellung erprobt:

Inwieweit sind Lernende in der Lage, die Wissenschaftskriterien Falsifizierbarkeit, Korrelation und Kausalität, Randomisierung und Reproduzierbarkeit zu erarbeiten und anzuwenden?

Damit wird an das Desiderat 1 angeknüpft, das eine Entwicklung von Unterrichtsmaterialien für weitere Prüfkriterien fordert.
Die Entwicklung und Erprobung jedes Prüfkriteriums bildet dabei einen Durchlauf des Mesozyklus und kann somit auch als *Mikrozyklus* (MiZ) bezeichnet werden. Es

ergeben sich somit vier MiZ, in denen Sub-Forschungsfragen für eine strukturierte Erforschung der übergeordneten Frage sorgen.
Die Daten der Material-Erprobungen werden hinsichtlich des Erreichens der zuvor formulierten Lernziele untersucht (Sub-Frage a)). Weitere Sub-Forschungsfragen sind spezifisch für die jeweiligen Prüfkriterien.
Zur Falsifizierbarkeit wird untersucht, welche Aspekte die Lernenden bei der Einschätzung der Überprüfbarkeit von Aussagen anführen und inwiefern eine Unzufriedenheit mit dem Konzept des Beweisens von Aussagen bei den Lernenden zu erkennen ist.
Bei den Erprobungen der Kriterien *Korrelation und Kausalität* sowie *Randomisierung* werden Impulse der Lehrkraft in Plenumsphasen auf *Passung* und *Anschlussfähigkeit* untersucht, außerdem werden *Verständnisschwierigkeiten*, *Aha-Erlebnisse* und *Off-Task-Interaktionen* analysiert.
Die Sub-Forschungsfragen zur Reproduzierbarkeit beziehen sich direkt auf die Lernziele.

7.2 Forschungsfragen des Mesozyklus 2

Aus dem Desiderat 3 ist die Idee entstanden, choice^{2}reflect auch in weiteren gesellschaftlich kontrovers diskutierten Themen anzuwenden.
Um den Fokus auf die *kriteriengeleitete Bewertung* der Kontroverse legen zu können, bedarf es eines Kontextes, der zwar kontrovers ist, aber nicht derart stark emotional diskutiert wird, wie es bei Homöopathie der Fall ist.
Im Mesozyklus 2 dieser Arbeit wird folgende Fragestellung geprüft:

In welcher Weise kann das Thema Nahrungsergänzungsmittel als gesellschaftlich kontroverser Kontext in choice^{2}reflect verwendet werden?

Der Mesozyklus ist durch vier Mikrozyklen gekennzeichnet.
Im MiZ 2.1 wird geprüft, inwiefern das Thema Nahrungsergänzungsmittel die Kriterien eines guten Kontexts erfüllt.

MiZ 2.2 enthält eine Fragebogenerhebung mit 237 Schüler:innen, die das Konsumverhalten von Lernenden bezüglich NEM untersucht.

Sie gliedert sich in Analysen zu a) Vorwissen und Alltagserfahrungen, b) Einnahmeprävalenzen, c) Verwendungsgewohnheiten, d) Bezugsquellen, e) Einnahmegründen und f) Informationsverhalten der Lernenden. Weiterhin wird untersucht, wie die Schüler:innen Risiken von NEM einschätzen und ob sie kritisch gegenüber NEM eingestellt sind.

MiZ 2.3 untersucht die Frage: *In welcher Weise können die Erkenntnisse aus der Fragebogenstudei in Lernmaterial der Phasen I und II integriert werden?* Dabei werden Möglichkeiten entwickelt, das Thema NEM im *Fachspezifischen Problemaufriss* und der *Positionierung* von choice^{2}reflect zu behandeln.
Es wird geprüft, a) inwiefern die Lernziele erreicht werden, b) wie sich die Lernenden zur Problemfrage positionieren und c) inwiefern die Lernenden bei der Beantwortung Unsicherheiten zeigen. Die Ergebnisse können verwendet werden, um die Eignung von NEM als Kontroverse in der Konzeption herauszustellen.

Schließlich wird im MiZ 2.4 auch Lernmaterial für die Phasen IV und V entwickelt. Dazu dient die Forschungsfrage: *In welcher Weise können die Wissenschaftskriterien auf das Thema NEM angewendet werden und zu einer kriteriengeleiteten Reflexion anregen?*
Die Anwendung und Reflexion sollte anhand von realen empirischen Studien durchgeführt werden. Daher wird mit der Sub-Frage a) eine geeignete Studie identifiziert: *Welche reale Studie zur in Phase I aufgeworfenen Problematik ist geeignet, um anhand von Prüfkriterien bewertet zu werden?*
Randomisierung und *Korrelation und Kausalität* sollen von den Lernenden auf eine reale Studie angewendet werden. Sub-Frage b) begleitet die Entwicklung und Erprobung passenden Lernmaterials: *Wie gelingt es Lernenden, die Wissenschaftskriterien Randomisierung und Korrelation und Kausalität auf die reale Studie anzuwenden?*
Inwiefern die Lernenden zum Ende des Unterrichtskonzepts eine rationale Entscheidung treffen, wird zum Ende des MiZ 2.4 untersucht.

7.3 Forschungsfragen des Mesozyklus 3

Jungkamp (2021) hat bereits angedeutet, dass die Prüfkriterien zueinander in Beziehung stehen und teils auch auf verschiedene Bereiche angewendet werden. Dazu unterscheidet er zwischen *übergeordneten Prinzipien* und *Goldstandard-Kriterien*.

Er expliziert die Notwendigkeit weitere Forschung auf dem Gebiet mit dem Desiderat 2. Der Forderung begegnet der dritte Mesozyklus dieses Projekts mit der Fragestellung:

Wie können Lernende dabei unterstützt werden, das methodische Wissen aus Phase III auf gesellschaftliche Kontroversen anzuwenden?

Der Mesozyklus ist dreigeteilt. Zunächst wird die Verortung der Prüfkriterien durch Literatur gestützt expliziert und genauer ausdifferenziert. Dazu wird die von Jungkamp angeregte Trennung betrachtet und in die Verortung einbezogen (Sub-Frage a)). Dann werden b) Abhängigkeiten zwischen den Prüfkriterien geklärt und schließlich c) die Anwendungsbereiche genauer aufgeschlüsselt.

Die theoretisch fundierte Verortung kann auch im Unterricht expliziert werden. Die Idee, ein methodisches Instrument zu entwickeln, das die Stellung der Kriterien visuell darstellt und sie damit in choice^{2}reflect nutzbar macht, wird durch den MiZ 3.2 ausgedrückt. Die zentrale Frage ist: *In welcher Weise kann die Verortung der Prüfkriterien als methodisches Instrument in das Unterrichtskonzept choice^2 reflect integriert werden?*
Die Sub-Fragen nähern sich kleinschrittig den Gestaltungsmerkmalen des zu entwickelnden Prüfschemas. Somit werden a) linke Seite, b) rechte Seite, c) Übergänge und d) Symbolsprache detailliert erforscht.

Das entwickelte Prüfschema wird im Anschluss als Mikrozyklus 3.3 erprobt: *Welche Herangehensweise verfolgen Studierende bei der Bewertung eines kontroversen Themas unter Nutzung des entwickelten Prüfschemas?*
Im Rahmen der Erprobung wird untersucht, a) welches Vorwissen die Studierenden zu den Prüfkriterien besitzen, b) ob die Studierenden die europäische Leserichtung beim Absolvieren des Prüfschemas befolgen, c) inwiefern die Studierenden eine rationale Entscheidung bei der Bewertung treffen und d) welche Reaktionen sie beim ersten Blick auf das Schema zeigen.

Teil V

Design-Experiment

Im Framing (siehe Kapitel IV) wurde ein bildungspraktisches Problem identifiziert. Gepaart mit der Analyse bereits bestehender Forschungserkenntnisse aus einem Vorgängerprojekt sowie einer intensiven Recherche und Darlegung fachspezifischer Literatur wurden bisher ungelöste Probleme offengelegt und Desiderata hervorgebracht. Auf dieser Grundlage wurden Zielsetzungen und Fragestellungen für dieses Projekt definiert.

Das folgende Design-Experiment dient der Beantwortung der übergeordneten Fragestellung mittels Entwicklung und Erprobung von Unterrichtsmaterial. Dieser iterative Prozess beschreibt einen Makrozyklus, der in einzelne Mesozyklen mit eigenen Fragestellungen unterteilt ist.

Das methodische Vorgehen innerhalb des Makrozyklus besteht in der Aufteilung der grundlegenden Fragestellung in Teilbereiche, die anhand von einzelnen Mesozkylen erforscht werden. Dabei werden stets die vier Phasen *Synthese und Vorbereitung*, *Design*, *Erprobung* und *Analyse* innerhalb eines Zyklus durchlaufen. Diese Phasierung wird im Laufe des Projekts verfolgt, allerdings in den Kapiteln der Mesozyklen nicht näher durch Kapitel-Gliederungen expliziert.

Eine detaillierte übergreifende Vorgehensweise wird an dieser Stelle nicht definiert, denn die jeweils passende Forschungsmethode wird zu Beginn jedes Mesozyklus festgelegt und beschrieben.

Im ersten Mesozyklus (Kapitel 9) werden, anknüpfend an das Vorgängerprojekt von Jungkamp, weitere Unterrichtsmaterialien zur Erarbeitung und Anwendung der wissenschaftlichen Prüfkriterien entwickelt und erprobt.

Die Unterrichtskonzeption um zusätzliche Prüfkriteriums-Materialien zu erweitern ist nur ein Teilbereich des Projekts. Des Weiteren steht die Aufgabe im Vordergrund, den Anwendungsbereich zu vergrößern. Konkret bedeutet das, ein weiteres kontroverses Thema zu beleuchten und für die Verwendung in choice^{2}reflect aufzuarbeiten. Dieser Aspekt wird im Kapitel 10 behandelt.

Jungkamp (2021) beschreibt in seiner Arbeit die Wissenschaftskriterien zunächst als gleichwertig und nebeneinander-stehend. Im späteren Verlauf führt er eine Trennung zwischen übergeordneten Prinzipien und Goldstandard-Kriterien ein. Das Kapitel zum dritten Mesozyklus dieser Arbeit (Kapitel 11) widmet sich der Betrachtung des Verhältnisses einzelner Prüfkriterien zueinander und einer Explizierung der Vorgehensweise bei der Prüfung auf Wissenschaftlichkeit im Unterricht.

8 Übersicht der Erprobungen und Forschungsmethodik

In den Kapiteln 9 bis 11 wird die genaue Vorgehensweise des Durchlaufens der Mesozyklen dargelegt. Jeder Mesozyklus behandelt eigene Fragestellungen und enthält die Entwicklung und Erprobung von Lernmaterialien zur Lösung der zuvor identifizierten bildungspraktischen Probleme.
Die Planung und Chronologie der in diesem Projekt durchgeführten Erprobungen folgt jedoch nicht strikt der Einteilung in Mesozyklen. In einigen Fällen wurden beispielsweise Materialien aus verschiedenen Mesozyklen mit derselben Lerngruppe direkt nacheinander erprobt.
In der Abbildung 8.1 befindet sich eine Übersicht über alle während des Projektverlaufs durchgeführten Erprobungen. Im folgenden Kapitel werden diese Erprobungen kurz erläutert und die jeweils verwendeten Forschungsmethoden werden angegeben.

8.1 Datenerhebung

8.1.1 Erprobung des Prüfkriteriums Falsifizierbarkeit

Lernmaterialien zum Prüfkriterium *Falsifizierbarkeit* wurden bereits im Jahr 2017 im Rahmen einer Masterarbeit von Kralisch (2017) entwickelt und erprobt. Es wurde untersucht, in welcher Weise Schüler:innen der 9. Klasse eines Gymnasiums in Münster das Unterrichtsmaterial erarbeiten und inwiefern sie im Zusammenhang mit dem Konzept der Falsifizierbarkeit kognitive Konflikte erleben. An der Erprobung am 12.09.2017 nahmen 14 Lernende teil, die Datenerhebung bestand in der Audiographie der Doppelstunde nach unten abgebildetem Raumplan (Abbildung 8.2). Die wichtigsten Erkenntnisse aus der Erprobung werden im Kapitel 9.1 erläutert

Abbildung 8.1: *Übersicht über alle durchgeführten Erprobungen. Horizontale Achse bildet den zeitlichen Projektverlauf ab. QIA = Qualitative Inhaltsanalyse.*

Projektverlauf	2017	2018	
Typ	Erprobung	Fragebogenerhebung	Erprobung
Inhalt	Prüfkriterium: Falsifizierbarkeit	Konsumverhalten Nahrungs-ergänzungsmittel	Phase I/II: Magnesium gegen Muskelkrämpfe?
Ort, Schule	Münster, Gymnasium	Münster, 2 Gymnasien	Münster, Gymnasium
Umfang	1x 90 min.	-	2x 90 min.
Teilnehmende	14 Schüler:innen, 9. Klasse	237 Schüler:innen, 8. bis 10. Klasse	26 Schüler:innen, 9. Klasse
Datenerhebung	Audiographie, Arbeitsblätter	Fragebogen	Audio- & Videographie, Arbeitsblätter
Datenauswertung	QIA	deskriptive Statistik, QIA	QIA
Mesozyklus; Forschungsfragen	1.1	2.2	2.3
Kapitel	9.1	10.2	10.3

2019	2020	2021	2022
Erprobung	Gesamterprobung	Pilotierende Erprobung im Hochschulkontext	
Prüfkriterium: Reproduzierbarkeit	Phase I/II; Prüfkriterien: Randomisierung & Korrelation und Kausalität; Phase IV/V	Erhebung Vorwissen zu Prüfkriterien; Bewertung der Kontroverse *Veganismus* anhand des Prüfschemas	
Münster, Gymnasium	Lotte-Westerkappeln, Gesamtschule	online-Konferenz; chemiedidaktisches Seminar an der Universität Münster	
1x 90 min.	2x 90 min. 1x 45 min.	2x 90 min.	
26 Schüler:innen, 9. Klasse	27 Schüler:innen, 9. Klasse	je 26 Studierende, Bachelor	
Audio- & Videographie, Arbeitsblätter	Audio- & Videographie, Arbeitsblätter, digital erstellte Notizen der Lernenden	Audio- & Videographie (Aufzeichnung der Konferenz), online-Fragebogen	
QIA	QIA	deskriptive Statistik, QIA	
1.4	1.2; 1.3 2.3; 2.4 (3.3)	3.3	
9.4	9.2; 9.3 10.3; 10.4 (11.4)	11.4	

und eine genaue Beschreibung mit tiefgehenden Einblicken in die theoriebasierte Entwicklung des Lernmaterials findet sich bei Kralisch (2017).

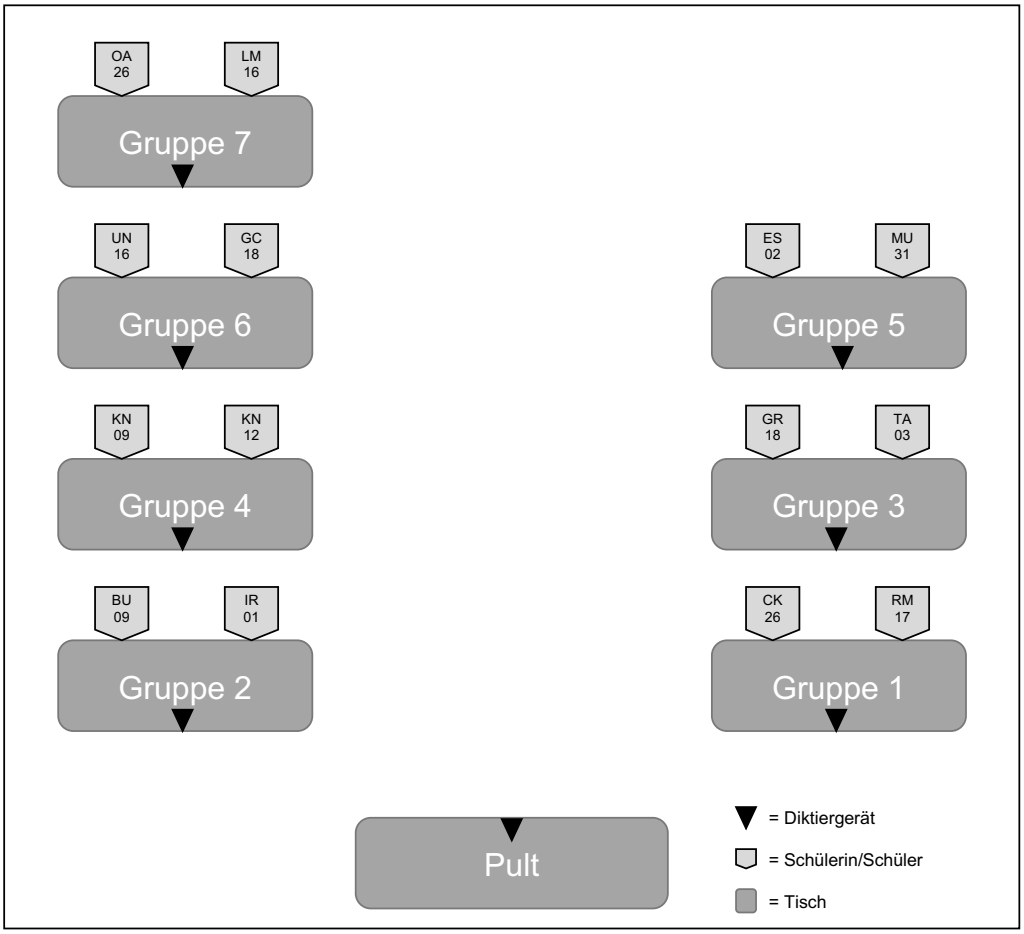


Abbildung 8.2: *Darstellung der Sitzpositionen der Schüler:innen in der Doppelstunde am 12.09.2017 zur Falsifizierbarkeit und Darstellung der verwendeten technischen Geräte zur Datenerhebung.*

8.1.2 Fragebogenerhebung zu Nahrungsergänzungsmitteln

Die Unterrichtskonzeption choice^{2}reflect ist fächer- sowie kontextübergreifend gestaltet und somit auch auf Kontroversen anwendbar, die nicht das von Jungkamp (2021) beschriebene Thema *Homöopathie* behandeln. Die Entwicklung und Erprobung von Lernmaterial im Kontext *Nahrungsergänzungsmittel* (NEM) ist ein wichtiger Teil dieses Projekts. Er findet im Rahmen des Mesozyklus 2 statt und wird im Ka-

pitel 10 behandelt. Dabei wird anhand einer Fragebogenerhebung ermittelt, welche Vorerfahrungen Lernende im Umgang mit NEM besitzen und wie häufig sie NEM mit welcher Intention konsumieren. Weiterhin äußern die Lernenden ihre Meinungen zu verschiedenen Aussagen mit Bezug zu NEM.
An der Erhebung haben 236 Lernende aus der achten, neunten und zehnten Jahrgangsstufe zweier Gymnasien in Münster teilgenommen. Die Auswertung der Fragebögen mit dem Ziel, die Eignung der Kontroverse um Nahrungsergänzungsmittel zu überprüfen, wird im Kapitel 10 behandelt.

8.1.3 Erprobung von Phase I und II

Im zweiten Mesozyklus dieses Projekts entstehen Arbeitsblätter und weitere Materialien im kontroversen Kontext *Nahrungsergänzungsmittel* mit dem Ziel, die Kontroverse um NEM zum zentralen Thema in den Phasen I, II, IV und V der Unterrichtskonzeption zu machen. Es werden Vorschläge für mögliche Stundenverlaufspläne diskutiert sowie Lernmaterial ausgearbeitet und im praktischen Unterrichtseinsatz erprobt.
Ein Arbeitsblatt zur Effektivität von Magnesium-Supplementierung bei Muskelkrämpfen wird im Rahmen einer Doppelstunde in einer achten Klasse eines Gymnasiums in Münster erprobt. Im Fokus der Untersuchung stehen die Fragen, wie die 26 Lernenden auf den fachspezifischen Problemaufriss reagieren und welche Position sie gegenüber Magnesium-Supplementen einnehmen. Die Unterrichtsstunde wurde durch Bild- und Tonaufnahmen erfasst. Diese werden im Kapitel 10.3 betrachtet und interpretiert.

8.1.4 Erprobung des Prüfkriteriums Reproduzierbarkeit

Die Entwicklung und Erprobung von Unterrichtsmaterial zur Erarbeitung wissenschaftlicher Prüfkriterien ist die zentrale Zielsetzung des ersten Mesozyklus. Die Prüfkriterien werden jeweils an eigenständigen Kontexten erarbeitet und das Kriterium *Reproduzierbarkeit* steht im Kontext der Nachhaltigkeit von E-Scootern. Eine Erprobung dieser Materialien fand im November 2019 mit 26 Lernenden einer neunten Klasse an einem Gymnasium in Münster statt. Die Unterrichtsstunde wurde videographiert sowie audiographiert und qualitativ-inhaltsanalytisch untersucht. Die detail-

lierte Beschreibung des Entwicklungsprozesses mit anschließender Planung, Durchführung und Analyse der Materialien zur ersten und zweiten Phase der Unterrichtskonzeption befindet sich im Kapitel 9.4.

8.1.5 Gesamterprobung

Im September 2020 wurde eine umfangreiche Erprobung durchgeführt, die sich über eine Unterrichtszeit von 225 Minuten in einer 2-2-1 bzw. 90-90-45 Aufteilung erstreckte. Das entspricht zwei Doppelstunden und einer Einzelstunde.
An der Unterrichtseinheit nahmen insgesamt 27 Schüler:innen der 9. Klasse einer Gesamtschule teil. In der ersten Doppelstunde waren 25 Lernende anwesend, in der zweiten Doppelstunde 18 Lernende und in der Einzelstunde ebenfalls 18 Lernende.
In den beiden Doppelstunden hat jede:r Schüler:in auf dem Arbeitsplatz ein Tablet der Marke Apple (Modell iPad Air 3. Generation) zur Verfügung gestellt bekommen. Die Tablets verfügten über einen Internetzugang und die App „Schoolwork". Über diese App konnten die Schüler:innen alle für den Unterricht benötigten Materialien abrufen. Die weitere Benutzeroberfläche der Tablets wurde durch die Lehrkraft gesperrt, sodass lediglich die benötigte Applikation verfügbar war.
Die Datenerhebung erfolgte durch Video- und Audiographie der gesamten Unterrichtsstunden. Eine Übersicht der verwendeten technischen Geräte sowie Sitzpläne befinden sich in der Abbildung 8.3. Dort ist auch erkennbar, dass die Datenerhebung in der ersten Doppelstunde mit 16 Lernenden und in der zweiten Doppelstunde mit 14 Lernenden stattgefunden hat. Lernende, die nicht an der Erhebung teilnahmen, haben dem Unterricht in einem abgetrennten Bereich des Raums beigewohnt.
In der Erprobung wurden alle Phasen der Konzeption choice^{2}reflect einmal durchlaufen. Es fand ein fachspezifischer Problemaufriss (Phase I) zum Thema „Magnesium gegen Muskelkrämpfe" mit anschließender Positionierung (Phase II) statt. In der Phase III wurden die Prüfkriterien *Randomisierung* und *Korrelation vs. Kausalität* erarbeitet. Den Abschluss bildeten Anwendung (Phase IV) und kriteriengeleitete Reflexion (Phase V). Ein Ablaufschema dieser Erprobung ist in der Tabelle 8.1 dargestellt.
Die Inhalte der Gesamterprobung lassen sich unterschiedlichen Mesozyklen des Projekts zuordnen und werden daher an verschiedenen Stellen dieser Arbeit dargestellt. Der erste Teil der ersten Doppelstunde sowie die Einzelstunde stehen im übergeordneten Kontext *Nahrungsergänzungsmittel*. Sie bilden den Rahmen für die Phasen I, II,

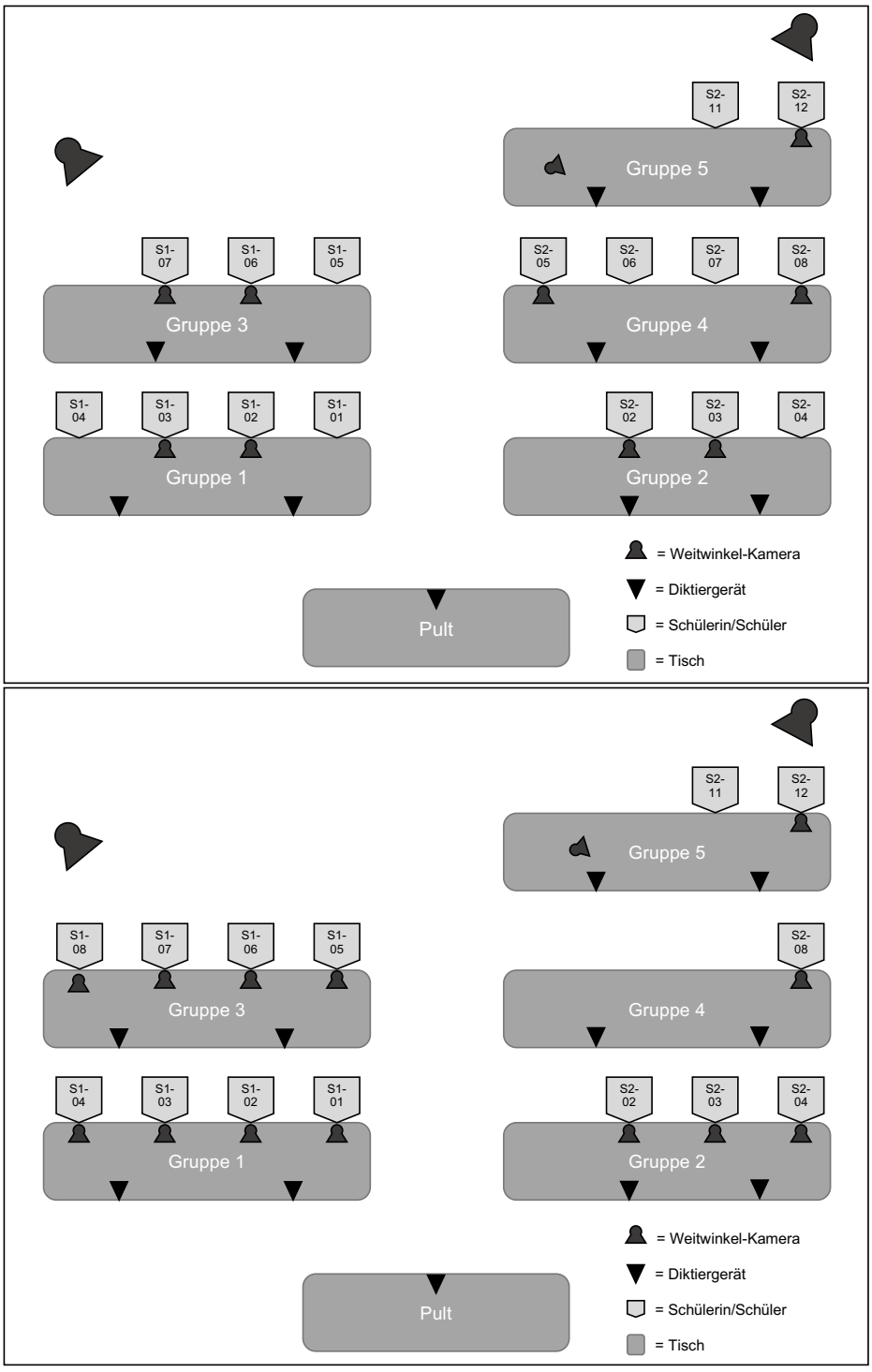


Abbildung 8.3: *Darstellung der Sitzpositionen der Schüler:innen in der ersten (oben) und zweiten (unten) Doppelstunde am 21. bzw. 28.09.2020 und der verwendeten technischen Geräte zur Datenerhebung.*

Tabelle 8.1: *Ablaufschema der Gesamterprobung im September 2020. Spaltenbreiten proportional zur Dauer der Unterrichtsstunden.*

	90 min	**90 min**	**45 min**
Phase	I, II und III	III	IV und V
Kontext	45 min: NEM 45 min: *Randomisierung*	Wdh. *Randomisierung* und Erarbeitung *Korrelation vs. Kausalität*	NEM

IV und V und sind somit dem Mesozyklus 2 zugehörig.
Die Erarbeitung der beiden Prüfkriterien lässt sich Mesozyklus 1 zuordnen. Die Entwicklung der Materialien und Analyse der erhobenen Daten werden in den Kapiteln 9 (Mesozyklus 1) und 10 (Mesozyklus 2) genauer beschrieben.

8.1.6 Erprobung des Prüfschemas

Jungkamp (2021) beschreibt, dass die wissenschaftlichen Prüfkriterien verschiedene Ausrichtungen besitzen. Während einerseits von *übergeordneten Prinzipien* die Rede ist, werden andere Prüfkriterien dem *Goldstandard naturwissenschaftlicher Untersuchungen* zugeschrieben (Jungkamp 2021, S. 171). Diese Verortung der Prüfkriterien wird im Mesozyklus 3 dieses Projekts weiter verfolgt und in Form eines Prüfschemas konkretisiert, das für den direkten Einsatz im Unterricht konzipiert ist.
Im Rahmen eines chemiedidaktischen Seminars an der Universität Münster wurde das Prüfschema im Juli 2021 mit Studierenden aus zwei Lerngruppen erprobt. Die Erprobung fand im Zusammenhang mit der Masterarbeit von Theresa Flach statt. Dabei wurde in Zusammenarbeit mit Flach untersucht, in welcher Weise die Studierenden das Prüfschema bei der Bewertung einer kontroversen Schlagzeile im Zusammenhang mit einer wissenschaftlichen Studie verwenden (Flach 2021). Die Erhebung wurde aufgrund der zum Erhebungszeitpunkt geltenden Beschränkungen durch die Corona-Pandemie in Form einer Online-Videokonferenz durchgeführt.
An der ersten Erprobung nahmen 26 Studierende teil (17 weiblich, acht männlich, einmal keine Angabe). 20 Teilnehmende befanden sich im sechsten Fachsemester, je zwei im achten und neunten Fachsemester und je ein:e Teilnehmende:r im vierten und achten Fachsemester. 15 Lernende kannten die Konzeption choice^{2}reflect vor der Erhebung noch nicht, elf gaben an, sie bereits zu kennen. Es wurden 13 Gruppen

mit folgender Aufteilung gebildet: Elf Zweiergruppen, eine Dreiergruppe, eine Einzelperson.
An der zweiten Sitzung nahmen ebenfalls 26 Studierende teil (13 weiblich, 13 männlich). Von den Teilnehmenden befanden sich 20 im sechsten Fachsemester, vier im vierten, eine:r im zweiten und eine:r im 14. Fachsemester. 21 gaben an, die Unterrichtskonzeption choice^{2}reflect noch nicht zu kennen. Zwei kannten sie bereits aus einem Seminar, eine:r aus einer Vorlesung und eine:r aus der eigenen Bachelorarbeit. Die Teilnehmenden wurden auf 13 Gruppen zu je zwei Personen verteilt.
Die online-Gruppenmeetings wurden aufgezeichnet.
Die Studierenden erhielten zunächst eine Instruktion in die Technik, füllten anschließend einen offenen Fragebogen aus und bearbeiteten schließlich zwei Arbeitsaufträge. Ein Ablaufplan ist in der Abbildung 11.4 (Seite 272) dargestellt, die detaillierten Hintergründe zur Konzeption der Seminarsitzung werden bei Flach (2021) erläutert.
Die Lernenden wurden instruiert, die Methode *lautes Denken* bei der Bearbeitung der Aufgaben anzuwenden. Der erste Arbeitsauftrag hatte die Funktion einer Übungsaufgabe, in der ein fachlich irrelevanter Kontext gewählt wurde, zu dem die Studierenden ihre Überlegungen laut aussprechen sollten. Anschließend führten die Lernenden die fachlich relevante Aufgabe durch, die die Bewertung einer Schlagzeile durch Prüfen einer Studie mithilfe des Prüfschemas beinhaltete. Die Erprobung wird im Kapitel zum dritten Mesozyklus (11) behandelt.

8.2 Datenauswertung

8.2.1 Qualitative Inhaltsanalyse nach Kuckartz

Die im Zuge des Mesozyklus 1 erhobenen Daten liegen in Form von Unterrichtsvideos in der Plenums-Perspektive, Kleingruppen-Perspektive und auch Individual-Perspektive vor. Weiterhin wurden die von den Lernenden angefertigten und ausgefüllten Lernmaterialien als Artefakte archiviert und zur Datenauswertung bereitgestellt.
Diese Fülle an Rohdaten systematisch zu analysieren gelingt durch die Methode der *inhaltlich strukturierenden qualitativen Inhaltsanalyse* (QIA), wie sie von Kuckartz (2018) beschrieben wurde. Dabei werden, ausgehend von einer zentralen Forschungsfrage, sieben Schritte zur Inhaltsanalyse durchgeführt:

1. Transformation der audiovisuellen Inhalte in eine Textform mittels Transkription und initiierende Textarbeit
2. Entwickeln von thematischen Hauptkategorien
3. Codieren des gesamten Materials mit den Hauptkategorien
4. Zusammenstellen aller mit der gleichen Hauptkategorie codierten Textstellen
5. Induktives Bestimmen von Subkategorien am Material
6. Codieren des kompletten Materials mit dem ausdifferenzierten Kategoriensystem
7. Analysen und Visualisierungen

Die Transformation der in diesem Projekt generierten audiovisuellen Inhalte geschieht in der Regel mit der Software *f4transkript*. Dabei werden aus Ton- und Videoaufnahmen nach vorgegebenen Transkriptionsregeln Textdokumente erstellt, die den Inhalt der Mediendateien schriftlich wiedergeben (vgl. Rädiker & Kuckartz 2019, S. 44f). Dabei werden einzelne Sinneinheiten in Absätze gegliedert und mit Zeitmarken der Form *##hh:mm:ss##* versehen.
Die in eine Textform transformierten Inhalte werden in eine Software zur computergestützten qualitativen Inhaltsanalyse, in diesem Fall das Programm *MAXQDA2020* eingelesen und mit der zugehörigen Mediendatei verknüpft. In MAXQDA lässt sich nahtlos zwischen den verschiedenen Darstellungsformen der Forschungsdaten wechseln, sodass weitreichende Codierungs- und Analyseverfahren möglich sind.
In den nächsten Schritten der QIA werden thematische Hauptkategorien für die Analyse des Materials entwickelt und diese Kategorien als Grundlage für nachfolgende Codierungen verwendet. Die Auswertungsverfahren in diesem Projekt verlaufen meist in niedrig-inferenten Event-Sampling-Verfahren. Dabei sind die Zeiteinheiten der Codes nicht festgelegt; Anfang und Ende der Codier-Einheiten müssen im Videomaterial individuell gesetzt werden. Das Setzen eines Codes wird durch das Auftreten von *Events* ausgelöst, die in der Code-Definition präzise festgelegt wurden (Hugener et al. 2006; Bakeman & Gottman 1997; Lotz 2016).
In einigen Fällen erfordern die Forschungsschwerpunkte ein höher inferentes Auswertungsverfahren, wobei Rating- und Schätzverfahren zum Einsatz kommen. Zuvor definierte Ereignisse werden in ihrer Qualität eingeschätzt (Hugener et al. 2006; Lotz 2016).

Die Einzelheiten der Kategorien-Genese und Codierabläufe werden in den jeweiligen Unterkapiteln des MZ 1 (Kapitel 9) dargestellt.

Auch die im Mesozyklus 3 erhobenen Daten werden durch eine strukturierende qualitative Inhaltsanalyse nach oben genanntem Schema ausgewertet. Dieses Vorgehen wird im Kapitel 11 beschrieben.

8.2.2 weitere Analysemethoden

Die Unterrichtskonzeption choice^{2}reflect auf einen weiteren Kontext zu erweitern ist die zentrale Aufgabe des MZ 2 (siehe Kapitel 10). Dabei wird die Kontroverse um *Nahrungsergänzungsmittel* thematisiert und durch eine Fragebogenerhebung auf Relevanz im schulischen Kontext geprüft. Bei der Auswertung der Fragebögen kommen größtenteils Methoden der deskriptiven Statistik zum Einsatz. Somit werden für verschiedene Fragebogenitems Häufigkeitsanalysen und Mittelwertbestimmungen durchgeführt. Offene Fragebogenitems werden zusätzlich mithilfe qualitativer inhaltsanalytischer Verfahren ausgewertet. Dabei erfolgt eine induktive Kategorienbildung mit anschließender leitfadengestützter Codierung. Die Analyse dient als Bestandsaufnahme und ist nicht auf eine Hypothesenbildung ausgerichtet.

9 Mesozyklus 1: Erarbeitung und Anwendung von Prüfkriterien

Im Vorgängerprojekt von Jungkamp (2021) wurde im Mesozyklus 2 herausgestellt, auf welche Weise Kriterien eines einheitlichen, wissenschaftlichen Weltbildes die Vermittlung von Bewertungskompetenz ermöglichen können. Dabei wurden Materialien zu den Prüfkriterien *Falsifizierbarkeit*, *Kontrollierte Bedingungen*, *Kausalität & Korrelation* und *Verblindung* entwickelt und erprobt.
Die Materialien zur *Falsifizierbarkeit* entstanden projektübergreifend in enger Zusammenarbeit, daher wird der Entstehungsprozess in dieser Arbeit detailliert aufgegriffen.
Für das Prüfkriterium *Kausalität & Korrelation* wird im Rahmen dieser Arbeit ein weiterer Ansatz zur unterrichtlichen Vermittlung und Einbettung in die Unterrichtskonzeption dargestellt.
Weiterhin werden die übrigen Kriterien *Randomisierung* und *Reproduzierbarkeit* ausgearbeitet.
Der Mesozyklus 1 (im Folgenden als *MZ 1* bezeichnet) steht unter der Fragestellung:

Inwieweit sind Lernende in der Lage, die Wissenschaftskriterien *Falsifizierbarkeit*, *Korrelation vs. Kausalität*, *Randomisierung* und *Reproduzierbarkeit* zu erarbeiten und anzuwenden?

Diese Forschungsfrage erfasst die Erarbeitung und Anwendung von vier Prüfkriterien. Somit sind zur Beantwortung der Frage mehrere Durchläufe des MZ 1 notwendig. Jedes Prüfkriterium bildet einen Mikrozyklus innerhalb des MZ 1. Die Mikrozyklen zeichnen sich durch eine einheitliche Erhebungsmethodik aus: Gemäß den oben angegebenen Phasen wird zunächst Lernmaterial zum Prüfkriterium entwickelt und dieses anschließend im Praxiseinsatz erprobt. Die Erprobungen finden in realen Lernumgebungen als Einbettungen in den Fachunterricht statt. Die Durchführungen der

Unterrichtsstunden werden videographiert und die erhobenen Videos anschließend im Hinblick auf die Fragestellung des MZ 1 ausgewertet.
Die Auswertung der Daten erfolgt, wie im Kapitel 8.2 beschrieben, angelehnt an die qualitative Inhaltsanalyse nach Kuckartz (2018). Sie wird nach folgendem Ablaufschema durchgeführt:

1. Transkription der audiovisuellen Daten zur Überführung in eine Textform
2. Entwickeln von thematischen Hauptkategorien
3. Codieren des gesamten Materials mit den Hauptkategorien
4. Zusammenstellen aller mit der gleichen Hauptkategorie codierten Textstellen
5. Induktives Bestimmen von Subkategorien am Material
6. Codieren des kompletten Materials mit dem ausdifferenzierten Kategoriensystem durch zwei Codierende
7. Analysen und Visualisierungen

Die meisten Schritte werden in den Durchläufen der Mesozyklen identisch durchgeführt. Der Anteil deduktiv gebildeter Kategorien in Schritt 2 kann jedoch variieren, außerdem verläuft in die Kategorien-Entwicklung teilweise stärker induktiv (Schritt 5).
Diese Variationen liegen darin begründet, dass zur Erarbeitung und Anwendung jedes Prüfkriteriums ein individueller Stundenverlaufsplan erstellt wird. Jede Unterrichtsstunde verläuft somit nach einem anderen Schema und ist auf das jeweilige Prüfkriterium sowie organisatorische Gegebenheiten abgestimmt. Es erscheint also nicht sinnvoll, an dieser Stelle bereits ein Kategoriensystem festzulegen, das zur Analyse aller MZ1-Durchläufe verwendet wird. Stattdessen erfolgt die Entwicklung thematischer Hauptkategorien jeweils innerhalb der MZ1-Durchläufe.

9.1 MZ 1.1: Falsifizierbarkeit

„*Ein empirisch-wissenschaftliches System muss an der Erfahrung scheitern können*“ (Popper 1989, S. 15). Diese Anforderung an Aussagensysteme formulierte Popper als Antwort auf die Unüberwindbarkeit des Induktionsproblems: Induktive Schlüsse sind

für eine endgültige Verifikation von Aussagen unzulässig, stattdessen kann anhand von deduktiv gewonnenen Einzelaussagen auf die Falschheit einer Aussage geschlossen werden. Damit gründete Popper den Falsifikationismus, dessen Hintergründe im Kapitel 2.3.2 detailliert dargestellt werden.
Aussagen müssen also *falsifizierbar* sein, um einen Beitrag zu einer rationalen Entscheidungsfindung leisten zu können. Im Unterrichtskonzept choice^{2}reflect ist Falsifizierbarkeit als Wissenschaftskriterium fest verankert. Jungkamp (2021, S. 294) ordnet Falsifizierbarkeit dabei den *übergeordneten Prinzipien* zu und grenzt sie damit von den *Goldstandard-Kriterien* ab. In dieser Arbeit wird das Verhältnis der Wissenschaftskriterien zueinander weiter ausgeschärft und in Form eines Prüfschemas ausgedrückt, dessen Entwicklung und Erprobung Teil des dritten Mesozyklus ist (Kapitel 11). Im Prüfschema erhält Falsifizierbarkeit die Position zwischen *These* und *Hypothese*, da erst mit der Widerlegbarkeit einer These auch von einer wissenschaftlichen Hypothese gesprochen werden kann.
Im Rahmen der Masterarbeit von Kralisch (2017) wurden die wesentlichen fachlichen Inhalte der Falsifizierbarkeit didaktisch aufbereitet und in Lernmaterialien für die Verwendung im naturwissenschaftlichen Unterricht umgesetzt. Die Entwicklung der Materialien fand in enger Zusammenarbeit mit Florian Jungkamp statt und wird auch im Rahmen seiner Dissertation beschrieben (Jungkamp 2021). Für die Erarbeitung des Kriteriums wurden folgende Lernziele formuliert:

Schüler:innen...

1. ... beschreiben Falsifizierbarkeit als ein Prüfkriterium für Wissenschaftlichkeit
2. ... beschreiben problematische Aspekte der Wissenschaftlichkeit von Aussagen, Behauptungen und Hypothesen und begründen diese
3. ... beurteilen die Falsifizierbarkeit von Aussagen, Behauptungen und Hypothesen anhand verschiedener lebenspraktischer Beispiele im Kontext von Fake News in sozialen Medien
4. ... verwenden folgende Fachbegriffe zielgerichtet: falsifizierbar, nicht falsifizierbar, beweisen, widerlegen, wissenschaftlich, unwissenschaftlich

9.1.1 Entwicklung des Lernmaterials

Die Erarbeitung des Prüfkriteriums Falsifizierbarkeit ist auf eine Doppelstunde ausgerichtet. Das Lernmaterial besteht aus vier Arbeitsblättern, die im Folgenden kurz

vorgestellt werden.

Das erste Arbeitsblatt (Abbildung 9.1) stellt den Charakter *Birte* vor, der die Behauptung aufstellt: „Haselnüsse enthalten weniger Fett als Nutella“. Ein Arbeitsauftrag leitet die Schüler:innen kleinschrittig an, ein kurzes Experiment zu planen, mit dem die Behauptung überprüft werden kann. Dabei antizipieren die Lernenden, welche Beobachtungen zur Bestätigung oder zur Widerlegung der Behauptung führen würden.

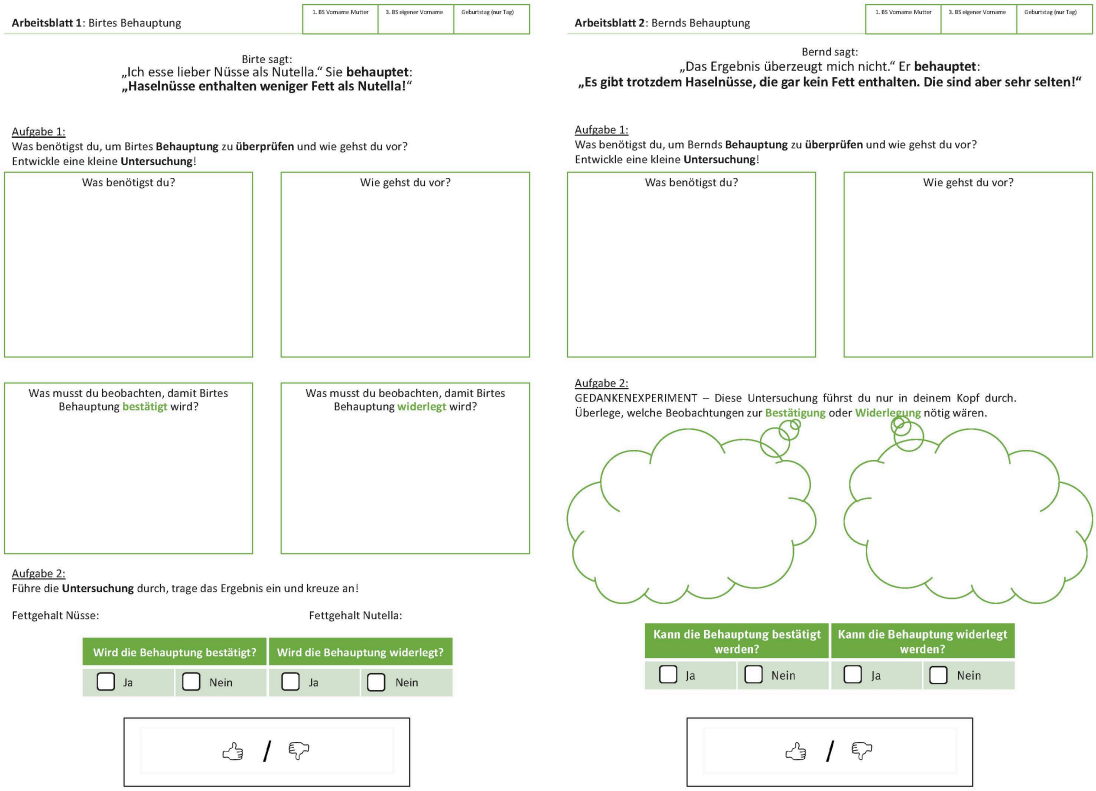


Abbildung 9.1: *Arbeitsblätter 1 (links) und 2 (rechts) zur Erarbeitung des Prüfkriteriums Falsifizierbarkeit (Kralisch 2017).*

In einer zweiten Aufgabe führen sie die Untersuchung selbst durch und kreuzen an, ob die Behauptung bestätigt oder widerlegt werden konnte. Es wird erwartet, dass die Schüler:innen die Nährwertangaben der Haselnüsse bzw. Nutella von den Verpackungen ablesen und daraufhin erkennen, dass die Behauptung widerlegt wurde - denn 100 g der im Unterricht eingesetzten Haselnüsse enthalten mehr Fett als 100 g Nutella.

Im Anschluss bearbeiten die Schüler:innen das Arbeitsblatt 2 (Abbildung 9.1). Dieses

enthält Bernds Gegenbehauptung: „Es gibt trotzdem Haselnüsse, die gar kein Fett enthalten. Die sind aber sehr selten!“ Die Lernenden planen ein Gedankenexperiment zu der Frage, welche Materialien nötig wären, um Bernds Behauptung zu überprüfen. Sie antizipieren Beobachtungen, die zu einer Bestätigung oder Widerlegung der Behauptung führen würden. Dabei sollten sie feststellen, dass das Widerlegen von Bernds Aussage gar nicht möglich ist. Theoretisch betrachtet müssten alle Haselnüsse des Universums auf ihren Fettgehalt geprüft werden - und das ist natürlich nicht möglich.
Die Aussage entzieht sich durch ihre logische Form als Existenzbehauptung jedem Versuch, sie zu widerlegen.

Anhand dieser beiden (Gedanken-)Experimente erleben die Schüler:innen, dass manche Aussagen unmöglich zu widerlegen sind. Nachdem dieses Problem im Unterricht aufgeworfen wurde, widmen die Lernenden sich dem dritten Arbeitsblatt (siehe Abbildung 9.2), das den theoretischen Hintergrund zur Falsifizierbarkeit erklärt und die wichtigsten Fachbegriffe einführt.
Bei der Entwicklung dieses Arbeitsblattes wurde ein besonderer Wert auf Sprachsensibilität gelegt. Auf Grundlage der Gestaltungskriterien nach Kulgemeyer & Starauschek (2014) kamen verständliche Formulierungen, kurze Textabschnitten mit mittlerer Satzlänge und mittlerer Silbenzahl, sowie eine hohe Textkohärenz zum Einsatz.

Das vierte Arbeitsblatt (Abbildung 9.2) dient der Anwendung des Prüfkriteriums auf die Thematik der Fake News in sozialen Medien. Lernende bewerten verschiedene Aussagen aus sozialen Medien anhand ihrer Falsifizierbarkeit und stufen sie als wissenschaftlich oder nicht-wissenschaftlich ein. Diese befinden sich im Anhang B.5.4.
In einem weiteren Arbeitsauftrag entwickeln die Lernenden Kriterien, aufgrund derer nicht-falsifizierbare Aussagen erkannt werden können. Dazu zählen: unpräzise Formulierungen, Tautologien oder auch Definitionen, Existenzbehauptungen (bzw. „Es-gibt-Sätze“), Formulierungen im Konjunktiv und Gedankengänge. Diese Kriterien sammeln die Lernenden auf einer Checkliste, die im anschließenden Plenumsgespräch diskutiert wird.

Die entwickelten Prüfkarten können in nachfolgenden Unterrichtsstunden thematisiert werden. Sie dienen nicht der direkten Erarbeitung des Prüfkriteriums - vielmehr sind sie zur Sicherung des Gelernten gedacht oder auch als Nachschlagewerk,

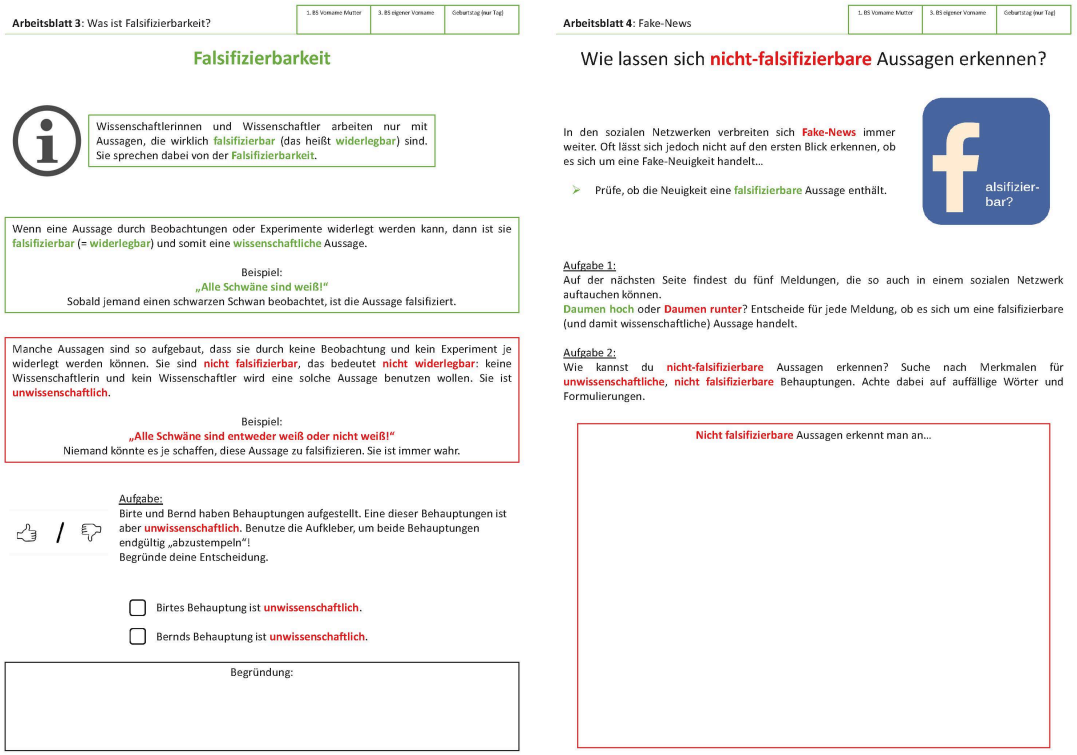


Abbildung 9.2: *Arbeitsblätter 3 (links) und 4 (rechts) zur Erprobung des Prüfkriteriums Falsifizierbarkeit (Kralisch 2017).*

das die wichtigsten Informationen zur Falsifizierbarkeit auf einen Blick bereit hält. Die Karten enthalten auch Vertiefungsaufgaben, die im nachfolgenden Unterricht eingesetzt werden können.

Sie folgen der von Jungkamp (2021) entwickelten Struktur: Eine Prüfkarte enthält generelle Informationen zur Falsifizierbarkeit („Was ist Falsifizierbarkeit?"), eine weitere Karte zeigt, wie auf Falsifizierbarkeit geprüft werden kann. Es folgt je eine Prüfkarte zur Falsifizierbarkeit in der Chemie, in der Forschung und im Alltag.

Die Prüfkarten sind in der Abbildung 9.3 dargestellt und befinden sich vergrößert auch im Anhang B.6.1.

Abbildung 9.3: *Prüfkarten zum Prüfkriterium Falsifizierbarkeit.*

9.1.2 Erprobung

Methodik

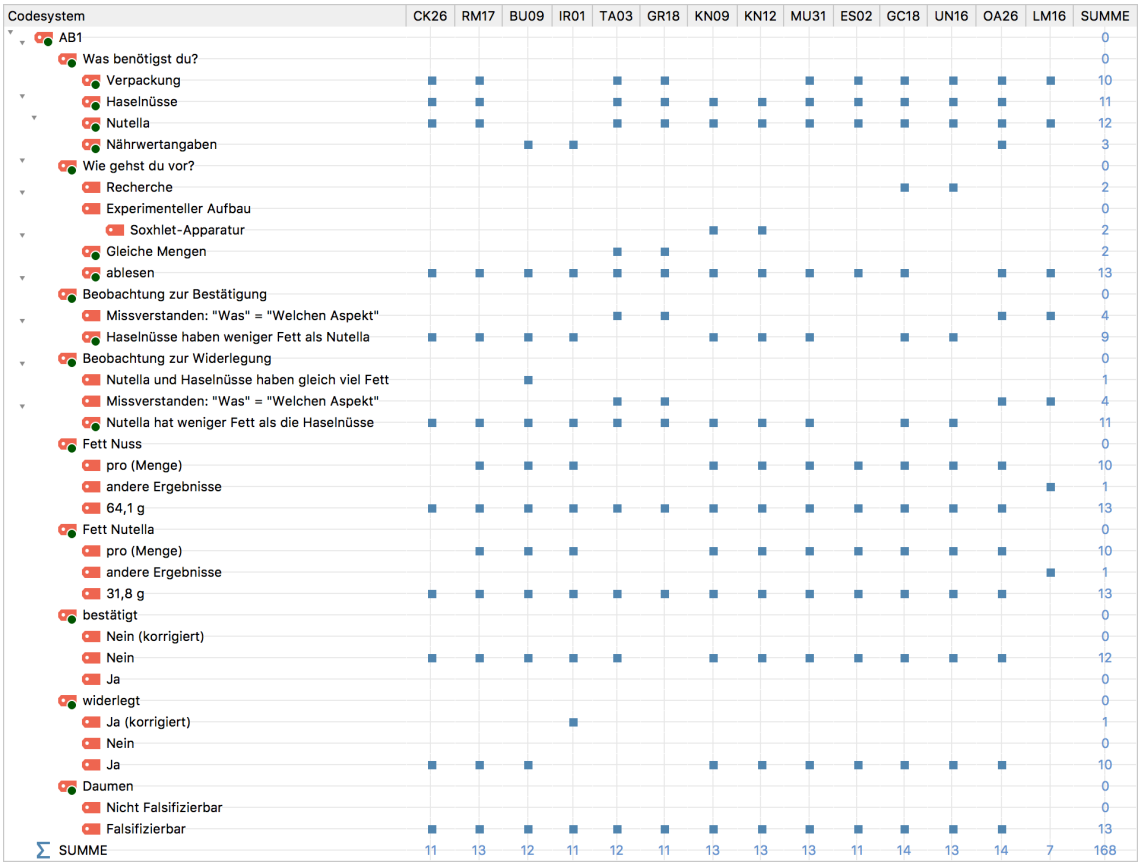
Kralisch (2017) erprobte die entwickelten Unterrichtsmaterialien mit 14 Lernenden der neunten Klasse eines Gymnasiums in Münster, die genauen Gegebenheiten sind im Kapitel 8.1.1 dargestellt.
Die audiographierte Unterrichtsstunde wurde durch qualitative Inhaltsanalyse anhand eines Kategoriensystems analysiert. Die Kategorien wurden zum Teil induktiv aus dem Material ermittelt, zu einem anderen Teil aber auch deduktiv und theoriegeleitet entwickelt. Letztere Kategorien dienten vor allem dem Aufdecken von Anzeichen auf kognitive Konflikte, die die Lernenden bei der Bearbeitung des Lernmaterials möglicherweise erfuhren. Die induktiv aufgestellten Kategorien wurden zur Analyse der von den Lernenden ausgefüllten Arbeitsblätter verwendet.
Die Kategorien richtete Kralisch (2017) entsprechend der Fragestellungen aus. Die notierten Antworten der Lernenden auf den Arbeitsblättern wurden eher explorativ unter der Fragestellung, welche Aspekte die Lernenden zur Problemlösung heranziehen untersucht, während die Audiodateien einer theoriegeleiteten Analyse unterzogen wurden (leicht angepasst für die Verwendung in diesem Projekt): „Inwiefern ist bei den Lernenden eine Unzufriedenheit mit ihrem Konzept des Beweisens von Aussagen zu erkennen?“
Das vollständige Kategoriensystem ist im Anhang B.1.1 abgebildet. Die Kategorien zur Auswertung der Arbeitsblätter gliedern sich in der ersten Hierarchieebene nach der Nummer des Arbeitsblatts und in der zweiten Ebene nach den Arbeitsaufträgen, die mit der Bearbeitung der AB zusammenhängen. Die dritte Ebene enthält die aus dem Material induktiv entwickelten Codes. Lediglich die Kategorien Aussagen_Gr1 und Aussagen_Gr2 wurden deduktiv entwickelt. Für jede Aussage des vierten AB kann somit codiert werden, ob die Lernenden die Falsifizierbarkeit der Aussage a. richtig zugeordnet mit richtiger Begründung, b. richtig zugeordnet mit falscher Begründung, c. richtig zugeordnet ohne Begründung oder d. falsch zugeordnet haben.
Codes, die (auch) direkt an den Audioaufnahmen eingesetzt werden, sind die Codes *AB2/Audiodatei* mit Subcodes, *Aussagen_Gr1* und *Aussagen_Gr2*. Bezüglich AB2 wird codiert, ob die Lernenden das Problem erkennen und inwiefern eine Hilfestellung benötigt wurde. Weiterhin gibt es eine Kategorie für selbstständig entwickelte Beispiele.

Ergebnisse

Der Ablauf der Unterrichtsstunde orientierte sich stark am zuvor geplanten Schema und wich nur selten davon ab. In der Plenumsphase zum AB1 offenbarte sich eine hohe Diskussionsbereitschaft in der Lerngruppe, denn es wurde über die Notwendigkeit diskutiert, Nährwertangaben auf Lebensmittelverpackungen als Massenanteile auszudrücken.
Noch deutlicher wurde diese Diskussionsbereitschaft in der zweiten Plenumsphase. In nahezu enthusiastischen Wortbeiträgen argumentierten die Schüler:innen, dass Bernds Aussage unmöglich zu widerlegen ist. Gruppe 5 brachte sogar ein selbstständiges Beispiel, das im späteren Material noch thematisiert werden sollte: „Es gibt Gott“ wurde als nicht widerlegbare Aussage erkannt. Kralisch (2017) vermutet, dass die Lernenden ihr Konzept des Beweisens von Aussagen um den Aspekt der Falsifikation erweitert haben.
Auf die Ankündigung, dass in folgender Unterrichtsphase *Fake News* thematisiert werden, reagierten die Lernenden mit Begeisterung. Manche riefen: „Donald Trump - You are Fake News“ spontan ins Plenum. Während der folgenden Arbeitsphase war zu beobachten, dass manchen Gruppen die Einteilung der Aussagen leicht fiel, während andere Gruppen Schwierigkeiten dabei zeigten. Die Schüleraktivität war jedoch durchgängig hoch und die meisten Gruppen arbeiteten effektiv am Material.
Im letzten Teil der Unterrichtsstunde stellten einzelne Lernende die Gruppenergebnisse vor und gaben Begründungen für die Einstufungen der Falsifizierbarkeit an. Die Diskussion der Lernenden untereinander über die verschiedenen Aussagen wurde von der Lehrkraft umgeleitet in ein Gespräch über die Kriterien für *Falsifizierbarkeit*. Die Prüfkarten wurden im Unterricht nicht mehr behandelt, sondern lediglich für den Einsatz in der nächsten Stunde ausgeteilt.

Der Code-Matrix-Browser als Visualisierung der Codierungen in MAXQDA zeigt in Abbildung 9.4, welche Codierungen für das Arbeitsblatt 1 vergeben wurden. Im Feld „Wie gehst du vor?“ gaben nahezu alle Lernenden an, die Verpackungen der Haselnüsse und Nutella abzulesen. Das widerspricht den Beobachtungen aus der Unterrichtsstunde; dort wurden diverse experimentelle Umsetzungen von mehreren Lernenden diskutiert. Auf dem Arbeitsblatt gaben nur zwei Schüler eine konkrete Versuchsvorschrift an. Neun Schülerinnen und Schüler antworteten zielführend und schrieben, man müsse Haselnüsse finden, die weniger Fett haben als Nutella, um Birtes Behaup-

tung zu bestätigen. Elf Lernende gaben korrekterweise an, man müsse zur Widerlegung herausfinden, dass eine bestimmte Menge Nutella weniger Fett hat als die gleiche Menge an Haselnüssen.



Codesystem	CK26	RM17	BU09	IR01	TA03	GR18	KN09	KN12	MU31	ES02	GC18	UN16	OA26	LM16	SUMME
AB1															0
Was benötigst du?															0
Verpackung	■	■			■	■			■	■	■	■	■	■	10
Haselnüsse	■	■			■	■	■	■	■	■	■	■	■		11
Nutella	■	■			■	■	■	■	■	■	■	■	■	■	12
Nährwertangaben			■	■									■		3
Wie gehst du vor?															0
Recherche											■	■			2
Experimenteller Aufbau															0
Soxhlet-Apparatur							■	■							2
Gleiche Mengen					■	■									2
ablesen	■	■	■	■	■	■	■	■	■	■	■		■	■	13
Beobachtung zur Bestätigung															0
Missverstanden: "Was" = "Welchen Aspekt"					■	■							■	■	4
Haselnüsse haben weniger Fett als Nutella	■	■	■	■			■	■	■		■	■			9
Beobachtung zur Widerlegung															0
Nutella und Haselnüsse haben gleich viel Fett			■												1
Missverstanden: "Was" = "Welchen Aspekt"					■	■							■	■	4
Nutella hat weniger Fett als die Haselnüsse	■	■	■	■	■	■	■	■	■		■	■			11
Fett Nuss															0
pro (Menge)		■	■	■			■	■	■	■	■	■	■		10
andere Ergebnisse														■	1
64,1 g	■	■	■	■	■	■	■	■	■	■	■	■	■		13
Fett Nutella															0
pro (Menge)		■	■	■			■	■	■	■	■	■	■		10
andere Ergebnisse														■	1
31,8 g	■	■	■	■	■	■	■	■	■	■	■	■	■		13
bestätigt															0
Nein (korrigiert)															0
Nein	■	■	■	■	■		■	■	■	■	■	■	■		12
Ja															0
widerlegt															0
Ja (korrigiert)				■											1
Nein															0
Ja	■	■	■				■	■	■	■	■	■	■		10
Daumen															0
Nicht Falsifizierbar															0
Falsifizierbar	■	■	■	■	■	■	■	■	■	■	■	■	■		13
SUMME	11	13	12	11	12	11	13	13	13	11	14	13	14	7	168

Abbildung 9.4: *Code-Matrix-Browser zur Visualisierung der Codierungen für Arbeitsblatt 1.*

Fast alle lasen die korrekten Nährwerte von den Verpackungen ab, zehn von 14 Lernenden erwähnten, dass der Fettgehalt als Anteil angegeben ist. Alle Lernenden klebten das richtige Etikett auf („falsifizierbar“).

Weniger deutlich fallen die Antworten zu AB2 aus. Acht Schüler:innen beschrieben verschiedenste experimentelle Untersuchungen, mit denen Bernds Behauptung überprüft werden könne - nur sechs Lernende gaben die korrekte Antwort an: Alle Haselnüsse der Welt (oder sogar des Universums) müssten geprüft werden, was unmöglich ist. Diese Verteilung ist in der Abbildung 9.5 erkennbar, denn die Punkte in der Matrix sind deutlich ungleicher verteilt als in der ersten Grafik.

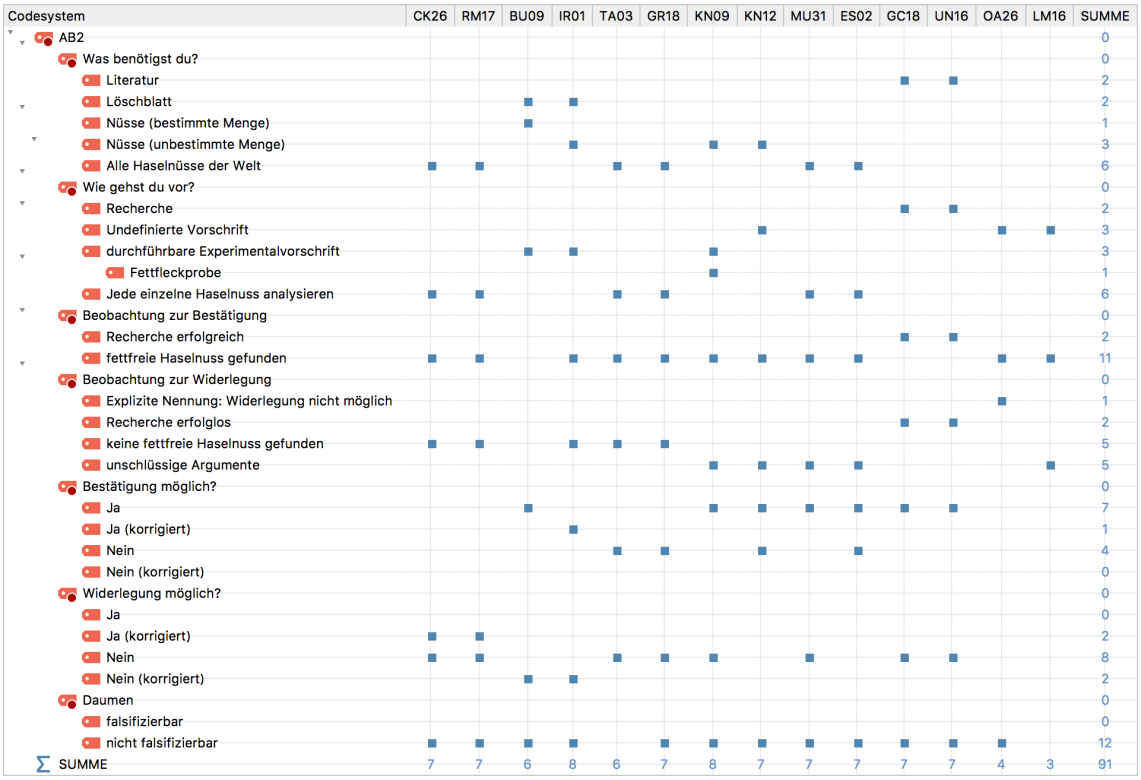


Abbildung 9.5: *Code-Matrix-Browser zur Visualisierung der Codierungen für Arbeitsblatt 2.*

Einigkeit herrscht bei der Frage, welche Beobachtungen Bernds Behauptung bestätigen würden: Die Entdeckung einer fettfreien Haselnuss oder auch eine erfolgreiche Literaturrecherche. Beide Antworten werden als fachlich korrekt angesehen.
Zehn Schülerinnen und Schüler erkannten, dass die Widerlegung von Bernds Aussage unmöglich ist.
Auch die Analyse der Audiodateien zeigt, dass in einigen Schülergruppen das zentrale Problem erkannt wurde (siehe Abbildung 9.6). Kralisch (2017, S. 54) macht diese Einschätzung an Äußerungen der Schüler:innen in den Arbeitsphasen fest:

CK26: Wir müssen noch etwas anders machen. Alle Haselnüsse der Welt, eigentlich!

KN12: Hä, aber… Das kann man nicht wirklich nachweisen. Weil, wenn er sagt die sind sehr selten, kann es ja sein, dass es nur zwei auf der Welt gibt

UN16: Ja, da müsste man eine große Menge/ Ja, aber dann sagt er ja: Joa, die sind ja trotzdem nicht dabei gewesen. Wenn man eine große Menge einfach nimmt,

aber, er kann ja einfach sagen, wenn man 1000 überprüft: Joa, das wäre die 1001. Nuss. Deswegen.

In den Gruppengesprächen entwickeln die Lernenden auch ihre eigenen Beispiele, um sich die Falsifizierbarkeit gegenseitig zu erklären:

BU09: Es gibt bestimmt auch einen Baum, der in Herzform gewachsen ist. Das [hat man bisher leider nicht gefunden], aber den gibt es bestimmt!

MU31: Weil Bernd kann halt immer sagen: Jo, irgendwo, meinetwegen im Amazonas, da ist noch eine Haselnuss und die hat kein Fett. Ich weiß zwar nicht wo, aber die ist da. Das weiß ich.

ES02: Ja, das ist ungefähr so eine Behauptung, wie wenn jemand sagt: Es gibt Gott.

UN16: Du könntest dann ja halt auch sagen: Ja, Menschen können fliegen oder so, du könntest ja alles sagen.

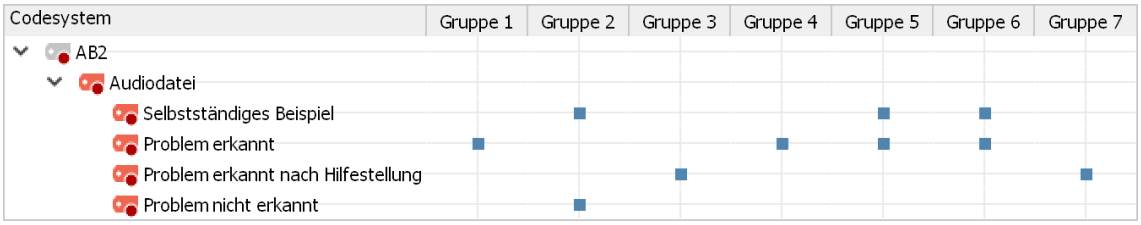


Abbildung 9.6: *Code-Matrix-Browser zur Visualisierung der Codierungen in den Audioaufnahmen der Unterrichtsstunde.*

Mit AB3 geben Lernende eine Einschätzung der Widerlegbarkeit von Bernds Aussage. Abbildung 9.6 zeigt, dass sie sich dabei einig sind. Bernds Aussage ist nicht widerlegbar. In den Begründungen erkennt man allerdings, dass nur selten der Fachbegriff *nicht falsifizierbar* genannt wird.
AB4 wurde detailliert analysiert. Für jede der zehn Beispielaussagen wurde sowohl an den Arbeitsblättern als auch im Audiomaterial codiert, wie die Lernenden sich dazu positionieren. Ausführlich beschrieben wird diese Auswertung bei Kralisch (2017). Zusammengefasst kann festgehalten werden, dass die meisten Kriterien für Falsifizierbarkeit von den Lernenden erkannt und auf der Checkliste notiert wurden.
Die Bewertungen der Beispielaussagen fielen unterschiedlich aus, es wurde aber mehr als die Hälfte der Zuordnungen korrekt vorgenommen und korrekt begründet. Problematische Aussagen konnten anhand der erhobenen Daten identifiziert werden.

Im Anhang B.4.1 sind Einzelaussagen der Lernenden aufgelistet, die von Kralisch als relevant herausgestellt wurden. Daran ist zum Beispiel erkennbar, dass die Schüler:innen die eingeführten Fachbegriffe häufig verwenden und auch nicht vor dem Fremdwort *falsifizierbar* zurückschrecken.

Kralisch (2017) resümiert, dass das entwickelte Unterrichtsmaterial gut geeignet ist, um Lernende dazu zu bewegen, sich mit Erkenntnistheorie auseinanderzusetzen. Viele Schülergruppen erkannten das zentrale Stundenproblem sofort oder nach kleineren Hilfestellungen; manche Schüler:innen entwickelten sogar eigene Beispiele zur Erklärung des Konzepts. Hohen Diskussionsbedarf gab es bei den Kriterien für Falsifizierbarkeit. Auf die Bemerkung der Lehrkraft, dass *Es-gibt-Sätze* nicht falsifizierbar sind, reagierten die Lernenden sofort und forderten Trennschärfe bei Abgrenzung der Kriterien ein.
Manche Teilaspekte von Arbeitsaufträgen konnten als nicht zielführend identifiziert werden. In einem Schritt des Re-Design schlägt Kralisch alternative Formulierungen vor.
Die umfangreiche qualitative Inhaltsanalyse der Daten hat einige Anzeichen auf kognitive Konflikte bei den Lernenden aufgedeckt. Es kann vermutet werden, dass die Schüler:innen sich auf dem Weg zu einem Konzeptwechsel befanden. Allerdings ist weitere Forschung nötig, um dies zu belegen.
Die Lernziele 1 und 2 werden vollständig erreicht, Anknüpfungspunkte für Teile des Lernziels 3 sind vorhanden. Lernziel 4 konnte nicht in voller Ausprägung erreicht werden.

9.1.3 Fazit und Ausblick

In gemeinschaftlicher Zusammenarbeit mit Florian Jungkamp wurden die wichtigsten Aspekte der Falsifizierbarkeit herausgearbeitet, Arbeitsblätter erstellt, Problemstellungen formuliert und der Stundenverlauf zur Erarbeitung von Falsifizierbarkeit im Chemieunterricht der neunten Klasse geplant.
Die Erprobung dieser Materialien mit einer Lerngruppe hat es ermöglicht, Aussagen über die Eignung für den realen Unterrichtseinsatz zu treffen. Die durchgehend hohe Diskussionsbereitschaft der Schülerinnen und Schüler sowie große Anteile korrekt gelöster Arbeitsaufträge zeigen, dass die Materialien von den Lernenden gut angenommen und auch verstanden werden. Die aufgeworfene Problemstellung auf Ar-

beitsblatt 2 scheint einen direkten Appell an die Intuition der Lernenden zu richten - denn die Aussage von Bernd wird nahezu einstimmig als „unmöglich zu widerlegen“ erkannt.
Nachdem die Lernenden das Konzept der Falsifizierbarkeit theoretisch erarbeitet hatten, konnten sie dieses auf Aussagen anwenden, die dem Stil sozialer Medien nachempfunden waren. Alle vier Lernziele wurden in der Unterrichtseinheit zumindest touchiert, zwei davon konnten vollständig erreicht werden.
Hinweise auf *Kognitive Konflikte* wurden im Verhalten und den schriftlichen Antworten der Schüler:innen gefunden. Das Lernmaterial lässt somit Konzeptwechsel notwendig erscheinen; ob dieser bei allen Schülerinnen und Schüler vollständig gelingt, konnte den erhobenen Daten allerdings nicht entnommen werden. Ein vollständiger Konzeptwechsel ist aufgrund der Hartnäckigkeit von Schülervorstellungen (Marohn & Egbers 2011) sogar eher unwahrscheinlich.
Es wurde an einigen Stellen erkannt, dass Korrekturen und Umformulierungen nötig sind, um die Verständlichkeit für die Schülerinnen und Schüler weiter zu erhöhen. Auch die Prüfkarten sollten sinnvoll in den Unterricht eingebunden werden, dazu eignen sich kurze Übungsphasen im nachfolgenden Unterricht, in denen jeweils eine der Prüfkarten mit ihrer Aufgabenstellung thematisiert und ausgewertet wird.
Insgesamt kann die Unterrichtseinheit zur *Falsifizierbarkeit* als geeignet für den Einsatz in der Phase III von choice^{2}reflect angesehen werden.

9.2 MZ 1.2: Korrelation vs. Kausalität

> Im Jahr 2019 wurde der Ausbau von 5G-Mobilfunknetzen gestartet. Alleine in Deutschland wurden etwa 510 neue Mobilfunkmasten bis Ende 2019 aufgestellt, die auf 5G-Frequenzen funken (Vodafone 2020; Habel 2019). Bis Ende 2020 waren in nur einem Netz bereits 3000 Masten aktiv (Krzossa 2020). Gleichzeitig mit dem Ausbau kam es im Frühjahr 2020 weltweit zu Infektionen mit dem Corona-Virus: Eine Pandemie entwickelte sich. Dadurch ist eindeutig bewiesen, dass 5G-Strahlung für Corona verantwortlich ist.

In der Debatte um den Ausbau der Mobilfunktechnologie 5G ist diese Argumentation immer wieder zu finden. Sie beruht auf zwei Prämissen, die zu einer Schlussfolgerung

kombiniert werden. Prämisse 1: *Im Jahr 2019 wurde in Deutschland 5G eingeführt und ausgebaut.* Prämisse 2: *Im Jahr 2020 brach der Corona-Virus aus.* Die daraus gezogene Schlussfolgerung lautet: *5G hat den Ausbruch der Corona-Pandemie ausgelöst.*

Die Schlussfolgerung ist aussagenlogisch gültig, enthält allerdings auf der inhaltlichen Ebene einen schwerwiegenden Fehlschluss. Zwischen dem Ausbau von 5G und dem Auftreten des Corona-Virus besteht eine zeitliche Korrelation, eine kausale Aussage darf hier jedoch nicht getroffen werden. Das Auftreten der Korrelation ist kein Beweis für ein Ursache-Wirkungs-Verhältnis. Da die oben genannte Aussage jedoch ein solches Verhältnis postuliert, ist sie nicht gültig und darf in einer wissenschaftlichen Diskussion nicht verwendet werden.

In diesem Kapitel wird geklärt, wie sich Korrelationen von Kausalitäten unterscheiden lassen und welche Möglichkeiten es gibt, diese Unterscheidung im naturwissenschaftlichen Unterricht zu thematisieren. Dafür werden Lernmaterialien für den Einsatz in der Unterrichtskonzeption choice^{2}reflect entwickelt, erprobt und reflektiert.

9.2.1 Theoretischer Hintergrund zu Korrelationen und Kausalitäten

Das Eingangsbeispiel enthält den Fehlschluss *Post hoc, ergo propter hoc* (kurz: *Post hoc*). Übersetzt in die deutsche Sprache bedeutet das: „Danach, also deswegen“ bzw. „Es passierte danach, also passierte es deswegen“. Es handelt sich dabei um einen Denkfehler, bei dem Korrelationen als Kausalitäten missverstanden werden. Wird ein Zusammenhang zwischen unabhängigen und abhängigen Variablen empirisch bestätigt, dann besteht eine Korrelation. Beide Faktoren treten miteinander oder nacheinander auf, sodass eine Wenn-Dann-Aussage oder Je-desto-Aussage möglich ist: „Je mehr 5G Mobilfunkmasten in einer Region stehen, desto mehr Corona-Infizierte gibt es in dieser Region“. Der Fehlschluss wurde schon von Aristoteles (2018) in Kapitel 24 des zweiten Buches von „Die Rhetorik“ beschrieben:

> Ein weiterer Topos des Trugschlusses besteht darin, etwas nicht Ursächliches als Ursache hinzustellen, z. B. wenn etwas gleichzeitig oder nach etwas anderem geschehen ist; denn die Leute sehen das *danach* oft für *deswegen* an, besonders im politischen Leben [...]

Seitdem wurde in zahlreichen Veröffentlichungen versucht, die genauen Hintergründe des Fehlschlusses *Post hoc, ergo propter hoc* zu erklären. Beispielsweise wurde von Woods & D. Walton (1977) ein Schema mit fünf Prämissen aufgestellt, die für eine Kausalität notwendig sind.
Ein Faktor A sei nur kausal für den Faktor B, wenn alle dieser fünf Prämissen erfüllt sind:

1. Es besteht eine positive Korrelation zwischen A und B
2. B ist nicht die Ursache von A
3. Es gibt keinen dritten Faktor C, der als Ursache für A oder B angesehen werden kann, außerdem ist A nicht die Ursache von C
4. A tritt nicht ohne B auf
5. B besitzt eine pragmatische Relevanz.

In späteren Veröffentlichungen bauen Woods & D. Walton (1977) allerdings nicht weiter auf diesem Schema auf und geben zu, dass eine Analyse des Fehlschlusses aufgrund fehlender Grundsätze zur Analyse des Konzepts Kausalität schwierig ist (Woods & D. N. Walton 1982). Eine Erklärung, was genau am „post hoc reasoning" falsch ist, können sie nicht liefern.
Der Kern des Konzepts bleibt jedoch bestehen: Dass ein Faktor auf den anderen folgt, ist kein Beweis für ein Ursache-Wirkungs-Verhältnis. Wird ein solches dennoch fälschlicherweise angenommen, dann liegt der oben beschriebene Fehlschluss vor. Für diese Art von Denkfehler haben sich weitere Bezeichnungen durchgesetzt. So ist auch von Kausalitätsirrtum, Kausalillusion oder Scheinkorrelation die Rede. Diese Bezeichnungen beleuchten den Fehlschluss jeweils aus einer leicht anderen Perspektive, können ansonsten jedoch weitgehend synonym verwendet werden.
Abzugrenzen ist das Phänomen *Post hoc, ergo propter hoc* jedoch vom eng verwandten *Cum hoc, ergo propter hoc*. Während *Post hoc* die zeitliche Abfolge der korrelierenden Faktoren betont, bezeichnet *Cum hoc* auch zusammen auftretende Korrelationen ohne zeitlichen Versatz. Beide Phänomene werden zur übergeordneten Gruppe der *Non causa pro causa*-Fehlschlüsse, auch als *false cause-fallacies* bezeichnet, zusammengefasst (Arp et al. 2018, S. 335).
Im weiteren Verlauf der Arbeit werden Fehlschlüsse, die sich auf das Verhältnis von Korrelationen zu Kausalitäten beziehen, als Scheinkausalitäten bezeichnet. Im Zuge der didaktischen Aufarbeitung für den Einsatz im naturwissenschaftlichen Unter-

richt wird dieser Begriff dem Wort Scheinkorrelationen vorgezogen, da er noch etwas intuitiver verständlich ist.

9.2.2 Entwicklung des Lernmaterials

Der vorherige Abschnitt hat eine Klasse von Fehlschlüssen dargestellt, die verschiedene Denkfehler zur Überinterpretation von Korrelationen zu Kausalitäten enthält. Eine Möglichkeit, das Wissen um derartige Fehlschlüsse für den naturwissenschaftlichen Unterricht nutzbar zu machen, wird im folgenden Abschnitt vorgestellt. Dabei wird das Prüfen auf Scheinkausalitäten als eigenständiges Wissenschaftskriterium in der Unterrichtskonzeption choice^{2}reflect angesehen.

Festlegen der Lernziele

Die Thematik zu Scheinkausalitäten, verbunden mit philosophisch-logischen Überlegungen zu den Verhältnissen verschiedener Prämissen und Konklusionen zueinander, ist ein komplexes Feld und innerhalb einer kurzen Unterrichtseinheit nicht komplett zu erschließen. Daher ist es notwendig, gezielte Elementarisierungen der Inhalte vorzunehmen. Die zu erlernenden Inhalte werden in Form von Lernzielen ausgedrückt:

1. Die Schüler:innen filtern Rohdaten einer wissenschaftlichen Studie zu einem lebensnahen Kontext nach verschiedenen untersuchten Faktoren und werten sie aus
2. Die Schüler:innen ziehen eigene Schlussfolgerungen aus den analysierten Daten
3. Die Schüler:innen bewerten Schlussfolgerungen, die von Dritten anhand der Studien-Daten getroffen wurden
4. Die Schüler:innen erarbeiten das Konzept der Scheinkausalität anhand von Prüfkarten

Nicht im Unterricht thematisiert wird die Unterscheidung zwischen *Cum hoc* und *Post hoc* sowie die Tatsache, dass Scheinkausalitäten nur einen Teil der übergeordneten Gruppe von *Non causa pro causa*-Fehlschlüssen bilden.

Einführung in das Prüfkriterium

Zur Einführung in die Thematik der Scheinkausalität lesen die Schüler:innen einen kurzen Text, der eine Studie zur Smartphone-Nutzung im Zusammenhang mit der Anwendung Instagram vorstellt. Die Einzelheiten zu dieser Studie wurden konstruiert und basieren nicht auf echten Forschungsdaten. Der Studie zugrunde liegt die Hypothese, dass Smartphones, auf denen die Instagram-App installiert ist, eine kürzere Akkulaufzeit haben als Smartphones ohne diese App. Der einleitende Text informiert, dass 2000 Personen zu ihrer Smartphone-Nutzung befragt wurden und nennt das Ergebnis der Studie: Die Akkuladung von Smartphones mit installierter Instagram-App sank rund 6 Stunden schneller auf unter 20 % als von Smartphones, auf denen die App nicht installiert ist. Die Schlussfolgerung der Forschenden sei die Aussage: „Die Instagram-App sorgt für eine kurze Akkulaufzeit und ist daher als Stromfresser einzustufen".
Die erste Aufgabe auf dem Arbeitsblatt, welches auch in Abbildung 9.7 abgebildet ist, lautet:

> Aufgabe Basis: Wechsle zu den Daten der Studie und führe die Datenauswertung durch. Wie das geht, steht auf dem ausgeteilten Arbeitsblatt.

Nach der Datenauswertung, die im nächsten Kapitel beschrieben wird, wechseln die Lernenden zurück zu dem ersten Arbeitsblatt und bearbeiten die nächste Aufgabe. Dabei bewerten sie die oben genannte Schlussfolgerung, die die Forschenden aus den Studien-Daten generiert haben.

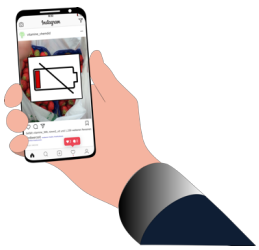
Analyse der Rohdaten

Den Lernenden wird auf einem Tablet eine Datei zur Verfügung gestellt, die mit der Anwendung „Numbers" (verfügbar für die Betriebssysteme iOS/iPadOS sowie MacOS) geöffnet wird. Dabei handelt es sich um eine Tabelle mit einem Ausschnitt aus den Rohdaten der Studie (siehe Abbildung 9.8). Aufgelistet werden die Ergebnisse der Befragung von 40 Personen. Dabei werden neben dem Alter der Person die Antworten der Befragten in den Spalten *Smartphone vorhanden?*, *Alter des Smartphones*, *Handynutzung pro Tag (Stunden)*, *Instagram installiert?* und *Akkulaufzeit (Stunden)* angegeben. Der Datensatz wird durch Symbole ergänzt, die die Antworten der Befragten visualisieren.

WWU MÜNSTER

Chemie Didaktik Münster

Wie hängt Instagram mit der Akkulaufzeit zusammen?



Forscherinnen und Forscher haben in einer Studie den Zusammenhang zwischen der Instagram-App und der Akkulaufzeit von Smartphones untersucht. **2000 Personen** zwischen 16 und 62 Jahren wurden zu ihrer Smartphone-Nutzung befragt. Dabei wurde herausgefunden, dass die Akkuladung von Smartphones mit installierter Instagram-App rund 6 Stunden schneller auf unter 20 % sank als von Smartphones, auf denen die App nicht installiert ist. Die Forschergruppe stufte die Instagram-App aufgrund dieser Ergebnisse als „Stromfresser“ ein.

Aufgabe Basis:
Wechsle zu den **Daten der Studie** und **führe die Datenauswertung durch.** Wie das geht, steht auf dem ausgeteilten Arbeitsblatt.

Aufgabe ★1:
Die Forschergruppe kommt nach dieser Analyse zu folgendem Ergebnis:

„Die Instagram-App sorgt für eine kurze Akkulaufzeit und ist daher als Stromfresser einzustufen“.

Bewerte dieses Ergebnis gemeinsam mit deinem Partner.

© C. Kralisch, Prof. Dr. A. Marohn: choice²reflect; Institut für Didaktik der Chemie, Münster

Abbildung 9.7: *Arbeitsblatt 1 zur Erarbeitung des Prüfkriteriums Korrelation und Kausalität.*

Untersuchung der Smartphone-Nutzung

Person	Alter der Person (Jahre)	Smartphone vorhanden?	Alter des Smartphones	Handynutzung pro Tag (Stunden)	Instagram installiert?	Akkulaufzeit (Stunden)
1	38	Ja	31-36 Monate	4-5 Stunden		6-12 Stunden
2	51	Ja	13-18 Monate	4-5 Stunden		13-19 Stunden
3	25	Ja	37-42 Monate	über 5 Stunden		<6 Stunden
4	55	Ja	19-24 Monate	1-2 Stunden		25-30 Stunden
5	19	Nein				
6	46	Ja	25-30 Monate	4-5 Stunden		6-12 Stunden
7	20	Ja	37-42 Monate	über 5 Stunden		6-12 Stunden
8	31	Ja	37-42 Monate	3-4 Stunden		6-12 Stunden

Abbildung 9.8: *Ausschnitt aus den Rohdaten der Befragung. Dargestellt werden die konstruierten Daten von acht Versuchspersonen.*

Die Lernenden erhalten ein weiteres Arbeitsblatt, das genaue Anweisungen enthält, wie die Daten analysiert werden sollen. Dabei führt jede:r Schüler:in die Analyse an einer eigenen Kopie der Datei durch, kann sich aber während der Arbeitsphase mit der Partnerin oder dem Partner besprechen.

Durch gezieltes Setzen von sogenannten Schnellfiltern in der App „Numbers“ werden die Daten übersichtlicher und nach den gewünschten Faktoren gefiltert. Die Schüler:innen filtern dabei zunächst nach Personen, die in der Spalte *Smartphone vorhanden?* „Ja“ eingetragen haben und blenden damit alle Datensätze von Personen ohne Smartphone aus. Anschließend wird nach der Akkulaufzeit und installierter Instagram-App gefiltert: Die Lernenden lassen sich nur Daten von Personen anzeigen, die ein Smartphone mit installierter Instagram-App und einer Akkulaufzeit von über 20 Stunden haben. Diese Personen werden ausgezählt und das Ergebnis wird in einer Tabelle eingetragen.

Die restliche Analyse führen die Lernenden eigenständig durch: Sie filtern nach den Faktoren *Handynutzung* und *Alter des Smartphones* und setzen jeweils die Akkulaufzeit in Beziehung zu den Faktoren. In Abbildung 9.9 ist die Tabelle mit bereits eingetragenen Ergebnissen dargestellt.

		Akkulaufzeit	
		kurz (unter 20 Stunden)	lang (über 20 Stunden)
Instagram App	installiert	12	4
	nicht installiert	5	9
Handynutzung pro Tag	unter 3 Stunden	4	9
	über 3 Stunden	13	4
☆ 1 Alter des Smartphones	unter 24 Monate (alt / neu)	6	10
	über 24 Monate (alt / neu)	11	3

Abbildung 9.9: *Ergebnistabelle der Analyse von Rohdaten durch Setzen von Schnellfiltern in der App „Numbers“ mit bereits eingetragenen Ergebnissen.*

Aus der Analyse lassen sich folgende Korrelationen ableiten:

- Smartphones, bei denen die Instagram-App installiert ist, haben eine kürzere Akkulaufzeit (AL) als Smartphones, bei denen die App nicht installiert ist.
- Smartphones, die mehr als drei Stunden am Tag genutzt werden, haben eine kürzere AL als solche, die weniger als drei Stunden am Tag genutzt werden.
- Smartphones, die über 24 Monate alt sind, haben eine kürzere AL als neuere Smartphones.

Diese Aussagen fassen die Befragungsergebnisse der 40 untersuchten Personen zusammen. Dabei beschreiben sie die Tendenzen, die innerhalb der untersuchten Stichprobe gefunden wurden. An dieser Stelle könnte eine Diskussion über die Repräsentativität der Stichprobe und die Generalisierbarkeit der Ergebnisse geführt werden, wie es die Schüler:innen bei der Erarbeitung des Prüfkriteriums Randomisierung erlernen (siehe Kapitel 9.3). Der Fokus soll in dieser Einheit jedoch auf der Schlussfolgerung liegen, die die Autor:innen der Studie auf Grundlage der Ergebnisse gezogen haben.

Bewertung der Schlussfolgerung

Die Schlussfolgerung: „Die Instagram-App sorgt für eine kurze Akkulaufzeit und ist daher als Stromfresser einzustufen“ bezieht sich nur auf den dichotomen Faktor *Vorhandensein der Instagram-App* und dessen Auswirkungen auf die Akkulaufzeit. Die Beobachtung, dass Smartphones mit Instagram-App eine kürzere Akkulaufzeit haben, wird dabei als kausaler Zusammenhang angesehen. Die verkürzte Laufzeit wird nur dem Vorhandensein der App zugeschrieben, ohne weitere Faktoren in den Blick zu nehmen.

Es handelt sich bei dem Zusammenhang um eine Scheinkausalität, denn es gibt weitere Faktoren, die die Akkulaufzeit beeinflussen können. Dazu gehören die in der Studie untersuchten Faktoren *Handynutzung pro Tag* und *Alter des Smartphones*, aber auch weitere Einflussfaktoren sind denkbar. Eine Kausalität darf daher hier nicht angenommen werden.

Die Schlussfolgerung zu bewerten und dabei den Fehlschluss zu erkennen, ist nach erfolgter Datenanalyse die Aufgabe der Lernenden und sollte in Partnerarbeit durchgeführt werden. In einem ersten Schritt wird die Bewertung intuitiv durchgeführt, ohne dass das Prüfkriterium *Korrelation vs. Kausalität* zuvor thematisiert wurde. Die Lernenden kennen somit den Begriff der Scheinkausalität noch nicht. Mithilfe der Aufgabe zur Bewertung der Schlussfolgerung soll diagnostiziert werden, inwiefern die Schüler:innen Scheinkausalitäten auf intuitive Weise erkennen können.

In einer Plenumsphase werden die Ergebnisse der Datenanalyse gesammelt. Dabei wird die Ergebnistabelle gemeinsam ausgefüllt, anschließend wird die Bewertung der Schlussfolgerung kurz diskutiert.

Erarbeitung des Prüfkriteriums *Korrelation vs. Kausalität*

Nach erfolgter Bewertung der Schlussfolgerung erhalten die Schüler:innen einen Satz an Prüfkarten (siehe Anhang B.6.5), bestehend aus vier einzelnen Prüfkarten. Diese werden zunächst in Einzelarbeit gelesen. Die erste Prüfkarte mit der Überschrift „Bringt der Storch die Kinder?“ dient zur Einführung der Begriffe *Korrelation* und *Kausalität*. Das in der Literatur häufig angeführte Beispiel, bei dem das gehäufte Vorkommen von Störchen mit einer erhöhten Geburtenrate in eine kausale Beziehung gesetzt wird, dient der Veranschaulichung der Begriffe.

Die Inhalte der Prüfkarte setzen den Begriff *Korrelation* mit dem Begriff *Zusammenhang* gleich und *Kausalität* mit *Ursache*. Eine schematische Grafik stellt dar, dass Korrelationen als „UND-Aussage“ formuliert werden können, während Kausalitäten als „WEIL-Aussage“ zu verstehen sind.

Die nächste Prüfkarte „Korrelation und Kausalität“ verdeutlicht das Konzept anhand von weiteren schematischen Darstellungen. Zusätzlich wird der Begriff *Scheinkausalität* eingeführt. Dabei wird der Kontext „Bringt der Storch die Kinder?“ auch auf dieser Prüfkarte zur Veranschaulichung verwendet.

Die dritte Prüfkarte thematisiert Korrelation und Kausalität in der Forschung. Es wird ein Rückbezug auf die zuvor behandelte Studie zur Smartphone-Nutzung hergestellt. Die einzelnen untersuchten Faktoren und daraus abgeleitete Korrelationen werden aufgelistet und die Schlussfolgerung der Forscher:innen erneut dargestellt.

Schließlich sollen die Lernenden auf der letzten Prüfkarte erklären, was die Forschenden bei ihrer Schlussfolgerung falsch gemacht haben. Dazu diskutieren die Schüler:innen mit einer Partnerin oder einem Partner und notieren anschließend ihre Lösung.

Nach Abschluss der Arbeitsphase werden die Ergebnisse im Plenum diskutiert. Die Lehrperson moderiert dabei die Wortbeiträge der Lernenden, gibt einzelne Impulse zur Diskussion und notiert die zentralen Ergebnisse auf dem projizierten Arbeitsblatt.

Den Abschluss der Unterrichtsstunde bildet die kurze Abfrage: „Was haben wir heute gelernt?“ Diese wird durch die Lehrkraft gestellt und von den Schüler:innen in kurzen Wortbeiträgen beantwortet.

9.2.3 Erprobung

Eine Erprobung der entwickelten Materialien zur Erarbeitung des Prüfkriteriums *Korrelation vs. Kausalität* wurde im Rahmen einer Gesamterprobung im September 2020 durchgeführt. Die Erprobung umfasste drei Doppelstunden und eine Einzelstunde im naturwissenschaftlichen Unterricht einer 9. Klasse, wobei eine Doppelstunde der Erarbeitung des Prüfkriteriums *Korrelation vs. Kausalität* gewidmet wurde. Es nahmen 18 Lernende an der Stunde teil. Eine genaue Beschreibung der Rahmenbedingungen dieser Erprobung befindet sich im Kapitel 8.1.5, der Stundenverlaufsplan ist im Anhang B.7.2 dargestellt.

Beschreibung der Unterrichtsstunde

Die durchgeführte Unterrichtsstunde mit einer Dauer von 90 Minuten begann mit einer leichten Verspätung, da um 07:55 Uhr noch nicht alle Schüler:innen ihre Sitzplätze eingenommen hatten. Technische Aspekte stellten sich zu Anfang der Stunde als zeitaufwendig heraus, sodass sowohl Lernende als auch Lehrende mehr Zeit als geplant mit der Verbindung der Tablets und Aktivierung der Kameras verbrachten.
Die Lehrperson erklärte die Wiederholungsaufgabe, woraufhin die Lernenden diese in Partnerarbeit durchführten. Während der Arbeitsphase konnten in mehreren Bereichen des Raumes Gespräche über die Fachbegriffe der letzten Stunde vernommen werden. Der Übergang zur ersten Plenumsphase erfolgte nach etwa 21 Minuten und somit etwa zehn Minuten später als geplant.
In der Plenumsphase erklärten einzelne Schüler:innen in Wortmeldungen die Begriffe *Randomisierung* und *repräsentativ* (siehe Kapitel 9.3). Diese Phase fiel mit etwa fünf Minuten halb so lang aus wie zuvor antizipiert (zehn Minuten).
Nach Abschluss der Wiederholungsphase erläuterte die Lehrkraft anhand von zwei Postern den weiteren Verlauf der Unterrichtsreihe. Zunächst wurde auf einem an die Tafel gehefteten Poster die Segmentierung der Reihe in drei Unterrichtseinheiten visualisiert. Der Andeutung, dass in der aktuellen Stunde eine Instagram-Studie zum Kriterium *Korrelation vs. Kausalität* behandelt werde, folgte ein Wechsel des gezeigten Posters.
Nach der Visualisierung der Unterrichtsreihe zeigte die Lehrperson das Prüfschema zur Bewertung empirischer Studien (siehe Kapitel 11) und erklärte in groben Zügen den Aufbau des Schemas. Die Präsentation beider Poster dauerte insgesamt

02:30 min und liegt damit im Planungsrahmen (03:00 min).

Mit einer Dauer von etwa drei Minuten unterschritt die Unterrichtsphase zur Einleitung des Arbeitsauftrages von Arbeitsblatt 1 (*Wie hängt Instagram mit der Akkulaufzeit zusammen?*) die geplante Zeit von sieben Minuten. In der Phase las eine Schülerin den Einleitungstext vor und die Lehrperson gab Hinweise zur Bearbeitung der Aufgabe.

Anschließend wurde eine Arbeitsphase eingeleitet, die von den Schüler:innen zur Analyse der Tabellendaten in der App *Numbers* genutzt wurde. Ein elementarer Bestandteil der Aufgabe ist die Erstellung von „Schnellfiltern" in der Tabelle. Im Laufe der Arbeitsphase stellte die Lehrkraft fest, dass die Lernenden Schwierigkeiten bei der Bedienung der App und dem Setzen der Schnellfilter hatten. Daraufhin wurde eine von der Planung abweichende Plenumsphase eingeschoben, in der die Lehrkraft an einem Beispiel die Verwendung von Schnellfiltern demonstrierte.

Die Arbeitsphase dauerte statt zuvor antizipierter 20 Minuten somit etwa 27 Minuten.

Nachdem die Lernenden die Arbeit in der App *Numbers* abgeschlossen hatten, wurden ihre Antworten im Plenum zusammengetragen. Die Lehrkraft unterstützte durch gezielte Impulse. Es wurde innerhalb von etwa 14 Minuten geklärt (geplant 10 Minuten), welche Aussagekraft die Daten der Tabelle besitzen.

Die Lehrperson leitete zur Erarbeitung der Prüfkarten (Arbeitsblatt 3) über. Die Lernenden erhielten 15 Minuten Bearbeitungszeit, die Lehrkraft beendete die Arbeitsphase jedoch nach 12 Minuten, da alle Lernenden ihre Arbeit abgeschlossen hatten.

Zum Ende der Stunde wurden die Lösungen zu den Aufgaben der Prüfkarten im Plenum zusammengetragen. Dabei wurde herausgestellt, dass es sich bei Korrelationen um Zusammenhänge handelt, während Kausalitäten Ursachen beschreiben. Bezugnehmend auf die von den Lernenden bearbeiteten Studiendaten wurde auch herausgearbeitet, dass die Faktoren Akku-Laufzeit und Instagram-App zwar statistisch zusammenhängen, dass jedoch keiner dieser Faktoren die Ursache des anderen ist.

Die letzte Aktion der Unterrichtsstunde bestand in der Einordnung der Instagram-Studie in das Prüfschema. Aufgrund der fehlenden Betrachtung weiterer Einflussfaktoren wurde die Studie als unwissenschaftlich deklariert und auf dem Prüfschema in die *Mülltonne* geleitet.

Auswertung der bearbeiteten Arbeitsblätter

Im Anschluss an die Unterrichtsstunde wurden die digital vorliegenden Arbeitsergebnisse der Schüler:innen an einem zentralen Ort gespeichert. Im folgenden Abschnitt werden die Arbeitsblätter ausgewertet.
Ein Kategoriensystem wurde entwickelt, um die eingesammelten Arbeitsblätter systematisch auswerten zu können. Die Grobstruktur des Kategoriensystems wurde deduktiv vordefiniert und an der Struktur der Arbeitsblätter ausgerichtet: Jedes Arbeitsblatt wird durch eine Oberkategorie erfasst, die jeweiligen Arbeitsaufträge bilden die erste Ebene der Subkategorien. Die Kategorien der nächsten Ebene werden im Falle einer geschlossenen Aufgabe (z. B. Ja/Nein-Frage) ebenfalls deduktiv festgelegt, bei offenen Freitext-Aufgaben werden die Kategorien jedoch induktiv am Material bestimmt. Das auf diese Weise entstandene Kategoriensystem ist im Anhang B.1.2 dargestellt.

Arbeitsblatt 1 enthält ein Freitextfeld, in dem die Lernenden die vorgegebene Aussage der Studie bewerten sollen. Die meisten Schüler:innen haben hier die Plenumsantwort notiert, also die Antwort, die im Laufe der Unterrichtsstunde gemeinsam im Plenum erarbeitet wurde. Lediglich eine Partnergruppe notierte einen in eigenen Worten formulierten Stichpunkt: „Instagram zieht viel Akku, aber geringe Akkulaufzeit liegt auch an hoher Nutzung oder einem älteren Smartphone“.
Das zweite Arbeitsblatt bearbeiteten die Schüler:innen in der Tabellenkalkulations-App *Numbers*. Die Ergebnisse dieser Arbeit notierten die Lernenden nicht in der Datei selbst, sondern auf einem beiliegenden Arbeitsblatt in Form einer Ergebnistabelle. Acht von Dreizehn Lernenden trugen hier alle Werte korrekt ein. Vier Lernende notierten einen falschen Wert in der Tabelle und ein:e Schüler:in notierte mehr als vier falsche Werte.
Arbeitsblatt 3, bestehend aus den Prüfkarten zum Prüfkriterium *Korrelation und Kausalität*, enthält eine offene Aufgabe. Die Schüler:innen diskutieren, welche Fehler die Forscher:innen bei der Formulierung ihrer Schlussfolgerung begangen haben. In 9 von 15 Fällen notierten die Lernenden die im Plenum erarbeiteten Stichworte als Antwort auf die Frage. Abseits von der Plenumsantwort benannten fünf Lernende den Faktor *Handynutzung* als fehlenden Faktor bei der Schlussfolgerung, drei Lernende nannten das *Alter des Smartphones* und drei Lernende *andere Apps* als weitere Einflussfaktoren. Zwei Lernende erwähnten, dass ein *dritter Faktor* beachtet werden

müsse.

Die verschriftlichten Antworten der Lernenden geben in diesem Fall keine gesicherte Auskunft darüber, welche Aspekte innerhalb der Gruppenarbeitsphasen wirklich erarbeitet wurden. Die Schüler:innen hatten auch während der Plenumsphase einen uneingeschränkten Zugriff auf die Arbeitsblätter und konnten somit ihre eigenen Arbeitsergebnisse durch die im Plenum besprochenen ergänzen oder gar die eigenen Ergebnisse entfernen.
Es lässt sich aufgrund der eigenen Wortwahl vieler Schülerantworten vermuten, dass diese Ergebnisse aus eigenen Überlegungen entstanden sind. Belegt werden kann diese Vermutung allerdings nicht.
Im folgenden Abschnitt werden die aufgenommenen Unterrichtsvideos analysiert und im Hinblick auf die Forschungsfrage ausgewertet.

Untersuchung von Lehrimpulsen in Plenumsphasen der Unterrichtsstunde

Die Plenumsphasen der Unterrichtsstunde dienen der Sammlung und Sicherung der Arbeitsergebnisse. Die Lehrkraft bittet nach jeder Arbeitsphase um Beendigung der Gruppengespräche und initiiert eine Diskussion im Plenum. Entscheidend für den Verlauf der Diskussionen sind gezielte Moderationen in Form von Impulsen durch die Lehrkraft.
Als *Impuls* werden im weiteren Verlauf alle verbalen Äußerungen der Lehrkraft angesehen, die in einer Plenumsphase der Interaktion mit Schüler:innen dienen. Dabei kann es sich um offene Fragen an das Plenum oder auch Operatoren-gestützte Arbeitsaufträge handeln.

Darstellung der Lehrimpulse

Im folgenden Abschnitt werden die einzelnen Impulse der Lehrkraft in den Plenumsphasen näher beleuchtet. Dabei wird die Art des Impulses bestimmt, antizipierte Schülerantworten vorgestellt und mit den tatsächlich im Unterricht aufgetretenen Äußerungen der Schüler:innen verglichen. Dadurch wird festgestellt, inwiefern die Reaktionen der Schüler:innen auf Lehreräußerungen in der Planung passend antizipiert wurden.
Weiterhin wird dieser Vergleich herangezogen, um erste Aussagen über das Verständnis der Aufgabenstellungen zu treffen.

Die Tabelle B.2.1 im Anhang zeigt eine Übersicht über alle in der Doppelstunde gegebenen Impulse der Lehrkraft, dazu antizipierte Schülerantworten sowie die tatsächlich aufgetretenen Äußerungen.

Impuls 28g

Die Lehrkraft erkundigt sich nach den Arbeitsergebnissen der Lernenden mit der Erwartung, dass jeder Tabellenwert von den Schüler:innen kurz genannt und beschrieben wird. Die Frage: „Was habt ihr da für Werte herausbekommen?" hat Aufforderungscharakter.

In der Situation meldeten sich einige Lernende zu Wort und lasen die von ihnen notierte Lösung vor. Dabei bestanden die Wortbeiträge jedoch nur aus wenigen Worten. Weiterhin formulierten die Lernenden im Plenum keine elaborierten Erklärungen. Stattdessen wurde auf einer oberflächlichen Ebene die Position der einzelnen Antworten genannt: „Unter die Vier kommt eine Neun". Diese Schüleraussage diente der Navigation auf dem Arbeitsblatt, verdeutlicht aber nicht die Bedeutung des Zahlenwerts an der entsprechenden Stelle der Tabelle.

Mithilfe des *Impuls 28h* versuchte die Lehrkraft, der Tendenz zu unreflektierten Ein-Wort-Antworten zu begegnen: „[...] einmal zusammen schauen was diese Werte jetzt überhaupt aussagen. [...] Was könnte uns das sagen?"

Anhand von dieser Frage sollten die Lernenden in ihrem strukturierten Denken unterstützt und zu deutenden Antworten ermutigt werden. Die Daten der zuvor untersuchten Studie lassen unterschiedliche Schlussfolgerungen zu, diese sollten von den Schüler:innen benannt werden.

Ein Schüler diskutierte daraufhin die Hauptaussage der Studie: „[…] Also Instagram zieht viel Akku, ja. Aber es hängt oft auch damit zusammen, wie lange man das nutzt oder wie alt das Handy ist" (S1-02 in Transkript 2020-09-28_Korrelation; Pos. 535). Weitere Wortbeiträge der Lernenden folgten jedoch nicht.

Impuls 28i ist eine offen gehaltene Frage mit Aufforderungscharakter. Die Lernenden sollten die im Arbeitsmaterial genannte Schlussfolgerung der Forschenden kritisch bewerten. Die Frage: „Was sagt ihr dazu? Kann man das so machen?" wurde von zwei Lernenden verneint. Als Begründung führten sie an, dass das Alter der Smartphones und andere Apps als Stromverbraucher nicht beachtet wurden.

Im Anschluss an die zweite Erarbeitungsphase, in der die Lernenden sich mit den Prüfkarten auseinandergesetzt hatten, initiierte die Lehrkraft eine Plenumsphase durch den *Impuls 28k*. Die Schüler:innen wurden aufgefordert, die Begriffe *Korrelation* und

Kausalität zu definieren. Ein Schüler bezeichnete Korrelationen als „[...] Zusammenhänge, also wenn etwas miteinander zusammenhängt“ (S2-16; Pos. 627). Ein:e andere:r Lernende:r bezeichnete Kausalität als „die Ursache“ (S1-12; Pos. 637).
Daraufhin folgte eine Nachfrage der Lehrkraft: „Genau, woran erkennt man eine Kausalität oder eine Korrelation?“ (*Impuls 28l*). Ein:e Lernende:r antwortete: „Bei Korrelation steht *und* und bei Kausalität steht *weil*“.
Der offene Bewertungsimpuls 28i wurde in dieser Plenumsphase erneut aufgegriffen. Auf Grundlage des durch die Prüfkarten erarbeiteten Wissens sollte eine erneute Einschätzung vorgenommen werden, ob bei der Instagram-Studie das Kriterium *Korrelation vs. Kausalität* eingehalten wurde. Die Lehrkraft formulierte den *Impuls 28m*. Eine Schülerin las daraufhin ihre auf dem Arbeitsblatt notierte Antwort vor: „[...], sie haben nicht darauf geachtet, dass vielleicht eine andere App den Akku fressen könnte. Dadurch haben sie eine Scheinkausalität erzeugt, denn der dritte Faktor muss auch beachtet werden, so wie die ersten zwei.“
Der *Impuls 28n* bildete den Abschluss der Plenumsphase zum Arbeitsblatt 3. Dabei fragte die Lehrkraft, ob man dieser Studie nun trauen könne oder nicht. Die verhaltenen Reaktionen der Lernenden bewegten die Lehrperson zur Umformulierung, woraufhin ein:e Schüler:in antwortete: „Nein, weil die anderen Faktoren nicht beachtet wurden“.
Abschließend fragte die Lehrkraft ins Plenum, welche Inhalte die Lernenden „aus der Stunde mitnehmen konnten“ (*Impuls 28o*). Ein:e Schüler:in wiederholte die Definition von Korrelation und Kausalität, ein:e andere:r Schüler:in antwortete: „Dass Instagram bei alten Handys Strom fressen kann, also viel“; relativierte diese Aussage im Anschluss an eine Nachfrage der Lehrkraft (*Impuls 28p*) jedoch mit dem anschließenden Satz: „Das kann auch noch an anderen Apps liegen“.

Anschlussfähigkeit und Passung der Reaktionen auf Lehrimpulse

Die Plenumsphasen der Unterrichtsstunde sind durch zahlreiche Impulse der Lehrkraft gekennzeichnet. Nachdem diese Impulse im vorherigen Abschnitt charakterisiert worden sind, folgt nun ein Blick auf die von den Lernenden gegebenen Antworten und deren Passung zur Intention der Lehrkraft.
Zur Untersuchung der Schülerantworten erfolgt ein genauer Vergleich zwischen der intendierten Wirkung eines Impulses und der tatsächlich eingetretenen Situation. In einem hoch-inferenten Codierungs- bzw. Rating-Verfahren (siehe auch Kapitel 8.2)

werden den Schülerantworten Werte für die Passgenauigkeit zugeordnet. Weiterhin wird die Anschlussfähigkeit der Antwort für die weitere Entwicklung des Unterrichtsverlaufs bestimmt und die Anzahl an Nachfragen ermittelt, die die Lehrkraft bis zum Eintreten der analysierten Schülerantwort gestellt hat.

Die Skala für die Passung erstreckt sich von 1 (schlecht) über 2 (wenig passend) und 3 (eher passend) bis zu 4 (passgenau). Die Anschlussfähigkeit wird mit den Werten 1 (nicht anschlussfähig), 2 (wenig anschlussfähig), 3 (eher anschlussfähig) oder 4 (sehr anschlussfähig) codiert.

Die Tabelle im Anhang B.2.1 zeigt die Ergebnisse der Codierung.

Die Reaktion der Lernenden auf den Impuls 28g während der Zwischensicherung zum AB 1 wird mit einer Passung von 2 als wenig passend codiert. Lernende nannten zunächst lediglich die Tabellenwerte ohne Angabe von Begründungen.

Auch nach mehrfachem Nachfragen traten nur Reaktionen auf, die zur Intention wenig passend und nicht anschlussfähig waren.

Zielführender wirkten die Impulse 28h und 28i zum Ende der Zwischensicherung, die die Lernenden zunächst zum Bilden einer eigenen Schlussfolgerung und anschließend zur Bewertung der in der Studie genannten Schlussfolgerung aufforderten. Lernende tätigten sehr anschlussfähige Aussagen, die den Übergang in die nächste Unterrichtsphase ermöglichten.

Die Sicherungsphase zum AB 3 enthält die Impulse 28k bis 28n. Jedem dieser Impulse folgte eine passgenaue Reaktion der Lernenden (Codierung der Passung mit dem Wert 4). Die Anschlussfähigkeit zu 28k und 28l, bei denen die Definition der Begriffe *Korrelation* und *Kausalität* gefordert waren, stellte sich als gering bis mittelmäßig heraus (Codierung mit dem Wert 2 bzw. 3).

Den weiterführenden Impulsen zur Bewertung der Instagram-Studie bzw. zur Formulierung von Schlussfolgerungen (28m und 28n) folgten passgenaue und anschlussfähige Reaktionen.

Im abschließenden Teil der Unterrichtsstunde reagierten die Lernenden gemischt auf die Abschlussimpulse 28o und 28p. Schüler:in 2-16 antwortete passgenau und anschlussfähig. Schüler:in 1-01 gab nach einem zusätzlichen Impuls 28p der Lehrkraft eine passgenaue Antwort.

Kategorisierung der Lehrimpulse

Die Betrachtung der Parameter Passung und Anschlussfähigkeit gibt Aufschluss darüber, inwiefern die im Unterricht durch die Lehrkraft gegebenen Impulse angemessen und dem Erreichen der Lernziele zuträglich waren.
Die vier möglichen Kombinationen der Parameter sind:

1. Hohe Passung mit hoher Anschlussfähigkeit
2. Hohe Passung ohne Anschlussfähigkeit
3. Niedrige Passung mit hoher Anschlussfähigkeit
4. Niedrige Passung ohne Anschlussfähigkeit

Kategorie 1 ist der gewünschte Effekt. Der Unterricht wird durch gezielte Impulse der Lehrkraft vorangebracht, auf die die Lernenden angemessen reagieren. Die Lehrer-Schüler-Interaktion wird in dem Fall als sinnstiftend und gelungen angesehen.
Geht eine hohe Passung mit geringer Anschlussfähigkeit einher (Kategorie 2), lässt sich darauf schließen, dass der Impuls von den Lernenden verstanden und entsprechend der Intention der Lehrkraft beantwortet wurde - allerdings eignet sich die Reaktion der Lernenden nicht, um den Unterrichtsverlauf sinnstiftend voranzubringen. Betroffene Impulse sollten in der Phase des Re-Design als Konsequenz dieser Bewertung mittels Umformulierung verbessert oder durch einen anschlussfähigen weiteren Impuls ergänzt werden.
Ähnlich verhält es sich mit den Kategorien 3 und 4. Eine niedrige Passung der Schülerantworten deutet darauf hin, dass der Impuls von den Lernenden nicht in einer Weise verstanden wurde, die von der Lehrkraft angedacht war. Folglich ist es unwahrscheinlich, dass das unmittelbar auf betroffene Impulse folgende Unterrichtsgeschehen sinnstiftend verläuft. Das Auftreten von Kategorie 3 wird als unwahrscheinlich angesehen, ist aber dennoch möglich: Wenn Schüler:innen einen neuen bis zu dem Zeitpunkt des Unterrichts nicht betrachteten Aspekt in die Diskussion einbringen, kann trotz geringer Passung zur Ausgangsfrage eine Anschlussfähigkeit für den weiteren Unterrichtsverlauf entstehen.
Folgende Impuls-Reaktionen wurden der Kategorie 1 zugeordnet: 28h-I, 28h-II, 28h-III, 28i-II, 28l-I, 28m-I, 28m-II, 28n-I und 28o-I. Das entspricht etwa zwei Dritteln aller Impuls-Reaktionen.
Eine hohe Passung mit niedriger Anschlussfähigkeit zeichnen die Impuls-Reaktionen 28i-I, 28k-I, 28k-II und 28p-I aus (ca. ein Viertel der Reaktionen).

Impuls-Reaktionen der Kategorie 3 traten in dieser Unterrichtsstunde nicht auf. Weder als passend noch als anschlussfähig (Kategorie 4) erwiesen sich die Reaktionen auf den ersten und den letzten Impuls der Stunde (28g-I und 28o-II).

Untersuchung von Gruppenarbeitsphasen: *Aha-Erlebnisse* und Verständnisschwierigkeiten

Die von Schüler:innen im Plenum getätigten Äußerungen haben Aufschluss darüber gegeben, inwiefern die von der Lehrkraft verwendeten Lehr- und Moderationsimpulse angemessen und sinnstiftend waren. Zur weiteren Analyse des Umgangs von Schüler:innen mit den Lernmaterialien werden im folgenden Abschnitt Schüleräußerungen betrachtet, die nicht innerhalb von Plenumsphasen, sondern unmittelbar während der Bearbeitung der Arbeitsblätter getätigt wurden.
Zur Analyse der Stillarbeitsphasen wird ein detaillierter Blick auf die Gespräche innerhalb der Kleingruppen geworfen. Diese können Aufschluss darüber geben, in welcher Weise die Schüler:innen mit dem Material umgehen. Verständnisschwierigkeiten können aufgedeckt und erste Hinweise auf die Entstehung kognitiver Konflikte gesammelt werden. Die Fokussierung dieser Analyse gelingt mithilfe von drei Schwerpunkten, die als Vorlage für die Entwicklung der thematischen Hauptkategorien dienen:

1. Anzeichen auf Aha-Erlebnisse in den Stillarbeitsphasen der Unterrichtsstunde
2. Anzeichen auf Verständnisschwierigkeiten in den Gesprächen der Lernenden
3. Erkennbare Off-Task-Handlungen der Lernenden

Die Analyse-Schwerpunkte werden angelehnt an die qualitative Inhaltsanalyse nach Kuckartz mithilfe eines Kategoriensystems, das am Videomaterial angesetzt wird, untersucht. Nähere Hinweise zur Forschungsmethodik dieser Arbeit im Hinblick auf die Datenauswertung befinden sich im Kapitel 8.2. Einige Kategorien des Kategoriensystem werden, bedingt durch die Schwerpunktsetzung, zunächst deduktiv vorformuliert. In einem zweiten Schritt erfolgt die induktive Ausdifferenzierung und Ergänzung durch weitere (Sub-)Kategorien.

Schwerpunkt 1 behandelt die Entstehung von *Aha-Erlebnissen* bei der Bearbeitung des Lernmaterials. Ein Aha-Erlebnis kann verstanden werden als durch plötzlich eintretende Erkenntnis ausgelöste Auflösung von Irritationen in Form einer Einsicht (Caspar & Berger 2007; Bühler 1907).
Das Verhalten der Schüler:innen wird demnach auf plötzliche Ausrufe und besondere Gesten untersucht, die auf eine Einsicht und damit ein *Aha-Erlebnis* hindeuten. Diese Ausrufe können ein erstes Anzeichen dafür sein, dass die Schüler:innen zuvor einen kognitiven Konflikt erlebt haben.
Beispielsweise könnte ein Schüler bisher davon ausgegangen sein, dass einzelne Beobachtungen (z. B. schlechte Akkulaufzeit seines Smartphones bei Nutzung von *Instagram*) für die Bildung einer Kausalaussage ausreichen. Wenn er durch das Lernmaterial bemerkt, dass weiterführende Daten für Kausalitäten notwendig sind, tätigt er womöglich einen erstaunten Ausruf – dieser kann als *Aha-Erlebnis* codiert werden und deutet auf die Auflösung eines kognitiven Konflikts hin.
Allerdings ist das Vorhandensein eines Aha-Erlebnisses keineswegs hinreichend für den Beleg eines Konflikts. Einige unter diese Kategorie fallende Äußerungen können sich auch auf das Text- oder Aufgabenverständnis beziehen und liegen keiner *Unzufriedenheit mit dem bestehenden Konzept* (siehe Kapitel 3.5) zugrunde.
Das Lernmaterial ist so ausgelegt, dass die Lernenden im Zuge der Erarbeitung des Prüfkriteriums *Korrelation vs. Kausalität Aha-Erlebnisse* erfahren sollen. Vor allem die Nähe der Problemsituation zur Lebenswelt der Schüler:innen soll einen intuitiven Zugang zur Problematik der Scheinkausalitäten ermöglichen: Schüler:innen sollen erkennen, dass einzelne isolierte Faktoren, wie das Vorhandensein einer Applikation auf einem Smartphone, nicht für die kausale Erklärung der Akkulaufzeit herangezogen werden dürfen. Stattdessen müssen weitere mögliche Einflussfaktoren in die Erklärung einbezogen werden.
Auf der Grundlage dieser Überlegungen zum Analyse-Schwerpunkt 1 wurde eine passende Kategorie deduktiv festgelegt, diese wurde mit dem Begriff *Aha-Erlebnis* benannt.
Der Schwerpunkt 2 zielt auf erkennbare Verständnisschwierigkeiten bei der Bearbeitung der Lernmaterialien ab. Es wird untersucht, welche verbalen und nonverbalen Aktionen der Lernenden darauf hinweisen, dass eine Schwierigkeit besteht, dem Unterrichtsgeschehen sinnstiftend zu folgen. Die deduktiv entwickelte Kategorie heißt *Verständnisschwierigkeiten*. In einem ersten Codierungsverfahren werden

Sub-Kategorien entwickelt, die einen differenzierten Blick auf die Verständnisschwierigkeiten erlauben.
Als dritter Aspekt dieser Untersuchung der Stillarbeitsphasen steht die Ermittlung von Off-Task-Interaktionen im Fokus. Dabei werden Situationen ermittelt, in denen Lernende offensichtlich nicht dem Unterrichtsgeschehen folgen. Die Off-Task-Interaktionen werden unterteilt in *Gespräche*, *Tablet-Interaktionen*, *Allgemeine Interaktionen* und *Verlassen des Sitzplatzes*. Dadurch wird ein differenzierterer Blick auf die Art der Off-Task-Handlung ermöglicht. Die Codierung erfolgt niedrig-inferent.

Ergebnisdarstellung nach Gruppe

In einem niedrig-inferenten Event-Sampling-Verfahren wurden die Videos der Fokusgruppen auf das Auftreten der zuvor definierten Kategorien untersucht. Der in Tabelle 9.1 dargestellte Auszug aus dem Codierleitfaden diente als Grundlage für die unabhängige Codierung des Materials durch zwei Personen. Der Leitfaden enthält die Kategorien *Aha-Erlebnisse*, *Verständnisschwierigkeiten* mit den induktiv ermittelten Unterkategorien *Technik*, *Fachinhalt* und *Aufgabenstellung* sowie *Off-Task* mit den oben genannten deduktiv bestimmten Unterkategorien.
Im Codierungs-Verfahren wurde nach diskursiver gemeinsamer Überprüfung von nicht-übereinstimmenden Codes eine Intercoder-Übereinstimmung von etwa 92 % erreicht. Von 447 vergebenen Codes sind 412 zwischen beiden Codierenden übereinstimmend und 35 nicht-übereinstimmend. Der größte Teil der nicht-übereinstimmenden Codes besteht in der Unterscheidung der Kategorien *Fachinhalt* und *Aufgabenstellung* in der Oberkategorie *Verständnisprobleme*. Von der Berechnung ausgenommen wurden die Codes *Interessante Einzel-Aussagen* und *effektives Arbeiten an Tabelle*, da diese nur von einer Person codiert wurden.

Tabelle 9.1: *Auszug aus dem Kategoriensystem zur Analyse der drei Schwerpunkte in der Unterrichtseinheit zum Prüfkriterium Korrelation und Kausalität.*

Kategorie			**Definition**
Kürzel		Bezeichnung	
allg-1		Off-Task	Mindestens ein:e Lernende:r der Fokusgruppe arbeitet nicht an der Aufgabenstellung.

	allg-1_1	Verlassen des Sitzplatzes	Die Schülerin oder der Schüler verlässt seinen Sitzplatz und bewegt sich im Klassenraum.
	allg-1_2	allgemeine Interaktion	Es wird mit der unbelebten Umwelt (Stifte, Tisch, Papier, Handy, ...) in einer Weise interagiert, die der Bearbeitung der aktuellen Aufgabenstellung nicht zweckdienlich ist.
	allg-1_3	Gespräch	Es wird ein Gespräch mit einer/einem oder mehreren Partner:innen geführt, das der Bearbeitung der aktuellen Aufgabenstellung nicht zweckdienlich ist.
	allg-1_4	Tablet-Interaktion	Es wird mit dem Tablet in einer Weise interagiert, die der Bearbeitung der aktuellen Aufgabenstellung nicht zweckdienlich ist.
allg-2		Aha-Effekt	Lernende äußern verbal oder mittels Mimik/Gestik, dass ihnen ein Sachverhalt „plötzlich“ klar wird.
allg-4		Verständnisprobleme	Oberkategorie für Verständnisprobleme jeglicher Art, die bei den Lernenden auftreten.
	allg-4_1	Technik	Die Lernenden tätigen einen Ausruf, führen eine Geste aus oder schreiben etwas auf, das auf ein fehlerhaftes oder unvollständiges Verständnis eines Fachinhaltes hinweist.
	allg-4_2	Fachinhalt	Die Lernenden tätigen einen Ausruf, führen eine Geste aus oder schreiben etwas auf, das auf ein fehlerhaftes oder unvollständiges Verständnis eines Fachinhaltes hinweist.
	allg-4_3	Aufgabenstellung	Die Lernenden tätigen einen Ausruf, führen eine Geste aus oder schreiben etwas auf, das auf ein fehlerhaftes oder unvollständiges Verständnis der Aufgabenstellung hinweist.

allg-5		Interessante Einzel-Aussagen	In dieser allgemeinen Kategorie werden forschungsmethodisch-interessante Einzel-Aussagen von Lernenden codiert, die im späteren Verlauf der Auswertung wichtig sein könnten.
28-3	28-3.1	Effektives Arbeiten an Tabelle	Wird in Zeiträumen codiert, in denen die Lernenden offensichtlich konzentriert an der Tabelle in der App Numbers arbeiten.

Die folgende Ergebnisdarstellung erfolgt zunächst gruppenweise und in einem zweiten Schritt auch zusammenfassend für die Gesamtheit aller Gruppen. Das Auftreten der codierten Ereignisse im zeitlichen Verlauf der Unterrichtsstunde wurde für alle Gruppen in Form einer *Codeline* (angelehnt an das Visualisierungs-Instrument in MAXQDA) visualisiert. Die Codelines befinden sich im Anhang dieser Arbeit (B.3.3).
Verständnisschwierigkeiten sind in Gruppe 1 vorwiegend in den ersten zwei Dritteln der *AP Tabelle* erkennbar. Technische Probleme treten während der AP immer wieder auf (Minuten 36, 42, 48, 58). Schwierigkeiten mit dem Fachinhalt werden teilweise geäußert (Minuten 40-41, 47-48) und auch die Aufgabenstellung ist nicht durchweg klar (Min. 37, 45-46, 49-52, 56). Erkennbar ist dies an Aussagen wie: „S1-02: Digga keine Ahnung was ich da hinschreiben soll“ (*Fachinhalt*, Pos. 233), „S1-01: Hä, ich verstehe nicht was wir machen sollen. Was sollen wir machen?“ (*Aufgabenstellung*, Pos. 98 bzw. Min. 45) oder „S1-02: Wie kann man das zurücksetzen?“ (*Technik*, Pos. 185).
Aha-Erlebnisse gibt es in drei Situationen (Min. 22, 55, 62), zwei davon liegen in der AP Tabelle. In der ersten Situation während der Wiederholungsphase erinnern sich die Schüler:innen, dass bei der Randomisierung der Begriff *zufällig* eine Rolle spielt. Das zweite *Aha-Erlebnis* ist eine Irritation eines Schülers, der beim Bearbeiten der Tabelle fehlende Daten bemerkt und im Anschluss erkennt, dass ein gesetzter Filter diese Daten ausgeblendet hatte (Gruppe 1, 00:54:44). Die letzte Aha-Situation zum Ende der AP Tabelle besteht in dem Ausruf eines Schülers, der Sachverhalt sei „ja easy“ (Gruppe 1, 01:01:35).
Off-Task-Handlungen sind über die gesamte Unterrichtsstunde verteilt in Gruppe 1 immer wieder zu beobachten. Teilweise dauern diese Handlungen mehrere Minuten

an. Sie konzentrieren sich auf Off-Task-Gespräche in der Arbeitsphase zu AB 3 (Min. 80-90, 92-95), sind aber auch in der AP Tabelle vorhanden (Min. 37-40, 43-44, 56-59). Off-Task-Interaktionen mit dem Tablet sind nur in der Wiederholungsphase (Min. 25-27) erkennbar.

In Gruppe 2 sind Verständnisschwierigkeiten selten. Zu Beginn des zweiten Drittels der AP Tabelle ist eine Aufgabenstellung unklar (Min. 45-46). In der nachfolgenden Plenumsphase äußern die Lernenden an zwei Stellen technische Probleme (Min. 64 und 69).

Aha-Erlebnisse sind in den Aktionen der Gruppe 2 nicht zu beobachten.

Off-Task-Interaktionen gibt es kaum, lediglich zum Ende der AP Tabelle, nachdem die Schüler:innen ihre Arbeit beendet hatten, befindet sich ein längerer Block an Off-Task-Gesprächen (Min. 60-64). In der letzten Arbeitsphase sind in zwei Situationen Tablet-Interaktionen erkennbar, die nicht der Bearbeitung der Aufgabenstellung dienten (Min. 78-79, 86-87).

Gruppe 3 hatte keine technischen Schwierigkeiten, äußerte einige Verständnisschwierigkeiten bezüglich des Fachinhalts in der AP Tabelle (Min. 43-45, 48, 55-56) und an drei Stellen Probleme mit der Aufgabenstellung (Min. 30, 45, 84).

Die Schüler:innen der Gruppe 3 zeigten an vielen Stellen *Aha-Erlebnisse*. Zu Beginn der Plenumsphase zur Wiederholungsaufgabe (Min. 27) und zum Ende der AP zu AB3 (Min. 89) gibt es jeweils ein *Aha-Erlebnis*, größtenteils sind diese jedoch in der AP Tabelle verortet (Min. 48-50, 53-54, 59, 63-65). Die Äußerungen der Lernenden betreffen dabei bestimmte Arbeitsschritte beim Ausfüllen der Tabelle, zum Beispiel von Schüler 1-05: „Ja, das können wir auch so durchzählen/ 1-07: Da muss jetzt ne vier hin“ (Gruppe 3, 00:47:30). Sie betreffen aber auch allgemeine Ausrufe, die auf eine plötzliche Erkenntnis hindeuten („1-08: Ich hab’s sogar verstanden!“, Gruppe 3, 00:52:15).

Off-Task Gespräche treten vermehrt auf, vor allem in der Wiederholungsphase (Min. 24-26), kurz in der AP Tabelle (Min. 45-46) und der AP zu AB3 (Min. 87-89) sowie in Plenumsphasen (Min. 31-32, 68-71, 89-92). Punktuell sind Off-Task-Tablet-Interaktionen erkennbar (Min. 21, 70-71).

Verständnisschwierigkeiten in Gruppe 5 gibt es nur punktuell. Im ersten Drittel der AP Tabelle (Min. 39-40) sowie in der AP zu AB3 (Min. 78) gibt es Probleme mit der Aufgabenstellung. Außerdem werden Probleme mit dem Fachinhalt geäußert (Min. 42-43 und 81).

Tabelle 9.2: *Auszug aus den mit Aha-Erlebnis codierten Segmenten aus der Unterrichtsstunde zum Prüfkriterium Korrelation und Kausalität.*

Quelle	Zeitmarke	codiertes Segment
28-gruppe1, Pos. 147	0:54:44	S1-01: Hä bei mir fehlen zwei. Bei mir fehlen zwei Leute. Bei dir sind auch nur 38.
28-gruppe1, Pos. 184	1:01:33	S1-01: Oh nein, jetzt hab ich es vergessen. Das ist ja richtig easy.
28-gruppe3, Pos. 166	0:47:18	1-08: Das, ohh, guck mal. Das sind ja jetzt die Installierten und das sind die nicht Installierten, müssen wir jetzt alles so einordnen. Das ja musst du wegmachen.
28-gruppe3, Pos. 186	0:52:13	1-08: Ich hab's sogar verstanden.
28-gruppe3, Pos. 354	1:28:17	1-08: Das hängt von Instagram selbst ab. 1-07: Hängt nicht. 1-08: Achso, ja dann.
28-gruppe5, Pos. 75	0:47:30	2-11: Ich glaub wir haben das hier falsch gemacht. Weißt du warum? Weil, warte. Wir haben/ ich hab nur das gemacht bei mir. Nur das, aber wir müssen ja das über 20 Stunden und nicht nur 20 bis 24 Stunden.
2020-09-28_Korrelation, Pos. 421	0:58:45	1-05: Ah, ich glaub ich habs ein bisschen verstanden.

Es gibt ein Aha-Erlebnis in der *AP Tabelle* (Min. 48-49).
Off-Task-Gespräche treten nur in der Plenumsphase zur Tabelle (Min. 68-72) und in der letzten Arbeitsphase auf (Min. 82-85). Außerdem gibt es allgemeine Off-Task-Interaktionen zum Ende der Stunde (Min. 87-90). Diese Gruppe zeichnet sich durch eine sehr effektive Arbeitsweise in der AP Tabelle aus (Min. 46-64).

Ergebnisdarstellung nach Kategorie

Die hier beschriebenen und im Anhang dargestellten Codeline-Visualisierungen geben einen Überblick über die Ereignisse innerhalb der Fokusgruppen. Auftretende Aha-Erlebnisse, Verständnisschwierigkeiten und Off Task-Handlungen wurden beschrieben. Lohnenswert ist jedoch auch die Betrachtung aus einer anderen Perspektive. Werden die Codelines nicht nach Gruppen sondern nach Codes getrennt, dann lässt sich das Auftreten der isolierten Codes über mehrere Gruppen vergleichen. Dieser Vergleich wird im folgenden Abschnitt dargestellt.
Die Codeline im Anhang B.3.2 stellt das Auftreten des Codes *Verständnisschwierigkeiten* im zeitlichen Verlauf über alle Fokusgruppen dar. Es wird verdeutlicht, dass der größte Teil auftretender Verständnisschwierigkeiten in der ersten Hälfte der AP Tabelle liegt. Von 83 codierten Verständnisschwierigkeiten handelt es sich in 28 Fällen (etwa ein Drittel) um technische Probleme, in 22 Fällen (etwa ein Viertel) um Schwierigkeiten beim Verständnis des Fachinhalts und 33 mal (etwa 40 %) wurde zum Ausdruck gebracht, dass die Aufgabenstellung nicht verstanden wurde.
Eine weitere Codeline (Anhang B.3.2) zeigt, dass *Aha-Erlebnisse* überwiegend im späteren Verlauf der AP Tabelle liegen und teilweise bis in die Plenumsphase hinein auftreten.
Off-Task-Handlungen sind hingegen über die gesamte Unterrichtsstunde verteilt und äußern sich überwiegend in Off-Task-Gesprächen zwischen den Lernenden (siehe Anhang B.3.2). Auffällig ist, dass die Gruppen 4 und 5 vor allem zum Ende der Stunde vermehrt *off-task* interagieren.

9.2.4 Auswertung

In einem Umfang von drei Doppelstunden wurden Unterrichtsmaterialien erprobt, davon wurde eine Doppelstunde zur Erarbeitung des Prüfkriteriums *Korrelation und*

Kausalität mit Lernenden der neunten Klasse einer Gesamtschule genutzt. Die Unterrichtsstunden wurden videographiert und einer qualitativen Inhaltsanalyse unterzogen. Die Schwerpunkte der Analysen in diesem Kapitel liegen in der Ermittlung von Verständnisschwierigkeiten, denen Lernende im Unterrichtsverlauf begegnen sowie dem Aufdecken von *Aha-Erlebnissen* bei den Schüler:innen. Weiterhin wurden Situationen beschrieben, in denen die Lernenden offensichtlich *off-task* gehandelt oder besonders effektiv an der Tabelle gearbeitet haben.
Die Analyseergebnisse zusammenfassend zu betrachten und aufeinander zu beziehen soll eine Aussage über das Erreichen der zuvor formulierten Lernziele ermöglichen.
Die Lernziele beschreiben, durch Operatoren gestützt, die in der Unterrichtsstunde durchzuführenden Handlungen und implizieren damit einen Kompetenzzuwachs seitens der Schüler:innen. Beispielsweise setzt das Erreichen des dritten Lernziels („Schüler:innen bewerten Schlussfolgerungen, die von Dritten anhand der Studien-Daten getroffen wurden“) einen Gewinn an Bewertungskompetenz voraus. Das Erreichen der Lernziele wird im folgenden Abschnitt geprüft.

Lernziel 1

Das erste Lernziel: „Die Schüler:innen filtern Rohdaten einer wissenschaftlichen Studie zu einem lebensnahen Kontext nach verschiedenen untersuchten Faktoren und werten sie aus“ wird durch das zweite Arbeitsblatt adressiert. Dabei setzen sich die Lernenden mit einer Studie zum Stromverbrauch der App *Instagram* auseinander und analysieren Rohdaten der Studie mithilfe eines Tabellenkalkulationsprogramms auf einem Tablet.
Es hat sich gezeigt, dass in der Gruppenarbeitsphase *AP Tabelle* die meisten Anzeichen auf Verständnisschwierigkeiten zum Ausdruck gebracht werden. Das bedeutet zwar, dass die Schüler:innen nicht unmittelbar an der Lösung des Arbeitsblattes arbeiten konnten, allerdings häufen sich in der zweiten Hälfte der Arbeitsphase die *Aha-Erlebnisse* und damit die ersten Anzeichen auf die Auflösung möglicher kognitiver Konflikte. Der geringe Off-Task-Anteil in dieser Arbeitsphase mit gleichzeitig häufig auftretenden Anzeichen für *effektives Arbeiten an der Tabelle* bestätigt die Aussage, dass die Lernenden sich intensiv mit den Rohdaten der Studie auseinander gesetzt haben.

Die Datenlage weist darauf hin, dass das erste Lernziel erreicht wurde.

Lernziel 2

Im Anschluss an die Auswertung der Rohdaten ziehen die Lernenden ihre eigenen Schlussfolgerungen und nutzen die Tabelle auf dem Arbeitsblatt 2b als Strukturierungshilfe. Dadurch wird das zweite Lernziel adressiert: „Die Schüler:innen ziehen eigene Schlussfolgerungen aus den analysierten Daten".
Das Ausfüllen der Tabelle sorgt allerdings nicht automatisch dafür, dass die Lernenden eigene Schlussfolgerungen ziehen. Das wurde in der Plenumsphase zur Zwischensicherung deutlich: Die Schüler:innen gaben als Reaktion auf den Impuls *28g* die zuvor eingetragenen Werte wieder, ohne sie in einen Sinn-Zusammenhang zu stellen. Erst danach, als die Lehrkraft explizit dazu aufrief, Schlussfolgerungen aus den Werten zu ziehen (Impuls 28h), reagierten die Lernenden entsprechend: „Also es liegt nicht nur/ es liegt nicht nur an der App sondern, wenn man viel mit dem Handy macht, müsst man auch länger/öfter laden" (2020-09-28_Korrelation, Pos. 544, 1:10:35).
Die Erkenntnisse aus der Auswertung der Plenumsphasen (Kapitel 9.2.3) legen nahe, dass das zweite Lernziel nicht allein anhand der Arbeitsblätter erreicht werden konnte. Erst die unterstützende Plenumsphase sorgt dafür, dass die Lernenden passende Schlussfolgerungen ziehen.
Der Blick auf Schüleraussagen innerhalb von Gruppenarbeitsphasen kann dieses Ergebnis etwas weiter ausschärfen. Während der Bearbeitung der Tabellendaten sagte ein Schüler aus Gruppe 2: „Ich weiß schon warum ich mir Instagram nicht hole. Meine Akku-Laufzeit ist sowieso schon scheiße" (S2-02, Gruppe 2, Pos. 95, 0:59:05). Zwar deuten derartige Aussagen darauf hin, dass die Lernenden aus den Ergebnissen eine eigene Schlussfolgerung ziehen, empirisch absichern lässt sich dies jedoch nicht.

Lernziel 3

Die Aufgabe 2 des AB 1 fordert die Lernenden zur Bewertung der angegebenen Schlussfolgerung auf, passend zum dritten Lernziel: „Die Schüler:innen bewerten Schlussfolgerungen, die von Dritten anhand der Studien-Daten getroffen wurden".
Eine wichtige Datengrundlage für dieses Lernziel bilden die von den Lernenden ausgefüllten Arbeitsblätter, speziell das AB 1. Die Auswertung dieser Artefakte hat je-

doch ergeben, dass die meisten Lernenden auf den Arbeitsblättern die *Plenumsantwort* notiert haben, also die in der Plenumsphase gemeinsam erarbeitete Antwort. Es wurde also keine eigene Wortwahl verwendet, daher eignen sich diese Daten nicht, um das flächendeckende Erreichen des Lernziels zu untersuchen.
Die Impuls-Reaktionen in der Plenumsphase geben jedoch Auskunft darüber, inwiefern einzelne Lernende die Schlussfolgerung bewerten. Dafür werden vor allem die offenen Bewertungsimpulse 28i und 28m betrachtet. Wie bereits beschrieben wurde, gehören fast alle Reaktionen auf diese Impulse der Kategorie 1 oder 2 an und sind damit passend und teils anschlussfähig. Diese Erkenntnis ist ein Anzeichen dafür, dass das dritte Lernziel von einigen Schüler:innen erreicht wurde.

Lernziel 4

Das Lernziel „Die Schüler:innen erarbeiten das Konzept der Scheinkausalität anhand von Prüfkarten“ fokussiert die Prüfkarten (AB 3), die im Verlauf der Unterrichtsstunde von den Lernenden bearbeitet werden. Den zentralen dabei zu erarbeitenden Inhalt stellt das Konzept der Scheinkausalität dar. Die Lernenden sollen erkennen, dass für valide Schlussfolgerungen stets alle relevanten Störfaktoren beachtet werden müssen.
Den Begriff der Scheinkausalität haben die Lernenden im Unterricht nicht häufig benutzt. In der Impuls-Reaktion 28m-I benutzt ein:e Schüler:in den Begriff in der Erklärung des Fehlers in der Studie: „[...], sie haben nicht darauf geachtet, dass vielleicht eine andere App den Akku fressen könnte. Dadurch haben sie eine Scheinkausalität erzeugt, denn der dritte Faktor muss auch beachtet werden, so wie ersten zwei“ (2020-09-28_Korrelation, Pos. 656, 1:32:43). Weiterhin passend und anschlussfähig sind die Impuls-Reaktionen 28n-I („2-16: Nein, weil die anderen Faktoren nicht beachtet wurden“) und 28o-I (2-16:[Ich habe aus dieser Stunde mitgenommen,] „dass Korrelation Zusammenhänge sind und Kausalität Ursachen“).
Die Betrachtung der Gruppenarbeitsphasen liefert keine genaueren Erkenntnisse über das Erreichen dieses Lernziels.

Kognitive Konflikte

Das Lernmaterial wurde vor dem Hintergrund entwickelt, dass Lernende bei der Bearbeitung einen kognitiven Konflikt erleben sollen (siehe Abschnitt 9.2.3). Dazu soll

eine Unzufriedenheit mit der eigenen Vorstellung, die der fachwissenschaftlichen Vorstellung widerspricht, hergestellt werden. In diesem Fall handelt es sich weniger um eine fachwissenschaftliche Schülervorstellung; stattdessen geht es um die Kompetenz, logische Schlussfolgerungen aus Studiendaten korrekt zu ziehen und dabei den Unterschied zwischen Korrelationen und Kausalitäten zu berücksichtigen.

Der im Unterricht behandelte Kontext stellt eine Schlussfolgerung dar, die „auf den ersten Blick" logisch erscheint und von den Schüler:innen vermutlich nachvollzogen werden kann. Im Laufe der Unterrichtsstunde wird jedoch klar, dass die Schlussfolgerung nicht wissenschaftlich korrekt aufgestellt wurde. An dieser Stelle sollten die Lernenden einen kognitiven Konflikt erfahren. Wie oben beschrieben (Abschnitt 9.2.3) sollen sie erkennen, dass einzelne isolierte Faktoren nicht für kausale Erklärungen verwendet werden können. In der aufgezeichneten Unterrichtsstunde wurden im Zusammenhang mit dieser Erkenntnis einige auftretende *Aha-Erlebnisse* gefunden.

Diese sind vor allem im späteren Verlauf der *AP Tabelle* verortet, nachdem der Arbeitsauftrag durch die Lehrperson im Plenum präzisiert wurde. Sie äußern sich zum Beispiel durch einen Ausruf, wie: „Ich hab's sogar verstanden!". Im Anschluss an auftretende Aha-Erlebnisse ist häufig zu beobachten, dass der betroffene Schüler seiner Mitschülerin den soeben verstandenen Sachverhalt erklärt (2-11: „Ich glaub wir haben das hier falsch gemacht. Weißt du warum? Weil, warte. Wir haben/ ich hab nur *das* gemacht bei mir. Nur das, aber wir müssen ja das über 20 Stunden und nicht nur 20 bis 24 Stunden").

Weiterhin konnte beobachtet werden, dass im späteren Verlauf der Stunde das Auftreten von Verständnisschwierigkeiten abnimmt. Diese Erkenntnis stützt die Auffassung, dass die Lernenden zuvor einen kognitiven Konflikt auflösen konnten – und dass die Auflösung möglicherweise durch die Aha-Erlebnisse begleitet wurde, die die qualitative Inhaltsanalyse aufdecken konnte.

9.2.5 Re-Design

Entwicklung von Leitimpulsen

Die Analysen der von der Lehrkraft gegebenen Impulse in Plenumsphasen der Unterrichtsstunde sowie der Vergleich aufgetretener Reaktionen der Lernenden mit intendierten Reaktionen haben gezeigt, dass Fragen und Impulse unterschiedliche Reak-

tionen hervorrufen können.
Die Lernenden antworten nicht immer zielführend auf Fragen der Lehrkraft oder reagieren auf Impulse in einer Weise, die einem stringenten Unterrichtsverlauf nicht zuträglich ist. Mit der Einteilung der Impuls-Reaktionen in Kategorien konnte dieser Blick auf das Unterrichtsgeschehen während der Plenumsphasen systematisiert werden.
In neun von insgesamt 15 untersuchten Fällen reagierten die Schüler:innen passend und anschlussfähig auf Lehr-Impulse. Das zeigt, dass die Kommunikation zwischen der Lehrkraft und den Lernenden in diesen Fällen gut funktionierte. Die sechs zugehörigen Impulse sind folglich angemessen und in den Plenumsphasen passend platziert. Im Zuge der Weiterentwicklung der Unterrichtskonzeption können die Impulse der Kategorie 1 beibehalten und in die Entwicklung gezielter Leitimpulse eingebracht werden.
Vier Impulse der Kategorie 2 und zwei Impulse der Kategorie 4 bringen keine anschlussfähigen Reaktionen hervor und sind demnach zu vermeiden. Sie spielen in der folgenden Entwicklung der Leitimpulse für die Plenumsphasen eine untergeordnete Rolle.

In der Erprobung der Unterrichtsmaterialien zum Prüfkriterium *Korrelation vs. Kausalität* bearbeiteten die Schüler:innen innerhalb von Gruppenarbeitsphasen verschiedene Aufgabenstellungen. Die 90-minütige Unterrichtsstunde zeichnete sich durch Wechsel zwischen Gruppenarbeits- und Plenumsphasen aus. Die Plenumsphasen wurden von der Lehrkraft moderiert und mithilfe von Impulsen geleitet, die die Lernenden zur Präsentation der erarbeiteten Ergebnisse sowie zur Diskussion der Unterrichtsinhalte innerhalb der Klassengemeinschaft aufforderten.
Durch die Analysen der Impulse in den Plenumsphasen wurden insgesamt neun Lehrimpulse in zwei Plenumsphasen identifiziert, die zur Moderation der Plenumsphasen geeignet sind. Die Impulse 28h und 28i treten in der Phase zur Zwischensicherung der Arbeitsergebnisse von Arbeitsblatt 1 auf, während 28k, 28l, 28m, 28n und 28o in der Sicherungsphase zum Arbeitsblatt 3 vorkommen. 28p ist in der Abschlussreflexion verortet.

Leitimpulse in der Zwischensicherung zum AB 1

Zu Anfang der Zwischensicherung werden die Ergebnisse des Arbeitsblatts 1, also

die Eintragungen in der Tabelle, im Plenum verglichen. Impuls 28g hat sich als nicht sinnstiftend erwiesen, da die Lernenden als Reaktion lediglich einzutragende Zahlen ohne weiterführende Erklärungen oder Schlussfolgerungen nannten (s. o.). Erst Impuls 28h zur gleichen Thematik schaffte Abhilfe, allerdings bedurfte es mehrfacher Nachfrage durch die Lehrperson, bis alle Tabellenwerte erklärt waren. Es bietet sich an, beide Impulse zu vereinen und gezielter zu formulieren, um die Anzahl benötigter Nachfragen zu verringern. Der folgende Leitimpuls für den ersten Teil dieser Plenumsphase wird definiert:
Jede:r von euch nennt bitte einen Tabellenwert, den sie oder er eingetragen hat - in einem vollständigen Satz, der zeigt, was uns der Wert sagt. Ich nehme die erste Person dran, dann wählt jeder eine weitere Person aus, die den nächsten Wert erklärt.
Im zweiten Teil der Plenumsphase ist es die Aufgabe der Lernenden, die in der Studie gegebene Schlussfolgerung auf der Grundlage der zuvor ausgefüllten Tabelle zu bewerten. Impuls 28i hat ohne weitere Nachfragen zu einer zielführenden, differenzierten Antwort geführt und wird als Grundlage für den nächsten Leitimpuls verwendet:
Nachdem wir unsere eigenen Schlussfolgerungen zu den Tabellenwerten gezogen haben, wollen wir nun einen Blick werfen auf die Schlussfolgerung der Forscherinnen und Forscher. Bewertet doch mal die Schlussfolgerung!

Leitimpulse in der Sicherungsphase zum AB 3

Die Sicherungsphase zum AB 3 ist im späteren Teil der Unterrichtsstunde verortet. Sie schließt unmittelbar an die Arbeitsphase an, in der die Lernenden das Prüfkriterium *Korrelation vs. Kausalität* erarbeiten. Es konnte gezeigt werden, dass die Impulse 28l, 28m, 28n und 28o passend in die Unterrichtsstunde integriert wurden. Die Plenumsphase zeichnet sich bei Verwendung all dieser Impulse jedoch durch ein stark leitendes Unterrichtsgespräch seitens der Lehrkraft aus und lässt wenig Raum für Diskussionen. Es bietet sich daher an, die Plenumsphase mit einem starken Leitimpuls zu versehen, der die Impulse 28l bis 28o ersetzt:
Bewertet mit diesem neuen Wissen bitte, ob die Forschergruppe ihre Ergebnisse richtig gedeutet hat!
Dieser Leitimpuls ist recht offen gehalten und bietet Spielraum für unterschiedliche Reaktionen der Lernenden. Der Operator *Bewerten* verdeutlicht den Lernenden, auf welche Weise sie auf den Impuls reagieren können. Da in der vorausgegangenen Arbeitsphase das Konzept *Korrelation vs. Kausalität* ausführlich behandelt wurde, soll-

ten den Lernenden genügend Anknüpfungspunkte für die Teilnahme an der Plenumsdiskussion präsent sein. Falls die Diskussion jedoch keine zielführende Richtung einschlägt, kann die Lehrkraft weitere stichwortartige Impulse zur Moderation des Gesprächs einsetzen.

Optimierung des Ablenkungspotentials

Off-Task Handlungen sind über die gesamte Unterrichtsstunde verteilt erkennbar, häufen sich jedoch im letzten Drittel der Stunde. Es handelt sich dabei meist um Off-Task-Gespräche zwischen den Lernenden. Auch Tablet-Interaktionen, die keinen direkten Bezug zum Arbeitsauftrag besitzen, waren immer wieder zu beobachten.

Um das Auftreten von Off-Task-Handlungen zu verringern, kann in Zukunft darauf geachtet werden, das von digitalen Medien ausgehende Ablenkungspotential, beschrieben u. a. bei Schaumburg (2015), Welling et al. (2014) oder Dellbrügge (2020, S. 32) zu minimieren. Eine denkbare Möglichkeit, dem Ablenkungspotential zu begegnen, sind z. B. Bildschirmsperren. Diese werden von der Lehrkraft punktuell eingesetzt, um die Bedienoberfläche der Tablets für Phasen zu sperren, in denen die Tablets nicht benutzt werden sollen. Die Methode wurde in der erprobten Unterrichtsstunde bereits in der Plenumsphase nach Bearbeitung der Tabelle eingesetzt (2020-09-28_Korrelation, Pos. 441, 1:03:36) und könnte auf weitere Phasen, in denen die Tablets nicht benötigt werden, ausgeweitet werden. Allerdings muss dabei beachtet werden, dass die Schüler:innen zur Teilnahme an der Plenumsdiskussion ihre Arbeitsergebnisse weiterhin aufrufen können.

Manche Off-Task-Aktionen resultieren daraus, dass Lernende ihrer Langeweile Raum geben (Bolkan & Griffin 2017). Langeweile lässt sich im Klassenraum nicht komplett vermeiden, jedoch können Anpassungen in Form von Differenzierung der Aufgabenstellungen oder noch engerer Verknüpfung mit der Lebenswelt der Schüler:innen Langeweile reduzieren (Qian & Li 2017; Dellbrügge 2020, S. 43). Als Konsequenz für die Unterrichtsstunde zum Prüfkriterium *Korrelation und Kausalität* lässt sich festhalten, dass die innere Differenzierung, die in den Arbeitsblättern in Form von *Basis*, *Stern1* und *Stern2*-Aufgaben gegeben ist, durch die Lehrkraft stärker hervorgehoben und präziser erklärt werden sollte.

Minimieren von Verständnisschwierigkeiten

Es wurde herausgestellt, dass der größte Teil codierter Verständnisschwierigkeiten in der ersten Hälfte der *AP Tabelle* auftritt. Dabei handelt es sich sowohl um technische Probleme mit der Bedienung der Tablets als auch um Probleme mit dem Verständnis der Fachinhalte oder der Aufgabenstellungen. Da die Verständnisschwierigkeiten von Lernenden in den meisten Fällen mit allgemeinen Sätzen wie: „Ich verstehe das einfach nicht“ zum Ausdruck gebracht wurden, ist es schwierig, die Hintergründe der Schwierigkeiten zu erfahren.
Während der *AP Tabelle* gab die betreuende Lehrkraft, die als stille Beobachterin an der Unterrichtsstunde teilnahm, der durchführenden Lehrperson den Hinweis, dass die Lernenden den Arbeitsauftrag offenbar nicht verstanden hatten: „Ich glaub die wissen nicht klar warum die Filtern sollen“ (2020-09-28_Korrelation, Pos. 326, 0:47:25). Die Beobachterin ergänzte, dass es ratsam wäre, die Arbeitsphase zu unterbrechen und die Vorgehensweise im Plenum zu demonstrieren. Nachdem das geschehen war (Minute 54), wurden weniger Verständnisschwierigkeiten codiert.
Aus diesen Ergebnissen lässt sich ableiten, dass bei der Erteilung des Arbeitsauftrags kurz im Plenum abgeklärt werden sollte, wie in der App *Numbers* die Schnellfilter gesetzt werden können. Es bietet sich an, dass die Lehrkraft ihren Tablet-Bildschirm an die Projektionsfläche spiegelt und dort mit den Lernenden Schritt für Schritt einen Beispiel-Filter setzt. Damit erhalten die Schüler:innen einen Orientierungspunkt, bei dem sie ihre Analyse starten können und werden bezüglich der technischen Anforderungen und dem Verstehen der Aufgabenstellung kognitiv entlastet. Somit können sich die Lernenden auf den Fachinhalt konzentrieren, und das Auftreten von Verständnisschwierigkeiten könnte minimiert werden.

9.2.6 Fazit und Ausblick

Das vorangegangene Kapitel beschreibt einen Durchlauf des ersten Mesozyklus in diesem Projekt. Es wurde detailliert dargelegt, auf welchen Überlegungen das Prüfkriterium *Korrelation und Kausalität* aufbaut und wie eine Umsetzung für den naturwissenschaftlichen Unterricht der neunten Klasse gestaltet werden kann.
In der entwickelten Unterrichtsstunde bearbeiten Lernende die Rohdaten einer Studie, die die Akkulaufzeit von Smartphones mit der Applikation *Instagram* in Bezie-

hung setzt. Sie filtern die Rohdaten nach verschiedenen Bedingungen und ziehen eigene Schlussfolgerungen aus ihrer Datenanalyse. Weiterhin erarbeiten die Lernenden das Prüfkriterium *Korrelation und Kausalität* anhand der entwickelten Prüfkarten und wenden dieses auf die untersuchte Studie an. Dabei bewerten sie auch die von einem Team aus Forschenden getroffene Schlussfolgerung, nach welcher die Applikation *Instagram* für eine verkürzte Akkulaufzeit verantwortlich sei.
Die entwickelten Unterrichtsmaterialien wurden in einer neunten Klasse einer Gesamtschule erprobt. Aus den videographierten Daten der Erprobung wurden mittels qualitativer Inhaltsanalyse Aussagen über Arbeitsweise der Lernenden in Plenums- und Arbeitsphasen generiert sowie das Lehrerverhalten in Plenumsphasen analysiert. Dabei konnte festgestellt werden, dass alle vier zuvor formulierten Lernziele in der Stunde adressiert wurden. Ein Erreichen oder Nicht-Erreichen der Lernziele konnte anhand der vorliegenden Daten nicht endgültig geklärt werden. In der Auswertung (Kapitel 9.2.4) wurden jedoch begründete Vermutungen geäußert, nach denen das erste Lernziel vollständig, sowie die Lernziele zwei, drei und vier teilweise erreicht werden konnten.
In den Reaktionen der Lernenden konnten *Aha-Erlebnisse* gefunden werden (Abschnitt 9.2.4). Die Häufung der Aha-Erlebnisse zum Ende der Haupt-Arbeitsphase mit gleichzeitig weniger auftretenden Verständnisschwierigkeiten deutet darauf hin, dass durch die Bearbeitung des Arbeitsmaterials kognitive Konflikte aufgelöst wurden. Dies ist allerdings eine Kausalaussage, die durch die vorliegenden Daten nicht ausreichend gestützt werden kann. Zukünftige Untersuchungen könnten an dieser Stelle ansetzen.
Ein detaillierter Blick auf die Lehrimpulse in Plenumsphasen hat gezeigt, dass die meisten der von der Lehrkraft gestellten Fragen und Impulse bei den Lernenden für passende und anschlussfähige Reaktionen gesorgt haben. Allerdings wurden die Plenumsphasen durch das Lehrerverhalten stark vorstrukturiert und wiesen einen fragend-entwickelnden Charakter auf. Daher wurden auf Grundlage der analysierten Impulse für jede Plenumsphase gezielte Leitimpulse formuliert, die für eine offenere Gestaltung der Diskussionen sorgen sollen.
Schließlich wurden auf der Grundlage der Analyseergebnisse für die Codes *Verständnisschwierigkeiten* und *Off-Task-Interaktionen* Ableitungen getroffen für ein Re-Design des entwickelten Arbeitsmaterials bzw. der Stundenplanung. Im Fokus standen dabei die Optimierung des Ablenkungspotentials, welches u. a. von der Nutzung digi-

taler Medien ausgeht, sowie die Minimierung von Verständnisschwierigkeiten durch Reformulierung von Arbeitsaufträgen und konkrete Handlungsanweisungen für die Unterrichtsgestaltung.

9.3 MZ 1.3: Randomisierung

9.3.1 Theoretischer Hintergrund zu Randomisierung

Randomisierte kontrollierte Studien sind der Goldstandard der wissenschaftlichen Forschung. Wird eine Studie sowohl korrekt randomisiert als auch kontrolliert, dann ist die Wahrscheinlichkeit des Auftretens von Verzerrungseffekten gering und die interne Validität ist gesichert (Westermann 2000, S. 311).
Kontrolliertheit bzw. kontrollierte Bedingungen liegen vor, wenn das Design der Studie eine Experimentalgruppe und eine Kontrollgruppe beinhaltet. Weiterhin darf in experimentellen Settings stets nur eine Variable verändert werden. Eine detailliertere Beschreibung befindet sich im Kapitel 2.4.2 sowie bei Jungkamp (2021).
Die Forderung nach Randomisierung als Grundprinzip experimenteller Forschung wurde bereits in den 1920er Jahren von Fisher entwickelt und 1935 beschrieben (Fisher 1935; Schulz & Grimes 2002). Dabei bezeichnet Randomisierung Aspekte des Studiendesigns rund um die Stichprobe der Studie. Gefordert wird die zufällige Zuordnung der Proband:innen zu den Untersuchungsgruppen (Döring & Bortz 2016). Es darf kein Muster bei der Zuteilung zu Experimental- und Kontrollgruppe zu erkennen sein. Ziel der Randomisierung ist es, den Einfluss personengebundener Störvariablen zu minimieren. Die verschiedenen Gruppen sollen sich hinsichtlich der Ausprägung aller möglichen psychologischen und sozialen Merkmale, die als personengebundene Störvariablen wirken könnten, nicht systematisch unterscheiden sondern weitgehend äquivalent sein (Döring & Bortz 2016, S. 196).
Der Vorgang des Randomisierens kann klar abgegrenzt werden von der Stichprobenauswahl. Letztere sollte ebenso randomisiert – also nach einem Zufallsverfahren – erfolgen wie die Zuordnung der Proband:innen zu den Untersuchungsgruppen, wird aber in der empirischen Forschung nicht klassisch zur Randomisierung gezählt. Die Rekrutierung von Proband:innen unter Anwendung von Zufallsverfahren ist also zu unterscheiden von der zufallsbedingten Zuordnung der Proband:innen zu Experimental- und Kontrollgruppe.

Die Stichprobenauswahl kann nach verschiedenen Verfahren durchgeführt werden. Es wird unterschieden zwischen probabilistischen (zufälligen) Auswahlverfahren und nicht-probabilistischen Auswahlverfahren. Probabilistische Verfahren folgen einer statistischen Zufallsauswahl und sind charakteristisch für groß angelegte quantitative Untersuchungen. Dabei kann es sich um reine Zufallsstichproben, systematische, geschichtete oder Klumpen-Stichproben handeln (Mossig 2012).
Nicht-probabilistische Stichprobenauswahlverfahren sind vorwiegend in der qualitativen Forschung anzutreffen. Sie beinhalten bewusste und willkürliche Auswahlverfahren. Zu ersteren gehört z. B. das Quotenverfahren, bei dem für infrage kommende Personen ein Quotenplan basierend auf verschiedenen persönlichen Merkmalen generiert und zur Auswahl verwendet wird. Ein willkürliches Verfahren ist z. B. die Selbstselektionsstichprobe, bei der alle Personen befragt werden, die sich auf eine Umfrageeinladung gemeldet haben (Döring & Bortz 2016, S. 294; Mossig 2012).
Nach der Auswahl der Stichprobe erfolgt die Zuteilung der einzelnen Individuen zur Experimentalgruppe oder Kontrollgruppe. Dieser Schritt bildet den Kern des Konzeptes Randomisierung. Die Zuteilung muss, um wissenschaftlichen Qualitätsstandards zu genügen, randomisiert erfolgen. Ein Zufallsverfahren ist zwingend erforderlich, da ansonsten Verzerrungseffekte auftreten können.
Vollständige Randomisierung kann nicht mithilfe von systematischen Verfahren erreicht werden. Vorgehensweisen, die z. B. nach Geburtsdatum des Probanden oder Datum der Registrierung für die Studie randomisieren, sind nicht zufallsbedingt und daher nicht geeignet für eine wissenschaftlich gültige Randomisierung (Schulz & Grimes 2002).

9.3.2 Entwicklung des Lernmaterials

Im vorangegangenen Kapitel wurde dargelegt, dass sich hinter dem Begriff Randomisierung ein komplexes Konzept verbirgt, das sowohl Stichprobenauswahlverfahren als auch Zuteilungsverfahren zu Experimental- und Kontrollgruppe umfasst. Je nach Anwendungsgebiet der Studie und erwarteter Tragweite der Ergebnisse gibt es verschiedene Verfahren zur Realisierung einer randomisierten Stichprobe. Die Bandbreite der Verfahren reicht von einfachem Münzwurf oder Würfeln bis zu systematischen Quotenverfahren.
Auch der Bereich, auf den Randomisierung angewendet werden kann, ist vielfältig.

Randomisiert werden können Personenzuteilungen zu Untersuchungsbedingungen, Reihenfolgen von Bedingungen, situationale Merkmale und auch andere Untersuchungseinheiten (Westermann 2000, S. 310).
Allen Aspekten gemeinsam ist jedoch das Konzept des Zufalls. Stichproben werden zufällig ausgewählt und Störvariablen sowie Personen werden nach einem Zufallsprinzip auf die Untersuchungsbedingungen verteilt.

Festlegen der Lernziele

Für den Einsatz im Unterricht als Teil der Konzeption choice^{2}reflect ist eine didaktische Reduktion des Prüfkriteriums Randomisierung sinnvoll und notwendig. Schüler:innen sollen die wichtigsten Punkte kennenlernen, die bei der Durchführung einer wissenschaftlichen Studie zu beachten sind. Eine detaillierte Behandlung von Randomisierungsverfahren wäre in diesem Fall nicht zielführend. Das Verständnis von *Randomisierung* für die Anwendung in der Unterrichtskonzeption lässt sich in einer Arbeitsdefinition festhalten: „Randomisierung ist eine Anforderung an wissenschaftliche Studien. Sie setzt voraus, dass bei der Stichprobenauswahl ein auf Zufall basierendes Auswahlverfahren verwendet wird. Eine korrekte Randomisierung sorgt dafür, dass die Repräsentativität der gezogenen Stichprobe für die Grundgesamtheit gegeben ist".
Das Unterrichtsmaterial zum Prüfkriterium Randomisierung wurde auf dieser Grundlage im Hinblick auf die folgenden Lernziele entwickelt:

1. Die Schüler:innen beschreiben, dass Vollerhebungen bei wissenschaftlichen Studien nicht praktikabel sind und stattdessen Stichproben aus der Grundgesamtheit gezogen werden
2. Die Schüler:innen erklären das Prinzip der Randomisierung als zufallsbedingte Stichprobenauswahl
3. Die Schüler:innen beschreiben die Randomisierung einer Stichprobe als Voraussetzung für dessen Repräsentativität
4. Die Schüler:innen prüfen die Stichprobe der Heinsberg-Studie auf Randomisierung und bewerten ihre Repräsentativität

Im Zuge der didaktischen Reduktion wurden folgende Aspekte der Randomisierung nicht in das Lernmaterial integriert:

- Randomisierung bedeutet zufallsbedingte Zuteilung der Proband:innen zu den Untersuchungsbedingungen
- in einigen Fällen ist es notwendig, systematischere Randomisierungsverfahren zu wählen
- enge Verzahnung von *Randomisierung* und *kontrollierte Bedingungen*

Randomisierung bei Studien zu COVID-19: Die Heinsberg-Studie

Die Einführung und Problemgewinnung zum Prüfkriterium Randomisierung wird anhand eines aktuellen Themas, das im Zuge der Corona-Pandemie im Jahr 2020 eine hohe mediale Aufmerksamkeit erlangte, durchgeführt. Die Schüler:innen erarbeiten die Inhalte der sogenannten Heinsberg-Studie und prüfen diese auf Randomisierung. Bei der Heinsberg-Studie handelt es sich um eine Untersuchung von Streeck et al. (2020) zur Ermittlung der Sterblichkeitsrate und Dunkelziffer von Infektionen mit dem Virus „SARS-CoV-2“, das auch Coronavirus genannt wird. Die Studie wurde im Zuge der Corona-Pandemie im Juni 2020 auf einem Preprint-Server veröffentlicht. Die in der Studie ermittelten Werte für Sterblichkeit und Dunkelziffer lassen sich auf die gesamte Bevölkerung der Gemeinde Gangelt übertragen, da die verwendete Stichprobe repräsentativ für Gangelt ist. Der Geltungsbereich der Studie lässt sich jedoch nicht auf die gesamte Bevölkerung Deutschlands ausweiten – denn für diese Gruppe ist die Stichprobe nicht repräsentativ. Diesen Sachverhalt haben die Autor:innen in der Studie zum Ausdruck gebracht. Dennoch gaben sie eine theoretische Hochrechnung an (Streeck et al. 2020):

> It is important to note that the infection rate in Gangelt is not representative for other regions in Germany or other countries. However, with the limitations discussed above, the IFR [infection fertility rate] calculated here remains a useful metric for other regions with higher or lower infection rates. If in a theoretical model the here calculated IFR is applied to Germany [...], the estimated number of infected in Germany would be higher than 1.8 Mio [...]. It will be very important to determine the true average IFR for Germany.

Die Formulierungen zeigen deutlich, dass die Forscher:innen sich des Geltungsbereichs der Studie bewusst waren. Die Angabe der Dunkelziffer für ganz Deutschland

erfolgte zu Anschauungszwecken und um eine grobe Einschätzung der Lage zur Diskussion zu stellen.

Dennoch wurden die Ergebnisse der Studie kontrovers diskutiert. Unterschiedliche Wissenschaftler:innen übten Kritik an der Hochrechnung der Sterblichkeitsrate (und damit der Dunkelziffer) auf ganz Deutschland (Kümmerer & Schmid-Johannsen 2020). Die Autor:innen reagierten auf diese Kritik mit einer Stellungnahme, in der erneut betont wird, dass es sich bei der Übertragung auf Deutschland um eine Beispielrechnung handele und die Ergebnisse nur auf Gangelt übertragbar seien (Universität Bonn 2020).

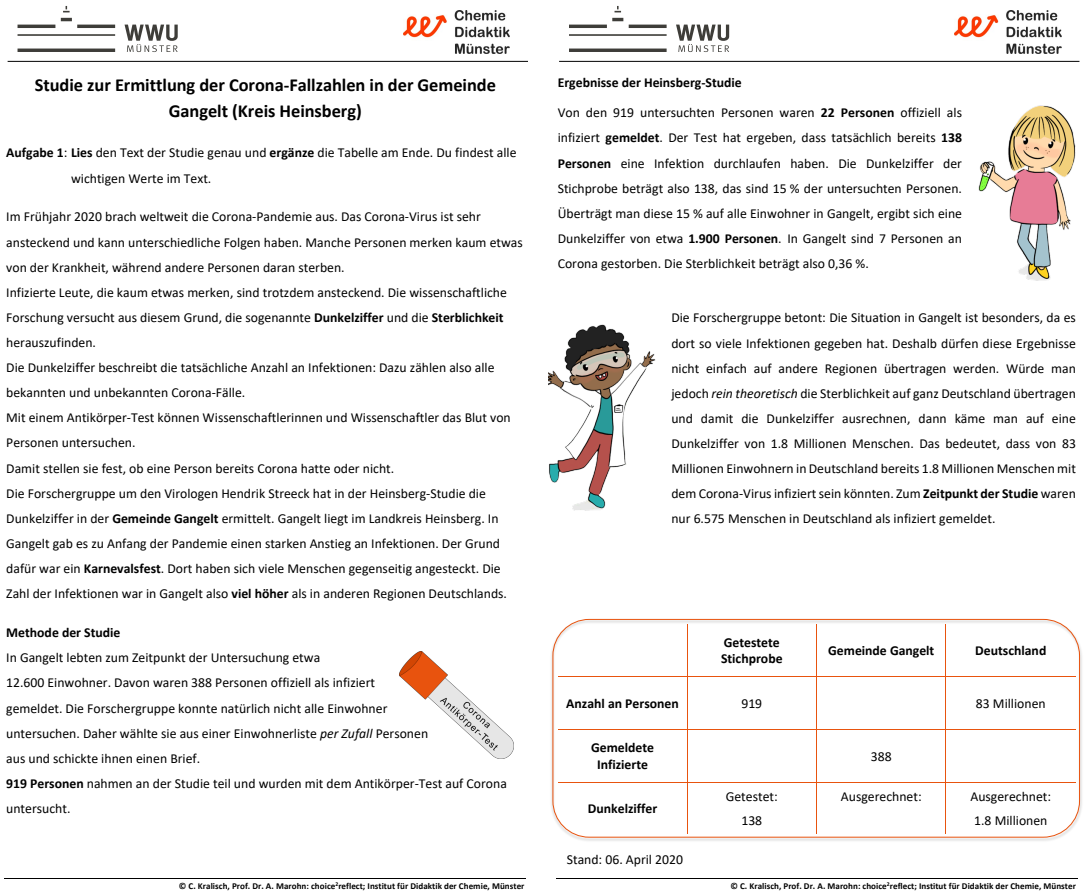
Die Diskussionen um die Methodik der Heinsberg-Studie zeigen, wie wichtig es ist, Randomisierung bei der Durchführung, Auswertung und auch der Interpretation wissenschaftlicher Studien zu beachten und Repräsentativitäten im Blick zu behalten. Daher bietet sich auch der Einsatz dieser Studie im Unterricht zur Einführung des Prüfkriteriums Randomisierung an.

Die Schüler:innen erarbeiten die Inhalte der Studie an einem zusammengefassten und vereinfachten Text, der sich über zwei DIN-A4 Seiten erstreckt (siehe Abbildung 9.10 sowie Anhang B.5.5). Der Text enthält einige Informationen zu Hintergrund, Methodik und Ergebnissen der Studie. Der Fokus liegt dabei zum Einen auf der Art und Weise, wie die Proband:innen rekrutiert wurden und weiterhin auf der ermittelten Anzahl der tatsächlichen Infektionen in der Gemeinde Gangelt. Der Text enthält auch den Hinweis, dass eine Hochrechnung auf die gesamte Bevölkerung Deutschlands einen rein theoretischen Nutzen hat und nicht so einfach durchgeführt werden darf.

Die erste Aufgabe des Arbeitsblatts bearbeiten die Lernenden in Einzelarbeit. Sie lautet:

Aufgabe 1: Lies den Text der Studie genau und ergänze die Tabelle am Ende. Du findest alle wichtigen Werte im Text.

Die Ergebnisse der Arbeitsphase werden im Plenum kurz zusammengetragen, sodass allen Lernenden eine kompakte Übersicht über die wichtigsten Daten der Heinsberg-Studie vorliegt. Im Anschluss beginnt die Erarbeitung des Prüfkriteriums *Randomisierung*, die anhand von Prüfkarten durchgeführt wird.



WWU MÜNSTER — Chemie Didaktik Münster

Studie zur Ermittlung der Corona-Fallzahlen in der Gemeinde Gangelt (Kreis Heinsberg)

Aufgabe 1: **Lies** den Text der Studie genau und **ergänze** die Tabelle am Ende. Du findest alle wichtigen Werte im Text.

Im Frühjahr 2020 brach weltweit die Corona-Pandemie aus. Das Corona-Virus ist sehr ansteckend und kann unterschiedliche Folgen haben. Manche Personen merken kaum etwas von der Krankheit, während andere Personen daran sterben.
Infizierte Leute, die kaum etwas merken, sind trotzdem ansteckend. Die wissenschaftliche Forschung versucht aus diesem Grund, die sogenannte **Dunkelziffer** und die **Sterblichkeit** herauszufinden.
Die Dunkelziffer beschreibt die tatsächliche Anzahl an Infektionen: Dazu zählen also alle bekannten und unbekannten Corona-Fälle.
Mit einem Antikörper-Test können Wissenschaftlerinnen und Wissenschaftler das Blut von Personen untersuchen.
Damit stellen sie fest, ob eine Person bereits Corona hatte oder nicht.
Die Forschergruppe um den Virologen Hendrik Streeck hat in der Heinsberg-Studie die Dunkelziffer in der **Gemeinde Gangelt** ermittelt. Gangelt liegt im Landkreis Heinsberg. In Gangelt gab es zu Anfang der Pandemie einen starken Anstieg an Infektionen. Der Grund dafür war ein **Karnevalsfest**. Dort haben sich viele Menschen gegenseitig angesteckt. Die Zahl der Infektionen war in Gangelt also **viel höher** als in anderen Regionen Deutschlands.

Methode der Studie

In Gangelt lebten zum Zeitpunkt der Untersuchung etwa 12.600 Einwohner. Davon waren 388 Personen offiziell als infiziert gemeldet. Die Forschergruppe konnte natürlich nicht alle Einwohner untersuchen. Daher wählte sie aus einer Einwohnerliste *per Zufall* Personen aus und schickte ihnen einen Brief.

919 Personen nahmen an der Studie teil und wurden mit dem Antikörper-Test auf Corona untersucht.

© C. Kralisch, Prof. Dr. A. Marohn: choice²reflect; Institut für Didaktik der Chemie, Münster

WWU MÜNSTER — Chemie Didaktik Münster

Ergebnisse der Heinsberg-Studie

Von den 919 untersuchten Personen waren **22 Personen** offiziell als infiziert **gemeldet**. Der Test hat ergeben, dass tatsächlich bereits **138 Personen** eine Infektion durchlaufen haben. Die Dunkelziffer der Stichprobe beträgt also 138, das sind 15 % der untersuchten Personen. Überträgt man diese 15 % auf alle Einwohner in Gangelt, ergibt sich eine Dunkelziffer von etwa **1.900 Personen**. In Gangelt sind 7 Personen an Corona gestorben. Die Sterblichkeit beträgt also 0,36 %.

Die Forschergruppe betont: Die Situation in Gangelt ist besonders, da es dort so viele Infektionen gegeben hat. Deshalb dürfen diese Ergebnisse nicht einfach auf andere Regionen übertragen werden. Würde man jedoch *rein theoretisch* die Sterblichkeit auf ganz Deutschland übertragen und damit die Dunkelziffer ausrechnen, dann käme man auf eine Dunkelziffer von 1.8 Millionen Menschen. Das bedeutet, dass von 83 Millionen Einwohnern in Deutschland bereits 1.8 Millionen Menschen mit dem Corona-Virus infiziert sein könnten. Zum **Zeitpunkt der Studie** waren nur 6.575 Menschen in Deutschland als infiziert gemeldet.

	Getestete Stichprobe	**Gemeinde Gangelt**	**Deutschland**
Anzahl an Personen	919		83 Millionen
Gemeldete Infizierte		388	
Dunkelziffer	Getestet: 138	Ausgerechnet:	Ausgerechnet: 1.8 Millionen

Stand: 06. April 2020

© C. Kralisch, Prof. Dr. A. Marohn: choice²reflect; Institut für Didaktik der Chemie, Münster

Abbildung 9.10: *Arbeitsblatt 1 zur Erarbeitung des Prüfkriteriums Randomisierung.*

Erarbeitung des Prüfkriteriums Randomisierung anhand von Prüfkarten

Die Prüfkarten wurden für den direkten Einsatz im Unterricht zur Erarbeitung des Prüfkriteriums Randomisierung entwickelt. Sie enthalten neben einem Deckblatt zwei Seiten mit Informationen zum Prüfkriterium und eine Seite mit Arbeitsaufträgen. Die Prüfkarten werden in drei Differenzierungsstufen bereitgestellt: Basis, Stern 1 und Stern 2. Die Abbildungen 9.11 und 9.12 zeigen exemplarisch die Prüfkarten in der Differenzierungsstufe Stern 2, alle weiteren entwickelten Prüfkarten sind in den Anhängen B.6.2 bis B.6.4 dargestellt.

Was ist Randomisierung?

Die erste Seite erklärt den Begriff Randomisierung anhand eines alltagsnahen Beispiels. Es wird dargelegt, dass bei Studien meist Stichproben aus der Grundgesamt-

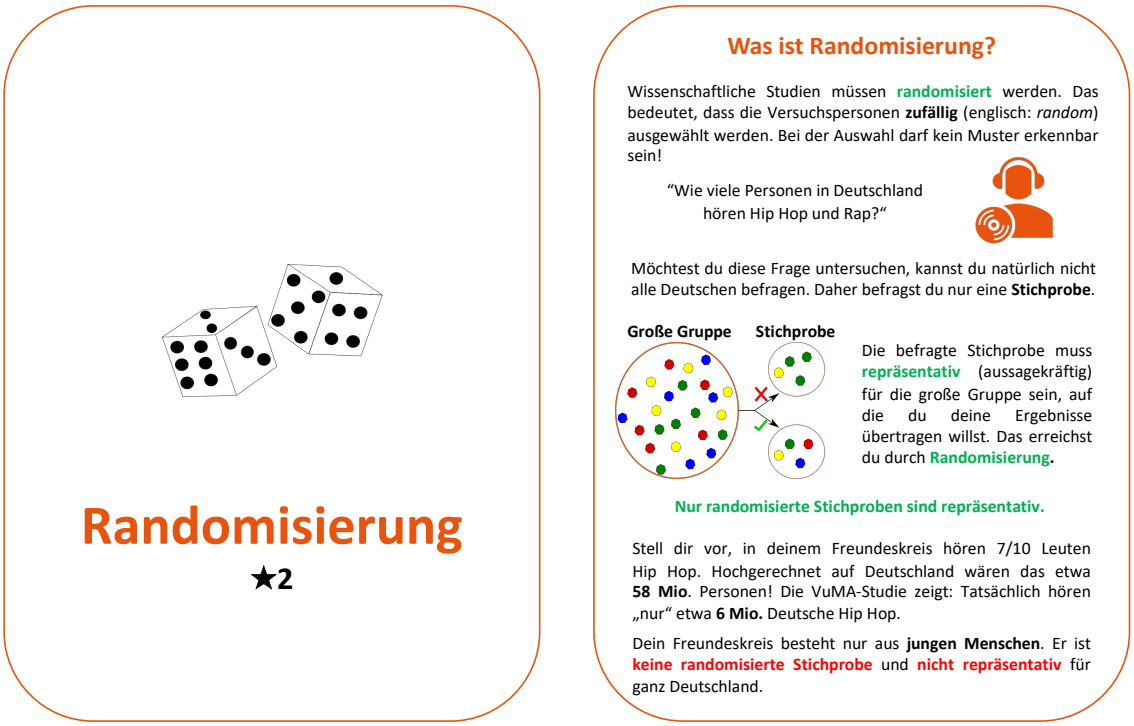


Abbildung 9.11: *Prüfkarten zur Erarbeitung des Prüfkriteriums Randomisierung in der Differenzierungsstufe Stern 2 (Seiten 1 und 2).*

heit (im Material: „große Gruppe") gezogen werden. Dies muss nach einem Zufallsverfahren geschehen, um zu gewährleisten, dass die Stichprobe repräsentativ für die Grundgesamtheit ist. Die genauen Inhalte und Formulierungen unterscheiden sich je nach Differenzierungsgrad.

Wie erreicht man Randomisierung?

Auf der zweiten Seite der Prüfkarten wird herausgestellt, welche zufallsbedingten Methoden für eine Randomisierung passend sind. Außerdem wird verdeutlicht, dass bei Stichprobenziehungen die besonderen Merkmale der Individuen beachtet werden müssen und dass eine genaue Erklärung der verwendeten Stichprobenauswahlverfahren nötig ist. Auch diese Karte bietet mit steigendem Differenzierungsgrad zusätzliche Inhalte.

Randomisierung der Heinsberg-Studie

Nachdem die Schüler:innen sich mit dem Konzept der Randomisierung vertraut ge-

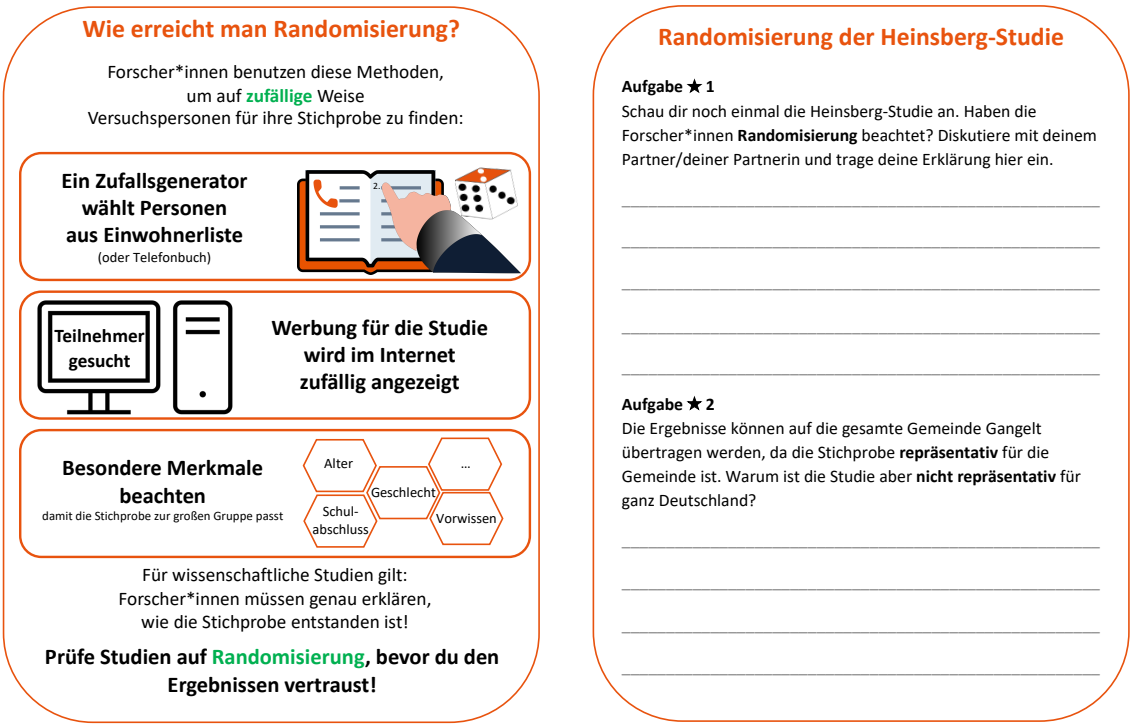


Abbildung 9.12: *Prüfkarten zur Erarbeitung des Prüfkriteriums Randomisierung in der Differenzierungsstufe Stern 2 (Seiten 3 und 4).*

macht haben, bearbeiten sie in Partnerarbeit die Aufgaben auf der letzten Seite der Prüfkarten. Die Basis-Aufgaben enthalten kurze Impulsfragen, die mit Ja oder Nein beantwortet werden, während für Stern 1 und Stern 2 auch Freitext-Felder ausgefüllt werden müssen. Allen Schwierigkeitsstufen gemeinsam ist die Bewertung der Heinsberg-Studie hinsichtlich Randomisierung. Die Lernenden geben an, ob die Forscher:innen Randomisierung beachtet haben. Die Stern 2-Aufgabe verlangt von den Lernenden eine Begründung dafür, dass die Stichprobe nicht repräsentativ für ganz Deutschland ist.

Prüfen der Heinsberg-Studie auf Randomisierung

Nachdem die Lernenden in Einzel- und Partnerarbeit die Einzelheiten zur Randomisierung erarbeitet haben, werden die Ergebnisse in einer Plenumsphase zusammengetragen. Die Lehrperson stellt dafür die Ergebnisse der Heinsberg-Studie zur Diskussion und moderiert die Redebeiträge der Schüler:innen. Dabei stehen die Fragen aus dem Arbeitsmaterial im Vordergrund: Die Schüler:innen diskutieren, ob in der

Studie korrekt randomisiert wurde und ob gültige Schlussfolgerungen unter Beachtung von Repräsentativität gezogen wurden.

In der Plenumsphase kommt zur Sprache, dass die Forschenden bei der Ziehung der Stichprobe korrekt vorgegangen sind, indem nach einem zufälligen Verfahren die Versuchspersonen rekrutiert wurden. Die Stichprobe ist also randomisiert und für Gangelt repräsentativ. Diese zentrale Erkenntnis sollte an der Tafel festgehalten werden. Dafür bieten sich auf Papier gedruckte Trichter an, die von der Lehrperson an das Prüfschema (siehe Kapitel 11) gepinnt werden und die Studie damit in die wissenschaftliche Schatzkiste einsortieren.

Bewerten von Boulevard-Presse-Artikeln über die Heinsberg-Studie

Die Autor:innen der Heinsberg-Studie haben die begrenzte Tragweite und Anwendbarkeit ihrer Ergebnisse im Blick gehabt und in der Veröffentlichung auch darauf hingewiesen. Dennoch wurde in zahlreichen Medien von dieser Einschränkung kaum Notiz genommen. Die Beispielrechnung, nach der zum Zeitpunkt der Studie bereits 1,8 Millionen Menschen in Deutschland mit dem Coronavirus infiziert gewesen sein könnten, wurde unreflektiert übernommen und für die Erstellung großer Schlagzeilen verwendet.

Derartige Schlagzeilen einordnen und bewerten zu können, ist unabdingbar. Daher sollte die Wirkung der Studienergebnisse auf die Öffentlichkeit explizit im Unterricht thematisiert werden. Dazu wird den Schüler:innen eine Schlagzeile aus der BILD-Zeitung präsentiert, die die Hochrechnung als bewiesen darstellt (siehe 9.13).

Die Lernenden nehmen Stellung zu der Schlagzeile und bewerten die Gültigkeit der vermittelten Informationen unter Rückbezug auf die zuvor diskutierte fehlende Repräsentativität der Ergebnisse für ganz Deutschland. Dabei ist es wichtig, herauszustellen, dass die Studie zwar nicht repräsentativ für die Bevölkerung Deutschlands ist, aber dass sie dennoch wissenschaftlich gültig ist. Die Schüler:innen sollen erkennen, dass die Forschenden korrekt vorgegangen sind, aber dass die Ergebnisse der Studie von den Medien fehlinterpretiert bzw. überinterpretiert wurden.

Bild

04.05.2020 - 10:48 Uhr

CORONA-KRISE

HEINSBERG-STUDIE:

Möglicherweise 1,8 Mio Corona-Infizierte in Deutschland

Prof. Dr. Hendrik Streeck leitet die Heinsberg-Studie
Foto: picture alliance / Flashpic

Heinsberg (NRW) – **In Deutschland könnten sich nach Ergebnissen der sogenannten Heinsberg-Studie mittlerweile möglicherweise 1,8 Millionen Menschen mit dem Coronavirus infiziert haben.**

Abbildung 9.13: *Schlagzeile der BILD-Zeitung am 04.05.2020 als Reaktion auf die Veröffentlichung der Heinsberg-Studie.*

9.3.3 Erprobung

Das Lernmaterial zur Erarbeitung des Prüfkriteriums Randomisierung wurde im Rahmen einer Gesamterprobung an einer Gesamtschule im September 2020 erprobt.
Die Erprobung erstreckte sich über zwei 90 minütige Doppelstunden und eine 45 minütige Unterrichtsstunde. Behandelt wurden dabei Materialien zum fachspezifischen Problemaufriss (Phase I) und Positionierung (Phase II), die zwei Prüfkriterien *Randomisierung* und *Korrelation vs. Kausalität* als Teile von Phase III sowie Elemente der Anwendung und Reflexion (Phasen IV und V). 25 Lernende der neunten Klasse nahmen an der Stunde teil. Eine Charakterisierung der Lerngruppe sowie weitere Angaben zu organisatorischen Gegebenheiten befinden sich im Kapitel 8.1.5.
Die erste Doppelstunde begann mit einem fachspezifischen Problemaufriss gemäß Phase I der Unterrichtskonzeption choice^{2}reflect und einer anschließenden Positionierung. Der genaue Ablauf des 45 minütigen Stundenabschnitts, verwendete Materialien und Erprobungsergebnisse dieser Phasen werden im Kapitel 10.3 dargestellt.
Im zweiten Teil der Doppelstunde fand die Erarbeitung des Prüfkriteriums Randomisierung am Beispiel der Heinsberg-Studie statt. Der Stundenverlaufsplan ist im Anhang B.7.1 dargestellt. Dieser Stundenabschnitt wird im Folgenden fokussiert. Weiterhin wird der erste Teil der zweiten Doppelstunde beleuchtet, in dem eine kurze Wiederholung der zentralen Inhalte stattgefunden hat.

Beschreibung der Unterrichtsstunde

Die Unterrichtsstunde begann wie geplant um 07:55 Uhr mit einer Begrüßung durch die betreuende Lehrkraft und einem Unterrichtseinstieg gemäß Phase I der Unterrichtskonzeption. Fachspezifischer Problemaufriss und anschließende Positionierung dauerten bis 08:30 Uhr und wurden somit etwa acht Minuten eher beendet als geplant. Eine genauere Analyse dieses Unterrichtsabschnitts befindet sich im Kapitel 10.3. Als zentrales Ergebnis lässt sich festhalten, dass sich zehn Lernende in der Positionierung für die Magnesium-Tablette entschieden haben und neun Lernende für die Antwort „Eine Banane am Tag“. Ein:e Schüler:in hat angegeben, sich nicht entscheiden zu können.
Anschließend verteilte die Lehrkraft das AB 2 zur Heinsberg-Studie und leitete die Lernenden zur Arbeitsphase an. Dafür wurden etwa drei Minuten beansprucht (geplant: fünf Minuten). Das Prüfschema wurde in dieser Phase nicht vorgestellt. Die

Arbeitsphase absolvierten die Lernenden entsprechend der Planung innerhalb von zehn Minuten. Es folgte eine kurze Plenumsphase mit einer Dauer von zwei Minuten zur Sicherung der Tabellenwerte, die jedoch nicht eingeplant gewesen war.
Um 08:46 Uhr wurden die Prüfkarten verteilt und der dazugehörige Arbeitsauftrag gestellt, den die Lernenden in 15 Minuten absolvierten (geplant: 17 Minuten). Im Plenum wurden anschließend die Ergebnisse gesichert und diskutiert. Diese Phase lag mit einer Dauer von 16 Minuten über dem Planungsrahmen von zehn Minuten. Ebenso dauerte die abschließende Diskussion zur Bewertung der BILD-Schlagzeile mit einer Länge von neun Minuten länger als geplant (fünf Minuten). Die Stunde wurde um 09:26 Uhr beendet.

Auswertung der bearbeiteten Arbeitsblätter

Die von den Lernenden digital ausgefüllten Arbeitsblätter wurden in Form von PDF-Dateien zentral gesammelt und zur Auswertung bereitgestellt.
Arbeitsblatt 2 enthält einen interaktiven Teil in Form einer Tabelle, in der fehlende Werte ergänzt werden müssen. Die Auswertung der Arbeitsblätter hat ergeben, dass 23 Lernende alle Werte korrekt aus dem Text übernommen haben. Zwei Lernende haben einen Wert falsch eingetragen und in einem Fall waren zwei Werte nicht korrekt.
Arbeitsblatt 3 wurde in drei Differenzierungsstufen bereitgestellt. Diese sind so ausgelegt, dass jede:r Lernende *eine* dieser Stufen auswählt und bearbeitet. Die Inhalte der Aufgaben sind ähnlich, die Bearbeitungsweise variiert jedoch zwischen der Basisstufe, Stern 1 und Stern 2.
AB 3, Basis: Zur Frage, ob die Stichprobe genau erklärt wird, kreuzten 22 Lernende *Ja* an. Auch die Frage, ob die Stichprobe zufällig ausgewählt wird, beantworteten 22 Lernende mit *Ja*. Drei Lernende gaben bei der zweiten Aufgabe an, das Ergebnis sei auf ganz Deutschland übertragbar, 18 Lernende verneinten dies und zwei Lernende bearbeiteten diese Aufgabe nicht.
Die erste Differenzierungsstufe des AB 3 (mit Stern 1 bezeichnet) enthält ebenfalls die Frage nach der genauen Erklärung der Stichprobe. Zwölf Lernende kreuzten *Ja* an, ein:e Lernende:r *Nein* und in einem Fall wurde keine Antwort angekreuzt. Die zweite Aufgabe erfordert eine Einschätzung, ob in der Studie Randomisierung beachtet wurde. Alle Lernenden stimmten zu, als Begründung wurde in fünf Fällen die Zufalls-

methode genannt, drei Mal fiel das Stichwort *repräsentativ*. Eine Antwort enthält eine elaborierte und sehr treffende Erklärung.
Die offene Frage der zweiten Differenzierungsstufe (Stern 2) wurde von allen 15 Lernenden mit *Ja* beantwortet. Davon wird in 13 Antworten das Stichwort *Zufall* genannt. Ein Mal wird erwähnt, dass die Stichprobe nur für Gangelt repräsentativ ist, eine Antwort enthält keine Begründung. Die zweite Aufgabe dieser Stufe ist ebenfalls offen gestaltet und fragt nach der Repräsentativität der Stichprobe. Hier führten die Lernenden verschiedene Aspekte an. In vier Fällen wurde genannt, dass „eine besondere Situation" in Gangelt vorlag, vier weitere Antworten bemerkten, dass „nur in Gangelt getestet wurde". Ein Mal wurde das Karnevalsfest als Grund für fehlende Repräsentativität über Gangelt hinaus erwähnt, ein:e Schüler:in beschrieb ohne weitere Begründung, die Stichprobe sei nicht repräsentativ. Weiterhin wurde genannt, dass die Stichprobe zu klein gewesen sei und dass die Forscher nur „grob geschätzt" hätten.
In der nachfolgenden Unterrichtsstunde wurde zu Beginn eine Wiederholungsaufgabe von den Lernenden bearbeitet. Dabei setzten sie sich mit folgenden Fragen auseinander: „Was ist Randomisierung?" und „Die Stichprobe in einer Studie muss repräsentativ sein. Was bedeutet das?"
Insgesamt beantworteten 16 Lernende die erste Aufgabe und 14 Lernende die zweite Aufgabe. Die Antworten der Lernenden zeigen, dass sie sich recht detailliert an die Fachbegriffe der vorherigen Unterrichtsstunde erinnern konnten. Elf Lernende nannten das Stichwort *zufällig* im Zusammenhang mit der Randomisierung. Zehn nannten *Studie* als zentralen Begriff, sieben erwähnten *Stichprobe*. Zwei Lernende führten an, dass es um die Befragung von Personen geht und zwei nannten als Beispiel den Schlüsselbegriff der *Dunkelziffer* von Infektionen in einem Ort.
In vielen Antworten wurden die Schlüsselbegriffe auch gemeinsam verwendet. Am häufigsten wurde der Begriff *zufällig* mit *Studie* kombiniert (sieben Nennungen), gefolgt von *Stichprobe* (sechs Nennungen). Auch Kombinationen von *Studie* und *Stichprobe* treten in vier Fällen auf. Anzumerken ist, dass in vier Fällen die notierte Antwort sehr ähnlich zu der Lösung ist, die in der anschließenden Plenumsphase gemeinsam erarbeitet wurde.
Die zweite Aufgabe brachte sechs Nennungen des Stichworts *aussagekräftig* als Erklärung des Begriffs *repräsentativ* hervor. Vier Lernende nannten *Stichprobe* als relevanten Begriff. In drei Fällen wurde auch hier das Wort *zufällig* zur Erklärung verwendet.

Vereinzelte Erklärungen besagten, dass eine Stichprobe *auf alle Menschen zutreffen* kann. In fünf Fällen ähnelten die notierten Antworten der im Plenum erarbeiteten Lösung.
Die transkribierten Antworten der Lernenden befinden sich in einer Tabelle im Anhang B.4.2.

Untersuchung von Lehrimpulsen in Plenumsphasen der Unterrichtsstunde

Die Plenumsphasen der Unterrichtsstunde sind dadurch gekennzeichnet, dass die Lehrkraft verbal Arbeitsaufträge formuliert, die von den Lernenden ad-hoc beantwortet werden. Häufig werden dabei die Lernenden aufgefordert, die eigenen Ergebnisse der vorangegangenen Arbeitsphase vorzustellen und mit dem Stundenthema in einen Zusammenhang zu bringen. Im folgenden Abschnitt werden die von der Lehrkraft gegebenen Lehrimpulse genauer untersucht. Dazu werden die Lehrimpulse zunächst identifiziert und mögliche Reaktionen der Lernenden auf diese Impulse antizipiert. Diese erwarteten Reaktionen werden mit den tatsächlich aufgetretenen Reaktionen verglichen. Die Passung der Schülerantworten sowie die Anschlussfähigkeit an den weiteren Unterrichtsverlauf werden bewertet, um zu einer abschließenden Einschätzung über die Güte des Lehrimpulses zu gelangen.

Darstellung der Lehrimpulse

Die Tabelle B.2.2 im Anhang zeigt eine Übersicht über alle in der Doppelstunde gegebenen Impulse der Lehrkraft, dazu antizipierte Schülerantworten sowie die tatsächlich aufgetretenen Äußerungen.
Die erste Plenumsphase zum AB 1 enthält Lehrimpulse zum fachspezifischen Problemaufriss und der Positionierung. Diese Impulse werden im Kapitel 10.3 genauer beleuchtet.
Die ersten Impulse des hier betrachteten Teils der Unterrichtsstunde, mit dem die Phase III der Unterrichtskonzeption beginnt, treten in der Plenumsphase zum AB 2 auf (ab Impuls 21k). Mit Impuls 21k fordert die Lehrperson die Schüler:innen auf, die zuvor eingetragenen Tabellenwerte im Plenum vorzulesen. Erwartet wurde, dass die Schüler:innen einzelne Zellenwerte kurz nennen und deren Bedeutung erklären. Die Reaktion einer Schülerin oder eines Schülers lautete: „Ja das erste ist 22“.
Die Lehrkraft fordert die Schülerin auf, die Bedeutung des Zellenwertes zu erklären

(Impuls 21l). Daraufhin entgegnet diese, dass von 900 untersuchten Personen nur 22 offiziell als infiziert gemeldet waren.
Nach diesem Plenumsgespräch begeben sich die Lernenden in eine weitere Arbeitsphase und widmen sich dem AB 3. In der folgenden Plenumsphase dienen die Lehrimpulse 21m, 21n und 21o dazu, die Ergebnisse des Arbeitsblattes zu vergleichen. Die Schüler:innen nennen jeweils das von ihnen angekreuzte Feld („S: Also bei der ersten Aufgabe habe ich ja angekreuzt. // S: Ja. // S: Das ist randomisiert.“).
Einen umfangreicheren Impuls stellt 21p dar. Mit der Frage: „Kann man das Ergebnis der Studie auf ganz Deutschland übertragen?“ erwartet die Lehrperson eine Begründung dafür, dass die Studienergebnisse nicht auf ganz Deutschland übertragbar sind. Die Lernenden sollten fehlende Repräsentativität anführen. Zwei Lernende geben daraufhin elaborierte Erklärungen ab.
Das AB 3 wurde in verschiedenen Differenzierungsstufen von den Lernenden bearbeitet, somit verwendete die Lehrkraft weitere Impulse, um auch die Aufgaben der Differenzierungsstufen Stern1 und Stern2 abzufragen. Der umfangreiche Impuls 21q beinhaltet kleinere Fragestellungen: Haben die Forscher:innen Randomisierung beachtet? Ist die Stichprobe repräsentativ für ganz Deutschland?
Die Lernenden sollten an dieser Stelle die zuvor erarbeitete Erkenntnis wiederholen, dass die Stichprobe der Heinsberg-Studie randomisiert ausgewählt wurde und damit repräsentativ ist, dies jedoch nur für Gangelt gilt. Die erste Frage wurde von einer Schülerin bzw. einem Schüler beantwortet mit: „S: Ja, weil es durch ein Zufallssystem ausgewählt wurde“, ein weiterer Lernender las vom Arbeitsblatt die Definition für Repräsentativität ab.
Impuls 21r erfragt den Grund dafür, dass das Ergebnis repräsentativ für Gangelt, jedoch nicht für Deutschland ist. Die Antwort eines Lernenden lautete: „S: Weil es ja nur in Gangelt also befragt wurde“.
Impuls 21s („Können wir der Studie trauen, ist sie gut gemacht [...]?“) erfordert eine Stellungnahme der Lernenden. Ein:e Schüler:in reagierte mit der Aussage: „S: Kann man trauen. Weil [die Stichprobe] durch ein Zufalls ausgewählt wurde“.
Die Intention von 21t bestand in der Abfrage, ob eine:r der Schüler:innen bereits Kenntnis von der Heinsberg-Studie hatte. Kein:e Lernende:r antwortete.
In der letzten Phase der Unterrichtsstunde (Anwendung) stellte die Lehrkraft einige anspruchsvolle Bewertungs-Impulse. Schüler:innen sollten die Schlagzeile der BILD-Zeitung bewerten und dabei herausstellen, dass diese Schlagzeile Fehlinterpretatio-

nen enthält (Impuls 21u). Ein:e Lernende:r erklärte daraufhin, dass es sich bei der Angabe „1,8 Mio. Infizierte“ um die Dunkelziffer handelt.
Auf die Frage, ob die BILD-Zeitung alles richtig gemacht habe (Impuls 21v), gaben die Lernenden keine Antwort.
Impuls 21w erfragt die Einschränkung, die beachtet werden muss, wenn die Studie zur Entscheidungsfindung verwendet werden soll. Eine Schülerin antwortete: „S: Also nur unter einer Einschränkung, dass das nur für die Gemeinde Gangelt genommen wird [...]“.
Anschließend fragte die Lehrkraft nach der Definition des Begriffs *Dunkelziffer* (Impuls 21x), die von eine:r Lernende:n geliefert wurde.
Der letzte Lehrimpuls der Stunde, 21y, erfragt erneut, warum das Studienergebnis nur für Gangelt und nicht die gesamte deutsche Bevölkerung gilt. In einer Reaktion wurde behauptet, es wurden 50 % der Bevölkerung in Gangelt getestet und ein Schüler betonte erneut, dass die Hochrechnung in der Studie von „rein theoretischer“ Natur ist.

Anschlussfähigkeit und Passung der Reaktionen auf Lehrimpulse

Die Lehrimpulse wurden nach dem Muster 21a bis 21y benannt. Die Reaktionen der Lernenden auf die jeweiligen Impulse werden ebenfalls durchnummeriert, dazu wird eine römische Zahl an das Kürzel des Impulses angehängt. Beispiel: *21m-II* bezeichnet die zweite Reaktion auf den Impuls 21m.
Für jede Impuls-Reaktion lässt sich in einem hoch-inferenten Codierungs- bzw. Rating-Verfahren feststellen, wie gut die Reaktion zur zuvor antizipierten Schülerantwort passt. Für die Passung werden die Werte 1 (schlecht), 2 (wenig passend), 3 (eher passend) und 4 (passgenau) vergeben.
Zur Charakterisierung der Impuls-Reaktionen wird außerdem die Anschlussfähigkeit der Reaktionen ermittelt. Damit wird zum Ausdruck gebracht, welchen Wert die Reaktion für den weiteren Unterrichtsverlauf besitzt. Es werden die Werte 1 (nicht anschlussfähig), 2 (wenig anschlussfähig), 3 (eher anschlussfähig) und 4 (sehr anschlussfähig) vergeben.
Die Impuls-Reaktion 21k-I ist nur wenig passend und wenig anschlussfähig. Sie entspricht nicht der Intention, dass die Lernenden die Bedeutung der Werte erklären. Stattdessen wird lediglich der Zahlenwert genannt. Erst durch gezielte Nachfrage der Lehrkraft wird eine passende und anschlussfähige Reaktion (21l-I) erreicht.

In der späteren Plenumsphase zu AB 3 treten fünf passende Impuls-Reaktionen auf. Mit 21m-I und 21n-I geben die Lernenden die von ihnen angekreuzten Lösungen wieder, diese Antworten sind wenig anschlussfähig. Mit 21o-I wird die anschlussfähige Schlussfolgerung, dass die Studie randomisiert ist, gezogen. Beide Reaktionen auf den Impuls 21p haben sich als passend und anschlussfähig herausgestellt, denn die Schüler:innen liefern stimmige Erklärungen zur Repräsentativität der Stichprobe.
Im weiteren Verlauf der Plenumsphase werden auch die Arbeitsergebnisse zur höheren Differenzierungsstufe von AB 3 verglichen. Der Impuls 21q bringt eine passende und anschlussfähige, aber auch eine unpassende, nicht anschlussfähige Reaktion hervor.
21r-I ist eine passende und anschlussfähige Reaktion mit der Erklärung, warum die Stichprobe nicht repräsentativ für ganz Deutschland ist.
Die Reaktion 21s-I ist eher passend, aber nur wenig anschlussfähig. Es wird passenderweise erkannt, dass die Studie vertrauenswürdig ist, allerdings wird als Begründung erneut die Zufallsauswahl genannt. Es wird nicht über Repräsentativitäten gesprochen.
Auf die Impulse 21t und 21v wird nicht reagiert, die eintretende Stille wird als unpassend und nicht anschlussfähig codiert.
Die Reaktionen 21u-I und 21u-II beziehen sich nicht auf die von der Lehrkraft formulierte Fragestellung und sind daher unpassend und nicht anschlussfähig.
Mit 21w-I und 21x-I geben die Lernenden elaborierte Erklärungen ab, die passend und anschlussfähig sind.
Der letzte Lehrimpuls der Unterrichtsstunde (21y) wurde von der Lehrkraft zur Sicherung des Begriffs Repräsentativität gegeben. Die Schüler:innen antworteten jedoch nur unpassend und wenig anschlussfähig.

Kategorisierung der Lehrimpulse

Analog zu der im Kapitel 9.2.3 beschriebenen Prozessierung der Forschungsdaten werden die Impuls-Reaktionen im folgenden Abschnitt in vier Kategorien unterteilt. Diese spiegeln die möglichen Kombinationen der Parameter Passung und Anschlussfähigkeit wider. In die erste Kategorie fallen alle Impuls-Reaktionen mit hoher Passung und hoher Anschlussfähigkeit. Kategorie 2 beschreibt passgenaue Reaktionen mit geringer Anschlussfähigkeit. In Kategorie 3 werden unpassende aber anschlussfähige Reaktionen codiert, während Kategorie 4 Reaktionen enthält, die sich weder

als passend noch als anschlussfähig erwiesen haben.
Folgende Impuls-Reaktionen werden der Kategorie 1 zugeordnet: 21l-I, 21o-I, 21p-I, 21p-II, 21q-I, 21r-I, 21w-I und 21x-I.
In Kategorie 2 fallen die Reaktionen: 21m-I, 21n-I und 21s-I.
Der dritten Kategorie zugehörig ist 21y-II.
Die übrigen Impuls-Reaktionen (21k-I, 21q-II, 21t, 21u-I, 21u-II, 21v, 21y-I) waren weder passend noch anschlussfähig und sind damit Kategorie 4 zuzuordnen.

Untersuchung von Gruppenarbeitsphasen: *Aha-Erlebnisse* und Verständnisschwierigkeiten

Neben der Analyse der Lehrimpulse und Schüleräußerungen in den Plenumsphasen ist es auch von Bedeutung, Gespräche zwischen Lernenden innerhalb von Gruppenarbeitsphasen zu erfassen und auf Besonderheiten zu untersuchen.
Wie bereits im Kapitel 9.2.3 zur Erforschung des Prüfkriteriums *Korrelation und Kausalität* geschehen, werden auch bei der Betrachtung dieser Unterrichtsstunde zum Prüfkriterium *Randomisierung* die folgenden drei Schwerpunkte bezüglich der Gruppenarbeitsphasen betrachtet:

1. Anzeichen auf *Aha-Erlebnisse* in den Stillarbeitsphasen der Unterrichtsstunde
2. Anzeichen auf Verständnisschwierigkeiten in den Gesprächen der Lernenden
3. Erkennbare Off-Task-Handlungen der Lernenden

In einer zunächst deduktiven und anschließend induktiven Vorgehensweise werden die Forschungsdaten einer qualitativen Inhaltsanalyse nach Kuckartz unterzogen, Näheres dazu befindet sich im Kapitel 8.2.
Der verwendete Codierleitfaden entspricht dem im Kapitel 9.2.3 beschriebenen Leitfaden und ist in Tabelle 9.1 dargestellt. Lediglich die letzte Kategorie (28-3.1) wurde in dieser Analyse nicht verwendet. Der Leitfaden enthält die Kategorien *Aha-Erlebnisse* und *Verständnisschwierigkeiten* mit den induktiv ermittelten Unterkategorien *Technik*, *Fachinhalt* und *Aufgabenstellung*. Außerdem gibt es die Kategorie *Off-Task* mit den deduktiv bestimmten Unterkategorien *Verlassen des Sitzplatzes*, *Allgemeine Interaktion*, *Gespräch* und *Tablet-Interaktion*.

Das Datenmaterial wurde durch zwei Personen codiert. Nach diskursiver gemeinsamer Überprüfung nicht-übereinstimmender Codes wurde eine Intercoder-Übereinstimmung von 92% erreicht. Die Unterschiede in den Codierungen der beiden Codierenden bestehen zumeist in den Zuteilungen zu den Codes *Aufgabenstellung* und *Fachinhalt* bei den Verständnisschwierigkeiten. Häufig wurden dieselben Stellen codiert, jedoch von einer Person dem Code *Aufgabenstellung* zugeordnet und von der anderen Person dem Code *Fachinhalt*. Weiterhin wurden einige *Aha-Erlebnisse* von der zweiten codierenden Person nicht gefunden.

Ergebnisdarstellung nach Gruppe

Auf der Basis des Codierleitfadens wurden die fünf Fokusgruppenvideos in einem niedrig-inferenten Event-Sampling-Verfahren auf das Auftreten der Kategorien *Verständnisschwierigkeiten*, *Aha-Erlebnisse* und *Off-Task Handlungen* mit ihren Sub-Kategorien untersucht. Zur Visualisierung der Ergebnisse wurden *Codelines* erstellt, diese Darstellungsform ist angelehnt an das Visualisierungs-Werkzeug in der Software MAXQDA. Die Codelines befinden sich im Anhang dieser Arbeit (B.3.3). In dieser Analyse werden die Codelines ab Minute 46 betrachtet, da im ersten Teil der Doppelstunde die Phasen I und II der Unterrichtkonzeption durchlaufen wurden. Diese werden an anderer Stelle (Kapitel 10.3) betrachtet.

Gruppe 1 zeigt in der Arbeitsphase zu AB 3 Schwierigkeiten bei dem Verständnis der Fachinhalte (Min. 64-69) und der Aufgabenstellung (Min. 69). Dabei fragt S1-07: „Machst du Stern2?“ Daraufhin antwortet S1-06: „Nein, das ist zu schwer. Also, das ist nicht an sich so schwer, aber irgendwie verstehe ich den Zusammenhang nicht“.

Ein *Aha-Erlebnis* tritt in derselben Arbeitsphase in der Minute 65 auf. S1-03 fragt die anderen Lernenden, was der Begriff *repräsentativ* bedeutet. S1-04 zieht einen Vergleich zu dem Wort *Repräsentation*, woraufhin das Aha-Erlebnis bei bei S1-03 eintritt: „Achso, also wenn du das wiedergeben kannst quasi“.

„Off-Task“-Handlungen sind in der Gruppe 1 kaum erkennbar. Lediglich zum Ende der Arbeitsphase zu AB 2, nachdem die Lernenden ihre Arbeit in der Gruppe abgeschlossen hatten, führten sie Off-Task-Gespräche (Min. 54-57).

In Gruppe 2 treten vier Verständnisschwierigkeiten auf. Ein technisches Problem zu Beginn der Arbeitsphase AB 3 (Min. 61), sowie fachinhaltliche Probleme treten um Minute 66 und 69-71 auf. Ein Lernender äußert, er wisse nicht, „was er dazu schreiben soll“ (20-09-21_Gruppe2, Pos. 55). Außerdem ist in der Minute 66 den Lernenden

die Aufgabenstellung unklar.
Ein *Aha-Erlebnis* ist zum Ende der Stunde in der letzten Plenumsphase erkennbar, dieses ist jedoch als recht schwach einzustufen. Ein:e Schüler:in erkennt, dass die Informationen über die Übertragbarkeit dem Arbeitsblatt zu entnehmen sind (S2-02 flüstert: „Das steht da doch schon!", Min. 85).
„Off-Task"-Handlungen treten in der Gruppe, bis auf eine kurze Interaktion in der Minute 65, nicht auf.
Die Codeline zu Gruppe 3 offenbart, dass es in der Arbeitsphase zu AB 3 einige Verständnisschwierigkeiten in den Bereichen Technik (Min. 63-65, 68), Fachinhalt (Min. 64-65) und Aufgabenstellung (Min. 70) gab.
Zwei *Aha-Erlebnis*-Situationen sind zum Ende der Stunde in den Minuten 79 und 90 erkennbar. S1-06 erkennt einen Fehler, den sie/er zuvor begangen hat: „Ich hab da/ ich hab das falsch gemacht, weil ich es nicht richtig verstanden hab. Ach, auf die Studie! Ups". Außerdem berichtet S1-06, dass man der Instagram-Studie nicht trauen kann.
Eine längere Phase, in der die Lernenden sich *off-Task* unterhalten und mit dem Tablet interagieren, ist in der zweiten Hälfte der Arbeitsphase zu AB 2 zu erkennen (Min. 52-58).
Gruppe 4 hat in der Arbeitsphase zu AB 3 einige technische Schwierigkeiten (Min. 62-63, 67) und Probleme mit dem Fachinhalt (Min. 69-72). Bei erstmaligem Öffnen des AB3, Stern1 reagiert S2-05 mit: „Hä, das hatten wir doch gerade", während S2-06 durch den Ausruf: „Ich bin überfordert" zu verstehen gibt, dass der Fachinhalt nicht verstanden wurde. Später äußert S2-05 während sie sich den Arbeitsauftrag auf AB 3 durchliest: „Worüber sollen wir jetzt diskutieren?" (20-09-21_Gruppe 4, Pos. 81).
Auffällig ist, dass in dieser Gruppe drei *Aha-Erlebnisse* auftreten. In der Plenumsphase zu AB 2 erkennt S2-08 einen Fehler im Gruppenergebnis und korrigiert diesen (Min. 59). In der darauffolgenden Arbeitsphase unterhalten sich die Lernenden über die Übertragbarkeit der Studienergebnisse auf ganz Deutschland. In der Diskussion ist erkennbar, dass ein *Aha-Erlebnis* durchlaufen wird (Min. 64-67). In der Minute 73 flüstert S2-08: „Ach, die haben das ja" und bezieht sich damit auf die Randomisierung der Stichprobe.
In der Arbeitsphase zu AB 2 führen die Lernenden Off-Task-Interaktionen mit dem Tablet durch (Min 52-56), weitere *Off-Task*-Handlungen wurden nicht erkannt.
In den Videos der Gruppe 5 fällt auf, dass es nahezu keine Verständnisschwierigkei-

ten gibt; lediglich in der Minute 63 äußert S2-11 flüsternd: „Ich verstehe das nicht“.
Aha-Erlebnisse sind an drei Stellen zu beobachten. In Minute 49 korrigiert S2-11 einen Tabelleneintrag, begleitet von einem nahezu euphorischen Ausruf, ebenso in Minute 52. In der Arbeitsphase AB 3 ist erkennbar, dass eine Schülerin nach längerem Überlegen ihre zuvor eingetragene Lösung entfernt und durch eine fachlich richtigere ersetzt (Min. 66-68).
Die Gruppe 5 führt öfter Off-Task-Handlungen aus, dabei vor allem Tablet-Interaktionen (Min. 52-54, 69-70, 75-76) und allgemeine Interaktionen (Min. 63-64, 85-56).

Ergebnisdarstellung nach Kategorie

Die zuvor beschriebenen Codelines geben für jede Fokusgruppe einen Überblick über auftretende Verständnisschwierigkeiten, Aha-Erlebnisse und Off-Task-Handlungen. Besser vergleichbar werden diese Ergebnisse jedoch mithilfe einer anders fokussierten Visualisierung: Die Trennung der Codelines nach Kategorien (nicht nach Gruppen) wird im folgenden Abschnitt beschrieben. Die zugehörigen Codelines befinden sich im Anhang B.3.4.
Der Blick auf die Codeline zur Kategorie *Verständnisschwierigkeiten* zeigt deutlich, dass bis auf eine Situation, in der die Aufgabenstellung in der Plenumsphase AB 3 unklar war, alle Schwierigkeiten in der Arbeitsphase zu AB 3 auftraten. In insgesamt sechs Situationen hatten die Schüler:innen fachinhaltliche Schwierigkeiten. In fünf Situationen traten technische Probleme auf und an vier Stellen konnte festgestellt werden, dass Lernende die Aufgabenstellung nur unzureichend verstanden.
Eine Visualisierung aller codierter *Aha-Erlebnisse* stellt die mittlere Codeline in B.3.4 dar. Dort ist erkennbar, dass in jeder Gruppe mindestens ein *Aha-Erlebnis* gefunden wurde. Diese liegen teilweise bereits in der AP zum AB 2, größtenteils jedoch später in der AP zum AB 3 oder auch in der darauffolgenden Plenumsphase. Ein Blick auf einzelne Situationen zeigt, dass die Arten der Aha-Erlebnisse sich stark unterscheiden. Zum Teil entdecken Lernende, dass sie in einer Arbeitsphase einen fehlerhaften Tabellenwert notiert haben (Gruppe 5, Minute 49), in manchen Fällen wird deutlich, dass ein:e Lernende:r plötzlich ein vertieftes Verständnis eines Sachverhalts erlangt, bspw. S1-06 (Gruppe 3) in der Minute 79. Insgesamt treten zehn Situationen auf, in denen ein *Aha-Erlebnis* codiert wurde.
Die rechtsstehende Codeline im Anhang B.3.4 zum Auftreten von *Off-Task*-Handlungen in der Unterrichtsstunde zeigt ein recht einheitliches Bild. In fast allen Gruppen

Tabelle 9.3: *Darstellung der mit Aha-Erlebnis codierten Segmente aus der Unterrichtsstunde zum Prüfkriterium Randomisierung.*

Quelle	Zeitmarke	codiertes Segment
21-gruppe5, Pos. 26	0:30:48	S2-11: Ha. Ah ja, ich weiß wie!
21-gruppe5, Pos. 46	0:48:32	S2-11: Ah ne, warte! 138 nicht 388. (korrigiert Eintrag in der Tabelle, korrigiert später erneut zu 22)
21-gruppe5, Pos. 52	0:51:11	S2-11 (flüstert): Oh Junge, ich bin so behindert, ja. (Korrigiert Tabelleneintrag)
21-gruppe5, Pos. 75	1:06:00	S2-12 radiert und notiert neue Lösung: Weil in anderen Bundesländern weniger Infektionen sind.
21-gruppe4, Pos. 46	0:58:53	S2-08: Oh, bei uns ist etwas falsch. (S2-05 bis S2-08 korrigieren ihre Eintragungen)
21-gruppe4, Pos. 62-64	1:04:00	S2-07: warte mal, ähm. War das zufällig? // S1-08 (liest Arbeitsblatt, stöhnt etwas): Haben die nicht einfach irgendwelche genommen? Zufällig? Ja. //S2-07: Kann man das Ergebnis der Studie auf ganz Deutschland übertragen? Nein, weil das nur/ obwohl doch. Haben die ja da geschrieben. Doch kann man.
21-gruppe3, Pos. 72	1:14:19	S1-06: Ich hab da/ ich hab das falsch gemacht, weil ich es nicht richtig verstanden hab. Ach, auf die Studie! Ups.
21-gruppe2, Pos. 68	1:24:29	S2-02 flüstert: Das steht da doch schon!
21-gruppe1, Pos. 62	1:04:05	S1-03 (überrascht): Also wenn du das wiedergeben kannst quasi? Oder wie?

sind zum Ende der Arbeitsphase AB 2 Off-Task-Gespräche und Tablet-Interaktionen erkennbar (ab Minute 52). Diese reichen teilweise bis in die nachfolgende Plenumsphase hinein. Während der Arbeits- und auch der Plenumsphase AB 3 finden kaum Off-Task-Handlungen statt. Lediglich Gruppe 5 zeigt an manchen Stellen Interaktionen mit dem Tablet (Min. 68-69, 75-76) und allgemeine Interaktionen (Min. 63-64, 85-86), die nicht dem Voranbringen des Unterrichtsgeschehens zuträglich sind.

Besondere Aussagen von Lernenden

In einigen Situationen tätigen Schüler:innen Aussagen, die sich direkt auf die Arbeitsaufgabe beziehen und Einblicke in die Denkweise der Schülerinnen und Schüler geben. Nicht alle fachbezogenen Aussagen der Lernenden können von dem zuvor erarbeiteten Kategoriensystem erfasst und analysiert werden, daher werden in diesem Kapitel Unterrichtssituationen vorgestellt, die eine besondere forschungsbezogene Relevanz besitzen. Die Selektion der Aussagen erfolgte anhand der subjektiven Einschätzung des Codierenden. Die Tabelle 9.4 zeigt vier selektierte *besondere Aussagen*. Die Aussage *21-bes-1* zeigt, dass eine Schülerin das Konzept der Repräsentativität anhand eines eigenen Beispiels begründet. Sie erklärt ihrer Mitschülerin, dass bei einer Stichprobe nur mit Leuten, die in Berlin wohnen, keine Rückschlüsse auf Personen in Hamburg gezogen werden können. Die Erklärung wird fortgeführt (*21-bes-2*) unter Einbezug des Beispiels aus den Prüfkarten.
Der kurze Ausruf *21-bes-3* in der letzten Plenumsphase zeigt durch die spontane, emotional geladene Reaktion, dass der Schüler das Medium *BILD-Zeitung* als nicht vertrauenswürdig einstuft.
In der nachfolgenden Stunde versuchten die Lernenden, sich an die Begriffe *Randomisierung* und *repräsentativ* zu erinnern. Dabei umschrieb S1-07 in eigenen Worten die Details, an die sie sich erinnern konnte (*21-bes-4*).

9.3.4 Auswertung

Das Lernmaterial zum Prüfkriterium *Randomisierung* wurde in einer 45-minütigen Stunde als Teil einer Gesamterprobung im September 2020 erprobt. Der Unterricht wurde videographiert, anschließend transkribiert und qualitativ-inhaltsanalytisch ausgewertet. Dabei wurden Lehrimpulse in den Plenumsphasen betrachtet und das Lernendenverhalten in den Gruppenarbeitsphasen analysiert.

Tabelle 9.4: *Darstellung der als besondere Aussagen eingestuften Unterrichtssituationen in der Erprobung des Prüfkriteriums Randomisierung.*

Kürzel	Aussage	Quelle
21-bes-1	S1-07: „Also ich habe Nein angekreuzt weil: Deutschland ist sehr groß. Und wenn jetzt zum Beispiel nur Leute stichprobenartig aus Berlin irgendwie nachgefragt werden, kann man es ja nicht auf ganz Deutschland beziehen. Man weiß ja nicht, ob die aus Hamburg das genauso sehen.“	20-09-21_Gruppe 3, Pos. 60
21-bes-2	S1-07: „Aber guck mal, hier steht. Aussagekräftige Stichprobe. Hm, ich glaube trotzdem nicht. Weil das wäre eigentlich voll doof. Wenn in Berlin die Leute HipHop hören, weil das stand da ja glaub ich. HipHop und Rap hören, kann man ja nicht sagen, dass alle das in Deutschland hören!“	20-09-21_Gruppe 3, Pos. 64
21-bes-3	S1-03: „Oah, der Bild kann man aber nicht vertrauen!“	20-09-21_Gruppe 1, Pos. 74
21-bes-4	1-07: „Die haben doch dieses Leute auszuwählen aus bestimmten z.B. einer Stadt rausgeholt und solche abfragen gemacht oder wie das war“.	20-09-28_Gruppe 3, Pos. 14

Im folgenden Abschnitt werden die Ergebnisse zusammengetragen, um Aussagen über das Erreichen der zuvor definierten Lernziele treffen zu können.

Lernziel 1

„Die Schüler:innen beschreiben, dass Vollerhebungen bei wissenschaftlichen Studien nicht praktikabel sind und stattdessen Stichproben aus der Grundgesamtheit gezogen werden“.

Die Inhalte dieses Lernziels werden durch die Arbeitsblätter AB 2 und AB 3 vermittelt. In beiden Texten ist enthalten, dass „nicht alle“ Einwohner Gangelts (AB 2) bzw. Deutschlands (AB 3, alle Differenzierungsstufen) befragt werden können und dass deshalb auf Stichprobenziehungen zurückgegriffen wird. Im Verlauf des Unterrichts bearbeiteten die Lernenden jedoch keine Aufgaben, die zum Explizieren dieses Sachverhaltes aufforderten. Daher ist es schwierig, das Erreichen des ersten Lernziels zu überprüfen. In einigen Formulierungen der notierten Antworten verwenden die

Lernenden die Begriffe *aussagekräftige Stichprobe* oder sie umschreiben, dass keine Vollerhebung an der Grundgesamtheit durchgeführt wurde („Man kann das Ergebnis nicht auf ganz Deutschland übertragen, da nur eine Stichprobe gemacht wurde [...]", CH1-07 - AB3_Basis_Seite_4). Auf dieser Grundlage wird vermutet, dass die Schüler:innen das Konzept einer Stichprobenziehung aus der Grundgesamtheit im Sinne des Lernziels 1 mithilfe des Lernmaterials zielführend erarbeiten konnten. Detailliertere Aussagen über ein tiefergehendes Verständnis können an dieser Stelle nicht getroffen werden.

Lernziel 2

„Die Schüler:innen erklären das Prinzip der Randomisierung als zufallsbedingte Stichprobenauswahl."

Das zweite Lernziel dient der Begriffsbildung. Die Schüler:innen erlernen den Begriff *Randomisierung* als Anforderung an die Stichprobenauswahl bei empirischen Studien. Diese muss nach einem zufälligen Verfahren erfolgen. Die Unterrichtsstunde wurde so geplant, dass dieser Aspekt immer wieder thematisiert wird - sowohl auf den Arbeitsblättern AB 2 und AB 3 als auch in Gruppenarbeitsphasen und Plenumsgesprächen. In den Antworten der Lernenden ist zu erkennen, dass die Begriffsbildung funktioniert. Immer wieder notieren sie das Stichwort *Zufall* im Zusammenhang mit der Stichprobenauswahl, wie im Abschnitt 9.3.3 beschrieben wurde.

Die Ergebnisse der Wiederholungsaufgabe in der nachfolgenden Stunde geben weitere Hinweise darauf, dass die Lernenden das Konzept der Randomisierung verstanden haben, denn im größten Teil der Antworten sind die relevanten Schlüsselbegriffe enthalten. Einschränkend zu erwähnen ist jedoch die Nähe mancher Antworten zu der im Plenum gemeinsam erarbeiteten Lösung. Vermutlich haben einige Lernende in der Sicherungsphase ihre eigene Antwort gelöscht und durch die Plenumsantwort ersetzt. Dadurch lässt sich nicht mehr nachvollziehen, wie die zuvor durch die Lernenden erarbeiteten Ergebnisse gestaltet waren.

Lernziel 3

„Die Schüler:innen beschreiben die Randomisierung einer Stichprobe als Voraussetzung für dessen Repräsentativität"

Die Prüfkarten (AB 3) beschreiben genau, dass in einer empirischen Studie die Versuchspersonen randomisiert ausgewählt werden müssen und dass nur in diesem Fall eine Repräsentativität für die „große Gruppe", also die Grundgesamtheit, gewährleistet ist. Weiterhin wird der Begriff *repräsentativ* mit *aussagekräftig* gleichgesetzt. Die zentrale Aussage: „Nur randomisierte Stichproben sind repräsentativ" wird grün hervorgehoben. Durch ein alltagsnahes Beispiel werden diese Inhalte verdeutlicht.
Hinweise auf Erreichen des dritten Lernziels lassen sich in den *besonderen Aussagen der Lernenden* (siehe Tabelle 9.4) finden. Einige Lernende übertragen ihre Erkenntnisse aus der Erarbeitung des Prüfkriteriums auf ein selbst gewähltes Beispiel und erklären sich damit das Konzept der Repräsentativität von Stichproben. Der auf den Prüfkarten äquivalent verwendete Begriff *aussagekräftig* wird jedoch von den Schüler:innen bevorzugt. Die Aussage *21-bes-4* deckt Probleme bei der Verbalisierung der Fachinhalte auf.
Ob das Lernziel bei dem überwiegenden Teil der Klasse erreicht wurde, ist nicht klar zu beantworten. Im Unterricht wurden die relevanten Inhalte thematisiert und die Arbeitsaufträge forderten die Lernenden auf, sich mit den Randomisierungsverfahren bei der Stichprobenauswahl vertraut zu machen. Es konnten jedoch nur wenige Situationen identifiziert werden, die Rückschlüsse auf dieses Lernziel erlauben.

Lernziel 4

„Die Schüler:innen prüfen die Stichprobe der Heinsberg-Studie auf Randomisierung und bewerten ihre Repräsentativität"
Ein zentrales Stundenziel bestand in der Bewertung der Heinsberg-Studie im Hinblick auf die Auswahl der Versuchspersonen. Diese Bewertung wurde explizit im Unterricht eingefordert. Sowohl die Aufgaben der Arbeitsblätter als auch die in Plenumsphasen gegebenen Lehrimpulse fokussierten darauf. Wie im Abschnitt 9.3.3 detailliert beschrieben wurde, haben die meisten Lernenden ihre Bewertungen auf den Arbeitsblättern notiert. Häufig fallen dabei die Begriffe *zufällig* und *repräsentativ*, teilweise liefern die Lernenden auch elaborierte Erklärungen als Antwort auf die Frage nach der Repräsentativität der Stichprobe. In der zweiten Differenzierungsstufe erkannten die Lernenden, dass in Gangelt eine „besondere Situation" vorlag und die Repräsentativität der Stichprobe daher nicht über Gangelt hinaus gegeben ist.
Das vierte Lernziel wurde erreicht.

Kognitive Konflikte

Die Förderung von Bewertungskompetenz wird als geeigneter Weg zu einem Konzeptwechsel angesehen (*conceptual change*, siehe Kapitel 3.5). Das Konzept der Randomisierung stellt ein Puzzle-Stück der Unterrichtskonzeption choice^{2}reflect dar und liefert einen Beitrag zur Förderung von Bewertungskompetenz. Eine wichtige Voraussetzung für einsetzende Konzeptwechsel ist das Auftreten eines kognitiven Konflikts bei den Lernenden. Dieser tritt ein, wenn die Schülervorstellung mit der wissenschaftlichen Sichtweise in Konflikt steht.

Das Lernmaterial zur Erarbeitung von Randomisierung ist auf die Entstehung kognitiver Konflikte ausgerichtet. Anhand eines aktuellen Kontextes wird eine im Alltag häufig auftretende Situation dargestellt: In den Medien wird eine Studie zitiert, die anhand einer Bürgerbefragung eine bestimmte Fragestellung untersucht hat. Vermutlich stellen sich Lernende jedoch nicht die Frage, wie genau die befragten Personen ausgewählt wurden oder es existieren intuitiv-generierte Vorstellungen über die Stichprobenauswahl. Anhand des Lernmaterials wird verdeutlicht, dass Vollerhebungen bei wissenschaftlichen Studien nur selten sinnvoll oder praktikabel sind und dass stattdessen ein randomisiertes Auswahlverfahren angewendet werden muss. Diese Feststellung kann eine Unzufriedenheit bei den Lernenden hervorrufen und sollte zu einem kognitiven Konflikt führen.

Kognitive Konflikte wurden im Rahmen der Analysen nicht direkt erfasst. Bei den Lernenden auftretende Irritationen, die sich in Form von *Aha-Erlebnissen* ausdrücken, werden als Hinweise für besondere kognitive Prozesse aufgefasst. Möglicherweise durchliefen die Lernenden einen kognitiven Konflikt bei der Bearbeitung des Lernmaterials, der sie zu erstaunten Ausrufen bewog. Die *Aha-Erlebnisse* können jedoch auch Ausdruck anderweitiger Irritationen sein, z. B. des plötzlichen Verstehens eines Aufgabenteils.

Bei der Codierung der Unterrichtsvideos wurden zehn Situationen identifiziert, in denen ein *Aha-Erlebnis* codiert wurde (siehe Kapitel 9.3.3). In den meisten Fällen ist erkennbar, dass die Lernenden nach einem plötzlichen Einfall ihre bis dahin erarbeiteten Lösungen streichen, überdenken oder ersetzen. Oft wird dieser Einfall durch einen Ausruf begleitet in der Art: „[...] bei uns ist etwas falsch!“ oder: „Ich hab das falsch gemacht!“.

Die codierten Aussagen zeigen, dass bei den Lernenden Konzeptwechsel im Bezug

auf die fachliche Vorstellung randomisierter Stichproben stattgefunden haben könnten. Eindeutig belegbar sind diese Konzeptwechsel nicht, aber aufgrund der Häufung der *Aha-Erlebnisse* in der Gruppenarbeitsphase zum AB 3 um Minute 65 wird auf das Auftreten kognitiver Konflikte geschlossen.

9.3.5 Re-Design

Entwicklung von Leitimpulsen

Die Plenumsphasen der analysierten Unterrichtsstunde enthalten 15 Lehrimpulse mit insgesamt 19 Impuls-Reaktionen. Es wurde eine recht hohe Dichte an durch die Lehrkraft initiierten Lehrimpulsen festgestellt. Weiterhin hat sich das Unterrichtsgespräch entlang der Impulse zu einem Frage-Antwort-Gespräch entwickelt, der Unterricht verlief fragend-entwickelnd. Wünschenswert wäre es jedoch, wenn die Schüler:innen auch in den Plenumsphasen ein hohes Maß an Eigenaktivität zeigen und sich selbständig in die Diskussionen einbringen. Eine Möglichkeit, dies zu erreichen, stellt die Formulierung gezielter Leitimpulse dar. Im folgenden Abschnitt wird die zuvor erläuterte Kategorisierung der Impuls-Reaktionen als Grundlage verwendet, um Impulse zu identifizieren, die bereits zielführend in den Unterricht integriert wurden.
Schließlich werden für jede Plenumsphase maximal zwei Leitimpulse formuliert, die zum Anfang einer Diskussion von der Lehrkraft ins Plenum gegeben werden. Diese sollten Operatoren-gestützt sein und leicht öffnende Elemente enthalten.
Von 19 Impuls-Reaktionen wurden acht Reaktionen der Kategorie 1 zugeordnet und somit als passgenau und anschlussfähig deklariert. Drei Reaktionen sind passend, aber nicht anschlussfähig (Kategorie 2). Der dritten Kategorie zugehörig ist eine Impuls-Reaktion, während sieben Reaktion in die Kategorie 4 fallen und damit weder passend noch anschlussfähig sind.

Leitimpulse in der Sicherungsphase zum AB 2

In der Sicherungsphase zum AB 2 steht das Vergleichen der Tabellenwerte im Vordergrund. Ein gezielter Lehrimpuls sollte dazu anregen, einzelne Werte nicht nur vorzulesen, sondern in den Gesamtzusammenhang einzuordnen. Der Impuls 21k war eher wenig zielführend und erst die Nachfrage 21l führte zu einer passenden und

anschlussfähigen Reaktion. Folgender Leitimpuls wird vorgeschlagen:
„Erklärt doch einmal, was die Reihen in der Tabelle bedeuten. Dafür gehen wir am besten so vor: Jede:r von euch nennt einen Wert und beschreibt, wie er den im Text gefunden hat."
Mit diesem Impuls werden die Intentionen von 21k und 21l vereint. Die Lernenden werden aufgefordert, die zuvor aus dem Text extrahierten Werte vorzulesen und deren Bedeutung zu erklären. Die Lehrkraft rückt ab dem Zeitpunkt in den Hintergrund und wird nur noch moderierend tätig.

Leitimpulse in der Sicherungsphase zum AB 3

Im Anschluss an die Bearbeitung der Prüfkarten (AB 3) wurde eine Sicherungsphase im Plenum durchgeführt. Das Ziel dieser Phase bestand in dem Vergleich der von den Lernenden angekreuzten Antworten bei der Differenzierungsstufe *Basis* sowie der in *Stern1* und *Stern2* geforderten Formulierung von Schlussfolgerungen bezogen auf die Übertragbarkeit der Studienergebnisse auf die deutsche Bevölkerung. Begleitet wurde die Phase durch die vier Impulse 21m, 21n, 21o und 21p. Die Impulse 21m und 21n erwiesen sich als passend, aber nicht anschlussfähig: Die in der Erprobung gewählte Formulierung war zu suggestiv, sodass die Lernenden mit einer sehr kurzen Antwort reagieren konnten. Um die Anschlussfähigkeit zu gewährleisten, sollten die Impulse offener gestaltet sein und mehr Raum für Erklärungen durch die Lernenden lassen. Dies gelingt durch die Impulse 21o und 21p bereits in angemessenem Maß, denn sie verlangen eine Begründung.
Auf diesen Erkenntnissen basierend wird folgender Leitimpuls formuliert:
„Wir möchten einmal gemeinsam schauen, ob die Heinsberg-Studie randomisiert ist. Dafür brauchen wir jemanden von euch, der Randomisierung genau erklärt und dann sagt, ob das in der Studie gegeben ist!"
Dieser Leitimpuls vereint die Differenzierungsstufen und sorgt dafür, dass alle Lernenden sich an der Diskussion beteiligen können, unabhängig davon, welche Differenzierungsstufe sie zuvor gewählt haben. Damit werden die in der Erprobung verwendeten Impulse 21q, 21r und 21s vollständig ersetzt. Das fragend-entwickelnde Unterrichtsgespräch kann stärker zu einer schülerzentrierten Diskussion umgewandelt werden und die Unterrichtszeit kann gewinnbringender genutzt werden, indem die Lehrkraft zu spezifischen Aspekten Nachfragen stellt oder tiefer gehende Transferaufgaben diskutieren lässt.

Impuls 21t diente als Überleitung in die Anwendungs- bzw. Bewertungsphase. Die Lehrkraft erkundigte sich, ob Lernende bereits von der Heinsberg-Studie gehört hatten. Die Lernenden antworteten darauf nicht. Hier könnte die Lehrkraft geschickter vorgehen, indem stärker an die Lebenswelt der Schüler:innen angeknüpft wird. Anstatt sich nach der Studie zu erkundigen, könnte die Lehrperson nach allgemeineren medienbezogenen Auffälligkeiten fragen, z. B.: „Welche Studien kennt ihr denn aus dem Alltag?“ oder „Kennt ihr Schlagzeilen, in denen wissenschaftliche Studien eine Rolle spielen?“

Leitimpuls in der Anwendungs- und Bewertungsphase

Die Impulse 21u, 21v und 21y wurden von den Lernenden nicht passend beantwortet. Von der Lehrkraft intendiert war eine differenzierte Bewertung der Art und Weise, wie öffentliche Medien die Faktenlage darstellen. Die Lernenden sollten explizieren, dass die Heinsberg-Studie wissenschaftlich korrekt randomisiert wurde, aber dass die Ergebnisse der Studie von der Presse zu stark vereinfacht und plakativ dargestellt wurde. Eine derartige Bewertung führten die Lernenden jedoch nicht durch. Der Leitimpuls für die Anwendungsphase muss demnach deutlicher hervorheben, dass die Schlagzeile im Hinblick auf die zuvor analysierte Randomisierung der Heinsberg-Studie zu bewerten ist. Einen passenden Ansatz zeigt Impuls 21w, denn die Reaktion der Schüler:innen zeigt, dass der Impuls einen stärkeren Aufforderungscharakter besaß. Die Lehrkraft betonte: „Wir können ihr [der Studie] vertrauen, aber nur unter einer Einschränkung!“ Ein Schüler erklärte daraufhin die Einschränkung, dass das Studienergebnis nur für Gangelt gültig ist. Der Fachbegriff *repräsentativ* wurde jedoch nicht erwähnt.
Aus der Bewertung von Passung und Anschlussfähigkeit der Impulse wird folgender Leitimpuls zur Verwendung in der Anwendungsphase vorgeschlagen:
„Irgendetwas stimmt doch mit diesem Zeitungsartikel nicht. Bewertet die Schlagzeile!“

Optimierung des Ablenkungspotentials

Die vorangegangene Analyse von *Off-Task*-Interaktionen hat gezeigt, dass diese in der untersuchten Unterrichtsstunde nur selten auftraten. Lediglich gegen Ende der Arbeitsphase zum AB 2 ist in vier von fünf Gruppen erkennbar, dass sie *off-task* ge-

wesen sind. Vermutlich hatten die Gruppen ihre Arbeitsaufträge zu dem Zeitpunkt bereits abgeschlossen und warteten auf die nächste Plenumsphase.
Im Kapitel 9.2.5 wurden bereits Möglichkeiten aufgezeigt, den Off-Task-Interaktionen zu begegnen, wie z. B. Bildschirmsperren oder stärkere innere Differenzierung. Diese sind auch zur Optimierung des Ablenkungspotentials in der Unterrichtsstunde zum Prüfkriterium Randomisierung geeignet und sollten daher in Zukunft umgesetzt werden.

Minimieren von Verständnisschwierigkeiten Die Codes *Fachinhalt*, *Aufgabenstellung* und *Technik* wurden als Unterkategorien zur Codierung von Verständnisschwierigkeiten verwendet. Es hat sich herausgestellt, dass während der Bearbeitung des AB 2 keine Probleme auftraten. Erst in der darauffolgenden Arbeitsphase brachten einige Lernende zum Ausdruck, dass sie Schwierigkeiten in allen drei Kategorien hatten. Es lässt sich jedoch nicht für jede Situation identifizieren, welche Inhalte den Lernenden unklar geblieben sind. Die Ausrufe sind oft eher unspezifisch („Ich bin überfordert", „Ich verstehe das nicht") und lassen keine Schlussfolgerungen auf die Ursache der Verständnisschwierigkeiten zu.
In manchen Fällen wird jedoch klar, welche Aspekte verbessert werden können, wie beispielsweise bei der Kategorie *Aufgabenstellung*. Ausrufe der Form: „Moment, sollen wir jetzt etwas schreiben?" (20-09-21_Gruppe3, Pos. 24) oder: „Sollen wir schon Aufgabe 2 machen?" fallen unter diese Kategorie; derartigen Situationen sollte durch ein bedachteres Anleiten der Arbeitsphasen vorgebeugt werden. Dabei sollte die Lehrkraft betonen, an welchen Stellen des Arbeitsmaterials die Lernenden etwas notieren sollen und wie genau der Arbeitsauftrag gestaltet ist.
In Gruppe 4 äußert eine Schülerin während sie sich den Arbeitsauftrag auf AB 3 durchliest: „Worüber sollen wir jetzt diskutieren?" (20-09-21_Gruppe 4, Pos. 81). Damit bringt sie zum Ausdruck, dass die Erarbeitung des Konzepts Randomisierung nicht unmittelbar mit dem Arbeitsauftrag in Zusammenhang gebracht wurde. Abhilfe schaffen könnte eine leichte Umformulierung der Aufgabe, sodass die Verknüpfung mit der zuvor analysierten Heinsberg-Studie deutlicher wird: „Diskutiere mit deinem Partner, ob die Forscher:innen bei der Heinsberg-Studie die Randomisierung beachtet haben".
Eine weitere Verbesserungsmöglichkeit besteht in der Abgrenzung der Differenzierungsstufen. Auffällig ist, dass viele Schüler:innen sich zwar zunächst für die leich-

teste Stufe (Basis) entschieden, jedoch anschließend zu Stern 1 wechselten und die dort aufgeführten Aufgaben durchführten; teilweise wurde danach sogar noch Stern 2 bearbeitet. Diese Vorgehensweise war jedoch nicht intendiert. Die Lernenden sollten sich zu Beginn für eine der Stufen entscheiden und bei Bedarf dann die innerhalb dieser Stufe vorhandenen „Stern-Aufgaben“ lösen. Ein Wechsel zwischen den Arbeitsblättern war nicht vorgesehen.
Das Vorgehen der Lernenden liegt vermutlich in ihrer Unterrichtsroutine begründet, die während des Regelunterrichts bei der betreuenden Lehrkraft aufgebaut wurde. Es ist somit notwendig, die Art der Differenzierung stärker den Gewohnheiten der Lerngruppe anzupassen, sodass die Lernenden hürdenlos die für sie passende Differenzierungsstufe wählen können.

9.3.6 Fazit und Ausblick

Der Zyklus 1.3 in diesem Kapitel 9.3 stellt einen Durchlauf des ersten Mesozyklus dieser Arbeit dar. Die umspannende Frage beschäftigt sich damit, inwieweit Lernende das Prüfkriterium *Randomisierung* erarbeiten und anwenden können.
Es wurden Lernmaterialien zur Erarbeitung von *Randomisierung* theoriebasiert entwickelt und für die Verwendung im naturwissenschaftlichen Unterricht vorbereitet. Ein aktueller gesellschaftlich kontrovers diskutierter Kontext wurde ausgewählt und Schüler:innen-gerecht aufgearbeitet. Im Rahmen einer Gesamterprobung im September 2020 wurden die Materialien im Unterrichtseinsatz erprobt. Das Kapitel 9.3.3 beinhaltet eine ausführliche Beschreibung der Erprobungsgegebenheiten sowie eine qualitative Inhaltsanalyse der erhobenen Daten anhand eines Kategoriensystems.
Bei der Analyse wurden zum einen die Lehrimpulse in Plenumsphasen und zum anderen das Verhalten der Lernenden in Gruppenarbeitsphasen betrachtet.
Die Reaktionen der Lernenden auf Lehrimpulse waren in einigen Fällen zum Lehrimpuls passend und auch anschlussfähig für den weiteren Unterrichtsverlauf, teilweise reagierten die Lernenden aber auch unpassend und kaum anschlussfähig. Somit wurden alle Impuls-Reaktionen hinsichtlich der Passung und Anschlussfähigkeit bewertet, um daraus Rückschlüsse auf die Eignung des Lehrimpulses zu ziehen. Zur Förderung der Eigenständigkeit von Lernenden in den Plenumsphasen und Vermeidung eines zu stark vorstrukturierten Lehrerverhaltens wurden aus den Rückschlüssen für jede Plenumsphase Leitimpulse entwickelt, an denen sich Lehrkräfte bei der

Planung der Stunde orientieren können.
Bei der Betrachtung der Gruppenarbeitsphasen wurden das Auftreten von Verständnisschwierigkeiten, Aha-Effekten und Off-Task-Interaktionen besonders fokussiert. Durch diese Vorgehensweise konnte ermittelt werden, an welchen Stellen des Unterrichts möglicherweise kognitive Konflikte auftraten oder gelöst wurden und zu welchen Zeitpunkten die Lernenden sich - z. B. aufgrund von Langeweile oder Ablenkung durch digitale Medien - nicht mit dem Stundenthema befassten. Im Re-Design wurden die Ergebnisse verwendet, um das Ablenkungspotential in der Stunde und Verständnisschwierigkeiten im Umgang mit den Lernmaterialien zu minimieren.
Weiterhin konnte festgestellt werden, dass die Unterrichtsmaterialien alle zuvor formulierten Lernziele adressieren. Auf ein vollständiges Erreichen der Lernziele konnte jedoch nicht mit Sicherheit geschlossen werden: Es wurden Hinweise auf das Erreichen der Lernziele 1, 2 und 3 gefunden, die jeweils geltenden Einschränkungen werden im Kapitel 9.3.4 beschrieben. Lernziel 4 konnte vollständig erreicht werden.
Mit den im Re-Design entwickelten Verbesserungen des Lernmaterials und den Vorschlägen für eine gezielte Gestaltung der Plenumsphasen ist der Weg für die weitere Verwendung der Arbeitsblätter und Stundenverlaufspläne im Rahmen der Unterrichtskonzeption choice^{2}reflect geebnet. Es konnte gezeigt werden, dass einige Lernende der neunten Klasse in der Lage sind, das Konzept der Randomisierung zu erarbeiten und zielführend auf die Heinsberg-Studie anzuwenden. Die Aufgabe weiterer Untersuchungen ist es, die Eignung der verbesserten Materialien im realen Unterrichtseinsatz zu überprüfen und das Material im größeren Einsatz mit weiteren Lerngruppen zu erproben.

9.4 MZ 1.4: Reproduzierbarkeit

9.4.1 Theoretischer Hintergrund zur Reproduzierbarkeit

Rationale Bewertungen erfordern verlässliche, wissenschaftlich abgesicherte Informationen. Im Rahmen der Phase III von choice^{2}reflect erarbeiten Lernende Wissenschaftskriterien, anhand derer Informationen auf ihre Wissenschaftlichkeit geprüft werden können. Diese Prüfkriterien zu entwickeln und im unterrichtlichen Einsatz zu erproben ist die Hauptaufgabe des ersten Mesozyklus im hier beschriebenen Projekt.

Die bis zu diesem Punkt ausgearbeiteten Wissenschaftskriterien fokussieren sich größtenteils auf die *Validität* empirischer Untersuchungen. Sie beleuchten die Notwendigkeit von randomisierten, kontrollierten und verblindeten Studiendesigns für eine adäquate Berücksichtigung von Störvariablen. Auch das wissenschaftlich korrekte Schlussfolgern unter Berücksichtigung möglicher Scheinkausalitäten ist Teil der Prüfkriterien.
Das Grundgerüst valider Untersuchungen stellen jedoch deren Reliabilität und Objektivität dar. Dieser Zusammenhang wird im Kapitel 2.4.1 genau beleuchtet. Reliabilität und Objektivität werden sichergestellt, indem eine empirische Untersuchung *reproduzierbar* gestaltet wird:
Reproduzierbar bedeutet, dass genau das gleiche Ergebnis erhalten werden muss, wenn das Experiment unter den gleichen Bedingungen (*reliabel*) und von anderen Wissenschaftler:innen (*objektiv*) erneut durchgeführt wird (Tetens 2018). Dabei geht der Begriff über die reine Wiederholbarkeit hinaus: Reproduzierbarkeit ist abhängig vom Versuchsergebnis.
Das Forschungsdesign muss dazu nicht nur prinzipiell reproduzierbar sein, sondern auch so dokumentiert werden, dass es von unabhängigen Forscher:innen nachvollzogen werden kann (Huber et al. 2014; Pernkopf 2006).
Die große Bedeutung reproduzierbar gestalteter Forschung wird auch von der Deutschen Forschungsgemeinschaft in einer Stellungnahme herausgestellt: Reproduzierbarkeit ist für alle wissenschaftlichen Teilbereiche von Bedeutung, in denen experimentalwissenschaftliche und quantitativ begründete empirische Wissensansprüche deklariert werden (DFG 2017).
Fehlende Reproduzierbarkeit entsteht durch das Fälschen oder Erfinden von Daten, gezieltes Verwerfen unerwünschter Daten und unbewusste Befangenheit (Bergman & Danheiser 2016). Erreicht wird Reproduzierbarkeit durch vollständige Aufzeichnung aller Gegebenheiten und einer ausreichend großen Stichprobe; außerdem sollten auch die anderen Prüfkriterien (Randomisierung, kontrollierte Bedingungen und Verblindung) erfüllt sein. Nur so wird prinzipielle Reproduzierbarkeit gewährleistet.

9.4.2 Entwicklung des Lernmaterials

Das Modul zur Erarbeitung des Wissenschaftskriteriums *Reproduzierbarkeit* entstand im Rahmen einer Masterarbeit in Zusammenarbeit mit Carolin Banse. Gemeinsame

Gespräche und Feedback-Runden trugen zur theoriebasierten Entwicklung von Kontext, Arbeitsblättern und Stundenverlauf bei. Die Zielgruppe bilden Schüler:innen der neunten Klasse.

Dabei wurde die Unterrichtseinheit in zwei Teile gegliedert: Zum einen erfahren die Lernenden die Wichtigkeit reproduzierbarer Versuchsdokumentation anhand eines eigens geplanten und durchgeführten Experiments. Zum anderen wird der Begriff der Reproduzierbarkeit auf wissenschaftliche Studien erweitert.

Folgende Lernziele werden durch das Material verfolgt:

Die Schülerinnen und Schüler...

1. erarbeiten selbstständig spezifische Kriterien für Reproduzierbarkeit am Beispiel des Experiments „Die Zitronenbatterie“ und benennen diese Kriterien
2. bewerten anhand von Kriterien für Reproduzierbarkeit das Versuchsprotokoll zum Experiment „Die Zitronenbatterie“ und begründen ihre Bewertung
3. bewerten anhand von Kriterien für Reproduzierbarkeit die Studie zum Thema „E-Scooter“ und begründen ihre Bewertung.

Die Diskussion um Elektro-Tretroller („E-Scooter“) hat in Deutschland im Laufe des Jahres 2019 Fahrt aufgenommen, als die Fahrzeuge für den Einsatz im Straßenverkehr zugelassen wurden. Es ist davon auszugehen, dass dieses Thema auch unter Jugendlichen ein lebensnahes und aktuelles Thema darstellt, da die Nutzung der Scooter bereits ab 14 Jahren zugelassen ist.

Die Nachhaltigkeit der Scooter wird viel diskutiert, dabei werden Argumente durch naturwissenschaftliche Fakten zur Funktionsweise und den Inhaltsstoffen der Batterien untermauert. Es bietet sich an, die verschiedenen Standpunkte im Unterricht zu thematisieren und die Herkunft der Informationen gemeinsam mit den Lernenden zu überprüfen. Dafür wurde das Arbeitsmaterial zur Reproduzierbarkeit im Kontext *Nachhaltigkeit von E-Scootern* entwickelt.

Die bereits erwähnte Zweiteilung der Unterrichtseinheit kann auch in diesem Kontext erwirkt werden: Auf der einen Seite können Schüler:innen durch Experimente die Funktionsweise der Batterien nachvollziehen, auf der anderen Seite ist es jedoch auch wichtig, die empirisch-quantitative Seite der Thematik zu betrachten und Studien zur Nachhaltigkeit einzubeziehen.

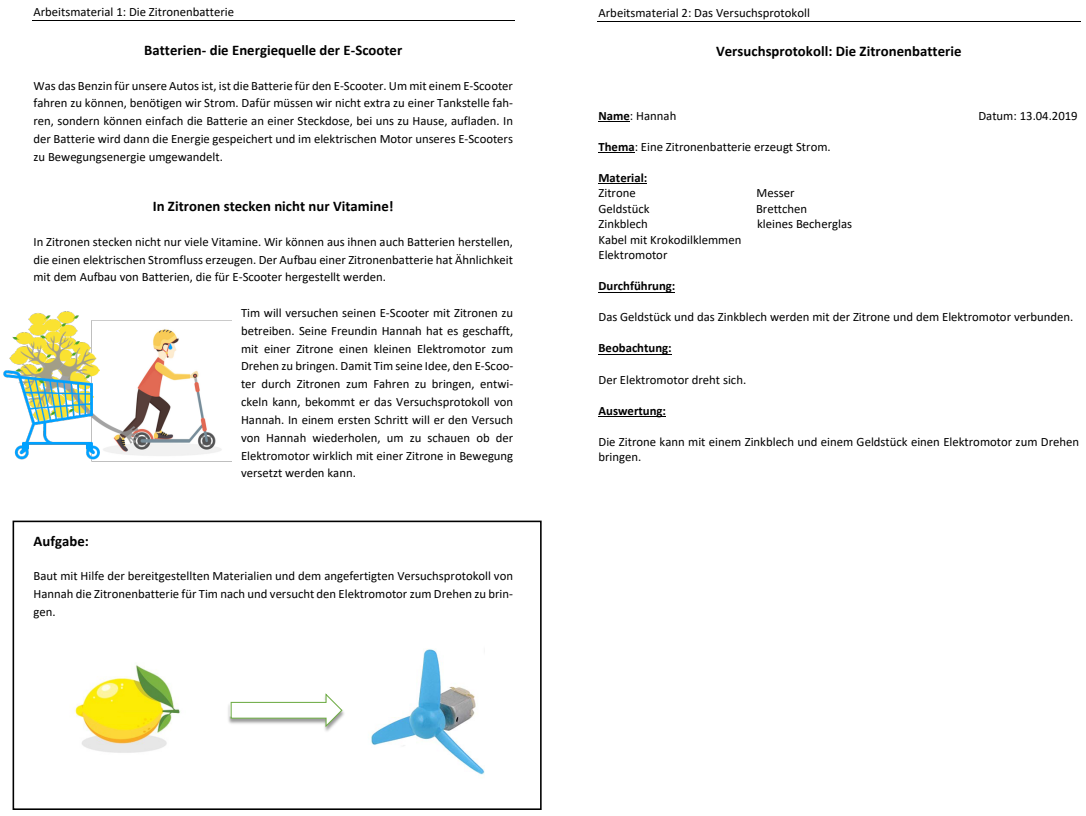
Das erste Arbeitsblatt (Abbildung 9.14) mit dem Titel: „Die Zitronenbatterie“ enthält einen kurzen auf die Batterietechnik fokussierten Einführungstext zu Elektro-

Scootern, dabei wird jedoch auf tiefgreifende Details der Elektrochemie verzichtet. Dass in Zitronen „nicht nur Vitamine stecken", wird im nächsten Abschnitt des AB verdeutlicht. Die Schüler:innen erfahren, dass die Herstellung einer Batterie aus Zitronen theoretisch möglich ist.
Die elektrochemischen Prozesse der Zitronenbatterie sind sehr komplex und werden viel diskutiert (Rosenberg et al. 2019; Rosenberg et al. 2018; Hagen-Schittkowski et al. 2015).
Manche Forschungs-Erkenntnisse widerlegen den Zusammenhang zur galvanischen Zelle, daher ist der Einsatz der Zitronenbatterie im Chemieunterricht umstritten. In der Unterrichtseinheit zur Reproduzierbarkeit wird die Batterie allerdings losgelöst von der fachwissenschaftlichen Erklärung chemischer Reaktionen in einem anderen Kontext verwendet.
Schüler:innen sollen versuchen, anhand eines vorgegebenen Versuchsprotokolls die Zitronenbatterie nachzubauen. Dazu wird eine fiktive Person namens Tim eingeführt, die eine „kreative Idee" umsetzen möchte. Tim möchte versuchen, seinen E-Scooter mit Zitronen zu betreiben. Ein kurzer Text mit angegliedertem Arbeitsauftrag erklärt diese Idee. Die Aufgabe lautet:

> Baut mit Hilfe der bereitgestellten Materialien und dem angefertigten Versuchsprotokoll von Hannah die Zitronenbatterie für Tim nach und versucht den Elektromotor zum Drehen zu bringen.

Mithilfe einer vorbereiteten Experimentierbox, die alle notwendigen Materialien enthält, und einem Versuchsprotokoll (Arbeitsblatt 2) führen die Lernenden die Aufgabe durch.
Das zweite Arbeitsblatt (Abbildung 9.14) besitzt die Struktur eines klassischen Versuchsprotokolls, das die Lernenden aus dem regulären Chemieunterricht bereits kennen sollten. Es beinhaltet Überschrift, Thema, Namen der Verfasserin oder des Verfassers, Material, Durchführung, Beobachtung und Auswertung.
Die Durchführung, Beobachtung und Auswertung bestehen jeweils nur aus kurzen Sätzen ohne hilfreiche Informationen darüber, wie die Batterie tatsächlich aufzubauen ist. Das Protokoll ist somit unvollständig - die Lernenden erhalten keine exakte Beschreibung der nötigen Versuchsschritte. Sie erfahren nur: „Das Geldstück und das Zinkblech werden mit der Zitrone und dem Elektromotor verbunden". Es gibt keine genaueren Angaben zur Beschaffenheit des Geldstücks.

Das fehlerhafte und unvollständige Versuchsprotokoll soll im Verlauf der Unterrichtseinheit den Schüler:innen eine Grundlage geben, an der sie die Kriterien für Reproduzierbarkeit selbst erarbeiten können: Sind nicht alle Materialien und Versuchsschritte genau dokumentiert, ist es für andere Wissenschaftler:innen nicht möglich, das Ergebnis zu reproduzieren.
Die Aufgabenstellung für die Experimentierphase orientiert sich am Konzept der Interaktionsboxen (Marohn 2014) und findet hier in abgewandelter Form Anwendung. Dabei bleibt der genaue Lösungsweg offen, die Lernenden können mit den angegebenen Materialien verschiedene Varianten ausprobieren.



Arbeitsmaterial 1: Die Zitronenbatterie

Batterien- die Energiequelle der E-Scooter

Was das Benzin für unsere Autos ist, ist die Batterie für den E-Scooter. Um mit einem E-Scooter fahren zu können, benötigen wir Strom. Dafür müssen wir nicht extra zu einer Tankstelle fahren, sondern können einfach die Batterie an einer Steckdose, bei uns zu Hause, aufladen. In der Batterie wird dann die Energie gespeichert und im elektrischen Motor unseres E-Scooters zu Bewegungsenergie umgewandelt.

In Zitronen stecken nicht nur Vitamine!

In Zitronen stecken nicht nur viele Vitamine. Wir können aus ihnen auch Batterien herstellen, die einen elektrischen Stromfluss erzeugen. Der Aufbau einer Zitronenbatterie hat Ähnlichkeit mit dem Aufbau von Batterien, die für E-Scooter hergestellt werden.

Tim will versuchen seinen E-Scooter mit Zitronen zu betreiben. Seine Freundin Hannah hat es geschafft, mit einer Zitrone einen kleinen Elektromotor zum Drehen zu bringen. Damit Tim seine Idee, den E-Scooter durch Zitronen zum Fahren zu bringen, entwickeln kann, bekommt er das Versuchsprotokoll von Hannah. In einem ersten Schritt will er den Versuch von Hannah wiederholen, um zu schauen ob der Elektromotor wirklich mit einer Zitrone in Bewegung versetzt werden kann.

Aufgabe:

Baut mit Hilfe der bereitgestellten Materialien und dem angefertigten Versuchsprotokoll von Hannah die Zitronenbatterie für Tim nach und versucht den Elektromotor zum Drehen zu bringen.

Arbeitsmaterial 2: Das Versuchsprotokoll

Versuchsprotokoll: Die Zitronenbatterie

Name: Hannah Datum: 13.04.2019

Thema: Eine Zitronenbatterie erzeugt Strom.

Material:

Zitrone	Messer
Geldstück	Brettchen
Zinkblech	kleines Becherglas
Kabel mit Krokodilklemmen	
Elektromotor	

Durchführung:

Das Geldstück und das Zinkblech werden mit der Zitrone und dem Elektromotor verbunden.

Beobachtung:

Der Elektromotor dreht sich.

Auswertung:

Die Zitrone kann mit einem Zinkblech und einem Geldstück einen Elektromotor zum Drehen bringen.

Abbildung 9.14: *Arbeitsblätter 1 (links) und 2 (rechts) zur Erarbeitung des Prüfkriteriums Reproduzierbarkeit.*

In einer an das AB 2 angegliederten Plenumsphase sollten die Schülergruppen kurz die erzielten Ergebnisse rückmelden und berichten, welche Vorgehensweise zum erfolgreichen Betreiben der Zitronenbatterie geführt hat. Die Lehrkraft bündelt die Rückmeldungen und nutzt diese für eine Überleitung zur Erarbeitung des Wissenschafts-

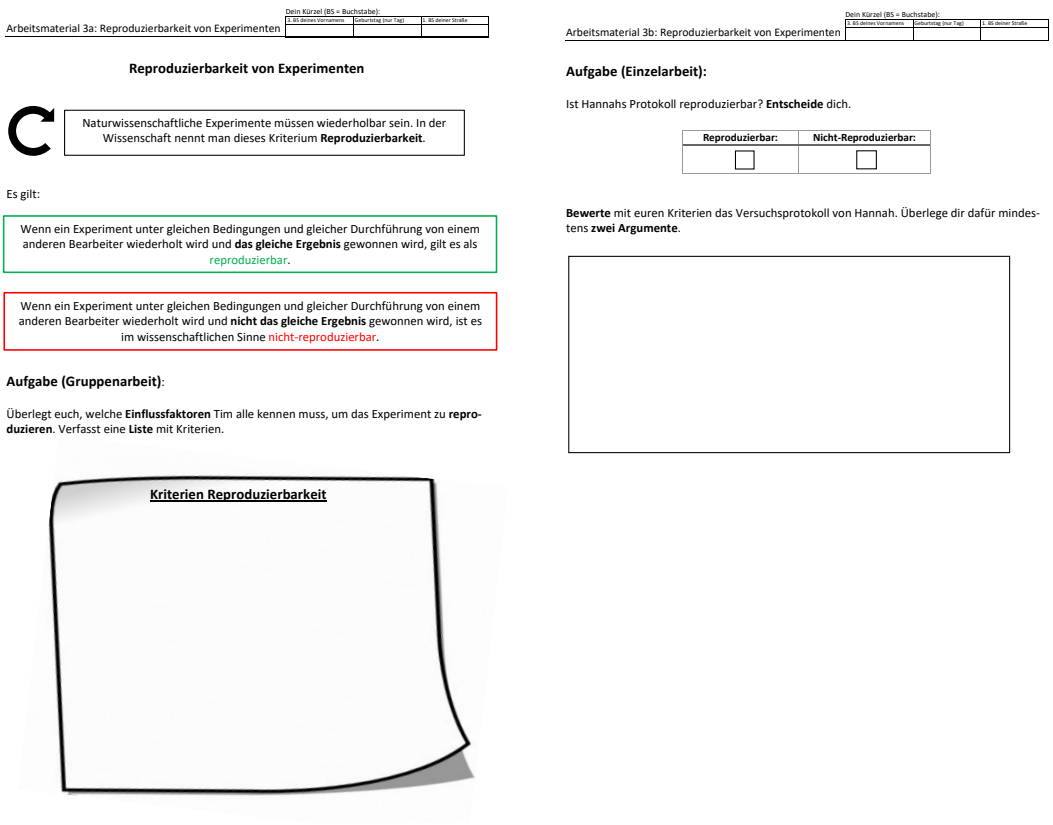
kriteriums *Reproduzierbarkeit*. Dabei stellt sie heraus, dass wissenschaftliche Experimente immer so angelegt sein müssen, dass sie bei einer Wiederholung stes das gleiche Ergebnis erzielen. Dann wird das dritte Arbeitsblatt ausgeteilt.
Arbeitsmaterial 3a (Abbildung 9.15) enthält theoretische Hintergrundinformationen zur Reproduzierbarkeit von Experimenten und definiert die Begriffe *Reproduzierbarkeit*, *reproduzierbar* und *nicht-reproduzierbar*. Der Arbeitsauftrag fordert die Lernenden auf, Kriterien für Reproduzierbarkeit festzulegen:

> Überlegt euch, welche **Einflussfaktoren** Tim kennen muss, um das Experiment zu **reproduzieren**. Verfasst eine Liste mit **Kriterien**.

Nach Beendigung der Gruppenarbeit tragen die Lernenden ihre Ergebnisse an der Tafel zusammen. Die Lehrkraft ordnet die Beiträge in Form einer Mind-Map um den Begriff *Reproduzierbarkeit*.
Mithilfe von AB 3b (Abbildung 9.15) bewerten die Schüler:innen das Versuchsprotokoll im Hinblick auf seine Reproduzierbarkeit. Die Antworten werden im Plenum diskutiert. Dabei kann die Lehrkraft moderieren und Fragen einwerfen wie z. B.: „Gibt es Begründungen, die für die Gegenposition sprechen?“ Das Ziel dieser Plenumsphase ist es, das Versuchsprotokoll als *nicht-reproduzierbar* einzustufen, da es nicht alle wichtigen Informationen zur Durchführung des Experiments enthält. Damit ist der erste Teil der Unterrichtseinheit abgeschlossen.
Im nächsten Teil findet eine Übertragung des erlernten Konzepts auf wissenschaftliche Studien mithilfe von Arbeitsblatt 4 (Abbildung **??**) statt. Dieses greift die vorherigen Begriffe auf und erklärt, welche Bedeutung sie im Kontext empirischer Studien besitzen:

> Wenn eine wissenschaftliche Studie mit der gleichen Methode und den gleichen Versuchsbedingungen, aber mit anderen Versuchsteilnehmern von einem anderen Bearbeiter, erneut wiederholt wird und das gleiche Ergebnis erhalten wird, spricht man von einer reproduzierbaren Studie.

Das Arbeitsblatt thematisiert auch die Außenwirkung wissenschaftlicher Forschung, Details dazu finden sich bei (Banse 2020). Es enthält keine direkten Aufgabenstellungen, legt jedoch die Grundlagen für AB 5 und 6.



Arbeitsmaterial 3a: Reproduzierbarkeit von Experimenten

Dein Kürzel (BS = Buchstabe):

1. BS deines Vornamens	Geburtstag (nur Tag)	1. BS deiner Straße

Reproduzierbarkeit von Experimenten

Naturwissenschaftliche Experimente müssen wiederholbar sein. In der Wissenschaft nennt man dieses Kriterium **Reproduzierbarkeit**.

Es gilt:

Wenn ein Experiment unter gleichen Bedingungen und gleicher Durchführung von einem anderen Bearbeiter wiederholt wird und **das gleiche Ergebnis** gewonnen wird, gilt es als reproduzierbar.

Wenn ein Experiment unter gleichen Bedingungen und gleicher Durchführung von einem anderen Bearbeiter wiederholt wird und **nicht das gleiche Ergebnis** gewonnen wird, ist es im wissenschaftlichen Sinne nicht-reproduzierbar.

Aufgabe (Gruppenarbeit):

Überlegt euch, welche **Einflussfaktoren** Tim alle kennen muss, um das Experiment zu **reproduzieren**. Verfasst eine **Liste** mit Kriterien.

Kriterien Reproduzierbarkeit

Arbeitsmaterial 3b: Reproduzierbarkeit von Experimenten

Dein Kürzel (BS = Buchstabe):

1. BS deines Vornamens	Geburtstag (nur Tag)	1. BS deiner Straße

Aufgabe (Einzelarbeit):

Ist Hannahs Protokoll reproduzierbar? **Entscheide** dich.

Reproduzierbar:	Nicht-Reproduzierbar:
☐	☐

Bewerte mit euren Kriterien das Versuchsprotokoll von Hannah. Überlege dir dafür mindestens **zwei Argumente**.

Abbildung 9.15: *Arbeitsblätter 3a (links) und 3b (rechts) zur Erarbeitung des Prüfkriteriums Reproduzierbarkeit.*

Arbeitsmaterial 5 (Abbildung **??**) besteht aus einer gekürzten und veränderten Studie des Marktforschungsunternehmens GfK („Growth for Knowledge“). Dabei handelt es sich um eine Pressemitteilung über eine Online-Umfrage zu E-Scootern (GfK-Pressemitteilung 2019). Diese bietet eine Diskussionsgrundlage für die Frage, inwiefern die Studie als reproduzierbar eingestuft werden kann. Eine abschließende Bewertung treffen Schüler:innen mithilfe von AB 6. Dazu dient folgender von Banse (2020) entwickelter Arbeitsauftrag:

> Die Studie soll reproduziert werden, um ihr Ergebnis zu überprüfen. Überlege und plane in Gedanken die erneute Durchführung der Studie. **Bewerte** die **Reproduzierbarkeit** der Studie. Begründe deine Antwort.

Die Lernenden sollten hier unter Rückbezug auf die zuvor entwickelten Kriterien für Reproduzierbarkeit erkennen, dass die Studie nicht reproduzierbar ist. Als Begrün-

dung kann z. B. das Fehlen wichtiger Hintergrundinformationen über die Stichprobe herangezogen werden.
In einer letzten Plenumsphase erstellt die Lehrkraft eine weitere Mind-Map an der Tafel, in der die Arbeitsergebnisse der Schülergruppen gesammelt werden. Damit gilt die Erarbeitung des Prüfkriteriums *Reproduzierbarkeit* als abgeschlossen.

9.4.3 Erprobung

Methodik

Banse (2020) beschreibt detailliert die organisatorischen Gegebenheiten bei der 90 minütigen Erprobung der Lernmaterialien mit 28 Schüler:innen der neunten Klasse eines Gymnasiums in Münster.
Es wurde untersucht, wie Lernende mit dem Material umgehen und inwiefern es sich für den Einsatz in der Schule als geeignet herausstellt. Dazu wurden sowohl Audio-/Videomitschnitte der Unterrichtsstunde angefertigt als auch schriftlich ausgefüllte Arbeitsblätter erfasst und analysiert.
Die Arbeitsblätter zeigen, welche Ergebnisse die Schüler:innen letztendlich notiert haben, während anhand der Mitschnitte der Arbeitsprozess verfolgt werden kann. Mögliche Diskussionen und Gespräche zwischen den Lernenden werden dazu analysiert. Jede:r Lernende generierte ein individuelles Kürzel, das zur anonymen Identifizierung dient und einen Abgleich zwischen Arbeitsblatt und Videomaterial erlaubt.
Materialien 3a, 3b und 6 wurden digitalisiert und in das Programm MAXQDA zur qualitativen Inhaltsanalyse eingepflegt. Aus den Antworten der Lernenden in den vorgegebenen Antwortfeldern wurden nach einer vollständigen Sichtung induktiv Kategorien erstellt. In einem zweiten Schritt wurden die induktiv gebildeten Kategorien auf das Material angewendet und durchkodiert. Drei Fragestellungen wurden dabei verfolgt, die sich mit den Lernzielen decken.
Um die Frage nach dem Erreichen der Lernziele zielführend beantworten zu können, wird eine Analyse des Audio- und Videomaterials hinzugezogen.
Alle Mitschnitte wurden zunächst gesichtet, um besonders aufschlussreiche Diskussionen der Lernenden untereinander zu identifizieren. Diese wurden einer wörtlichen Transkription unterzogen. Ausgewählt wurden die Situationen nach erkennbaren Problemen oder Verständnisschwierigkeiten sowie nach anderen Merkmalen,

Abbildung 9.16: *Arbeitsblätter 4 (oben) und 5 (unten) zur Erarbeitung des Prüfkriteriums Reproduzierbarkeit.*

die Aufschluss über den Arbeitsprozess geben können.

Ergebnisse

Die durchgeführte Unterrichtsstunde wurde von Banse (2020) einem subjektiven Rückblick unterzogen. Sie beschreibt, dass die zeitlichen Vorgaben des geplanten Stundenverlaufs weitgehend eingehalten wurden. Der Kontext E-Scooter wurde von den Lernenden angenommen und das Vorwissen der Schüler:innen konnte gut aktiviert werden. In der Experimentalphase schafften einige Gruppen es, eine funktionsfähige Zitronenbatterie zu bauen, während andere Gruppen nicht erfolgreich waren. Die Gruppenphase zum AB 2 verlief ohne größere Hilfestellungen der Lehrkraft, in der zweiten Plenumsphase beteiligten sich viele Schüler:innen. Bei der Erstellung der zentralen Mind-Map nannten sie alle wichtigen Kriterien.
Bezüglich Arbeitsmaterial 3b gab es ein paar Nachfragen und nicht alle Lernenden schienen die Aufgabenstellung problemlos bearbeiten zu können.
Im zweiten Teil der Unterrichtsstunde wurde deutlich, dass die Schüler:innen nicht durchgängig wussten, wie die Aufgabe zu lösen war. Auch empfand Banse die Schulklasse als unruhig. Dennoch konnten die Arbeitsergebnisse im Plenum zügig zusammengetragen und in ein zielführendes Fazit umgewandelt werden.

Das Arbeitsmaterial 3a diente der eigenständigen Formulierung von Einflussfaktoren des Zitronenbatterie-Experiments. Der Code-Matrix-Browser in Abbildung 9.17 stellt dar, wie die Schülerantworten auf die jeweiligen Kategorien verteilt sind. Es ist erkennbar, dass mehr als die Hälfte der Schüler:innen das Material der Münze und die Beschaffenheit der Zitrone als Einflussfaktoren angaben. Auch die Menge des Zitronensafts, die Funktionalität der Materialien oder das Motormodell wurden häufig genannt.
Die Transkripte der Gruppenarbeitsphase liefern keine tiefergehenden Erkenntnisse über diese Aufgabe und werden daher an dieser Stelle nicht weiter behandelt.
Das Versuchsprotokoll bewerteten die Lernenden mithilfe von Arbeitsmaterial 3b. Sie sollten zwischen *reproduzierbar* und *nicht-reproduzierbar* wählen und anschließend eine Begründung abgeben. Die Verteilung der Antworten zeigt sich in Abbildung 9.18. Die unteren fünf Argumente (orange markiert) stellen fachlich korrekte Argumente dar. Die drei oberen (grau markierten) Argumente sind nicht korrekt.
Es zeigt sich, dass die Mehrheit der Schüler:innen fachlich korrekte Antworten wähl-

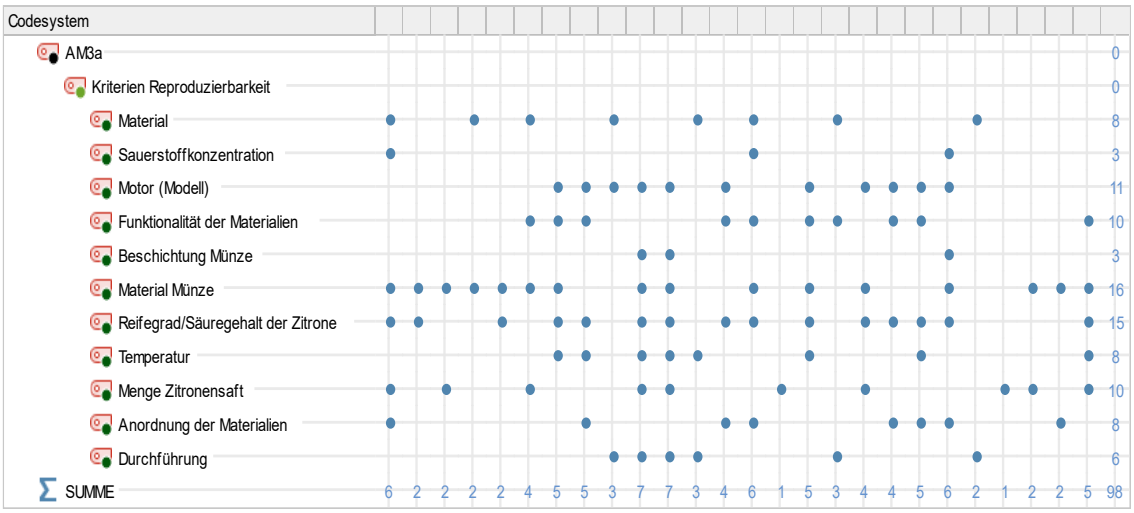


Abbildung 9.17: *Code-Matrix-Browser zur Visualisierung der Codierungen für Arbeitsmaterial 3a des Prüfkriteriums Reproduzierbarkeit.*

te und das Versuchsprotokoll damit als nicht-reproduzierbar erkannte. Die Kategorie *unvollständige/fehlende Durchführung* wurde von zehn Lernenden mit einem Beispiel untermauert. Größtenteils blieben die Argumente jedoch allgemeiner: Es wurde mit *unvollständiger Materialangabe*, *unzureichender Ausführlichkeit* oder *fehlender Durchführung* ohne Angabe eines Beispiels argumentiert.
Die Schülerinnen und Schüler argumentieren teils auch widersprüchlich. Sie begründen ihre Entscheidung, das Protokoll sei reproduzierbar damit, dass das Ergebnis erfolgreich war. Sie setzen damit das Gelingen eines Experiments mit Reproduzierbarkeit gleich. Das ist auch an der folgenden Aussage einer Schülerin erkennbar:„S 1: Das heißt, wenn wir den Versuch eben geschafft haben und sich der Propeller gedreht hat, können wir das im Grunde mit ja beantworten?“
Diese Aufgabe wurde von den Lernenden in manchen Gruppen ausführlich diskutiert:

S1: „Natürlich! Sie hat es nicht detailliert geschrieben, aber sie hat geschrieben was man tun soll“

S2: „Ja. Auch wenn es länger dauert“

S5: „Aber es fehlt ja in der Durchführung. Man muss ja schreiben was man macht damit es funktioniert. Aber in der Durchführung steht *mit der Zitrone*, nicht dass man den Zitronensaft pressen muss.“

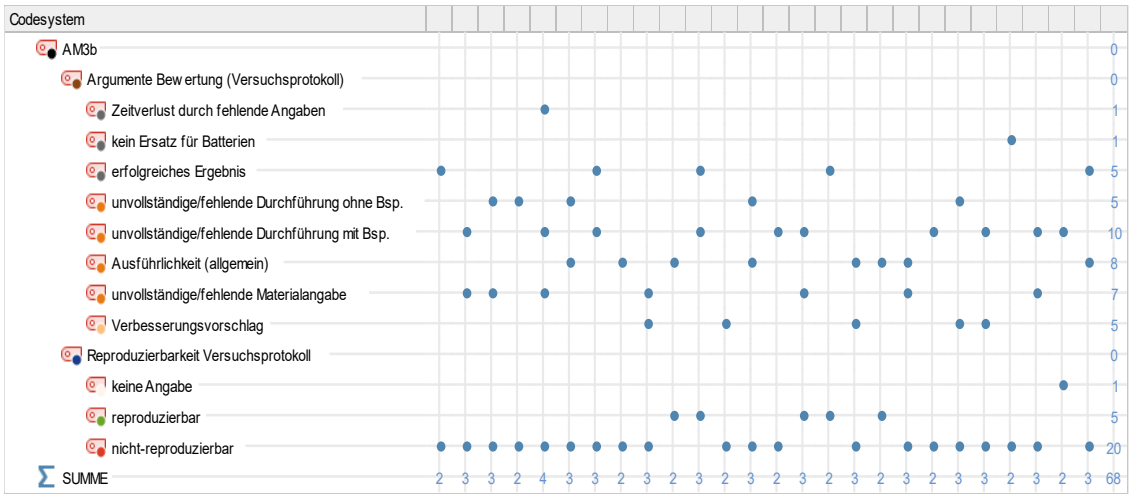


Abbildung 9.18: *Code-Matrix-Browser zur Visualisierung der Codierungen für Arbeitsmaterial 3b des Prüfkriteriums Reproduzierbarkeit.*

Ein Schüler hält daran fest, dass das Protokoll reproduzierbar ist, wenn mit genügend verfügbarer Zeit eine passende Vorgehensweise gefunden wird:

S 3: „Ich musste ja noch mehr machen als sie geschrieben hat."

S 1: „Aber grundsätzlich hat sie es beschrieben."

S 4: „Aber wenn du dumm bist, dann weißt du es ja nicht genau."

S 3: „Du musstest ja mehr machen als sie gesagt hat."

S 2: „Ja kommt drauf an, ob du halt Zeit hast oder nicht."

S 1: „Man müsste sich hier tatsächlich auf ein Mittel einigen."

S 2: „Wenn du Zeit hast, reicht das."

S 1: „Eigentlich schon, ne? Eigentlich müsste man sagen: Ja, es ist mittel-reproduzierbar, weil man ja selber auch noch was experimentieren muss, um das herauszufinden."

In dieser Gruppe wird also gefordert, eine mittlere Kategorie einzuführen, da für beide Seiten Argumente gefunden wurden. Im späteren Verlauf der Diskussion bilden sich immer deutlicher die Meinungen von Schüler:in 1 und 3 auf der einen Seite und die Meinung von Schüler:in 2 auf der anderen Seite heraus. S 1 scheint noch einmal über die Definition der Reproduzierbarkeit nachgedacht zu haben und erkennt plötzlich, dass das Versuchsprotokoll unvollständig ist. S 3 unterstützt:

S 3: „Weil wenn ich nur das Protokoll lese, dann ist es nicht-reproduzierbar. Wenn ich dann beschreiben soll wie das geht, wie ich es durchführe, dann würde ich es ja falsch machen. Weil ich würde den Saft ja dann in der Zitrone lassen und das klappt ja hier dann nicht."

Schüler 1 kommentiert die Diskussion mit einem Ausruf über seine kognitive Auslastung: „Oh Jungs in meinem Kopf zischt's schon!"

In anderen Wortbeiträgen innerhalb der Tischgruppe ist erkennbar, dass die Lernenden ihre Diskussion auf eine Meta-Ebene heraufsetzen und antizipieren, inwieweit die Reproduzierbarkeit des Versuchsprotokolls *für andere Lernende* gegeben ist. S 4 sagt, für sich selbst „sei es klar":

S 3: „Also ich habe das daran wirklich nicht verstanden, wie ich das machen soll. Also ist halt einfach so."

S 4: „Ich überlege halt, ob ich reproduzierbar oder nicht-reproduzierbar mache. Weil ich finde für mich war es klar, aber ich könnte mir vorstellen, dass andere damit Probleme haben, weißt du?"

S 3: „Ich finde es ist halt allgemein nicht reproduzierbar. Weil halt als Erstes: Hier werden noch nicht mal alle Materialien benannt. Wie willst du wissen, dass das irgendwas mit dem Becherglas ist? Natürlich, das steht in den Materialien, aber da kommst du doch nicht drauf."

S 4: „Ich habe halt geschrieben, dass die Durchführung zu ungenau ist und da sollte halt nicht nur Geldstück, sondern Kupfermünze stehen."

S 3: „Ja, aber dann kannst du das nicht so ankreuzen. Wenn du das ankreuzt und das so begründest, dann macht das keinen Sinn. Wirklich nicht."

S 4: „Ja, ich habe hier aber auch noch geschrieben, dass es für mich klar war."

S 3: „Also für mich war das nicht klar."

Diese Unterscheidung zwischen Reproduzierbarkeit „für sich selbst" und „für andere" entspricht nicht der fachlichen Vorstellung von Reproduzierbarkeit, ist allerdings ein nachvollziehbarer Gedanke.

Das Arbeitsmaterial 6 steht nicht mehr im Kontext der experimentellen Durchführung der Zitronenbatterie sondern fordert die Lernenden zur Bewertung der vorliegenden Studie zum Thema E-Scooter auf. Dazu bearbeiten die Schüler:innen eine

Freitext-Aufgabe, um anschließend erneut *reproduzierbar* oder *nicht-reproduzierbar* anzukreuzen. Die Antworten der Lernenden wurden anhand des induktiv gebildeten Kategoriensystems kodiert, der Code-Matrix-Browser ist in Abbildung 9.19 dargestellt.

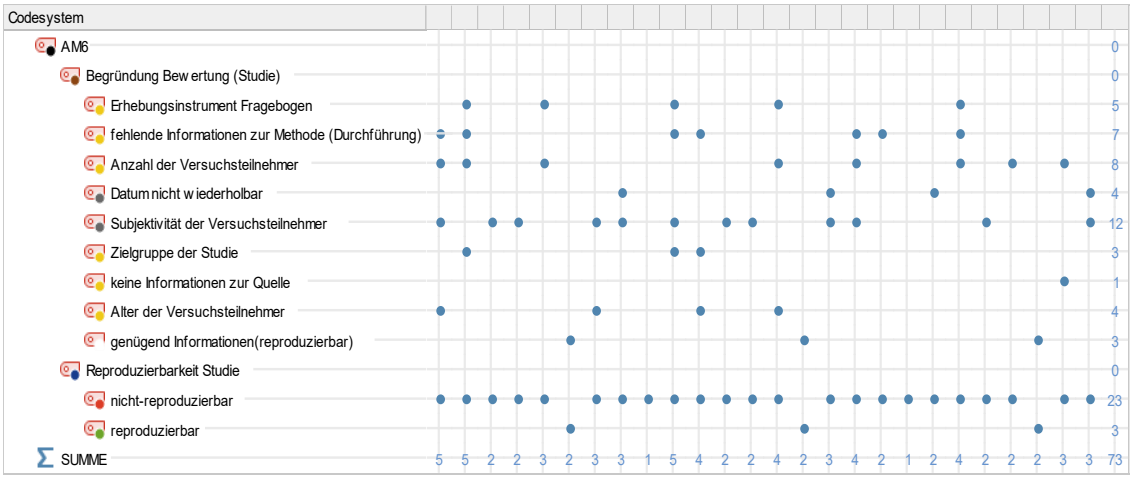


Abbildung 9.19: *Code-Matrix-Browser zur Visualisierung der Codierungen für Arbeitsmaterial 6 des Prüfkriteriums Reproduzierbarkeit.*

Fast alle Lernenden erkannten, dass die Studie nicht reproduzierbar gestaltet ist. Die drei Lernenden, die hier reproduzierbar wählten, untersuchte Banse (2020) genauer und erkannte anhand des Videomitschnitts, dass die Schüler:innen die Aufgabenstellung nicht konzentriert bearbeitet haben. Sie setzten ein Kreuz, ohne die Studie gelesen zu haben.

Unter den Begründungen befinden sich sechs fachlich korrekte Argumente: Lernende erkannten, dass keine Versuchsteilnehmer (8 Nennungen) oder Erhebungsinstrumente (5 Nennungen) genannt wurden und dass Durchführungsschritte fehlen (7 Nennungen). Sie benannten fehlende Informationen zu Alter der Teilnehmenden (4 Nennungen), Zielgruppe der Studie (3 Nennungen) oder Auftraggeber (1 Nennung). Auffällig ist, dass die Mehrheit der Lernenden auf die Diversität von persönlichen Meinungen einging. In 12 Fällen wurde als Argument angeführt, dass jeder Mensch eine eigene Meinung hat und diese Tatsache dazu führt, dass das Ergebnis nicht reproduzierbar ist. Weiterhin wurde genannt, dass „ein Datum nicht reproduzierbar ist“.

Wenngleich die letztgenannten Argumente nicht den Kern der Reproduzierbarkeit

treffen, zeigen sie doch sehr anschaulich, wie tiefgehend die Schüler:innen über das Konzept nachdenken und nach Begründungen suchen.

9.4.4 Diskussion und Re-Design

Die erprobte Unterrichtsstunde zeichnet sich durch eine hohe Schüleraktivität aus. Arbeitsaufträge wurden erfüllt, in den Plenumsphasen beteiligten sich die Lernenden und innerhalb der Schülergruppen entstanden zahlreiche Diskussionen zum Thema. Anhand der gewonnenen Daten lassen sich die zuvor gesetzten Lernziele überprüfen. Arbeitsmaterial 3a ruft dazu auf, Einflussfaktoren zu formulieren, die Einfluss auf ein Gelingen des Experiments haben. Die Analyse der Schülerantworten mithilfe des Code-Matrix-Browsers hat gezeigt, dass die Lernenden eine große Breite an Faktoren nennen. Ingesamt wurden elf Faktoren herausgearbeitet. Die Lernenden setzen sich demnach intensiv mit dem Thema auseinander und erfüllen das erste Lernziel. Banse (2020) fand in den Gesprächen der Schüler:innen keine Anzeichen auf Verständnisschwierigkeiten, daher ist für das AB 3a kein Re-Design notwendig.
Etwas anders sieht es bezüglich Arbeitsmaterial 3b aus. Bei der Bewertung des Versuchsprotokolls im Hinblick auf Reproduzierbarkeit reagierten die Lernenden widersprüchlich. Dabei scheint unklar gewesen zu sein, welcher Aspekt bewertet werden sollte. Manche Schüler:innen bewerteten das Gelingen des Experiments und setzten dieses mit Reproduzierbarkeit gleich. Um dieser Unklarheit zu begegnen, könnten die Arbeitsaufträge umformuliert und etwas offener gestaltet werden. Die Lernenden sollten zunächst das Versuchsprotokoll kritisieren und erst danach eine Einschätzung der Reproduzierbarkeit geben. Der Arbeitsauftrag lautet dann: „Berücksichtigt Hannah die Kriterien für Reproduzierbarkeit in ihrem Versuchsprotokoll? Entscheide dich für ein Fazit“.
Insgesamt kann das zweite Lernziel jedoch bei vielen Schüler:innen als erreicht angesehen werden, denn die Mehrheit stufte das Versuchsprotokoll korrekt ein und gab eine stimmige Begründung an.
Die Übertragung des Konzepts auf die Studie zu E-Scootern klappte sehr gut. Nahezu alle Schüler:innen äußerten begründete Bewertungen zur Reproduzierbarkeit der Umfrage. Allerdings waren einige Argumente nicht ganz zielführend (z. B. „jeder Mensch hat eine eigene Meinung“). Eine Sensibilisierung der Schülerinnen und Schüler dafür, dass wissenschaftliche Studien stets umfangreich dokumentiert werden

müssen, fand aber in jedem Fall statt. Die Lernenden erkannten, dass die Studie nicht ausreichend beschrieben wurde und legten diese Einschätzung ihren Bewertungen zugrunde.
Auf die Sensibilisierung kommt es letztendlich an - die genaue fachlich korrekte Begründung war hier nicht unbedingt notwendig zum Erreichen des letzten Lernziels. Dennoch könnte dem zuletzt beschriebenen Missverständnis noch entgegengewirkt werden, indem auf Arbeitsmaterial 4 die für die Studie relevanten Begriffe deutlicher hervorgehoben werden. Es sollte klar gemacht werden, dass Reproduktion von einer *anderen Person*, mit *anderen Versuchsteilnehmern* an einem *anderen Tag* wissenschaftlich gültig ist und zulässig für das Erfüllen der Reproduzierbarkeit.

Zusammenfassend konnte in der Zusammenarbeit mit Banse (2020) gezeigt werden, dass Schülerinnen und Schüler der untersuchten Lerngruppe durch die Thematisierung von Elektro-Scootern einen Zugang zur Bewertung der Versuchsprotokolle naturwissenschaftlicher Experimente erlangen und diese Protokolle auf Reproduzierbarkeit prüfen können. Auch das Übertragen auf empirisch-wissenschaftliche Studien zur Thematik fällt ihnen nicht schwer. Die mit den Lernmaterialien eingeführten Fachbegriffe finden sich in den Erklärungen der Lernenden wieder. Somit ist davon auszugehen, dass das Lernmaterial einen großen Beitrag zur erfolgreichen Prüfung von Experimenten und Studien auf ihre Reproduzierbarkeit leistet. Die Lernenden werden für die Einhaltung wissenschaftlicher Standards sensibilisiert - und nach dem Durchlaufen der hier beschriebenen Unterrichtseinheit können sie die *Reproduzierbarkeit* ihrem Repertoire an wissenschaftlichen Prüfkriterien hinzufügen.

10 Mesozyklus 2: *Nahrungsergänzungsmittel* als Kontroverse

10.1 Bewertung nach Kriterien eines guten Kontextes

Das Unterrichtskonzept choice^{2}reflect funktioniert Kontext-unabhängig. Es schult Bewertungskompetenz und soll dazu befähigen, bei gesellschaftlich kontrovers diskutierten Sachverhalten rationale und wissenschaftlich fundierte Entscheidungen zu treffen. Das dafür notwendige Wissen über die Prinzipien empirischer Wissenschaft erlangen die Lernenden in der Phase III, während die Phasen I und II sowie IV und V den kontextuellen Rahmen schaffen, in dem sich die Lernenden mit einem kontroversen Thema vertraut machen.

Jungkamp (2021) hat Lernmaterial zur Kontroverse *Homöopathie* entwickelt. Die Auswahl dieses Kontext wurde auf der Grundlage ausführlicher Überlegungen und Recherchen zu Vorgaben der Lehrpläne, Gütekriterien geeigneter Kontexte und Präsentationsmöglichkeiten getroffen. Dabei beschreibt Jungkamp in Form von *Essenzen*, welche Anforderungen ein geeigneter Kontext erfüllen muss. Er orientiert sich am gesellschaftskritisch-problemorientierten Unterrichtsverfahren, welches die Kriterien *Relevanz*, *Authentizität*, *Multidimensionalität*, *Bezug zur Chemie* und *Anschlussfähigkeit* beinhaltet (Marks et al. 2014). Jungkamp führt aber auch die damit einhergehenden Schwierigkeiten wie z. B. wachsende Emotionalisierung auf.

Die Essenzen stellen dar, dass *socio-scientific-issues* (SSI) nach Sadler & Donnelly (2006) den *Kriterien geeigneter Kontexte* gerecht werden und sich daher gut für den Einsatz in choice^{2}reflect eignen. Allerdings dürfen sie nicht durch das bloße Bereitstellen von Informationen präsentiert werden, sondern sollten sinnstiftend unter Berücksichti-

gung der Inhalts- und Bedeutungsdimension in den Unterricht integriert werden. Die Vermittlung sei so anzupassen, dass für die Lernenden eine „individuelle Bedeutsamkeit bzw. Betroffenheit“ (Jungkamp 2021, S. 66) hergestellt werde. Auch eine gesellschaftliche Bedeutung sollte den Lernenden verdeutlicht werden.

Weiterhin ist es wichtig, dass der Kontext *offen diskutierbar* ist, also dass keine eindeutig richtige Lösung zur Thematik existiert. Die alternierende Präsentation von Texten, die zu den Überzeugungen der Lernenden passen und Texten, die nicht passen, führe zu einer strategischeren und rationaleren Verarbeitung der Texte und sei daher anzustreben (Maier & Richter 2014).

Authentisch - relevant - multidimensional - anschlussfähig. Dabei auch zur Steigerung der Relevanz emotionsauslösend; aber nicht zu stark polarisierend, damit das Auftreten kognitiver Verzerrungen im Rahmen gehalten wird. Das ist eine ganze Reihe an Anforderungen, die an unterrichtliche Kontexte gestellt wird. Im Rahmen dieses Projekts ist die Wahl des Kontextes auf die Kontroverse *Nahrungsergänzungsmittel* gefallen.

Nahrungsergänzungsmittel, kurz NEM, erfreuen sich einer immer größeren Beliebtheit. Im Jahr 2020 wurden in Deutschland über 180.000 Tonnen produziert (Statistisches Bundesamt (Destatis) 2021), der Jahresumsatz betrug rund 1,35 Mrd. Euro. Schätzungen zufolge wird der Umsatz bis 2025 bis auf 1,7 Mrd. Euro ansteigen (Mansour 2021). Das zeigt, dass die Idee der Supplementierung einzelner Nährstoffe durch Kapseln, Tabletten, Pulver oder auch andere Darreichungsformen für viele Leute attraktiv erscheint. In der nationalen Verzehrsstudie (Max Rubner-Institut & Bundesforschungsinstitut für Ernährung und Lebensmittel 2008a) wurde herausgefunden, dass 27,6 % der Befragten regelmäßig NEM konsumieren. Diese Studie liegt bereits über 13 Jahre zurück, basierend auf den stark gestiegenen Absatzzahlen ist auch von einer erhöhten Einnahmeprävalenz auszugehen. Aktuelle Angaben zum Konsum aller NEM liegen derzeit nicht vor, aber allein der Konsum von Vitaminpräparaten (als Teil der Gruppe NEM) lag im Jahr 2020 bei 53 % der deutschen Bevölkerung. 33 % gaben an, Mineralstoffpräparate einzunehmen und lediglich 30 % der Befragten konsumierten keine NEM in den letzten 12 Monaten (Statista 2021). Diese Daten belegen deutlich die hohe gesellschaftliche *Relevanz* der Nährstoff-Supplementierung. Inwiefern NEM jedoch von Schüler:innen konsumiert werden und damit auch für sie eine persönlich-individuelle Relevanz besitzen, gilt es zu ermitteln. Dieser Aufgabe

widmet sich das nächste Kapitel 10.2.

Authentisch ist die Thematik der NEM in jedem Fall. Die mannigfaltigen gesundheitsbezogenen Angaben, die von einigen NEM-Herstellern auf ihre Produkte gedruckt oder in Werbung integriert werden, dürften den meisten Lernenden bereits begegnet sein. Das Material muss daher für den Einsatz im Unterricht nicht erst entworfen, konstruiert oder stark verändert werden und wird damit den Anforderungen der Authentizität gerecht. Wird das Material dann auch angemessen und realitätsnah präsentiert, erhöht dies die Glaubwürdigkeit auf Schülerebene erneut (Glaesser et al. 2009).

Die Fachinhalte, die im Zusammenhang mit NEM erarbeitet werden, können als *anschlussfähig* angesehen werden, denn NEM und deren Wirkungen sind unmittelbar gesundheitsbezogen. Im Zuge des Durchlaufens der Konzeption choice^{2}reflect lernen die Schüler:innen, die gesundheitsbezogenen Angaben von NEM-Herstellern auf Wissenschaftlichkeit zu prüfen. Weiterhin bewerten sie die Einstufung von NEM als „Lebensmittel“ gemäß Lebensmittelinformationsverordnung (LMIV) und beschäftigen sich mit der Versorgungslage mit Nährstoffen in der deutschen Bevölkerung. Die damit erworbenen Kompetenzen sind nicht spezifisch für NEM, denn auch in anderen gesundheitsbezogenen Bereichen helfen sie den Lernenden weiter, bspw. bei der Bewertung von medialen Schlagzeilen bezüglich der Corona-Pandemie. Wenn dort von „möglichen Spätfolgen einer Impfung“ berichtet wird, können die Lernenden die im Zusammenhang mit NEM erworbenen Kenntnisse zur Bewertung einsetzen. Der Kontext NEM ist also anschlussfähig.

Das Thema Nahrungsergänzung weist auch eine *Multidimensionalität* (Feierabend & Eilks 2010, S. 179; Marks et al. 2014)auf. Das Thema wird anhand des Prüfschemas zwar eher aus einer rein naturwissenschaftlichen Sicht betrachtet, bietet jedoch Spielraum für individuelle Implikationen. Unter Berücksichtigung ökonomischer Faktoren, z. B. der Gewinnabsicht der Hersteller, kann sich ein anderes Entscheidungsverhalten ergeben. Dadurch entsteht die ebenfalls geforderte *offene Diskutierbarkeit* und eine eindeutig richtige Einstellung gegenüber NEM kann nicht definiert werden. Der in choice^{2}reflect eingesetzte Kontext sollte weiterhin einen Bezug zu Naturwissenschaften bzw. zu den Inhalten der Curricula in naturwissenschaftlichen Schulfächern haben. Der Kernlehrplan für die gymnasiale Sekundarstufe I in Nordrhein-Westfalen im Fach Chemie (KLP 2019) unterscheidet Inhaltsfelder von Kompetenzbereichen und führt beides zu übergeordneten Kompetenzerwartungen zusammen

(siehe dazu auch Kapitel 3.4.2). Das Thema Nahrungsergänzungsmittel ist nicht klar einem Inhaltsfeld zuzuordnen, touchiert aber Bereiche aus verschiedenen Inhaltsfeldern:

- Inhaltsfeld 1: Stoffe und Stoffeigenschaften

  „Ein fundiertes Wissen über Einsatzbereiche, Anwendungen und mögliche Gefahren verschiedener Stoffe ist Voraussetzung, um beim alltäglichen Konsum sinnvolle Entscheidungen zu ihrer Verwendung treffen zu können". Diese Formulierung des KLP zeigt, dass NEM für das Inhaltsfeld 1 relevant sind und deren Konsum im Alltag reflektiert werden sollte.

- Inhaltsfeld 5: Elemente und ihre Ordnung

  Viele NEM, vor allem im Bereich der Mineralstoffpräparate, bestehen aus oder enthalten Mineralsalzverbindungen, also Metall-Kationen mit ihren Gegenionen. Für Lernende ist es wichtig zu erkennen, welchen Platz das jeweilige Metall im Periodensystem einnimmt.

- Inhaltsfeld 6: Salze und Ionen

  Es wird im KLP expliziert, dass „die richtige Dosierung und Zusammensetzung von Salzgemischen bezüglich der Gesunderhaltung" von entscheidender Bedeutung sind. Wenn sich Lernende mit NEM beschäftigen, beziehen sie Angaben zu Dosierung und Zusammensetzung in ihre Überlegungen mit ein. Konkretisiert wird im KLP: „Die Schülerinnen und Schüler können unter Umwelt- und Gesundheitsaspekten die Verwendung von Salzen im Alltag reflektieren (B1)".

Als gesellschaftlich kontrovers diskutiertes Thema bieten NEM auch Ansatzpunkte für die Vermittlung der übergeordneten Kompetenzerwartungen, doch die Schulung von Bewertungs-, Kommunikations- und Erkenntnisgewinnungskompetenzen hängt vor allem von der Art der Präsentation im Unterricht und der Vorbereitung durch die Lehrkraft ab. Es kann also nicht per se festgelegt werden, dass „NEM in Einklang mit den übergeordneten Kompetenzerwartungen" stehen - erst eine passende Einbindung in den Unterricht sorgt für diesen Einklang.

10.2 Fragebogenerhebung zu Nahrungsergänzungsmitteln

53 % von 1041 Befragten einer Studie nahmen im Zeitraum von Mitte 2020 bis Mitte 2021 Vitamin-Präparate zu sich, etwa ein Drittel konsumierte Mineralstoffpräparate (Statista 2021). Damit ist ein deutlicher Anstieg im Vergleich zu den Vorjahren zu verzeichnen - Nahrungsergänzung ist für viele Deutsche ein aktuelles Thema. Die zitierten Studien blicken dabei jedoch hauptsächlich auf Erwachsene, über den Konsum von NEM unter Jugendlichen gibt es kaum belastbare Ergebnisse. Aus diesem Grund werden im Rahmen einer Fragebogenerhebung Schüler:innen der achten bis zehnten Klassenstufen zu ihren Einnahmegewohnheiten und Einnahmegründen befragt. Es soll ermittelt werden, wie häufig die Jugendlichen NEM konsumieren und welche Präparate dabei bevorzugt werden. Weiterhin werden sie zu den erhofften Wirkungen der Präparate befragt. Da innerhalb der Gesellschaft viele verschiedene Positionen und Meinungen zu NEM bestehen, bewerten die Lernenden in einem Abschnitt des Fragebogens kontroverse Aussagen zur Thematik. Aus den dadurch gewonnenen Antworten lässt sich ein Meinungsbild generieren und eine Aussage darüber treffen, ob Lernende der Nahrungsergänzung kritisch gegenüberstehen.

10.2.1 Entwicklung des Fragebogens

Fragestellung

Richtung der Analyse

Im Vordergrund der Analyse stehen die Vorerfahrungen von Schülerinnen und Schülern der Sekundarstufe I mit dem Thema Nahrungsergänzungsmittel. Durch die Analyse der Schülerantworten werden Aussagen über das Konsumverhalten, das Vorwissen über Wirkungen von Nahrungsergänzungsmitteln und die persönlichen Einstellungen gegenüber NEM getroffen.

Theoriegeleitete Differenzierung der Fragestellung

Die Nationale Verzehrs-Studie II aus dem Jahr 2008 hat gezeigt, dass etwa 27 % der Deutschen zumindest gelegentlich NEM konsumieren (Max Rubner-Institut & Bundesforschungsinstitut für Ernährung und Lebensmittel 2008b). Andere Untersuchun-

gen berichten von Werten zwischen 18 und 50 %, dabei unterscheiden sich z. B. die Altersstufen der Befragten oder die zugrundeliegenden Definitionen für Nahrungsergänzungsmittel (Reinert et al. 2007; Schwab et al. 2014; Heinemann et al. 2015). Die aktuellsten, oben bereits zitierten Daten von Statista (2021) ergeben noch höhere Einnahmeprävalenzen. Den größten Anteil der supplementierten Produkte bilden Magnesium-, Calcium- und Vitamin-C-Präparate. Im Hinblick auf diese Datenlage ist es nun von Interesse, wie die genannten Faktoren bei SuS ausgeprägt sind. Es soll vor allem untersucht werden, wie hoch der Anteil Jugendlicher ist, die NEM konsumieren, welche Produkte dabei bevorzugt werden und welche Wirkungen sich die SuS erhoffen. Darüber hinaus werden die persönlichen Einstellungen der SuS gegenüber NEM analysiert.
Im Mittelpunkt der Datenauswertung stehen daher folgende Fragestellungen:

- Fragestellung 1: Mit welcher Intention werden welche NEM von Schülerinnen und Schülern der gymnasialen Sekundarstufe I (Jahrgang 8 bis 10) konsumiert?
- Fragestellung 2: Wie bewerten SuS verschiedene Aspekte zu NEM im Zusammenhang mit der eigenen Ernährung?

Itementwicklung

Der Fragebogen (siehe Anhang B.8.1) enthält neun Items, die das Konsumverhalten, Vorwissen und persönliche Einstellungen der Teilnehmer im Bezug auf NEM untersuchen. Ihm liegt eine Informationsseite mit der Definition von NEM sowie einigen Beispielen für verschiedene Kategorien von NEM bei, damit die Befragten eine einheitliche Vorstellung des Begriffs Nahrungsergänzungsmittel bekommen. Der Einstieg in den Fragebogen mit der offenen Frage: „Wo sind dir NEM bereits begegnet?“ zielt darauf ab, Alltagserfahrungen der Schülerinnen und Schüler mit NEM zu erfassen. Die Frage enthält keinen besonderen Fokus und ist allgemein gehalten, um nicht abschreckend zu wirken bzw. keine zu persönlichen Informationen direkt am Anfang zu erfragen. Anschließend werden die Teilnehmer um eine Einschätzung gebeten, welche Wirkung NEM aus ihrer Sicht haben. Die Frage: „Was bewirken NEM?“ soll mit einem Freitext beantwortet werden.
Das dritte Item dient der Erhebung aller NEM, die von den Fragebogenteilnehmern eingenommen werden. Ein kurzer Informationstext erklärt die untenstehende Tabelle: Produktname und die Häufigkeit der Einnahme werden dort notiert. Zur Einstu-

fung der Häufigkeiten wird eine vierstufige Skala verwendet („täglich", „mehrmals pro Woche", „einmal pro Woche" und „nach Bedarf").
Das folgende Item erfragt die Bezugsquellen der konsumierten NEM: „Woher bekommst du bzw. bekommt deine Familie die NEM?" Die Teilnehmer können folgende Antwortmöglichkeiten ankreuzen: „Apotheke", „Drogeriemarkt", „Supermarkt", „Internethandel", „Weiß nicht" und „Sonstige"; Mehrfachantworten sind möglich. Item 5 enthält wieder eine Tabelle, in der für jedes konsumierte Produkt die erhoffte Wirkung eingetragen wird.
Interessant ist auch, wo sich die NEM-Konsumierenden über die Produkte informieren (Item 6) und wie stark die Konsumenten selbst den Einfluss der jeweiligen Faktoren auf das eigene Konsumverhalten einschätzen (Item 7). Als Informationsquellen bzw. Einflussfaktoren können die Teilnehmer folgendes angeben: Arzt, Eltern, Freunde, Trainer, Werbung, Bücher/Zeitschriften, Schule und soziale Medien. Außerdem gibt es die Möglichkeit, weitere Personen/Faktoren zu ergänzen. Item 7 hält eine vierstufige endpunktverbalisierte Skala von „Kein Einfluss" bis „Starker Einfluss" bereit; auch „Keine Angabe" ist möglich.
Auf der nächsten Seite des Fragebogens befindet sich Item 8. Dabei handelt es sich um eine Auflistung elf verschiedener Aussagen zur Thematik der NEM. Dabei werden verschiedene Aspekte thematisiert. Die Aussagen a. bis c. (siehe Auflistung unten) beziehen sich auf Wirkungen von NEM, die von Teilnehmern anderer Studien häufig genannt wurden (Rehaag et al. 2013). Aussagen d. und e. verlangen von den Schülerinnen und Schülern eine Einschätzung, ob NEM dem Körper schaden können. Anhand der Aussagen f. und k. wird überprüft, ob NEM von den Lernenden eher als Lebensmittel oder als Medikament angesehen werden, denn gemäß §1 der Nahrungsergänzungsmittelverordnung (NemV) handelt es sich um Lebensmittel. NEM unterliegen dabei den gültigen Vorschriften zur Kennzeichnung von Lebensmitteln, nicht jedoch denen für Medikamente. Die Aussagen g. bis j. decken Ernährungsaspekte ab: Die Teilnehmer sollen einschätzen, ob eine Supplementierung zusätzlich zu einer gesunden Ernährung nötig ist und ob in der deutschen Bevölkerung ein Vitaminmangel herrscht.
Die Aufgabenstellung lautet: „Bitte kreuze an, wie sehr du der jeweiligen Aussage zustimmst. Beziehe dich dabei auf einen durchschnittlich gesunden Menschen in Deutschland". Für jede Aussage können die Teilnehmer ihre Zustimmung auf einer vierstufigen endpunktverbalisierten Skala von „Stimme überhaupt nicht zu" bis „Stimme voll

und ganz zu“ angeben oder „Weiß nicht“ wählen. Folgende Aussagen sollen bewertet werden:

a. NEM können die Gesundheit verbessern.
b. NEM können die Fitness verbessern.
c. NEM verhindern, dass es dem Körper an Nährstoffen mangelt.
d. Je mehr NEM der Körper bekommt, desto besser ist es für ihn.
e. NEM können dem Körper nicht schaden.
f. NEM sind für mich Medikamente.
g. Man kann den Bedarf an Vitaminen und Mineralstoffen allein durch eine gesunde Ernährung decken.
h. Zusätzlich zur gesunden Ernährung sollte man auch NEM nehmen.
i. Wenn man sich mal ungesund ernährt, lässt sich das durch NEM ausgleichen.
j. Die meisten Deutschen haben einen Vitaminmangel.
k. NEM sind für mich Lebensmittel.

Item 9 enthält folgende Aufgabenstellung: „Bewerte auf der Skala: Aus deiner Sicht sind NEM...“. Es folgen vier Begriffspaare, zwischen denen in vier Stufen gewählt oder „Weiß ich nicht“ angegeben werden kann:

- ungesund - gesund
- gefährlich - ungefährlich
- teuer - günstig
- sinnlos - sinnvoll

Abschließend werden durch den Fragebogen einige demographische Daten (wie Geschlecht, Jahrgang und Alter) der Schülerinnen und Schüler sowie die Häufigkeit von sportlicher Aktivität erhoben. NEM sind besonders bei Sportlern beliebt und es gibt in diesem Bereich viele Einsatzmöglichkeiten für NEM, daher kann es von Interesse sein, die sportliche Aktivität der Lernenden mit deren NEM-Konsum in Beziehung zu setzen.

Statistische Auswertung

Die Auswertung der erhobenen Daten erfolgte mittels Microsoft Excel 2016 für Mac. Die Analyse ist nicht auf die Hypothesenbildung ausgerichtet, sie dient lediglich einer Bestandsaufnahme des Umgangs von Lernenden der Sekundarstufe I mit NEM gemäß den im Abschnitt 10.2.1 formulierten Fragestellungen. Daher werden bei der Auswertung der Daten Methoden der deskriptiven Statistik angewendet, sodass eine Angabe von Häufigkeiten und Mittelwerten erfolgt. Offene Fragebogenitems wurden zusätzlich mithilfe von induktiver Kategorienbildung und anschließender leitfadengestützter Codierung qualitativ ausgewertet.

10.2.2 Ergebnisse

Charakterisierung der Stichprobe

Im Rahmen der Fragebogenerhebung wurden 236 Schülerinnen und Schüler von zwei Gymnasien im Stadtgebiet Münster befragt. Davon sind 162 weiblich (70 %), 57 männlich (24 %) und 14 Teilnehmer (6 %) machten keine Angabe zu ihrem Geschlecht. Die Stichprobe setzt sich aus 104 Schülerinnen und Schülern (44 %) der 8. Jahrgangsstufe, 102 (44 %) der 9. sowie 28 Schülerinnen und Schüler (12 %) der 10. Jahrgangsstufe zusammen. Etwa die Hälfte der Teilnehmer (111, 48 %) war zum Zeitpunkt der Erhebung 14 Jahre alt, 75 Teilnehmer (32 %) 13 Jahre und 39 Teilnehmer (17 %) 15 Jahre alt. 7 Teilnehmer (3 %) waren älter oder jünger.

Vorwissen und Alltagserfahrungen

Mit den Items eins und zwei werden Alltagserfahrungen und das Vorwissen zu NEM erhoben. 26 % der Befragten nannten im Freitextfeld der ersten Frage mindestens ein Produkt, das als NEM klassifiziert wird. 19 % gaben an, NEM würden ihnen in Nahrungsmitteln begegnen, dabei vor allem in Fruchtsäften. 14 % kennen NEM aus dem Einzelhandel, 12 % bringen sie mit Sport in Verbindung. Außerdem wurden die eigene Familie (7 %) und Freunde (7 %) genannt, die Erfahrungen mit NEM besitzen. Werbung und soziale Medien wurden nur selten erwähnt. Lediglich 3 % der Befragten gaben an, keine Erfahrung mit NEM zu haben.
21 % der Antworten auf die Frage: „Was bewirken NEM?“ gaben an, NEM gleichen Mängel aus. Mit je 11 % wurde genannt, dass NEM dem Körper Energie geben und

dass sie die allgemeine Gesundheit verbessern. Außerdem wurden NEM Wirkungen auf die körperliche Fitness zugeschrieben: Sie seien für Muskelaufbau (8 %) und zusätzliche Kraft (7 %) verantwortlich. Insgesamt behaupten 48 % der Antworten, dass NEM kurativ wirken; 33 % beziehen sich auf die körperliche Fitness und 4 % auf eine präventive Wirkung. 15 % der Antworten enthalten unspezifische Nennungen von Wirkungen.

Einnahmeprävalenz und Verwendungsgewohnheiten

Die Auswertung der Antworten zu Item drei hat ergeben, dass 61 % der Befragten mindestens ein NEM nach Bedarf oder häufiger verwenden. Dabei besteht ein geringer Unterschied zwischen den Geschlechtern: Während 60 % der Schülerinnen NEM konsumieren, liegt der Konsum bei Schülern mit 63 % etwas höher. 6 % der Befragten, die NEM konsumieren, gaben kein Geschlecht an.
Die mit 35 % aller Nennungen von den Befragten am häufigsten konsumierte Nährstoffgruppe ist die Gruppe der Mineralstoffe, wobei 21 % aller Nennungen auf den Mineralstoff Magnesium fallen, 6 % auf Eisen, 4 % auf Calcium und 2 % auf sonstige Mineralstoffe. Weitere 2 % nannten nur den Begriff Mineralstoffe, spezifizierten die Einnahme aber nicht.
Bei 29 % der Nennungen handelt es sich um Vitaminpräparate. Meist wurde kein konkretes Vitamin genannt (14 %), teilweise wurden Vitamin C (8 %), Vitamin D (6 %) und sonstige Vitamine (2 %) erwähnt.
Proteine (6%) und Energie-Drinks bzw. -Riegel (je 5 %) spielen ebenfalls eine Rolle bei der Supplementierung. Insgesamt wurden 25 verschiedene Produkte, die als NEM klassifiziert werden, genannt. Häufig erwähnt wurden Fruchtsäfte (13 %), diese fallen jedoch nicht unter die Definition von NEM und werden daher nicht weiter betrachtet.
Die Mehrheit der NEM wird nur nach Bedarf konsumiert (60 %). 15 % werden täglich, 13 % mehrmals pro Woche und 8 % einmal pro Woche eingenommen.

Bezugsquellen

Aus der Gruppe der 236 befragten Lernenden gaben 172 an, woher sie Nahrungsergänzungsmittel beziehen. Die am häufigsten genannte Bezugsquelle ist mit 55 % die Apotheke, gefolgt von Drogerie- und Supermarkt (je 45 %). 8 % der Lernenden

besaßen keine Kenntnis über die Bezugsquellen, je 5 % beziehen NEM aus dem Internet oder sonstigen Quellen. Als sonstige Quellen werden Sporttrainer, Reformhaus, Reitsportfachhandel und Arzt genannt.

Einnahmegründe

Magnesium als der am häufigsten supplementierte Nährstoff wird von den Befragten hauptsächlich eingenommen, um Muskelkrämpfe und -schmerzen zu behandeln oder diesen vorzubeugen (70 % der Nennungen für Magnesium). Außerdem sollen durch die Supplementierung von Magnesium weitere Effekte wie die Steigerung der körperlichen Fitness (8 %), ein Energiegewinn (6 %), Verbesserung äußerlicher Körpermerkmale (6 %), positive Effekte auf das Blut (4 %) oder das Immunsystem (2 %) sowie der Ausgleich eines Mangels (4 %) hervorgerufen werden.
Proteine werden fast ausschließlich zur Steigerung der körperlichen Fitness (94 %) und für einen Energiegewinn (6 %) supplementiert.
Vitamin C wird zur Vorbeugung bzw. Behandlung von Mangelerscheinungen (33 %), zur Verbesserung von Haut und Nägeln (33 %) sowie zur Stärkung des Immunsystems (17 %) supplementiert. Einige Befragte erwarten einen besseren Zellschutz (8 %) sowie eine bessere körperliche Fitness (8 %).

Informationsverhalten

Die wichtigsten Quellen, die rund zwei Drittel der Lernenden nutzen, um sich über Nahrungsergänzungsmittel zu informieren, sind die eigenen Eltern (39 %) und der behandelnde Arzt (28 %). Das weitere Drittel entfällt auf die übrigen Antwortmöglichkeiten: Je 8 % informieren sich in sozialen Medien und bei dem eigenen Sporttrainer, 6 % gaben an, Informationen durch die Schule zu erhalten. Je 4 % der Befragten nutzen Freunde bzw. Bücher als Informationsquelle, 3 % informieren sich durch Werbung. Unter dem Punkt „Sonstiges“ (5 %) wurden Apotheke, Internet-Webseiten, die eigene Familie, Lebensmittelverpackungen, Radio und Psychotherapeutin genannt.

Risikoeinschätzung

Der Fragebogen enthält unter Item 8 elf verschiedene Aussagen zur Thematik der NEM. Die Befragten werden aufgefordert, zu jeder dieser Aussagen auf einer vier-

stufigen endpunktverbalisierten Skala zwischen „stimme überhaupt nicht zu“ (zugeordneter Wert 0) und „stimme voll und ganz zu“ (zugeordneter Wert 3) eine eigene Einschätzung zu geben. Zur Auswertung der Antworten wurde für jede Aussage das arithmetische Mittel über alle Antworten gebildet, sodass der durchschnittliche Standpunkt gegenüber der Aussage deutlich wird. Die dazu errechnete Standardabweichung für die Grundgesamtheit zeigt, wie sehr die Meinungen innerhalb der Antworten um den Mittelwert streuen. Die errechneten Werte sind in Tabelle 10.1 angegeben.

Mit einem Mittelwert kleiner als 1 erfahren die Aussagen d, e, f, h, i und k eine große Ablehnung. Im Bereich zwischen leichter Ablehnung und leichter Zustimmung liegen die Aussagen b und j. Die Lernenden stimmen den Aussagen a, c und g mit Mittelwerten über 2 stark zu.

Die Standardabweichungen erlauben eine Einschätzung der Heterogenität der Meinungslage unter den Befragten. Am wenigsten Streuung herrscht bei den Aussagen a, b, c und d (St.abw zwischen 0,68 und 0,78), stärker streuen die Antworten bei e, g, h und k (0,82 bis 0,86) und die höchsten Streuungen ergeben sich für f (0,96), i (0,94) und j (1,13).

Aussagekräftiger werden diese Ergebnisse jedoch, wenn ein detaillierterer Blick auf die Verteilung der einzelnen Antworten geworfen wird - denn diese Verteilung wird durch Mittelwerte nicht abgebildet. Dazu zeigt das Diagramm in Abbildung 10.1 für jede mögliche Antwort (0 bis 3) den relativen Anteil der Optionen an den Gesamtantworten. Die fettgedruckte Linie zeigt die Abgrenzung zwischen den unteren und den oberen beiden Segmenten der Säulen. Die unteren Segmente stellen Zustimmung dar, die oberen Segmente Ablehnung. Liegt die fettgedruckte Linie also unter 50 %, dann haben mehr als die Hälfte der Lernenden „Stimme überhaupt nicht zu“ oder „Stimme eher nicht zu“ gewählt. Liegt die Linie oberhalb von 50 %, wurde häufiger „Stimme voll und ganz zu“ oder „Stimme eher zu“ gewählt.

Die Unterschiede zwischen der Mittelwertsbetrachtung und der Darstellung in Abbildung 10.1 zeigen sich in den Aussagen d, i und k: Während die fettgedruckten Linien im Verteilungs-Diagramm auf nahezu derselben Höhe liegen (um 15 %) und damit starke Ablehnung ausgedrückt wird, unterscheiden sich die *Mittelwerte* der Aussagen mit 0,71 zu 0,58 zu 0,57. Mit dem Verteilungs-Diagramm ist es besser möglich, bei allgemeiner Zustimmung laut Mittelwert zwischen „Stimme eher zu“ und „Stimme voll und ganz zu“ zu unterscheiden.

Den Aussagen a, b, c und g wurde mit mehr als 75 % aller Antworten stark zugestimmt. Dabei haben c und g den größten Anteil an 3 („Stimme voll und ganz zu“) und die Zustimmung bei a und b ergibt sich eher aus der Wahl von 2 („Stimme eher zu“). In der Klasse herrscht also ein großer Konsens darüber, dass NEM Nährstoffmangel verhindern und dass eine gesunde Ernährung reicht, um den Bedarf an Vitaminen und Mineralstoffen zu decken.

Etwas unsicherer bei insgesamt ähnlich hohem Zustimmungsanteil sind sich die Befragten bei der Wirkung von NEM auf Gesundheit und Fitness: Beides kann ihrer Meinung nach durch NEM verbessert werden.

Große Unsicherheit herrscht bezüglich der Aussage j: „Die meisten Deutschen haben einen Vitaminmangel“. Der Mittelwert liegt genau zwischen den Extrema und die Standardabweichung demonstriert hohe Streuungen. Auch das Verteilungs-Diagramm zeigt, dass nur selten volle Zustimmung oder Ablehnung gewählt wurde.

Die meisten Aussagen des Item 8 wurden von den Befragten abgelehnt. Das Verhältnis von leichter Ablehnung (Wert 1) zu starker Ablehnung (Wert 0) unterscheidet sich jedoch. Bei d, e und f wurde nahezu gleich häufig der Wert 0 vergeben (etwa 45 %), Wert 1 wurde für Aussagen d und e jedoch häufiger gewählt als für f, wodurch sich eine insgesamt höhere Ablehnung ergibt. Thematisch liegen d und e auch nah beieinander. Die Befragten glauben insgesamt nicht, dass es umso besser für den Körper ist, je mehr NEM diesem zugeführt werden. Sie erkennen, dass NEM auch schädlich für den Körper sein können - wobei ein hoher Anteil der Ablehnung aus dem Wert 1 (nur leichte Ablehnung) stammt.

Die Ablehnung der Aussage f: „NEM sind für mich Medikamente“ setzt sich zu einem größeren Anteil aus Extrema zusammen. Die Werte 0 und 3 werden häufiger gewählt als 1 und 2. Noch größer ist dieser Effekt bei der kontrastierenden Aussage k („NEM sind für mich Lebensmittel“), denn dort ist der Anteil der Werte 1 und 2 besonders klein.

Aussage h, als Gegenpol zu g, erfragt eine Einschätzung, ob zusätzlich zur gesunden Ernährung auch NEM genommen werden sollten. Die starke Ablehnung ist etwas unsicherer als beispielsweise bei k, denn der Anteil an 1 und 2 ist deutlich größer.

Die Aussage i erfährt zusammen mit k die größte Ablehnung, die auch ähnlich einstimmig gewählt wird. Der Standpunkt: „Wenn man sich mal ungesund ernährt, lässt sich das durch NEM ausgleichen“ wird also von den Befragten nicht vertreten.

Tabelle 10.1: *Mittelwerte (MW) und Standardabweichungen (St.abw) über alle Einschätzungen zu den Aussagen a. bis k. des Items 8. 4-stufige Skala: 0 = stimme überhaupt nicht zu, 1 = stimme eher nicht zu, 2 = stimme eher zu, 3 = stimme voll und ganz zu.*

	Aussage	**MW**	**St.abw**
a.	NEM können die Gesundheit verbessern.	2,141	0,685
b.	NEM können die Fitness verbessern.	1,956	0,741
c.	NEM verhindern, dass es dem Körper an Nährstoffen mangelt.	2,304	0,785
d.	Je mehr NEM der Körper bekommt, desto besser ist es für ihn.	0,707	0,760
e.	NEM können dem Körper nicht schaden.	0,835	0,860
f.	NEM sind für mich Medikamente.	0,912	0,959
g.	Man kann den Bedarf an Vitaminen und Mineralstoffen allein durch eine gesunde Ernährung decken.	2,203	0,819
h.	Zusätzlich zur gesunden Ernährung sollte man auch NEM nehmen.	0,926	0,848
i.	Wenn man sich mal ungesund ernährt, lässt sich das durch NEM ausgleichen.	0,577	0,943
j.	Die meisten Deutschen haben einen Vitaminmangel.	1,700	1,130
k.	NEM sind für mich Lebensmittel.	0,573	0,859

allgemeine Bewertung von NEM

Item 9 enthält vier Begriffspaare, die die Befragten auf einer vierstufigen Skala bewerten. Die linksstehenden Begriffe (ungesund, gefährlich, teuer, sinnlos) werden je mit dem Wert 1 codiert, während die rechtsstehenden Gegenteile (gesund, ungefährlich, günstig, sinnvoll) dem Wert 4 zugeordnet werden. Die über alle Befragten errechneten Mittelwerte und Standardabweichungen sind in Tabelle 10.2 dargestellt.
In der Tabelle ist erkennbar, dass die Mittelwerte aller vier Begriffspaare zwischen 2,0 und 3,0 liegen. Die Paare a, b und d liegen sogar zwischen 2,5 und 3, während c bei etwa 2,3 liegt. Damit gibt es zu keiner der Begriffspaare (im Mittel) ein klares Meinungsbild.
Um die tatsächliche Meinungslage in der Gruppe der Befragten genauer untersuchen zu können, wird eine weitere Abbildung herangezogen. Diese ist angelehnt an das bereits in Item 8 verwendete Anteils-Diagramm. In diesem Fall werden horizontale Bal-

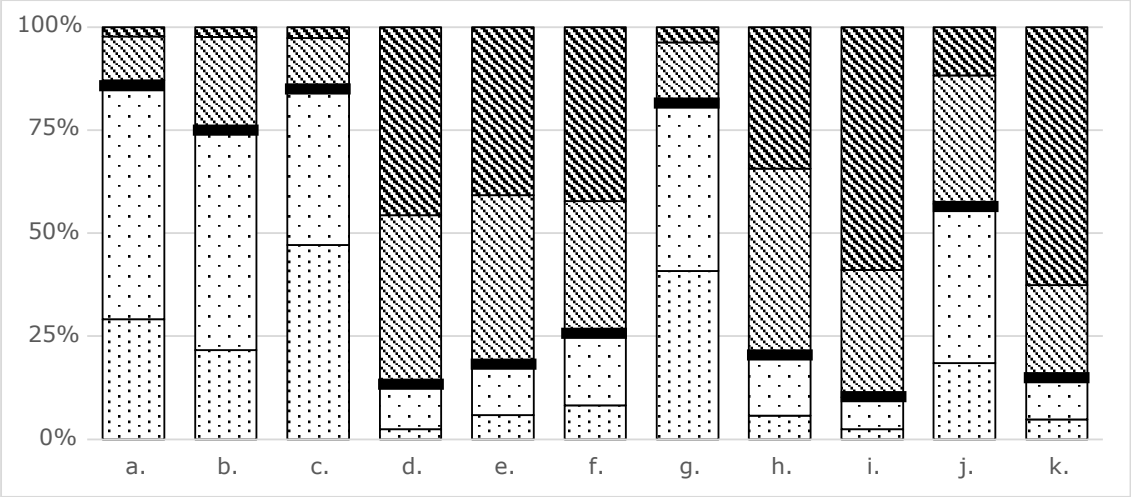


Abbildung 10.1: *Darstellung der relativen Anteile der Antwortmöglichkeiten 0 bis 3 bei Item 8 des Fragebogens. Fettgedruckte Linie als Abgrenzung zwischen Zustimmung und Ablehnung. Dicht gepunktet = „Stimme voll und ganz zu", leicht gepunktet = „Stimme eher zu", leicht schraffiert = „Stimme eher nicht zu", dicht schraffiert = „Stimme überhaupt nicht zu".*

Tabelle 10.2: *Mittelwerte (MW) und Standardabweichungen (St.abw) über alle Bewertungen der Begriffspaare a. bis d. des Items 9. 4-stufige endpunktverbalisierte Skala: 1 = linksstehender Begriff, 4 = rechtsstehender Begriff.*

	Begriffspaar	**MW**	**St.abw**
a.	ungesund - gesund	2,871	0,671
b.	gefährlich - ungefährlich	2,750	0,864
c.	teuer - günstig	2,328	0,911
d.	sinnlos - sinnvoll	2,925	0,844

ken dargestellt, deren Aufteilung die relativen Anteile der gewählten Optionen visualisiert. Dicht schraffierte Balkenabschnitte zeigen die jeweiligen Anteile der Option 4, also desr rechtsstehenden Begriffe der Paare. Leicht schraffierte Abschnitte zeigen die Anteile der dritten Option, leicht gepunktete zeigen die Anteile der zweiten Option und dicht gepunktete Bereiche stellen dar, wie häufig der linksstehende Begriff gewählt wurde. Eine fettgedruckte Linie verdeutlicht, ob die Mehrheit der Befragten eher zum linken oder zum rechten Begriff tendiert.

Bereits bei der ersten Betrachtung des Diagramms fällt auf, dass die Kernaussage der Mittelwerte durch das Anteils-Diagramm nicht bestätigt wird. Bei den Begriffspaaren

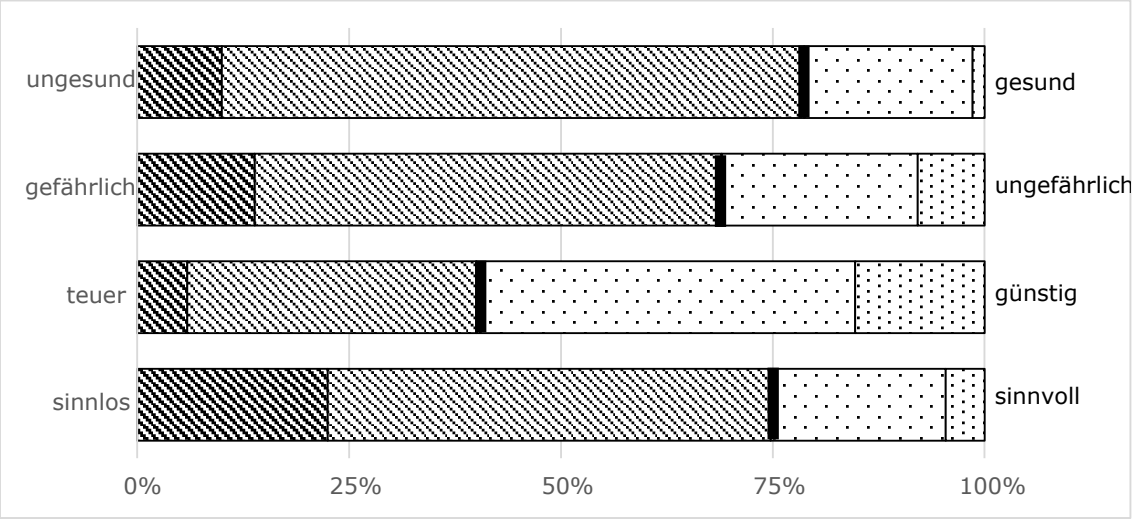


Abbildung 10.2: *Darstellung der relativen Anteile der Antwortmöglichkeiten 1 bis 4 bei Item 9 des Fragebogens. Fettgedruckte Linie als Abgrenzung zwischen dem linken und rechten Begriff. Dicht gepunktet = Option 1, leicht gepunktet = Option 2, leicht schraffiert = Option 3, dicht schraffiert = Option 4.*

a und d entschieden sich über 75 % der Lernenden für die Option 3 oder 4 (Schraffur) und stuften damit NEM als eher gesund sowie eher sinnvoll ein. Das Verhältnis zwischen Option 4 und 3 variiert jedoch. Bei Paar d wählten im Vergleich zu Paar a etwa doppelt so viele Befragte Option 4 und drückten somit eine stärkere Position aus. Bei Paar a ist der Anteil an Option 3 höher.

Bei Begriffspaar b wurden von etwa 70 % der Schüler:innen die Optionen 4 und 3 gewählt (14 % zu 55 %), von 23 % Option 2 und 7 % Option 1. Die Bewertung fällt damit leicht zugunsten des Begriffs *ungefährlich* aus.

Etwa 60 % kreuzten bei Begriffspaar c die Option 1 oder 2 an (gepunktet), also sehen sie NEM als eher teuer an. Bei diesem Paar ist die Meinungslage jedoch am wenigsten eindeutig.

10.2.3 Diskussion

Die Ergebnisse der Befragung von Lernenden der 8. bis 10. Klassen zeigen, dass Schülerinnen und Schüler Nahrungsergänzungsmittel aus ihrem Alltag kennen. Sie nannten spezielle Präparate und gaben verschiedene Anwendungsfelder an. Etwa ein Fünf-

tel der Befragten erkannte, dass NEM hauptsächlich den Ausgleich eines Nährstoffmangels bewirken. Einige Schülerinnen und Schüler schreiben ihnen jedoch Wirkungen zu, die über die Vermeidung von Mangelerscheinungen hinaus gehen. So wird bspw. erwartet, dass die allgemeine Gesundheit des Körpers verbessert wird oder dass die Präparate einen Energiegewinn liefern. Die Auswertung des ersten Fragebogenitems zeigt zudem, dass die Befragten oft keine klare Vorstellung davon haben, welche Produkte zu NEM gezählt werden und welche durch die Definition ausgenommen sind. So gaben 19 % an, NEM seien in Lebensmitteln (z. B. Fruchtsäften) enthalten. Es zeigt sich, dass nicht zwischen dem Nährstoff an sich und dem Nährstoffpräparat unterschieden werden konnte.

61 % der befragten Schülerinnen und Schüler verwenden NEM nach Bedarf oder häufiger, die Einnahme von Nährstoffpräparaten ist demnach unter Lernenden weit verbreitet. Am häufigsten supplementieren sie den Nährstoff Magnesium mit der Intention, Muskelkrämpfe und -schmerzen zu behandeln oder diesen vorzubeugen. Außerdem besteht eine hohe Einnahmeprävalenz von Vitaminen zur Vermeidung von Mangelerscheinungen und für die Verbesserung von Haut, Nägeln und Immunsystem.

Auffällig ist, dass einige erhoffte Wirkungen (z. B. die Vermeidung von Muskelkrämpfen) speziell einem Produkt zugeordnet werden (Magnesium), während andere Wirkungen von verschiedenen NEM erwartet werden. Beispielsweise erwarten die Lernenden einen Energiegewinn sowohl von Energie-Drinks und -Riegeln, als auch von Magnesium- und Vitaminpräparaten. Auch die allgemeine Verbesserung der Gesundheit wird von einigen Vitaminen und von Magnesium erwartet. Ein anderer Blickwinkel auf die Ergebnisse zeigt auch, dass manche Präparate spezifische Wirkungen zugeschrieben bekommen, während von anderen Nährstoffen erwartet wird, an vielerlei Orten im Körper zu wirken: Magnesium wird in den meisten Fällen mit Muskeln in Verbindung gebracht, nur wenige Lernende nannten hier andere erhoffte Wirkungen. Vitamin C oder D hingegen sollen die Gesundheit verbessern, das Immunsystem stärken, Energie liefern, positiv auf Psyche und Sinne wirken sowie für Zellschutz, Sonnenersatz, Fitness, bessere Haare und Nägel sorgen.

Die befragten Lernenden beziehen NEM mehrheitlich aus Apotheke, Drogerie und Supermarkt, informieren sich jedoch meist bei den eigenen Eltern und dem behandelnden Arzt über die einzunehmenden Präparate.

Die Items 8 und 9 verlangten eine Positionierung von den Befragten. Diese sind der Ansicht, dass Nahrungsergänzungsmittel einen Nährstoffmangel verhindern können

und dass eine gesunde Ernährung den Nährstoffbedarf decken kann. Ob dies in der deutschen Bevölkerung jedoch auch erreicht wird, können die Lernenden nicht einheitlich beantworten. Sie sind sich nicht einig, ob in Deutschland ein Vitaminmangel herrscht. Vermutlich ist aus diesem Grund die Einschätzung bezüglich Aussage h nicht ganz eindeutig: Die Befragten sollten bewerten, ob eine Supplementierung allgemein nötig ist. Zwar wurde hier im Mittel abgelehnt, dies jedoch mit erhöhter Unsicherheit.

Zu der lebensmittelrechtlich bedeutsamen Frage nach der Klassifizierung von NEM als Medikamente oder als Lebensmittel konnte ebenfalls kein Konsens innerhalb der Gruppe der Befragten herausgestellt werden, denn beide Aussagen werden im Mittel abgelehnt. Allerdings waren sich die Lernenden bei ihren abgegebenen Einschätzungen recht sicher, denn der Anteil an Extrem-Positionen ist gegenüber den mittleren Antworten erhöht.

Die Meinungslage bezüglich der Aussage i zeigt, dass die Befragten die Bedeutung einer gesunden Ernährung grundsätzlich erkennen und nicht glauben, dass sich eine ungesunde Ernährung durch Supplementierung ausgleichen lässt. Der hohe Anteil der stark ablehnenden Position zeigt, dass hier von einem Konsens gesprochen werden kann.

Auch die ebenfalls kritisch ausgelegten Aussagen d und e werden abgelehnt. Das zeigt, dass die Schüler:innen bei einer Konfrontation mit spezifischen Vorurteilen gegenüber Supplementen („Je mehr NEM, desto besser für den Körper“) eine kritische Haltung einnehmen können und diese auch vertreten. Dennoch sehen sie viele Vorteile in dem Konsum von NEM, beispielsweise stimmen sie den Aussagen a und b zu, nach denen NEM die allgemeine Gesundheit und Fitness verbessern können. Weiterhin zeigt sich anhand der anderen Items, dass weitere, teils unspezifische Intentionen und Gründe hinter der Einnahme von NEM stecken. Energiegewinn, Stärkung des Immunsystems, Verbesserung von äußeren Körpermerkmalen usw. erhoffen sich die befragten Schüler:innen von verschiedenen Nahrungsergänzungsmitteln.

Es bleibt in dieser Befragung offen, in welcher fachlichen Tiefe sich die Lernenden mit NEM auseinandersetzen. Sie geben als Informationsquellen hauptsächlich die eigenen Eltern an und an zweiter Stelle die behandelnde Ärztin bzw. den behandelnden Arzt. In Verbindung mit umfangreicher Beratung durch diese Bezugspersonen könnte ein rationaler, sachlich begründeter Umgang mit NEM hergestellt werden. Problematischer sind die Situationen, in denen sich die Lernenden in sozialen Medien, Wer-

bung oder bei Freund:innen informieren. Denn dort ist nicht immer sichergestellt, dass sie korrekt über Chancen und Risiken von NEM aufgeklärt werden.

Die Fragebogenerhebung hat einen Einblick in den alltäglichen Umgang von Lernenden mit Nahrungsergänzungsmitteln ermöglicht und dabei aufgedeckt, dass fast zwei Drittel der Befragten zumindest gelegentlich ihre Ernährung mit verschiedenen Präparaten supplementieren. Insgesamt herrschen dabei große Unsicherheiten bei Bewertungsaspekten: Wenngleich die Befragten NEM im Durchschnitt als eher sinnvoll, gesund und ungefährlich einstufen, zeigt die Detailanalyse der Antworten, dass hauptsächlich mittlere Positionen gewählt wurden. Volle Zustimmung oder Ablehnung wurde selten gewählt. Diese Erkenntnisse zeigen die Kontroversität des Themas bei gleichzeitig geringem Informiertheitsgrad der Lernenden und stellen den besonderen Bedarf einer unterrichtlichen Intervention dar, mit der den Lernenden Werkzeuge zur Bewertung von NEM vermittelt werden.

10.3 Fachspezifischer Problemaufriss und Positionierung

Das Unterrichtskonzept choice^{2}reflect wurde entwickelt, um Lernende zu einer rationalen Bewertung in gesellschaftlichen Kontroversen zu befähigen. Dazu ist ein fünfphasiges Vorgehen geplant, das mit einem fachspezifischen Problemaufriss zur Kontroverse beginnt und in eine Positionierungsphase übergeht.
Im folgenden Kapitel wird die Entwicklung der Lernmaterialien für die ersten beiden Phasen beschrieben und – dem mesozyklischen Vorgehen des Design-Based Research-Ansatzes folgend – Erprobung, Re-Design und eine erneute Erprobung ausführlich dargestellt.

10.3.1 Entwicklung des Lernmaterials

Die ersten im Rahmen dieses Promotionsprojektes entstandenen Unterrichtsmaterialien behandeln die erste Phase der Unterrichtskonzeption choice^{2}reflect: Den fachspezifischen Problemaufriss zum Thema Nahrungsergänzungsmittel. Die Vorstudie hat ergeben, dass ein recht großer Teil der Lernenden bereits in Kontakt mit Nahrungsergänzungsmitteln gekommen ist. Der am häufigsten supplementierte Nähr-

stoff ist Magnesium, daher bietet sich dieser besonders an, um an die Erfahrungen der Lernenden anzuknüpfen und die Unterrichtseinheit mit der Thematisierung von Magnesium zu beginnen. Ein durch die SuS häufig genannter erwünschter Effekt der Magnesium-Supplementierung ist die Vermeidung von Muskelkrämpfen. Die Assoziation von Magnesium mit Muskelkrämpfen ist in der Gesellschaft sehr geläufig (Garrison et al. 2020; DAZ 2007), sie bildet daher in dieser Unterrichtseinheit ein zentrales Element des Kontextes.

Der fachspezifische Problemaufriss dient dazu, die Lernenden an die Thematik der Nahrungsergänzungsmittel heranzuführen und dabei sachlich und neutral zu bleiben. Kontroverse Standpunkte werden aufgezeigt, um zu demonstrieren, dass eine Entscheidung häufig nicht eindeutig zu fällen ist und es weiterer Auseinandersetzung mit Aspekten der Wissenschaftstheorie bedarf.

Der entwickelte Unterrichtsentwurf sieht vor, zunächst kurz das Thema der Stunde vorzustellen und dann den einführenden Kontext zu präsentieren. Dieser behandelt die Assoziation von Magnesiummangel mit Muskelkrämpfen. Eine Aufgabenstellung fordert die Lernenden auf, der im Text vorgestellten fiktiven Person Paul ein Produkt zu empfehlen. Zur Auswahl steht die tägliche Einnahme entweder einer Magnesium-Tablette oder einer Banane. Weiterhin gibt es die Option der Enthaltung („Ich kann mich nicht entscheiden"). Die Entscheidung fällen die SuS anhand von zwei Bildern, die sie auf einem Tablet betrachten können. Es handelt sich dabei um selbst erstellte Beiträge im sozialen Netzwerk „Instagram". Beide Beiträge sind mit ähnlichen Stichwörtern - sogenannten Hashtags - versehen, wodurch symbolisiert wird, dass sie die gleiche inhaltliche Aussage haben sollen. Dabei bewirbt ein Beitrag Magnesium-Tabletten als Nahrungsergänzungsmittel und stellt die positive Wirkung auf Muskelfunktion und Leistungsfähigkeit in den Vordergrund. Der andere Beitrag bewirbt eine Banane als aufgrund ihres hohen Kalium- und Magnesiumgehaltes gut geeignetes Nahrungsmittel für Sportler. Beide Beiträge beschreiben außerdem die Inhaltsstoffe der Tabletten bzw. von Bananen.

Bei Betrachtung der Instagram-Beiträge sollte den Lernenden auffallen, dass eine Magnesium-Tablette 300 mg Magnesium enthält, während in 100 g Banane nur etwa 30 mg enthalten sind. Die Kalium-Gehalte sind mit 400 mg in der Tablette und 370 mg in 100 g Banane vergleichbar hoch.

Die Angaben der Instagram-Beiträge reichen aus wissenschaftlicher Sicht nicht aus, um eine eindeutige Entscheidung zu treffen, wie am besten gegen Muskelkrämpfe

vorzugehen ist. Das Arbeitsmaterial suggeriert jedoch, dass eine Empfehlung abgegeben werden kann. Mithilfe der Erprobung des Materials in einer Unterrichtssituation soll explorativ erforscht werden, in welcher Weise Schülerinnen und Schüler mit dem entwickelten Material umgehen und ob erkannt wird, dass eine eindeutige Produktempfehlung im beschriebenen Zusammenhang wenig sinnvoll ist.
Nach der Bearbeitung des Arbeitsblattes wird die Meinungslage in der Klasse mithilfe einer „Kahoot-Umfrage“ analysiert. Die Lernenden geben auf dem Tablet ihre Entscheidung ein und auf der Projektionsfläche im Klassenraum wird die Verteilung der abgegebenen Antworten direkt als Balkendiagramm angezeigt. Darauf folgt eine Diskussion im Plenum, wobei die Meinungslage in der Klasse thematisiert und über das Problem diskutiert wird. Dabei steht im Vordergrund, dass die Lernenden ihre Gedanken zur Thematik verbalisieren und miteinander in den Austausch treten. Im Anschluss an die Diskussion wird eine erneute Kahoot-Umfrage mit der gleichen Fragestellung durchgeführt um zu erheben, ob sich die Meinungen der Lernenden verändert haben.
Die Meinungslage nach der Diskussion wird im Plenum kurz angesprochen und in einem abschließenden Unterrichtsgespräch findet eine kurze Reflexion darüber statt, dass zur wissenschaftlichen Aufbereitung der Problematik weitere Kriterien erarbeitet werden müssen.
Folgende Lernziele werden für diesen Unterrichtsabschnitt festgelegt:

a. Lernziel 1: Schüler:innen vergleichen Nährstoffgehalte des Präparats *Magnesium/Kalium 400* mit denen von 100 g frischer Bananen
b. Lernziel 2: Schüler:innen nehmen Stellung zu der Problematik, ob und in welcher Form ein:e Sportler:in zur Vermeidung von Muskelkrämpfen Magnesium supplementieren sollte

10.3.2 Erste Erprobung des Lernmaterials

Der fachspezifische Problemaufriss mit anschließender Positionierung wurde im November 2018 mit 26 Lernenden der neunten Klasse eines Gymnasiums in Münster erprobt. Die Unterrichtsstunde mit einer Dauer von 45 Minuten wurde audio- und videographiert.

Die Analyse des Datenmaterials erfolgt unter der Fragestellung: „Inwiefern eignet sich der Kontext ‚Magnesium-Präparate gegen Muskelkrämpfe' für den Einsatz im *Fachspezifischen Problemaufriss* und der *Positionierung*?" Weiterhin wird das Erreichen der Lernziele überprüft.

Untersuchungsmethode

Zur Beantwortung der Fragestellung müssen Anforderungen definiert werden, die das Lernmaterial bzw. die konzipierte Unterrichtsstunde erfüllen muss, um als geeignet eingestuft werden zu können.
Der *Fachspezifische Problemaufriss*, also Phase I der Unterrichtskonzeption, dient einer neutralen Einführung in das kontrovers diskutierte Thema. Die Lernenden sollen mit einer Problemstellung konfrontiert werden und nachvollziehen, dass es unterschiedliche Sichtweisen auf die Kontroverse gibt. Das Material kann als geeignet angesehen werden, wenn auch unter den Lernenden keine klare Einigkeit über eine Sichtweise existiert und es Diskussionsanlässe im Unterricht gibt. Wird in der Positionierungs-Phase (Phase II) ein geteiltes Meinungsbild in der Klasse vorgefunden, kann die Lehrkraft optimal moderieren und genau diese Meinungsverschiedenheiten zum Diskussionsgegenstand machen. Dabei sollen die Schüler:innen erkennen, dass Entscheidungen auf der Grundlage von Informationen getroffen werden - und dass es essentiell ist, die Herkunft der Informationen zu überprüfen.
Nach Abschluss der Phase II sollte in der Klasse ein Konsens darüber geschaffen sein, dass ohne zusätzliche, wissenschaftlich-abgesicherte Informationen keine rationale Entscheidung in der Kontroverse möglich ist. Diese Notwendigkeit, wissenschaftliche Fakten zu erarbeiten, wird als Übergang in die Phase III verwendet.

Beschreibung der Unterrichtsstunde

Zu Beginn der Unterrichtsstunde wurde eine Bildschirmpräsentation gestartet, die eine Folie mit der Überschrift „Nahrungsergänzungsmittel" und einem Foto von verschiedenen Präparaten enthielt. Dieser initiale Impuls wurde von der Lehrkraft mit der Frage nach eigenen Erfahrungen bezüglich Nahrungsergänzungsmitteln (NEM) unterstützt. Die Lernenden nannten daraufhin in ihren Wortmeldungen, dass sie Vitamin-Tabletten (Video 2018-Problemaufriss; Zeitmarke 11:15), B-Sprint-Tabletten

(11:30) und Tabletten zur Verbesserung von Eisen- bzw. Magnesiummangel (11:52) kennen.
Die Lehrkraft erklärte kurz, was unter NEM verstanden wird und leitete über zur ersten Arbeitsphase. Diese beinhaltete das gemeinsame Vorlesen des Arbeitsblattes (14:20-15:30), Austeilen der Tablets (16:30-17:30) und die Arbeit der Lernenden an den Foto-Dateien auf den Tablets (18:00-26:30).
Die Positionierung mittels Kahoot-Quiz folgte auf die Arbeitsphase. Aufgrund von technischen Schwierigkeiten konnten nicht alle Tablets in die Kahoot-Umgebung eingewählt werden, sodass lediglich zwölf Antworten abgegeben wurden. Davon haben acht *Magnesium* (Mg) und vier *Banane* ausgewählt. Keine:r der Lernenden entschied sich für *weiß ich nicht*.
In einer Plenumsphase begründeten einzelne Lernende ihre Antworten. Ein Schüler habe Mg gewählt, da er selbst auch Mg einnehme und „das nimmt man halt" (31:40). Außerdem würden 30 Milligramm nicht ausreichen, um Muskelkrämpfe zu bekämpfen und Tabletten seien dafür perfektioniert, Muskelkrämpfe zu lösen. Als weitere Argumente für Mg wurde genannt, dass Mg etwas für Erwachsene sei (33:12) und dass Mg für den Bereich Sport besonders sinnvoll sei (33:27). Weiterhin wurde den Angaben der Instagram-Posts vertraut, nach denen das 1:1 Verhältnis von Magnesium und Kalium eine gute Muskelfunktion hervorrufe (33:50).
Die Gegenseite argumentierte, die Einnahme von Tabletten sei auf Dauer nicht gut (32:41) und Banane sei als „natürliches Produkt" besser geeignet (33:02). Außerdem wären bei der Einnahme von Mg-Tabletten eventuell Nebenwirkungen zu erwarten (35:11).
Die Lehrkraft stellte zur Diskussion, ob die Instagram-Beiträge eine gute Quelle für den Bezug von Informationen sind. Die Lernenden verneinten dies unter Angabe von drei Begründungen: Die fehlende Nennung von Nebenwirkungen (36:49) oder das gezielte Verschweigen „schlechter Details" (37:04) sowie der Umstand, dass „jeder schreiben kann, was er möchte" und kein Wissenschaftler sein muss (37:21) diskreditieren aus Sicht der Schüler:innen die Instagram-Beiträge.
Auf welchem Wege eine Informationssuche sonst stattfinden könnte, wurde von einigen Schüler:innen vorgeschlagen: Man könne „einfach im Internet googeln" (38:55), einen Experten (38:40) oder Arzt (38:59) fragen. Möglicherweise hätten auch eine Apotheke (38:52) oder der Fußballtrainer (39:05) genauere Informationen zur Thematik.

Da die Lernenden Ärztinnen und Ärzte als Haupt-Informationsquellen benannten, verfolgte die Lehrkraft den Gedankengang und fragte, woher Ärzte ihre Informationen beziehen. Daraufhin antworteten die Lernenden, der Arzt „müsse das wissen" (39:39), er lerne das in der Ausbildung (39:41) oder lese Studien (39:33).
Im Anschluss an die Plenumsdiskussion rief die Lehrkraft zu einer erneuten Positionierung zu derselben Fragestellung wie am Anfang der Stunde auf. Diese wurde nicht digital sondern per Handzeichen durchgeführt. Von 24 Lernenden positionierten sich zwölf für *Magnesium-Tabletten*, zwölf für *Bananen* und zwei wählten *weiß ich nicht*.
Dass das Meinungsbild auch bei der zweiten Positionierung derart auf die verschiedenen Optionen verteilt war, begründeten die Schüleri:nnen mit fehlenden Informationen. Dabei gaben sie an, dass vor allem Informationen zu Nebenwirkungen (47:35), dem Tagesbedarf an Magnesium (47:55) und darüber, wie viel Magnesium für einen Menschen über einen langen Zeitraum gesund ist, fehlen.
Zum Ende der Stunde kommt das Plenum zu dem Ergebnis, dass „man sich damit auseinandersetzen und sich informieren" (52:38) muss.

Auswertung der Arbeitsblätter

Anhand der Arbeitsblätter lässt sich neben der Positionierung jeder Schülerin und jedes Schülers auch die jeweilige Begründung ermitteln. 25 der 26 Arbeitsblätter wurden nach Ende des Unterrichts eingesammelt und untersucht. 15 Lernende wählten die Magnesium-Tablette und 10 Lernende die Banane. In keinem Fall wurde *weiß ich nicht* gewählt.
Die Wahl der Magnesium-Tablette wurde in 13 von 15 Nennungen damit begründet, dass die Tablette deutlich mehr Magnesium enthält als eine Banane und daher zu bevorzugen ist. Weiterhin wurde das Verhältnis von Magnesium und Kalium im Präparat als Vorteil aufgeführt.
Zehn Lernende entschieden sich für die Banane und begründeten ihre Wahl damit, dass Bananen allgemein gesünder (fünf Nennungen) und ein natürliches Produkt (zwei Nennungen) seien. Weiterhin wurde angeführt, dass Bananen auch Vitamine enthalten (zwei Nennungen), vor allem für Sportler geeignet sind (zwei Nennungen) und für Kinder verträglicher sind als Tabletten (zwei Nennungen).
Die notierten Begründungen zeigen an einigen Stellen, dass die Lernenden über keine gesicherten Informationen verfügen und dass sie sich unsicher sind. Im Hinblick

Tabelle 10.3: *Darstellung aller als „unsicher" eingestuften Antworten auf dem Arbeitsblatt zur Positionierung. Indikatoren für Unsicherheit sind fettgedruckt. Spalte * zeigt die von den Lernenden angekreuzte Option (Mg = Magnesium; Ban = Banane).*

Kürzel	*	Antwort
NP42	Mg	Eine Magnesium-Tablette bietet 400 mg Magnesium. Die Banane dagegen nur 30 mg. Zudem hat die Tablette ein 1:1 Verhältnis im Zusammenhang mit Kalium. Die Banane besitzt 370 mg Kalium. **Die Frage ist nur**, wie viel der Körper braucht.
KI24	Ban	Ich würde die Banane nehmen, weil eine Banane gesünder als eine Tablette ist, auch ist es **glaub ich** auch nicht so gut für Kinder wenn sie täglich Tabletten nehmen.
MW37	Ban	Für mich hört sich beides ganz gut an aber **vielleicht** ist die Banane ein bisschen gesünder. Weil auf Dauer sind Tabletten nicht gut.
BL12	Mg	Eine Magnesium-Tablette, weil sie **wahrscheinlich** viel mehr Magnesium hat als die Banane. Die Banane ist **wahrscheinlich** gesünder, aber die Tablette hat mehr Magnesium.
LS14	Mg	Ich denke, dass eine Banane oder eine Tablette eigentlich gleich viel bringen, für seinen Bereich die Tablette **wahrscheinlich** sinnvoller ist.
EG31	Ban	Ich treibe selber Aktiv Sport und wir bekommen auch oft Bananen. Regelmäßige Tabletten sind selten gut und **ich glaube**, Bananen sind gesünder und besser.

auf unsichere Formulierungen wurden die Arbeitsblätter codiert und die gefundenen Antworten in der Tabelle 10.3 dargestellt. Es fällt auf, dass die Ausdrücke *vielleicht*, *wahrscheinlich* oder *ich glaube, dass...* vermehrt vorkommen.

Auswertung der Gruppengespräche

Die Wortbeiträge in Plenumsphasen zu betrachten oder notierte Antworten auf den Arbeitsblättern zu analysieren bringt Erkenntnisse über den Endstand der Überlegungen von Lernenden. Es zeigt, auf welche Lösungen und Gedankengänge sich die Schüler:innen während der Arbeitsphase einigen konnten. Einen tiefergehenden Einblick in den Entstehungsprozess der Arbeitsergebnisse bietet jedoch die Auswertung von Gruppengesprächen in den Arbeitsphasen.

Der Einsatz von Weitwinkel-Kameras und Diktiergeräten in den Schülergruppen sorgt dafür, dass jede Gruppe über die Dauer der Unterrichtsstunde isoliert betrachtet werden kann. Im Folgenden werden alle mit dem Stundenthema verknüpften Aussagen, Ausrufe und Diskussionen der Lernenden dargestellt. Für die Beantwortung der Fragestellung irrelevante Äußerungen werden hier nicht betrachtet.
Ein Schüler aus Gruppe 1 äußerte in der Positionierung: „Ich bin mir schon sicher. Ich nehme ja selber Magnesium-Tabletten“ (S8; 19:38). Etwas später beginnt eine kurze Diskussion zwischen S4, S2 und S7. S4 bringt ein Argument für die Banane: „Aber ich finde das andere [die Option Banane] natürlicher“ (S4; 22:49). S2 entgegnet: „Aber das [die Tablette] ist gesünder“. S7 führt seine Gedanken daraufhin aus: „Ich habe geschrieben, dass es für Paul speziell besser ist mit der Magnesium-Tablette. Weil, äh, eine Banane hat vielleicht noch andere wichtige Wirkstoffe. Aber halt für Paul sind halt die Tabletten besser, um schnell die Wirkung zu kriegen“ (S7; 23:03).
In der Gruppe 2 wird ersichtlich, dass ein Schüler sich nicht für eine Option entscheiden kann: „Jo, ich kann mich nicht zwischen diesen beiden entscheiden. Echt“ (S14, 16:48). S10 sagt: „Banane! Das gibt Energie, aber ich hab Magnesium genommen“ (S10; 17:18). Etwas später entsteht eine Diskussion. S13 meint: „Das Ding ist halt, Bananen sind halt wirklich gut“ (S13; 18:28), S11 stimmt zu (18:32). S13 führt weiter aus: „Ich glaube wenn du Sportler bist, kannst du Magn- eh kannst du Bananen nehmen um die [Muskelkrämpfe] zu vermeiden“ (18:34). Anschließend ruft S13 leicht euphorisiert: „Hey, ich glaub ich würds so machen. Ich würde immer Bananen essen um die [Muskelkrämpfe] zu vermeiden“ (18:44).
In Gruppe 3 wurde nur eine fachlich passende Aussage von S24 codiert: „Also ich habe jetzt hier die Magnesium-Tablette genommen. Aber theoretisch wäre es doch eigentlich besser, wenn er vor dem Sport eine Banane essen würde, weil die hat ja viel mehr Eiweiße und Kohlenhydrate“ (S24; 23:28).

10.3.3 Diskussion und Re-Design

Die Reaktionen auf den Einstiegsimpuls zeigen deutlich, dass das Thema NEM bei den Lernenden präsent und in ihrem Alltag relevant ist. Diese Einschätzung wird dadurch gestützt, dass alle Lernenden in der Positionierung eine klare Meinung zum Thema abgegeben haben und in ihren Begründungen Vor- und Nachteile der einzelnen Optionen gegeneinander abwägen konnten. Einige Schüler:innen haben auch angegeben,

selbst Magnesiumpräparate einzunehmen.
Es hat sich allerdings auch gezeigt, dass die Lernenden trotz der eigenen Erfahrungen mit Supplementen nicht umfangreich über Dosierungen, Wirkungen und Hintergründe der Präparate informiert sind. Sie vertrauen den Instagram-Beiträgen und übernehmen die dort gegebenen Informationen in ihre eigenen Argumentationsstrukturen. Dabei ist erkennbar, dass sich die Lernenden bei der Angabe der Begründung nicht immer sicher sind (siehe Tabelle 10.3).
Beide Lernziele können als erreicht angesehen werden, denn in der durchgeführten Stunde haben alle Lernenden sowohl die Nährstoffgehalte von Magnesium-Tabletten mit denen von Bananen verglichen, als auch Stellung zur Problematik bezogen. Von größerer Bedeutung ist an dieser Stelle jedoch die Beantwortung der Forschungsfrage. Dabei wird eine Aussage getroffen, inwiefern das entwickelte Material für den Einsatz in Phase I und II geeignet ist.
Wie zuvor beschrieben, werden das konzipierte Lernmaterial und dessen Einsatz in der geplanten Unterrichtsstunde als geeignet angesehen, wenn das Meinungsbild in der Klasse nicht eindeutig ist und im Unterrichtsgespräch Diskussionen über die Positionierungs-Optionen entstehen. Die Untersuchungen haben gezeigt, dass beide Anforderungen in der erprobten Stunde zutreffen. Weiterhin zeigt sich, dass das Thema der Magnesium-Supplementierung eine hohe Relevanz für die Lernenden besitzt, denn im Plenum haben mehrere Schüler:innen angegeben, selbst Magnesium-Präparate einzunehmen. Die jeweils angeführten Begründungen erwiesen sich als schwammig und nicht wissenschaftlich. Das ist an Formulierungen wie: „das nimmt man halt“ oder an den Gruppengesprächen zu erkennen.
Im späteren Verlauf zeigte die Lehrkraft auf, dass die vorhandenen Informationen nicht ausreichen und erörterte mit den Schüler:innen gemeinsam verschiedene Wege zur Informationsbeschaffung. Das Ergebnis der Diskussionsphase ist, dass „man sich damit auseinandersetzen und sich informieren“ muss. Damit wird auch die letzte Anforderung an das Unterrichtsmaterial erfüllt, sodass es als geeignet für die Verwendung in Phase I und II der Unterrichtskonzeption choice^{2}reflect angesehen werden kann.
An einigen Stellen des Unterrichts könnten noch Optimierungen vorgenommen werden. Es hat sich gezeigt, dass das digitale Werkzeug *Kahoot* unter den technischen Gegebenheiten vor Ort nur unzuverlässig und instabil funktionierte. Die Positionierung sollte von jeder Schülerin und jedem Schüler durchgeführt werden können, daher

bietet es sich an, auf ein zuverlässigeres Werkzeug zu setzen. Beispielsweise könnte per Handmeldung die Positionierung erfasst werden, welche dann allerdings weniger anonym ist. Dazu wurde im Rahmen des *Re-Design* das Arbeitsblatt um eine weitere Seite ergänzt. Auch die Aufgabenstellung wird leicht verändert: „Betrachte das Video vom Gruppenchat der Fußball-Mannschaft" lautet die erste Aufgabe. Dazu erhalten die Lernenden eine Video-Datei, die sie auf einem digitalen Endgerät abspielen können. Dort ist eine Messenger-Unterhaltung abgebildet, die einem Gruppenchat nachempfunden ist. Es wird eine Gruppenunterhaltung zwischen Mitgliedern eines Sportvereins simuliert. Ein Mitglied schreibt: „Hey Leute. Mein Trainer empfiehlt mir Magnesium gegen Muskelkrämpfe. Was meint ihr – brauche ich das?" Daraufhin antworten die anderen Mitglieder:

Björn „Heutzutage sind die Lebensmittel so nährstoffarm, dass man unbedingt Magnesium-Tabletten nehmen muss."

Kim „Aber früher gab es keine Nahrungsergänzungsmittel und den Menschen ging es trotzdem gut. Deswegen brauchen wir das heute auch nicht."

Noel „Das wird doch auf Instagram immer viel diskutiert…"

Noel, der letzte Gesprächsteilnehmer, teilt anschließend die zwei Beiträge des sozialen Netzwerks Instagram, die auch in der ersten Erprobung schon verwendet wurden. Die Konversation ist auch in Abbildung 10.3 dargestellt. Mit dieser Art der Einbindung des Fachspezifischen Problemaufrisses wird der Unterricht näher an der Lebenswelt der Lernenden orientiert und enthält etwas mehr Abwechslung als ein Einführungstext.
Die nächste Arbeitsaufgabe lautet: „Unten siehst du zwei Instagram-Beiträge, die Noel unter #magnesiummangel gefunden hat. Lies dir beide Beiträge genau durch".
Auf einer zweiten Seite des Arbeitsblatts ist erneut die Messenger-Situation abgebildet, allerdings befinden sich dort Ankreuz-Felder, die die Lernenden zu einer Positionierung auffordern. Die Lernenden können entscheiden:

- Gegen die Krämpfe sollte Paula Magnesium nehmen.
- Gegen die Krämpfe sollte Paula ausreichend Bananen essen.
- Ich weiß nicht, was Paula gegen die Krämpfe tun kann.

Abbildung 10.3: *Ausschnitt aus der Video-Datei „Gruppenchat", die die Lernenden im Rahmen des Fachspezifischen Problemaufrisses verwenden.*

Diese Anpassungen an dem Arbeitsblatt können in weiteren Erprobungen und im zukünftigen Unterrichtseinsatz verwendet werden.

Als weiteres Element im Re-Design sollte der Rahmen, in dem Diskussionen zugelassen werden, zuvor genau abgesteckt werden. Die Lehrkraft sollte eine gewisse Diskussionszeit einplanen, dann jedoch an geeigneter Stelle das Plenumsgespräch beenden, um zur nächsten Phase überzugehen. Wird die Diskussion nicht an einem festgelegten Punkt beendet, könnte sie sich unter Wiederauftreten der schon vorher erörterten Punkte im Kreis drehen, ohne einen Mehrgewinn für die Lerngruppe zu bieten. Sobald also deutlich wird, dass die Lerngruppe weitere Informationen verlangt und wissenschaftliches Prüfen fordert, kann zur dritten Phase übergeleitet werden.

10.3.4 Zweite Erprobung des Unterrichtsmaterials

Das Unterrichtsmaterial zu Phase I und II wurde auf der Grundlage der Erkenntnisse aus der ersten Erprobung leicht angepasst. Die zweite Erprobung dieser Phasen fand im Rahmen einer Gesamterprobung im September 2020 in der neunten Klasse einer Gesamtschule statt. Der Umfang betrug drei Doppelstunden, davon wurden 45 Minuten der ersten Doppelstunde den Phasen I und II gewidmet. 25 Schüler:innen nahmen an der Stunde teil. Details zu den Gegebenheiten bei dieser Erprobung werden im Kapitel 8.1 beschrieben.

Beschreibung der Unterrichtsstunde

Zu Beginn der Stunde (Start der Stunde an Position 08:35 min. des Videos) stellte sich die Lehrkraft vor und erklärte in einigen Sätzen den Ablaufplan für die folgenden Doppelstunden. Lernende wurden angewiesen, die Kameras aufzusetzen und die Tablets mit dem Internet zu verbinden. Etwa ab Minute 23 begann die inhaltliche Arbeit an dem Arbeitsblatt. Gemeinsam wurde der kurze Einführungstext gelesen und die Lehrkraft erklärte das Wechseln zur Videodatei an dem Tablet. Die Lernenden begaben sich in die Stillarbeitsphase, diese dauerte bis zur Minute 33 an.
Die Abfrage im Plenum ergab, dass sich zehn Leute für die Magnesium-Tabletten und neun für die Banane entschieden hatten. Ein:e Lernende:r kreuzte „weiß ich nicht“ an.
In der Diskussion begründete ein Schüler seine Wahl des Magnesium-Präparats folgendermaßen: „Ja. Also ich hätte jetzt/ also von mir aus hätte ich als erstes [...] gesagt, keine Ahnung geh zum Arzt. Und dann wäre die wahrscheinlich zum Arzt gegangen und dann hätte der Arzt wahrscheinlich gesagt: Musst mehr Magnesium oder so zu dir nehmen [...]. Und deswegen habe ich die erste Antwort genommen“ (34:35). Weitere Lernende führten an, dass „das entzündungshemmend ist und die Muskeln eventuell entspannt“ (35:38) und „dass Magnesium im Körper wieder zu normalen Leistungen macht“ (36:04).
Eine Schülerin nannte als Begründung für die Banane: „Bananen sind natürlicher, sie enthalten nicht nur Magnesium sondern auch andere Nährstoffe, die der Körper benötigt“ (37:14). Ein:e weitere Schüler:in sagte: „Also meine Begründung ist, dass Bananen zwar auch viele wichtige Stoffe noch zusätzlich enthalten, aber sehr viel weniger Magnesium als jetzt eine Magnesiumtablette“ (37:56).

Die Lehrkraft wies darauf hin, dass ein sehr gemischtes Meinungsbild vorliegt und fragte, wie denn nun eine Entscheidung getroffen werden könne. Daraufhin entgegneten die Lernenden, dass sie auf die Mehrheit hören würden, zum Arzt gehen könnten (39:07) oder einfach ausprobieren, was hilft (39:22). Ein Lernender schlug vor, auf Google nachzuschauen (40:12).
Auf die Nachfrage, woher denn ein Arzt wisse, „was richtig sei", antworteten die Lernenden mit: „Weil der auf der Uni war und das studiert hat [...] oder weil der auf vertrauenswürdigen Seiten unterwegs ist" (40:42).
Die Lehrkraft leitete über zur Phase III, indem die Bedeutung wissenschaftlicher Studien erklärt wurde. Dabei nannte die Lehrkraft auch die Verschwörungstheorie, dass 5G-Strahlung für den Ausbruch der Corona-Pandemie verantwortlich sei und stellte heraus, dass wissenschaftliche Studien zur Klärung nötig sind. Damit endete die Phase II.

Auswertung der Gruppengespräche

Die Verteilung der Kameras und Diktiergeräte im Raum wurde so gewählt, dass jede Gruppe auch einzeln betrachtet und angehört werden kann. Daher werden im folgenden Abschnitt die Gruppengespräche ausgewertet. Zitiert werden Aussagen, die unmittelbar mit der Thematik der Magnesium-Supplementierung zusammenhängen. Andere Äußerungen der Lernenden werden an dieser Stelle nicht betrachtet.
In Gruppe 1 ist ersichtlich, dass ein Schüler versucht, seine Mitschülerin zum Ankreuzen einer bestimmten Lösung zu überreden: „S1-03: Du musst das obere nehmen, Magnesium! Weil, Magnesium, ähm, ich nehme auch immer Magnesium!" (Gruppe 1, 29:44). Weiterhin äußerte diese:r Lernende: „S1-03: Würde das jetzt jemand aus meiner Klasse schreiben, dann würde ich sagen: KP, geh zum Arzt, alter. Und der würde wahrscheinlich sagen, dass er Magnesium schlucken sollte". Ein anderes Kind in dieser Gruppe tätigt eine ähnliche Aussage: „S1-02: Keine Ahnung, ich hab einfach Magnesium, weil ich nehme auch immer Magnesium. Aber ich weiß nicht, was ich als Begründung hinschreiben soll".
Diese Aussage ist auch in Gruppe 3 zu finden. S1-06 sagt: „Ich hab noch nie Bananen/ Ich habe immer Magnesium gegen Muskelkrämpfe genommen. Ich kreuze das jetzt an. Ist das nicht entzündungshemmend?" Weitere Aufgaben-bezogene Aussagen sind in dieser Gruppe nicht zu finden.

Gruppe 2 ist während der Arbeitsphase nahezu still.
In Gruppe 4 gibt es ebenso kaum Diskussionen über die Aufgabe. Gegen Ende der Arbeitsphase sagt S2-08: „Warte. Magnesium enthält viele Nährstoffe, die gegen Muskelkrämpfe helfen. Da schreibe ich das hin" (Gruppe 4, 30:30) und später auch: „S2-06: Die sollte lieber Magnesium nehmen. (…) Weil, das hilft" (Gruppe 4, 32:50).
Ein:e Schüler:in aus Gruppe 5 erklärt kurz seinen Mitschülern, was er oder sie notiert hat: „S2-11: Weil in der Banane noch andere Nährstoffe sind" (Gruppe 5, 31:05).
Weitere Gespräche sind nicht erkennbar.

Diskussion

Im ersten Erprobungsdurchlauf des Lernmaterials für die Phasen I und II konnte bereits gezeigt werden, dass die Thematik der Magnesium-Supplementierung für einen Problemaufriss und die Positionierung geeignet ist. Diese Einschätzung wird durch die Erkenntnisse der zweiten Erprobung unterstützt. Lernende einer anderen Schulform (Gesamtschule statt Gymnasium) und Klassenstufe (9 statt 8) reagieren ähnlich auf das Lernmaterial. Sie erzählen von eigenen Erfahrungen mit Magnesium-Präparaten und fällen daraufhin ihre Entscheidung. Sie besitzen unterschiedliche Vorstellungen über die Wirkungen von Magnesium-Präparaten im Körper: Lernende sprechen von entzündungshemmender Wirkung, von Nährstoffreichtum der Tabletten, Natürlichkeit der Bananen und bringen weitere ungesicherte Informationen in die Unterrichts-Diskussion ein.
Auch in dieser Erprobung wird deutlich, dass die Lernenden ihre Entscheidungen kaum auf der Basis wissenschaftlich gesicherter Informationen treffen. Die Frage der Lehrkraft nach möglichen Informationsquellen lässt diese Erkenntnis besonders hervortreten. Die Lernenden schlagen vor, auf die Mehrheit zu hören, einfach ein Präparat auszuprobieren oder auf eine Ärztin oder einen Arzt zu hören, die/der „das weiß, weil sie/er studiert hat".
Es ist somit von großer Bedeutung, die Schüler:innen in dieser Hinsicht weiterzubilden und Bewertungskompetenzen zu schulen. Das konzipierte Lernmaterial für die ersten beiden Phasen der Unterrichtskonzeption bereitet die Lernenden ideal auf die folgenden Phasen vor. Es sensibilisiert die Schüler:innen für die Kontroverse der Nahrungsergänzungsmittel und knüpft an die Alltagsvorstellungen an. Die Positionierung sorgt dafür, dass auch die eigene Haltung zum Thema reflektiert bzw. zunächst

nur bewusst gemacht wird, um am Ende der Konzeption in der Phase V dann auf der Basis wissenschaftlicher Prüfkriterien reflektiert werden zu können.

10.4 Anwendung und Reflexion

Die Unterrichtskonzeption choice^{2}reflect besteht aus fünf Phasen. Die erste, zweite, vierte und fünfte Phase stehen im Kontext eines gesellschaftlich kontrovers diskutierten Themas, in diesem Projekt handelt es sich dabei um das Thema *Nahrungsergänzungsmittel* (NEM). Die dritte Phase dient, losgelöst von der übergeordneten Kontroverse, der Erarbeitung wissenschaftlicher Prüfkriterien an weiteren kleinen Kontexten.
In dem diesem Kapitel übergeordneten Mesozyklus 2 wird die Eignung des Themas NEM überprüft und es werden Lernmaterialien für die vier Kontext-basierten Phasen entwickelt und erprobt. Im Abschnitt 10.3 wurden die Phasen I und II bereits ausgeschärft. Daher soll es im Folgenden darum gehen, den Bogen zu schließen und die letzten beiden Phasen zu entwickeln und zu erproben.

10.4.1 Ziele der Anwendungs- und Reflexionsphase

Die Phase III besteht aus einer umfangreichen Erarbeitung wissenschaftlicher Prüfkriterien und bildet den größten Teil der Konzeption. Die Lernenden denken sich dabei in unterschiedliche Kontexte und Sachverhalte ein. Im Anschluss an diese Phase sollen die Schüler:innen das erlernte Wissen auf eine Kontroverse anwenden. Dafür wird das Thema wieder in den Vordergrund geholt, mit dem sich die Schüler:innen schon in den ersten Phasen auseinandergesetzt haben: Es geht in Phase IV wieder um NEM.
Das Ziel der Anwendungsphase besteht darin, dass die Lernenden eine echte Studie über NEM im Hinblick auf wissenschaftliche Prüfkriterien bewerten. Sie schauen genau auf die Formulierungen der Studie und prüfen auf *Korrelation und Kausalität*, *Randomisierung*, *Verblindung*, *Kontrollierte Bedingungen* und/oder *Reproduzierbarkeit*. Der Umfang der Phase richtet sich danach, welche Prüfkriterien in Phase III erarbeitet wurden, denn die Auswahl der zu erarbeitenden Kriterien obliegt der Lehrkraft.

Wurde die echte Studie auf ihre Wissenschaftlichkeit geprüft, dann wird das Ergebnis der Studie bei bestandener Prüfung zur eigenen Entscheidungsfindung herangezogen oder bei Feststellung mangelnder Wissenschaftlichkeit verworfen. Das ist Teil der *Kriteriengeleiteten Reflexion* in Phase V. Die Lernenden positionieren sich erneut zur Kontroverse, beziehen dieses Mal jedoch ihre Erkenntnisse der vorangegangenen Phasen in die Entscheidung ein. Jungkamp (2021) beschreibt, dass die Lernenden dafür eine hohe Ambiguitätstoleranz benötigen, denn in den meisten Fällen lässt sich keine endgültige, vollständig „richtige“ Entscheidung treffen.

10.4.2 Entwicklung des Lernmaterials

In der Anwendung beschäftigen sich Lernende mit Studien zur Kontroverse um den Konsum von NEM. Speziell geht es in den Phasen I und II um den Nutzen einer Magnesium-SSupplementierung bei regelmäßig auftretenden Muskelkrämpfen. Es wurde geklärt, dass wissenschaftlich gesicherte Informationen zum Thema nötig sind, um eine rationale Entscheidung treffen zu können. Diese Informationen werden den Lernenden in der Anwendungsphase bereitgestellt. Dafür gilt es jedoch zunächst zu klären, wie die Studienlage zum Thema aussieht und auf welchem Stand die Wissenschaft sich derzeit befindet.
Die Entwicklung des Unterrichtsmaterials geschieht auf der Grundlage eines aktuellen Reviews, das den Einfluss von Magnesium auf Skelettmuskelkrämpfe untersucht und dabei sieben empirische Studien in den Blick nimmt (Garrison et al. 2020). Die Auswahlkriterien sind klar definiert, es wurden nur Studien inkludiert, die angaben, randomisiert, kontrolliert und verblindet zu sein. Die Meta-Analyse der sieben Studien kam zu dem Ergebnis, dass es unwahrscheinlich ist, dass Magnesium-Supplementierung in klinisch bedeutsamen Maße zur Prophylaxe von Muskelkrämpfen bei älteren Personen eingesetzt werden kann. Im Bezug auf Schwangere, die Wadenkrämpfe erleiden, wurden konfligierende Daten gefunden. Die Forschenden schreiben, dass daher in dem Bereich weitere Forschung nötig ist.
In dem Review wird auch beschrieben, nach welchen Kriterien die Studien ausgewählt wurden, die die Forschenden zum Bilden einer Schlussfolgerung einbezogen haben. Außerdem werden Studien beschrieben, die aus verschiedenen Gründen exkludiert wurden. Diese Studien sind besonders interessant für eine Anwendungsphase in choice^{2}reflect . In einer Tabelle stellen die Forschenden sieben exkludierte Stu-

Tabelle 10.4: *Darstellung der im Review von Garrison et al. (2020) beschriebenen exkludierten Studien mit Angabe des zugehörigen Prüfkriteriums.*

Studie	Grund für Ausschluss	zugeh. Prüfkriterium
Aagaard et al. (2005)	Es wurden lediglich Muskelmasse, -stärke und -Magnesiumgehalt gemessen, keine Muskelkrämpfe.	*Korrelation und Kausalität*
Bachem et al. (1986)	Keine Kontrollgruppe	*Kontrollierte Bedingungen*
Bartl & Riss (1982)	Vermutlich nicht randomisiert, außerdem wurde die Änderung der Häufigkeit von Krämpfen nicht erhoben	*Randomisierung*
Hammar et al. (1987)	Keine Gruppe mit Magnesium-Behandlung	*Kontrollierte Bedingungen*
Häringer (1981)	Keine Kontrollgruppe	*Kontrollierte Bedingungen*
Riss et al. (1983)	Unkontrolliert	*Kontrollierte Bedingungen* bzw. *Korrelation und Kausalität*
Weller et al. (1998)	Es wurden lediglich Aktivitätsleistung und Magnesiumkonzentration gemessen, keine Muskelkrämpfe.	*Korrelation und Kausalität*

dien dar und beschreiben die Gründe für den Ausschluss. Es bietet sich an, diese Studien für den Einsatz in Phase IV aufzubereiten, damit die Lernenden nicht erfüllte Prüfkriterien an diesen Studien erkennen können.

In Tabelle 10.4 sind die von Garrison et al. (2020) beschriebenen exkludierten Studien aufgelistet. Die rechte Spalte zeigt an, welches Prüfkriterium nicht erfüllt wurde.

Für den Einsatz im Rahmen der Gesamterprobung im September 2020 wurde die Studie von Bartl & Riss (1982) ausgewählt, in der nach Ansicht von Garrison et al. (2020) nicht (bzw. nur unzureichend) randomisiert wurde.

In der Studie wurden 67 schwangere Frauen untersucht, davon litten 39 unter regelmäßigen Wadenkrämpfen. Nach einer initialen Messung des Gehalts von Magnesium-Verbindungen im Blut bekam eine Hälfte der Teilnehmerinnen zwei Magnesium-Tab-

letten pro Tag, während die andere Gruppe nicht behandelt wurde.
Nach vier Wochen wurde erneut gemessen, dabei stellten die Forschenden in der mit Magnesium behandelten Gruppe einen Anstieg des Magnesium-Serum-Spiegels fest. Die Forschenden schlossen aus diesen Ergebnissen, dass mit einer Magnesium-Substitutionstherapie eine Verbesserung der klinischen Symptome (Wadenkrämpfe) bei Schwangeren eintreten wird.
Die Studie wurde für den Einsatz im Unterricht gekürzt, vereinfacht und mit schematischen Abbildungen versehen. Dadurch ist gewährleistet, dass die Lernenden auch bei wenig zur Verfügung stehender Unterrichtszeit die wichtigsten Informationen aus der Studie herausarbeiten können. Einige Details wurden aus didaktischen Gründen abgeändert, es wurde jedoch darauf geachtet, möglichst nah an der Original-Studie zu bleiben.
Der vereinfachte Studien-Text kann in Phase IV von den Schüler:innen auf kontrollierte Bedingungen (Kontrollgruppendesign) und auf Randomisierung geprüft werden. Es ist zu beachten, dass die Lernenden im Rahmen von Phase III die Randomisierung im Sinne einer zufallsbasierten Stichprobenauswahl kennenlernen, während die Aufteilung auf Test- und Kontrollgruppe eher den kontrollierten Bedingungen zugeordnet wird. Dieses Verständnis der Kriterien wird auch auf dem Arbeitsblatt zur Anwendung vertreten.
Hinzugefügt wurde eine weitere Schlussfolgerung, die von den Forschenden nicht getroffen wurde: *Magnesium-Tabletten erhöhen den Magnesium-Gehalt im Blut und reduzieren dadurch Wadenkrämpfe*. Diese Annahme wurde ergänzt, um mit dem Arbeitsblatt auch die Anwendung des Prüfkriteriums *Korrelation und Kausalität* zu ermöglichen. Sofern zuvor erarbeitet, können die Lernenden an dieser vereinfachten Studie also drei Prüfkriterien anwenden.
Die Aufgabenstellungen auf der dritten Seite des Arbeitsblatts greifen die Differenzierungsstufen Basis, Stern1 und Stern2 erneut auf (siehe dazu auch Kapitel 9.2). Sie zielen auf die Bewertung der Studie hinsichtlich Randomisierung; und die Aufgabe Stern2 lässt auch Deutungen in Richtung *Korrelation und Kausalität* zu. Außerdem werden die Lernenden aufgefordert, auf dem Prüfschema einzutragen, welchen Weg die Studie durch das Schema nimmt. Weitere Details zu dem Prüfschema werden im Mesozyklus 3 geklärt (Kapitel 11).
Die Unterrichtsstunde zu den Phasen IV und V wird im folgenden Abschnitt geplant. Aus der Erarbeitung der Prüfkriterien kommend, bietet es sich an, die Kontroverse

NEM zurück in das Gedächtnis der Schüler:innen zu bringen, indem das Meinungsbild aus Phase II im Plenum erneut gezeigt wird (z. B. an der Projektionsfläche). Die Lehrkraft erinnert daran, dass vor Beginn der dritten Phase zuletzt gefordert wurde, wissenschaftlich gesicherte Informationen in die Entscheidung einzubeziehen.

Mit diesem Hinweis als Gedächtnisaktivierung leitet die Lehrkraft zur Verteilung des Arbeitsblattes über. Sie erklärt, dass sich die Lernenden nun eine „echte Studie" anschauen, die auf Wissenschaftlichkeit geprüft werden soll. Das erledigen sie in einer ca. 15 minütigen Arbeitsphase. Anschließend werden die Arbeitsergebnisse im Plenum gesichert, die Lehrkraft kann z. B. das Arbeitsblatt projizieren und die Antworten der Lernenden mitschreiben. Dabei wird zum einen darauf fokussiert, dass die Stichprobe nicht randomisiert ausgewählt wurde und zum anderen wird die Schlussfolgerung der Forschenden in den Blick genommen. Diese ist zu allgemein gestaltet, denn die Studienergebnisse sind nur repräsentativ für die Gruppe der Schwangeren - sie sind nicht gültig für alle Menschen.

Zum Ende geführt werden kann die Unterrichtsstunde mit einem Blick auf die Meta-Studie. Die Lernenden sollten erkennen, dass die im Unterricht behandelte Studie lediglich einen kleinen Ausschnitt des Reviews darstellt, und dass die Forschenden eine größere Anzahl von Studien auf Wissenschaftlichkeit geprüft haben. Die Lehrkraft kann die Ergebnisse der Meta-Studie kurz vorstellen, um den Bogen sinnstiftend zu schließen. Eine kurze Bildschirmpräsentation mit Hintergründen zur Meta-Studie und den Ergebnissen in Form der folgenden Stichpunkte kann dafür verwendet werden:

- Bei älteren Personen hat Magnesium **keinen Effekt** auf Krämpfe
- Bei schwangeren Personen verringert Magnesium ein **ganz kleines Bisschen** das Auftreten von Krämpfen
- Bei Sportlern gibt es vermutlich **keinen Effekt** auf Krämpfe

Die Phase V bildet in diesem Format nur eine kurze Abschluss-Positionierung. Nachdem die Lernenden eine oder mehrere Studien in der Anwendungsphase in den Blick genommen haben, beziehen sie abschließend Stellung zum Thema. Das kann z. B. in Form eines kurzen „Blitzlichts" geschehen.

10.4.3 Erprobung

Die *Anwendung* und *Kriteriengeleitete Reflexion* wurden im Rahmen der Gesamterprobung im September 2020 mit 18 Lernenden der neunten Klasse einer Gesamtschule erprobt. Details zu den Gegebenheiten der Erprobung befinden sich im Kapitel 8.1.5.

Beschreibung der Unterrichtsstunde

Die Unterrichtsstunde startete mit einem Rückblick auf die beiden vorherigen Doppelstunden. Die Lehrkraft initiierte eine spielerische Wiederholungsphase, in der die Lernenden die wichtigsten Begriffe im Zusammenhang mit den Prüfkriterien im Plenum erklärten. Diese Phase dauerte von der Position 01:27 bis 14:04 (Videodatei) bzw. Absatz 2 bis 86 (Transkript 2020-10-09_Anwendung) an.
Die inhaltliche Arbeit an Phase IV begann anschließend mit der Projektion des Arbeitsergebnisses aus der ersten Doppelstunde, also der Positionierung in Phase II in Form des auf einem Smartphone abgebildeten Gruppenchats. Ein Schüler las die erste Nachricht vor, in der ein Mitglied des Gruppenchats nach einem Mittel gegen Muskelkrämpfe fragt. Die Lehrkraft stellte kurz das Meinungsbild dar, das in der Klasse ermittelt wurde (15:10, Pos. 89). Daraufhin teilte die Lehrkraft das Prüfschema aus und leitete die Arbeitsphase ein.
Die Lernenden lasen nacheinander die Arbeitsaufträge vor (18:07, Pos. 99; 18:37, Pos. 101; 18:50, Pos. 103). In der Stillarbeitsphase lösten die Schüler:innen die Aufgaben gemeinsam mit einer Partnerin oder einem Partner. Dabei konnten die folgenden Diskussionen in den Gruppen beobachtet werden.

Diskussion 1 (22:50-23:15, Pos. 117-123):

1-02: Ich finde es ist nicht randomisiert, weil es wurden ja nur Frauen ausgewählt.

1-01: Sind aber alles verschiedene [Frauen].

1-03: Also willst du schwangere Männer austesten?

1-02: Warum denn schwangere Männer? Ich hab gedacht es geht um Muskelkrämpfe.

1-03: Bei denen hast du auch Muskelkrämpfe.

1-02: Machen die Sport oder was?

1-03: Nein, aber die Muskeln brauchen auch Magnesium.

Diskussion 2 (24:24-24:48, Pos. 133-135):

2-02: Ich habe aufgeschrieben: Es wurde eine zufällige Stichprobe aus allen schwangeren Frauen gewählt, die diese Krämpfe haben. Soll ich nochmal langsamer vorlesen?

1-01: Warte, ich habe geschrieben: Ja es wurde randomisiert, da 67 schwangere Frauen verschieden getestet wurden.

2-02: Also verschieden getestet wurden die nicht. Ähm, einfach zufällig ausgewählt.

Diskussion 3 (28:00-28:43, Pos. 152-166):

S2: Die Studie ist nicht randomisiert (engl.), weil die sowieso schon auf einer Klinik waren, und da Medikamente gekriegt haben.

S1: Das steht doch drin oder nicht?

S2: Nein, da steht nur, dass sie untersucht worden sind

S1: Sowieso schon in eine Klinik gegangen.

S2: Aber das heißt ja nicht, dass die da schon Tabletten und so gekriegt haben.

C: Ne genau, das heißt es nicht. Aber das heißt einfach, dass die Frauen, die da jetzt untersucht wurden, sowieso regelmäßig da in der Klink sind/

S1: Also sind die nicht zufällig.

S2: Aber warum ist das dann nicht zufällig? Das sind doch trotzdem zufällige Frauen.

Im späteren Verlauf der Arbeitsphase beschäftigten sich die Schülerinnen und Schüler mit dem Prüfschema. Mit einem Stift kennzeichneten sie gemäß der Aufgabenstellung den Weg, den die Studie „durch das Prüfschema nimmt".
Eine nachfolgende Plenumsphase diente der Sicherung der Arbeitsergebnisse und bat Anlässe zur Diskussion. S1-01 las seine notierte Antwort vor: „Ich hab aufgeschrieben: Ja es wurde randomisiert, da 67 schwangere Frauen stichprobenartig und

zufällig getestet wurden“ (33:52, Pos. 195). S2-14 äußerte in einer Wortmeldung: „Es wurde nicht randomisiert, weil die Testpersonen schwanger waren“ (34:29, Pos. 199). S1-04 schloss sich der Meinung an: „Ähm, ich bin der Meinung, dass auch nicht randomisiert wurde, weil alle schwanger waren und alle auf der selben Klink waren, da alle öfter schon mal“ (34:52, Pos. 200).

Ein weiterer Punkt wurde von S1-12 angeführt: „Ähm es wurde eigentlich auch nicht randomisiert, weil wir wissen nicht, ob hier per Zufall ausgewählt wurde oder nicht. Also die Forscher haben das in dieser Studie nicht genannt“ (36:34, Pos. 216).

Die Lehrkraft fasste die Aussagen der Lernenden kurz zusammen (36:42, Pos. 218) und forderte von den Lernenden ein, die Studie im Prüfschema einzuordnen. Ein Schüler berichtete, dass bei Randomisierung *Nein* gewählt werden müsse, damit lande die Studie in der Mülltonne (38:37, Pos. 225).

Mit einem längeren Redebeitrag und anhand einer kurzen Bildschirmpräsentation erklärte die Lehrkraft die Bedeutung von Meta-Studien und stellte die Metastudie von Garrison et al. (2020) vor. S1-12 meldete sich und fragte, ob diese Meta-Studie denn jetzt nicht auch nochmal geprüft werden müsste auf Randomisierung und so etwas (42:09, Pos. 231).

Die abschließende Aktion der Unterrichtsstunde bildete die Methode *Blitzlicht*, bei der jede:r Lernende ein kurzes Statement dazu abgibt, was sie oder er denn nun, nach der Unterrichtseinheit, Paula raten würde: Die Einnahme von Magnesium-Präparaten oder anderweitige Behandlungen. Die Antworten der Lernenden werden nun kurz aufgelistet.

2-14: Ich würde Magnesium sagen.

2-13: Ich würde ihr raten Bananen zu essen.

2-12: Ich würde ihr raten, dass sie kein Magnesium nimmt, weil das ja keinen Effekt bringt.

2-11: Ich habe keine Ahnung.

1-13: Ich würde ihr auch kein Magnesium empfehlen, weil es nichts bringt.

1-12: Ich würde auch kein Magnesium empfehlen.

1-11: Ich würde ihr kein Magnesium empfehlen, weil die Chance, das das was bringt, ist nicht hoch. Man kann ja auch andere Sachen essen oder so.

2-10: Ich würde ihr empfehlen Bananen zu essen, weil wir in der Studie gesehen haben, dass Magnesium nichts bringt.

2-09: Ich würde ihr gar nichts empfehlen, weil nichts hilft irgendwie.

2-08: Ich würde ihr kein Magnesium empfehlen, weil es halt nichts bringt.

1-10: Ich würde ihr auch kein Magnesium empfehlen, weil es halt nichts bringt.

1-09: Ich würde ihr auch kein Magnesium empfehlen, weil es nichts bringt.

1-08: Ich würde ihr Magnesium empfehlen, aber ich würde auch was anderes ausprobieren.

2-07: Ich denke es geht irgendwann so von alleine weg, deshalb würde ich eigentlich nichts machen.

?-?: Ich würde kein Magnesium empfehlen. Ich denk Bananen.

?-?: Ich auch nicht, ich würde eher auch nur (unv.) nehmen

2-02: Ich würde es auch nicht empfehlen, da es einfach unsicher ist, dass es was bringt.

?-?: Ich würde ihr auch nichts empfehlen, weil es ja gar nicht hilft und ja/

1-03: Die soll halt selber entscheiden was hilft oder nicht.

1-02: Ich würde ihr raten vorm Sport richtig aufwärmen oder einfach Sport ausfallen lassen.

1-01: Ich würde ich raten Bananen zu essen und Bananensaft zu trinken.

Auswertung der Arbeitsblätter

In der ersten Aufgabe (Aufgabe Basis) sollten die Lernenden erklären, ob und wie bei der Studie randomisiert wurde (im Sinne einer zufallsbasierten Stichprobenauswahl).
Erwartet wurde hier die Aussage, dass die Stichprobe nicht randomisiert ausgewählt wurde, da die Testpersonen nur aus der Menge der schwangeren Frauen, die regelmäßig in die Klinik kommen, gewählt wurden. Weiterhin gibt es im Text keine Informationen darüber, ob bei der Auswahl Zufallsverfahren angewendet wurden. Bei der Studie wurde also nicht randomisiert.

Zehn Lernende erkannten die fehlende Randomisierung. Den wichtigsten Punkt, dass nur innerhalb der eigenen Patientinnen ausgewählt wurde, erkannten fünf Lernende. Zwei schrieben, dass das Aufteilungsschema nicht ersichtlich ist und daher von fehlender Randomisierung auszugehen ist. Die Schwangerschaften der Testpersonen nannten zwei Lernende als alleinige Gründe, ein:e Schüler:in gab keine Begründung an.
Sechs Schüler:innen hielten die Studie für randomisiert und begründeten diese Annahme damit, dass die Stichprobe (innerhalb der Gruppe der Schwangeren) zufällig ausgewählt sei.
Die Aufgabe Stern2 verlangte den Lernenden eine Einschätzung der Wissenschaftlichkeit der Studie ab. Sie sollten den Weg durch das Prüfschema beschreiben und die Studie bewerten.
Zwei Schüler:innen stuften die Studie als wissenschaftlich ein, da alle Kriterien erfüllt seien. 15 Mal wurde jedoch geschrieben, dass die Studie nicht wissenschaftlich sei. Davon wurde in sieben Fällen keine Begründung abgegeben und sieben Lernende schrieben, dass aufgrund von fehlender Randomisierung bereits der erste Punkt im Schema nicht erfüllt ist. Ein Lernender erwähnte, dass die Studie in der Mülltonne landet.

10.4.4 Diskussion und Re-Design

Die erprobte Unterrichtsstunde bildete den Abschluss einer dreiteiligen Unterrichtsreihe zur Bewertung kontroverser Themen. Das Ziel bestand darin, die Erkenntnisse aus der Gewinnungsphase der Wissenschaftskriterien zusammenzutragen und in einen Leitfaden zur Vorgehensweise bei der Prüfung von Aussagen und Studien umzuwandeln. Dazu wurde das Prüfschema eingeführt und von den Lernenden zur Bewertung von Nahrungsergänzungsmitteln verwendet. Aus den Diskussionen während der Gruppenarbeitsphase ist erkennbar, dass die Schüler:innen intensiv darüber nachgedacht haben, inwiefern die untersuchte Studie randomisiert wurde. Dabei sind sie sich nicht immer einig darüber, ob die Stichprobenauswahl zufällig war oder nicht. Einige argumentieren, dass als Probandinnen nur Frauen ausgewählt wurden, die ohnehin in die Klinik kamen. Dadurch treten Verzerrungseffekte auf und es kann nicht mehr von einer randomisierten Stichprobe gesprochen werden.
Andere Lernende halten dagegen, dass die Stichprobe zwar aus Schwangeren besteht,

die regelmäßig die Klinik besuchen - dass aber innerhalb dieser Gruppe zufällige Frauen ausgewählt wurden. Damit sei Randomisierung gegeben.
Beide Gruppen vertreten nachvollziehbare Standpunkte, allerdings beziehen sie den Aspekt der Repräsentativität in ihre Argumentationen nicht mit ein. Sie müssten darauf schauen, welche Schlussfolgerungen die Forschenden aus den Studiendaten ziehen und prüfen, ob diese überhaupt wissenschaftlich korrekt aus den Daten abgeleitet werden können.
Die Analyse der Gruppendiskussionen hat also gezeigt, dass die Lernenden das Konzept der Randomisierung zwar verstanden haben und dass sie Stichproben von Studien darauf prüfen können. Allerdings betrachten nicht alle Schüler:innen auch die Repräsentativität der Stichprobe, wodurch der Referenzrahmen von jeder und jedem Lernenden unterschiedlich definiert wird.
Zur Wissenschaftlichkeit der Studie gibt es unter den Schüler:innen ebenfalls gemischte Ansichten. Manche erkennen, dass die Studie direkt bei der Randomisierung durchfällt und daher unwissenschaftlich ist. Andere sehen die Studie als wissenschaftlich an, da alle Prüfkriterien, auf die getestet werden konnte, bestanden wurden.
Im Unterrichtsgespräch konnte dann geklärt werden, dass unter Beachtung der Repräsentativität das Prüfkriterium Randomisierung nicht erfüllt wird. Aus diesem Grund wurde die untersuchte Studie auch aus der Meta-Studie exkludiert.
Das Arbeitsblatt mit abgedrucktem Prüfschema wurde von den Schüler:innen zumeist korrekt verstanden. Das ist an den Begründungen zur Aufgabe Stern2 erkennbar, denn dort wird in sieben Fällen genannt, dass „bereits der erste Punkt“ des Schemas nicht erfüllt ist, wodurch die Studie in der Mülltonne landet. Lernende erkennen, dass das weitere Prüfen dieser Studie nicht relevant für die eigene Entscheidungsfindung wäre und daher entfällt. Da allerdings sieben Lernende keine Begründung für ihre Einstufung als *nicht wissenschaftlich* angaben, lässt sich nicht nachvollziehen, inwieweit diese Lernenden das Schema verstanden haben.
Das Blitzlicht zum Ende der Stunde gibt einen guten Überblick über die Entscheidungen der Lernenden bezüglich der Supplementierung von Magnesium zur Vermeidung von Muskelkrämpfen. Nur zwei Lernende entscheiden sich für eine Magnesium-Supplementierung, während 17 sich klar dagegen positionieren. Davon schlagen vier als Alternative Bananen vor; vier sagen, es sollte nichts unternommen werden oder jede:r müsse es selbst wissen. Acht Lernende geben keine Alternativen an. Die am häufigsten genannte Begründung gegen Magnesium-Präparate war, dass die-

se „nichts bringen“. Es ist aufgefallen, dass zum Ende der Blitzlicht-Methode mehrere Schüler:innen die gleiche Antwort gaben. Sie könnten also durch ihre Vorredner:innen beeinflusst worden sein - möglicherweise wurden zu dem Zeitpunkt auch bereits alle für die Lernenden relevanten Punkte genannt. Das wird als Nachteil der Methode gesehen, denn so findet nicht jede:r Lernende ihre oder seine eigene Wortwahl. Die im Rahmen des Re-Design überarbeitete Unterrichtsplanung sieht an dieser Stelle eine andere Methode vor, die eine Formulierung in eigenen Worten zulässt und gleichzeitig einen schnellen Überblick über das Meinungsbild ermöglicht. Geeignet sind digitale Tools wie z. B. *Mentimeter*, wobei jede:r Lernende eigenständig eine Entscheidung einträgt und diese anschließend an der Projektionsfläche dargestellt wird.

Anwenden und Reflektieren - Das sind die Hauptziele der vierten und fünften Phase im Unterrichtskonzept choice^{2}reflect . Angewendet wird das in Phase III erworbene Wissen auf die in Phase I und II eingeführte Kontroverse, dafür nutzen die Lernenden das Prüfschema. Reflektiert wird die auf der Grundlage des Prüfschemas getroffene Entscheidung gegenüber der Kontroverse.
Es konnte gezeigt werden, dass die entwickelten Lernmaterialien für einen stimmigen Abschluss der Unterrichtsreihe sorgen. Lernende *wenden* die Prüfkriterien *an* und *reflektieren* ihre Entscheidungen zielführend. Das übergeordnete Ziel der Konzeption, Bewertungskompetenz zu schulen und eine rationale Entscheidung zu ermöglichen, scheint damit erreicht werden zu können.
Eine derart umfangreiche Schlussfolgerung wird an dieser Stelle jedoch aus folgenden Gründen nicht gezogen: Zum einen wurden keine Bewertungskompetenz-Tests mit den Lernenden durchgeführt, sodass genaue Aussagen über die Veränderung der Bewertungskompetenz nicht möglich sind. Außerdem wurden im Rahmen der dritten Phase lediglich zwei Prüfkriterien erarbeitet. Die Lernenden konnten in der Anwendung somit nicht das gesamte Prüfschema durchlaufen.
Auf das Erreichen der umfangreichen übergeordneten Zielsetzung kann also anhand der beschriebenen Analysen nicht sicher geschlossen werden. Die vorliegenden Daten sind jedoch gut geeignet, um die oben genannten qualitativen Aussagen über die Eignung des Lernmaterials zu treffen. Darüber hinaus sind weiterführende Forschungsarbeiten nötig, in denen auch Bewertungskompetenzen erhoben und weitere Prüfkriterien in der Anwendungsphase verwendet werden.

11 Mesozyklus 3: Prüfschema zur Unterstützung der Prüfkriterien

11.1 Stellung der Prüfkriterien

Jungkamp identifizierte sieben Wissenschaftskriterien. Schüler:innen erlernen die Kriterien in der dritten Phase des Unterrichtskonzepts choice^{2}reflect und wenden das erworbene Wissen in Phase IV auf ein kontroverses Thema an. Lehrkräfte können aus den Lernmaterialien die für ihren Unterricht passenden auswählen und sich somit eine Unterrichtsreihe aus einzelnen Lehrmodulen konstruieren.

Die Prüfkriterien wurden bislang allerdings als gleichwertige, nebeneinanderstehende und voneinander isolierte Konzepte betrachtet. Aus der unsortierten Vielfalt an Materialien einzelne Inhalte auszuwählen, gestaltet sich daher als schwierig: Ohne vorgegebene Reihenfolgen oder Hierarchien ist es nicht klar, welche Prüfkriterien sich für einen alleinigen Einsatz eignen.

Auch für Schüler:innen sollte klar herausgestellt werden, welche Prüfkriterien es gibt und wie sie zueinander in Beziehung stehen.

Eine Strukturierungshilfe formuliert Jungkamp (2021) im Kapitel 4.2.9 seiner Arbeit: Er unterscheidet zwischen *Goldstandard-Kriterien* und *übergeordneten Prinzipien*. Die Goldstandard-Kriterien bilden die in der Literatur häufig beschriebenen Anforderungen an empirische naturwissenschaftliche Untersuchungen ab: Diese müssen randomisiert, kontrolliert und verblindet sein, um in der Wissenschaft anerkannt zu werden (Mad et al. 2008; Akobeng 2005; Döring & Bortz 2016, S. 102). Auch die *Reproduzierbarkeit* verortet Jungkamp hier, trennt diese jedoch etwas von den anderen Kriterien ab.

Bei den übergeordneten Prinzipien geht es vor allem um das Überprüfen von Aussagen und Behauptungen, daher fallen *Falsifizierbarkeit*, *Korrelation und Kausalität*

und *Plausibilität* unter diese Kategorie.

Für die unterrichtliche Praxis muss die Verortung der Prüfkriterien allerdings noch weiterentwickelt werden. Folgende Fragen sind bisher offen und werden im weiteren Verlauf dieses Kapitels behandelt:

- In welchem Verhältnis stehen einzelne Kriterien zueinander?
- Gibt es Abhängigkeiten?
- Wie unterscheiden sich die Anwendungsbereiche?

11.1.1 Abgrenzung der Goldstandard-Kriterien von übergeordneten Prinzipien

Bei der Suche nach Fakten zu einer wissenschaftlichen Kontroverse gibt es verschiedene Möglichkeiten, mit Informationen in Berührung zu kommen. Sie können in persönlichen Gesprächen in Form von Behauptungen oder Faktenaussagen geäußert werden, in sozialen Medien als Beiträge veröffentlicht oder auch in der Werbung plakativ herausgestellt werden.
In manchen Fällen begegnet man den originalen empirischen Studien, die einen Teilaspekt der Kontroverse behandeln. Möglicherweise findet man auch Aussagen *über* empirische Studien und deren Ergebnisse – diese werden wiederum häufig in sozialen Medien angetroffen.
Die Wissenschaftskriterien befassen sich damit, Informationen aus allen möglichen Informationsquellen auf Wissenschaftlichkeit zu prüfen.
Der Anwendungsbereich sieht für jedes Prüfkriterium etwas anders aus, das wird in der Phase III bisher jedoch nicht unterschieden.
Die von Jungkamp als *übergeordnete Prinzipien* bezeichneten Prüfkriterien sind vor allem auf *Behauptungen* anwendbar. Sie behandeln aussagenlogische Aspekte und beziehen sich direkt auf Syntax und Inhalt der Aussage an sich. Das Kriterium *Falsifizierbarkeit* beinhaltet zunächst eine Prüfung auf die prinzipielle Widerlegbarkeit der Aussage. Gibt es ein (Gedanken-)Experiment, mit dem die Aussage direkt widerlegt werden könnte? Trifft das zu, muss auch die inhaltliche Dimension der Behauptung auf Widerlegbarkeit geprüft werden (ausführlich beschrieben im Kapitel 9.1).
Als weiteres übergeordnetes Prinzip nennt Jungkamp die *Plausibilität*. Es wird jedoch bereits angedeutet, dass eine Ausarbeitung von Lernmaterial für dieses Kriterium

nur wenig sinnvoll erscheint, da es nicht die nötige Trennschärfe besitzt. Im Laufe dieses Projekts wurde die Plausibilität aus dem Pool der Prüfkriterien vollständig entfernt, denn sie hat sich als für Schüler:innen schlecht handhabbar erwiesen.
Zwar sollte eine Behauptung, die in ernsthaften wissenschaftlichen Diskursen Bestand haben soll, plausibel sein - wie genau jedoch eine plausible Aussage definiert ist, kann nicht trennscharf auf den Punkt gebracht werden. Es ist ein gewisses Maß an „Bauchgefühl“ nötig, um plausible von nicht-plausiblen Aussagen zu unterscheiden. Diese Verwendung von Bauchgefühl widerspricht jedoch dem Kerngedanken des Unterrichtskonzepts choice^{2}reflect , dass Wissenschaft klar kriteriengeleitet ist. Ein „schwammiges“ Prüfkriterium hilft hier nicht weiter.
Klarer definiert ist der Unterschied zwischen „Korrelationen und Kausalitäten“ (siehe Kapitel 9.2). Eine Korrelations-Aussage bringt zwei Variablen miteinander in Zusammenhang und drückt dabei aus, dass ein Faktor *zusammen* mit einem anderen auftritt. Ob sich die Faktoren gegenseitig bedingen, also einer die Ursache des anderen darstellt, ist häufig nicht eindeutig bestimmbar. „Menschen mit niedrigen Magnesium-Serum-Leveln leiden häufiger an Muskelkrämpfen“ ist zunächst die Beschreibung eines statistischen Zusammenhangs - die Variablen korrelieren. Die Aussage: „Ein niedriges Magnesium-Serum-Level löst Muskelkrämpfe aus“ behauptet jedoch, dass ein niedriger Magnesiumspiegel der Auslöser für Muskelkrämpfe sei. Es handelt sich um eine Kausalitäts-Aussage. Inwiefern hier jedoch tatsächlich eine Kausalität vorherrscht, lässt sich nur durch eine einwandfreie Datenlage belegen.
Um auf *Korrelation und Kausalität* prüfen zu können, reicht es nicht aus, die reine Behauptung zu betrachten. Es sind stets auch empirische Ergebnisse heranzuziehen, z. B. aus Experimenten oder Studien.
Genau an dieser Stelle zeigt sich die Grenze zwischen den *übergeordneten Prinzipien* auf der einen, und den *Goldstandard-Kriterien* auf der anderen Seite. Behauptungen werden auf Falsifizierbarkeit geprüft, die inhaltliche Aussage hinter der Behauptung muss durch Empirie gestützt werden.

11.1.2 Goldstandard-Kriterien

Der erste Schritt in Richtung einer rationalen Entscheidungsfindung ist getan - eine Behauptung ist prinzipiell falsifizierbar und besteht damit die aussagenlogische Prüfung. Aber kann die Behauptung, dessen Falsifizierbarkeit festgestellt wurde, auch

empirisch gestützt werden? Gibt es empirische Studien oder Experimente, die die Behauptung bestätigen?
Die Antwort darauf wird entweder von der Quelle der Behauptung mitgeliefert, z. B. in einer Pressemitteilung, die sich auf eine bestimmte Studie bezieht („Eine Studie hat herausgefunden, dass Magnesiummangel Muskelkrämpfe auslöst!"); oder muss ansonsten selbst recherchiert werden.
Liegt die Studie vor, auf die sich die Behauptung bezieht, muss zunächst geprüft werden, inwiefern der Inhalt der Behauptung mit der Schlussfolgerung der Studie übereinstimmt.
Stimmen Behauptung und Studien-Ergebnis überein, ist der nächste Schritt ein Blick in den Methodenteil der Studie, um herauszufinden, inwiefern die Schlussfolgerung (also die Hauptaussage der Untersuchung) aus den vorhandenen Daten gezogen werden darf. Das kann, vor allem für Schüler:innen, eine schwierige Aufgabe darstellen, denn es ist ein umfangreiches Prüfen der gesamten Methodik nötig.
Im Rahmen des Prüfkriteriums *Korrelation und Kausalität* werden mögliche Störfaktoren antizipiert, die als weitere Einflussfaktoren auf die untersuchte Variable infrage kommen könnten. Erarbeitet wird dieses Kriterium mit dem Kontext: „Einfluss der App *Instagram* auf die Akkulaufzeit von Smartphones", siehe dazu Kapitel 9.2. Wird dabei festgestellt, dass keine Scheinkausalität vorhanden ist, kann die methodische Prüfung anhand der Goldstandard-Kriterien fortgesetzt werden.
Die Goldstandard-Kriterien erlauben eine strukturierte Herangehensweise an die Prüfung der Studien-Methodik. Sie beleuchten jeweils Teilaspekte empirischer Untersuchungen und erlauben zusammengenommen eine Einschätzung, inwiefern die geprüfte Studie wissenschaftlichen Standards genügt. Zu diesen Kriterien gehören: *Randomisierung*, *kontrollierte Bedingungen*, *Verblindung* und *Reproduzierbarkeit*. Im folgenden Abschnitt wird die Verortung der Kriterien konkretisiert und es werden Beziehungen zwischen den Prüfkriterien aufgezeigt.

Randomisierung, Verblindung und kontrollierte Bedingungen

Randomisierung wird klassisch verstanden als zufallsbasierte Verteilung von Testpersonen auf Experimental- und Kontrollgruppe (siehe Kapitel 9.3). Im Rahmen des Unterrichtskonzepts choice^{2}reflect fallen unter die Randomisierung jedoch eher die zufallsbasierte Stichprobenauswahl und die Betrachtung von Repräsentativitäten bei

der Bildung von Schlussfolgerungen. Der Fokus des Prüfkriteriums ist also in den Lernmaterialien etwas anders gesetzt als in der Literatur.

Dies hängt damit zusammen, dass die Lernenden beim Prüfen einer Studie den Weg nachvollziehen sollen, wie eine Studie überhaupt entsteht: Zu Beginn einer Untersuchung wird die Stichprobe festgelegt. Dabei ist es elementar, auf welche Weise die Proband:innen rekrutiert werden (möglichst randomisiert und ohne bewusste Selektion).

Der Aspekt einer randomisierten Zuteilung der Testpersonen wird dann im Rahmen des Prüfkriteriums *Verblindung* aufgegriffen. Dabei erlernen die Schüler:innen zum einen, dass sowohl Versuchspersonen als auch Versuchsleitende und -durchführende nicht wissen dürfen, welche Personen der Experimentalgruppe zugeordnet wurden und welche sich in der Kontrollgruppe befinden. Zum anderen wird auch thematisiert, wie Verblindung in experimentellen Settings sichergestellt werden kann.

Ebenfalls eng verbunden mit der Diskussion über die Versuchsbedingungen ist das Prüfkriterium *kontrollierte Bedingungen*. Dieses besagt, dass bei Experimenten stets nur ein Faktor variiert werden darf, während (möglichst) alle anderen konstant gehalten werden. Das Konzept erarbeiten die Lernenden anhand der Herstellung eines möglichst stabilen Milchschaums. Sie erstellen dazu Versuchsreihen und variieren Faktoren wie die Temperatur der Milch, Technik oder Dauer des Aufschäumens (Marohn & Jungkamp 2016; Jungkamp 2021).

Hier findet ein Rückbezug zum Prüfkriterium *Korrelation und Kausalität* statt: Damit möglichst viele Einflussfaktoren konstant gehalten werden können, ist es wichtig, auch möglichst viele Faktoren zu identifizieren. Davon hängt wiederum die Güte der Schlussfolgerung ab. Nur bei Betrachtung und Konstanthaltung aller Einflussfaktoren sind Kausalaussagen möglich.

Diese Ausführungen zeigen, dass eine genaue Abgrenzung der Prüfkriterien voneinander sehr schwierig ist. Bei der Erklärung eines Kriteriums sind Inhalte eines anderen anzuwenden, welche ebenfalls die Einhaltung von Teilen eines weiteren Prüfkriteriums voraussetzen. Das erschwert die Verbalisierung der Prüfkriterien und lässt Grenzen zwischen ihnen verschwimmen. Allerdings ist dieser Sachverhalt auch als Chance zu verstehen: Durch die umfangreiche Erarbeitung *eines* der Goldstandard-Kriterien erhalten Lernenden auch bereits Einblicke in die anderen Kriterien, die implizit im Material mitschwingen und in Teilen intuitiv verständlich sind.

Alle Prüfkriterien nacheinander abzuarbeiten, ist also nicht obligatorisch für die Weiterentwicklung des Wissenschaftsverständnisses von Lernenden. Im Sinne des *Cluster Approaches* baut sich das Verständnis wissenschaftlichen Arbeitens jedoch Stück für Stück auf, je mehr Wissenschaftstheorie vermittelt wird (Jungkamp 2021, S. 104). Randomisierung führt zu kontrollierten Bedingungen, allerdings nur bei konsequent durchgeführter Verblindung. Kontrollierte Bedingungen - sofern „alle" wichtigen Störfaktoren in die Kontrolle einbezogen werden - erlauben die Bildung von Kausalaussagen. Kausalaussagen müssen allerdings falsifizierbar formuliert werden, um auch überprüfbar zu sein. Sie werden schließlich von Medien aufgegriffen und an die Öffentlichkeit kommuniziert.
Öffentlich kommunizierte Kausalaussagen müssen von den Empfänger:innen wieder auf Wissenschaftlichkeit geprüft werden. Denn woher soll jemand wissen, inwiefern die soeben empfangene Aussage auf wissenschaftlich gültigen Informationen beruht?

Ein unendlicher Regress deutet sich an. Ein Regress, der jedoch den wissenschaftlichen Diskurs vorantreibt und rationale Entscheidungen ermöglicht! Eine Behauptung zu einem Sachverhalt wird aufgestellt, unter Einhaltung des Goldstandards empirisch untersucht, unter Berücksichtigung von Scheinkausalitäten bestätigt und die Bestätigung in Form einer falsifizierbaren Aussage weiter kommuniziert. Die Aussage wird erneut geprüft, das Ergebnis kommuniziert, usw.
Für das Gelingen einer unendlichen Fortführung dieses Kreislaufs ist das letzte Prüfkriterium elementar.

Unabdingbar: Die Reproduzierbarkeit

> „Wird ein Experiment unter gleichen Bedingungen und von anderen Wissenschaftler:innen wiederholt, so muss stets das gleiche Ergebnis erhalten werden" (siehe Kapitel 9.4 dieser Arbeit).

Forschende zielen auf eine hohe interne wie auch externe Validität ihrer Versuchsergebnisse ab. Dafür begeben sie sich mit ihren Forschungen in den oben beschriebenen Kreislauf und befolgen den *Goldstandard*. Für die Gesellschaft wenig hilfreich sind die Forschungen jedoch, wenn ihnen Reliabilität und Objektivität fehlt - andere Wissenschaftler:innen müssen die Ergebnisse unter gleichen Bedingungen reproduzieren können, um am Kreislauf teilhaben zu können. Dafür müssen Forschungen

umfangreich und *reproduzierbar* dokumentiert sein. Auch diese Anforderung findet Platz in der Phase III des Unterrichtskonzepts choice^{2}reflect . Sie wird durch das Prüfkriterium *Reproduzierbarkeit* ausgedrückt. Aufgrund der hohen Relevanz wird die Reproduzierbarkeit im Rahmen dieser Arbeit auch als Goldstandard-Kriterium angesehen.

11.2 Explizieren der Verortung von Prüfkriterien im Unterricht: Ein Prüfschema

Die Verortung der sieben Prüfkriterien hat eine neuartige Struktur hervorgebracht. Erstmals lassen sich die Prüfkriterien in unterschiedliche Sinnabschnitte gliedern, in eine Reihenfolge bringen und in einem Flussdiagramm darstellen. Aussagenlogische Prüfkriterien stehen empirischen Prüfkriterien gegenüber, damit werden die im Alltag relevanten Informationsquellen abgedeckt und für wissenschaftliche Prüfverfahren zugänglich gemacht. Dieses Verfahren zur Prüfung von Behauptungen, experimentellen Ergebnissen und empirischen Studien sollte auch für Lernende aufbereitet werden. Das explizite Thematisieren der Vorgehensweise im Unterricht sorgt dafür, dass die Schüler:innen zu jeder Zeit nachvollziehen können, an welchem Punkt der Prüfung sie sich befinden und welchen wissenschaftlichen Status der zu prüfende Gegenstand aktuell hat.
Schüler:innen erarbeiten damit nicht mehr unzusammenhängende Prüfkriterien; stattdessen können sie jedes Modul des Unterrichtskonzepts in einem Schema verorten.
Die visuelle Gestaltung dieses Prüfschemas wird im folgenden Abschnitt entwickelt. Dabei wird der aktuelle Stand des Schemas zum Zeitpunkt des Verfassens dieser Arbeit beschrieben und als *Version 1* bezeichnet. Damit unterscheidet sich die hier gewählte Vorgehensweise von den Darstellungen der ersten beiden Mesozyklen: Dort wurden ausgehend von einer „Erstversion“ des Materials die Erprobung und Analyse beschrieben, um aus den Ergebnissen Implikationen für ein Re-Design abzuleiten.
Die Entwicklung des Prüfschemas hat jedoch mit der Version 1 bereits mehrere Mesozyklen durchlaufen. Diese lassen sich hier nicht im Einzelnen beschreiben, da die Änderungen sehr vielfältig waren und die Schritte der Mesozyklen teilweise mit starkem zeitlichen Versatz bewältigt wurden. Grundsätzlich folgten auf einen Entwurf (*Design*-Phase) stets persönliche Gespräche mit Projekt-Angehörigen und Kolleg:in-

nen, teils im Rahmen interner Kolloquien. Gestaltungselemente wurden durchgesprochen, Einsatzmöglichkeiten im Unterricht antizipiert und auf dieser Grundlage verbessert.

Dies entspricht gemäß dem DBR-Ansatz (siehe Kapitel 5.4) den Phasen *Erprobung*, *Analyse* und *Re-Design*; wobei die Erprobung hier eher als Vorerprobung bzw. Antizipation eines praktischen Einsatzes zu verstehen ist.

Im Folgenden werden die Gedankengänge, die zur Entwicklung der Version 1 (V1) geführt haben, ausführlich dargelegt. Weiterhin wird ein Abriss über die Entwicklungsstadien bis hin zur V1 gegeben, um einen Einblick in die zuvor absolvierten Durchläufe des Mesozyklus zu geben.

Die erste richtungsweisende Entscheidung zur Entwicklung des Prüfschemas bestand darin, kategorisch zwischen aussagenlogischen und empirischen Kriterien zu unterscheiden. Da die Aussagenlogik zuerst geprüft wird, wird sie auf der linken Seite des Schemas angeordnet. Angewendet wird die „linke Seite" auf Behauptungen, denen man im Alltag begegnet. Hier bietet sich der Begriff *These* als „behauptend aufgestellter Satz, der als Ausgangspunkt für die weitere Argumentation dient" (Duden 2020b) an. Dieser wird oben links in großer Schriftart angeordnet und eingekreist, um den Startpunkt zu markieren.

Wie bereits geklärt wurde, muss eine These *widerlegbar* sein, um in der Wissenschaft Bestand haben zu können. Daher wird vom Begriff These eine direkte Verbindung zu einem größeren Rahmen mit dem zentralen fettgedruckten Begriff *Falsifizierbarkeit* gezogen. Dieser Rahmen symbolisiert ein Prüfkriterium. Eine kurze, Schlagwort-ähnliche Frage fasst den zentralen Inhalt des Kriteriums zusammen: *These widerlegbar?* Zusätzlich befindet sich in dem Rahmen das Symbol für die Falsifizierbarkeit (entwickelt im Kapitel 9.1 bzw. bei Kralisch 2017; Jungkamp 2021).

Der Rahmen hat zwei Ausgänge. Zum einen kann die These das Prüfkriterium nach links verlassen und „fällt" durch einen Trichter, der mit dem Wort: „NEIN" versehen ist. Von dort führt eine rote Linie direkt in eine Mülleimer-Grafik. Dieses drastische Bild wurde gewählt, um besonders deutlich zu machen, dass eine nicht-falsifizierbare These unwissenschaftlich ist.

Der andere mögliche Ausgang aus dem Prüfkriterium ist ein Trichter mit dem grün geschriebenen Wort: „JA". Dies steht für eine bestandene Falsifizierbarkeits-Prüfung. Per Definition kann eine falsifizierbare These auch als Hypothese bezeichnet werden

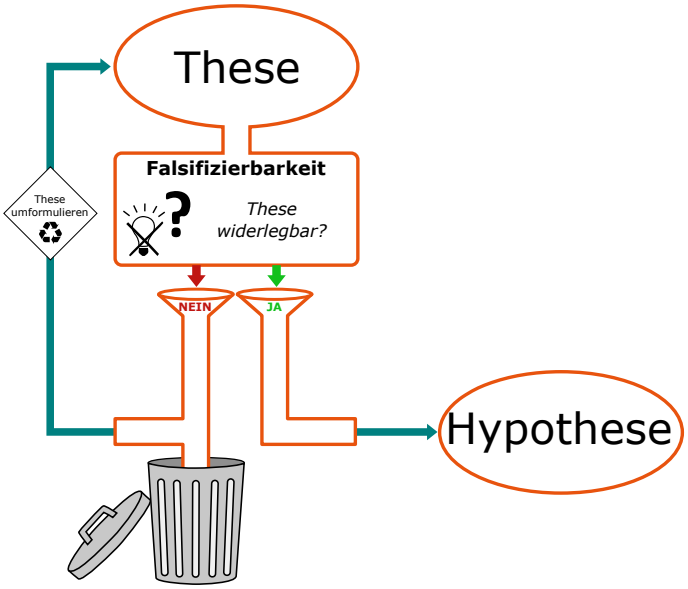


Abbildung 11.1: *Ausschnitt aus dem Prüfschema. Gezeigt wird die linke Seite mit aussagenlogischer Prüfung auf Falsifizierbarkeit.*

(Duden 2020a, siehe auch Kapitel 2.3.3). Daher führt eine Linie zum elliptisch eingerahmten Wort *Hypothese*.

Diese linke Seite des Prüfschemas ist in Abbildung 11.1 dargestellt. Dort fällt auf, dass es eine Rückschleife für den Fall gibt, dass die Falsifizierbarkeit nicht erfüllt ist. Eine nicht-falsifizierbare These kann umformuliert und erneut auf Widerlegbarkeit geprüft werden. Damit wird die drastische Mülleimer-Symbolik etwas relativiert. In manchen Fällen lohnt es sich, den Inhalt hinter nicht-falsifizierbaren Aussagen trotzdem in naturwissenschaftlichen Diskussionen zu verwenden.

Popper (1989) schreibt dazu, dass auch „kühne Hypothesen“ geprüft werden sollten. Daher bietet die Rückschleife im Prüfschema die Möglichkeit, eine Aussage doch noch falsifizierbar zu gestalten und in den wissenschaftlichen Diskurs zu integrieren.

Das Bild eines Trichters wird im gesamten Prüfschema wiederzufinden sein. Es symbolisiert das „Reduzieren“ der großen Vielfalt möglicher Thesen auf diejenigen, die wissenschaftlichen Standards entsprechen. In ersten Entwürfen des Prüfschemas wurde die Trichter-Symbolik deutlich präsenter eingesetzt, wie es in Abbildung 11.2 erkennbar ist. Diese Darstellungsweise wurde jedoch u. a. zugunsten der Lesbarkeit und für eine klarere Struktur verworfen. Auch wäre ein Trichter mit zwei Ausgängen irreführend gewesen.

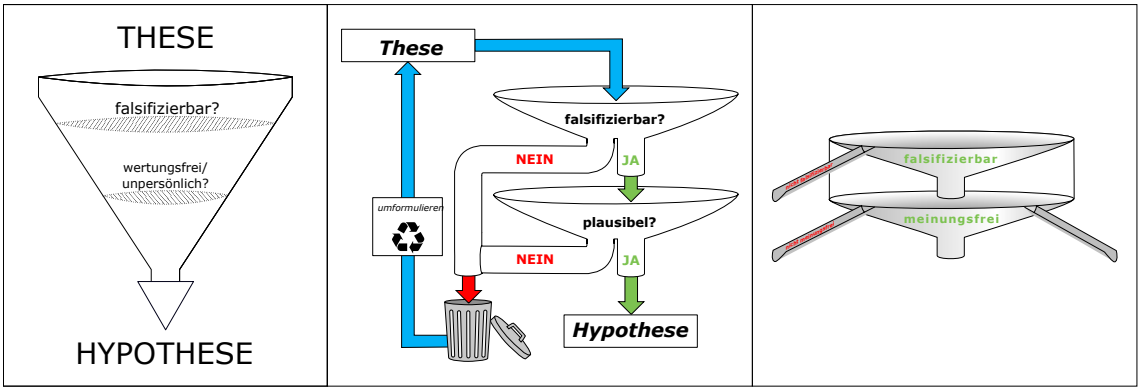


Abbildung 11.2: *Darstellung früherer Versionen der linken Seite des Prüfschemas. Erkennbar ist ein starker Wandel der Verwendung der Trichter-Symbolik.*

Auf dem Weg zu einer wissenschaftlich geprüften Aussage durchläuft die Hypothese anschließend die rechte Seite des Prüfschemas. Diese enthält die Goldstandard-Kriterien in der oben erläuterten Reihenfolge. Allerdings muss in dem Schema deutlich werden, dass der Anwendungsbereich der Prüfkriterien sich von der linken zur rechten Seite ändert. Links werden *Thesen* geprüft, rechts die empirischen Untersuchungen. Als Startpunkt für die rechte Seite dient somit der elliptisch eingerahmte Begriff *Studie*.

Es ergibt sich ein Dreieck zwischen These, Hypothese und Studie. Es verdeutlicht den Lernenden, welcher Gegenstand „von oben nach unten fällt", bzw. das Schema durchläuft. Den Übergang von der Hypothese zur Studie bildet eine Linie, die sich in der Mitte aufspaltet und in zwei Rauten mit der Aufschrift „Studie planen" bzw. „Studie recherchieren" mündet.

Um den Weg der naturwissenschaftlichen Erkenntnisgewinnung lückenlos nachvollziehen zu können, müssten Lernende idealerweise eine Studie selbst planen, mit der die Hypothese untersucht werden kann. Im Fall von kontrovers diskutierten Themen ist die Lage allerdings meist sehr komplex und würde die Kontrolle vieler Störvariablen erfordern. Handelt es sich um ein gesundheitsbezogenes Thema, würde eine Studienplanung noch komplizierter werden.

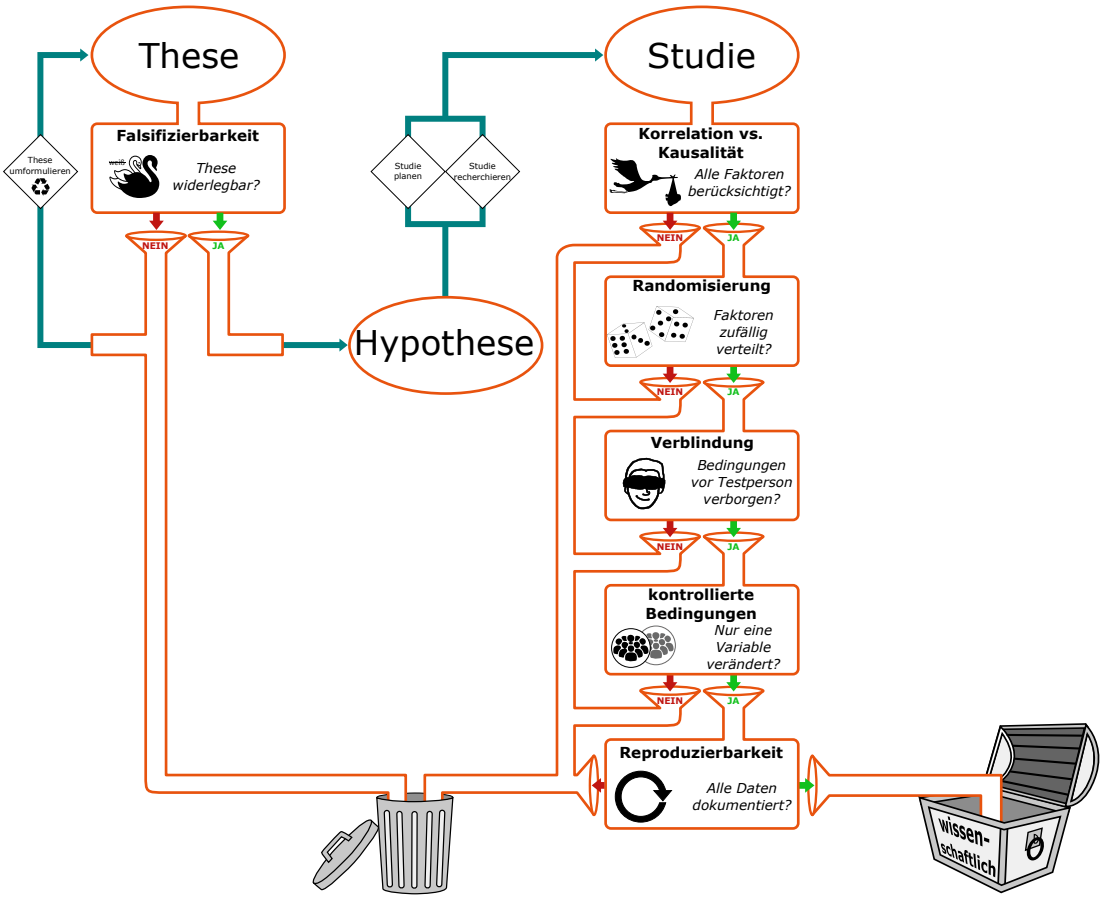
Zur Untersuchung von Gesundheitsfragen eignen sich meist nur Kohortenstudien, in denen Gruppen von Testpersonen über einen langen Zeitraum beobachtet werden. Experimentelle Designs, bei denen gezielt Variablen beeinflusst werden, eignen sich unter anderem aus ethischen Gründen nicht: Bei der Frage, ob ein bestimmter Nähr-

stoffmangel zu schweren Organschäden führt, ist es nicht möglich, eine Gruppe von Testpersonen gezielt einer Mangelernährung auszusetzen.
Für die Lernenden bleibt also die Möglichkeit, bereits vorhandene Studien zur Thematik zu recherchieren und diese anhand der rechten Seite des Prüfschemas zu prüfen.
Auf der rechten Seite begegnen die Schüler:innen zuerst dem Prüfkriterium *Korrelation und Kausalität*. Dieses wird von Jungkamp (2021) zwar bei den übergeordneten Prinzipien verortet. Allerdings ist es sowohl auf Aussagen als auch auf empirische Studien anwendbar, wie oben bereits beschrieben wurde.
Im hier beschriebenen Prüfschema wurde *Korrelation und Kausalität* der rechten Seite zugeordnet, da es am besten zu überprüfen ist, wenn bereits die empirischen Studiendaten vorliegen. So können Lernende besser erkennen, ob eventuell weitere Störfaktoren existieren, die nicht in die Schlussfolgerung der Studie eingeflossen sind, als wenn sie dies nur anhand einer alleinstehenden Behauptung prüfen sollen.
Die Darstellung der Prüfkriterien auf der rechten Seite ist identisch mit der Rahmen- und Trichterdarstellung bei der Falsifizierbarkeit. Nichterfüllen eines Prüfkriteriums resultiert in dem Aussortieren der Studie als unwissenschaftlich („Mülltonne"), während ein bestandenes Kriterium die Studie weiterrücken lässt zum nächsten Rahmen.
Gemäß den oben erörterten Zusammenhängen zwischen den Wissenschaftskriterien folgen im Prüfschema die *Randomisierung*, *Verblindung*, *kontrollierte Bedingungen* und schließlich die *Reproduzierbarkeit*. Das letzte Element des Prüfschemas ist die Abbildung einer Schatzkiste mit der Aufschrift „wissenschaftlich". Wurden alle Prüfkriterien bestanden, landet die geprüfte Studie in der Schatzkiste und kann somit zur Entscheidungsfindung verwendet werden. Die Prüfung auf Wissenschaftlichkeit ist damit abgeschlossen.
Das fertig entwickelte Prüfschema befindet sich in der Abbildung 11.3.

11.3 Entscheidungsfindung nach Absolvieren des Prüfschemas

Im Anschluss an das vollständige Durchqueren des Prüfschemas ist eine Studie entweder als unwissenschaftlich deklariert worden oder in der wissenschaftlichen Schatzkiste gelandet. Erst wenn Letzteres eintritt, darf dem Fazit der Studie vertraut wer-

Abbildung 11.3: *Ein Prüfschema zur Bewertung kontroverser Themen in den Phasen IV und V der Unterrichtskonzeption choice² reflect , Version 1.*



den. Nun wird herausgearbeitet: Gelangt die Studie zu einem Ergebnis, das die Hypothese unterstützt? Oder widerspricht sie der Eingangs-Hypothese?
Möglicherweise stellt sich dabei heraus, dass weitere Forschung nötig ist, um eine eindeutige Aussage treffen zu können. Bei gesundheitsbezogenen Themen oder auch anderen naturwissenschaftlich komplexen Themen kann das häufig auftreten. Studien zu diesen Themen sind meist nicht allgemeingültig, da sie nur bestimmte Personengruppen untersuchen oder ganz spezifische Aspekte des Themas abdecken.
In jedem Fall sollte eine Entscheidung nicht auf der Grundlage *einer* Studie getroffen werden. Es sollten mehrere Untersuchungen mit ähnlichen Hypothesen betrachtet und mithilfe des Prüfschemas geprüft werden, um letztendlich eine breite Aufstellung von Studienergebnissen „in der Schatzkiste“ zu erzielen.

11.4 Erprobung des Prüfschemas

11.4.1 Methodik und Aufgabenentwicklung

Das Prüfschema wurde rein theoriebasiert entwickelt. Es enthält Implikationen aus der Entscheidungs- und Wissenschaftstheorie, implementiert die von Jungkamp identifizierten Prüfkriterien und behält die Trennung zwischen übergeordneten Prinzipien und Goldstandard-Kriterien bei.
Bis zur aktuellen Version 1 hat das Schema bereits Verbesserungen erfahren, die durch kleinere Vorerprobungen zustande gekommen sind. Im Fachunterricht mit Schülerinnen und Schülern wurde es jedoch noch nicht eingesetzt.
Im Rahmen einer Masterarbeit von Theresa Flach wurde das Prüfschema in zwei Durchläufen eines universitären Seminars zu naturwissenschaftlichen Unterrichtskonzepten zusammen mit je 26 Studierenden erprobt. Details zu den Gegebenheiten finden sich im Kapitel 8.1.6.
Flach (2021) untersucht die Fragestellung: „Welche Herangehensweise verfolgen Studierende bei der Bewertung eines kontroversen Themas unter Nutzung des von Kralisch und Marohn entwickelten Prüfschemas?“
Von besonderem Interesse ist dabei, welche Wirkung das Prüfschema auf Studierende hat, wenn es ihnen zum ersten Mal begegnet. Die Intention bei der Entwicklung des Schemas war es, ein intuitiv verständliches Ablaufschema für das Prüfen auf Wissenschaftlichkeit zu erstellen. Inwiefern Studierende damit direkt zielführend arbei-

ten können, wurde untersucht. Dazu wurden die Studierenden vor Durchführung der Studie nicht instruiert - sie erhielten weder eine Einführung in das Unterrichtskonzept choice^{2}reflect noch wurde ihnen das Prüfschema erklärt.
Um die Herangehensweise der Lernenden strukturiert analysieren zu können, unterteilte Flach die Forschungsfrage in folgende Subfragen, dessen Formulierung für die Verwendung in diesem Projekt leicht angepasst wurde:

1 Welches Vorwissen haben die Studierenden bereits zu den einzelnen Prüfkriterien?
2 Inwiefern gehen die Studierenden gemäß der europäischen Leserichtung den richtigen Weg des Schemas?
3 Inwiefern treffen die Studierenden mithilfe des Prüfschemas eine rationale Entscheidung bei der Bewertung einer These/Studie?
4 Welche Reaktionen zeigen Studierende beim ersten Blick auf das Prüfschema?

Abbildung 11.4 zeigt einen Ablaufplan der Seminarsitzung. Das Vorwissen (Subfrage 1) wird anhand eines offenen Fragebogens zu Beginn des Seminars abgefragt. Bereitgestellt wird der Fragebogen in einer digitalen Form über die Lernplattform *Learnweb*. Die Studierenden werden aufgefordert, zu jedem der sechs Prüfkriterien die aufgeführte Aussage zu erklären, beispielsweise sollen sie den Satz: „Eine Korrelation ist keine Kausalität“ erklären. Die im Fragebogen verwendet Aussagen befinden sich in der Tabelle 11.1.

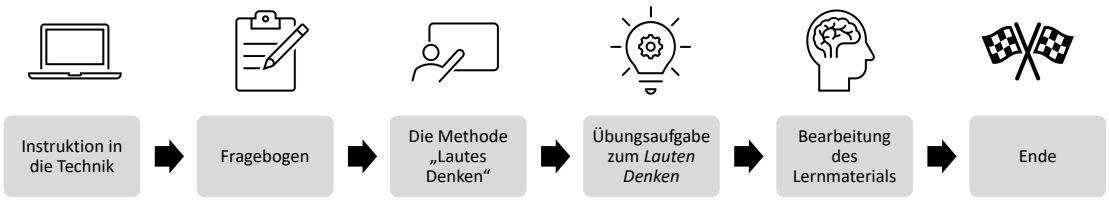


Abbildung 11.4: *Ablaufplan der Seminarsitzungen im Juli 2021 zur Erprobung des Prüfschemas.*

Im Anschluss an die Fragebogenerhebung erfolgt die Erprobung des Prüfschemas. Die Studierenden werden dazu angehalten, ab diesem Punkt des Seminars die Methode *Lautes Denken* durchzuführen. Das ermöglicht den Forschenden, „Einblicke in

Tabelle 11.1: *Auflistung über die im Fragebogen verwendeten Aussagen zur Erklärung der Prüfkriterien.*

Prüfkriterium	Aussage
Falsifizierbarkeit	Eine Aussage muss falsifizierbar sein.
Korrelation & Kausalität	Eine Korrelation ist keine Kausalität.
Randomisierung	Eine Studie muss randomisiert durchgeführt werden.
Verblindung	Eine Studie muss verblindet durchgeführt werden.
Kontrollierte Bedingungen	Eine Studie muss unter kontrollierten Bedingungen durchgeführt werden.
Reproduzierbarkeit	Eine Studie muss reproduzierbar sein.

Gedanken, Gefühle und Absichten einer lernenden und/oder denkenden Person“ zu erhalten (Konrad 2010, S. 374). Die Teilnehmenden verbalisieren möglichst ausführlich ihre Gedanken und äußern auch Wahrnehmungen, Gefühle und Handlungen. Dabei erfüllen sie eine Primäraufgabe. Das Verbalisieren der Tätigkeiten macht den Forschenden die kognitiven Prozesse in Form eines Protokolls zugänglich (Bilandzic & Trapp 2000; Heine 2014).

Die Primäraufgabe besteht in der Bewertung einer Studie zum Thema *Veganismus* anhand des Prüfschemas. Flach (2021) identifiziert Veganismus als aktuelles und für Studierende relevantes *socio-scientific issue*. Die Studierenden befassen sich mit der provokativen Aussage: „Veganer haben mehr sportliche Ausdauer und genauso viel Kraft wie Fleischesser“. Diese fungiert als Aufmacher eines Online-Artikels, der die Ergebnisse einer Studie aus dem *European Journal of Clinical Nutrition* beschreibt und die zentrale Aussage der Studie als Überschrift verwendet. Ein Ausschnitt aus dem Artikel befindet sich in der Abbildung 11.5.

Die im Artikel zitierte Studie bzw. der Abstract der Studie wird den Studierenden ebenfalls als Arbeitsmaterial zur Verfügung gestellt. Eine Vereinfachung des Studien-Texts erfolgt nicht, da es sich bei den Teilnehmenden um Studierende des fünften Hochschulsemesters handelt. Lediglich eine Übersetzung in die deutsche Sprache wurde vorgenommen.

Die Aufgabenstellung der Primäraufgabe (siehe Abbildung 11.6) fordert die Studierenden auf, die Aussage der Schlagzeile mithilfe des Prüfschemas zu bewerten. Sie starten im Schema oben links entweder mit der Schlagzeile als *These* oder stellen ei-

Veganer haben mehr sportliche Ausdauer und genauso viel Kraft wie Fleischesser — zeigt eine Studie

Gabby Landsverk
28 Feb 2021

Das Vorurteil, Veganer seien schwächer als Fleischesser, ist nicht neu und bleibt weiterhin bestehen, obwohl Dokumentarfilme wie „The Game Changers", in denen vegane Spitzensportler gezeigt werden, diese Annahme auf den Kopf stellen.

Laut einer neuen Studie, die im *European Journal of Clinical Nutrition* veröffentlicht wurde, gibt es wenig Anhaltspunkte dafür, dass eine vegane Ernährung die Leistungsfähigkeit beeinträchtigt. Sie fand sogar vielmehr das Gegenteil heraus: Vegane Sportler verfügen teilweise über eine etwas bessere Ausdauer als ihre Allesesser-Mitstreiter.

Quelle: Gabby Landswerk (2021, Februar 28). Veganer haben mehr sportliche Ausdauer und genauso viel Kraft wie Fleischesser – zeigt eine Studie. In: Buisnessinsider. Abgerufen: 02.05.2021, von https://www.businessinsider.de/wissenschaft/ernaehrung/veganer-haben-mehr-sportliche-ausdauer-und-genauso-viel-kraft-wie-fleischesser-zeigt-eine-studie-r2/

Abbildung 11.5: *Schlagzeile und Einleitung des in der Erprobung verwendeten Artikels zum Thema Veganismus. Quelle des Artikels: Landswerk (2021).*

ne eigene These zur Studie auf. Die rechte Seite des Schemas sollen die Studierenden nutzen, um die *Studie* auf Wissenschaftlichkeit zu prüfen.
Bezüglich der Bewertung mittels Prüfschema ist anzumerken, dass von den Studierenden nicht erwartet wird, alle Prüfkriterien zielsicher und korrekt anwenden zu können - denn über das Unterrichtskonzept choice^{2}reflect oder die Bedeutung wissenschaftlicher Prüfkriterien wurden die Studierenden bislang nicht aufgeklärt. Der Einsatz der Methode *Lautes Denken* bei der Bearbeitung der Primäraufgabe soll Einblicke in die Herangehensweise der Studierenden an das Prüfschema ermöglichen. Es wird nicht betrachtet, ob die Prüfkriterien korrekt angewendet werden.

11.4.2 Entwicklung des Kategoriensystems

Die Daten aus den Fragebögen sowie die Aufzeichnungen der Gruppenmeetings werden, dem methodischen Rahmen dieser Arbeit folgend, einer umfangreichen inhalt-

Aufgabe:	Lösen Sie die Aufgabe gemeinsam. Bewerten Sie gemeinsam die Aussage der Schlagzeile und sprechen Sie dabei jeden Gedanken, der Ihnen in den Sinn kommt, laut aus. Nutzen Sie dazu das unten abgebildete Prüfschema und die Studie, auf welche sich die Journalistin bezieht.

Abbildung 11.6: *Aufgabenstellung der Primäraufgabe in der Erprobung des Prüfschemas.*

lich strukturierenden qualitativen Inhaltsanalyse nach Kuckartz (2018) unterzogen. Die Fragebogen-Antworten liegen bereits digitalisiert in Textform vor. Die Protokolle des *Lauten Denkens* wurden mit der Software *f4transkript* wörtlich transkribiert; Sprache und Interpunktion wurde geglättet. Auch nonverbale Interaktionen wurden bei der Transkription berücksichtigt, sofern sie besonders auffällig hervortraten.
Die transkribierten Daten (in Verknüpfung mit den Videodateien) wurden in die Software MAXQDA eingepflegt.
Die Antworten auf den Fragebögen wurden mit einer Mischung aus induktiver und deduktiver Kategorienbildung ausgewertet. Jedes Item des Fragebogens stellt eine Kategorie (deduktiv). Die Sub-Kategorien wurden teilweise deduktiv vorentwickelt, indem a priori mögliche Antworten antizipiert wurden (Flach 2021, S. 37). Die Ausdifferenzierung des Kategoriensystems erfolgte dann induktiv direkt am Material. Ein Ausschnitt aus dem erstellten System zeigt Abbildung 11.7, vollständig befindet es sich im Anhang B.1.3.

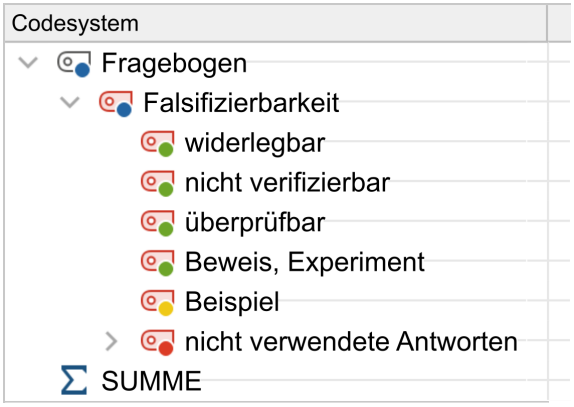


Abbildung 11.7: *Ausschnitt aus dem in MAXQDA deduktiv-induktiv erstellen Kategoriensystem für das Prüfkriterium Falsifizierbarkeit.*

Für das Item *Falsifizierbarkeit* werden die Begriffe *widerlegbar*, *nicht verifizierbar*, *überprüfbar* und *Beweis/Experiment* als zielführend codiert. Auch Beispiele, die die Studierenden angeben, werden erfasst.
In ähnlicher Weise wurden auch die Kategorien für die anderen Prüfkriterien aufgestellt; jeweils die zentralen Begriffe wurden als zielführende Codes definiert und am Material kodiert.

Der Kodierung der Transkripte liegt ein weiteres Kategoriensystem zugrunde. Bei der Entwicklung wurde ebenfalls eine induktiv-deduktive Mischform eingesetzt. Dazu wurden zu den Sub-Fragestellungen (siehe oben) deduktiv Oberkategorien erstellt: *Leserichtung des Prüfschemas* passend zur Subfrage 2 und *rationale Entscheidung* gemäß Subfrage 3. Auch die Kategorie *Kriterien* wurde im Voraus angelegt, dort werden alle Aussagen der Lernenden kodiert, die sich auf eines der Prüfkriterien beziehen. Abgestimmt auf die Subfrage 4 ist auch die Kategorie *Reaktionen der Studierenden* aufgeführt.
Nach einer Sichtung des Materials wurden induktiv Unterkategorien hinzugefügt. Das vollständige Kategoriensystem zur Analyse der Gruppenmeetings befindet sich im Anhang B.1.4.

11.4.3 Ergebnisse

Beschreibung der Seminarsitzung

Zu Beginn der Sitzung wurde den Lernenden vorgestellt, welche Inhalte und Methoden im Seminar behandelt werden. Dann wurde im *Learnweb*-Kurs der Fragebogen freigeschaltet und von den Studierenden ausgefüllt (ca. 12 Minuten).
Anschließend führte Theresa Flach die Methode *Lautes Denken* ein und erklärte die folgenden Übungs- und Primäraufgaben. Das Verteilen der Studierenden auf Gruppen-Meetings war mit einigen technischen Problemen behaftet. Somit haben nicht alle Gruppen die Übungsaufgabe *vor* der Primäraufgabe bearbeitet. Nur die Gruppen 3, 4, 7, 9 und 12 bewältigten die Übungsaufgabe und konnten die Methode kennenlernen. Die anderen Gruppen bearbeiteten direkt die Primäraufgabe, ohne das *Laute Denken* vorher zu üben.
In den Bild- und Tonaufnahmen ist erkennbar, dass die Studierenden zu Beginn der Gruppenarbeit sichtlich verwirrt waren, dies wird vor allem auf die technischen Pro-

bleme zurückgeführt. Nach den anfänglichen Schwierigkeiten konnten jedoch alle Gruppen die Primäraufgabe lösen. Dafür benötigten sie zwischen 12 und 25 Minuten.

Schlagzeile und Kontroverse zu *Veganismus*

Die Studierenden starteten mit dem Lesen der Schlagzeile, erst danach lasen sie sich die Aufgabenstellung durch. Manche Studierende äußerten sich sehr euphorisch über den Kontext, auch bei anderen Gruppen werden direkte Diskussionen angeregt (Flach 2021, S. 49):

- S12_1: „Okay, das erste was mir dazu in den Sinn kommt, glaube ich nicht unbedingt."
- S12_1: „Also, ich habe auch schon Veganer erlebt, die müssen ja zusätzlich sehr viele Vitamine nehmen oder Ausgleichsstoffe, wie auch immer, damit sie bei Kräften bleiben. Wer das nicht macht, kann auch nicht so wirklich viel Ausdauer haben wie ein Fleischesser."
- S12_2: „Ja, das stimmt. Aber ist ja auch/ Meine Schwester und ihr Freund, die probieren das gerade so ein bisschen aus alles. Die haben sich da wohl auch relativ viel eingelesen irgendwie und ich meine, wenn du halt alles passend quasi ersetzt, dann wüsste ich nicht, was dagegenspricht. Also wenn du halt alle Stoffe trotzdem zu dir nimmst, die du halt brauchst. Keine Ahnung."
- S12_1: „Das sage ich ja auch nicht. Also wenn du das raffst, dann kannst (unv.). Aber wer es halt nicht macht und halt auch nicht so darauf achtet/ Kommt dann ja auch drauf an, was für ein Fleischesser du siehst. Wenn du da so ein Fitnessjunkie siehst, der solche Muskeln hat, der/ da ist ein Veganer vielleicht nicht unbedingt genauso kräftig."
- S12_2: „Ja klar, kommt immer drauf an. Da müsste man natürlich halt Gleiches mit Gleichem vergleichen, ne."

In Gruppe 8 kommt die Meinung auf, dass die These der Schlagzeile in jedem Fall infrage gestellt werden kann. Gruppe 3 sieht es ähnlich und erkennt, dass die Schlagzeile eventuell zu verkürzt ist, und dass die Formulierung „zeigt eine Studie" künstliche Glaubwürdigkeit aufbauen soll:

- S3_1: „Ich finde, dass „zeigt eine Studie“, das ist immer sofort so/ keine Ahnung, dass das mehr Glaubwürdigkeit schenken soll. Also, weiß ich nicht. Aber ich glaube, dass viele Leute halt auch irgendwie das dahin schreiben, damit die Aussage mehr Gewichtung hat, als wenn das da nicht stehen würde.“

In Gruppe 2 und 9 wird genannt, dass die Studierenden einer Studie, wenn sie zitiert wird, grundsätzlich vertrauen würden.
Flach (2021) resümiert, dass die meisten Gruppen der Studie an sich nicht kritisch gegenüber stehen, aber sich skeptisch gegenüber der Schlagzeile „Veganer haben mehr sportliche Ausdauer und genauso viel Kraft wie Fleischesser“ zeigen. Sie stellt auch heraus, dass viele Studierende private Berührungspunkte mit der Kontroverse haben und diese sich dadurch gut für den Einsatz in der Seminarsitzung eignet.
Genauere Analysen des ersten Teils der Primäraufgabe werden an dieser Stelle nicht beschrieben, da hier nur zum Prüfschema hingeführt werden sollte. Im Folgenden werden die Daten hinsichtlich der zuvor aufgestellten Subfragen analysiert, um die Forschungsfrage beantworten zu können.

Subfrage 1: Vorwissen zu den Prüfkriterien

Die Kodierung der Fragebogen-Antworten erfolgte anhand des zuvor entwickelten Kategoriensystems. Jede der gegebenen Antworten wurde dabei einer Kategorie zugeordnet. Die Ergebnisse sind im Code-Matrix-Browser aus MAXQDA gut darstellbar (siehe Abbildung 11.8), diese Visualisierung erlaubt eine Detail-Auswertung der Antworten.
Eine weitere Visualisierung der Daten befindet sich in der Abbildung 11.9. Dabei handelt es sich um ein Diagramm, das die Verteilung der Antworten zeigt: Der *Anteil* zielführender Antworten wird dem Anteil nicht-zielführender Antworten gegenübergestellt. Dadurch ist direkt ersichtlich, wie groß das Vorwissen der Studierenden zu den einzelnen Prüfkriterien ist.
Die von den Studierenden gegebenen Antworten werden im folgenden Abschnitt beschrieben. Die genannten Zahlen beziehen sich auf den ersten Durchlauf mit einer Lerngruppe. In Klammern werden die Ergebnisse für die zweite Lerngruppe angegeben.

Zur Erklärung von *Falsifizierbarkeit* verwendeten 16 (10) Studierende Begriffe aus der Wortfamilie *widerlegen* und beschrieben sinngemäß, dass eine Hypothese wider-

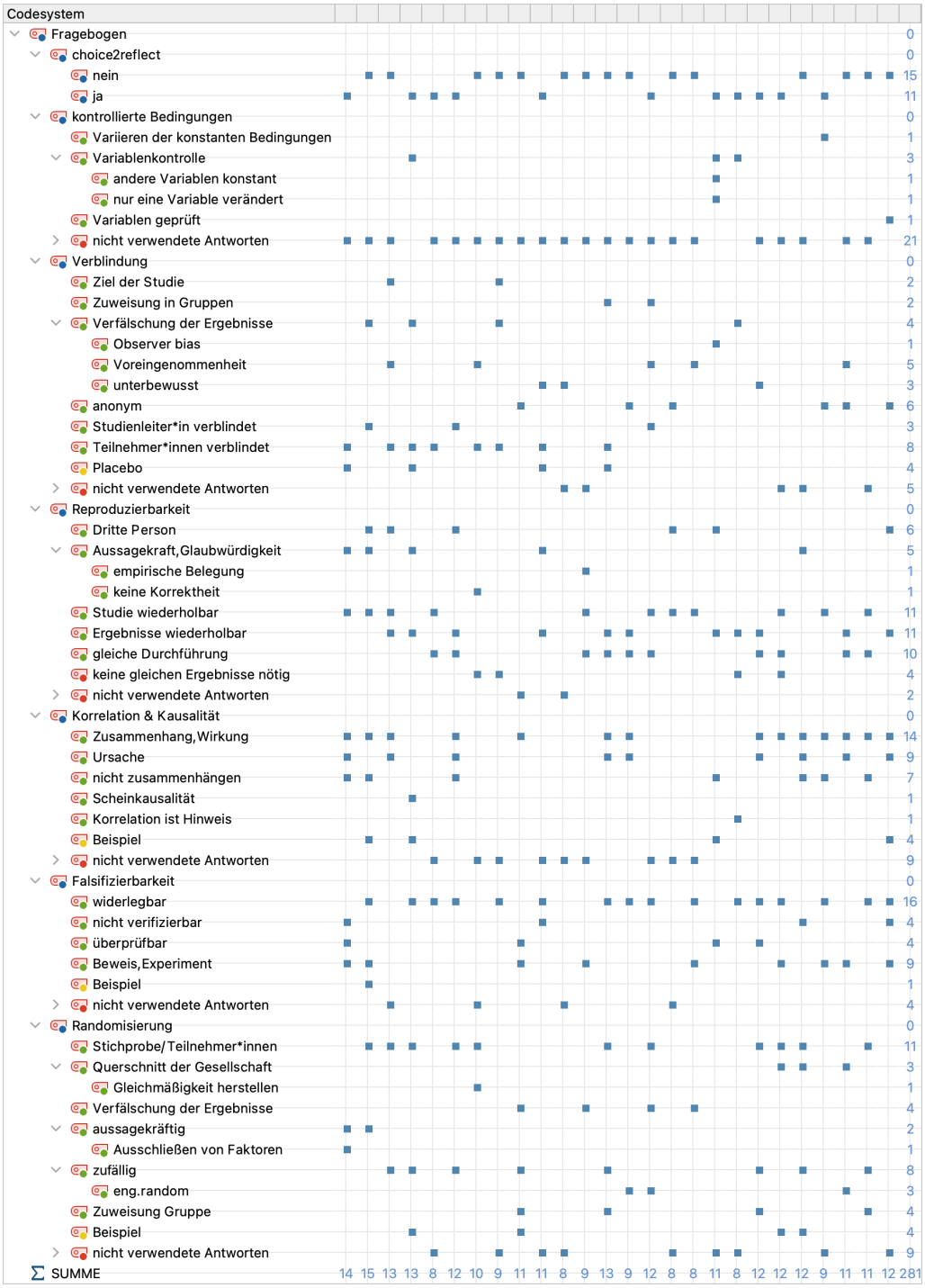


Abbildung 11.8: *Code-Matrix-Browser der Kodierung von Fragebogen-Antworten zur Beantwortung von Subfrage 1. Jede Spalte stellt die Antworten einer Person dar.*

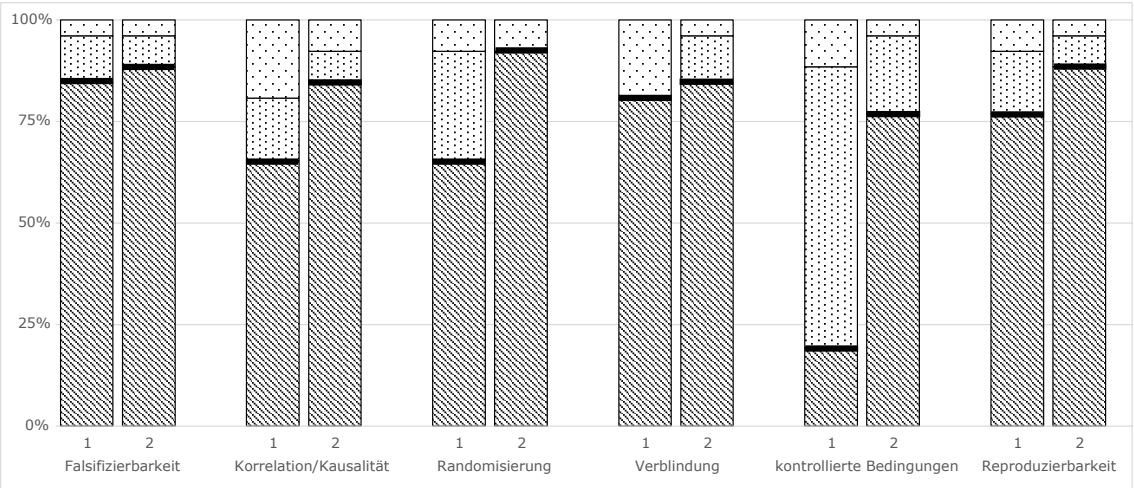


Abbildung 11.9: *Darstellung der Anteile zielführender und nicht-zielführender Antworten auf die Fragebogen-Items. 1 = Lerngruppe 1 (erster Durchlauf). 2 = Lerngruppe 2 (zweiter Durchlauf). Diagonal schraffiert = zielführende Antworten. Dicht gepunktet = nicht-zielführende Antworten. Leicht gepunktet = keine Antwort. Fettgedruckte Linie markiert Grenze zwischen zielführenden und nicht-zielführenden Antworten.*

legbar sein muss. Ein Student (5) untermauerte seine Antwort durch ein Beispiel, er stellte die Aussagen „Es gibt einen Gott" und „Eine Kerze benötigt Sauerstoff zum Brennen" gegenüber und demonstrierte daran den Unterschied zwischen Verifikation und Falsifikation. Insgesamt bemerkten neun (12) Studierende, dass es eines wissenschaftlichen Beweises oder eines Experiments bedarf, um Hypothesen zu überprüfen bzw. zu falsifizieren. Vier Antworten (3) waren nicht zielführend. In der zweiten Lerngruppe ging ein Studierender oder eine Studierende auf die Begriffe Theorie und nicht-falsifizierbare Theorie ein und nannte den Namen Popper im Zusammenhang mit Falsifizierbarkeit.

Zur zweiten Aussage „Eine Korrelation ist keine Kausalität" nannten 14 (16) Lernende den Begriff *Zusammenhang*, davon bezogen neun (11) auch die *Ursache* mit ein. Sieben (10) Studierende berichteten davon, dass Faktoren nicht zwingend *miteinander zusammenhängen*. Den Fachbegriff *Scheinkausalität* nannte nur eine Person (diese kannte choice^{2}reflect bereits). Neun (6) Antworten waren nicht zielführend.

In den Erklärungen der Randomisierung nannten elf (13) Studierende den Begriff Stichprobe. Der Zufall als zentrales Element wurde in elf (18) Fällen benannt. Vier (10) gaben an, dass Randomisierung wichtig ist, damit *Ergebnisse nicht verfälscht*

werden. Ebenfalls vier Mal (8) wurde erkannt, dass Randomisierung mit der *Zuweisung von Gruppen* zusammenhängt. Ein eigenes Beispiel führten vier (1) Studierende an. Neun Antworten (2) wurden keiner zielführenden Kategorie zugeordnet.

Bei der Verblindung gab es fünf (4) nicht-zielführende Antworten, die anderen Lernenden konnten das Prinzip zumindest in Grundzügen erklären. Dabei gaben acht (8) Studierende an, dass im Bezug auf *Teilnehmende* verblindet werden müsse, davon führten vier (0) Personen den Begriff *Placebo* an. Drei (8, aber nicht die gleichen wie bei *Teilnehmenden*) nannten auch die *Studienleitung*, die nicht über die Zuteilung der Variablen informiert werden dürfe. Sechs Mal (3) wurde angegeben, dass Verblindung die *Anonymisierung* persönlicher Daten der Studienteilnehmer:innen beinhalte. Neun Antworten (18) enthielten drei Konfundierungs-Faktoren, die zur Verfälschung von Ergebnissen im Zusammenhang mit fehlender Verblindung führen können (z. B. *Observer Bias* und *Voreingenommenheit*).

Kontrollierte Bedingungen zu erklären stellte für die Studierenden das schwierigste Item dar. In 21 Fällen waren die Antworten nicht zielführend. Zwar wurde häufiger genannt, dass die Forscher „Alles unter Kontrolle haben müssen“ (oder ähnliche Formulierungen), den Kern der Variablenkontrolle trafen sie damit jedoch nicht. Nur vier Studierende gingen auf das Konstanthalten von Störvariablen ein.
Die folgende Auswahl nicht-zielführender Antworten gibt einen Überblick über die unterschiedlichen Denkmuster der Studierenden:

- Es gibt kontrollierbare Bedingungen, wie z.B. den Ort, die Personengruppe. Bei der Personengruppe kann man die Personen auch nach verschiedenen Merkmalen, wie z.B. Geschlecht, Alter zuordnen. (S1_1)
- Die Aussage bedeutet, dass die Bedingungen durch die Forscher:innen, unter denen geforscht werden soll, vorher festgelegt werden müssen, und diese sich auch darüber im klaren sein sollten, unter welchen Bedingungen sie forschen wollen. (S3_1)
- Es sollten alle Parameter kontrolliert einstellbar sein, da sonst keine Kausalität ermittelt werden kann. (S4_2)
- Die Studie sollte in einem Rahmen durchgeführt werden, in dem alles erfasst und kontrolliert werden kann, sodass diese nicht unkontrolliert abläuft. (S6_1)

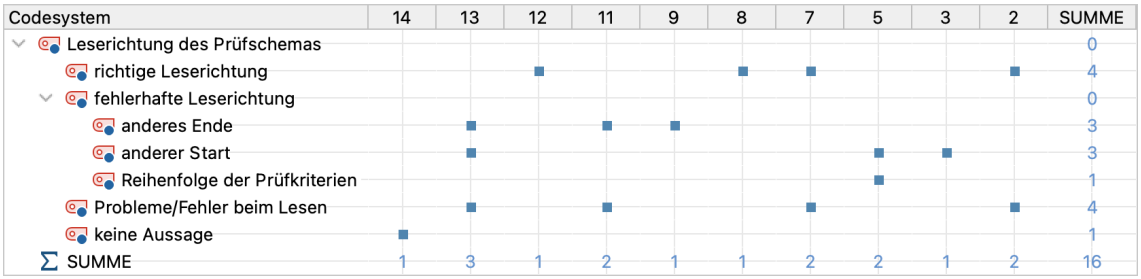
- Es ist wichtig, dass für verschiedene Studienteilnehmer die gleichen Grundvoraussetzungen herrschen, bzw. wenn man bei verschiedenen Voraussetzungen etwas prüfen möchte, müssen diese klar definiert, überprüft und festgehalten werden. (S9_1)
- Die Bedingungen unter denen eine Studie durchgeführt wird müssen permanent erfasst und kontrolliert werden damit sie reproduzierbar ist. (S12_2)

In der zweiten Lerngruppe waren nur sechs Antworten nicht zielführend. Variablenkontrolle wurde von den meisten Studierenden genannt, neun erkannten, dass möglichst alle Variablen konstant gehalten werden müssen, während nur eine variiert wird. Sieben Mal wurde genannt, dass ein *valides Ergebnis* erzielt werden muss.

„Eine Studie muss reproduzierbar sein“ setzten elf (10) Studierende mit der Wiederholbarkeit einer *Studie* gleich, weitere elf (12) bezogen die Wiederholbarkeit auf die *Ergebnisse*. Zehn Mal (7) wurde genannt, dass die Durchführung bei einer Reproduktion gleich sein muss. Sechs (5) nannten eine *Dritte Person*, die bei Reproduzierbarkeit eine Rolle spielt. Sieben (8) Studierende schrieben, dass nur reproduzierbare Studien auch *glaubwürdig* und *aussagekräftig* sind. Allerdings antworteten auch vier Lernende (0), dass *keine gleichen Ergebnisse* nötig seien. Zwei Mal (3) wurde nicht geantwortet.

Subfrage 2: Vorgehen entsprechend der Leserichtung

Ein Teil des Kategoriensystems eignet sich zur Untersuchung der zweiten Subfrage. Es wird kodiert, inwiefern die Studierenden mit der Bearbeitung des Prüfschemas oben links beginnen und sich nach unten rechts durcharbeiten - gemäß der intendierten und in Europa gängigen Leserichtung.
Die Ergebnisse, visualisiert durch den Code-Matrix-Browser aus MAXQDA, befinden sich in der Abbildung 11.10.
In den Videos von zehn Zweiergruppen konnte die Leserichtung bestimmt werden. Vier Gruppen befolgen die korrekte Leserichtung. Sieben Gruppen gehen jedoch nicht wie intendiert vor: Die Gruppen 3 und 5 starten an einer anderen Stelle, Gruppe 9 und 11 enden an einer anderen Stelle und Gruppe 13 startet *und* endet nicht korrekt. Die Gruppe 5 wählt auch eine andere Reihenfolge der Prüfkriterien. Probleme und Fehler



Codesystem	14	13	12	11	9	8	7	5	3	2	SUMME
Leserichtung des Prüfschemas											0
richtige Leserichtung			■			■	■			■	4
fehlerhafte Leserichtung											0
anderes Ende		■		■	■						3
anderer Start		■						■	■		3
Reihenfolge der Prüfkriterien								■			1
Probleme/Fehler beim Lesen		■		■			■			■	4
keine Aussage	■										1
Σ SUMME	1	3	1	2	1	1	2	2	1	2	16

Abbildung 11.10: *Code-Matrix-Browser der Kodierung von Gruppenvideos zur Beantwortung von Subfrage 2. Jede Spalte stellt eine Gruppe von Studierenden (zwei Studierende pro Gruppe) dar.*

beim Lesen der Aufgabenstellung sowie des Prüfschemas gab es bei den Gruppen 2, 7, 11 und 13.

Der Vergleich mit der zweiten Lerngruppe zeigt, dass diese deutlich häufiger die korrekte Leserichtung befolgte (elf Gruppen). Nur die Gruppe 14 beging Fehler beim Befolgen des „Rohrsystems“ im Prüfschema: Sie verfolgten den Weg von dem Prüfkriterium *Korrelation und Kausalität* durch das Fallrohr bis zur Reproduzierbarkeit. Die Gruppe 4 verwendete eine andere Reihenfolge der Prüfkriterien und Gruppe 12 beschäftigte sich nur mit der *rechten Seite* des Schemas.

Subfrage 3: Treffen rationaler Entscheidungen

Von den insgesamt zehn Gruppen treffen am Ende der Sitzung vier eine rationale Entscheidung bei der Bewertung der Studie. Das erkennt Flach (2021) an den Gruppendiskussionen. In der Gruppe 2 stellen die Studierenden fest, dass alle Kriterien bis auf die Verblindung erfüllt wurden. Sie gelangen zu dem Schluss, dass die Studie wissenschaftlich ist, aber dass der Aspekt der Verblindung noch genauer geprüft werden müsse.

Die Gruppen 3, 5 und 14 kommen zu dem Ergebnis, dass die Studie nicht wissenschaftlich ist, da einige Kriterien nicht erfüllt sind. Sie unterscheiden dabei auch zwischen der Schlagzeile und den tatsächlich erhaltenen Ergebnissen. Die Schlagzeile sei zu verallgemeinert und „unfassbar stark eingegrenzt“ (S14-1, Pos. 36). In den Gruppen 5 und 14 wird erwähnt, dass die Studie nicht „absolut nichts-aussagend ist“ (S14-1, Pos. 34), dass jedoch einzelne Faktoren noch verändert werden müssen.

Gruppe 8 entscheidet letztendlich nicht rational. Die Studierenden decken zwar auf,

dass die Studie „in den Müll gehört“ (S8-2, Pos. 65), stufen sie dennoch zum Ende der Gruppenphase als wissenschaftliche Studie ein (S8-1, Pos. 94).
Zu einem großen Teil der Gruppen konnte Flach keine Aussage treffen, inwiefern die Studierenden am Ende rational entscheiden. In den Gruppenvideos wird nicht ersichtlich, inwiefern eine Entscheidung getroffen wurde. Die Studierenden explizieren das nicht.

Anders sieht die Verteilung in der zweiten Lerngruppe aus: Zehn von dreizehn Gruppen entscheiden am Ende rational und begründen dies auch anhand des Prüfschemas. Gruppe 1 prüft nacheinander die Wissenschaftskriterien und entscheidet unter Berücksichtigung dieser (siehe Transkript zu Gruppe 1, Pos. 50-75). Ein Student aus Gruppe 3 sagt: „Also ich weiß nicht, ob die jetzt direkt/ ob man die direkt in den Müll schmeißen muss. Aber spätestens, wenn du dann genauer in diese ganzen Studienfaktoren reingehst, musst du früher oder später in den Müll.“ (S3-2, Pos. 98).
Auch in den anderen rational entscheidenden Gruppen sieht die Entscheidungsfindung ähnlich aus:
Gruppe 5, Pos. 160-162

- S5_1: Ist das dann da mit wissenschaftlich oder nicht? (zeigt auf Schatzkiste)
- S5_2: Ich glaube nicht, ehrlich gesagt.
- S5_1: Nicht so ganz. Wenn man vielleicht alle Augen zudrückt, dann vielleicht. Aber nicht so 100 Prozent.

Gruppe 7, Pos. 99

- S7_1: Wir hätten das bei ganz vielen Punkten schon in die Tonne kloppen können, wie es hier so schön steht.

Gruppe 8, 135-139

- S8_2: Wahrscheinlich nicht so ganz abdeckt.
- S8_1: Nicht so ganz richtig (zeigt auf Mülltonne). Deshalb weiß ich nicht, wie gut die/ Also, wenn es nicht kontrolliert ist, dann würde ich/ kann man dem auch nicht so ganz glauben, glaube ich.
- S8_2: Ja.

- S8_1: Da wäre ich glaube ich zu kritisch. Deshalb weiß ich nicht, ob die Studie so mega aussagekräftig ist.
- S8_2: Ja, ich weiß auch nicht.

Gruppe 9, Pos. 99-101

- S9_2: Ich kann mir auch vorstellen, dass diese Studie nicht beweist, dass vegane Menschen sportlicher sind als nicht-vegane Menschen. Sondern vielmehr widerlegt, dass es andersherum ist. Es wird ja oft gesagt: Veganer haben Eisenmangel und so und alles. Und sind viel ungesünder unterwegs. Und dass diese Studie vielmehr zeigen kann, dass diese Aussage falsch ist, aber nicht Behauptung aufstellen darf: Vegane Menschen sind viel sportlicher.
- S9_1: Ja.
- S9_2: Also die Studie an sich ist, glaube ich, wirklich wissenschaftlich und gut, aber die Schlussfolgerung der Reporterin ist komplett verdreht.

Gruppe 10, Pos. 118-119

- S10_1: Okay. Also ich würde sagen so als Fazit, dass bei allen dieser 5 Punkte man eh bei „nein“ abbiegen müsste (zeichnet die Wege von „nein“ nach) und dann ist das Ganze (zeigt auf die Tonne)
- S10_2: Für die Tonne (lachen).

Gruppe 12, Pos. 84-85

- S12_2: Ich überlege gerade schon die ganze Zeit. Ahja, guck mal bei dem Bild da am Ende - bei dem Prüfschema. Scrolle mal runter, genau. (Scrollt zu Prüfschema). Da ist ja am Ende noch diese Kiste mit wissenschaftlich. Ich glaube wir würden/ wir sind uns da einig, dass es für wissenschaftlich da nicht ganz reicht, ne?
- S12_1: Auf jeden Fall.

Gruppe 14, Pos. 122

- S14_2: Okay, Aussage bewerten und dafür die Studie und das Prüfungs/. Okay wie bewerten wir denn einfach nur die Schlagzeile? Dieses/ Wir bewerten, das ist falsch.

Gruppe 2 entscheidet ohne Zuhilfenahme des Prüfschemas, dass in der Studie die wirklichen Gegebenheiten nicht deutlich dargestellt wurden („...und ich finde, das wird halt in diesem Artikel irgendwie so gar nicht rübergebracht“ (S2-2, Pos. 49)) und kritisieren damit die Studie. Gruppe 4 verwendet für die Entscheidung ebenfalls nicht das Prüfschema, sie stuft die Schlagzeile als zu reißerisch ein: „[...] dass die Überschrift einfach viel zu reißerisch ist, weil ich meine im Text dahinter gehen die ja erstmal auf diese Dokumentation ein, die ja einfach keine wissenschaftliche Aussagekraft erstmal hat“ (S4-1, Pos. 55).
Lediglich über die Entscheidung von Gruppe 13 kann keine Aussage getroffen werden.

Subfrage 4: Reaktionen auf das Prüfschema

Mit der vierten Subfrage wird untersucht, welche Reaktionen die Studierenden bei dem ersten Betrachten des Prüfschemas zeigen. Dazu werden verbale Äußerungen und non-verbale Ausdrücke in Gestik und Mimik der Studierenden anhand der aufgezeichneten Gruppenvideos erfasst.
Das Kategoriensystem ist zur Untersuchung dieser Subfrage nicht weiter ausdifferenziert. Kodiert wird im niedrig-inferenten *Event-Sampling*, bei dem das Setzen eines Codes durch das Auftreten von Events ausgelöst wird (siehe auch Kapitel 8.2), in diesem Fall durch das erstmalige Betrachten des Prüfschemas. Dazu wird der Abschnitt jedes Videos betrachtet, in dem die Studierenden zum ersten Mal zu dem Prüfschema scrollen (erkennbar an der aufgezeichneten Bildschirmfreigabe).
Vier Gruppen scrollen nach dem Erfassen der Aufgabenstellung zum Prüfschema. Gruppe 3 bei ca. 7 Minuten, Gruppe 5 bei ca. 4 Minuten, die Gruppen 8 und 12 bereits nach 2 Minuten. Allerdings schauen sie sich das Schema nicht lange an und beginnen dann mit dem Lesen des Studien-Textes. Lediglich Gruppe 8 verharrt etwas länger auf der Grafik des Prüfschemas.
Die unmittelbar eintretenden Reaktionen sind in transkribierter Form in der Tabelle 11.2 aufgelistet.
Auch im zweiten Durchlauf der Erprobung wurden die Reaktionen der Lernenden erfasst. Allerdings konnten nur bei fünf Gruppen aussagekräftige Reaktionen festgestellt werden. Bei Gruppe 1 trat sofortige Stille ein, die für etwa 14 Sekunden andauerte. Eine Studentin begann daraufhin sofort, das Schema in der korrekten Leserich-

tung laut durchzulesen.
In Gruppe 5 wird S5_1 von S5_2 darauf hingewiesen, dass anhand des Schemas bewertet werden soll. S5_1 entgegnet: „Äh, da war ein Schema? Ach das! Oh Gott!“ und reißt die Augen auf (Gruppe 5, Pos. 31).
Gruppe 8 wirkt bei der Betrachtung etwas verlegen und beide Studierende lachen (Gruppe 8, Pos. 22).

Tabelle 11.2: *Auflistung über die Reaktionen der Studierenden (erste Lerngruppe) auf das erstmalige Betrachten des Prüfschemas.*

Gruppe, Position	**Reaktion**
3, Pos. 78	S3_1: „Oha!“
3, Pos. 180	S3_1: „Ähm ja. (Beide schauen fragend beim Betrachten des Prüfschemas.)“
5, Pos. 24-25	S5_1: „Aber ich glaube das ist halt ein allgemeines Schema, ne.“ S5_2: „Ja. Ja, so arbeitet man halt generell naturwissenschaftlich. Was war jetzt nochmal genau die Aufgabe? Irgendwie sollten wir die Studie prüfen oder so?“
8, Pos. 17-18	S8_2: „Boah“ (zu Prüfschema) S8_1: „Das ist auf jeden Fall sehr erschlagend auf den ersten Blick.“ (sieht Prüfschema)
8, Pos. 59	S8_2: „Okay, also sind wir jetzt bei diesem Schema hier“ (lacht). S8_1: „Wir sollen es ja mitbenutzen“ (lachen, wirken verlegen).
11, Pos. 46	Studierende schauen erst ein wenig verwirrt und sind still.
12, Pos. 23	S12_2: „Um Gottes Willen!“ (Reaktion von Studierenden auf Prüfschema - schlägt sich die Hand vor den Kopf).
13, Pos. 32-33	S13_2: Ich muss jetzt erstmal schauen, dass ich das Schema verstehe. Es ist ja schon ein bisschen wirr. (schauen alle etwas verwirrt)

Eine Studierender aus Gruppe 12 betrachtet das Prüfschema intensiv und setzt einen sehr konzentriert wirkenden Blick auf. Verbale Äußerungen tätigt er jedoch nicht (Gruppe 12, 04:10 min).
Emotionaler fällt die Reaktion bei S14_1 aus, denn unmittelbar bei dem Scrollen auf das Schema sagt sie: „Um Gottes Willen!“ (Gruppe 14, Pos. 19).
Weitere auffällige Reaktionen wurden in dieser Lerngruppe nicht gefunden.

11.5 Analyse und Diskussion

11.5.1 Vorwissen zu den Prüfkriterien

Die Kodierung der Fragebogen-Antworten hat gezeigt, dass ein großer Teil der Studierenden bereits Vorwissen zu den Prüfkriterien besitzt. In der ersten Lerngruppe ist der Anteil zielführender Antworten besonders hoch bei den Prüfkriterien *Falsifizierbarkeit*, *Verblindung* und *Reproduzierbarkeit* (erkennbar in Abbildung 11.9).
Bei diesen Items sind mehr als 75 % der Antworten als zielführend eingestuft worden.
Die Prüfkriterien *Korrelation und Kausalität* und *Randomisierung* wurden weniger häufig, aber dennoch von mehr als der Hälfte korrekt erklärt. Lediglich bezüglich der *kontrollierten Bedingungen* liegt der Anteil nicht-zielführender Antworten bei über 75 %.
Auffällig ist, dass die Studierenden dieses Item nicht einfach übersprangen. In den meisten Erklärungen sind Ansätze vorhanden, die eher eines der anderen Prüfkriterien erklären. In den oben aufgelisteten Antworten gehen die Studierenden z. B. auf Reproduzierbarkeit ein: „Die Bedingungen [...] müssen permanent erfasst und kontrolliert werden damit [die Studie] reproduzierbar ist“. Auch Aspekte der Randomisierung oder der Problematik von Scheinkausalitäten finden sich in den Texten der Studierenden.
Die Variablenkontrolle, die mit dem Prüfkriterium eigentlich angesprochen werden soll, wird nur selten genannt. Diese Antwortverteilung deutet darauf hin, dass entweder der Name des Prüfkriteriums irreführend ist (bzw. nicht an intuitive Denkmuster der Studierenden anknüpft) oder dass das Fragebogen-Item nicht treffsicher formuliert wurde. Flach (2021) schlägt als alternative Benennung des Wissenschaftskriteriums die Bezeichnung *Variablenkontrolle* vor.
Diese neue Bezeichnung wurde bereits in der nächsten Erprobung mit der zweiten Lerngruppe eingesetzt. Im Fragebogen-Item wurde die Aussage verwendet: „Eine Studie muss unter kontrollierten Variablen durchgeführt werden“. Hier wurde nur das Wort *Bedingungen* durch *Variablen* ersetzt, um (zulasten der grammatikalischen Korrektheit) größtmögliche Vergleichbarkeit mit der ersten Lerngruppe zu gewährleisten.
Bei der Darstellung des Prüfschemas im Rahmen der Primäraufgabe wurde auch die

Stichwort-Frage angepasst: „Nur eine Variable verändert?“

Die Betrachtung der Fragebogen-Ergebnisse für die zweite Lerngruppe zeigt deutlich, dass bei dem Item *kontrollierte Bedingungen* (bzw. neu: Variablenkontrolle) eine höhere Leistung erzielt wurde. Der Anteil nicht-zielführender Antworten liegt mit unter 25 % deutlich niedriger als bei der ersten Lerngruppe.

Selbstverständlich ist hier nicht von einer *Verbesserung* zu sprechen, da es sich um unterschiedliche Lerngruppen handelt. Auch ist nicht erklärbar, inwiefern die höhere Leistung auf die Umbenennung des Prüfkriteriums zurückzuführen ist. Auch andere Effekte, die durch das Forschungsdesign nicht erhoben wurden, können hier eine Rolle gespielt haben.

Auch bei den anderen Prüfkriterien schnitt die zweite Lerngruppe etwas besser ab, die Unterschiede sind jedoch weitaus kleiner als bei der *Variablenkontrolle*.

Flach (2021) erkennt, dass die Ergebnisse für Reproduzierbarkeit einer Einschränkung unterliegen. Als *zielführend* kodiert wurden alle Antworten, die in eine der im Codesystem aufgeführten grünen Kategorien passen. Dazu gehört, neben den fachlich korrekten Kategorien *Dritte Person*, *Aussagekraft*, *gleiche Durchführung* und *Ergebnisse wiederholbar* jedoch auch die Kategorie *Studie wiederholbar*.

Diese wurde kodiert, wenn Studierende beschrieben, dass eine *Studie* wiederholt durchführbar sein muss, um als reproduzierbar zu gelten. Diese Erklärung ist jedoch nur teilweise korrekt, denn Reproduzierbarkeit orientiert sich am Ausgang des Experiments oder der Studie: Die *Ergebnisse* müssen wiederholt erhalten werden.

Die Kategorie wurde elf Mal vergeben, davon enthielten sechs Erklärungen zusätzlich zielführende Inhalte. Die übrigen fünf Studierenden haben hier nicht ganz zielführend geantwortet, damit ergibt sich ein Anteil zielführender Antworten von 58 % anstelle der zuvor angegebenen 77 %. In der zweiten Lerngruppe ist keine Korrektur nötig, da alle mit *Studie wiederholbar* kodierten Antworten zusätzlich auch zielführende Codes erhielten.

Insgesamt kann festgehalten werden, dass die an der Erprobung teilnehmenden Studierenden ein moderates bis gutes Vorwissen zu den Prüfkriterien haben. Lediglich bei der *Variablenkontrolle* gibt es Schwierigkeiten mit dem Ausschärfen der Begriffe.

11.5.2 Umgang mit dem Prüfschema

Die Untersuchung der Leserichtung hat ergeben, dass in der ersten Lerngruppe nur vier von zehn Zweiergruppen die korrekte Leserichtung wählen. Manche Gruppen starten bzw. enden an einer anderen Stelle oder befolgen in der Diskussion nicht die vorgegebene Reihenfolge der Prüfkriterien. In der zweiten Lerngruppe ergibt sich ein anderes Bild, dort befolgten 11 von 13 Gruppen die Leserichtung von oben links (*These*) über die *Hypothese* nach unten rechts.

Das verdeutlicht die Relevanz einer Instruktion in die Vorgehensweise beim Verwenden des Prüfschemas. Diese Erprobung mit Lehramts-Studierenden erfolgte ohne jegliche Art einer Erklärung der Kontroverse oder des Schemas. Die Studierenden wurden mit dem Arbeitsmaterial konfrontiert und führten die Methode *Lautes Denken* durch. Gepaart mit der vorangegangenen Fragebogenerhebung ergab sich für viele Teilnehmende vermutlich ein *Overload* an neuen Informationen und Visualisierungen.

An den Reaktionen bei erstmaligem Betrachten des Prüfschemas ist erkennbar, dass viele Studierende zunächst erschrocken oder verwirrt bzw. unsicher sind: Manche rufen „Oha“ oder „Um Gottes Willen“, andere verbalisieren ihren Gedanken: „Das ist auf jeden Fall sehr erschlagend auf den ersten Blick“ (Beispiele siehe 11.2).

Allerdings konnten am Ende - trotz fehlender Instruktion und teils erschrockenen Reaktionen - 10 von 13 Gruppen eine klar rationale Entscheidung auf Basis des Prüfschemas treffen (zweite Lerngruppe). Die Gespräche der Lernenden untereinander zeigen, dass sie das Prüfschema in ihre Argumentationen einbeziehen. Sie nutzen die plakativen Bilder der Mülltonne und der Schatzkiste. Auch den sehr wichtigen Aspekt, dass bereits das Nicht-Erfüllen *eines* Prüfkriteriums die Unwissenschaftlichkeit der geprüften Studie bedeutet, erkennen und befolgen die Studierenden.

Weiterhin erkennen die Lernenden den Unterschied zwischen der Wissenschaftlichkeit der Studie und der Angemessenheit der Schlagzeile, zum Beispiel S9_2: „Also die Studie an sich ist, glaube ich, wirklich wissenschaftlich und gut, aber die Schlussfolgerung der Reporterin ist komplett verdreht“.

11.5.3 Beantworten der Forschungsfrage

Zusammenfassend wird festgehalten, dass viele Studierende mit einem moderaten bis guten Vorwissen zu den Prüfkriterien in die Erprobung gestartet sind - vor allem jene, die bereits Teile des Unterrichtskonzepts choice^{2}reflect kannten.
In der Erprobung haben sie sich mit einer kontroversen Schlagzeile zum Thema *Veganismus* beschäftigt und diese auf Wissenschaftlichkeit geprüft. Dabei ist aufgefallen, dass viele Studierende sich zunächst nicht mit dem Prüfschema beschäftigen, sondern in ihrer Gruppe über die Schlagzeile diskutieren. Nach Lesen des Studientextes werden die Diskussionen konkreter, die Studierenden stellen einzelne Aspekte der Methodik infrage.
Die Aufgabenstellung fordert dazu auf, auch das Prüfschema zur Bewertung zu nutzen. In dem Zuge beschäftigen sich die meisten Gruppen dann auch mit dem Schema. Die Herangehensweise ist im Allgemeinen positiv zu bewerten, denn auch ohne Instruktion erkennen die Studierenden viele relevante Punkte der Prüfkriterien und wenden diese sowohl auf die vorliegende Studie als auch die Schlagzeile an.
Schließlich schaffen es einige Gruppen, auf der Grundlage des Prüfschemas eine rationale Entscheidung zu treffen. Diese Erkenntnis ist für das Voranbringen des dritten Mesozyklus dieses Projekts von entscheidender Bedeutung.
Zur Beantwortung der zusammen mit Flach (2021) entwickelten Forschungsfrage wurde das Prüfschema in einem „Hands-On"-Ansatz an die Studierenden gegeben, ohne zuvor die gesamte Konzeption mit allen Phasen zu durchlaufen. Dass auch in diesem Setting rationale Entscheidungen die Folge waren ist ein Hinweis auf die Intuitivität und Verständlichkeit des Schemas.
Die *Lautes Denken*-Protokolle haben jedoch auch offenbart, dass nicht für alle Lernenden das Prüfschema direkt eingängig und intuitiv verständlich ist. Einige waren im ersten Moment überfordert oder verfolgten nicht den richtigen Weg durch das Schema. Im Re-Design werden einzelne Elemente des Prüfschemas auf Grundlage der empirisch gewonnenen Erkenntnisse optimiert.

11.6 Re-Design

Als methodisches Instrument zur Bewertung von Thesen und Studien wurde ein Prüfschema literaturbasiert entwickelt und in einem ersten *Hands On*-Ansatz mit Studie-

renden des Lehramts im Fach Chemie erprobt. Dabei wurde herausgestellt, dass das Prüfschema durchaus intuitiv verständlich sein *kann* und von den Studierenden auch zur Entscheidungsfindung verwendet wird.

Allerdings haben sich im praktischen Einsatz auch ein paar Schwächen in der Gestaltung des Schemas gezeigt.

Zum einen wurde herausgefunden, dass nicht alle Studierenden die korrekte bzw. intendierte Leserichtung einhalten. Teilweise starten sie an einer anderen Stelle, wie z. B. die Gruppe 13 der ersten Lerngruppe. Sie scrollen zur Abbildung des Schemas, dabei sagt S13_2: „Ich muss jetzt erstmal schauen, dass ich das Schema verstehe. Es ist ja schon ein bisschen wirr. [Aber] es sind ja quasi die Fragen, die wir gerade schonmal beantwortet haben. Ich weiß nicht, ob dir alle Begriffe da was gesagt haben." (Gruppe 13, Pos. 32). Damit beziehen sie sich auf die Prüfkriterien und starten auf der rechten Seite des Schemas mit dem Begriff *Studie*.

Der veränderte Startpunkt könnte daraus resultieren, dass die Studierenden sich direkt vor der Bearbeitung der Primäraufgabe mit den Prüfkriterien im Fragebogen auseinandergesetzt haben. Bei Betrachten des Schemas fallen ihnen diese Begriffe eher ins Auge und sie starten dort.

Gruppe 3 startet bei *Hypothese* mit der Begründung: „Wir haben eine Hypothese, ne? [Studierender meint damit die Schlagzeile]" (Gruppe 3, Pos. 185). Hier wird also Bezug genommen auf die zuvor gelesene Schlagzeile; diese wird mit einer Hypothese gleichgesetzt und als Startpunkt angenommen.

Teilweise „verirren" sich die Studierenden auch in dem Rohrsystem des Schemas, wie z. B. Gruppe 13 des ersten Durchlaufs: „So. Wir haben ja oft nein gesagt, ne. Dann müssen wir hier oben These umformulieren [zeigt auf linken oberen Teil des Prüfschemas]" (Gruppe 13, Pos. 108).

Oder auch in Gruppe 9: „[Maus auf Reproduzierbarkeit] Okay und dann, wenn es/ also wenn es *Ja* ist, dann ist es wissenschaftlich [zeigt auf die Schatzkiste] und wenn es *Nein* ist [folgt dem Weg zur Falsifizierbarkeit], dann muss man das halt nochmal irgendwie quasi alles über These umformulieren zurück zu These. Also das ist quasi wie so ein Kreislauf. So, weißt du? [Zeichnet einen Kreislauf, nimmt die Mülltonne nicht wahr]." (Gruppe 9, Pos. 138f). Hier wird das Schema eher als Kreislauf verstanden, da das Rohrsystem auf dem Weg zur Mülltonne auch an dieser vorbeiführt, wieder zurück zur Falsifizierbarkeit.

Diesen Tendenzen in Zukunft entgegenwirken könnte eine strukturierte *Einführung* in die Verwendung des Prüfschemas. Im Rahmen des Unterrichtskonzepts wird das Schema nicht von Studierenden sondern von Lernenden der Sekundarstufe I verwendet. Dabei sollte beim Übergang von der Phase III in die Phase IV ausführlich erklärt werden, wie das Prüfschema zu verstehen ist und an welcher Stelle mit welchem Inhalt gestartet wird. Die kognitive Belastung bei Konfrontation mit dem Schema wird sich damit vermutlich anders gestalten als in der Erprobung mit den (nicht instruierten) Studierenden - auch aus dem Grund, dass die Schüler:innen sich bereits in den Phasen I und II mit der Kontroverse beschäftigt haben und somit bereits „eingedacht" sind, wohingegen die Studierenden nur an einer Schlagzeile und einem kurzen Informationstext ein neues Thema erarbeitet haben.

Dennoch gibt es an diesem Punkt Anlässe für kleinere Re-Design-Schritte. Das Verirren im Rohrsystem sowie die Wahrnehmung des Schemas als Kreislauf könnten verhindert werden, indem das durchgängige Rohrsystem am Mülleimer unterbrochen wird: Je eine Zuleitung von links und von rechts der Mülltonne zuzuführen, sorgt für eine stärkere optische Trennung der beiden Seiten.
Eine weitere Optimierung wird von Flach (2021) wie oben beschrieben vorgeschlagen: Die Umbenennung des Prüfkriteriums *kontrollierte Bedingungen* in *Variablenkontrolle*. Diese Verbesserung wird in Teilen übernommen. Zwar bleibt die Bezeichnung des Kriteriums gleich, aber die zugeordnete Stichwort-Frage wird angepasst. In der aktuellen Version des Prüfschemas steht anstelle von: „Nur ein Faktor verändert?" nun: „Nur eine Variable verändert?"
Damit wird eindeutiger definiert, worum es bei dem Prüfkriterium geht. Der Begriff *Faktor* wurde zuvor einheitlich in jeder Stichwort-Frage verwendet, die Diskussionen der Studierenden haben jedoch gezeigt, dass das zu Missverständnissen führt und die Grenzen zwischen den Prüfkriterien verschwimmen lässt. Der Begriff *Faktor* eignet sich also nicht für jedes Kriterium. Im Zusammenhang mit der Verblindung wird nun von *Bedingungen* gesprochen, bei den kontrollierten Bedingungen von *Variablen* und bei der Reproduzierbarkeit von *Daten*. Es resultieren folgende Stichwort-Fragen:

- Verblindung: **Bedingungen** vor Testperson verborgen?
- kontrollierte Bedingungen: Nur eine **Variable** verändert?
- Reproduzierbarkeit: Alle **Daten** dokumentiert?

Weiterhin bietet es sich an, das Prüfschema als großes Poster im Klassenraum aufzuhängen und an verschiedenen Stellen der Unterrichtsreihe darauf zu verweisen. In den Phasen, in denen die Schüler:innen konkret mit dem Schema arbeiten, sollte es ihnen direkt zugänglich sein; dafür kann es in kleinerer Form für jede Schülergruppe vervielfältigt werden (z. B. als Arbeitsblatt oder digital auf einem Tablet). In jedem Fall ist eine hohe Verfügbarkeit zu gewährleisten, sodass von jeder und jedem Lernenden stets nachvollzogen werden kann, an welcher Stelle des Bewertungsprozesses sie oder er sich befindet.

Teil VI

Re-Framing und Reflexion

Im folgenden Kapitel findet ein Rückbezug auf die im Framing entwickelten Forschungsfragen statt. Die übergreifende Fragestellung lautete:

„Wie kann die Unterrichtskonzeption choice^{2}reflect als praxistaugliche Intervention zur Bewertung kontroverser Kontexte weiterentwickelt werden?“

Das zunächst ungerichtet wirkende Fragewort „Wie“ erhält bei der Beantwortung der Frage jedoch zwei klare Zielrichtungen bzw. Auslegungen. Es kann auf der einen Seite verstanden werden als Frage nach neuen „äußeren“ Gestaltungselementen, die der Konzeption hinzugefügt werden können oder auch als Frage nach der Veränderung und Ergänzung bereits bestehender Elemente - jeweils vor dem Hintergrund einer Erhöhung der Praxistauglichkeit.

Auf der anderen Seite fokussiert die Fragestellung aber auch „innenliegende“ Gestaltungsmerkmale. Nicht nur die Frage nach dem hinzuzufügenden Element steht im Vordergrund, sondern auch *in welcher Weise* dieses Element zu einer Verbesserung der Praxistauglichkeit beiträgt. Zusätzlich zum *praktischen Output* in Form von Elementen der Konzeption wird auch ein *theoretischer Output* in Form von Gestaltungskriterien der Konzeptions-Elemente generiert. Der Weg dieser Ergebnis-Genese wird in folgenden Abschnitten erläutert.

Zu Beginn des Projekts erfolgte eine Bestandsaufnahme zum Unterrichtskonzept. Im Vorgängerprojekt aufgetretene Desiderata wurden aufgegriffen und in das Design-Experiment aufgenommen. Zusätzlich wurden auch eigene Ideen zur Weiterentwicklung der Konzeption eingebracht und umgesetzt.

So kristallisierten sich drei Forschungsschwerpunkte heraus: Der Pool an Unterrichtsmaterial zu den wissenschaftlichen Prüfkriterien sollte erweitert werden, um zu allen Prüfkriterien praxistaugliches Material zur Verfügung stellen zu können. Die Weiterentwicklung der Prüfkriterien bildete den ersten Mesozyklus in diesem Projekt.

Den kontextuellen Rahmen für die Phasen I, II, IV und V bildete bisher die Kontroverse Homöopathie. Anwendbar ist die Konzeption jedoch Kontext-übergreifend, sodass auch andere gesellschaftlich kontrovers diskutierte Themen anhand der Wissenschaftskriterien bewertet werden können. Das Ziel des zweiten Mesozyklus bestand darin, eine weitere Kontroverse für den Einsatz in choice^{2}reflect aufzubereiten.

Einen weiteren Forschungsschwerpunkt stellt die Entwicklung eines methodischen Instruments dar, das Lernende in der Anwendung der Prüfkriterien auf Aussagen und Studien unterstützt. Im Rahmen des dritten Mesozyklus wurde ein Prüfschema

entwickelt und im praktischen Einsatz mit Studierenden erprobt.
Die Ausdifferenzierung dreier Mesozyklen beantwortet bereits einen kleinen Teil der übergeordneten Forschungsfrage:
Die Unterrichtskonzeption choice^{2}reflect kann durch 1. Ergänzen des Prüfkriterien-Pools, 2. den Einsatz einer weiteren Kontroverse und 3. die Konzeption eines Prüfschemas weiterentwickelt werden.
Damit ist festgelegt, wie sich der *praktische Output* dieser Arbeit gestaltet. Zur Beantwortung der DBR-typischen Frage „*What works?*“ reicht diese Aussage allerdings nicht aus. Im folgenden Abschnitt wird detailliert dargelegt, welche Materialien in den Mesozyklen entstanden sind und inwiefern deren Praxistauglichkeit festgestellt wurde.
Im Anschluss daran werden Gestaltungsmerkmale der ausgearbeiteten Materialien herausgestellt, die für die Einstufung als *praxistauglich* verantwortlich gemacht werden *könnten*.

12 Praktischer Output des ersten Mesozyklus

An verschiedenen Stellen dieser Arbeit wird beschrieben, welche Aufgabe der erste Mesozyklus besitzt: Im Rahmen mehrfacher Durchläufe wird der Prüfkriterien-Pool durch weiteres empirisch erprobtes Unterrichtsmaterial ergänzt.
Die Entwicklung des Mesozyklus mit Beschreibung der Iterationen und der Genese einer MZ-spezifischen (Sub-)Fragestellung befindet sich im Kapitel 9.
Die untersuchte Fragestellung lautet: „Inwieweit sind Lernende in der Lage, die Wissenschaftskriterien *Falsifizierbarkeit*, *Korrelation und Kausalität*, *Randomisierung* und *Reproduzierbarkeit* zu erarbeiten und anzuwenden?“
Jedes Prüfkriterium stellt eine eigene Iteration (oder auch einen Mikrozyklus) des Mesozyklus dar. Das methodische Vorgehen bei der Entwicklung und Erprobung der Lernmaterialien wurde innerhalb des Zyklus möglichst konstant gehalten. Zentrale Ergebnisse der Mikrozyklen werden im Folgenden zusammengetragen.

12.1 Falsifizierbarkeit

Im Zuge der Erarbeitung von *Falsifizierbarkeit* (Kapitel 9.1) erlernen die Schüler:-innen, dass das endgültige Beweisen wissenschaftlicher Hypothesen unmöglich ist. Stattdessen müssen Aussagen so strukturiert sein, dass sie mithilfe von Beobachtungen und Experimenten widerlegt werden können.
In der konzipierten Unterrichtsstunde werden die Lernenden mit einer nicht-falsifizierbaren Aussage konfrontiert, dessen Unwissenschaftlichkeit sie festzustellen haben. Es folgt eine kurze theoretische Einführung in die Falsifizierbarkeit, woraufhin die Schüler:innen verschiedene Aussagen im Stil sozialer Medien hinsichtlich der Widerlegbarkeit prüfen.

In der Erprobung zeigte die Lerngruppe eine insgesamt hohe Diskussionsbereitschaft mit teils von den Schüler:innen selbstständig gewählten Beispielen für nicht-falsifizierbare Aussagen.
In den Gruppendiskussionen und schriftlichen Antworten der Schüler:innen konnten Hinweise auf kognitive Konflikte gefunden werden, wenngleich auftretende Konzeptwechsel nicht mit Sicherheit bestätigt werden konnten. Alle Lernziele fanden Beachtung und mindestens zwei dieser Lernziele wurden in der Erprobung erreicht. Die Gesamtheit der empirischen Ergebnisse lässt darauf schließen, dass das Unterrichtsmaterial zur *Falsifizierbarkeit* für den Einsatz in Phase III des Unterrichtskonzepts gut geeignet ist.
Bei den erprobten Materialien handelt es sich um einen Satz Arbeitsblätter für den direkten Einsatz in der Stunde, Prüfkarten für nachfolgende Sicherungs- und Vertiefungsphasen sowie den Stundenverlaufsplan.

12.2 Korrelation und Kausalität

Störche bringen keine Neugeborenen. 5G-Mobilfunk hat nicht die Corona-Pandemie ausgelöst. Und dass die App *Instagram* nicht für eine kurze Akkulaufzeit verantwortlich sein muss, wird in der Unterrichtsstunde zur *Korrelation und Kausalität* geklärt.
Im zweiten Durchlauf des ersten Mesozyklus (siehe Kapitel 9.2) geht es darum, dass das Auftreten statistischer Korrelationen nicht zwangsläufig auf Ursache-Wirkungs-Beziehungen schließen lässt.
Die Lernenden setzen sich mit einer Studie (fiktiver Natur) auseinander, die den Zusammenhang zwischen der App *Instagram* und einer verkürzten Akkulaufzeit von Smartphones untersucht. Die Studie gelangt zu dem Ergebnis, dass *Instagram* ein „Stromfresser“ sei - diese Schlussfolgerung prüfen die Lernenden auf Scheinkausalitäten.
Dabei agieren sie zunächst ohne Kenntnis der fachlichen Hintergründe zu Scheinkausalitäten: Sie setzen sich mit Rohdaten der zitierten Studie auseinander und führen eine eigenständige Datenauswertung durch, aufgrund derer sie eine Schlussfolgerung ziehen.
Es folgt ein fachlicher Input zu den Hintergründen des Prüfkriteriums.
Abschließend bewerten die Schüler:innen die zuvor zitierte Schlussfolgerung und

stellen fest, dass es sich dabei um eine Scheinkausalität handelt.

In der Erprobung der Unterrichtseinheit wurde herausgefunden, dass alle Lernziele adressiert wurden. Vermutlich wurde das erste Lernziel vollständig, die anderen drei teilweise erreicht. Das Verhalten der Lernenden wurde auf *Aha-Erlebnisse* untersucht. Diese wurden als erste Hinweise für die Auflösung kognitiver Konflikte angesehen, wobei nicht abschließend bestätigt werden konnten, inwiefern den Irritationen tatsächlich kognitive Konflikte zugrunde lagen.
Die von der Lehrkraft gegebenen Lehrimpulse wurden genau analysiert. Dabei wurde einerseits festgestellt, dass die meisten Impulse für passende und anschlussfähige Reaktionen sorgten. Andererseits fiel auf, dass diese den Unterricht auch stark vorstrukturierten. Anlässlich dieser Erkenntnisse wurden Leit-Impulse für jede Plenumsphase formuliert.
Auch Verständnisschwierigkeiten im Umgang mit den Arbeitsmaterialien sowie Off-Task-Interaktionen wurden untersucht. Die Ergebnisse wurden in einem anschließenden Re-Design zur Optimierung des Ablenkungspotentials und Minimierung von Verständnisschwierigkeiten verwendet.
Insgesamt hat sich die entwickelte Unterrichtseinheit, bestehend aus einigen Arbeitsblättern für die Studie, den Prüfkarten für die Erarbeitungsphase und einem Stundenverlaufsplan, als geeignet erwiesen.

12.3 Randomisierung

Kapitel 9.3 stellt dar, wie eine Unterrichtsreihe zur *Randomisierung* entwickelt, erprobt und weiterentwickelt wurde.
Als eines der sogenannten *Goldstandard-Kriterien* (Jungkamp 2021; Akobeng 2005) beschreibt Randomisierung die zufallsbasierte Zuteilung von Testpersonen auf die Experimental- und Kontrollgruppe. Im Rahmen des Unterrichtskonzepts choice^{2}reflect lernen die Schüler:innen neben dieser Bedeutung auch die Relevanz einer zufälligen Stichprobenziehung zur korrekten Angabe von Repräsentativitäten unter dem Stichwort Randomisierung kennen.
Vier Lernziele werden in der Unterrichtseinheit verfolgt. Sie bauen aufeinander auf und beschreiben die Progression bei der Erarbeitung der Randomisierung.
Die Schüler:innen beschreiben zunächst die Notwendigkeit der Stichprobenziehung

aus Grundgesamtheiten, erklären daraufhin das Prinzip der Randomisierung und beschreiben diese als Voraussetzung für Repräsentativität.
Das Lernmaterial umfasst eine gekürzte und vereinfachte Version der *Heinsberg-Studie* mit Nennung der zentralen Methoden. Diesen Text durchsuchen die Lernenden im Hinblick auf Ergebnisse der Studie und tragen diese in eine Tabelle ein.
Im Anschluss erhalten die Lernenden die entwickelten Prüfkarten und nutzen diese zur Erarbeitung des Prüfkriteriums *Randomisierung*. Dabei erfahren sie, was Randomisierung ist und wie man sie bei der Stichprobenziehung einsetzt. Die Schüler:innen prüfen dann die Heinsberg-Studie auf randomisierte Stichprobenauswahl und ziehen Rückschlüsse auf die Repräsentativität der Stichprobe.
Schließlich nutzen sie die Erkenntnisse aus der Erarbeitungsphase, um eine Schlagzeile der BILD-Zeitung zu bewerten.

Die Erprobung des Lernmaterials hat gezeigt, dass die Lernziele 1 bis 3 vermutlich mit Einschränkungen erreicht wurden, während das Lernziel 4 vollständig erreicht wurde. Einige Lernende der untersuchten Lerngruppe waren in der Lage, das Konzept der *Randomisierung* zielführend zu erarbeiten und auf die Heinsberg-Studie anzuwenden. Allerdings trifft diese Beobachtung nicht auf alle Lernenden zu; in einigen Gruppen und Unterrichtsphasen wurden Verständnisschwierigkeiten und Off-Task-Interaktionen beobachtet. Im Re-Design wurden daher Verbesserungsmöglichkeiten vorgeschlagen.

12.4 Reproduzierbarkeit

Wenn ein Experiment unter den gleichen Bedingungen von anderen Wissenschaftler:innen erneut durchgeführt wird, muss das gleiche Ergebnis erhalten werden - das Experiment muss *reproduzierbar* sein (Tetens 2018). Im Kapitel 9.4 werden (Weiter-)Entwicklung und Erprobung der Lernmaterialien zur *Reproduzierbarkeit* erläutert.
Drei Lernziele definieren den zu erwartenden Kompetenzzuwachs und ziehen sich als Leitfaden durch die entwickelte Unterrichtsstunde. Die Schüler:innen erarbeiten spezifische Kriterien für Reproduzierbarkeit und bewerten anhand dieser Kriterien sowohl ein (Labor-)Experiment als auch eine empirische Studie.
Dabei kennen sie die fachlichen Hintergründe des Prüfkriteriums zunächst nicht und sollen die Notwendigkeit reproduzierbar gestalteter Experimente selbst erfahren:

Anhand einer unvollständigen Versuchsbeschreibung versuchen die Schüler:innen, eine „Zitronenbatterie“ nachzubauen. Die Erfolgsaussichten sind dabei gering, da die Beschreibung keine genauen Details zur Durchführung enthält. Die Lernenden formulieren daraufhin ihre Probleme bei der Reproduktion der Zitronenbatterie.
Die ausformulierten Probleme werden in eine Gruppenarbeitsphase überführt, in der die Lernenden anhand eines Satzes an Arbeitsblättern die Reproduzierbarkeit erlernen und direkt auf die vorherige Experimentalphase anwenden.
Auf die Erkenntnis, dass die Versuchsbeschreibung *nicht reproduzierbar* ist, folgt die gemeinsame Betrachtung eines Zeitungsartikels, in dem über eine Studie zum Thema *E-Scooter* berichtet wird. In dem Bericht fehlen allerdings wichtige Kenndaten der zitierten Studie, weshalb diese als *nicht reproduzierbar* eingestuft wird.

Die Unterrichtseinheit wurde in einer neunten Klasse erprobt. Bei durchgängig hoher Schüleraktivität und zahlreichen Diskussionen in Plenums- und Gruppenarbeitsphasen erlangten die Lernenden einen guten Zugang zur Bewertung von Versuchsprotokollen im Hinblick auf Reproduzierbarkeit. Auch das Übertragen des Konzepts auf wissenschaftliche Studien hat bei vielen Schülergruppen gut funktioniert.
Die Analysen im Kapitel 9.4 haben gezeigt, dass das erste und das dritte Lernziel vollständig erreicht wurden, während das Lernziel 1 nur von einigen Schüler:innen erreicht werden konnte. Ein paar Unklarheiten im Umgang mit dem Material wurden aufgedeckt und zur Verbesserung in das Re-Design übernommen.

13 Praktischer Output des zweiten Mesozyklus

13.1 Kriterien eines guten Kontexts

In der Vorbereitungsphase des zweiten Mesozyklus wurden literaturbasierte Anforderungen an „gute Kontexte“ herausgearbeitet. Es wurde herausgestellt, dass *socio-scientific-issues* (SSI) diesen Anforderungen im Allgemeinen gerecht werden und für den Einsatz in choice^{2}reflect gut geeignet sind. *Homöopathie* als SSI und kontrovers diskutiertes Thema hat sich im Rahmen der Forschungsarbeiten von Jungkamp (2021) als gut geeignet erwiesen.
Im Zuge der Weiterentwicklung von choice^{2}reflect gemäß der übergeordneten Fragestellung besteht ein großes Forschungsinteresse an der Etablierung eines neuen Kontextes, also eines neuen SSI mit gesellschaftlicher Tragweite und naturwissenschaftlichem Hintergrund. Das neue Thema bildet den kontextuellen Rahmen der ersten und letzten beiden Phasen; es stellt den „Haupt-Aufhänger“ und die fachlichen Inhalte bei der Bewertung mithilfe von Prüfkriterien.
In der Design-Phase des zweiten MZ wurde das gesellschaftlich kontrovers diskutierte Thema *Nahrungsergänzungsmittel* als möglicherweise passender („guter“) Kontext identifiziert.

13.2 Validierung des Themas *Nahrungsergänzungsmittel* als guter Kontext

Die Hypothese, NEM eigne sich als guter Kontext, wurde durch eine umfangreiche Fragebogenerhebung zu Einnahmeprävalenz, Verwendungsgewohnheiten und Bewertung von NEM mit 237 Schüler:innen empirisch überprüft. Es hat sich herausgestellt,

dass etwa 60 % der Befragten NEM konsumieren. Dabei sind Magnesium-Präparate die am häufigsten genannten NEM, gefolgt von verschiedenen Vitamin-Präparaten.
Die Schüler:innen schreiben den Präparaten Wirkungen zu, die über die reine Vermeidung von Mangelerscheinungen weit hinausgehen. Teils sind die erhofften Wirkungen auch sehr unspezifisch und unreflektiert.
NEM-bezogene Bewertungsfragen werden von den Lernenden unterschiedlich beantwortet. Grundsätzlich erkennen sie die Bedeutung einer gesunden Ernährung und lehnen die These ab, dass eine ungesunde Ernährung durch NEM ausgeglichen werden könne. Andererseits stufen die Befragten NEM als eher sinnvoll, gesund und ungefährlich ein. Es bleibt offen, in welcher Tiefe sich die Schüler:innen mit NEM auseinandersetzen. Allerdings werden als Informationsquelle die eigenen Eltern benannt, weiterhin soziale Medien und Werbung.
Die Fragebogenerhebung konnte zeigen, dass Nahrungsergänzungsmittel ein für Schüler:innen der Sekundarstufe I relevantes und alltagsnahes Thema darstellen. Es wird keineswegs einheitlich beurteilt und nicht immer kritisch hinterfragt. Die Kontroversität des Themas bei gleichzeitig geringem Informationsgrad der Lernenden wurde herausgestellt und damit der Kontext *Nahrungsergänzungsmittel* als geeignet für den Einsatz in choice^{2}reflect eingestuft.

13.3 Entwicklung und Erprobung der Phasen I und II

Die ersten beiden Phasen des Unterrichtskonzepts bestehen aus einem fachspezifischen Problemaufriss mit anschließender Positionierung.
Das Lernmaterial für diese Phasen entstand im Kapitel 10.3 aufbauend auf den Ergebnissen der Fragebogenerhebung: Die am häufigsten genannten Supplemente waren Magnesium-Präparate mit der Intention, Muskelkrämpfe zu vermeiden. Daher bildet die Magnesium-Supplementierung als Teil der Kontroverse um Nahrungsergänzungsmittel den zentralen Kontext für Problemaufriss und Positionierung; später dann auch für die Anwendung und Reflexion.
Lernende werden mit einer Problemstellung konfrontiert, die in ähnlicher Weise auch im Alltag der Schüler:innen auftreten könnte. In dem Gruppenchat einer Fußballmannschaft klagt ein Mitglied über vermehrt auftretende Muskelkrämpfe und erkundigt sich in der Gruppe nach Behandlungsmöglichkeiten. Der Trainer schlage die Ein-

nahme von Magnesium-Tabletten vor, die Gruppe wird nach ihrer Meinung zu diesem Vorschlag befragt.

Anschließend werden in dem Gruppenchat zwei Beiträge aus dem sozialen Netzwerk *Instagram* bereitgestellt. Einer dieser Beiträge informiert über ein bestimmtes *Magnesium-Kalium*-Präparat und dessen Wirkungen auf den Körper. Der andere bewirbt Bananen als gute Magnesium-Quelle.

Die Schüler:innen positionieren sich zu der Frage, ob zur Behandlung von Wadenkrämpfen täglich eine Magnesium-Tablette eingenommen oder täglich eine Banane verzehrt werden sollte.

In einer anschließenden Plenumsphase wird die Meinungslage in der Lerngruppe diskutiert. Dabei wird herausgestellt, dass eine fundierte Entscheidung nicht ohne zusätzliche, wissenschaftlich-gesicherte Informationen getroffen werden kann. Damit wird zur Phase III übergeleitet.

Erprobt wurden die Phasen I und II einmal als alleinstehende Unterrichtsstunde im Jahr 2018 und ein anderes Mal im Rahmen einer Gesamterprobung im September 2020.

Das Ziel der ersten Erprobung war es, Erkenntnisse über den Umgang von Schüler:-innen mit Positionierungsfragen zum Kontext Magnesium-Supplementierung zu erhalten. Es wurde ein Blick auf die in der Klasse vorherrschende Meinungslage sowie die verschiedenen Standpunkte und Begründungen der Lernenden geworfen, siehe dazu auch Kapitel 10.3.2.

Der Einstiegsimpuls wurde von den Lernenden gut angenommen, sie nannten im Plenum verschiedene Supplemente, die sie selbst konsumieren und bemaßen an den eigenen Erfahrungen die Notwendigkeit der Nahrungsergänzung. Alle Schüler:innen positionierten sich zur Problemfrage und gaben meist auch Begründungen für ihre Meinungen an. Es hat sich gezeigt, dass diese Begründungen jedoch häufig nur wenig reflektiert waren:

Lernende vertrauten den Angaben der Instagram-Beiträge, außerdem ist in den Gruppengesprächen erkennbar, dass eine hohe Unsicherheit bei Angabe der Begründungen herrschte. Weiterhin ist das Meinungsbild in der Klasse nicht eindeutig, beide Positionen wurden in ähnlichem Maße gewählt.

Die uneindeutige Meinungslage gepaart mit hoher Diskussionsbereitschaft und den eigenen Erfahrungen der Lernenden bei gleichzeitig geringem Informationsstand hat

für die Einschätzung gesorgt, dass das Unterrichtsmaterial als Hinführung zur Phase III gut geeignet ist.
Manche Elemente der entwickelten Einheit haben im Praxiseinsatz nicht wie intendiert funktioniert, daher wurden sie im Re-Design verbessert bzw. durch Alternativen ersetzt. So wurde beispielsweise der Einsatz des digitalen Tools *Kahoot* durch ein zusätzliches Arbeitsblatt ersetzt, auf dem die Lernenden ihre Position ankreuzen können. Dadurch wurde die Wahrscheinlichkeit von technischen Problemen reduziert und gleichzeitig wurden die Anforderungen für die Durchführung gesenkt (kein Internetzugang mehr nötig).
Weiterhin wurde eine Videodatei erstellt, die den Gruppenchat der Mannschaft wiedergibt.
Für die Moderation der Diskussionsphase durch die Lehrkraft wurde eine Art „Stopp-Signal“ definiert; bei Auftreten bestimmter Argumentationen in der Lerngruppe ist das Ziel der Diskussion erreicht und es kann zur Phase III übergeleitet werden.

Die vorgenommenen Optimierungen des Lernmaterials wurden in der zweiten Erprobung eingesetzt. Eine Auswertung der videographierten Unterrichtsstunde befindet sich im Kapitel 10.3.4. Die Einschätzung aus der ersten Erprobung, dass Nahrungsergänzung in Form von Magnesium-Präparaten ein für Schüler:innen relevantes Thema darstellt, bestätigte sich in der zweiten Erprobung.
Auch das Meinungsbild in der Lerngruppe fiel erneut uneindeutig aus und die Begründungen ließen darauf schließen, dass die Entscheidungen kaum auf rationaler Basis getroffen wurden. In der Diskussion tritt noch deutlicher hervor, dass die Lernenden unwissenschaftliche Informationsquellen nutzen, bspw. schlugen sie vor, „auf die Mehrheit zu hören“ oder „einfach Präparate auszuprobieren“.
Damit hat sich bestätigt, dass das entwickelte Lernmaterial die Schüler:innen ideal auf die Phase der Gewinnung wissenschaftlicher Prüfkriterien vorbereitet.

13.4 Entwicklung und Erprobung der Phasen IV und V

Nach Erarbeitung der Prüfkriterien wird die Phase IV eingeleitet: *Anwendung auf Kontroverse*. Diese Phase steht, nachdem in Phase III für jedes Kriterium ein eigener

Kontext behandelt wurde, wieder im Kontext NEM. Die in Phase I aufgeworfene Problemstellung wird hier erneut aufgegriffen; den Lernenden wird signalisiert, dass im folgenden Abschnitt des Unterrichts wissenschaftliches Prüfen gefragt ist.

Das Prüfschema übernimmt in der Phase eine bedeutungsvolle Funktion: Es bildet das zentrale Instrument, mit dem Aussagen und Studien zu der in Phase I aufgetretenen Problemstellung auf ihre Wissenschaftlichkeit geprüft werden.

Begonnen wird die vierte Phase mit einem Rückbezug auf die in Phase II herausgearbeiteten Desiderate, nach denen wissenschaftlich gesicherte Informationen für eine rationale Entscheidungsfindung vonnöten sind. Den Desideraten wird entsprochen, indem die Lehrkraft eine echte Studie thematisiert.

Als echte Studie eignet sich z. B. eine aus der Meta-Studie von Garrison et al. (2020) exkludierte Untersuchung zur Supplementierung von Magnesium in der Schwangerschaft zur Vermeidung von Wadenkrämpfen.

In dieser Studie (mit ergänzter fiktiver Schlussfolgerung) werden drei Prüfkriterien nicht beachtet (siehe Kapitel 10.3.4), sodass sie eine große Angriffsfläche für die Anwendung des Prüfschemas bietet.

Lernende wenden das Prüfschema auf die Studie an und sollen herausfinden, dass die Studie wissenschaftlichen Standards nicht genügt und daher in der symbolischen *Mülltonne* landet.

In der Phase V resümieren die Lernenden, welche wirklich belastbaren Studien und Aussagen nach der Prüfung übrig geblieben sind - nur diese können in der abschließenden Entscheidungsfindung verwendet werden. Ihre begründete Entscheidung explizieren sie in der Klasse z. B. im Rahmen der Methode *Blitzlicht*.

Die Gesamterprobung im September 2020 schloss mit einer Anwendungs- und Reflexionsphase ab. Die Unterrichtsmaterialien wurden hier eingesetzt, um die Thematik der NEM zurück in das Gedächtnis der Lernenden zu rufen und sie anschließend zu bewerten.

Die Lernenden prüften das Stichprobenauswahlverfahren der vorgelegten Studie. Einige bemerkten, dass es nicht randomisiert durchgeführt wurde. Allerdings erkannten das nicht alle Schüler:innen, außerdem bezogen sie in ihre Argumentationen fehlende Repräsentativitäten nicht mit ein.

Auch die Beurteilungen der Wissenschaftlichkeit der geprüften Studie fielen unterschiedlich aus. Erst in der Plenumsphase konnte ein Konsens in der Lerngruppe über

die Beurteilung der Studie generiert werden.
Die fünfte Phase wurde in der Erprobung in Form einer kurzen *Blitzlicht*-Runde durchgeführt, in der jede Schülerin und jeder Schüler ein kurze Statement mit der eigenen Entscheidung zum Thema präsentierte. Dabei positionierten sich die meisten Lernenden *gegen* die Magnesium-Supplementierung zur Vermeidung von Muskelkrämpfen.
Die Ergebnisse aus der erprobten Unterrichtsstunde sprechen dafür, dass die entwickelten Lernmaterialien für die Phasen IV und V förderlich sind für das Treffen rationaler Entscheidungen. Allerdings wurden auch durch die Untersuchungsmethode bedingte Einschränkungen formuliert - dadurch kann auf das Erreichen der übergeordneten Zielsetzung nicht sicher geschlossen werden.

14 Praktischer Output des dritten Mesozyklus

Der dritte große Forschungsschwerpunkt behandelt die Entwicklung eines methodischen Instruments zur Bewertung von Aussagen, Experimenten und Studien im Hinblick auf dessen Wissenschaftlichkeit.
Das Kapitel 11 befasst sich mit der zugehörigen Fragestellung. Es stellt theoretische Hintergründe dar, die zu einer kategorischen Trennung aussagenlogischer Prüfkriterien von den Goldstandard-Kriterien geführt haben. Es wird beschrieben, wie aus dieser zunächst gedanklichen Trennung ein Prüfschema entwickelt wurde, das für den direkten Einsatz in den Phasen IV und V der Konzeption choice^{2}reflect konzipiert ist. Es expliziert die Trennung und visualisiert die unterschiedlichen Anwendungsbereiche der Wissenschaftskriterien: Neben empirischen Studien können auch weniger konkrete, im Alltag auftretende Thesen, Behauptungen und Vermutungen im Prüfschema eingesetzt werden.

Erprobt wurde das Prüfschema mit Lehramts-Studierenden im Rahmen eines chemiedidaktischen Seminars. Die Studierenden wurden nicht in die Unterrichtskonzeption eingeführt. Sie wurden mit einer Schlagzeile und einem kurzen Artikel zur Kontroverse *Veganimsus* konfrontiert. Dort wurde eine Studie zitiert, die die Aussage der Schlagzeile unterstützt.
Die Studierenden erhielten die zitierte Studie und das Prüfschema. Sie wurden angewiesen, anhand des Schemas die Studie zu bewerten und ihre Gedanken zu verbalisieren.
Die aufgezeichneten Videos sowie die ausgefüllten Fragebögen wurden hinsichtlich der Analyseschwerpunkte *Vorwissen*, *Leserichtung*, *rationale Entscheidungen* und *Reaktionen* untersucht. Die Analyse hat ergeben, dass bei vielen der teilnehmenden Studierenden bereits ein ausgeprägtes Vorwissen zu einzelnen Prüfkriterien vorhanden

war. Lediglich zu den Prüfkriterien *kontrollierte Bedingungen* und *Reproduzierbarkeit* wurden wenige zielführende Erklärungen abgegeben.
Bezüglich des Umgangs mit dem Prüfschema wurde herausgefunden, dass im ersten Durchlauf der Erprobung nur wenige Zweiergruppen die korrekte Leserichtung befolgten. In dem zweiten Durchlauf, in dem eine optimierte Version des Prüfschemas zum Einsatz kam, befolgten nahezu alle Gruppen die korrekte Leserichtung.
Bei gemeinsamer Betrachtung mit den Reaktionen, die Studierende beim ersten Betrachten des Schemas äußern, wurde festgestellt, dass die erste Version des Prüfschemas ohne zuvor erfolgte Instruktionen nur bedingt intuitiv verständlich ist. Zur Erhöhung der Intuitivität wurden im Kapitel 11.6 Verbesserungen vorgeschlagen und umgesetzt. Damit wurden auch die Einbindungsmöglichkeiten in den naturwissenschaftlichen Unterricht der Sekundarstufe I erhöht und die kognitiven Hürden für den Umgang mit dem Schema etwas herabgesetzt. I
m Zuge des Re-Designs wurde auch die Benennung des Kriteriums *kontrollierte Bedingungen* überdacht; hier wurde als Alternative das Prüfkriterium *Variablenkontrolle* vorgeschlagen. Auch die Stichwort-Fragen, die auf dem Prüfschema neben jedem Kriterium abgedruckt sind, wurden für die rechte Seite des Schemas überarbeitet.
In den Gruppendiskussionen wurde deutlich, dass die Studierenden die Elemente des Prüfschemas in ihre Argumentationen einbeziehen und auch die plakativen Bilder der unwissenschaftlichen Mülltonne und der wissenschaftlichen Schatzkiste übernehmen. Letztlich trafen fast 80 % der Studierenden eine rationale Entscheidung auf Basis des Prüfschemas.
Aus dem dritten Mesozyklus dieses Projektes geht somit eine bereits optimierte Version des durch die Forschungsfrage geforderten methodischen Instruments hervor: Ein Prüfschema zur Bewertung von Thesen, Hypothesen und Studien.

15 Theoriebildung

„Wie kann die Unterrichtskonzeption choice^{2}reflect als praxistaugliche Intervention zur Bewertung kontroverser Kontexte weiterentwickelt werden?“

Die zentrale, dieses Projekt umspannende Forschungsfrage kann an diesem Punkt der Arbeit bereits als beantwortet angesehen werden; allerdings nur in einer Teildimension.

Die zweidimensionale Bedeutung des Frageworts *Wie* wurde bereits zu Beginn des Re-Framings herausgestellt. Die Erforschung von *Elementen*, die zur Weiterentwicklung der Konzeption beitragen, ist abgeschlossen und in Form des *praktischen Outputs* ausformuliert. Die Wirksamkeit dieser Elemente, also das „Funktionieren“ im realen Bildungskontext, wurde in zahlreichen Durchläufen drei großer Mesozyklen erforscht.

Es schließt sich eine Theoriebildung an, die im Rahmen des Design-Based-Research-Ansatzes häufig mit der Fragestellung „*How, when, and why it works?*“ angegeben wird (Rohrbach-Lochner 2019, S. 19; Cobb et al. 2003).

Dabei werden Theorien entwickelt, die nach ihrem Geltungsbereich aufgeteilt werden. Häufig werden *lokale Theorien* für Interventionen angegeben, die „mehrfach unter klar definierten, aber die Aussagekraft begrenzenden Bedingungen durchgeführt“ wurden (McKenney & Reeves 2012). Theorien *mittlerer Reichweite* zu entwickeln ist bei wiederholter Untersuchung z. B. in verschiedenen Versuchsgruppen möglich. Selten werden auch Theorien *höchster Aussagekraft* gebildet. Diese entstehen jedoch nur, wenn Phänomene mehrfach in unterschiedlichen Kontexten unter Kombination von Theorien mittlerer Reichweite untersucht werden.

15.1 Ausgangstheorien des Vorgängerprojektes

Jungkamp (2021) hat den didaktischen Rahmen des Unterrichtskonzepts bereitgestellt. Er entwickelte das Vorgehen in fünf Phasen, bereitete den Kontext Homöopathie für den Einsatz in dem Konzept auf und erprobte einige Prüfkriterien. Er wählte den *Cluster Approach* zur schrittweisen Erarbeitung und Anwendung der Kriterien und beschreibt, dass sich dieser als zielführend erwiesen hat.
Die Erarbeitung der Prüfkriterien verläuft nach Jungkamp zielführend, wenn die Materialien so präsentiert werden, dass die darin vermittelten wissenschaftlichen Denk- und Arbeitsweisen ein (für die Schüler:innen) besseres Ergebnis liefern als eine intuitive, Bauchgefühl-gesteuerte Herangehensweise. Außerdem stellte es sich im Rahmen seiner Erprobungen als förderlich heraus, Lernende für eine kurze Zeit „bewusst zu verwirren“ und somit für eine intensivere Beschäftigung mit Behauptungen und ambivalenten Informationen zu sorgen.
Weiterhin stellt Jungkamp die Hypothese auf, dass Lerninhalte mit einem zu hohen oder zu niedrigen Grad an Emotionalität hinderlich seien für eine rationale Entscheidungsfindung, da bei Auftreten von Emotionalitäts-Extrema intuitive Heuristiken die Entscheidungen übernehmen.

Es handelt sich dabei um *lokale Theorien*, die als Anregungen für weitere Forschungsprojekte dienen können, jedoch zunächst nur eine begrenzte Aussagekraft bieten. Im Zuge der Materialentwicklung in diesem Projekt wurden von Jungkamp formulierte Erkenntnisse und Desiderate berücksichtigt und im Material implementiert (beschrieben im Kapitel 5.5). Einige der im folgenden Abschnitt gebildeten Theorien können aufgrund des neuen Forschungskontext gegebenenfalls eine *mittlere Reichweite* erreichen.

15.2 Gestaltungsmerkmale für Prüfkriterien

Von Jungkamp herausgestellt wurde die Erkenntnis, dass Prüfkriterien so gestaltet werden sollten, dass die zu vermittelnden Denk- und Arbeitsweisen ein besseres Ergebnis liefern als intuitive Problemlösestrategien.
Das daraus resultierende Gestaltungskriterium zur Berücksichtigung bei der Materialentwicklung ist eine Thematisierung der eigenen intuitiven Herangehensweise

(der Schüler:innen). Um zu erkennen, dass diese keine zur wissenschaftlichen Denkweise gleichwertigen Ergebnisse erzielen, sollten die Lernenden zunächst ohne Kenntnis der fachlichen Hintergründe mit einem Problem konfrontiert werden und dieses Problem intuitiv zu lösen versuchen.
Diese Herangehensweise wird im Unterrichtsmaterial zu allen Prüfkriterien berücksichtigt:

- Bei der *Falsifizierbarkeit* begegnen die Lernenden zunächst Birtes und anschließend Bernds Aussage, die intuitive Argumentationsmuster enthalten. Die Schüler:innen prüfen auf Überprüfbarkeit und schließlich auf Widerlegbarkeit und werden dann in die Hintergründe der Falsifizierbarkeit eingeführt.
- Die Unterrichtseinheit zur *Korrelation und Kausalität* beginnt mit einer Studie zum Akkuverbrauch von Instagram, dessen Schlussfolgerung von den Lernenden bewertet wird. Erst danach erhalten sie Informationen zu Scheinkausalitäten und erkennen, dass mit Kenntnis dieses Kriteriums die Schlussfolgerungen empirischer Untersuchungen besser gedeutet werden können.
- *Randomisierung* ist eines der wenigen Prüfkriterien, in denen die Prüfkarten das direkt im Unterricht thematisierte Material darstellen. Diese sind so entwickelt, dass sie von einem intuitiveren Beispiel („HipHop in Berlin“) ausgehend die Randomisierung enthüllen.
- *Verblindungs*-Materialien sind bereits bei Jungkamp (2021) entstanden und berücksichtigen die Intuition bei der Blind-Verkostung von Joghurt-Sorten.
- Im Rahmen der von Jungkamp entwickelten Erarbeitung von *Variablenkontrolle* (bzw. *kontrollierten Bedingungen*) wird betrachtet, wie Schüler:innen bei der Planung einer Experimentalreihe intuitiv vorgehen.
- *Reproduzierbarkeit* wird anhand eines unvollständigen Versuchsprotokolls beim Herstellen einer Zitronen-Batterie erarbeitet. Genau beobachtet wurden die intuitiv auftretenden Reaktionen der Lernenden auf die unvollständige Versuchsbeschreibung.

Jedes der Prüfkriterien bildet einen Durchlauf des ersten Mesozyklus in diesem Projekt. Somit besteht eine mittlerweile gut ausdifferenzierte Datenbasis im Bezug auf das Verwenden von Intuitions-Appellen im Unterrichtsmaterial. Die Durchläufe fanden unter sehr unterschiedlichen Bedingungen mit unterschiedlichen Lerngruppen

in einem vergleichsweise langen Zeitraum statt. Damit sind die Voraussetzungen für die Genese einer *Theorie mittlerer Reichweite* gegeben.

Die Theorie, dass das Aufrufen intuitiver Problemlösestrategien bei den Lernenden mit anschließendem Vergleich zum wissenschaftlichen Hintergrund des Prüfkriteriums zu einer zielführenden Erarbeitung des Prüfkriteriums führt, wird durch die vorliegenden Daten dieses Projekts unterstützt. Das ist beispielsweise an der Art und Weise erkennbar, mit der die Lernenden die *Reproduzierbarkeit* erarbeiten (siehe Kapitel 9.4.3): Sie versuchen zunächst, anhand des unvollständigen Versuchsprotokolls eine funktionsfähige Zitronenbatterie herzustellen. Dann stellen sie jedoch selbstständig fest, dass die Angaben des Protokolls nicht genügen und fordern in der folgenden Plenumsphase, dass weitere Einflussfaktoren wie das Material der verwendeten Münze oder die Beschaffenheit der Zitrone genannt werden.

Diese Forderungen stellen die Lernenden, ohne zuvor mit den fachlichen Hintergründen der Reproduzierbarkeit konfrontiert worden zu sein. Die Hintergründe erlernen sie erst im Anschluss an die intuitive Problemlösung. Danach wenden sie dieses Wissen auf eine empirische Studie an. In der Erprobung gelang diese Anwendung, weshalb die Vorgehensweise als angemessen angesehen wird.

Die im Rahmen des ersten Mesozyklus erhaltenen Erkenntnisse lassen eine weiterführende Theoriebildung zu. Die entwickelten Unterrichtseinheiten zu den Prüfkriterien enthalten in den meisten Fällen eine deutliche Zweiteilung: Auf der einen Seite wird das Prüfkriterium (unter Verwendung intuitiver Herangehensweisen) *experimentell* erarbeitet. Auf der anderen Seite findet im Anschluss an den theoretischen Input zum Kriterium meist eine Anwendung des PK auf eine *empirische Untersuchung* statt.

Das bedeutet, dass die Lernenden im intuitiven Teil eigenständig ein Experiment durchführen, anhand dessen das Nicht-Beachten eines Prüfkriteriums deutlich wird. Auch hier lässt sich die Reproduzierbarkeit als Beispiel heranziehen. Wie beschrieben, bauen die Schüler:innen eine Zitronen-Batterie anhand einer unvollständigen nicht-reproduzierbar gestalteten Versuchsbeschreibung. Dann erhalten sie den fachlichen Input zur Reproduzierbarkeit. Zuletzt behandeln sie eine real existierende Studie zur Thematik der Elektro-Scooter und prüfen diese auf Reproduzierbarkeit.

Auch die anderen PK werden in ähnlicher Weise erarbeitet. Lediglich zur Randomisierung existiert kein praktisches Experiment, dort beschäftigen sich die Lernenden

direkt mit der Heinsberg-Studie.
Es hat sich gezeigt, dass die Lernenden mit dieser Zweiteilung gut zurechtkommen und das Prüfkriterium meist zielführend erarbeiten. Durch den Weg über die Durchführung von Experimenten ist der Bruch zum „klassischen" Chemieunterricht nicht so stark, als wenn direkt empirische Studien thematisiert würden. Als *lokale Theorie* wird daher aufgestellt, dass eine zweigliedrig geplante Unterrichtseinheit eine zielführende Methode zur Gewinnung von wissenschaftlichen Prüfkriterien darstellt.

15.3 Präsentation eines kontroversen Kontexts in choice^{2}reflect

Den praktischen Output des zweiten Mesozyklus bildet die Übertragung der Konzeption auf eine weitere Kontroverse. Durch eine Fragebogenerhebung mit Schüler:innen der Sekundarstufe I zum Konsum von Nahrungsergänzungsmitteln wurde die Kontroverse Nahrungsergänzungsmittel als relevantes Thema identifiziert.
Für die Phasen I und II sowie IV und V wurden Unterrichtsmaterialien im Kontext NEM entwickelt. Dabei wurde vor allem auf den Mineralstoff *Magnesium* zur Vermeidung von Muskelkrämpfen fokussiert, da dieser als am häufigsten supplementierter Nährstoff aus der Befragung hervortrat.
Die Phasen I und II wurden zwei Mal erprobt, die Phasen IV und V ein Mal im Rahmen der Gesamterprobung (zur Übersicht der Erprobungen siehe 8.1). Im Folgenden entwickelte Theorien sind daher als lokale Theorien mit beschränktem Geltungsbereich anzusehen. Sie sollten in zukünftiger Forschung weiter untersucht werden.

Die Einführung der Thematik erfolgte mithilfe von Beiträgen des sozialen Netzwerks *Instagram*. Die Präsentation der Problemfrage wurde durch eine Videodatei vorgenommen, die einen Gruppenchat simuliert. Es wird vermutet, dass der Einsatz dieser Gestaltungsmittel eine Identifikation der Lernenden mit der Problemfrage begünstigt. In den Wortbeiträgen der Lernenden in den Erprobungsstunden ist erkennbar, dass eine derartige Identifikation bei manchen Schüler:innen eingetreten ist. Sie begründen ihre Positionierungen mit eigenen Erfahrungen (siehe Kapitel 10.3.2), z. B. begründet ein Schüler seine Wahl von Magnesium-Tabletten als geeignetes Mittel zur Vermeidung von Muskelkrämpfen damit, dass er selbst gelegentlich diese Tabletten

einnehme.

Die ersten beiden Phasen verfolgen das Ziel, das zur Kontroverse bestehende Meinungsbild in der Klasse aufzudecken und mögliche Erklärungsmuster der Lernenden offenzulegen.

Das Meinungsbild stellte sich in beiden untersuchten Lerngruppen als sehr heterogen heraus (siehe Kapitel 10.3). Diesbezüglich wird die lokale Theorie aufgestellt, dass eine heterogene Meinungslage mit unterschiedlich elaborierten Erklärungsmustern optimale Voraussetzungen für den weiteren Verlauf von choice^{2}reflect bietet.

Begründet wird die Annahme darin, dass die Kontroversität der Problemstellung in dieser Weise auch für die Lernenden kenntlich gemacht wird; sie erfahren, inwiefern es unterschiedliche Meinungslagen zur Problemfrage gibt.

Möglicherweise erleben die Lernenden in dem Zusammenhang auch eine Unzufriedenheit mit ihrer eigenen Position und verspüren den Bedarf an empirisch gesicherten Fakten zur Klärung. Das hat sich in den Antworten der Schüler:innen gezeigt. Diese waren von Unsicherheiten geprägt (erkennbar in der Tabelle 10.3 auf der Seite 239). Zur Klärung schlugen die Lernenden z. B. vor, Ärzte zu konsultieren oder im Internet nach Informationen zu suchen (siehe Kapitel 10.3.4).

Die Lehrkraft kann in ihrer moderierenden Funktion verdeutlichen, dass weiterführende Informationen zur Klärung des Problems nötig sind. Mit diesem *Cliffhanger*, der ebenfalls für erhöhte Lernbereitschaft verantwortlich sein könnte, wird zur *Gewinnung wissenschaftlicher Prüfkriterien* (Phase III) übergeleitet.

Aufgelöst wird die Problematik erst mit Eintritt der Phase IV. Dort begegnen die Lernenden dem *Prüfschema*, dessen Gestaltungskriterien im nächsten Abschnitt behandelt werden. In der letzten Phase treffen die Lernenden eine eigene begründete Entscheidung. An dieser Stelle ist das *selbstständige* Treffen der Entscheidung besonders hervorzuheben, denn es erhöht die persönliche Betroffenheit der Lernenden und signalisiert, dass sie sich nicht auf „der Wissenschaft“ ausruhen können, sondern tatsächlich eine eigene faktenbasierte Meinung herausbilden müssen.

Diese Theorie steht im Einklang mit Jungkamps Ausführungen auf Seite 301, dass die Selbstständigkeit hervorzuheben von zentraler Bedeutung sei und dass Eigenaktivität und persönliche Auseinandersetzung zur Bildung des eigenen Standpunktes nötig seien.

15.4 Gestaltungsmerkmale eines Prüfschemas

Wie können Lernende dabei unterstützt werden, das methodische Wissen aus Phase III auf gesellschaftliche Kontroversen anzuwenden?

Die Beantwortung dieser Frage wurde im dritten Mesozyklus verfolgt (Kapitel 11. Auf der *praktischen* Seite hat der dritte Mesozyklus das Prüfschema als methodisches Instrument hervorgebracht und Erkenntnisse über die Eignung des Schemas generiert. Lehramts-Studierende, die an der Erprobung teilgenommen haben, können mithilfe des Schemas rationale Entscheidungen fällen. Auch der Einsatz in schulischem Kontext im Rahmen der Gesamterprobung hat Hinweise darauf geliefert, dass Schüler:innen der Sekundarstufe I das Schema (zumindest als Ablaufschema bei der Prüfung von Studien auf Wissenschaftlichkeit) verstehen können.

Welche Gestaltungskriterien wurden bei der Entwicklung des Prüfschemas beachtet? Wie wirken sich diese auf das Verständnis als intuitiv verwendbares methodisches Instrument aus?
Allgemeingültige Theoriensysteme und weitreichende Aussagen über die Gestaltung von Prüfschemata können in diesem Rahmen mit der vorhandenen Datenlage nicht generiert werden. Vielmehr geht es um die Sensibilisierung für Gestaltungsentscheidungen und deren Einflüsse auf den realen Lehr-Lern-Kontext.

Das Prüfschema wurde stark optisch vorstrukturiert. Das bedeutet, dass wichtige Start- und Knotenpunkte in großer Schrift hervorgehoben und mit einem Rahmen versehen wurden. Weiterhin wurde die im europäischen Raum verwendete Leserichtung von oben links nach unten rechts berücksichtigt. Der Einsatz des Trichter-Symbols und der Rohrleitungen verleiht dem Schema einen plastischen Eindruck: Bei der Betrachtung soll der Eindruck entstehen, dass *etwas* beim Durchlaufen des Schemas „nach unten fällt“ und dabei physikalischen Gesetzen folgt. Eine Reduzierung des *cognitive load* wird damit angestrebt.
Die Bedeutung der Rohrleitungen wurde bei Betrachtung der Erprobungsergebnisse deutlich: Die erste Lerngruppe „verirrte“ sich deutlich häufiger in den Rohrleitungen und folgte den Rohren wieder zurück zum Ursprung; während diese Probleme in der zweiten Lerngruppe mit dem optimierten Prüfschema nicht mehr auftraten. Die Rohre sollten also stets einen klar definierten Anfangs- und Endpunkt haben.

Weiterhin haben sich die *Stichwort-Fragen* als nützlich erwiesen. In der Erprobung mit Studierenden stellten sie die einzigen Ankerpunkte dar, an denen sich die Studierenden orientieren konnten, da zuvor keine Erarbeitung der Prüfkriterien stattfand.

Im Einsatz mit Schüler:innen können die Fragen nach Abschluss der Phase III zur Reaktivierung des erlernten Wissens über Prüfkriterien dienen. Eine zielgenaue Formulierung dieser Fragen ist daher unabdingbar. Durch das Re-Design (11.6) wurden sie bereits einmal verbessert und können in zukünftiger Forschung weiter ausgeschärft werden.

Nach Jungkamp (2021) unterstützen Prüfkriterien den schrittweisen Aufbau eines Wissenschaftsverständnisses im Sinne eines *Cluster Approaches*. Die Schüler:innen erlernen nach-und-nach die Denk- und Arbeitsweisen der (empirischen) Naturwissenschaften und etablieren ein einheitliches empirisch-wissenschaftliches Weltbild.

Bisherige Forschung konnte allerdings nicht aufzeigen, in welcher Weise der *Cluster Approach* im Unterricht umgesetzt werden kann. Diese Lücke vermag das Prüfschema zu schließen. Es gibt sowohl Lehrenden als auch Lernenden die Details der Erarbeitung vor. Das Schema sollte eine zentrale Rolle beim Durchführen der Konzeption choice^{2}reflect erhalten und für die Schüler:innen stets zugänglich sein. Dadurch können sie an jedem Punkt der Erarbeitung einen Blick auf das Prüfschema werfen und sehen dort die komplexen theoretischen Hintergründe der Bewertung empirischer Studien als Ablaufschema visualisiert.

Findet der Unterricht unter beschriebenen Voraussetzungen statt, fördert das Prüfschema vermutlich das schrittweise Etablieren eines einheitlichen empirisch-wissenschaftlichen Weltbildes und sorgt dafür, dass der Prüfkriterien-Pool nicht nur als „Sammlung von Anforderungen“ verstanden wird. Stattdessen entdecken Lehrende und Lernende die Systematik hinter dem Prüfen von alltäglichen Aussagen, wissenschaftlichen Hypothesen und empirischen Studien.

Damit wird die lokale Theorie aufgestellt, dass das Prüfschema einen Beitrag dazu leisten kann, den *Cluster-Approach* im Unterricht zu realisieren.

An dieser Stelle ist es wichtig, herauszustellen, dass der komplette Prozess nicht für *jede* alltägliche Aussage vollzogen werden kann. Es ist nicht praktikabel, in jeder persönlichen Konversation innezuhalten und Wissenschaftskriterien „auszupacken“. Auch bei der täglichen Rezeption von Nachrichten und Schlagzeilen können Lernende keine vollständige Bewertung im Sinne des Prüfschemas vornehmen. Haben sie die Hin-

tergründe der Wissenschaftskriterien jedoch nach einer Unterrichtseinheit von choice^2-reflect verstanden, so sollten die intuitiven Heuristiken der Lernenden um wissenschaftliche Denkmuster erweitert worden sein.
Bei der Entwicklung von choice^{2}reflect wurde ein großer Wert darauf gelegt, an intuitive Herangehensweisen anzuknüpfen und diese zu erweitern. Mit den erweiterten Denkmustern sollte sich im Alltag eine stärkere *kritische Grundhaltung* etablieren, wie sie beispielsweise auch in den Lehrplänen gefordert wird. Damit wird ein Bewusstsein für Wissenschaftlichkeit geschaffen, sodass täglich auftretende Konversationen oder Schlagzeilen in den Nachrichten mit einer gewissen Skepsis betrachtet werden.

16 Fazit und Abschlussreflexion

S9_2: Also die Studie an sich ist, glaube ich, wirklich wissenschaftlich und gut, aber die Schlussfolgerung der Reporterin ist komplett verdreht.

In gesellschaftlichen Kontroversen den Überblick zu behalten, ist eine herausfordernde Aufgabe. Vermeintliche Fakten begegnen uns von allen Seiten: In den sozialen Medien werden Beiträge zu allen möglichen Themen geteilt, weiterverbreitet und in manchen Fällen den Rezipienten aufgezwungen. Einige Printmedien versprechen, „alle Fakten" zu liefern und auch in persönlichen Gesprächen werden Informationen vermittelt.
Aus welcher Quelle diese Informationen stammen, ob sie wissenschaftlich abgesichert sind und inwiefern ein Studienergebnis überhaupt in einer aktuellen Diskussion verwendet werden darf, wird in vielen Fällen nicht deutlich.
Zweifelt ein Rezipient die Informationen an, erfolgt häufig der Verweis auf die Meinungsfreiheit: „Man wird doch noch seine Meinung sagen dürfen."
Allerdings sind in naturwissenschaftlichen Kontroversen keine Meinungen gefragt. Zum Treffen einer Entscheidung müssen empirisch abgesicherte Fakten betrachtet werden, das wurde im Laufe dieser Arbeit mehrfach herausgestellt.
Die von Jungkamp (2021) entwickelte Unterrichtskonzeption choice^{2}reflect adressiert die Unterscheidung zwischen unwissenschaftlichen Pseudo-Fakten und wissenschaftlich gültigen Informationen. Schülerinnen und Schüler lernen, anhand von Wissenschaftskriterien gesellschaftlich kontrovers diskutierte Themen zu bewerten.
Das in dieser Arbeit beschriebene Projekt umfasst die Weiterentwicklung der Konzeption nach dem Design-Based Research-Ansatz: Im Rahmen von drei groß angelegten Mesozyklen wurden Lernmaterialien zu den Wissenschaftskriterien entwickelt und erprobt, *Nahrungsergänzungsmittel* als passende Kontroverse herausgestellt und ein *Prüfschema* zur Anwendung der Wissenschaftskriterien entwickelt.
Mehrere Einsätze der Konzeption im realen Lehr-Lern-Kontext haben gezeigt, dass

die entwickelten Unterrichtsmaterialien praxistauglich sind und Lernende beim Treffen rationaler Entscheidungen in Kontroversen unterstützen.
Die Basis für die Generierung eines theoretischen Outputs haben qualitativ orientierte Analysen der Unterrichtseinsätze gelegt. So konnten Gestaltungskriterien für Lernmaterialien zu wissenschaftlichen Prüfkriterien formuliert werden.
Außerdem wurden für die schülernahe Kontroverse *Nahrungsergänzungsmittel* Handlungsanweisungen für eine passende Unterrichtseinbindung im Rahmen der Konzeptionsphasen I, II, IV und V generiert.
Der letzte Schritt der Theoriebildung hat Gestaltungsmerkmale für das Prüfschema hervorgebracht.
Im Rahmen dieses Projekts wurden Forschungsdesiderate erfüllt und die Unterrichtskonzeption choice^{2}reflect wurde abgerundet. Der „Materialien-Pool“ wurde erweitert, sodass Lehrkräfte aus dem Lernmaterial einzelne Module entnehmen und für den eigenen naturwissenschaftlichen Unterricht verwenden können.
Für jede Phase und für jedes Prüfkriterium bestehen nun Stundenverlaufspläne und ausgearbeitete Arbeitsblätter sowie Prüfkarten. Mit dem Prüfschema wurde die Handhabbarkeit des Prüfkriterien-Pools erhöht und eine klare Struktur für Lehrende und Lernende geschaffen, die die Vorgehensweise beim Prüfen von Aussagen und Untersuchungen auf Wissenschaftlichkeit verdeutlicht.

In weiterführender Forschung sollte untersucht werden, wie der Einsatz der Unterrichtskonzeption von Lehrkräften beurteilt wird. Da die Förderung von Bewertungskompetenz mithilfe von Wissenschaftskriterien nicht in dieser Form in den Lehrplänen der naturwissenschaftlichen Fächer verankert ist, wirft die Durchführung von choice^{2}reflect vermutlich neue Herausforderungen für die Lehrkräfte auf.
Diese Herausforderungen aufzudecken und zu analysieren, kann weiterführende Erkenntnisse und Handlungsanweisungen für den unterrichtlichen Einsatz liefern.

Das Eingangszitat dieses Abschnitts zeigt eine gut gelungene und rational getroffene Entscheidung: Der Lernende prüfte eine empirische Studie zu einem kontroversen Thema mithilfe des Prüfschemas auf Wissenschaftlichkeit und stellte fest, dass die Studie wissenschaftlichen Standards gerecht wird.
Er erkannte jedoch die in einer Schlagzeile fehlerhaft gezogene Schlussfolgerung „der Reporterin“.
Diese Art der Entscheidungsfindung in gesellschaftlichen Kontroversen ist wünschens-

wert und sollte in den „Alltag“ Einzug finden. Der Alltag wird hier bewusst akzentuiert: Selbstverständlich ist es nicht möglich, in jedem alltäglich auftretenden kontroversen Szenario alle Informationen mithilfe eines Prüfschemas auf sechs Wissenschaftskriterien zu prüfen.
Aber durch die Behandlung von Prüfkriterien und dem Prüfschema im naturwissenschaftlichen Unterricht der Sekundarstufe I sollte der Weg geebnet sein, Lernende für eine kritisch-rationale Haltung in Kontroversen zu sensibilisieren.

Teil VII

Literatur

Aagaard, N. K., Andersen, H., Vilstrup, H., Clausen, T., Jakobsen, J. & Dørup, I. (2005). Magnesium supplementation and muscle function in patients with alcoholic liver disease: A randomized, placebo-controlled trial. *Scandinavian Journal of Gastroenterology*, 40(8), S. 972–979.

Abd-El-Khalick, F., Bell, R. L. & Lederman, N. G. (1998). The nature of science and instructional practice: Making the unnatural natural. *Science Education*, 82(4), S. 417–436.

Ajzen, I. & Fishbein, M. (1969). The prediction of behavioral intentions in a choice situation. *Journal of experimental social psychology*, 5(4), S. 400–416.

Akker, J. van den, Gravemeijer, K., McKenney, S. & Nieveen, N. (2006). *Educational design research*. London, New York: Routledge.

Akobeng, A. K. (2005). Understanding randomised controlled trials. *Archives of Disease in Childhood*, 90(8), S. 840–844.

Alt, J. A. (1994). *Miteinander diskutieren. Eine Einführung in die Praxis vernünftiger Argumentation*. Frankfurt: Campus-Verlag.

American Museum of Natural History (2021). *What Is a Theory?* (Besucht am 08. 12. 2021).

Anderson, C. J. (2003). The psychology of doing nothing: Forms of decision avoidance result from reason and emotion. *Psychological Bulletin*, 129(1), S. 139–167.

Aristoteles (2018). *Die Rhetorik. Griechisch/Deutsch*. Hrsg. von G. Krapinger. Ditzingen: Reclam.

Arp, R., Barbone, S. & Bruce, M. (2018). *Bad Arguments: 100 of the Most Important Fallacies in Western Philosophy*. Wiley.

Bachem, M., Scheffler, K., Jastram, U. & Pfeiffer, E. (1986). Effektivität einer peroralen Magnesiumsubstitution bei Typ-I-Diabetikern mit nächtlichen Wadenkrämpfen. *Magnesium-Bulletin*, 8(3), S. 280–283.

Bakeman, R. & Gottman, J. M. (1997). Recording behavioral sequences. In: *Observing Interaction: An Introduction to Sequential Analysis*. 2. Aufl. Cambridge University Press, S. 38–55.

Banse, C. (2020). Entwicklung von Unterrichtsmaterialien zum wissenschaftlichen Prüfkriterium Reproduzierbarkeit. Masterarbeit. Münster.

Barke, H.-D., Harsch, G., Marohn, A. & Krees, S. (2015). *Chemiedidaktik kompakt*. 2. Auflage. Berlin/Heidelberg: Springer Spektrum.

Baron, J. (2000). Can we use human judgments to determine the discount rate? *Risk Analysis*, 20(6), S. 861–868.

Bartl, W. & Riss, P. (1982). Zur Pathophysiologie und Therapie des Magnesiummangels in der Schwangerschaft. *Zeitschrift für Geburtshilfe und Perinatalogie*, 186(6), S. 335–337.

Baumann, P. (2015). *Erkenntnistheorie*. Berlin/Heidelberg: Springer.

Bea, F. X. & Göbel, E. (2002). *Organisation: Theorie und Gestaltung*. Grundwissen der Ökonomik : Betriebswirtschaftslehre. Utb Fuer Wissenschaft.

Bell, R. L. & Lederman, N. G. (2003). Understandings of the Nature of Science and Decision Making on Science and Technology Based Issues. *Science Education*, 87(3), S. 352–377.

Ben-Ari, M. (2011). *Just a theory: Exploring the nature of science*. Prometheus Books.

Bereiter, C. & Scardamalia, M. (2003). Learning to work creatively with knowledge. *Powerful learning environments: Unravelling basic components and dimensions*, S. 55–68.

Berekoven, L., Eckert, W. & Ellenrieder, P. (2009). *Marktforschung - Methodische Grundlagen und praktische Anwendung*. 12. Wiesbaden: Gabler.

Bergman, R. G. & Danheiser, R. L. (2016). Reproduzierbarkeit in der chemischen Forschung. *Angewandte Chemie*, 128(41), S. 12736–12737.

Betsch, T., Funke, J. & Plessner, H. (2011). *Denken - Urteilen, Entscheiden, Problemlösen. Lesen, Hören, Lernen im Web.* Springer-Lehrbuch. Berlin/Heidelberg: Springer.

Betsch, T. & Haberstroh, S. (2014). *The Routines of Decision Making*. The Routines of Decision Making. Taylor & Francis.

Betsch, T., Haberstroh, S., Glöckner, A., Haar, T. & Fiedler, K. (2001). The effects of routine strength on adaptation and information search in recurrent decision making. *Organizational behavior and human decision processes*, 84(1), S. 23–53.

Bilandzic, H. & Trapp, B. (2000). Die Methode des lauten Denkens: Grundlagen des Verfahrens und die Anwendung bei der Untersuchung selektiver Fernsehnutzung bei Jugendlichen. In: Hrsg. von I. Paus-Haase & B. Schorb. *Qualitative Kinder- und Jugendmedienforschung*. München: KoPäd Verlag.

Bingle, W. H. & Gaskell, P. J. (1994). Scientific literacy for decisionmaking and the social construction of scientific knowledge. *Science Education*.

BLK (1998). *Gutachten zur Vorbereitung des Programms: Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts*. Techn. Ber. Bonn.

Bögeholz, S. (2006). Explizites Bewerten und Urteilen-Beispielkontext Streuobstwiese. *Praxis der Naturwissenschaften - Biologie in der Schule*, 55(1), S. 17.

Bögeholz, S. & Barkmann, J. (2005). Rational choice and beyond: Handlungsorientierende Kompetenzen für den Umgang mit faktischer und ethischer Komplexität. *Lehr-und Lernforschung in der Biologiedidaktik*, 2, S. 211–224.

Bögeholz, S., Hößle, C., Langlet, J., Sander, E. & Schlüter, K. (2004). Bewerten - Urteilen - Entscheiden im biologischen Kontext: Modelle in der Biologiedidaktik. *Zeitschrift für Didaktik der Naturwissenschaften*.

Bögeholz, S., Böhm, M., Eggert, S. & Barkmann, J. (2014). Education for Sustainable Development in German science education: Past - present - future. *Eurasia Journal of Mathematics, Science and Technology Education*, 10(4), S. 231–248.

Bögeholz, S., Eggert, S., Ziese, C. & Hasselhorn, M. (2017). Modeling and fostering decision-making competencies regarding challenging issues of sustainable development. In: *Competence Assessment in Education*. Springer, S. 263–284.

Bögeholz, S., Hößle, C., Höttecke, D. & Menthe, J. (2018). Bewertungskompetenz. In: *Theorien in der naturwissenschaftsdidaktischen Forschung*.

Bolkan, S. & Griffin, D. J. (2017). Students' use of cell phones in class for off-task behaviors: The indirect impact of instructors' teaching behaviors through boredom and students' attitudes. *Communication Education*, 66(3), S. 313–329.

Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human und Sozialwissenschaftler*. Berlin/Heidelberg.

Brosius, H.-B., Haas, A. & Koschel, F. (2008). *Methoden der empirischen Kommunikationsforschung*. Springer.

Brown, A. L. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *Journal of the Learning Sciences*, 2(2), S. 141–178.

Bühler, K. (1907). *Tatsachen und Probleme zu einer Psychologie der Denkvorgänge*. Engelmann.

Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices.* ERIC.

Campbell, D. T. & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. *Handbook of research of teaching*, S. 171–246.

Caspar, F. & Berger, T. (2007). Insight and Cognitive Psychology.

Cavanaugh, L. A., Bettman, J. R., Luce, M. F. & Payne, J. W. (2007). Appraising the appraisal-tendency framework. *Journal of Consumer Psychology*, 17(3), S. 169–173.

Chalmers, A. F., Bergemann, N., Bergemann, N., Altstötter-Gleich, C. & Altstötter-Gleich, C. (2007). *Wege der Wissenschaft: Einführung in die Wissenschaftstheorie*. Hrsg. von N. Bergemann & C. Altstötter-Gleich. 6. Berlin/Heidelberg: Springer.

Cobb, P., Confrey, J., Disessa, A., Lehrer, R. & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher*, 32(1), S. 9–13.

Collins, A. (1992). Toward a Design Science of Education. *New Directions in Educational Technology*, S. 15–22.

Collins, A., Joseph, D. & Bielaczyc, K. (2004). Design Research: Theoretical and Methodological Issues. *Journal of the Learning Sciences*, 13(1), S. 15–42.

Combe, A. & Gebhard, U. (2012). Verstehen im Unterricht. In: *Verstehen im Unterricht: Die Rolle von Phantasie und Erfahrung*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 7–14.

Cook, J., Ellerton, P. & Kinkead, D. (2018). Deconstructing climate misinformation to identify reasoning errors. *Environmental Research Letters*, 13(2), S. 24018.

Czerlinski, J., Gigerenzer, G. & Goldstein, D. G. (1999). How good are simple heuristics? *Simple Heuristics that make us smart*, S. 97–118.

DAZ (2007). Magnesium macht Muskeln munter. *Deutsche Apotheker Zeitung*, 22, 80f.

Dede, C. (2004). If design-based research is the answer, what is the question? A commentary on Collins, Joseph, and Bielaczyc; diSessa and Cobb; and Fishman, Marx, Blumenthal, Krajcik, and Soloway in the JLS special issue on design-based research. *The Journal of the Learning Sciences*, 13(1), S. 105–114.

Dellbrügge, B. (2020). *Choice2interact: interaktiv Lernen mit Tablets im Chemieunterricht*. Logos Verlag Berlin.

DFG (2017). *Stellungnahme Replizierbarkeit von Forschungsergebnissen*. URL: https://www.dfg.de/download/pdf/dfg%7B%5C_%7Dim%7B%5C_%7Dprofil/reden%7B%5C_%7Dstellungnahmen/2017/170425%7B%5C_%7Dstellungnahme%7B%5C_%7Dreplizierbarkeit%7B%5C_%7Dforschungsergebnisse%7B%5C_%7Dde.pdf.

Dittmer, A., Gebhard, U., Höttecke, D. & Menthe, J. (2016). Ethisches Bewerten im Naturwissenschaftlichen Unterricht: Theoretische Bezugspunkte. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), S. 97–108.

Döring, N. & Bortz, J. (2016). Forschungsmethoden und Evaluation.

Duden (2020a). *Hypothese*. URL: https://www.duden.de/node/69610/revision/69646 (besucht am 08. 12. 2020).

Duden (2020b). *These*. URL: https : / / www . duden . de / node / 182128 / revision / 182164 (besucht am 08. 12. 2020).

Duit, R. (1995). Zur Rolle der konstruktivistischen Sichtweise in der naturwissenschaftsdidaktischen Lehr- und Lernforschung. *Zeitschrift für Pädagogik*, 41(6), S. 905–923.

Edwards, W. (1954). The theory of decision making. *Psychological bulletin*, 51(4), S. 380.

Egbers, M. & Marohn, A. (2013). Konzeptwechseltexte - eine Textart zur Veränderung von Schülervorstellungen. *Chemkon*, 20(3), S. 119–126.

Eggert, S. (2008). Bewertungskompetenz für den Biologieunterricht. Diss. Niedersächsische Staats-und Universitätsbibliothek Göttingen.

Eggert, S. & Bögeholz, S. (2006). Göttinger Modell der Bewertungskompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, S. 177–197.

Eiser, J. R. (1990). *Social judgment.* Thomson Brooks/Cole Publishing Co.

Feierabend, T. & Eilks, I. (2010). Raising students' perception of the relevance of science teaching and promoting communication and evaluation capabilities using authentic and controversial socio-scientific issues in the Framework of climate change. *Science Education International*, 21(3), S. 176–196.

Fischer, F., Waibel, M. & Wecker, C. (2005). Nutzenorientierte Grundlagenforschung im Bildungsbereich. *Zeitschrift für Erziehungswissenschaft*, 8(3), S. 427–442.

Fisher, R. A. (1935). *The design of experiments.* Oxford, England: Oliver & Boyd, S. xi, 251–xi, 251.

Flach, T. (2021). Einsatzmöglichkeiten eines Prüfschemas zur Bewertung kontroverser Themen im naturwissenschaftlichen Unterricht im Rahmen der Unterrichtskonzeption choice2reflect. Masterarbeit. Münster.

Franke, N. (2002). *Realtheorie des Marketing: Gestalt und Erkenntnis*. Bd. 124. Mohr Siebeck.

Garrison, S. R., Korownyk, C. S., Kolber, M. R., Allan, G. M., Musini, V. M., Sekhon, R. K. & Dugré, N. (2020). Magnesium for skeletal muscle cramps. *Cochrane Database of Systematic Reviews*, 2020(9).

GfK-Pressemitteilung (2019). *Studie zeigt: E-Scooter, Freizeitspaß oder New Mobility?* URL: https://www.presseportal.de/pm/31692/4377120 (besucht am 10. 11. 2021).

Gigerenzer, G. & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, 62(1), S. 451–482.

Glaesser, J., Gott, R., Roberts, R. & Cooper, B. (2009). Underlying success in open-ended investigations in science: using qualitative comparative analysis to identify necessary and sufficient conditions. *Research in Science & Technological Education*, 27(1), S. 5–30.

Gropengießer, H. & Marohn, A. (2018). Schülervorstellungen und Conceptual Change. In: *Theorien in der naturwissenschaftsdidaktischen Forschung*. Hrsg. von D. Krüger, I. Parchmann & H. Schecker. Berlin/Heidelberg: Springer, S. 49–67.

Habel, P. (2019). *Telekom bei 5G im Plan*. URL: https://www.telekom.com/de/medien/medieninformationen/detail/telekom-bei-5g-im-plan--587232 (besucht am 10. 12. 2021).

Hagen-Schittkowski, C., Vorwerg, A., Habekost, A. & Maulbetsch, C. (2015). Zur Elektrochemie der Zitronenbatterie. *CHEMKON*, 22(2), S. 69–74.

Hammann, M. (2004). Kompetenzentwicklungsmodelle. *MNU, 57 (4)*, S. 196–203.

Hammar, M., Berg, G., Solheim, F. & Larsson, L. (1987). Calcium and magnesium status in pregnant women. A comparison between treatment with calcium and vitamin C in pregnant women with leg cramps. *International journal for vitamin and nutrition research*, 57(2), S. 179–183.

Han, S., Lerner, J. S. & Keltner, D. (2007). Feelings and consumer decision making: The appraisal-tendency framework. *Journal of Consumer Psychology*, 17(3), S. 158–168.

Häringer, E. (1981). Wadenkrämpfe in der Nacht: Fehlt Magnesium? *Ärtzliche Praxis*, 33, S. 2653–2654.

Heine, L. (2014). Introspektion. In: Hrsg. von S. Demirkaya, A. Feldmeier, N. Gültekin-Karakoç, C. Riemer & J. Settinieri. *Empirische Forschungsmethoden für Deutsch als Fremd-und Zweitsprache: Eine Einführung*. Stuttgart: UTB.

Heinemann, M., Willers, J., Bitterlich, N. & Hahn, A. (2015). Verwendung von Nahrungsergänzungsmitteln mit Vitaminen und Mineralstoffen – Ergebnisse einer deutschlandweiten Verbraucherbefragung. *Journal für Verbraucherschutz und Lebensmittelsicherheit*, 10(2), S. 131–142.

Heitmann, P. & Tiemann, R. (2011). Aspekte von Bewertungskompetenz im naturwissenschaftlichen Unterricht. *CHEMKON*, 18(3), S. 129–133.

Hoadley, C. M. (2004). Methodological alignment in design-based research. *Educational Psychologist*, 39(4), S. 203–212.

Hogan, K. (2002). Small groups' ecological reasoning while making an environmental management decision. *Journal of Research in Science Teaching*, 39(4), S. 341–368.

Hostenbach, J. & Walpuski, M. (2013). Untersuchung der Einflussfaktoren auf die Bewertungskompetenz im Fach Chemie. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, S. 129–157.

Hostenbach, J., Fischer, H. E., Kauertz, A., Mayer, J., Sumfleth, E. & Walpuski, M. (2011). Modellierung der Bewertungskompetenzen in den Naturwissenschaften zur Evaluation der Nationalen Bildungsstandards. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, S. 261–285.

Huber, F., Meyer, F. & Lenzen, M. (2014). Konzeption und Durchführung von Experimenten. In: *Grundlagen der Varianzanalyse: Konzeption - Durchführung - Auswertung*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 23–42.

Hugener, I., Pauli, C. & Reusser, K. (2006). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie Ünterrichtsqualität, Lernverhalten und mathematisches Verständnis". 3. Videoanalysen*. 15. Frankfurt, Main: GFPF u.a.

Hume, D. (2019). *An Enquiry Concerning Human Understanding*. London: Electric Book.

Jungkamp, F. (2021). Kontroversen mit Hilfe wissenschaftlicher Prüfkriterien bewerten: Die Unterrichtskonzeption choice2reflect. In: Hrsg. von A. Marohn. *Lernen in Naturwissenschaften*. Logos Berlin. Kap. Band 9.

Kabisch, M., Ruckes, C., Seibert-Grafe, M. & Blettner, M. (2011). Randomisierte kontrollierte Studien: Teil 17 der Serie zur Bewertung wissenschaftlicher Publikationen. *Deutsches Ärzteblatt*, 108(39), S. 663–668.

Kahneman, D. (2011). *Thinking, Fast and Slow.* Hrsg. von A. Lane. London, UK: Penguin Books Ltd.

Kahneman, D., Slovic, P. & Tversky, A., (Hrsg.) (1982). *Judgment under Uncertainty*. Cambridge University Press.

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), S. 263–291.

Klafki, W. (1994). *Konturen eines neuen Allgemeinbildungskonzepts*.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. J. (2003). Zur Entwicklung nationaler Bildungsstandards: eine Expertise. In: Hrsg. von Bundesministerium für Bildung und Forschung (BMBF). *Bildungsreform*. Bd. 1. Berlin.

KLP (2013). *Kernlehrplan für die Gesamtschule – Sekundarstufe I in Nordrhein-Westfalen - Naturwissenschaften*. 2. Aufl. Düsseldorf: Ministerium für Schule und Weiterbildung des Landes NRW.

KLP (2014). *Kernlehrplan für die Sekundarstufe II – Gymnasium/Gesamtschule in Nordrhein-Westfalen - Chemie*. 1. Aufl. Düsseldorf: Ministerium für Schule und Weiterbildung des Landes NRW.

KLP (2019). *Kernlehrplan für die Sekundarstufe I - Gymnasium in Nordrhein-Westfalen - Chemie*. 1. Aufl. Düsseldorf: Ministerium für Schule und Weiterbildung des Landes NRW.

KMK (2005a). *Beschlüsse der Kultusministerkonferenz - Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss*. Techn. Ber. Luchterhand.

KMK (2005b). Beschlüsse der Kultusministerkonferenz Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss.

KMK (2005c). Beschlüsse der Kultusministerkonferenz Bildungsstandards im Fach Physik für den Mittleren Schulabschluss.

Koch, J., Gebhardt, P. & Riedmüller, F. (2016). *Marktforschung*. 7. überarb. Berlin, Boston: De Gruyter Oldenbourg.

Koller, H.-C. (1999). *Bildung und Widerstreit: zur Struktur biographischer Bildungsprozesse in der (Post-) Moderne*. Fink.

Kolstø, S. D., Bungum, B., Arnesen, E., Isnes, A., Kristensen, T., Mathiassen, K., Mestad, I., Quale, A., Tonning, A. S. V. & Ulvik, M. (2006). Science students' critical examination of scientific information related to socioscientific issues. *Science Education*, 90(4), S. 632–655.

Konrad, K. (2010). Lautes Denken. In: Hrsg. von G. Mey & K. Mruck. *Handbuch qualitative Forschung in der Psychologie*. Springer, S. 476–490.

Kornmeier, M. (2007). *Wissenschaftstheorie und wissenschaftliches Arbeiten*.

Kötter, M. & Hammann, M. (2018). *Epistemische Kompetenz: Wissenschaftsreflexion im naturwissenschaftlichen Unterricht*. Hrsg. von M. Hammann & M. Lindner. Innsbruck: StudienVerlag, S. 69–84.

Kralisch, C. (2017). Entwicklung von Unterrichtsmaterial zum wissenschaftlichen Prüfkriterium „Falsifizierbarkeit" im Rahmen der Unterrichtskonzeption choice^{2}reflect. Masterarbeit. Münster.

Kralisch, C. & Marohn, A. (o.D.). Wissenschaft contra Fake News - Kontroversen prüfen mithilfe von Wissenschaftskriterien. *MNU, MINT-Unterricht vs. Fake-News* ().

Krüger, M. (2010). Das Lernszenario VideoLern: Selbstgesteuertes und kooperatives Lernen mit Vorlesungsaufzeichnungen: Eine Design-Based-Research Studie, S. 1–271.

Krzossa, T. (2020). *Vodafone erhöht 5G-Ziele. Ausbau schneller als geplant*. URL: https://5g.nrw/5g-ausbau-von-vodafone-schneller-als-geplant/ (besucht am 10. 12. 2021).

Kuckartz, U. (2018). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung.* 4. Auflage. Grundlagentexte Methoden. Weinheim; Basel: Beltz Juventa, 240 S.

Kulgemeyer, C. & Starauschek, E. (2014). Analyse der Verständlichkeit naturwissenschaftlicher Fachtexte. In: Hrsg. von D. Krüger, I. Parchmann & H. Schecker. *Methoden in der naturwissenschaftsdidaktischen Forschung*. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 241–253.

Kümmerer, F. & Schmid-Johannsen, J. (Mai 2020). *Heinsberg-Studie zu Coronavirus - Falsche Rechnung mindert Aussagekraft*. URL: https://www.tagesschau.de/investigativ/swr/heinsberg-studie-103.html (besucht am 27. 10. 2021).

Landswerk, G. (2021). *Veganer haben mehr sportliche Ausdauer und genauso viel Kraft wie Fleischesser - zeigt eine Studie.* (Besucht am 22. 11. 2021).

Langer, A. & Rogowski, W. (2009). Deskriptive Entscheidungstheorie. de. In: Hrsg. von M. Schwaiger & A. Meyer. *Theorien und Methoden der Betriebswirtschaft: Handbuch für Wissenschaftler und Studierende*. 1. Aufl. München: Verlag Franz Vahlen, S. 179–193.

Laux, H., Gillenkirch, R. M. & Schenk-Mathes, H. Y. (2018). *Entscheidungstheorie*. Berlin/Heidelberg: Springer.

Leisen, J. (2011). Kompetenzorientiert unterrichten. *Unterricht Physik*, 123/124, S. 4–10.

Lerner, J. S. & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition and Emotion*, 14(4), S. 473–493.

Lerner, J. S., Li, Y., Valdesolo, P. & Kassam, K. S. (2015). Emotion and Decision Making. *Annual Review of Psychology*, 66(1), S. 799–823.

Lerner, J. S., Small, D. A. & Loewenstein, G. (2004). Heart strings and purse strings: Carryover effects of emotions on economic decisions. *Psychological Science*, 15(5), S. 337–341.

Lerner, J. S. & Tiedens, L. Z. (2006). Portrait of the angry decision maker: How appraisal tendencies shape Anger's influence on cognition. *Journal of Behavioral Decision Making*, 19(2), S. 115–137.

Lingnau, V. (1995). Kritischer Rationalismus und Betriebswirtschaftslehre. *Wirtschaftswissenschaftliches Studium*, 24(3), S. 124–129.

Lotz, M. (2016). Beschreibung und Beurteilung von Unterricht durch Videoanalysen. In: *Kognitive Aktivierung im Leseunterricht der Grundschule: Eine Videostudie zur Gestaltung und Qualität von Leseübungen im ersten Schuljahr*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 149–159.

Luther-Kirner, B. (2009). Ethik im Biologieunterricht. Eine Kritik didaktischer Konzeptionen. *Bioethik in der Schule. Grundlagen und Gestaltungsformen*. Hrsg. von U. Manz, S. 39–52.

Macbeth, D. (2000). On an actual apparatus for conceptual change. *Science Education*, 84(2), S. 228–264.

Mad, P., Felder-Puig, R. & Gartlehner, G. (2008). Randomisiert kontrollierte Studien. *Wiener Medizinische Wochenschrift*, 158(7), S. 234–239.

Maier, J. & Richter, T. (2014). Verstehen multipler Texte zu kontroversen wissenschaftlichen Themen: Die Rolle der epistemischen Validierung. *Unterrichtswissenschaft*, 42(1), S. 24–38.

Mansour, H. (2021). *Nahrungsergänzungsmittel - Deutschland - 2020*. Techn. Ber. Mintel Group Ltd.

Mäntele, W. (2021). Naturwissenschaftliche Arbeitsweise - naturwissenschaftliche Sichtweise. In: *Elektrosmog und Ökoboom: Ein naturwissenschaftlicher Blick auf populäres Halbwissen*. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 1–7.

Marks, R. & Eilks, I. (2009). Promoting scientific literacy using a sociocritical and problem-oriented approach to chemistry teaching: Concept, examples, experiences. *International Journal of Environmental and Science Education*, 4(3), S. 231–245.

Marks, R., Stuckey, M. & Eilks, I. (2014). Die gesellschaftliche Dimension naturwissenschaftlich-technischer Sachfragen. 134, S. 19–28.

Marohn, A. (2014). Interaktionsboxen. *Praxis der Naturwissenschaften, Chemie in der Schule*, 6, S. 63.

Marohn, A. & Egbers, M. (2011). Vorstellungen verändern - Lernmaterialien zum Thema Verdampfen im Rahmen der Unterrichtskonzeption choice2learn. *Praxis der Naturwissenschaften, Chemie in der Schule*, 60(3), S. 5–9.

Marohn, A. & Jungkamp, F. (2016). choice2reflect - Kontrovers diskutierte Themen mit Hilfe wissenschaftlicher Prüfkriterien reflektieren - Beispiel Homöopathie. *PdN - ChiS*, 5(65), S. 38–43.

Max Rubner-Institut & Bundesforschungsinstitut für Ernährung und Lebensmittel (2008a). Nationale Verzehrs Studie II, Teil 1. *Report*, 1, S. 173.

Max Rubner-Institut & Bundesforschungsinstitut für Ernährung und Lebensmittel (2008b). Nationale Verzehrs Studie II, Teil 2. *Report*, 2, S. 280.

McKenney, S. & Reeves, T. C. (2012). *Conducting educational design research*. Routledge.

Meffert, H. (1992). *Marketingforschung und Käuferverhalten*. 2. Aufl. Wiesbaden: Gabler.

Mittelsten Scheid, N. & Hößle, C. (2008). Bewerten im Biologieunterricht: Niveaus von Bewertungskompetenz. *Erkenntnisweg Biologiedidaktik*, 6, S. 87–104.

Mittelstraß, J. (1997). Vom Nutzen des Irrtums in der Wissenschaft. *Naturwissenschaften*, 84(7), S. 291–299.

Mossig, I. (2012). *Stichproben, Stichprobenauswahlverfahren und Berechnung des minimal erforderlichen Stichprobenumfangs*. Beiträge zur Wirtschaftsgeographie und Regionalentwicklung 1-2012. Bremen.

Naeem, S. B., Bhatti, R. & Khan, A. (2021). An exploration of how fake news is taking over social media and putting public health at risk. *Health Information and Libraries Journal*, 38(2), S. 143–149.

Nguyen-Kim, M. T. (2021). *Die kleinste gemeinsame Wirklichkeit*. München: Droemer Verlag.

Nienhüser, W. (1989). Die praktische Nutzung theoretischer Erkenntnisse. *Betriebswirtschaftslehre, Stuttgart*.

Nienhüser, W. & Magnus, M. (2003). Die wissenschaftliche Bearbeitung personalwirtschaftlicher Problemstellungen. Eine Einführung. *Essener Beiträge zur Personalforschung*, 2(2), S. 1–32.

Pedretti, E. & Hodson, D. (1995). From rhetoric to action: Implementing STS education through action research. *Journal of Research in Science Teaching*, 32(5), S. 463–485.

Pernkopf, E. (2006). *Unerwartetes erwarten: zur Rolle des Experimentierens in naturwissenschaftlicher Forschung*. Bd. 413. Königshausen & Neumann.

Pfister, H.-R., Jungermann, H. & Fischer, K. (2017). *Die Psychologie der Entscheidung*. Berlin/Heidelberg: Springer.

Popper, K. R. (1989). *Logik der Forschung*. Bd. 9. JCB Mohr Tübingen.

Prediger, S., Link, M., Hinz, R., Hußmann, S., Thiele, J. & Ralle, B. (2012). Lehr-Lernprozesse initiieren und erforschen–fachdidaktische Entwicklungsforschung im Dortmunder Modell. *Mathematischer und Naturwissenschaftlicher Unterricht*, 65(8), S. 452–457.

Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E. & Pekrun, R., (Hrsg.) (2007). *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*. 1. Aufl. Münster u.a.: Waxmann.

Qian, Y. & Li, L. (Mai 2017). Student Off-Task Electronic Multitasking Predictors: Scale Development and Validation. *Journal of the Scholarship of Teaching and Learning*, 17(2), S. 53–73.

Rädiker, S. & Kuckartz, U. (2019). *Analyse qualitativer Daten mit MAXQDA*. Berlin/Heidelberg: Springer.

Raffée, H. (1974). *Grundprobleme der Betriebswirtschaftslehre*. Vandenhoeck & Ruprecht.

Rehaag, R., Tils, G., Röder, B., Ulbig, E., Kurzenhäuser-Carstens, S., Lohmann, M. & Böl, G.-F. (2013). *Zielgruppengerechte Risikokommunikation zum Thema Nahrungsmittel*. Techn. Ber., S. 1–135.

Reinert, A., Rohrmann, S., Becker, N. & Linseisen, J. (Apr. 2007). Lifestyle and diet in people using dietary supplements. *European Journal of Nutrition*, 46(3), S. 165–173.

Reinmann, G. (2005). Innovation ohne Forschung? Ein Plädoyer für den Design-Based Research-Ansatz in der Lehr-Lernforschung; Innovation without Research? Arguments for Design-Based Research in Educational Research. *Unterrichtswissenschaft*, 33(1), S. 52–69.

Reinmann, G. (2007). Innovationskrise in der Bildungsforschung: Von Interessenkämpfen und ungenutzten Chancen einer Hard-to-do-Science. *Der Nutzen wird vertagt.... Bildungswissenschaften im Spannungsfeld zwischen wissenschaftlicher Profilbildung und praktischem Mehrwert*, S. 198–220.

Reitschert, K., Langlet, J., Hössle, C., Mittelstein Scheid, N. & Schlüter, K. (2007). Dimensionen Ethischer Urteilskompetenz - Dimensionierung und Niveaukonkretisierung. *MNU*.

Riss, P., Bartl, W. & Jelincic, D. (1983). Zur Klinik und Therapie von Wadenkrämpfen in der Schwangerschaft TT - Clinical and Therapeutic Aspects of Muscle Cramps during Pregnancy. DE. *Geburtshilfe Frauenheilkd*, 43(05), S. 329–331.

Rohrbach-Lochner, F. (2019). Design-Based Research zur Weiterentwicklung der chemiedidaktischen Lehrerausbildung zu Schülervorstellungen: Entwicklung und Evaluation eines an Forschendem Lernen orientierten Seminarkonzepts. In: Hrsg. von A. Marohn. *Lernen in Naturwissenschaften*. Logos Berlin. Kap. Band 3.

Rohrbach-Lochner, F. & Marohn, A. (2018). How research-based learning can increase teacher students' knowledge and abilities: a design-based research project in the context of pupils'(mis) conceptions in science. *Research in Subject-matter Teaching and Learning*, 1, S. 35–50.

Rosenberg, D., Rautenstrauch, H., Pöhls, C., Böttger, S., Busker, M., Fittschen, U. & Jansen, W. (2018). Lemon battery again – a clarification. *CHEMKON*, 25(8), S. 343–349.

Rosenberg, D., Wachholz, M., Busker, M. & Jansen, W. (2019). Organic Batteries with ascorbic acid and the some different lemon battery. *CHEMKON*, 26(3), S. 119–123.

Ross, K. G., Lussier, J. W. & Klein, G. (2005). From the recognition primed decision model to training. *The routines of decision making*, S. 327–332.

Rost, J., Lauströer, A. & Raack, N. (2003). Kompetenzmodelle einer Bildung für Nachhaltigkeit. *Praxis der Naturwissenschaften - Chemie in der Schule*, 52(8), S. 10–15.

Rott, L. & Marohn, A. (2016). Inklusiven Unterricht entwickeln und erproben – Eine Verbindung von Theorie und Praxis im Rahmen von Design-Based Research. *Zeitschrift für Inklusion*(, 4).

Rundgren, S. & Rundgren, C. (2010). SEE-SEP: From a separate to a holistic view of socioscientific issues. In: *Asia-Pacific Forum on Science Learning & Teaching*. Bd. 11. 1.

Sadler, T. D. & Donnelly, L. A. (2006). Socioscientific argumentation: The effects of content knowledge and morality. *International journal of science education*, 28(12), S. 1463–1488.

Sadler, T. D. & Zeidler, D. L. (2005). The significance of content knowledge for informal reasoning regarding socioscientific issues: Applying genetics knowledge to genetic engineering issues. *Science Education*, 89(1), S. 71–93.

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching*, 41(5), S. 513–536.

Schaumburg, H. (2015). Chancen und Risiken digitaler Medien in der Schule. *Individuell fördern mit digitalen Medien. Chancen, Risiken, Erfolgsfaktoren.*, S. 20–94.

Scheffler, H. (2000). Stichprobenbildung und Datenerhebung. *Marktforschung*, 2. Hrsg. von A. Herrmann & C. Homburg, S. 59–77.

Schlick, M. (1931). Die Kausalität in der gegenwärtigen Physik. *Naturwissenschaften*, 19(7), S. 145–162.

Schnell, R., Hill, P. & Esser, E. (2005). *Methoden der empirischen Sozialforschung*. 7. Aufl. München, Wien: Oldenbourg.

Schulz, K. F. & Grimes, D. A. (2002). Generation of allocation sequences in randomised trials: chance, not choice. *The Lancet*, 359(9305), S. 515–519.

Schwab, S., Heier, M., Schneider, A., Fischer, B., Huth, C., Peters, A. & Thorand, B. (2014). The use of dietary supplements among older persons in Southern Germany - Results from the kora-age study. *Journal of Nutrition, Health and Aging*, 18(5), S. 510–519.

Shamos, M. H. (1995). The myth of scientific literacy.

Smith, E. R. & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4(2), S. 108–131.

Statista (2021). *Welche Arten von Nahrungsergänzungsmitteln haben Sie in den letzten 12 Monaten eingenommen?* URL: https://de.statista.com/prognosen/722012/umfrage-zum-konsum-von-nahrungsergaenzungsmitteln-in-deutschland (besucht am 27. 10. 2021).

Statistisches Bundesamt (Destatis) (2021). *Nahrungsergänzungsmittel boomen: Produktion im Jahr 2020 um 11 % höher als im Vorjahr*. URL: https://www.destatis.de/DE/Presse/Pressemitteilungen/Zahl-der-Woche/2021/PD21%7B%5C_%7D23%7B%5C_%7Dp002.html (besucht am 26. 10. 2021).

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Brookings Institution Press.

Streeck, H. et al. (2020). Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. *medRxiv*.

Strike, K. A. & Posner, G. J. (1982). Conceptual change and science teaching. *European Journal of Science Education*, 4(3), S. 231–240.

Tetens, H. (2018). Reproduzierbarkeit. In: *Enzyklopädie der Philosophie und Wissenschaftstheorie*. Band 7. Springer Berlin Heidelberg.

Tobias, V. (2010). Newton'sche Mechanik im Anfangsunterricht: Die Wirksamkeit einer Einführung über die zweidimensionale Dynamik auf das Lehren und Lernen.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79(4), S. 281–299.

Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), S. 453–458.

Tytler, R. (2001). Dimensions of evidence, the public understanding of science and science education. *International Journal of Science Education*, 23(8), S. 815–832.

Universität Bonn (2020). *Ergebnisse der Heinsberg-Studie veröffentlicht*. URL: https://www.uni-bonn.de/neues/111-2020 (besucht am 17. 11. 2020).

Vodafone (2020). *Vodafone-Netzausbau: Der Stand der Dinge*. URL: https://www.vodafone.de/business/featured/technologie/fortschritt-ausbau-vodafone-netz/ (besucht am 10. 12. 2021).

Vollmer, G. (1995). Wir irren uns empor. Zum Tode des Philosophen Karl Raimund Popper. *Skeptiker*, 8(1), S. 4–6.

Waismann, F. (1930). Logische analyse des Wahrscheinlichkeitsbegriffs. *Erkenntnis*, S. 228–248.

Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen - eine umstrittene Selbstverständlichkeit. In: Hrsg. von F. E. Weinert. *Leistungsmessungen in Schulen*. Weinheim: Beltz, S. 17–32.

Weller, E., Bachert, P., Meinck, H. M., Friedmann, B., Bärtsch, P. & Mairbäurl, H. (1998). Lack of effect of oral Mg-supplementation on Mg in serum, blood cells, and calf muscle. *Medicine and Science in Sports and Exercise*, 30(11), S. 1584–1591.

Welling, S., Averbeck, I., Stolpmann, B. E., Karbautzki, L., Appelt, R., Schwalbe, C. & Kammerl, R. (2014). Paducation. Evaluation eines Modellversuchs mit Tablets am Hamburger Kurt-Körber-Gymnasium. 2, S. 2017.

Westermann, R. (2000). *Wissenschaftstheorie und Experimentalmethodik*. Hogrefe Verlag für Psychologie.

Wilhelm, T. & Hopf, M. (2014). Design-Forschung. In: *Methoden in der naturwissenschaftsdidaktischen Forschung*. Hrsg. von D. Krüger, I. Parchmann & H. Schecker. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 31–42.

Wilson, M. & Sloane, K. (2000). From Principles to Practice: An Embedded Assessment System. *Applied Measurement in Education*, 13(2), S. 181–208.

Winko, S. (2015). Zur Plausibilität als Beurteilungskriterium literaturwissenschaftlicher Interpretationen. In: Hrsg. von A. Albrecht & E. Al. *Theorien, Methoden und Praktiken des Interpretierens*. Berlin, Boston, S. 483–511.

Wirth, U. (1995). Abduktion und ihre Anwendungen. *Zeitschrift für Semiotik*, 17(3-4), S. 405–424.

Woods, J. & Walton, D. (1977). Post Hoc, Ergo Propter Hoc. *The Review of Metaphysics*, 30(4), S. 569–593.

Woods, J. & Walton, D. N. (1982). *Argument, the logic of the fallacies*. McGraw-Hill Ryerson Ltd.

Yang, F.-Y. (2004). Exploring high school students' use of theory and evidence in an everyday context: the role of scientific thinking in environmental science decision-making. *International Journal of Science Education*, 26(11), S. 1345–1364.

Zeidler, D. L., Walker, K. A., Ackett, W. A. & Simmons, M. L. (2002). Tangled Up in Views: Beliefs in the Nature of Science and Responses to Socioscientific Dilemmas. *Science Education*, 86(3), S. 343–367.

Teil VIII

Anhang

A Theoretischer Hintergrund

Abbildung A.0.1: *Tabellarische Übersicht der Definitionen und Konzeptualisierungen von Bewertungskompetenz in verschiedenen schulischen Lehrplänen, bezugnehmend auf das Kapitel 3.4.2.*

	Gymnasium S1			Gymnasium S2			Gesamt	Realschule			Bildungsstandards mittl. Schulabschluss			Bildungsstandards Abitur		
	Bio	Chemie	Physik	Bio	Chemie	Physik	NaWi	Bio	Chemie	Physik	Bio	Chemie	Physik	Bio	Chemie	Physik
Die Fähigkeit, ...	auf die Fähigkeit, in Problemsituationen, in denen es mehrere denkbare Lösungen ohne ein klares Richtig oder Falsch gibt, sachlich fundiert und wertebasiert zu begründeten Entscheidungen zu kommen.			(ethisch) überlegt zu urteilen				überlegt zu urteilen.			Biologische/Chemische/Physikalische Sachverhalte in verschiedenen Kontexten zu erkennen und bewerten.			Die Bewertungskompetenz der Lernenden zeigt sich in der Kenntnis von fachlichen und überfachlichen Perspektiven und Bewertungsverfahren und in der Fähigkeit, diese zu nutzen, um Aussagen bzw. Daten anhand verschiedener Kriterien zu beurteilen, sich dazu begründet Meinungen zu bilden, Entscheidungen auch auf ethischer Grundlage zu treffen und Entscheidungsprozesse und deren Folgen zu reflektieren.		
Kriteriengeleitetes Bewerten?	- Handlungsmöglichkeiten zu entwickeln und auf der Grundlage von Kriterien gegeneinander abzuwägen - Bewertungskriterien und Handlungsoptionen benennen. - kriteriengeleitet eine Entscheidung für eine Handlungsoption treffen - in einer Bewertungssituation relevante chemische und naturwissenschaftlich-technische Sachverhalte und Zusammenhänge identifizieren, fehlende Informationen beschaffen sowie ggf. gesellschaftliche Bezüge beschreiben. - Bewertungskriterien auch unter Berücksichtigung der Energiespeicherung festlegen			- Kriterien und Handlungsmöglichkeiten sorgfältig zusammenzutragen und gegeneinander abzuwägen. - bei der Bewertung von Sachverhalten in nat. Zusammenhängen fachliche, gesellschaftliche und moralische Bewertungskriterien angeben, - Entscheidungsmöglichkeiten kriteriengeleitet abwägen, gewichten und einen begründeten Standpunkt beziehen - Bewertungskriterien angeben und begründet gewichten.				- Kriterien und Handlungsmöglichkeiten sorgfältig zusammenzutragen und gegeneinander abzuwägen. - in einfachen Zusammenhängen eigene Bewertungen und Entscheidungen unter Verwendung biologischen Wissens begründen (B1). - bei gegensätzlichen Ansichten Sachverhalte nach vorgegebenen Kriterien und vorliegenden Fakten beurteilen. - für Entscheidungen in Zusammenhängen Bewertungskriterien angeben und begründet gewichten.			- Kriterien für Bewertungen liefern Grundsätze einer nachhaltigen Entwicklung sowie zwei grundlegende ethische Denktraditionen. (Würde des Menschen, Wohlergehen des Menschen) Anforderungsbereich I -Bewerten			- Aussagen bzw. Daten anhand verschiedener Kriterien zu beurteilen, sich dazu begründet Meinungen zu bilden - Punkt: Kriteriengeleitet Meinungen bilden und Entscheidungen treffen		
Rationale Entscheidungen	- Faktenlage einschließlich der Interessen der Handelnden und Betroffenen sorgfältig zu analysieren - Entscheidungen zu finden, deren Tragweite zu reflektieren - Bewertungen und Entscheidungen begründen. Chemie: Handlungsoptionen durch Gewichten und Abwägen von Kriterien und nach Abschätzung der Folgen für die Natur, das Individuum und die Gesellschaft auswählen.			- rationale und begründete Entscheidungen zu treffen und dafür zielführend Position zu beziehen. Es ist jedoch auch notwendig, die Chancen für Problemlösungen einschätzen zu können und zu erkennen, wo nat. Erkenntnis an ihre Grenzen stößt. Gesamtschule: - Auf dieser Grundlage ist es möglich, Entscheidungen zu finden und dafür zielführend zu argumentieren und Position zu beziehen.				-[...] auf dieser Grundlage ist es möglich, Entscheidungen zu finden			Nur Chemie: Bei der Betrachtung gesellschaftsrelevanter Themen aus unterschiedlichen Perspektiven erkennen die Lernenden, dass Problemlösungen von Wertentscheidungen abhängig sind. Sie sollen befähigt sein, Argumente auf ihren sachlichen und ideologischen Anteil zu prüfen und Entscheidungen sachgerecht, selbstbestimmt und verantwortungsbewusst zu treffen.			Keine explizite Erwähnung		
Ethische Aspekte	diesbezüglich sind die Kenntnis und Berücksichtigung von Bewertungsmaßstäben bedeutsam, nach denen Interessen und Folgen naturwissenschaftlich-technischer Forschung und Entwicklung beurteilt werden können			- die Kenntnis und Berücksichtigung von normativen und ethischen Maßstäben bedeutsam, nach denen Interessen und Folgen naturwissenschaftlicher Forschung beurteilt werden können. - in bekannten Zusammenhängen ethische Konflikte bei Auseinandersetzungen mit Fragestellungen sowie mögliche Lösungen darstellen Biologie: Konfliktsituationen erkennen und bei Entscheidungen ethische Maßstäbe sowie Auswirkungen eigenen und fremden Handelns auf Natur, Gesellschaft und Gesundheit berücksichtigen.				- Kenntnis und Berücksichtigung von normativen und ethischen Maßstäben bedeutsam, nach denen Interessen und Folgen naturwissenschaftlicher Forschung beurteilt werden können. - Wertvorstellungen, Regeln und Vorschriften in Zusammenhängen hinterfragen und begründen. Konfliktsituationen erkennen und bei Entscheidungen ethische Maßstäbe sowie Auswirkungen eigenen und fremden Handelns auf Natur, Gesellschaft und Gesundheit berücksichtigen.			Nur Biologie: - ethischen Urteilsbildung sind Themen, die das verantwortungsbewusste Verhaltendes Menschen gegenüber sich selbst und anderen Personen sowie gegenüber der Umwelt betreffen. Beispiele dafür sind die Forschung an Embryonen, die gentechnische Veränderung von Lebewesen, Massentierhaltung und Eingriffe in Ökosysteme.			- Entscheidungen auch auf ethischer Grundlage zu treffen und Entscheidungsprozesse und deren Folgen zu reflektieren - Punkt: Entscheidungsprozesse und Folgen reflektieren Physik: Dazu tragen sie relevante physikalische, aber auch nicht physikalische (z. B. ökonomische, ökologische, soziale, politische oder ethische) Kriterien zusammen, sammeln geeignete Belege und wägen sie unter Berücksichtigung von Normen, Werten und Interessen gegeneinander ab.		

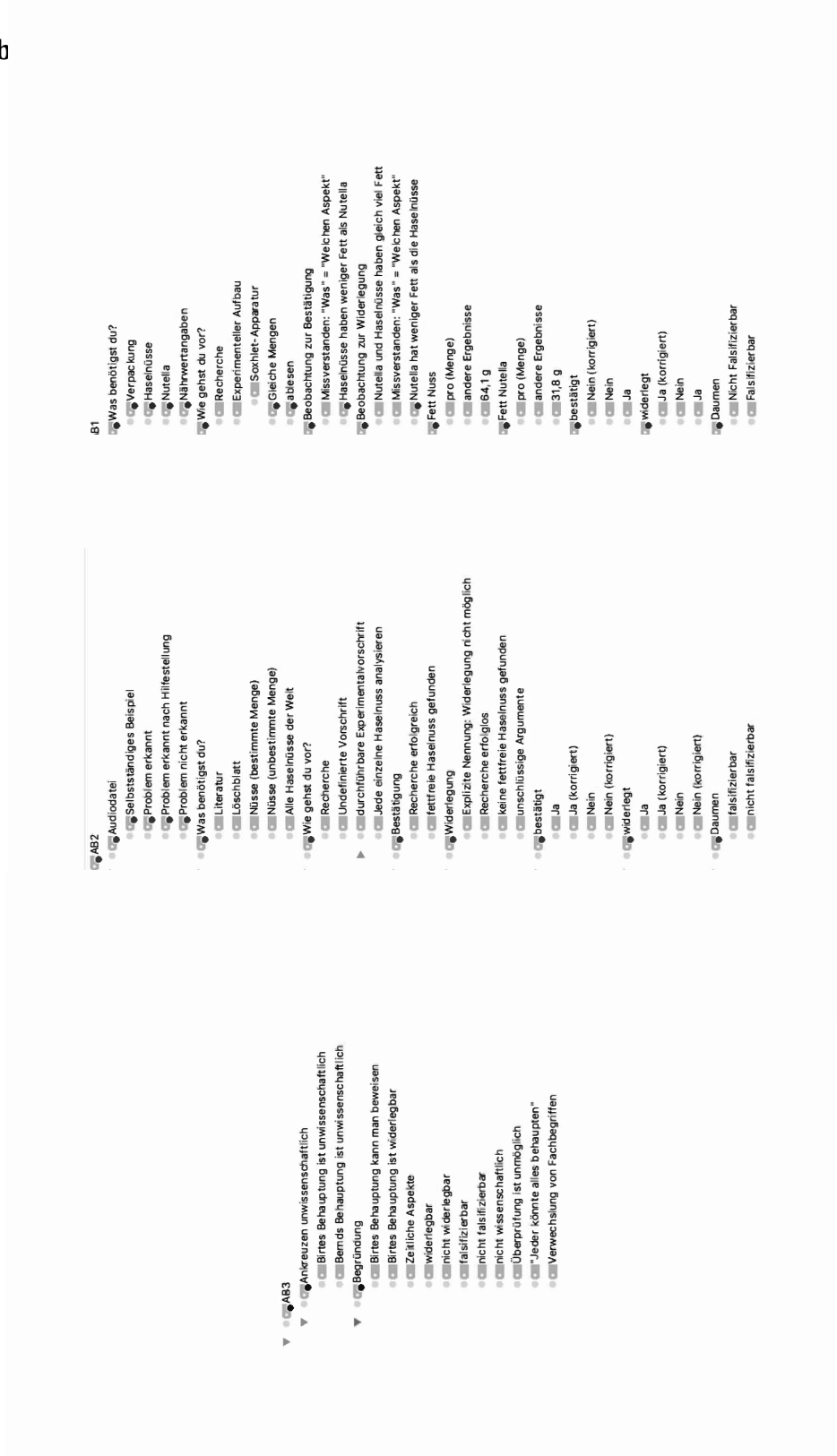
Argumentation	- auf der Grundlage Erkenntnisse und naturwissenschaftlicher Denkweisen faktenbasiert, rational und schlüssig argumentieren sowie zu Beiträgen anderer respektvolle, konstruktiv-kritische Rückmeldungen geben. - zielführend zu argumentieren und Positionen darzustellen - Fähigkeit zum rationalen, faktenbasierten Argumentieren bei der Darstellung eigener Überlegungen, der Diskussion und Reflexion von Ideen und Untersuchungsergebnissen sowie divergierender Positionen. - eigene Aussagen fachlich sinnvoll begründen, faktenbasierte Gründe von intuitiven Meinungen unterscheiden sowie bei Unklarheiten sachlich nachfragen. - begründet für eine Handlung entscheiden Bewertungen und Entscheidungen argumentativ vertreten und reflektieren	- bewährte Regeln der fachlichen Argumentation einzuhalten. - Aussagen und Behauptungen mit sachlich fundierten und überzeugenden Argumenten begründen bzw. kritisieren. - sich mit anderen über Sachverhalte kritisch-konstruktiv austauschen und dabei Behauptungen oder Beurteilungen durch Argumente belegen bzw. widerlegen. - Gesamtschule: - Entscheidungen zu finden und dafür zielführend zu argumentieren und Position zu beziehen. - Regeln der fachlichen Argumentation einzuhalten - bei gegensätzlichen Ansichten Sachverhalte nach vorgegebenen Kriterien und vorliegenden Fakten beurteilen. -in Situationen mit mehreren Entscheidungsmöglichkeiten kriteriengeleitet Argumente abwägen, einen Standpunkt beziehen und diesen gegenüber anderen Positionen begründet vertreten.	- zielführend zu argumentieren und Position zu beziehen. - in Situationen mit mehreren Entscheidungsmöglichkeiten kriteriengeleitet Argumente abwägen, einen Standpunkt beziehen und diesen gegenüber anderen Positionen begründet vertreten. Bewährte Regeln der fachlichen Argumentation einzuhalten.	- In den Anforderungsbereichen Kommunikation tritt Argumentieren auf	- Bei der Betrachtung gesellschaftsrelevanter Themen aus unterschiedlichen Perspektiven erkennen die Lernenden, dass Problemlösungen von Wertentscheidungen abhängig sind. Sie sollen befähigt sein, Argumente auf ihren sachlichen und ideologischen Anteil zu prüfen und Entscheidungen sachgerecht, selbstbestimmt und verantwortungsbewusst zu treffen. Anforderungsbereiche Bewertung Physik: Keine Erwähnung	- Nat. zu argumentieren und damit am gesellschaftlichen Diskurs teilhaben zu können - Kommunikationskompetenz - Dem kriterien- und theoriegeleiteten Argumentieren und dem Strukturieren fachwissenschaftlicher Erkenntnisse kommt dabei eine besondere Bedeutung zu. - Im Bereich der Bewertungskompetenz zeigt sich das erhöhte Anforderungsniveau darin, dass mehr und komplexere Argumente mit Belegen zur Bewertung naturwissenschaftlicher Sachverhalte herangezogen werden. - argumentieren wissenschaftlich zu biologischen Sachverhalten kriterien- und evidenzbasiert sowie situationsgerecht.
Gesellschaftliche Teilhabe	- fehlende Informationen beschaffen sowie ggf. gesellschaftliche Bezüge beschreiben. Chemie: SuS in die Lage, Phänomene der Lebenswelt (…) zu erklären, zu bewerten, Entscheidungen zu treffen, Urteile zu fällen	Für eine gesellschaftliche Teilhabe ist daher eine nat. Grundbildung unverzichtbar.			- Naturwissenschaftliche Bildung ermöglicht dem Individuum eine aktive Teilhabe an gesellschaftlicher Kommunikation und Meinungsbildung über technische Entwicklung und naturwissenschaftliche Forschung und ist deshalb wesentlicher Bestandteil von Allgemeinbildung. - Dies ist die Grundlage für ein gesundheitsbewusstes und umweltverträgliches Handeln sowohl in individueller als auch in gesellschaftlicher Verantwortung. Physik: Grundlage für die Auseinandersetzung der jungen Menschen mit nat. Themen und ihren gesellschaftlichen Zusammenhängen gelegt. Zudem leistet er einen Beitrag zu anderen Fächern und zur Vorbereitung auf technische Berufe bzw. weiterführende Bildungsgänge und ermöglicht damit ein anschlussfähiges Orientierungswissen	- Nicht explizit - Naturwissenschaftliche Kompetenz bietet Orientierung in der durch Naturwissenschaften und Technik geprägten Lebenswelt, eröffnet Perspektiven für die berufliche Orientierung und schafft Grundlagen für selbstgesteuertes, lebenslanges, globales und soziales Lernen. - Das Erkennen, Einordnen, Bewerten und Berücksichtigen möglicher Folgen für ökologische, ökonomische und soziale Systeme ist für eine verantwortungsvolle gesellschaftliche Teilhabe notwendig und erfordert naturwissenschaftliche Kompetenz. Physik: - Ihre rationale und analytische Sichtweise, Exaktheit der Sprache und planvollen, strukturierten Herangehensweisen haben eine zentrale Bedeutung in einer Vielzahl von Berufsfeldern und für die aktive Teilhabe an gesellschaftlicher Kommunikation. - Das Verständnis dieser Art der Kommunikation und die Fähigkeit, sie mitzugestalten, ermöglichen die selbstbestimmte Teilhabe an wissenschaftlichen und gesellschaftlich relevanten Diskussionen.
Naturwissenschaftl. Grundbildung	- Beitrag zum Bildungsziel einer vertieften nat. Grundbildung - bietet die Grundlage für fundierte Urteile in Entscheidungsprozessen über erwünschte oder unerwünschte Entwicklungen. - Phänomene erfahrbar zu machen, die Sprache und Geschichte der Naturwissenschaften zu verstehen, ihre Erkenntnisse	Keine Erwähnung	Siehe Gymnasium Sek I		- Naturwissenschaftliche Bildung siehe oben. - Ziel naturwissenschaftlicher Grundbildung ist es, Phänomene erfahrbar zu machen, die Sprache und Historie der Naturwissenschaften zu verstehen, ihre Ergebnisse zu kommunizieren sowie sich mit ihren spezifischen Methoden der	Keine Erwähnung

	zu kommunizieren sowie sich mit ihren spezifischen Methoden der Erkenntnisgewinnung und deren Grenzen auseinanderzusetzen. - Typische theorie- und hypothesengeleitete Denk- und Arbeitsweisen ermöglichen eine analytische und rationale Betrachtung der Welt. - Naturwissenschaftliche Bildung ermöglicht eine aktive Teilhabe an gesellschaftlicher Kommunikation und Meinungsbildung über technische Entwicklungen und nat. Forschung und ist deshalb wesentlicher Bestandteil von Allgemeinbildung.										Erkenntnisgewinnung und deren Grenzen auseinanderzusetzen. - Dazu gehört das theorie- und hypothesengeleitete nat. Arbeiten, das eine analytische und rationale Betrachtung der Welt ermöglicht. Darüber hinaus bietet nat. Grundbildung eine Orientierung für naturwissenschaftlich-technische Berufsfelder und schafft Grundlagen für anschlussfähiges berufsbezogenes Lernen.					
Hinweis auf **Fake-News, Falschmeldungen, Verschwörungen?**	nein	nein	nein	nein	nein	nein	nein	nein	nein	nein	nein	nein	nein	nein	nein	nein
Hinweis auf **Unwissenschaft-lichkeit?**	nein	nein	nein	nein	nein	nein	nein	nein	nein	nein	nein	nein	nein	Biologie und Physik: Keine Erwähnung. - Im Bereich der Bewertungskompetenz ist es wichtig, sich nicht darauf zu beschränken, Fakten zu vergleichen, sondern Sachverhalte und Informationen fachlich zu beurteilen und ggf. ethisch zu bewerten. - Quellen hinsichtlich ihrer Qualität beurteilt. - Die Unterscheidung von wissenschaftlichen und nicht-wissenschaftlichen Aussagen erfordert Kenntnisse formaler und inhaltlicher Kriterien zur Prüfung der Glaubwürdigkeit und zur Beurteilung des Einflusses von Werten, Normen und Interessen.		

B Design-Experiment

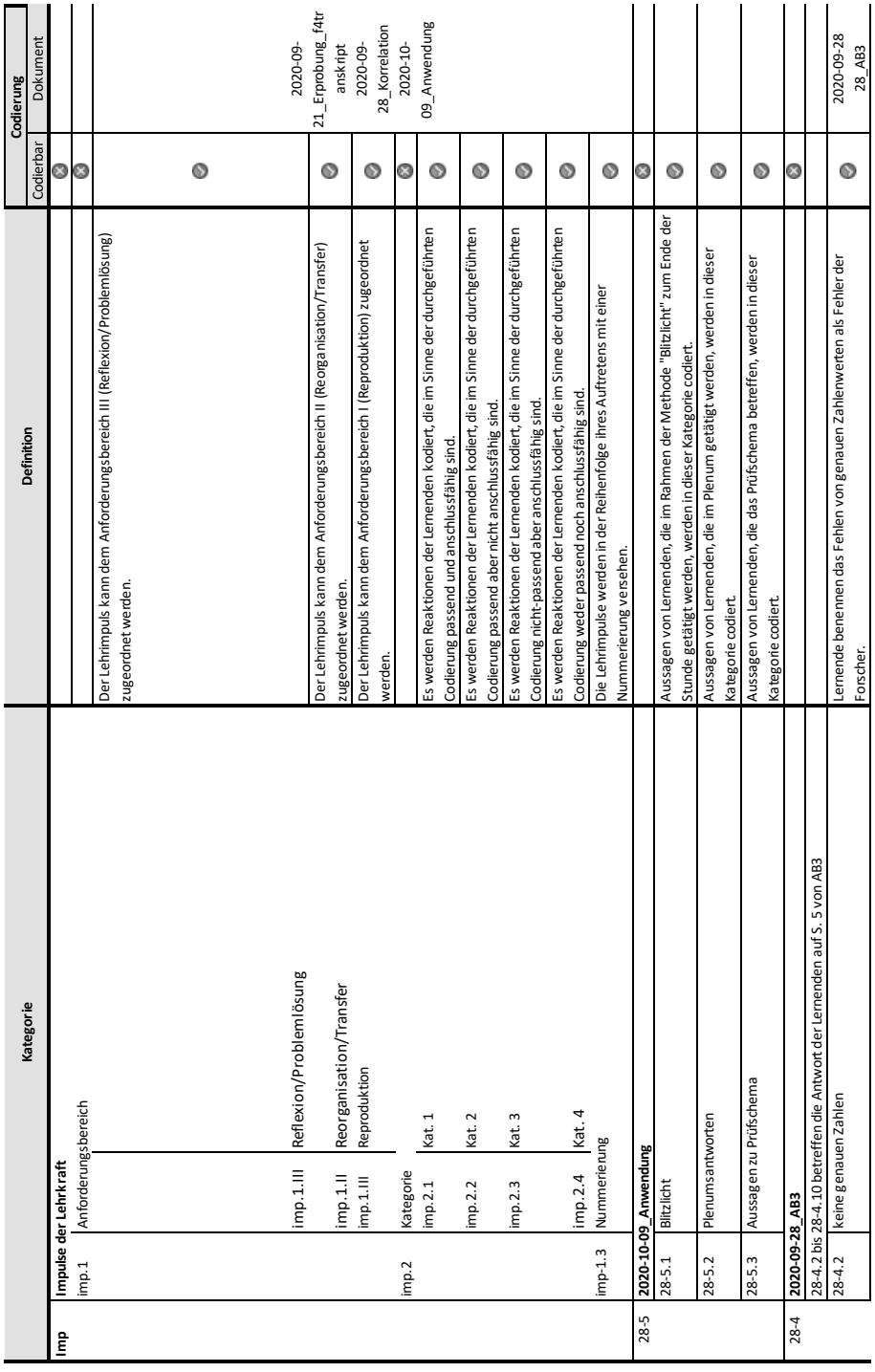
B.1 Kategoriensysteme

Ab *keit.*


B1
Was benötigst du?
Verpackung
Haselnüsse
Nutella
Nährwertangaben
Wie gehst du vor?
Recherche
Experimenteller Aufbau
Soxhlet-Apparatur
Gleiche Mengen
ableiten
Beobachtung zur Bestätigung
Missverstanden: "Was" = "Welchen Aspekt"
Haselnüsse haben weniger Fett als Nutella
Beobachtung zur Widerlegung
Nutella und Haselnüsse haben gleich viel Fett
Missverstanden: "Was" = "Welchen Aspekt"
Nutella hat weniger Fett als die Haselnüsse
Fett Nuss
pro (Menge)
andere Ergebnisse
64,1 g
Fett Nutella
pro (Menge)
andere Ergebnisse
31,8 g
bestätigt
Nein (korrigiert)
Nein
Ja
widerlegt
Ja (korrigiert)
Nein
Ja
Daumen
Nicht Falsifizierbar
Falsifizierbar
AB2
Audiodatei
Selbstständiges Beispiel
Problem erkannt
Problem erkannt nach Hilfestellung
Problem nicht erkannt
Was benötigst du?
Literatur
Löschblatt
Nüsse (bestimmte Menge)
Nüsse (unbestimmte Menge)
Alle Haselnüsse der Welt
Wie gehst du vor?
Recherche
Undefinierte Vorschrift
durchführbare Experimentalvorschrift
Jede einzelne Haselnuss analysieren
Bestätigung
Recherche erfolgreich
fettfreie Haselnuss gefunden
Widerlegung
Explizite Nennung: Widerlegung nicht möglich
Recherche erfolglos
keine fettfreie Haselnuss gefunden
unschlüssige Argumente
bestätigt
Ja
Ja (korrigiert)
Nein
Nein (korrigiert)
widerlegt
Ja
Ja (korrigiert)
Nein
Nein (korrigiert)
Daumen
falsifizierbar
nicht falsifizierbar
AB3
Anbreuzen unwissenschaftlich
Birtes Behauptung ist unwissenschaftlich
Bernds Behauptung ist unwissenschaftlich
Begründung
Birtes Behauptung kann man beweisen
Birtes Behauptung ist widerlegbar
Zeitliche Aspekte
widerlegbar
nicht widerlegbar
falsifizierbar
nicht falsifizierbar
nicht wissenschaftlich
Überprüfung ist unmöglich
"Jeder könnte alles behaupten"
Verwechslung von Fachbegriffen

Aussagen_Gr2
Shake
Richtig zugeordnet mit richtiger Begründung
Richtig zugeordnet mit falscher Begründung
Richtig zugeordnet ohne Begründung
Falsch zugeordnet
Gott
Richtig zugeordnet mit richtiger Begründung
Richtig zugeordnet mit falscher Begründung
Richtig zugeordnet ohne Begründung
Falsch zugeordnet
Vitamin C
Richtig zugeordnet mit richtiger Begründung
Richtig zugeordnet mit falscher Begründung
Richtig zugeordnet ohne Begründung
Falsch zugeordnet
Aluminium
Richtig zugeordnet mit richtiger Begründung
Richtig zugeordnet mit falscher Begründung
Richtig zugeordnet ohne Begründung
Falsch zugeordnet
Wasser
Richtig zugeordnet mit richtiger Begründung
Richtig zugeordnet mit falscher Begründung
Richtig zugeordnet ohne Begründung
Falsch zugeordnet
ussagen_Gr1
Zweistein
Richtig zugeordnet mit richtiger Begründung
Richtig zugeordnet mit falscher Begründung
Richtig zugeordnet ohne Begründung
Falsch zugeordnet
Trampel
Richtig zugeordnet mit richtiger Begründung
Richtig zugeordnet mit falscher Begründung
Richtig zugeordnet ohne Begründung
Falsch zugeordnet
Kriminell
Richtig zugeordnet mit richtiger Begründung
Richtig zugeordnet mit falscher Begründung
Richtig zugeordnet ohne Begründung
Falsch zugeordnet
Muskelaufbau
Richtig zugeordnet mit richtiger Begründung
Richtig zugeordnet mit falscher Begründung
Richtig zugeordnet ohne Begründung
Falsch zugeordnet
Bravo
Richtig zugeordnet mit richtiger Begründung
Richtig zugeordnet mit falscher Begründung
Richtig zugeordnet ohne Begründung
Falsch zugeordnet
AB4
Checkliste
nicht beweisbar
Konjunktiv
Gedanken
Existenzaussagen
Wahrscheinlichkeitsaussagen
Unendlichkeit
Schwammigkeit
wichtig
oft
Rückbezug
Verwechslung mit richtig/falsch

Abbildung B.1.2: *Vollständiges Kategoriensystem zur Codierung der zweiten Doppelstunde der Gesamterprobung.*



Kategorie					Definition	Codierung	
						Codierbar	Dokument
Imp	**Impulse der Lehrkraft**					⊗	
imp.1	Anforderungsbereich					⊗	
					Der Lehrimpuls kann dem Anforderungsbereich III (Reflexion/Problemlösung) zugeordnet werden.		
		imp.1.III	Reflexion/Problemlösung			⊘	2020-09-21_Erprobung_f4transkript
					Der Lehrimpuls kann dem Anforderungsbereich II (Reorganisation/Transfer) zugeordnet werden.		
		imp.1.II	Reorganisation/Transfer			⊘	2020-09-28_Korrelation
		imp.1.III	Reproduktion		Der Lehrimpuls kann dem Anforderungsbereich I (Reproduktion) zugeordnet werden.		
	imp.2	Kategorie				⊗	2020-10-09_Anwendung
		imp.2.1	Kat. 1		Es werden Reaktionen der Lernenden kodiert, die im Sinne der durchgeführten Codierung passend und anschlussfähig sind.	⊘	
		imp.2.2	Kat. 2		Es werden Reaktionen der Lernenden kodiert, die im Sinne der durchgeführten Codierung passend aber nicht anschlussfähig sind.	⊘	
		imp.2.3	Kat. 3		Es werden Reaktionen der Lernenden kodiert, die im Sinne der durchgeführten Codierung nicht-passend aber anschlussfähig sind.	⊘	
		imp.2.4	Kat. 4		Es werden Reaktionen der Lernenden kodiert, die im Sinne der durchgeführten Codierung weder passend noch anschlussfähig sind.	⊘	
	imp-1.3	Nummerierung			Die Lehrimpulse werden in der Reihenfolge ihres Auftretens mit einer Nummerierung versehen.	⊘	
28-5	**2020-10-09_Anwendung**					⊗	
	28-5.1	Blitzlicht			Aussagen von Lernenden, die im Rahmen der Methode "Blitzlicht" zum Ende der Stunde getätigt werden, werden in dieser Kategorie codiert.	⊘	
	28-5.2	Plenumsantworten			Aussagen von Lernenden, die im Plenum getätigt werden, werden in dieser Kategorie codiert.	⊘	
	28-5.3	Aussagen zu Prüfschema			Aussagen von Lernenden, die das Prüfschema betreffen, werden in dieser Kategorie codiert.	⊘	
28-4	**2020-09-28_AB3**					⊗	
	28-4.2 bis 28-4.10 betreffen die Antwort der Lernenden auf S. 5 von AB3						
	28-4.2	Keine genauen Zahlen			Lernende benennen das Fehlen von genauen Zahlenwerten als Fehler der Forscher.	⊘	2020-09-28_AB3

	28-4.3	Antwort nicht zielführend		Lernende notieren eine Antwort, die nicht zur Aufgabe passt.	✓	2020-09-28 28_AB3
	28-4.4	Zusammenhang, aber nicht Ursache		Lernende benennen als Fehler der Forscher, dass zwischen der App und Akkulaufzeit ein Zusammenhang besteht, aber keine Ursache.	✓	2020-09-28 28_AB3
	28-4.5	Kausalität		Lernende verwenden den Fachbegriff "Kausalität" in ihrer Erklärung.	✓	2020-09-28 28_AB3
	28-4.6	dritter Faktor muss beachtet werden		Lernende benennen das Nichtbeachten weiterer "dritter" Faktoren als Fehler der Forscher.	✓	2020-09-28 28_AB3
	28-4.7	Faktor Andere Apps		Lernende erklären, dass die Forscher nicht beachtet haben, dass auch andere Apps Einfluss auf die Akkulaufzeit haben können.	✓	2020-09-28 28_AB3
	28-4.8	Faktor Alter des Smartphones nicht beachtet		Lernende erklären, dass die Forscher nicht beachtet haben, dass auch das Alter des Smartphones Einfluss auf die Akkulaufzeit haben könnte.	✓	2020-09-28 28_AB3
	28-4.9	Faktor Handynutzung nicht beachtet		Lernende erklären, dass die Forscher nicht beachtet haben, dass auch die Dauder der Handynutzung Einfluss auf die Akkulaufzeit haben könnte.	✓	2020-09-28 28_AB3
	28-4.10	Plenumsantwort		Lernende notieren die Antwort, die von der Lehrkraft aufgeschrieben wurde: - Korrelation = Zusammenhang -> Und - Kausalität = Ursache -> Weil, desto - Schein-Kausalität hergestellt, da der 3. Faktor nicht beachtet wurde	✓	2020-09-28 28_AB3
28-3	**2020-09-28_tabelle (AB2)**				✗	
	28-3.eff	Effektives Arbeiten an Tabelle		wird an Zeiträumen codiert, in denen die Lernenden offensichtlich konzentriert an der Tabelle in der App Numbers arbeiten.	✓	2020-09-28_tabelle
	28-3.ausf	Ausfüllen des Fragebogens		wird codiert, wenn die Lernenden den im Arbeitsblatt dargestellten Fragebogen mit eigenen Daten ausfüllen, statt ihn nur zu betrachten.	✓	gruppenvideos
	28-3.1	keine Anmerkungen ergänzt		Lernende ergänzen rechts neben der Tabelle keine Anmerkungen.	✓	2020-09-28_tabelle
	28-3.2	Anmerkungen ergänzt		Lernende ergänzen rechts neben der Tabelle die Anmerkungen, die die Lehrkraft im Plenum notiert.	✓	2020-09-28_tabelle
	28-3.3	nicht korrekt			✗	2020-09-28_tabelle
		28-3.3.1	>4 Werte falsch	In der Tabelle sind mehr als 4 Werte nicht korrekt eingetragen.	✓	2020-09-28_tabelle
		28-3.3.2	4 Werte falsch	In der Tabelle sind mehr 4 Werte nicht korrekt eingetragen.	✓	2020-09-28_tabelle
		28-3.3.3	3 Werte falsch	In der Tabelle sind mehr 3 Werte nicht korrekt eingetragen.	✓	2020-09-28_tabelle
		28-3.3.4	2 Werte falsch	In der Tabelle sind mehr 2 Werte nicht korrekt eingetragen.	✓	2020-09-28_tabelle
		28-3.3.5	1 Wert falsch	In der Tabelle sind mehr 1 Werte nicht korrekt eingetragen.	✓	2020-09-28_tabelle

	28-3.4	Korrekturen			⊗	2020-09-28_tabelle
		28-3.4.1	keine Korrekturen	Es sind **keine** Korrekturen (Durchstreichen, Schwärzen, etc.) in der Tabelle erkennbar.	✓	2020-09-28_tabelle
		28-3.4.2	> 6 Korrekturen	Es sind mehr als 6 Korrekturen (Durchstreichen, Schwärzen, etc.) in der Tabelle erkennbar.	✓	2020-09-28_tabelle
		28-3.4.3	5-6 Korrekturen	Es sind 5-6 Korrekturen (Durchstreichen, Schwärzen, etc.) in der Tabelle erkennbar.	✓	2020-09-28_tabelle
		28-3.4.4	3-4 Korrekturen	Es sind 3-4 Korrekturen (Durchstreichen, Schwärzen, etc.) in der Tabelle erkennbar.	✓	2020-09-28_tabelle
		28-3.4.5	1-2 Korrekturen	Es sind 1-2 Korrekturen (Durchstreichen, Schwärzen, etc.) in der Tabelle erkennbar.	✓	2020-09-28_tabelle
	28-3.5	korrekt		Alle Eintragungen in der Tabelle sind korrekt. 12-4-5-9-4-9-13-4-6-10-10/11-3	✓	2020-09-28_tabelle
28-2	**2020-09-28_AB1**				⊗	
	28-2.4	Arbeitsauftrag			⊗	
		28-2.4.1	Plenumsantwort	In dieser Kategorie werden alle Aussagen codiert, die Lernende in einer Plenumsphase verbal äußern.	✓	
		28-2.4.2	Intervention durch Eva	Die betreuende Lehrkraft Eva Kolbeck greift in das Unterrichtsgeschehen ein.	✓	
		28-2.4.3	Spontanreaktion	Die Lernenden zeigen beim Öffnen des AB oder beim Lesen der Texte eine unmittelbare Reaktion (Lachen, Kommentieren, besondere Ausrufe).	✓	
		28-2.4.4	Fragen von Lernenden		✓	
	28-2.5	Alter des Smartphones		wird codiert, wenn im Textfeld auf AB1 das Alter des Smartphones als wichtiger Faktor aufgeführt wird.	✓	
	28-2.6	Nutzungsdauer		wird codiert, wenn im Textfeld auf AB1 die Nutzungsdauer als wichtiger Faktor aufgeführt wird.	✓	
	28-2.7	notiert: Plenumsantwort		wird codiert, wenn auf dem AB nur die Anmerkungen vermerkt sind, die von der Lehrkraft im Plenum aufgeschrieben wurden: "- es kommt auch darauf an, wie alt das Handy ist - es kommt auch darauf an, wie viel man es nutzt - vielleicht liegt es an einer anderen App"	✓	2020-09-28 28_AB1
28-1	**2020-09-28_1**				⊗	
	28-1.2	Plenumsphase			⊗	
		28-1.2.1	Konflikt/Korrektur eigener Lösung	Lernende erkennen in der Plenumsphase zur Wiederholungs-Aufgabe, dass die eigene Lösung nicht korrekt ist.	✓	
		28-1.2.2	Prüfschema	Lernende beziehen sich in der Wiederholungsphase auf das Prüfschema oder Teile davon, wie z. B. die Schatztruhe.	✓	
		28-1.2.3	Fragen von Lernenden	Lernende äußern in der Wiederholungsphase Fragen.	✓	
	28-1.3	Aufgabe Basis, Randomisierung			⊗	

		28-1.3.1	Schätzung	Lernende notieren, dass Randomisierung mit Schätzungen zu tun hat.	✓	2020-09-28 28_Wdh
		28-1.3.2	Dunkelziffer	Lernende notieren, dass Randomisierung mit der Dunkelziffer zusammenhängt.	✓	2020-09-28 28_Wdh
		28-1.3.3	Prüfen von Studien	Lernende notieren, dass Randomisierung zum Prüfen von Studien verwendet wird.	✓	2020-09-28 28_Wdh
		28-1.3.4	ohne Merkmale zu beachten	Lernende notieren, dass Randomisierung mit dem Beachten von Merkmalen zusammenhängt.	✓	2020-09-28 28_Wdh
		28-1.3.5	gerechte Auswertung	Lernende notieren, dass Randomisierung zu einer gerechten Auswertung führt .	✓	2020-09-28 28_Wdh
		28-1.3.6	befragen	Lernende verwenden in der Erklärung das Wort "befragen" oder "Befragung".	✓	2020-09-28 28_Wdh
		28-1.3.7	Plenumsantwort	Lernende vermerken auf dem AB die Antworten, die von der Lehrkraft in der Plenumsphase aufgeschrieben wurden: - Leute werden zufällig gewählt - Stichprobe von verschiedenen Menschen für eine Studie	✓	2020-09-28 28_Wdh
		28-1.3.8	Stichprobe	Lernende verwenden in der Erklärung das Wort "Stichprobe".	✓	2020-09-28 28_Wdh
		28-1.3.9	Studie	Lernende verwenden in der Erklärung das Wort "Studie".	✓	2020-09-28 28_Wdh
		28-1.3.10	zufällig	Lernende verwenden in der Erklärung das Wort "zufällig".	✓	2020-09-28 28_Wdh
	28-1.4	Aufgabe Stern1, repräsentativ			✗	
		28-1.4.1	passend	Lernende verwenden das Wort "passend".	✓	2020-09-28 28_Wdh
		28-1.4.2	nicht verallgemeinerbar	Lernende erklären repräsentativ als nicht verallgemeinerbar.	✓	2020-09-28 28_Wdh
		28-1.4.3	Plenumsantwort	Lernende notieren die Antwort, die von der Lehrkraft aufgeschrieben wurde: "Repräsentativ bedeutet aussagekräftig für eine bestimmte Gruppe von Menschen"	✓	2020-09-28 28_Wdh
		28-1.4.4	zutreffend auf alle Menschen	Lernende erklären repräsentativ dadurch, dass etwas (Stichprobe etc.) zutreffend für alle Menschen oder eine ganze Gruppe ist.	✓	2020-09-28 28_Wdh
		28-1.4.5	zufällig	Lernende verwenden das Wort "zufällig".	✓	2020-09-28 28_Wdh
		28-1.4.6	[aussagekräftig] für …	Lernende erklären repräsentativ als aussagekräfig für etwas.	✓	2020-09-28 28_Wdh
		28-1.4.7	Stichprobe	Lernende verwenden das Wort "Stichprobe".	✓	2020-09-28 28_Wdh
		28-1.4.8	aussagekräftig	Lernende erklären repräsentativ als aussagekräftig. (abzugrenzen von: aussagekräftig für eine bestimmte Gruppe)	✓	2020-09-28 28_Wdh
21-4	**2020-09-21_AB4 (BILD)**				✗	

	21-4.1	Spontanreaktionen			Die Lernenden zeigen bei Betrachtung des AB4 (Zeitungsartikel) eine unmittelbare Reaktion (Lachen, Kommentieren, besondere Ausrufe).	✓	
21-3	**2020-09-21_AB3**				Was interessiert mich bzgl. AB3? Allgemeines spontane Reaktionen bei Betrachtung des AB bzw. beim Lesen der Texte (Lesen die Schüler*innen wirklich den ganzen Text? (beeinflusst durch Sitznachbar*in (Korrektur der Lösung (durch späteren Einfall/Impuls durch Plenumsphase Unsicherheiten/Unzufriedenheit, verbal geäußert (Aha-Effekt (Wie viel Zeit benötigen die Lernenden zur Bearbeitung der Aufgabe? (AB3_Basis, Stern1 und Stern2 Kategorien werden induktiv entwickelt	✗	
	21-3_1	Plenumsphase				✗	
		21-3_1.1	Plenumsantworten			✗	
			21-3_1.1.1	Impuls: Was bedeutet Dunkelziffer?	Lernende äußern sich zum Impuls der Lehrkraft: "Was bedeutet Dunkelziffer?" in der Plenumsphase zu AB3.	✓	
			21-3_1.1.2	Impuls: Bewertung des BILD-Artikels	Lernende äußern sich zum Zeitungsartikel in der Plenumsphase zu AB3.	✓	
			21-3_1.1.3	Impuls: Können wir der Studie trauen?	Lernende äußern sich zum Impuls der Lehrkraft: "Können wir der Studie trauen?" in der Plenumsphase zu AB3.	✓	
			21-3_1.1.4	Impuls: Wurde Randomisierung beachtet?	Lernende äußern sich zum Impuls der Lehrkraft: "Wurde Randomisierung beachtet?" in der Plenumsphase zu AB3.	✓	
			21-3_1.1.5	Impuls: Was bedeutet repräsentativ?	Lernende äußern sich zum Impuls der Lehrkraft: "Was bedeutet repräsentativ?" in der Plenumsphase zu AB3.	✓	
			21-3_1.1.6	Impuls: Wieso kann man das nicht übertragen?	Lernende äußern sich zur Frage der Lehrkraft, warum die Ergebnisse nicht auf ganz Deutschland übertragen werden können	✓	

					Beschreibung		
		21-3_1.1.7		zu AB3_Basis, Aufgabe Basis	Lernende nennen ihre Lösung zur Aufgabe Basis des Arbeitsblatt 3_Basis.	✓	
		21-3_1.1.8		zu AB3_Basis, Aufgabe Stern1	Lernende nennen ihre Lösung zur Aufgabe Stern1 des Arbeitsblatt 3_Basis.	✓	
		21-3_1.1.9		zu AB3_Stern2, Aufgabe Stern2	Lernende nennen ihre Lösung zur Aufgabe Stern2 des Arbeitsblatt 3_Stern2.	✓	
	21-3_1.2	Spontanreaktion			Die Lernenden zeigen während der Plenumsphase eine unmittelbare Reaktion (Lachen, Kommentieren, besondere Ausrufe) auf Äußerungen der Lehrkraft.	✓	
21-3_2	Diskussion über Fachbegriffe				Lernende diskutieren untereinander über Fachbegriffe. Zu betrachten sind hier vor allem die Gruppenarbeitsphasen zum AB3.	✓	
21-3_3	Basis				Subcodes dieser Kategorie basieren auf den eingesammelten Arbeitsblättern und werden direkt an den Dokumenten codiert.	✗	
	21-3_3.1	Aufg. Basis				✗	
		21-3_3.1.1		ja	Lernende kreuzen bei der Frage, ob die Stichprobe genau erklärt wird, in der Differenzierungsstufe "Basis" des AB3 "Ja" an.	✓	2020-09-21 21_AB3 **nur Basis**
		21-3_3.1.2		nein	Lernende kreuzen bei der Frage, ob die Stichprobe genau erklärt wird, in der Differenzierungsstufe "Basis" des AB3 "Nein" an.	✓	2020-09-21 21_AB3 **nur Basis**
	21-3_3.2	zufällig?				✗	
		21-3_3.2.1		ja	Lernende kreuzen bei der Frage, ob die Stichprobe zufällig ausgewählt wird, in der Differenzierungsstufe "Basis" des AB3 "Ja" an.	✓	2020-09-21 21_AB3 **nur Basis**
		21-3_3.2.2		nein	Lernende kreuzen bei der Frage, ob die Stichprobe zufällig ausgewählt wird, in der Differenzierungsstufe "Basis" des AB3 "Nein" an.	✓	2020-09-21 21_AB3 **nur Basis**
	21-3_3.3	Aufg.*1				✗	
		21-3_3.3.1		ja	Lernende kreuzen bei der Frage, ob das Ergebnis auf Deutschland übertragen werden kann, in der Differenzierungsstufe "Stern1" des AB3 "Ja" an.	✓	2020-09-21 21_AB3 **nur Basis**
			21-3_3.3.1a	Übertragen durch Hochrechnung	wird codiert, wenn als Begründung aufgeführt wird, dass mithilfe einer Rechnung die Übertragung auf ganz Deutschland möglich ist.	✓	2020-09-21 21_AB3 **nur Basis**
			21-3_3.3.1.b	erneute Studie	wird codiert, wenn als Begründung aufgeführt wird, dass eine Befragung von ganz Deutschland nötig wäre; man die Ergebnisse dann aber übertragen könnte.	✓	2020-09-21 21_AB3 **nur Basis**

			21-3_3.3.2	nein			Lernende kreuzen bei der Frage, ob das Ergebnis auf Deutschland übertragen werden kann, in der Differenzierungsstufe "Stern1" des AB3 "Nein" an.	2020-09-21 21_AB3 **nur Basis**
				21-3_3.3.2a	Schätzung		wird codiert, wenn als Begründung aufgeführt wird, dass es sich nur um eine Schätzung handelt.	2020-09-21 21_AB3 **nur Basis**
				21-3_3.3.2b	nur in Gangelt getestet		wird codiert, wenn als Begründung aufgeführt wird, dass die Tests nur in Gangelt durchgeführt wurden.	2020-09-21 21_AB3 **nur Basis**
				21-3_3.3.2c	besondere Situation in Gangelt		wird codiert, wenn als Begründung aufgeführt wird, dass es sich bei der Situation in Gangelt um eine besondere Situation handelt.	2020-09-21 21_AB3 **nur Basis**
				21-3_3.3.2d	nicht repräsentativ		wird codiert, wenn in der Begründung der Fachbegriff "nicht repräsentativ" genannt wird.	2020-09-21 21_AB3 **nur Basis**
						höherer Anteil an Infizierten	wird codiert, wenn als "Unter-Begründung" im Obercode "nicht repräsentativ" ein höherer Anteil an Infizierten als Grundlage für nicht-Repräsentativität herangezogen wird.	2020-09-21 21_AB3 **nur Basis**
				21-3_3.3.2e	die Position anderer Forscher ist nicht bekannt		wird codiert, wenn als Begründung die fehlenden Meinungen anderer Forscher aufgeführt wird.	2020-09-21 21_AB3 **nur Basis**
				21-3_3.3.2f	nur eine Stichprobe		wird codiert, wenn als Begründung aufgeführt wird, dass es sich nur um eine Stichprobe handelt.	2020-09-21 21_AB3 **nur Basis**
		21-3_3.4	Interaktion mit Legende				wird codiert, wenn an der Legende, die eigentlich nicht-interaktiv ist, etwas eingetragen wird.	2020-09-21 21_AB3 **nur Basis**
	21-3_4	Stern1						
		21-3_4.1	Aufg. Basis					
			21-3_4.1.1	ja			Lernende kreuzen bei der Frage, ob die Stichprobe genau erklärt wird, in der Differenzierungsstufe "Stern1" des AB3 "Ja" an.	2020-09-21 21_AB3 **nur Stern1**
			21-3_4.1.2	nein			Lernende kreuzen bei der Frage, ob die Stichprobe genau erklärt wird, in der Differenzierungsstufe "Stern1" des AB3 "Nein" an.	2020-09-21 21_AB3 **nur Stern1**
		21-3_4.2	Aufg. *1 - Begründung					
			21-3_4.2.1	Ja			Lernende kreuzen bei der Frage, ob die Forscher Randomisierung beachtet haben, "Ja" an. Begründungen werden in Subcodes codiert.	
				21-3_4.2.1a	alles beachtet		wird codiert, wenn als Begründung angegeben wird, dass die Forscher*innen alle Aspekte beachtet haben.	2020-09-21 21_AB3 **nur Stern1**

		21-3_4.2.1b	genau erklärt	wird codiert, wenn als Begründung angegeben wird, dass die Forscher*innen ihre Vorgehensweise genau erklärt haben.	⊘	2020-09-21 21_AB3 **nur Stern1**
		21-3_4.2.1c	Stichprobe ist repräsentativ	wird codiert, wenn als Begründung angegeben wird, dass die Stichprobe repräsentativ ist.	⊘	2020-09-21 21_AB3 **nur Stern1**
		21-3_4.2.1d	Gangelt ist repräsentativ	wird codiert, wenn Gangelt als repräsentative Stadt herausgestellt wird. Der Begriff repräsentativ muss nicht genannt werden, es reicht die inhaltliche Aussage (JCH1-04 iPad - AB3_Stern1_Seite_4, 39 | 621)	⊘	2020-09-21 21_AB3 **nur Stern1**
		21-3_4.2.1e	Zufallsmethode	wird codiert, wenn als Begründung angegeben wird, dass eine zufällige Methode zur Stichprobenauswahl verwendet wurde.	⊘	2020-09-21 21_AB3 **nur Stern1**
		21-3_4.2.1f	zufällige Stadt getestet	wird codiert, wenn als Begründung angegeben wird, dass eine zufällige Stadt getestet wurde.	⊘	2020-09-21 21_AB3 **nur Stern1**
		21-3_4.2.2	Nein	Lernende kreuzen bei der Frage, ob die Forscher Randomisierung beachtet haben, "Nein" an. Begründungen werden in Subcodes codiert.	⊘	2020-09-21 21_AB3 **nur Stern1**
	21-3_5	Stern2			⊗	
		21-3_5.1	Aufg. *1		⊗	
			21-3_5.1.1	Ja	Lernende kreuzen bei der Frage, ob die Forscher Randomisierung beachtet haben, "Ja" an. Begründungen werden in Subcodes codiert.	⊘	2020-09-21 21_AB3 **nur Stern2**
				21-3_5.1.1a	Gangelt ist repräsentativ	wird codiert, wenn Gangelt als repräsentative Stadt herausgestellt wird. Der Begriff repräsentativ muss nicht genannt werden, es reicht die inhaltliche Aussage	⊘	2020-09-21 21_AB3 **nur Stern2**
				21-3_5.1.1b	Zufall	Zufall wird als Begründung genannt	⊘	2020-09-21 21_AB3 **nur Stern2**
					Zufallsverfahren: Briefe	Zufall wird als Begründung genannt und "Briefe" als Zufallsverfahren präzisiert.	⊘	2020-09-21 21_AB3 **nur Stern2**
			21-3_5.1.2	Nein	Lernende kreuzen bei der Frage, ob die Forscher Randomisierung beachtet haben, "Nein" an.	⊘	2020-09-21 21_AB3 **nur Stern2**
		21-3_5.2	Aufg. *2		Subcodes behandeln die Frage, warum die Studie nicht repräsentativ für Deutschland ist (AB3_Stern2, Aufgabe Stern2).	⊗	

			21-3_5.2.1	Schätzung	wird codiert, wenn als Begründung angegeben wird, dass es sich nur um eine Schätzung handelt.	✓	2020-09-21 21_AB3 **nur Stern2**
			21-3_5.2.2	weniger Infektionen in anderen Städten	wird codiert, wenn als Begründung angegeben wird, dass es in anderen Städten weniger Infektionen gibt.	✓	2020-09-21 21_AB3 **nur Stern2**
			21-3_5.2.3	Stichprobe zu klein	wird codiert, wenn als Begründung angegeben wird, dass die Stichprobe zu klein war.	✓	2020-09-21 21_AB3 **nur Stern2**
			21-3_5.2.4	Nennung Karneval	wird codiert, wenn in der Begründung das Karnevalsfest als relevanter Faktor genannt wird.	✓	2020-09-21 21_AB3 **nur Stern2**
			21-3_5.2.5	Nennung "nicht repräsentativ"	wird codiert, wenn in der Begründung der Begriff "nicht repräsentativ" genannt wird.	✓	2020-09-21 21_AB3 **nur Stern2**
			21-3_5.2.6	besondere Situation in Gangelt	wird codiert, wenn als Begründung angegeben wird, dass die Situation in Gangelt besonders ist.	✓	2020-09-21 21_AB3 **nur Stern2**
			21-3_5.2.7	andere Interessen in Gangelt	wird codiert, wenn als Begründung angegeben wird, dass in Gangelt andere Interessen vorherrschen.	✓	2020-09-21 21_AB3 **nur Stern2**
			21-3_5.2.8	nur Gangelt	wird codiert, wenn als Begründung angegeben wird, dass nur in Gangelt getestet wurde.	✓	2020-09-21 21_AB3 **nur Stern2**
	21-3_6	Spontanreaktion			Die Lernenden zeigen beim Öffnen des AB oder beim Lesen der Texte eine unmittelbare Reaktion (Lachen, Kommentieren, besondere Ausrufe)	✓	
	21-3_8	Korrektur der Lösung			Lernende korrigieren das Niedergeschriebene nach einiger Zeit und füllen das AB neu aus. Wird in Subcodes codiert.	✗	
		21-3_8.1	durch späteren Einfall		Dem Schüler fällt nach einiger Zeit eine andere Lösung ein und korrigiert das Niedergeschriebene.	✓	
		21-3_8.2	durch Plenumsphase		In der Plenumsphase besprochene Inhalte weichen von der eigenen Lösung ab, wodurch der Schüler seine Lösung korrigiert.	✓	
	21-3_11	Bearbeitungszeit				✓	
	21-3_12	Text übersprungen			Hinweise, die darauf deuten, dass Schüler*innen Teile des Texts nicht lesen/überfliegen	✓	
	21-3_13	Spontanreaktion			Die Lernenden zeigen beim Öffnen des AB oder beim Lesen der Texte eine unmittelbare Reaktion (Lachen, Kommentieren, besondere Ausrufe)	✓	

21-2	**2020-09-21_AB2**				Was interessiert mich bzgl. AB2? spontane Reaktionen bei Betrachtung des AB bzw. beim Lesen der Texte (x) Lesen die Schüler*innen wirklich den ganzen Text? (x) beeinflusst durch Sitznachbar*in (x) Korrektur der Lösung (x) durch späteren Einfall/Impuls durch Plenumsphase Unsicherheiten/Unzufriedenheit, verbal geäußert (x) Aha-Effekt (x) Wie viel Zeit benötigen die Lernenden zur Bearbeitung der Aufgabe? (x)	⊗	
	21-2_1	Plenumsphase				⊗	
		21-2_1.1	Plenumsantwort			⊗	
			21-2_1.1.1	Impuls: Was sagt dieser Wert aus?	Lernende äußern sich zum Impuls der Lehrkraft: "Was sagt dieser Wert aus?" in der Plenumsphase zu AB2.	✓	
		21-2_1.2	Spontanreaktion		Die Lernenden zeigen beim Öffnen des AB oder beim Lesen der Texte eine unmittelbare Reaktion (Lachen, Kommentieren, besondere Ausrufe)	✓	
	21-2_3	Werte korrekt			Alle Werte in der Tabelle auf AB2 sind korrekt.	✓	2020-09-21 21_AB2
	21-2_4	1 Wert falsch			Ein Wert in der Tabelle auf AB2 ist falsch.	✓	2020-09-21 21_AB2
	21-2_5	2 Werte falsch			Zwei Werte in der Tabelle auf AB2 sind falsch.	✓	2020-09-21 21_AB2
	21-2_6	zusätzliche Notiz			Lernende fügen dem Arbeitsblatt eine Notiz hinzu, die Anmerkungen enthält.	✓	
	21-2_7	off-task-Notiz			Lernende fügen dem Arbeitsblatt eine Notiz hinzu, die keinen fachlichen Inhalt hat.	✓	
	21-2_9	Korrektur der Lösung			Lernende korrigieren das Niedergeschriebene nach einiger Zeit und füllen das AB neu aus.	⊗	
		21-2_9.1	durch späteren Einfall		Dem Schüler fällt nach einiger Zeit eine andere Lösung ein und korrigiert das Niedergeschriebene.	✓	
		21-2_9.2	durch Plenumsphase		In der Plenumsphase besprochene Inhalte weichen von der eigenen Lösung ab, wodurch der Schüler seine Lösung korrigiert.	✓	
	21-2_14	keine Gespräche, Stille			wird codiert, wenn die Schüler*innen in der gesamten Arbeitsphase nicht über die Aufgabe sprechen	✓	

21-1	**2020-09-21_AB1**					Was interessiert mich bei AB1? spontane Reaktionen bei Betrachtung des AB bzw. beim Lesen der Texte Positionierung, schriftlich und mündlich (zum Vergleich mit dem eingesammelten AB) Begründung, schriftlich und mündlich (zum Vergleich mit dem eingesammelten AB) beeinflusst durch Sitznachbar*in Unsicherheiten/Unzufriedenheit, verbal geäußert Aha-Effekt Wie viel Zeit benötigen die Lemenden zur Bearbeitung der Aufgabe? Korrektur der Lösung durch späteren Einfall/Impuls durch Plenumsphase		
	21-1_1	Plenumsphase				Alle Codes in der Kategorie 21-1_1 beziehen sich auf die Plenumsphase zum AB1.		
		21-1_1.1	Plenumsantworten			In Subcodes dieser Oberkategorie werden alle Antworten von Schüler*innen codiert, die während der Plenumsphase zu AB1 verbal im Plenum geäußert werden.		
			21-1_1.1.1	Impuls: Woher bekommt der Arzt die Infos?		In der Plenumsphase zu AB1 gibt die LP den Impuls: "[…]Aber woher kriegt der Arzt denn seine Infos?". Plenumsantworten werden in Subcodes codiert.		
				21-1_1.1.1a	Laboruntersuchung	Lernende äußern, dass Ärzte über Laboruntersuchungen informiert werden.		
				21-1_1.1.1b	Studium	Lernende äußern, dass Ärzte durch ihr Studium über die nötigen Kenntnisse verfügen.		
			21-1_1.1.2	Impuls: Was müssten wir für eine Entscheidung tun?		In der Plenumsphase zu AB1 gibt die LP den Impuls: "[…] Wo könntet ihr euch denn informieren?[…]". Plenumsantworten werden in Subcodes codiert.		
				21-1_1.1.2a	Google	Lernende äußern, dass sie sich bei Google informieren würden.		
				21-1_1.1.2b	ausprobieren, was hilft	Lernende äußern, dass sie verschiedene Produkte ausprobieren würden.		
				21-1_1.1.2c	Arzt konsultieren	Lernende äußern, dass sie eine Ärztin oder einen Arzt zu Rate ziehen würden.		
				21-1_1.1.2d	Auf die Mehrheit hören	Lernende äußern, dass sie der Mehrheit vertrauen würden.		
		21-1_1.2	Spontanreaktionen			In Subcodes dieser Oberkategorie werden alle Reaktionen von Schüler*innen codiert, die während der Plenumsphase zu AB1 **spontan** und ohne Wortmeldung geäußert werden, z. B. als Reaktion auf eine Aussage der Lehrperson.		

		21-1_1.2.1	Studien kann man nicht prüfen		Lernende reagieren auf die Aussage der LP: "Wie prüft man Studien, die man irgendwo findet?" mit: "kann man nicht"	✓	
		21-1_1.2.2	Eine Studie hat gezeigt, …		Lernende reagieren auf die Aussage der LP: "Im Radio heißt es ja immer: Eine Studie hat gezeigt, dass…"	✓	
		21-1_1.2.3	Technik		Lernende reagieren auf etwas, das zur technischen Ausstattung gehört.	✓	
		21-1_1.2.4	spontane Positionierung		Lernende positionieren sich zu der Frage Tablette oder Banane durche eine spontane Äußerung.	✓	
		21-1_1.2.5	LP: Wo könntet ihr euch informieren?		Lernende reagieren spontan auf den Plenums-Impuls der LP: "Was meint ihr denn, müssten wir tun um uns jetzt entscheiden zu können für eine Lösung?". Hier werden Äußerungen codiert, die nicht im Plenum getätigt werden, sondern z. B. flüsternd in den Gruppen.	✗	
			21-1_1.2.5a	Arzt	Lernende äußern, dass sie eine Ärztin oder einen Arzt zu Rate ziehen würden.	✓	
			21-1_1.2.5b	Internet	Lernende äußern, dass sie im Internet nachschauen würden.	✓	
21-1_2	Positionierung (verbal)				In Subcodes dieser Oberkategorie werden verbale Aussagen codiert, die eine Positionierung der Schülerin oder des Schülers verdeutlichen.	✗	
	21-1_2.1	Magnesium-Tabletten			Äußerungen in der Gruppenarbeit, die verdeutlichen, dass der Schüler Magnesium wählt.	✓	
		21-1_2.1.1	Meldungen im Plenum		Äußerungen im Plenum, die verdeutlichen, dass der Schüler Magnesium wählt.	✓	
	21-1_2.2	Bananen			Äußerungen in der Gruppenarbeit, die verdeutlichen, dass der Schüler Bananen wählt.	✓	
		21-1_2.1.2	Meldungen im Plenum		Äußerungen im Plenum, die verdeutlichen, dass der Schüler Bananen wählt.	✓	
	21-1_2.3	weiß nicht			Äußerungen in der Gruppenarbeit, die verdeutlichen, dass der Schüler keine Entscheidung fällt.	✓	
		21-1_2.1.3	Meldungen im Plenum		Äußerungen im Plenum, die verdeutlichen, dass der Schüler keine Entscheidung fällt.	✓	
21-1_3	Positionierung (AB)					✗	
	21-1_3.1	Magnesium-Tabletten			Lernende kreuzen auf AB1 "Magnesium-Tabletten" an.	✓	2020-09-21 21_AB1

	21-1_3.2	Bananen	Lernende kreuzen auf AB1 "Bananen" an.	✓	2020-09-21 21_AB1
	21-1_3.3	weiß nicht	Lernende kreuzen auf AB1 "weiß nicht" an.	✓	2020-09-21 21_AB1
21-1_4	Begründung für Position			✗	
	21-1_4.1	Plenumsantwort	Lernende haben die im Plenum erarbeitete Lösung abgeschrieben und keine eigenen Notizen stehengelassen.	✓	
	21-1_4.2	es ist egal	Lernende notieren, dass es egal ist.	✓	
	21-1_4.3	keine Begründung	Lernende notieren keine Begründung.	✓	
	21-1_4.4	ausprobieren, was hilft	Lernende notieren, dass sie ausprobieren würden, welche Option ihnen hilft.	✓	
	21-1_4.5	eigene Erfahrung: Magnesium	Lernende haben schon eigene Erfahrungen mit der Supplementierung von Magnesium gesammelt und begründen darauf ihre Entscheidung.	✓	
	21-1_4.6	Banane ist gesund	Lernende notieren, dass Bananen gesund/gesünder seien.	✓	
	21-1_4.7	Trainer-Empfehlung	Lernende notieren, dass sie auf Empfehlung ihres (Sport-)Trainers ein Produkt ausprobieren würden.	✓	
	21-1_4.8	Banane enthält mehr Nährstoffe	Lernende notieren, dass Bananen mehr Nährstoffe enthalten.	✓	
	21-1_4.9	Banane ist für Sportler geeignet	Lernende notieren, dass Bananen für Sportler geeignet sind.	✓	
	21-1_4.10	Banane enthält Kalium und Magnesium	Lernende notieren, dass Bananen auch Kalium und Magnesium enthalten.	✓	
	21-1_4.11	Banane enthält weniger Magnesium	Lernende notieren, dass Bananen weniger Magnesium im Vergleich zur Tablette enthalten.	✓	2020-09-21 21_AB1
	21-1_4.12	bessere Leistung	Lernende notieren als Begründung eine verbesserte (körperliche) Leistung.	✓	
	21-1_4.13	Banane ist natürlich	Lernende notieren, dass Bananen natürliche Produkte/Naturprodukte sind.	✓	
	21-1_4.14	nach der Mehrheit entscheiden	Lernende notieren, dass sie nach der Mehrheit entscheiden würden.	✓	
	21-1_4.15	Arzt konsultieren	Lernende notieren, dass sie die Entscheidung auf der Grundlage einer ärztlichen Meinung treffen würden.	✓	
	21-1_4.16	Magnesium enthält mehr Nährstoffe	Lernende notieren als Begründung, dass in Magnesium-Tabletten (mehr)wichtige Nährstoffe enthalten sind.	✓	
	21-1_4.17	Muskel-entspannend	Lernende notieren, dass Magnesium Muskel-entspannend wirkt.	✓	
	21-1_4.18	entzündungshemmend	Lernende notieren, dass Magnesium entzündungshemmend wirkt.	✓	
	21-1_4.19	Banane hilft nicht	Lernende notieren als Begründung für Magnesium, dass Bananen nicht gegen Muskelkrämpfe helfen.	✓	
	21-1_4.20	normale Leistung	Lernende führen eine normale Leistung als Begründung an.	✓	
	21-1_4.21	1:1 Verhältnis	Lernende rezitieren das 1:1-Verhältnis der Nährstoffe, das auch im Instagram-Post angegeben ist.	✓	
	21-1_4.22	gute Muskelfunktion	Lernende notieren eine gute Muskelfunktion als Begründung.	✓	

	21-1_5	off-task-Notiz		Lernende benutzen Tablet-Funktionen, um auf dem Arbeitsblatt off-task-Notizen zu erstellen (z. B. Striche oder Formen außerhalb der zu bearbeitenden Bereiche).	✓	AB1, AB2, AB3
	21-1_7	Korrektur der Lösung		Lernende korrigieren das Niedergeschriebene nach einiger Zeit und füllen das AB neu aus.	✗	Videos: 20-09-21_Gruppe 1/2/3/4/5
		21-1_7.2	durch Plenumsphase	In der Plenumsphase besprochene Inhalte weichen von der eigenen Lösung ab, wodurch der Schüler seine Lösung korrigiert.	✓	
		21-1_7.3	durch späteren Einfall	Dem Schüler fällt nach einiger Zeit eine andere Lösung ein und korrigiert das Niedergeschriebene.	✓	
allg-1	Off-Task			Mindestens ein*e Lernende*r der Fokusgruppe arbeitet nicht an der Aufgabenstellung.	✗	
	allg-1_1	Verlassen des Sitzplatzes		Die Schülerin oder der Schüler verlässt seinen Sitzplatz und bewegt sich im Klassenraum.	✓	20-09-21_Gruppe 1/2/3/4/5 20-09-28_Gruppe 1/2/3/5
	allg-1_2	allgemeine Interaktion		Es wird mit der unbelebten Umwelt (Stifte, Tisch, Papier, Handy, …) in einer Weise interagiert, die der Bearbeitung der aktuellen Aufgabenstellung nicht zweckdienlich ist.	✓	
	allg-1_3	Gespräch		Es wird ein Gespräch mit einer/einem oder mehreren Partner*innen geführt, das der Bearbeitung der aktuellen Aufgabenstellung nicht zweckdienlich ist.	✓	
	allg-1_4	Tablet-Interaktion		Es wird mit dem Tablet in einer Weise interagiert, die der Bearbeitung der aktuellen Aufgabenstellung nicht zweckdienlich ist.	✓	
allg-2	Aha-Effekt			Lernende äußern verbal oder mittels Mimik/Gestik, dass ihnen ein Sachverhalt "plötzlich" klar wird.	✓	20-09-21_Gruppe 1/2/3/4/5 20-09-28_Gruppe 1/2/3/5
allg-3	Verwendung von Fachbegriffen			Die Lernenden tätigen einen Ausruf, der einen Fachbegriff der Unterrichtseinheit enthält oder schreiben einen Fachbegriff auf.	✗	
	allg-3_1	Wissenschaft [Wortfamilie]		Die Lernenden benutzen einen der folgenden Fachbegriffe: Wissenschaft Wissenschaftler*in wissenschaftlich Forscher*in	✓	

	allg-3_2	Kausalität [Wortfamilie]	Die Lernenden benutzen einen der folgenden Fachbegriffe: Kausalität kausal Schein-Kausalität	✓	20-09-21_Gruppe 1/2/3/4/5 20-09-28_Gruppe 1/2/3/5
	allg-3_3	Korrelation [Wortfamilie]	Die Lernenden benutzen einen der folgenden Fachbegriffe: Korrelation korrelieren [+ Deklination]	✓	
	allg-3_4	Repräsentativität [Wortfamilie]	Die Lernenden benutzen einen der folgenden Fachbegriffe: Repräsentativität repräsentativ repräsentieren [+ Deklination] repräsent~	✓	
	allg-3_5	Randomisierung [Wortfamilie]	Die Lernenden benutzen einen der folgenden Fachbegriffe: Randomisierung randomisiert randomisieren random~	✓	
	allg-3_6	Sonstige	Die Lernenden benutzen sonstige Fachbegriffe, die über normale Alltagssprache hinausgehen.	✓	
allg-4	Verständnisprobleme		Oberkategorie für Verständnisprobleme jeglicher Art, die bei den Lernenden auftreten.	✗	
	allg-4_1	Technik	Die Lernenden tätigen einen Ausruf, führen eine Geste aus oder schreiben etwas auf, das auf ein fehlerhaftes oder unvollständiges Verständnis eines Fachinhaltes hinweist.	✓	20-09-21_Gruppe 1/2/3/4/5 20-09-28_Gruppe 1/2/3/5
	allg-4_2	Fachinhalt	Die Lernenden tätigen einen Ausruf, führen eine Geste aus oder schreiben etwas auf, das auf technische Schwierigkeiten im Umgang mit den Tablets, Tablet-Stiften o.ä. hinweist.	✓	
	allg-4_3	Aufgabenstellung	Die Lernenden tätigen einen Ausruf, führen eine Geste aus oder schreiben etwas auf, das auf ein fehlerhaftes oder unvollständiges Verständnis der Aufgabenstellung hinweist.	✓	
allg-5	Interessante Einzel-Aussagen		In dieser allgemeinen Kategorie werden forschungsmethodisch-interessante Einzel-Aussagen von Lernenden codiert, die im späteren Verlauf der Auswertung wichtig sein könnten.	✓	20-09-21_Gruppe 1/2/3/4/5 20-09-28_Gruppe 1/2/3/5
allg-6	Skipping/Skimming		In dieser allgemeinen Kategorie wird das Verhalten des Überspringens oder Überfliegens von Texten codiert. Dabei wird im Video deutlich ersichtlich, dass Lernende schnell durch einen Text oder ein Arbeitsblatt scrollen, ohne alle Inhalte zu lesen.	✓	20-09-21_Gruppe 1/2/3/4/5 20-09-28_Gruppe 1/2/3/5

allg-7	Fragen von Lernenden	Lernende stellen eine Frage an die Lehrkraft.	20-09-21_Gruppe 1/2/3/4/5 20-09-28_Gruppe 1/2/3/5
allg-8	Beeinflussung des Sitznachbarn	Wenn die Positionierung eines Lernenden von dem/der Sitznachbar*in beeinflusst wird.	Videos: 20-09-21_Gruppe 1/2/3/4/5 sowie 20-09-21_Erprobung_f4tr anksript

Abbildung B.1.3: *Kategoriensystem zur Analyse der Fragebogen-Antworten in der Erprobung des Prüfschemas.*

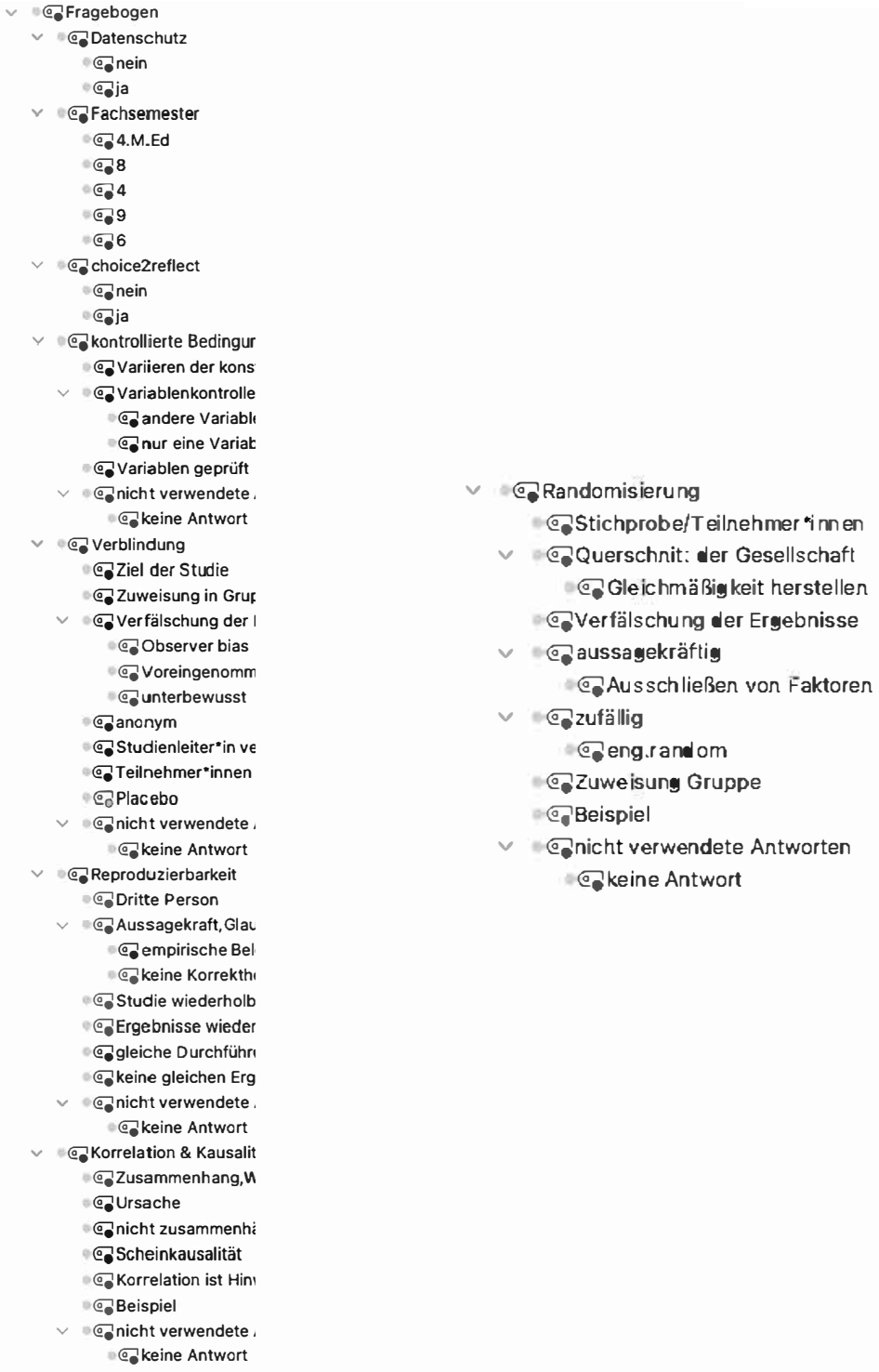


Abbildung B.1.4: *Kategoriensystem zur Analyse der Gruppen-Meetings in der Erprobung des Prüfschemas.*

B.2 Impulse in Plenumsphasen

Abbildung B.2.1: *Auflistung aller in der zweiten Doppelstunde der Gesamterprobung identifizierten Impulse mit Angabe antizipierter sowie gegebener Antworten. * = Passung. ** = Anschlussfähigkeit.*

Korrelation und Kausalität
Unterrichtsstunde am 28.09.2020

	Phase	Impuls	Antizipierte Antwort		Gegebene Antwort	Passung 1-4 1=schlecht	Wert der Antwort
	Problemaufriss	-					
	Erarbeitung 1	-					
28g	Zwischensicherung	Hier, so. Genau, und zwar auf dieser Tabelle was habt ihr da für Werte rausbekommen? Pos. 442	(für jede Zelle einen Tabellenwert, mit Antwortsatz)	I	(jede*r Schüler*in nennt einen Wert, ohne Kontext, nur „unter die 5 kommt eine 4" usw.)	2	2
28h		Einmal zusammen schauen was diese Werte jetzt überhaupt aussagen. Die erste Zeile hier: Instagram-App installiert und dann viele Leute mit einer kurzen wenige mit einer langen Akkulaufzeit. Was könnte uns das sagen? Pos. 521	Kurze Laufzeit bei... „installierter App" „langer Nutzung" „altem Handy" Nicht die Instagram-App allein ist schuld an einer kurzen Akkulaufzeit. Es gibt weitere Faktoren.	I	Zunächst keine Antworten der SuS. Nach Erklärung: 2-02: [...] Also Instagram zieht viel Akku, ja. Aber es hängt oft auch damit zusammen, wie lange man das nutzt oder wie alt das Handy ist.	4	4

		Und wenn wir uns jetzt nicht die Instagram-App, sondern wie du sagtest die Handy-Nutzung anschauen, was könnten wir daraus ziehen? Pos. 539 [LP liest Ergebnisse zum Faktor Nutzungszeit] C: Was wäre dann so die Schlussfolgerung, die wir ziehen könnten? Pos. 543		II	1-12: Also es liegt nicht nur/ es liegt nicht nur an der App sondern, wenn man viel mit dem Handy macht, müsst man auch länger/öfter laden.	4	4
				III	1-12: Vielleicht ist das Handy älter und der Akku ist deswegen nicht mehr so gut oder er ist schon kaputt gemacht worden, irgendwie.	4	4
28i		[…] die Forschergruppe, die jetzt diese Studie geschrieben hat, kommen nach der Analyse zu dem Ergebnis, die Instagram-App sorgt für eine kurze Akku-Laufzeit und ist daher als Stromfresser einzustufen. Pos. 560	(SuS bewerten kritisch die Schlussfolgerung) Pro: Daten zeigen, dass ein Zusammenhang besteht. Contra: Aber! Es gibt auch noch andere Faktoren, die auf die Akkulaufzeit wirken	I	2-16: Ähm, nein, weil es doch darauf ankommt, wie alt das Handy zum Beispiel ist oder wie lange es genutzt wird.	3	2
		Was sagt ihr dazu? Kann man das so machen? Pos. 561		II	1-12: Also da gab es ja nicht die Frage, ob es auch andere Apps gibt. Vielleicht liegt der Akku-Verbrauch auch an einer anderen App.	4	4
	Erarbeitung 2 (AB3)	-					

28k	Sicherung (zu AB3)	C: Ich möchte von euch einmal kurz hören in eigenen Worten was ist Korrelation und dann danach, was ist Kausalität. Vielleicht einmal Korrelation, 2-16? Pos. 626	Korrelation: Ein Zusammenhang zwischen zwei Faktoren. Durch Studien herausgefunden. Kausalität: ein Faktor ist die Ursache des anderen Faktors.	I	2-16: Das sind Zusammenhänge, also wenn etwas miteinander zusammenhängt.	4	2
				II	1-12: Kausalität ist die Ursache.	4	2
28l		Genau, woran erkennt man eine Kausalität oder eine Korrelation? Pos. 639	Korrelation: UND-Aussage Kausalität: WEIL-Aussage	I	2-13: Bei Korrelation steht und und bei Kausalität weil.	4	3
28m		Und dann jetzt noch kurz zum Abschluss haben die Forscherinnen und Forscher bei der Instagram-Studie Korrelation und Kausalität beachtet? Pos. 648	Studie stellt mehrere Korrelationen her. Die Forschenden gehen allerdings von einer Kausalität bei einem Faktor aus. Das ist jedoch nicht möglich.	I	1-12: Also ich hab geschrieben, sie haben nicht darauf geachtet, dass vielleicht eine andere App den Akku fressen könnte. Dadurch haben sie eine Scheinkausalität erzeugt, denn der dritte Faktor muss auch beachtet werden, so wie ersten zwei.	4	4
		Die Instagram-Studie, […] wurde es da beachtet oder haben die da was falsch gemacht?		II	1-12: Der dritte Faktor wurde nicht beachtet.	4	4
28n		C: Okay, dann jetzt die abschließende Frage:	Der Schlussfolgerung dieser Studie können wir nicht	I	2-16: Nein, weil die anderen Faktoren nicht beachtet wurden.	4	4

		Instagram-Studie können wir dieser Studie trauen oder können wir ihr nicht trauen? Pos. 686 (umformuliert) Das Ergebnis der Studie war ja, die Instagram-App ist ein Stromfresser. Aber ist die Studie richtig gut gemacht gewesen oder nicht?	trauen, denn es wurden nicht alle Faktoren beachtet. Dennoch: Die Tendenzen/Korrelationen können wir im Hinterkopf behalten. - Nicht alle Faktoren beachtet - Nur auf die App geachtet - Schein-Kausalität aufgestellt So getan, als wäre es eine Kausalität				
28o	Abschluss	Okay, zum Abschluss was habt ihr aus dieser Stunde mitgenommen? Pos. 698	-	I	2-16: Das Korrelation Zusammenhänge sind und Kausalität Ursachen.	4	4
				II	1-01: Dass Instagram bei alten Handys Strom fressen kann, also viel.	2	1
28p		C: Genau, kann Strom fressen, aber können wir das auf Instagram zurückführen? Pos. 706		I	1-01: Nein, das kann auch noch an anderen Apps liegen.	4	2

Abbildung B.2.2: *Darstellung der identifizierten Lehrimpulse in der ersten Doppelstunde der Gesamterprobung am 21.09.2020. * = Passung. ** = Anschlussfähigkeit.*

	Phase	Impuls	Antizipierte Antwort	Gegebene Antwort	*	**
21a	Plenumsphase AB 1	Wer von euch hatte sich für die erste Antwort entschieden?	(Schüler*innen melden sich entsprechend ihrer Auswahl)	(Schüler*innen melden sich entsprechend ihrer Auswahl)	4	
21b		So, wer hat sich für die zweite Antwort entschieden?	(Schüler*innen melden sich entsprechend ihrer Auswahl)	(Schüler*innen melden sich entsprechend ihrer Auswahl)	4	
21c		Hat auch jemand die dritte Antwort genommen?	(Schüler*innen melden sich entsprechend ihrer Auswahl)	(Schüler*innen melden sich entsprechend ihrer Auswahl)	4	
21d		Okay, dann wollen wir das einmal kurz diskutieren und ich möchte einmal eure Gründe dafür hören. Warum meintet ihr denn, Paula soll Magnesium nehmen gegen die Krämpfe?	(Schüler*innen geben in eigenen Worten die Begründung für ihren Standpunkt an)	S: „Ja. Also ich hätte jetzt/ also von mir aus hätte ich als erstes, wenn es jetzt in meinem Gruppenchat geschrieben hätte ich gesagt, keine Ahnung geh zum Arzt. Und dann wäre die wahrscheinlich zum Arzt gegangen und dann hätte der Arzt wahrscheinlich gesagt: Musst mehr Magnesium oder so zu dir nehmen. Und dann hätte die sich wahrscheinlich Magnesiumtabletten geholt oder wahrscheinlich hätte sie mehr Bananen gegessen. Also, ja. Und deswegen habe ich die erste Antwort genommen, weil es eigentlich egal ist, weil sie sowieso Magnesium zu sich nehmen sollte.“	4	4
				S: „Da stand, dass Magnesium im Körper wieder zu normalen Leistungen macht.“	4	
21e		Wer hat sich denn für die Bananen entschieden oder kann dafür einmal eine Begründung abgeben?	(Schüler*innen geben in eigenen Worten die Begründung für ihren Standpunkt an)	S: „Bananen sind natürlicher, sie enthalten nicht nur Magnesium sondern auch andere Nährstoffe, die der Körper benötigt.“	4	
				S: Also meine Begründung ist, dass Bananen zwar auch viele wichtige Stoffe noch zusätzlich enthalten, aber sehr viel weniger Magnesium als jetzt eine Magnesiumtablette.	4	

21f		C: Okay, gut. Wir haben ja hier jetzt ein ziemlich gemischtes Bild. Was meint ihr denn, müssten wir tun um uns jetzt entscheiden zu können für eine Lösung? C: Oder was würdet ihr jetzt als nächstes tun, wenn ihr eine Entscheidung treffen wollt?	Schüler*innen stellen fest, dass man zusätzliche Informationen benötigt: Studien zur Thematik, Expert*innen.	S: „Ähm, ich würde auf die Mehrheit hören oder sonst zum Arzt gehen“	4	2
				S: Ähm, ich würde einfach ausprobieren was hilft. Also ich würde beides ausprobieren und gucken was besser ist.	4	2
21g		C: So, wenn ihr euch jetzt vorstellt, ihr müsst jetzt eine Entscheidung fällen. Wollt das auch nicht alles ausprobieren, weil vielleicht mögt ihr gar nicht so gerne Bananen. Und wenn jetzt Bananen aber irgendwie super gut sind, dann würdet ihr die probieren, aber wollt auch nicht immer nur ausprobieren. Wo könntet ihr euch denn informieren?	Schüler*innen nennen verschiedene Informationsquellen; im Laufe des Gesprächs erkennen Sie Studien/Experimente/Umfragen als passende Quellen	S: Ja, ich würde auf Google nachschauen, auf Google nachschauen.	4	2
				S: Wie Marvin schon gesagt hat, zum Arzt gehen und nachfragen.	4	2
21h		C: Genau, das kann man alles tun. Aber woher kriegt der Arzt denn seine Infos? Wieso weiß der das?	Schüler*innen nennen wiss. Studien als Informationsquelle	S: Vielleicht, weil der auf der Uni war und der das da studiert hat. Deswegen weiß der das oder, weil er auf vertrauenswürdigen Seiten unterwegs ist.	4	3
21i		C: Ja, genau. Also der hat auf jeden Fall sein Fach studiert. Aber woher wissen denn die Leute, die ihn unterrichten was besser ist, oder wie treffen die Leute Entscheidungen?	Schüler*innen nennen wiss. Studien als Informationsquelle	S: Erstmal wird überhaupt geguckt, ob es überhaupt ein Magnesiummangel ist. Also meistens wird das Blut abgenommen und ins Labor geschickt und dann kontrollieren die das.	2	1
21k	Plenums-phase AB 2	C: Okay, dann wollen wir einmal kurz die Werte vergleichen. Damit wir da auch von den selben Zahlen alle zusammen reden. Habt ihr	Schüler*innen geben in ganzen Sätzen die eingetragenen Tabellenwerte wieder. Dabei	S: Ja, das Erste ist 22. Gemeldete Infizierte und getestete Stichprobe.	2	2

		doch alle ausgefüllt bestimmt, dann vielleicht aus dieser Richtung mal eine Meldung.	stellen sie auch einen inhaltlichen Bezug her.			
21l		Magst du noch einmal sagen, was das jetzt sagt dieser Wert?	Schüler*in präzisiert die Angabe und nennt auf inhaltlicher Ebene die Bedeutung der Zahlenwerts.	S: Also von den 900 Leuten, untersuchten Personen, waren halt 22 Personen offiziell als infiziert gemeldet.	4	4
21m	Plenums-phase zu AB 3	C: Hat jemand die Basisaufgabe gemacht? Und kann hier die Antwort geben, oder habt ihr alle mit Stern1 angefangen? / Okay, was habt ihr denn angekreuzt beim Ersten?	Schüler*innen geben die von ihnen angekreuzte Antwort an.	S: Also bei der ersten Aufgabe habe ich ja angekreuzt.	4	2
21n		C: Das genau/ Also erstmal was hast du da angekreuzt: Wird die Stichprobe zufällig ausgewählt?	Schüler*innen geben die von ihnen angekreuzte Antwort an.	S: Ja	4	1
21o		C: Also was können wir jetzt hieraus schließen?	Schüler*innen ziehen eine Schlussfolgerung aus der Kombination der vorherigen Antworten.	S: Das ist randomisiert.	4	4
21p		Kann man das Ergebnis der Studie auf ganz Deutschland übertragen?	Schüler*innen begründen, dass das Ergebnis nicht auf ganz Deutschland übertragbar ist mit der fehlenden Repräsentativität der Stichprobe.	S: Nein, das geht nicht. Also, ich hab jetzt Stern2 gemacht. Ich hab in der Erklärung geschrieben, die Strich/Stichprobe ist nicht repräs/ präsentativ für Deutschland, weil in Gangelt eine viel höhere Prozent/ ein viel höherer Prozentsatz infiziert war aufgrund der Karnevalsfeier.	4	4
				S: Ich hab geschrieben, also es ist nicht repräsentativ für ganz Deutschland, weil nur in der Gemeinde Gangelt quasi die Coronafälle gemessen wurden. Also es gilt das ja nicht für ganz Deutschland.	4	4

21q	Plenums-phase AB 3, Stern2	Und ich frag nochmal in die ganze Runde rein, haben die Forscherinnen und Forscher Randomisierung beachtet oder nicht in der Studie? Und ist die Stichprobe repräsentativ für die Gemeinde und für Deutschland? Das sind jetzt drei Sachen die wir klären wollen. Zuerst wurde Randomisierung beachtet?	Schüler*innen wiederholen die zuvor erarbeitete Erkenntnis, dass die Stichprobe randomisiert ausgewählt wurde, aber nur für Gangelt repräsentativ ist. Die Erklärung wird in mehreren Schritten von der Lehrkraft erfragt.	S: Ja, weil es durch ein Zufallssystem ausgewählt wurde.	4	4
		Da wird ja jetzt zwischen zwei Sachen unterschieden, nämlich ist die Stichprobe repräsentativ für die Gemeinde Gangelt und für ganz Deutschland. Gerade hat Kyell das ja für ganz Deutschland gesagt, aber wie sieht es bei Gangelt aus? C: Okay, dann gehen wir vielleicht einmal einen kurzen Schritt noch zurück und fragen uns, was dieses furchtbare Fremdwort überhaupt bedeutet. Kyell du hast das ja vorhin einmal benutzt. Repräsentativ, was versteht man darunter?		S (liest von AB): Das heißt, dass das in diesem Fall für die ganze Gemeinde beziehungsweise fürs ganze Deutschland steht. Die befragte Stichprobe muss repräsentativ in Klammern aussagekräftig für die große Gruppe sein, auf die du deine Ergebnisse übertragen willst. Das erreichst du durch Randomisierung.	1	1
21r		[…]Also es ist repräsentativ also aussagekräftig für Gangelt. Aber warum jetzt nicht für Deutschland?	Schüler*innen erklären, warum die Stichprobe nicht für ganz Deutschland repräsentativ ist.	S: Weil es ja nur in Gangelt also befragt wurde.	4	4
21s		Können wir der Stichprobe trauen/ der Studie trauen, ist sie gut	Schüler*innen nehmen begründet dazu Stellung, dass	S: Kann man trauen. Weil, wir durch ein Zufalls aus/ ausgewählt wurde.	3	2

		gemacht oder ist sie nicht gut gemacht?	die Studie vertrauenswürdig ist, allerdings nur geringe Aussagekraft über Gangelt hinaus besitzt.			
21t		Habt ihr mal mitgekriegt irgendwie in den Medien so dieses Wort Heinsberg-Studie von dem Virologen Henrik Streeck, habt ihr da irgendwie eine Meinung mitgekriegt? Wurde das diskutiert vielleicht?	Schüler*innen berichten, ob sie bereits von der Heinsberg-Studie gehört haben. Ggf. gibt es hier keine Antworten.	(keine Antworten)	1	1
21u	Anwend-ung, Bewerten des BILD-Artikels	C: Was sagt ihr denn dazu dann jetzt, wenn ihr euch die Tabelle anschaut und dann das, was die Bild daraus gemacht hat?	Schüler*innen erklären, dass die Heinsberg-Studie zwar vertrauenswürdig ist und wissenschaftlich durchgeführt wurde, dass allerdings die BILD-Zeitung die Ergebnisse falsch interpretiert.	S: 1,8 Millionen ist die Dunkelziffer, die sein könnte. Also das ist jetzt nicht bewiesen, sondern das ist nur eine Schätzung.	1	1
				S: Das ist nicht die/ also das sind nicht insgesamt die ganzen Infektionen, sondern nur eben die Dunkelziffer, die sein könnte.	1	1
21v		Meint ihr, das hat die Bild so richtig gemacht - so aus der Studie zu zitieren?	Schüler*innen stellen heraus, dass die BILD-Zeitung nicht korrekt vorgegangen ist.	(keine Antwort)	1	1
21w		C: Wir haben geklärt, die Studie ist randomisiert. Also wir können ihr vertrauen, aber nur unter einer Einschränkung.	Schüler*innen benennen die Einschränkung, dass die Ergebnisse nur für Gangelt gelten.	S: Also nur unter einer Einschränkung, dass das nur für die Gemeinde Gangelt genommen wird. Da steht auch, dass es eine Dunkelziffer ist. Also ist das nicht quasi/ quasi ist es nicht richtig. Also nicht ganz richtig.	4	4
21x		C: Ja, genau. Aber was bedeutet Dunkelziffer? Die Dunkelziffer, die ausgerechnet wurde?	„Die Dunkelziffer gibt an, wie viele Leute tatsächlich erkrankt sind“	S: Wie viele Leute ungefähr an Corona erkrankt sind, aber es wurden nicht so viele gemeldet.	4	4
21y		C: Warum noch mal, warum gilt das nur für Gangelt und nicht für Deutschland? Der eine Fachbegriff, Ja?	Nicht repräsentativ für ganz Deutschland.	S: Ich weiß nicht ob das genau das ist, aber hier steht doch nur 50% der Leute aus Gangelt (unv.)	1	1
				S: Würde man jedoch rein theoretisch (betonend) die Dunkelziffer in ganz Deutschland ausrechnen.	2	3

B.3 Codelines

Abbildung B.3.1: *Codeline-Darstellung der Arbeitsphasen-Auswertung zum Prüfkriterium Korrelation und Kausalität für die Gruppen 1-5.*

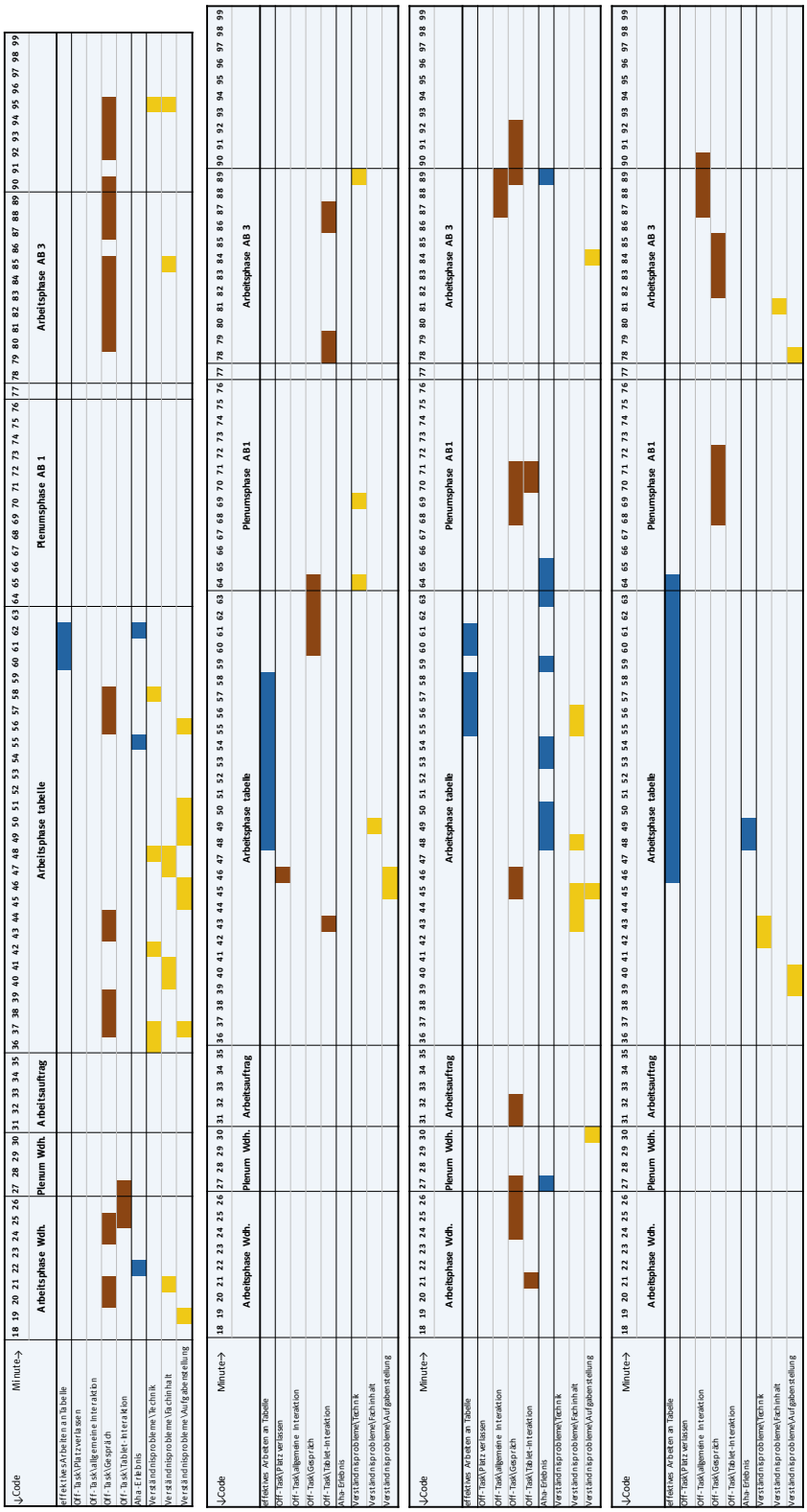


Abbildung B.3.2: *Codeline der Arbeitsphasen-Auswertung zum Prüfkriterium Korrelation und Kausalität in der Gesamtdarstellung aller Gruppen bezogen auf die Codes Verständnisschwierigkeiten, Aha-Erlebnisse und Off-Task.*

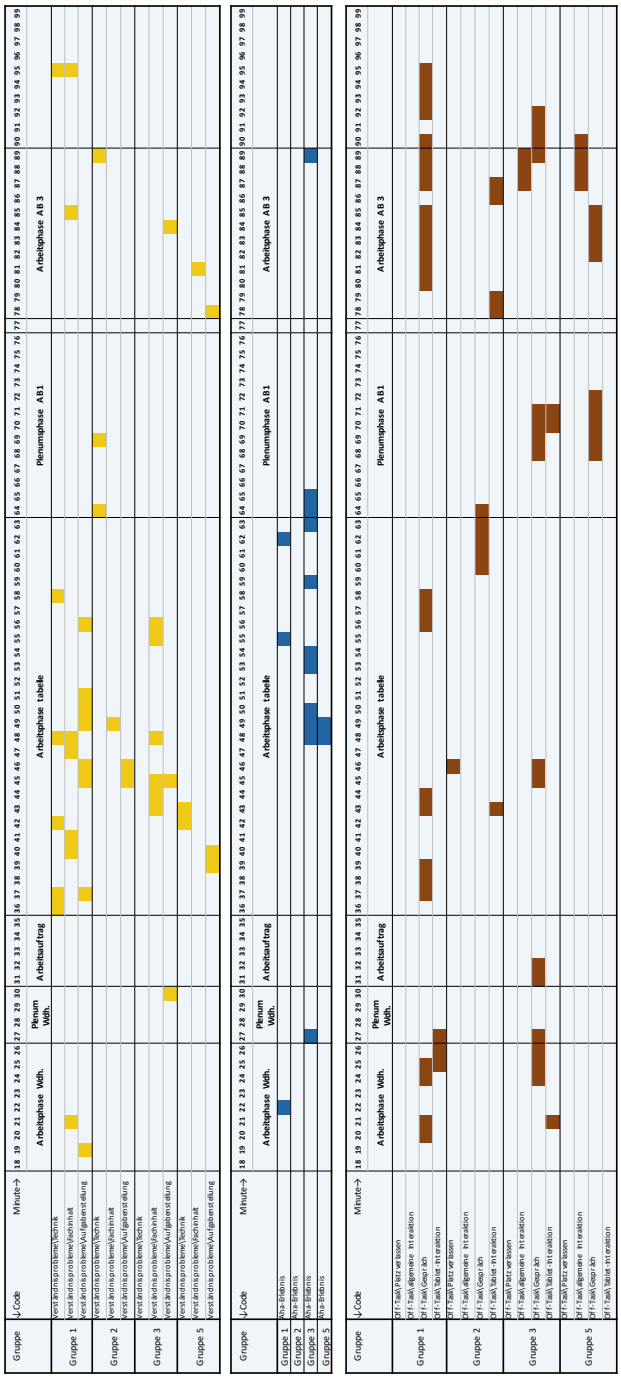


Abbildung B.3.3: *Codeline-Darstellung der Arbeitsphasen-Auswertung zum Prüfkriterium Randomisierung für die Gruppen 1-5.*

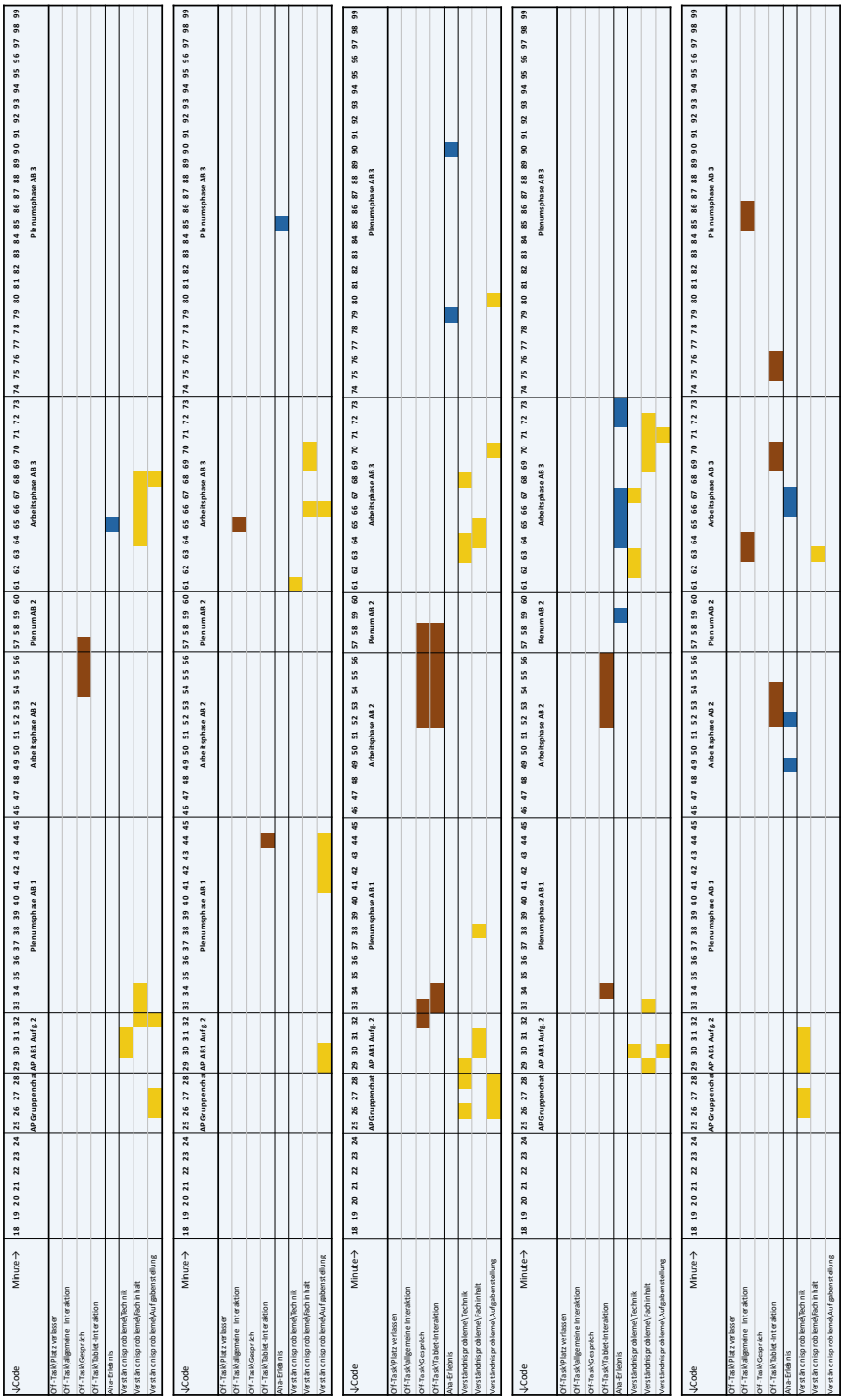
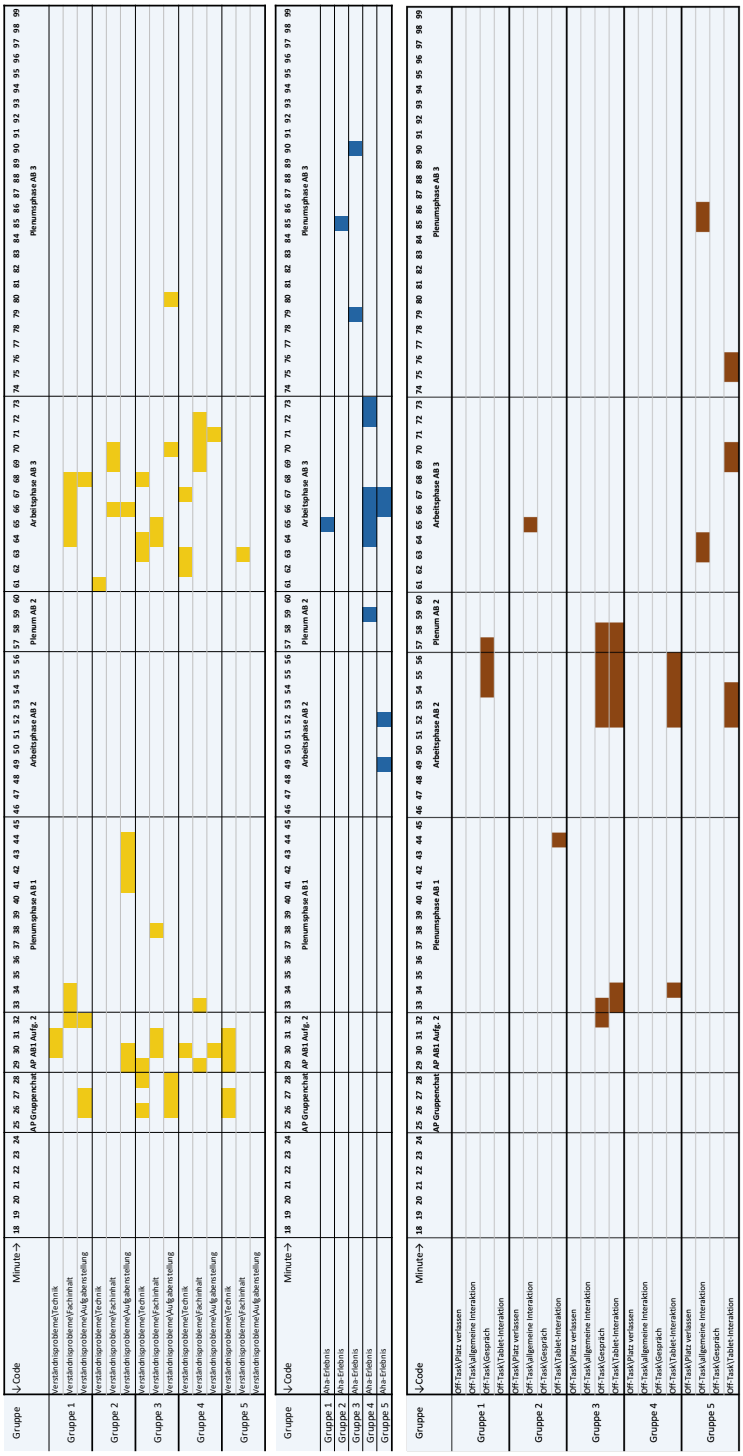


Abbildung B.3.4: *Codelines der Arbeitsphasen-Auswertung zum Prüfkriterium Randomisierung in der Gesamtdarstellung aller Gruppen bezogen auf die Codes Verständnisschwierigkeiten, Aha-Erlebnisse und Off-Task.*



B.4 Einzelaussagen

Abbildung B.4.1: *Auflistung von Einzelaussagen aus der Erprobung zum Prüfkriterium Falsifizierbarkeit, die als relevant für die Datenanalyse angesehen werden.*

Zitat	Unterrichtsphase	Nachweis	Nummer
Zur Bestätigung, was wäre da nötig? Keine/ ach nee. Es kommen keine/ nein, schon welche. Das ist ja die Bestätigung dafür. […] Wir müssen noch etwas anders machen. Alle Haselnüsse der Welt, eigentlich!	Arbeitsphase AB 2	CK26; M 9, Absatz 66-68	**1**
Hä, es gibt keine Haselnüsse, die kein Fett enthalten. Ja das denke ich gerade auch. Ich checke das nicht. Komm, wir nehmen jetzt einfach auch alle Haselnüsse, weil wir keine eigenen Ideen haben.	Arbeitsphase AB 2	GR18 und TA03; M 11, Absatz 76-79	**2**
Hä, aber.. Das kann man nicht wirklich nachweisen. Weil, wenn er sagt die sind sehr selten, kann es ja sein, dass es nur zwei auf der Welt gibt. Und die gibt‘s nur auf Mallorca, 3km unter der Insel.	Arbeitsphase AB 2	KN12; M 12, Absatz 61	**3**
Eigentlich müsste Bernd das irgendwie begründen können, dass er so etwas glaubt. Denn eigentlich kann man eine Behauptung ohne Begründung nicht ernst nehmen.	Plenumsphase zu AB 2	UN16; M 9, Absatz 87	**4**
Ja, das ist ungefähr so eine Behauptung, wie wenn jemand sagt: "Es gibt Gott". Denn unabhängig davon ob es einen gibt oder nicht, kann man halt in dem Sinne auch nicht widerlegen, weil man findet ihn halt nicht. Und man kann dann aber sagen, er ist trotzdem da. Auch wenn du ihn noch nicht gefunden hast. Heißt also, man kann nur maximal/ den dann finden oder es wird halt nicht widerlegt.	Plenumsphase zu AB 2	ES02; M 9, Absatz 91	**5**
Dass es etwas „gibt“. Dass es um die Existenz von etwas geht. Was halt nur schwer, bzw. fast unmöglich ist.	Arbeitsphase AB 4	MU31; M 13 Absatz 199	**6**
Ich habe eine andere Frage zu der davor mit dem Periodensystem. Wenn Aluminium in der dritten Hauptgruppe ist, wie ist das denn dann falsifizierbar, diese Aussage? Weil eigentlich kann es ja nicht falsch sein. Also kann diese Aussage ja nicht falsch sein, wenn Aluminium da nicht drin ist.	Plenumsphase zu AB 4	UN16; M 14; Absatz 177	**7**

Abbildung B.4.2: *Transkribierte Antworten der Wiederholungsaufgabe in der zweiten Doppelstunde der Gesamterprobung.*

Kürzel	Antwort Basis	Antwort Stern
1-01	Forscher untersuchen zufällige Menschen mit einer Stichprobe. - Stichprobe von verschiedenen Menschen für eine Studie.	Das die Stichprobe aussagekräftig sein sollte.
1-02	Randomisierung ist zufällige Menschen und (unv.) Stichprobe.	repräsentativ ist aussagekräftig.
1-03	Zufällig ausgesuchte Personen für eine Studie	aussagekräftig für eine andere Gruppe
1-04	Ist wenn man eine zufällige Gruppe an Menschen befragt und somit Ergebnisse zu einer Studie festhalten kann.	Das eine Studienstichprobe representativ sein muss, heißt dass eine Stichprobe zufällig und somit auf alle Menschen zutreffen kann (Interessen, Alter, usw
1-06	Leute werden zufällig gewählt - Stichprobe von verschiedenen Menschen	/
1-07	Eine gerechte Auswertung - Leute werden zufällig gewäht "random" - Stichprobe von verschiedenen Menschen für eine Studie	Repräsentativ bedeutet aussagekräftig für eine bestimmte Gruppen von Menschen
1-12	Das bedeutet, dass durch ein Zufalls System eine Stichprobe von verschiedenen Menschen ausgewählt wird.	Es muss für diese bestimmte Gruppe aussagekräftig sein. Das heißt aber nicht, dass das Ergebnis für alle Gruppen zählt sondern nur für die randomisierte Gruppe in dem ausgewählten Gebiet.
2-01	Bei einer Randomisierung bei einer Studie werden alle Beteiligten an Hintergrund(unv.) für die Studie ausgewählt. Also zufällig.	/
2-02	Bei einer Studie spricht man von Randomisierung, wenn alle Beteiligten zufällig ausgewählt wurden, ohne irgendwelche Merkmale zu beachten.	Das bedeutet, dass die Stichprobe für die repräsentierte Gruppe passen muss, sie repräsentiert die gesamte Gruppe. Nicht nur die Stichprobe.
2-03	Randomisierung ist ein Begriff mit dem man Studien nachprüfen kann. Dabei werden von zufälligen Leuten Stichproben genommen von versch. Menschen.	Repräsentativ bedeutet aussagekräftig.
2-08	Wenn verschiedene Personen befragt oder getestet werden. Für eine Studie	Das die Stichprobe nicht aus der Gleichen Familie oder Freundeskreis stamt sonder von verschieden Personen - Aussagekräftig für eine bestimmte gruppe von Menschen
2-11	Ist wenn z.b. eine Dunkelziffer von Infektionen in einem Ort heraus gefunden wird in dem auf die einwohner Randomisierung ist für eine Studie	Das Personen zufälig ausgesucht werden
2-12	Ist wenn zb. Eine Dunkelziffer von Infektionen in einem Ort heraus gefunden wird in dem auf die Einwohnerzahl hochgerechnet wird. Dies erfolgt zufällig. - für eine Studie - Stichprobe von verschiedenen Menschen - "random" = zufällig	Das zb. Personen zufällig ausgesucht werden. - Repräsentativ bedeutet aussagekräftig für eine bestimmte Gruppe von Menschen
2-13	Randomisierung ist z.B Coronafälle in einer Stadt z.B Osnabrück sind 20.000 erkrankte an Corona aber es werde Random gemacht also einfach geschätzt genauer kann man es nicht wissen - für eine Studie - Stichprobe von Menschen	Also das die Werte einigermaßen Stimmen sollten also sie sollten representativ sein. - repräsentativ = Aussagekräftig
2-15	Wenn die Leute zufällig gewählt werden Random heißt irgendwelche Mensch	Repräsentativ bedeutet aussagekräftig für eine bestimmte Gruppen von Menschen
2-16	Randomisierung ist, wenn z.b. bei einer Studie mehrere Leute zufällig (in Englisch --> random --> deshalb Randomisierung) ausgewählt wurden. - Stichprobe von verschiedenen Menschen für eine Studie	Repräsentativ heißt aussagekräftig für eine bestimmte Gruppen von Menschen

B.5 Arbeitsblätter

B.5.1 Phase I und II

Abbildung B.5.1: *Erstellte Instagram-Posts zu a) den Inhaltsstoffen einer Banane und b) einem Magnesium-Präparat zur Verwendung im Fachspezifischen Problemaufriss und der Positionierung.*

Abbildung B.5.2: *Arbeitsblatt 1 in der Unterrichtsstunde zu den Phasen I und II zur Problemfrage: Helfen Magnesium-Präparate gegen Muskelkrämpfe?*

WWU

Dein Kürzel (BS = Buchstabe):

3. BS deines Vornamens	1. BS der Straße in der du wohnst	Deine Hausnummer

Chemie Didaktik Münster

Magnesium gegen Muskelkrämpfe?

Aufgabe 1:

Paula hat beim Fußballtraining häufig Muskelkrämpfe. In ihrem Mannschafts-Chat fragt sie ihre Freundinnen und Freunde um Rat.

a) Betrachte das Video vom Gruppenchat der Fußball-Mannschaft.

b) Unten siehst du zwei Instagram-Beiträge, die Noel unter **#magnesiummangel** gefunden hat. Lies dir beide Beiträge genau durch.

B.5.2 Phase III

Abbildung B.5.3: *Arbeitsblätter für die Unterrichtseinheit zur Falsifizierbarkeit.*

Arbeitsblatt 1: Birtes Behauptung

1. BS Vorname Mutter	3. BS eigener Vorname	Geburtstag (nur Tag)

Birte sagt:
„Ich esse lieber Nüsse als Nutella.“ Sie **behauptet**:
„Haselnüsse enthalten weniger Fett als Nutella!“

Aufgabe 1:
Was benötigst du, um Birtes **Behauptung** zu **überprüfen** und wie gehst du vor?
Entwickle eine kleine **Untersuchung**!

Was benötigst du?	Wie gehst du vor?

Was musst du beobachten, damit Birtes Behauptung **bestätigt** wird?	Was musst du beobachten, damit Birtes Behauptung **widerlegt** wird?

Aufgabe 2:
Führe die **Untersuchung** durch, trage das Ergebnis ein und kreuze an!

Fettgehalt Nüsse: Fettgehalt Nutella:

Wird die Behauptung bestätigt?		Wird die Behauptung widerlegt?	
☐ Ja	☐ Nein	☐ Ja	☐ Nein

Arbeitsblatt 2: Bernds Behauptung

1. BS Vorname Mutter	3. BS eigener Vorname	Geburtstag (nur Tag)

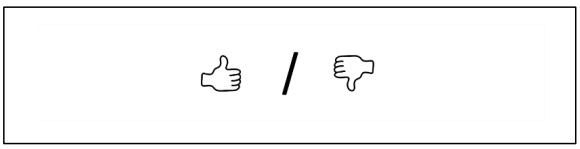
Bernd sagt:
„Das Ergebnis überzeugt mich nicht.“ Er **behauptet**:
„Es gibt trotzdem Haselnüsse, die gar kein Fett enthalten. Die sind aber sehr selten!“

Aufgabe 1:
Was benötigst du, um Bernds **Behauptung** zu **überprüfen** und wie gehst du vor?
Entwickle eine kleine **Untersuchung**!

Was benötigst du?	Wie gehst du vor?

Aufgabe 2:
GEDANKENEXPERIMENT – Diese Untersuchung führst du nur in deinem Kopf durch. Überlege, welche Beobachtungen zur **Bestätigung** oder **Widerlegung** nötig wären.

Kann die Behauptung bestätigt werden?		Kann die Behauptung widerlegt werden?	
☐ Ja	☐ Nein	☐ Ja	☐ Nein



Arbeitsblatt 3: Was ist Falsifizierbarkeit?

1. BS Vorname Mutter	3. BS eigener Vorname	Geburtstag (nur Tag)

Falsifizierbarkeit

Wissenschaftlerinnen und Wissenschaftler arbeiten nur mit Aussagen, die wirklich **falsifizierbar** (das heißt **widerlegbar**) sind. Sie sprechen dabei von der **Falsifizierbarkeit**.

Wenn eine Aussage durch Beobachtungen oder Experimente widerlegt werden kann, dann ist sie **falsifizierbar** (= **widerlegbar**) und somit eine **wissenschaftliche** Aussage.

Beispiel:
„Alle Schwäne sind weiß!“
Sobald jemand einen schwarzen Schwan beobachtet, ist die Aussage falsifiziert.

Manche Aussagen sind so aufgebaut, dass sie durch keine Beobachtung und kein Experiment je widerlegt werden können. Sie sind **nicht falsifizierbar**, das bedeutet **nicht widerlegbar**: keine Wissenschaftlerin und kein Wissenschaftler wird eine solche Aussage benutzen wollen. Sie ist **unwissenschaftlich**.

Beispiel:
„Alle Schwäne sind entweder weiß oder nicht weiß!“
Niemand könnte es je schaffen, diese Aussage zu falsifizieren. Sie ist immer wahr.

Aufgabe:
Birte und Bernd haben Behauptungen aufgestellt. Eine dieser Behauptungen ist aber **unwissenschaftlich**. Benutze die Aufkleber, um beide Behauptungen endgültig „abzustempeln“!
Begründe deine Entscheidung.

☐ Birtes Behauptung ist **unwissenschaftlich**.

☐ Bernds Behauptung ist **unwissenschaftlich**.

Begründung:

Arbeitsblatt 4: Fake-News

1. BS Vorname Mutter	3. BS eigener Vorname	Geburtstag (nur Tag)

Wie lassen sich **nicht-falsifizierbare** Aussagen erkennen?

In den sozialen Netzwerken verbreiten sich **Fake-News** immer weiter. Oft lässt sich jedoch nicht auf den ersten Blick erkennen, ob es sich um eine Fake-Neuigkeit handelt…

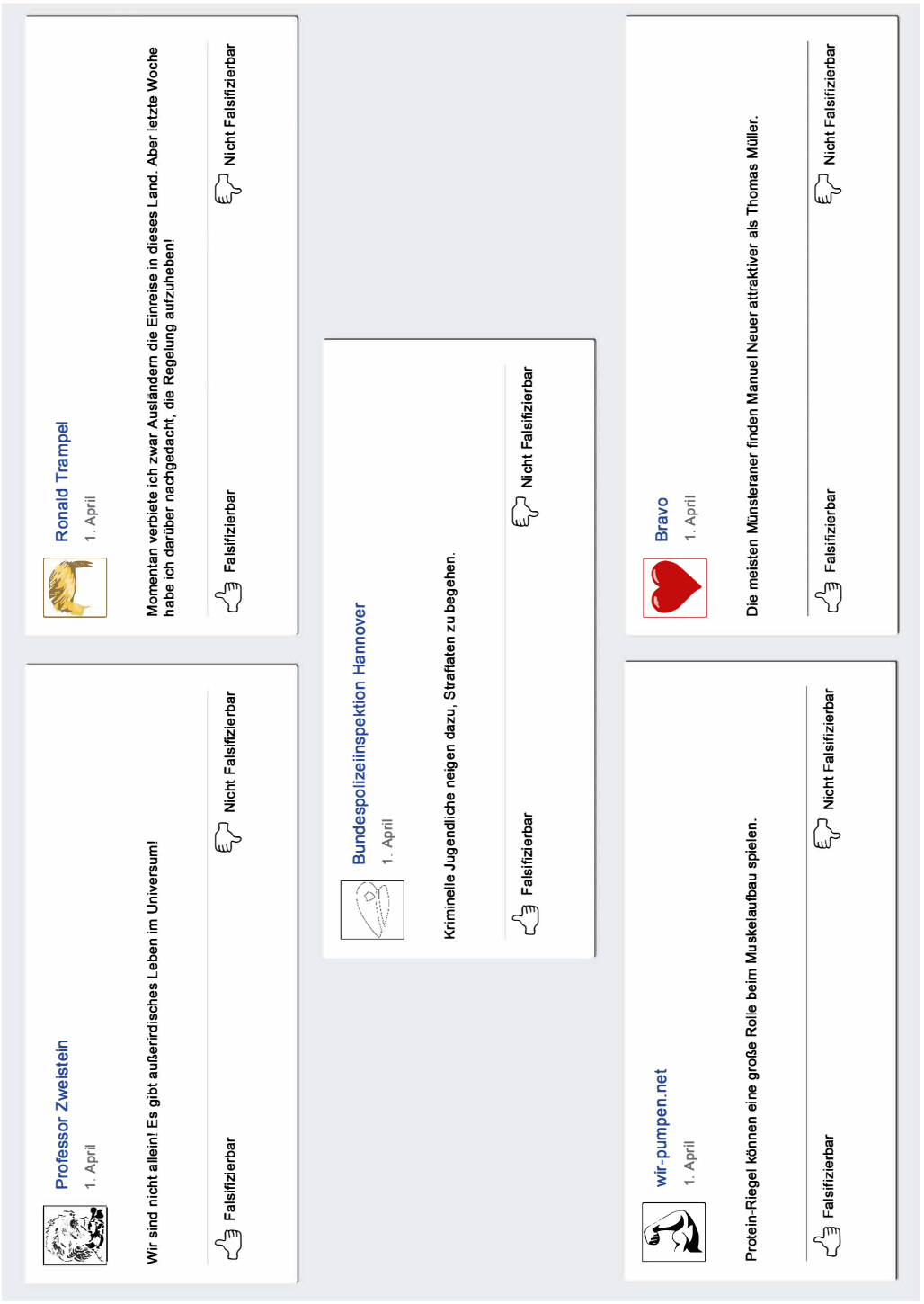
- Prüfe, ob die Neuigkeit eine **falsifizierbare** Aussage enthält.

Aufgabe 1:
Auf der nächsten Seite findest du fünf Meldungen, die so auch in einem sozialen Netzwerk auftauchen können.
Daumen hoch oder **Daumen runter**? Entscheide für jede Meldung, ob es sich um eine falsifizierbare (und damit wissenschaftliche) Aussage handelt.

Aufgabe 2:
Wie kannst du **nicht-falsifizierbare** Aussagen erkennen? Suche nach Merkmalen für **unwissenschaftliche**, **nicht falsifizierbare** Behauptungen. Achte dabei auf auffällige Wörter und Formulierungen.

Nicht falsifizierbare Aussagen erkennt man an…

Abbildung B.5.4: *Nach Falsifizierbarkeit zu bewertende Aussagen für die Unterrichtseinheit zur Falsifizierbarkeit.*



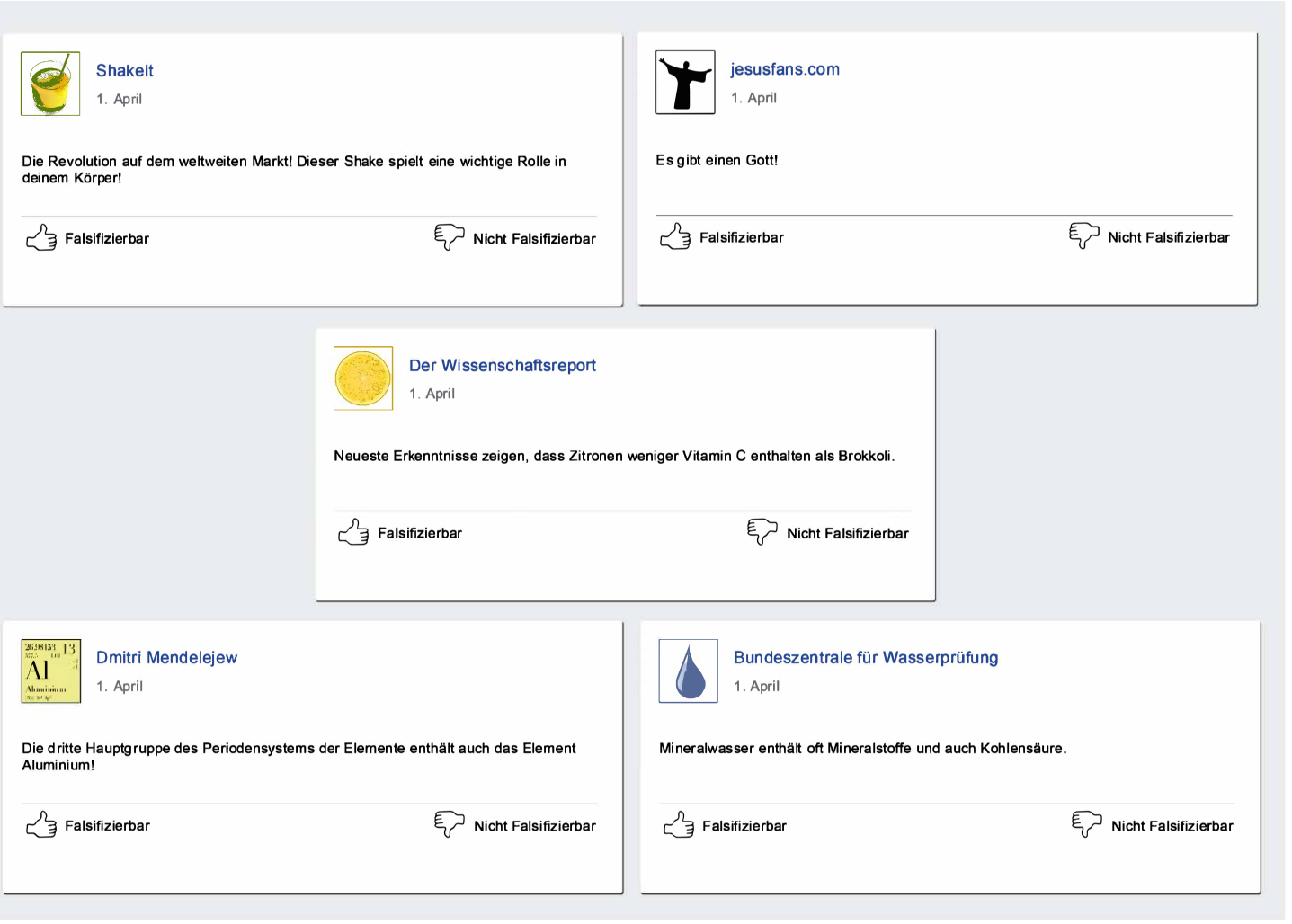

Shakeit
1. April
Die Revolution auf dem weltweiten Markt! Dieser Shake spielt eine wichtige Rolle in deinem Körper!
Falsifizierbar
Nicht Falsifizierbar
jesusfans.com
1. April
Es gibt einen Gott!
Falsifizierbar
Nicht Falsifizierbar
Der Wissenschaftsreport
1. April
Neueste Erkenntnisse zeigen, dass Zitronen weniger Vitamin C enthalten als Brokkoli.
Falsifizierbar
Nicht Falsifizierbar
Dmitri Mendelejew
1. April
Die dritte Hauptgruppe des Periodensystems der Elemente enthält auch das Element Aluminium!
Falsifizierbar
Nicht Falsifizierbar
Bundeszentrale für Wasserprüfung
1. April
Mineralwasser enthält oft Mineralstoffe und auch Kohlensäure.
Falsifizierbar
Nicht Falsifizierbar

Abbildung B.5.5: *AB 2 zur Erarbeitung von Randomisierung. Enthalten ist eine vereinfachte Zusammenfassung der Heinsberg-Studie nach Streeck et al. (2020)*

WWU MÜNSTER

Chemie Didaktik Münster

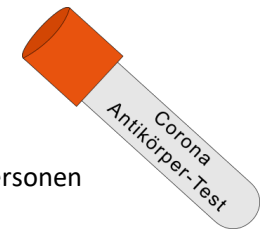
Studie zur Ermittlung der Corona-Fallzahlen in der Gemeinde Gangelt (Kreis Heinsberg)

Aufgabe 1: **Lies** den Text der Studie genau und **ergänze** die Tabelle am Ende. Du findest alle wichtigen Werte im Text.

Im Frühjahr 2020 brach weltweit die Corona-Pandemie aus. Das Corona-Virus ist sehr ansteckend und kann unterschiedliche Folgen haben. Manche Personen merken kaum etwas von der Krankheit, während andere Personen daran sterben.
Infizierte Leute, die kaum etwas merken, sind trotzdem ansteckend. Die wissenschaftliche Forschung versucht aus diesem Grund, die sogenannte **Dunkelziffer** und die **Sterblichkeit** herauszufinden.
Die Dunkelziffer beschreibt die tatsächliche Anzahl an Infektionen: Dazu zählen also alle bekannten und unbekannten Corona-Fälle.
Mit einem Antikörper-Test können Wissenschaftlerinnen und Wissenschaftler das Blut von Personen untersuchen.
Damit stellen sie fest, ob eine Person bereits Corona hatte oder nicht.
Die Forschergruppe um den Virologen Hendrik Streeck hat in der Heinsberg-Studie die Dunkelziffer in der **Gemeinde Gangelt** ermittelt. Gangelt liegt im Landkreis Heinsberg. In Gangelt gab es zu Anfang der Pandemie einen starken Anstieg an Infektionen. Der Grund dafür war ein **Karnevalsfest**. Dort haben sich viele Menschen gegenseitig angesteckt. Die Zahl der Infektionen war in Gangelt also **viel höher** als in anderen Regionen Deutschlands.

Methode der Studie

In Gangelt lebten zum Zeitpunkt der Untersuchung etwa 12.600 Einwohner. Davon waren 388 Personen offiziell als infiziert gemeldet. Die Forschergruppe konnte natürlich nicht alle Einwohner untersuchen. Daher wählte sie aus einer Einwohnerliste *per Zufall* Personen aus und schickte ihnen einen Brief.



919 Personen nahmen an der Studie teil und wurden mit dem Antikörper-Test auf Corona untersucht.

Abbildung B.5.6: *AB 0 zur Wiederholung der Begriffe Randomisierung und repräsentativ, eingesetzt in der zweiten Doppelstunde der Gesamterprobung.*

WWU MÜNSTER

Chemie Didaktik Münster

Aufgabe Basis:
Denke zurück an die letzte Stunde. Was ist **Randomisierung**?
Besprich dich mit deiner Partnerin oder deinem Partner und schreibe deine Erklärung in eigenen Worten auf!

Aufgabe ★1:
„Die Stichprobe in einer Studie muss **repräsentativ** sein".
Was bedeutet das?

Umblättern für Hilfe

© C. Kralisch, Prof. Dr. A. Marohn: choice²reflect; Institut für Didaktik der Chemie, Münster

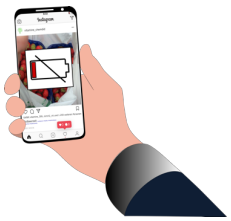
Abbildung B.5.7: *Arbeitsblätter zur Erarbeitung des Prüfkriteriums Korrelation und Kausalität.*

WWU MÜNSTER

Chemie Didaktik Münster

Wie hängt Instagram mit der Akkulaufzeit zusammen?

Forscherinnen und Forscher haben in einer Studie den Zusammenhang zwischen der Instagram-App und der Akkulaufzeit von Smartphones untersucht. **2000 Personen** zwischen 16 und 62 Jahren wurden zu ihrer Smartphone-Nutzung befragt. Dabei wurde herausgefunden, dass die Akkuladung von Smartphones mit installierter Instagram-App rund 6 Stunden schneller auf unter 20 % sank als von Smartphones, auf denen die App nicht installiert ist. Die Forschergruppe stufte die Instagram-App aufgrund dieser Ergebnisse als „Stromfresser“ ein.



Aufgabe Basis:
Wechsle zu den **Daten der Studie** und **führe die Datenauswertung durch.** Wie das geht, steht auf dem ausgeteilten Arbeitsblatt.

Aufgabe ★1:
Die Forschergruppe kommt nach dieser Analyse zu folgendem Ergebnis:

„Die Instagram-App sorgt für eine kurze Akkulaufzeit und ist daher als Stromfresser einzustufen“.

Bewerte dieses Ergebnis gemeinsam mit deinem Partner.

© C. Kralisch, Prof. Dr. A. Marohn: choice²reflect; Institut für Didaktik der Chemie, Münster

Alter: ____ Jahre **Geschlecht:** _____

Frage 1	**Antwort**	
Besitzen Sie ein Smartphone?	□ Ja	□ Nein

Frage 2	**Antwort**	
Wie alt ist Ihr Smartphone?	☐ 0-6 Monate ☐ 13-18 Monate ☐ 25-30 Monate ☐ 37-42 Monate	☐ 7-12 Monate ☐ 19-24 Monate ☐ 31-36 Monate ☐ über 42 Monate

Frage 3	**Antwort**	
Wie lange nutzen Sie Ihr Smartphone pro Tag?	☐ 0-1 Stunde ☐ 2-3 Stunden ☐ 4-5 Stunden	☐ 1-2 Stunden ☐ 3-4 Stunden ☐ über 5 Stunden

Frage 4	**Antwort**	
Haben Sie die Instagram-App auf Ihrem Smartphone installiert?	□ Ja	□ Nein

Frage 5	**Antwort**	
Wie lange hält der Akku Ihres Smartphones im Durchschnitt, bis Sie wieder laden müssen?	☐ unter 6 Stunden ☐ 13-19 Stunden ☐ 25-30 Stunden	☐ 6-12 Stunden ☐ 20-24 Stunden ☐ über 30 Stunden

Datenauswertung

Auf der Homepage der Forschungsgruppe steht die komplette Studie mit einem Ausschnitt aus den Daten, die erhoben wurden:

Die Ergebnisse von **40 befragten Personen** sind online abrufbar.

a. Betrachte den Fragebogen der Studie.

b. Wechsle zu den Original-Daten.

c. Setze einen **Filter**, damit die Daten übersichtlicher werden. Für diese Studie sind nur Personen wichtig, die ein Smartphone besitzen. Blende alle Personen aus, die kein Smartphone haben. Das machst du so:

i. Tippe auf die Spalte „Smartphone vorhanden?“

ii. Tippe unten in der rechten Ecke auf das „**Zelle**“ Symbol und wähle **Schnellfilter**. Setze einen Haken für die Option „Ja“.

iii. Kontrolliere, ob bei allen angezeigten Personen in der Spalte „Smartphone vorhanden“ „Ja“ steht.

d. Setze noch einen Filter, der nur Smartphones mit einer **langen Akkulaufzeit** größer als 20 Stunden anzeigt.

e. Setze einen dritten Filter, der nur Smartphones mit installierter **Instagram-App** anzeigt.

f. Zähle, wie viele Personen nun noch angezeigt werden und trage das Ergebnis unten in die Tabelle ein (App installiert + lange Laufzeit).

g. Setze Filter in den Spalten „Instagram installiert“ und „Handynutzung“ so, dass du die Tabelle unten ausfüllen kannst.

☆**1** Untersuche auch den Faktor „Alter des Smartphones“. Setze passende Filter und ergänze die Tabelle.

		Akkulaufzeit	
		kurz (unter 20 Stunden)	lang (über 20 Stunden)
Instagram App	installiert		
	nicht installiert		
Handynutzung pro Tag	unter 3 Stunden		
	über 3 Stunden		
☆**1** Alter des Smartphones	unter 24 Monate		
	über 24 Monate		

Wenn du die Tabelle ausgefüllt hast, wechsle zurück zum Arbeitsblatt und bearbeite **Aufgabe 2**.

Abbildung B.5.8: *Arbeitsblätter zur Erarbeitung des Prüfkriteriums Reproduzierbarkeit.*

Arbeitsmaterial 1: Die Zitronenbatterie

Batterien- die Energiequelle der E-Scooter

Was das Benzin für unsere Autos ist, ist die Batterie für den E-Scooter. Um mit einem E-Scooter fahren zu können, benötigen wir Strom. Dafür müssen wir nicht extra zu einer Tankstelle fahren, sondern können einfach die Batterie an einer Steckdose, bei uns zu Hause, aufladen. In der Batterie wird dann die Energie gespeichert und im elektrischen Motor unseres E-Scooters zu Bewegungsenergie umgewandelt.

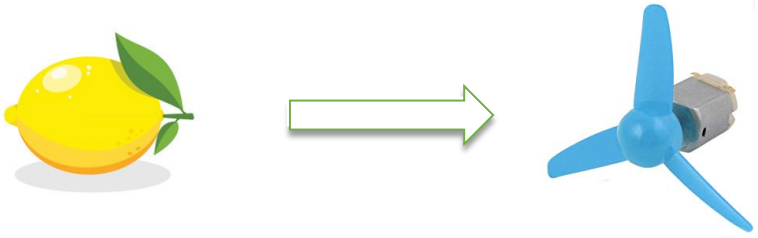
In Zitronen stecken nicht nur Vitamine!

In Zitronen stecken nicht nur viele Vitamine. Wir können aus ihnen auch Batterien herstellen, die einen elektrischen Stromfluss erzeugen. Der Aufbau einer Zitronenbatterie hat Ähnlichkeit mit dem Aufbau von Batterien, die für E-Scooter hergestellt werden.

Tim will versuchen seinen E-Scooter mit Zitronen zu betreiben. Seine Freundin Hannah hat es geschafft, mit einer Zitrone einen kleinen Elektromotor zum Drehen zu bringen. Damit Tim seine Idee, den E-Scooter durch Zitronen zum Fahren zu bringen, entwickeln kann, bekommt er das Versuchsprotokoll von Hannah. In einem ersten Schritt will er den Versuch von Hannah wiederholen, um zu schauen ob der Elektromotor wirklich mit einer Zitrone in Bewegung versetzt werden kann.

Aufgabe:

Baut mit Hilfe der bereitgestellten Materialien und dem angefertigten Versuchsprotokoll von Hannah die Zitronenbatterie für Tim nach und versucht den Elektromotor zum Drehen zu bringen.



Arbeitsmaterial 2: Das Versuchsprotokoll

Versuchsprotokoll: Die Zitronenbatterie

Name: Hannah Datum: 13.04.2019

Thema: Eine Zitronenbatterie erzeugt Strom.

Material:

Zitrone	Messer
Geldstück	Brettchen
Zinkblech	kleines Becherglas
Kabel mit Krokodilklemmen	
Elektromotor	

Durchführung:

Das Geldstück und das Zinkblech werden mit der Zitrone und dem Elektromotor verbunden.

Beobachtung:

Der Elektromotor dreht sich.

Auswertung:

Die Zitrone kann mit einem Zinkblech und einem Geldstück einen Elektromotor zum Drehen bringen.

Arbeitsmaterial 3a: Reproduzierbarkeit von Experimenten

Dein Kürzel (BS = Buchstabe):

3. BS deines Vornamens	Geburtstag (nur Tag)	1. BS deiner Straße

Reproduzierbarkeit von Experimenten

Naturwissenschaftliche Experimente müssen wiederholbar sein. In der Wissenschaft nennt man dieses Kriterium **Reproduzierbarkeit**.

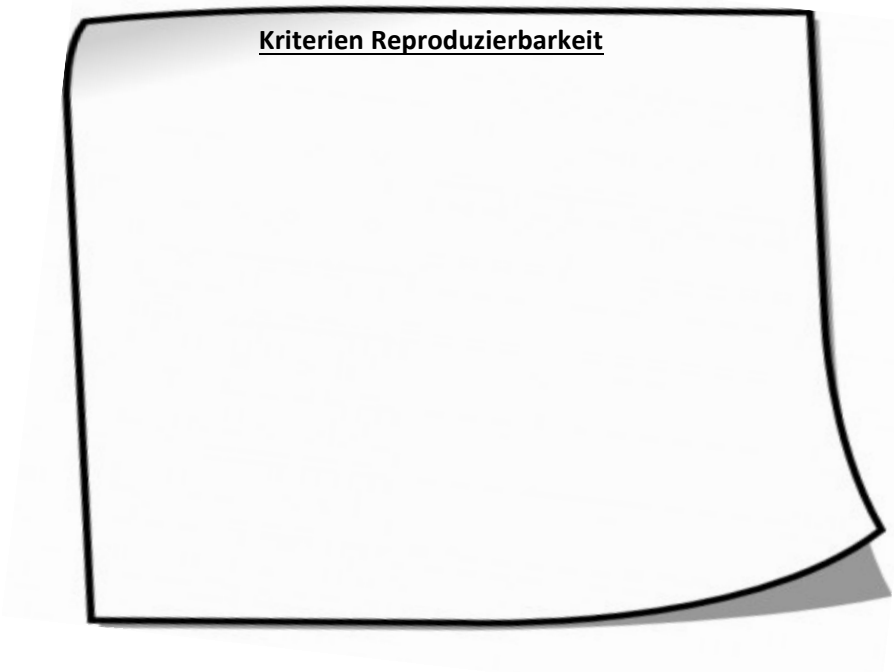
Es gilt:

Wenn ein Experiment unter gleichen Bedingungen und gleicher Durchführung von einem anderen Bearbeiter wiederholt wird und **das gleiche Ergebnis** gewonnen wird, gilt es als reproduzierbar.

Wenn ein Experiment unter gleichen Bedingungen und gleicher Durchführung von einem anderen Bearbeiter wiederholt wird und **nicht das gleiche Ergebnis** gewonnen wird, ist es im wissenschaftlichen Sinne nicht-reproduzierbar.

Aufgabe (Gruppenarbeit):

Überlegt euch, welche **Einflussfaktoren** Tim alle kennen muss, um das Experiment zu **reproduzieren**. Verfasst eine **Liste** mit Kriterien.



Arbeitsmaterial 3b: Reproduzierbarkeit von Experimenten

Dein Kürzel (BS = Buchstabe):

3. BS deines Vornamens	Geburtstag (nur Tag)	1. BS deiner Straße

Aufgabe (Einzelarbeit):

Ist Hannahs Protokoll reproduzierbar? **Entscheide** dich.

Reproduzierbar:	**Nicht-Reproduzierbar:**
☐	☐

Bewerte mit euren Kriterien das Versuchsprotokoll von Hannah. Überlege dir dafür mindestens **zwei Argumente**.

Arbeitsmaterial 4: Reproduzierbarkeit von Studien

Reproduzierbarkeit von Studien

Wenn eine wissenschaftliche Studie mit der gleichen Methode und den gleichen Versuchsbedingungen, aber mit anderen Versuchsteilnehmern von einem anderen Bearbeiter, erneut wiederholt wird und **das gleiche Ergebnis** erhalten wird, spricht man von einer **reproduzierbaren** Studie.

Ein Problem für die Wissenschaft:

Die Anforderungen an die Veröffentlichung einer Studie können mit den Anforderungen an ein Versuchsprotokoll verglichen werden.

Versuchsprotokoll

- Name und Datum (Bearbeiter)
- Thema
- Material
- Aufbau
- Durchführung
- Beobachtung
- Ergebnis

Aufbau Studie

- Name und Datum (Bearbeiter)
- Thema
- Versuchsteilnehmer
- Methode
- Durchführung
- Beobachtung
- Ergebnis

Arbeitsmaterial 5: Studie E-Scooter

Aufgabe: Lies dir die vorliegende Studie zum Thema E-Scooter durch.

Studie zeigt: E-Scooter, Freizeitspaß oder New Mobility?

Nürnberg, 17.09.2019

Seit Juni sind E-Scooter in Deutschland offiziell zugelassen und stehen seitdem an vielen Ecke in unseren Städten. Wer keinen eigenen besitzt, lädt sich einfach eine App der zahlreichen Sharing-Dienste herunter und kann für wenig Geld durch deutsche Innenstädte rollen. Das Interesse scheint groß. Aktuelle Befragungswerte zeichnen ein gemischtes Bild. Laut GfK haben 5 Prozent der Befragten in Deutschland Anfang Juli einen E-Scooter getestet und 14 Prozent der Befragten gaben an, dieses Fortbewegungsmittel zumindest einmal ausprobieren zu wollen.

Einen E-Scooter anschaffen kommt für 2 Prozent in Frage, weitere 13 Prozent haben schon darüber nachgedacht- und etwa 1 Prozent besitzen bereits einen. 70 Prozent zeigen hingegen gar kein Interesse. Bei den Nutzungsmotiven treffen zwei starke Trends aufeinander: Spaß-Faktor und Umweltbewusstsein.

Spaß-Faktor und Umweltbewusstsein stehen im Vordergrund

Über 70 Prozent der Befragten, die E-Scooter per se gut finden, verbinden mit dem Fahrerlebnis puren Spaß. Aber auch der Umweltgedanke bleibt nicht aus. 71 Prozent haben bei der Nutzung das Gefühl, etwas Gutes für die Umwelt zu tun. Wenngleich E-Scooter im Moment in erster Linie in der Freizeit gefahren werden (65 Prozent), geben bereits 33 Prozent der interessierten Zielgruppen an, sie künftig auch als Transportmittel zur Arbeit oder Ausbildungsstätte nutzen zu wollen.

Zu diesen Ergebnissen kam das Forschungsunternehmen Gfk (Growth from Knowledge“ nach Auswertung ihrer Umfrage.

Quelle verändert nach: https://www.gfk.com/de/insights/press-release/e-scooter-freizeitspass-oder-new-mobility/

Arbeitsmaterial 6: Bewertung Studie

Dein Kürzel (BS = Buchstabe):

3. BS deines Vornamens	Geburtstag (nur Tag)	1. BS deiner Straße

Aufgabe (Einzelarbeit):

Die Studie soll reproduziert werden, um ihr Ergebnis zu überprüfen. Überlege und plan in Gedanken die erneute Durchführung der Studie.

Bewerte die **Reproduzierbarkeit** der Studie. **Begründe** deine Antwort.

Fazit: Welche Bewertung würdest du der Studie geben?

Reproduzierbar:	Nicht-Reproduzierbar:
☐	☐

B.5.3 Phase IV und V

Abbildung B.5.9: *Arbeitsblätter zur Anwendung der Prüfkriterien Randomisierung und Korrelation und Kausalität auf eine (Meta-)Studie, die den Effekt von Magnesium auf Wadenkrämpfe bei schwangeren Frauen untersucht hat.*

WWU MÜNSTER | Code | Chemie Didaktik Münster

Der Effekt von Magnesium auf Wadenkrämpfe

– eine klinische Studie –

Autor*innen:

W. Bartl und *P. Riss*

Veröffentlicht in der Fachzeitschrift:

Zeitschrift „Magnesium-Bulletin“

im Jahr 1984, Heft 6, Seite 60-62

Fragestellung:

Welchen Effekt haben Magnesium-Tabletten auf Wadenkrämpfe?

Methoden:

- Die Forscher arbeiteten in einer Klinik in Wien. Auf einer Station in dieser Klinik untersuchten die Forscherinnen und Forscher **67 schwangere Frauen.** Diese Frauen kamen sowieso regelmäßig in die Klinik.

Stichprobe

67 schwangere Frauen

- **Vor der Therapie** wurden bei allen Frauen der Magnesium-Gehalt im Blut gemessen:

1. Messung
Magnesium-Gehalt
im Blut

- Eine Hälfte der Stichprobe bekam dann 2-mal am Tag eine Magnesium-Tablette

34 Frauen

33 Frauen

- **Nach 4 Wochen** wurde erneut gemessen:

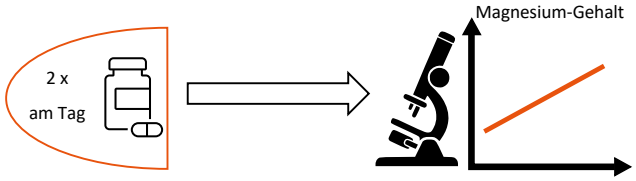
2. Messung
Magnesium-Gehalt
im Blut
nach 4 Wochen Therapie

WWU MÜNSTER | Code | Chemie Didaktik Münster

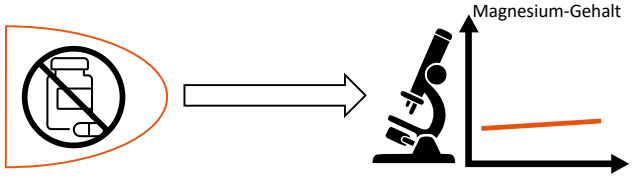
Ergebnisse:

Nach 4 Wochen Therapie

- Frauen, die **Tabletten** bekamen, hatten am Ende einen **höheren Magnesium-Gehalt** im Blut als vorher.

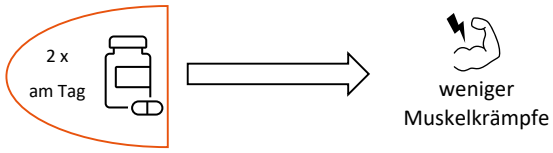


- Frauen, die **keine Tabletten** bekamen, hatten am Ende den gleichen Magnesium-Gehalt im Blut wie vorher.



Schlussfolgerung:

Magnesium-Tabletten erhöhen den Magnesium-Gehalt im Blut und reduzieren dadurch Wadenkrämpfe.



WWU MÜNSTER | Code | Chemie Didaktik Münster

Aufgabe Basis

Lies die Studie „Effekte von Magnesium auf Wadenkrämpfe“ genau durch.

Prüfe diese Studie gemeinsam mit deinem Partner **auf Randomisierung**:

Erkläre, ob und wie bei dieser Studie randomisiert wurde.

_____

_____

_____

_____

_____

_____

Aufgabe ★1

Zeichne mit dem großen Stift in das Prüfschema ein, welchen Weg die Studie nimmt.

Aufgabe ★2

Beschreibe, welche Abzweigungen die Studie auf dem Weg durch das Prüfschema nimmt und **bewerte**: Wie wissenschaftlich ist die Studie?

_____

_____

_____

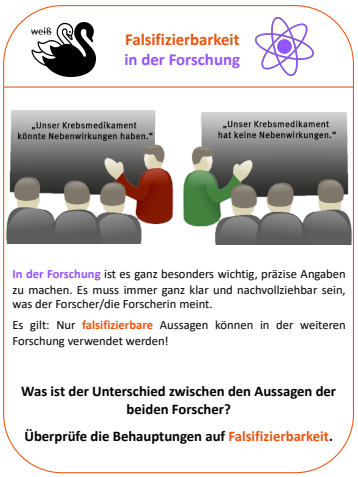
_____

_____

_____

B.6 Prüfkarten

Abbildung B.6.1: *Prüfkarten zur Erarbeitung des Prüfkriteriums Falsifizierbarkeit.*



B.7 Stundenverlaufspläne

Abbildung B.6.2: *Prüfkarten zur Erarbeitung des Prüfkriteriums Randomisierung in der Differenzierungsstufe Basis.*

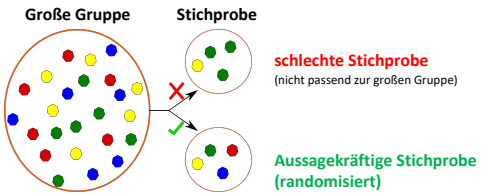
Was ist Randomisierung?

Wissenschaftliche Studien müssen **randomisiert** werden.
Das heißt: Versuchspersonen werden **zufällig** ausgewählt.

"Wie viele Personen in Deutschland hören Hip Hop und Rap?"

Möchtest du diese Frage untersuchen, kannst du natürlich nicht alle Deutschen befragen. Daher befragst du nur eine **Stichprobe**.

Diese Stichprobe muss **zufällig** gezogen werden. Nur dann passt sie zur großen Gruppe und ist **aussagekräftig**. Man spricht von einer **randomisierten** Stichprobe.



Wie erreicht man Randomisierung?

Forscher*innen benutzen diese Methoden,
um auf **zufällige** Weise Versuchspersonen zu finden:

Ein Zufallsgenerator wählt Personen aus Einwohnerliste
(oder Telefonbuch)

oder

Werbung für die Studie wird im Internet zufällig angezeigt

Für wissenschaftliche Studien gilt:
Forscher*innen müssen genau erklären,
wie die Stichprobe entstanden ist!

Prüfe Studien auf Randomisierung, bevor du den Ergebnissen vertraust!

Randomisierung der Heinsberg-Studie

Aufgabe Basis
Schau dir noch einmal die Heinsberg-Studie an.
Wird die Stichprobe genau erklärt?

❍ **ja** ❍ **nein**

Wird die Stichprobe **zufällig** ausgewählt?

❍ **ja** ❍ **nein**

2 x ja = **randomisiert**
1 x ja = **nicht randomisiert**
0 x ja = **nicht randomisiert**

Aufgabe *1
Kann man das Ergebnis der Studie auf ganz Deutschland übertragen?

❍ **ja** ❍ **nein**

Warum?

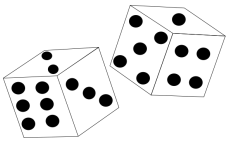
_____

_____

_____

_____

Abbildung B.6.3: *Prüfkarten zur Erarbeitung des Prüfkriteriums Randomisierung in der Differenzierungsstufe Stern 1.*



Randomisierung

★1

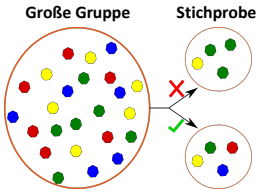
Was ist Randomisierung?

Wissenschaftliche Studien müssen **randomisiert** werden.
Das heißt: Versuchspersonen werden **zufällig** ausgewählt.

"Wie viele Personen in Deutschland hören Hip Hop und Rap?"

Möchtest du diese Frage untersuchen, kannst du natürlich nicht alle Deutschen befragen. Daher befragst du nur eine **Stichprobe**.

Diese Stichprobe muss **zufällig** gezogen werden. Nur dann ist sie **aussagekräftig**. Man spricht von einer **randomisierten** Stichprobe.



Wie erreicht man Randomisierung?

Forscher*innen benutzen diese Methoden,
um auf **zufällige** Weise Versuchspersonen zu finden:

Ein Zufallsgenerator wählt Personen aus Einwohnerliste
(oder Telefonbuch)

oder

Werbung für die Studie wird im Internet zufällig angezeigt

Für wissenschaftliche Studien gilt:
Forscher*innen müssen genau erklären,
wie die Stichprobe entstanden ist!

Prüfe Studien auf Randomisierung, bevor du den Ergebnissen vertraust!

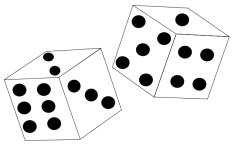
Randomisierung der Heinsberg-Studie

Aufgabe Basis
Schau dir noch einmal die Heinsberg-Studie an.
Wird die Stichprobe genau erklärt?

❍ **ja** ❍ **nein**

Aufgabe *1
Haben die Forscher/innen **Randomisierung** beachtet?
Diskutiere mit deinem Partner/deiner Partnerin und trage deine Erklärung hier ein.

Abbildung B.6.4: *Prüfkarten zur Erarbeitung des Prüfkriteriums Randomisierung in der Differenzierungsstufe Stern 2.*



Randomisierung

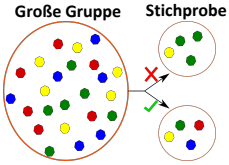
★2

Was ist Randomisierung?

Wissenschaftliche Studien müssen **randomisiert** werden. Das bedeutet, dass die Versuchspersonen **zufällig** (englisch: *random*) ausgewählt werden. Bei der Auswahl darf kein Muster erkennbar sein!

"Wie viele Personen in Deutschland hören Hip Hop und Rap?"

Möchtest du diese Frage untersuchen, kannst du natürlich nicht alle Deutschen befragen. Daher befragst du nur eine **Stichprobe**.



Die befragte Stichprobe muss **repräsentativ** (aussagekräftig) für die große Gruppe sein, auf die du deine Ergebnisse übertragen willst. Das erreichst du durch **Randomisierung.**

Nur randomisierte Stichproben sind repräsentativ.

Stell dir vor, in deinem Freundeskreis hören 7/10 Leuten Hip Hop. Hochgerechnet auf Deutschland wären das etwa **58 Mio**. Personen! Die VuMA-Studie zeigt: Tatsächlich hören „nur" etwa **6 Mio.** Deutsche Hip Hop.

Dein Freundeskreis besteht nur aus **jungen Menschen**. Er ist **keine randomisierte Stichprobe** und **nicht repräsentativ** für ganz Deutschland.

Wie erreicht man Randomisierung?

Forscher*innen benutzen diese Methoden, um auf **zufällige** Weise Versuchspersonen für ihre Stichprobe zu finden:

Ein Zufallsgenerator wählt Personen aus Einwohnerliste
(oder Telefonbuch)

Werbung für die Studie wird im Internet zufällig angezeigt

Besondere Merkmale beachten
damit die Stichprobe zur großen Gruppe passt

Für wissenschaftliche Studien gilt:
Forscher*innen müssen genau erklären, wie die Stichprobe entstanden ist!

Prüfe Studien auf Randomisierung, bevor du den Ergebnissen vertraust!

Randomisierung der Heinsberg-Studie

Aufgabe ★ 1
Schau dir noch einmal die Heinsberg-Studie an. Haben die Forscher*innen **Randomisierung** beachtet? Diskutiere mit deinem Partner/deiner Partnerin und trage deine Erklärung hier ein.

Aufgabe ★ 2
Die Ergebnisse können auf die gesamte Gemeinde Gangelt übertragen werden, da die Stichprobe **repräsentativ** für die Gemeinde ist. Warum ist die Studie aber **nicht repräsentativ** für ganz Deutschland?

Abbildung B.6.5: *Prüfkarten zur Erarbeitung des Prüfkriteriums Korrelation und Kausalität.*

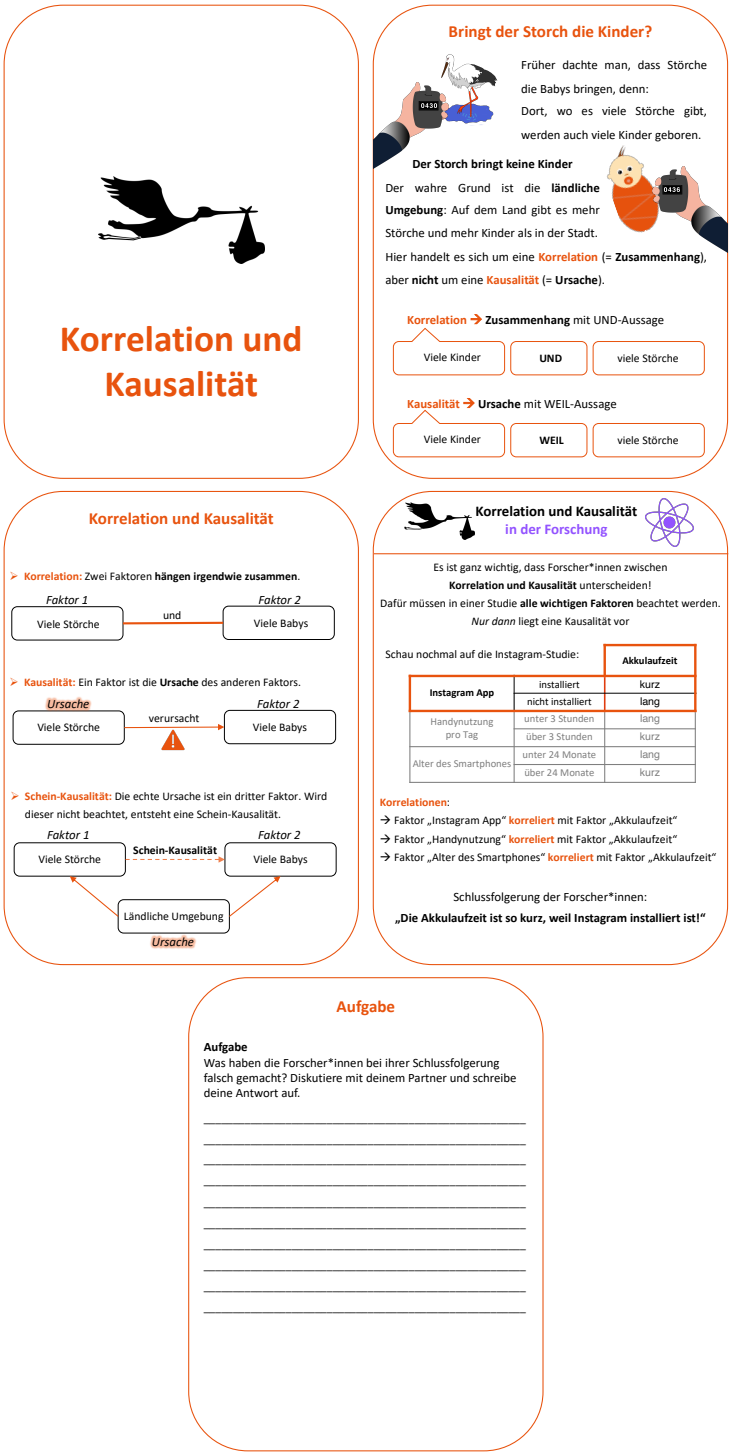


Abbildung B.6.6: *Prüfkarten zur Erarbeitung des Prüfkriteriums Korrelation und Kausalität.*

Reproduzierbarkeit

Was ist Reproduzierbarkeit?

Experimente und Studien können beliebig oft **wiederholt** werden. Um sie als **wissenschaftliche** Experimente und Studien zu bezeichnen, müssen ihre Ergebnisse dabei aber auch **reproduzierbar** sein.

Das Ergebnis wird auf Reproduzierbarkeit geprüft, indem die Untersuchung erneut durchgeführt wird:

Die Untersuchung wird unter **gleichen** Bedingungen und **gleicher** Durchführung von einem **anderen** Bearbeiter **wiederholt.**

Das **gleiche Ergebnis** wird gewonnen.
Die Untersuchung ist reproduzierbar.

Es wird nicht **das gleiche Ergebnis** gewonnen.
Die Untersuchung ist nicht-reproduzierbar.

Um die Wissenschaftlichkeit einer Untersuchung zu überprüfen, muss ihr Ergebnis jederzeit und überall reproduziert werden können!

Wie schafft man Reproduzierbarkeit?

Um die Reproduzierbarkeit einer Untersuchung zu sichern, ist es wichtig alles sehr genau aufzuschreiben und erneut zu prüfen:

Beispiel: Naturwissenschaftliches Experiment

1. Schritt: Alle **Einflussfaktoren** vor und während des Experiments müssen **genau** aufgeschrieben werden.

Fragen: *Welche Materialien benutze ich? Welche Mengen verwende ich? Bei welcher Temperatur arbeite ich?*

2. Schritt: Alle Schritte vor und während der **Durchführung** des Experiments müssen **genau** aufgeschrieben werden. Notiere neue **Änderungen**!

Fragen: *Welche Reihenfolge benutze ich? Wie viel Zeit benötige ich zwischen den Arbeitsschritten? Wie sieht der Aufbau des Experiments aus?*

Die Untersuchung muss jetzt mehrfach durchgeführt werden.

*Erhalte **Ich** immer wieder dasselbe Ergebnis? Erhält auch ein **anderer Bearbeiter** immer wieder dasselbe Ergebnis?*

Erst eine mehrfache Durchführung kann die Reproduzierbarkeit wirklich bestätigen!

Reproduzierbarkeit in der Chemie

In einer wissenschaftlichen Zeitschrift steht:

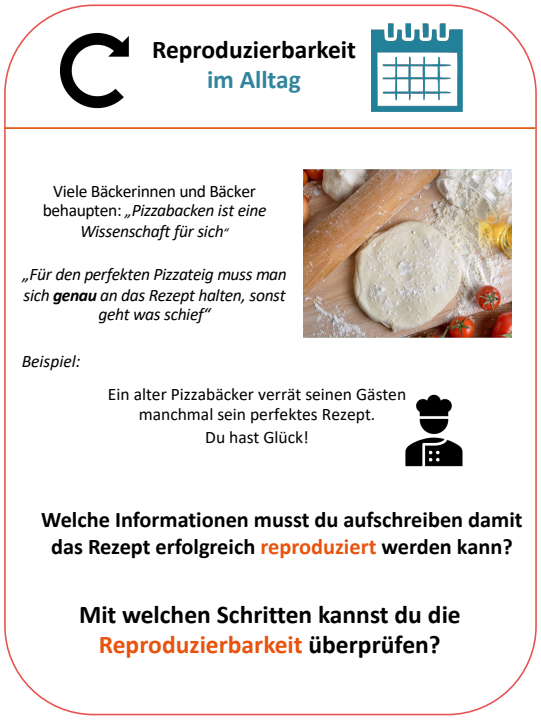
Ein neuer Farbstoff wurde im Labor entwickelt!

„ Der chemischen Forschung ist es in einem Experiment gelungen, durch das Mischen von rotem Obst- oder Gemüsesaft mit Natronlauge eine gelbe Farbe herzustellen.“

Führe das Experiment selber durch und überprüfe die Reproduzierbarkeit der gelben Farbe!

Welche **Einflussfaktoren** für den Ausgang eines Experiments wichtig sind und beachtet werden müssen hängt immer von dem jeweiligen Versuch ab.

Kannst du das Experiment reproduzieren?



B.8 Fragebogen

Abbildung B.7.1: *Stundenverlaufsplan der Erprobungsstunde am 21.09.2020. Im ersten Teil der Doppelstunde wurden die Phasen I und II durchgeführt. Der zweite Teil beinhaltet die Erarbeitung des Prüfkriteriums Randomisierung.*

Stundenverlaufsplan „Paul und seine Muskelkrämpfe“
(21.09.2020, Gesamtschule Lotte-Westerkappeln)

	Phase/Zeit	**Inhalt**	**Sozialform**	**Medien/Material**
7:55	Begrüßung 5min	Begrüßung und Vorstellung Visualisierung der Einheit durch Poster	UG	 Poster
8:00	Einleiten der Arbeitsphase 10 min	Verteilen der iPads, Einloggen in WLAN, gemeinsames Lesen der Aufgabenstellungen	UG	Datei „Muskelkrämpfe bei Magnesium“ *Starten der „Lektion 1“* iPads
8:10	Arbeitsphase 10 min	Bearbeitung des Arbeitsblattes.	EA	Arbeitsblatt, iPads
8:20	 3 min	Abfrage, welche Schüler*innen welche Antwort gewählt haben	UG	iPads Tafel
8:23	Besprechung und Diskussion 5-10min	Umfrageergebnisse an der Tafel (/Beamer) sammeln; geben die Einleitung zur Diskussion. Fiel euch die Entscheidung leicht?	UG	Tafel oder Beamer mit AppleTV
8:33	 5 min	Abbruch der Diskussion nach einigen Minuten, Überleitung zu Prüfschema **Vorstellung Prüfschema (rechte Seite)** „Studien“ anpinnen; dadurch die Vorgehensweise demonstrieren	UG UG	 Prüfschema-Poster

		Verdeutlichen, dass wir jetzt von der Magnesium-Problematik weggehen		
	Zweiter Teil			
8:38	Einleiten der Arbeitsphase 5min	AB2 wird digital verteilt. Schüler/innen lesen im Plenum den Text vor: Bis „Methode“. Ab dort EA	UG EA	iPad mit AB2
8:43	Arbeitsphase 10 min	Schüler/innen füllen Tabelle aus	PA; Flüster-regel	iPad mit AB2
8:53	Arbeitsphase 17 min	AB3 (Prüfkarten) wird digital verteilt. Schüler/innen bearbeiten Aufgaben	PA (Diff. nach *1/*2)	iPad mit AB3 (Prüfkarten)
9:10	Sicherung 10 min	Ergebnisse der Arbeitsphase werden im Plenum diskutiert. Impuls: **Haben die Forscher/innen in der Heinsbergstudie Randomisierung beachtet?** → Ja, die Stichprobe ist repräsentativ für Gangelt → Hochrechnung auf Deutschland nicht repräsentativ, aber das schreiben die Forscher auch!		
9:20	Anwendung, Abschluss 5 min	Überschrift aus der Bild-Zeitung wird gezeigt („Möglicherweise 1,8 Mio Corona-Infizierte“) Arbeitsauftrag (Plenum): Bewertet diese Überschrift. Impuls: Die Forscher/innen haben das in ihrer Studie richtig beschrieben! Die Presse hat die Studie falsch ausgelegt.	UG	Beamer

Abbildung B.7.2: *Stundenverlaufsplan der Erprobungsstunde am 28.09.2020 zur Erarbeitung des Prüfkriteriums Korrelation und Kausalität.*

Stundenverlaufsplan „Akkufresser Instagram“
(28.09.2020)

Zeit	Phase/Dauer	Inhalt	Sozialform	Medien/Material
7:55	Begrüßung, Wiederholung 10 min	Begrüßung, Einloggen der Tablets ins WLAN Wiederholung von Randomisierung - Murmelphase 5 min: Was ist Randomisierung? Tauscht euch mit eurem Partner aus und schreibt eure Erklärung in eigenen Worten auf das AB.	UG PA AB 0	 Tablets
8:05	Besprechen der Wiederholungsaufgabe 10 min	LP projiziert AB0; Schüler*innen präsentieren ihre Antworten - Was ist Randomisierung? - Könnt ihr euch an die Studie vom letzten Mal erinnern? - In Gangelt wurde eine Stichprobe gezogen - Zufallsverfahren Bitte ergänzt die Lösung bei euch LP schreibt stichpunktartig mit	UG	Beamer, Tablet
8:15	Einführung des Prüfschemas 3 min	Bitte schaut auf dieses Schema. Das ist eine Checkliste mit der wir Studien prüfen können. Wenn wir komplett prüfen wollen, dann alle Kriterien! Randomisierung: fertig	UG	Prüfschema-Poster
	Problemaufriss	Zur Bewertung von Studien werden weitere Kriterien benötigt:		

		Heute **Korrelation & Kausalität** Wir prüfen heute eine Studie zu Instagram. Dazu behandeln wir nun die Studie und danach lernt ihr Korrelation & Kausalität kennen.		
8:18	Einleiten der Erarbeitung 7 min	AB1 wird digital auf die Tablets verteilt Lesen im Plenum bis Aufgabe 1 LP gibt Hinweise zur Bearbeitung. Zettel wurde ausgeteilt, dort sind alle Anweisungen, wie ihr mit der Tabelle umgehen müsst. Ich zeige euch einmal, wie man einen Filter setzt.	 UG	Tablets, AB1 (digital) Beamer
8:25	Erarbeitung 20 min	Schüler/innen bearbeiten Aufgabe 1 und nutzen dafür die Numbers-Tabelle (AB2) → Arbeitsauftrag nur auf Papier	PA	Tablets, AB1, AB2 (digital)
8:45	Zwischen-sicherung 5 min	Sammeln der Ergebnisse: Tabellenwerte Was habt ihr denn hier eingetragen? (von oben links nach unten rechts) Die anderen ergänzen die Werte bei sich Was bedeutet diese Zeile? -› daneben schreiben „kurze Laufzeit bei installierter App“ „kurze Laufzeit bei langer Nutzung“	UG	Projektion an Beamer (?)

		„Kurze Laufzeit bei altem „Handy“		
8:50	Zwischen-sicherung ★1 5 min	Hat jemand ★1 geschafft? Was habt ihr da geschrieben?	UG	
	2. Stunde			
8:55	Einleiten der Arbeitsphase	Überleitung zu den Begriffen Korrelation+Kausalität; AB3 ankündigen Alle Seiten lesen und die Aufgabe am Ende bearbeiten. Besprecht euch gerne mit eurem Partner!	UG	AB3
	Erarbeitung 15 min	Schüler/innen lesen Prüfkarten (AB3) und bearbeiten die Aufgabe.	PA	
9:10	Sicherung, Diskussion, Abschluss… 8min	Was haben die Forscher*innen falsch gemacht? -› nicht alle Faktoren beachtet -› nur auf die App geachtet -› Schein-Kausalität aufgestellt -› so getan, als wäre es eine Kausalität -› Zusammenhang zwischen App und Akkulaufzeit ist nur eine Korrelation	UG	Beamer, Tablet
9:18	Sicherung	Was habt ihr heute gelernt?	UG	

Abbildung B.7.3: *Stundenverlaufsplan der Erprobungsstunde zur Falsifizierbarkeit.*

WWU MÜNSTER — Chemie Didaktik Münster

Stundenentwurf zum Thema „Falsifizierbarkeit“

Phase	**Unterrichtsgeschehen**	**Sozial-form**	**Medien/Material**
Unterrichts-einstieg 2 min	LP stellt sich vor, einleitende Worte zum Thema	UG	
Problemge-winnung 15 min	Nüsse und Nutella auf Pult stellen AB 1 wird ausgeteilt SuS lesen AB, bearbeiten Aufg. 1	PA	Nüsse, Nutella Arbeitsblatt 1
Problemge-winnung 10 min 5 min	LP fordert Plenum auf, Ergebnisse vorzustellen (einzelne SuS, Wortmeldungen) LP hält Ergebnisse mündlich fest; Experimente werden nicht durchgeführt, nur kleinste Untersuchungen Fettgehalte werden überprüft	UG	Nüsse, Nutella (Klassensatz)
Überlegungen zur Prob-lemlösung 15 min	Austeilen des AB 2 SuS lesen AB und bearbeiten Aufg 1+2	PA	Arbeitsblatt 2
Vorstellen der Ergebnisse 10 min	SuS stellen Ergebnisse vor (mündlich, keine schriftliche Sicherung) LP interveniert mit gezielten Einwänden, Fragen etc.	UG	
Erarbeitung und Abstraktion 10 min	Austeilen des AB 3 SuS lesen und bearbeiten AB 3		Arbeitsblatt 3

© C. Kralisch, Prof. Dr. A. Marohn: choice²reflect; Institut für Didaktik der Chemie, Münster

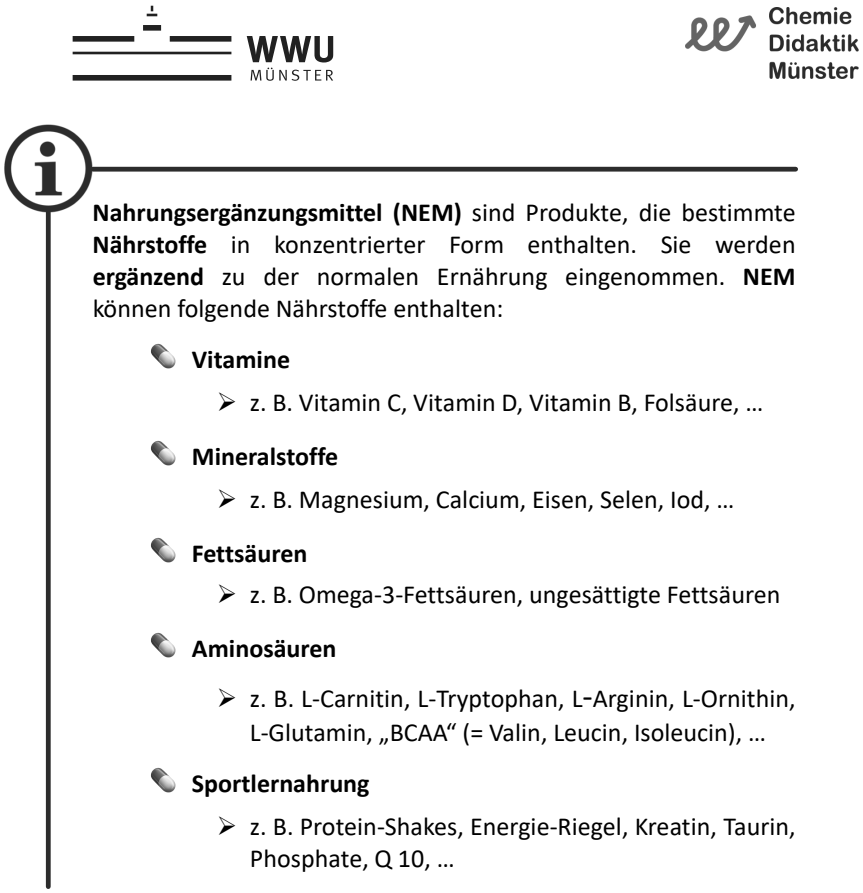
Abbildung B.7.4: *Stundenverlaufsplan der dritten Stunde der Gesamterprobung. Inhalte der Stunde sind Anwendung und Reflexion der Wissenschaftskriterien.*

Stundenverlaufsplan „Effekte von Magnesium auf Muskelkrämpfe“
(09.10.2020, Gesamtschule Lotte-Westerkappeln)

Zeit	**Phase/Dauer**	**Inhalt**	**Sozial-form**	**Medien/Material**
11:30	Begrüßung Wiederholung 10 min	Begrüßung Wiederholungsaufgabe: Tabu Begriffe: - Randomisierung - Repräsentativ - Scheinkausalität - Dunkelziffer	UG, einzelne Schüler vorne	Kärtchen
11:40	Problemgewinnung 2 min	Meinungsbild der ersten Stunde projizieren Wir konnten uns nicht entscheiden und hatten gesagt, dass wir Studien prüfen müssen. Ihr kennt nun 2 „Kriterien“ (!). **(Prüfschema umdrehen)** Eigentlich müssten wir alle anderen auch testen, um sicher sagen zu können, wie wissenschaftlich die Studie ist. Jetzt wollen wir eine echte Studie zu Magnesium prüfen. Studie (die von Meta-Studie ausgeschlossen wurde) austeilen	UG	Beamer+Tablet

11:42	Erarbeitung 15 min	Bitte schaut zuerst auf das Aufgabenblatt. Wer liest die Aufgaben vor? Schüler bearbeiten die Aufgaben Basis, Stern1 und Stern2.	PA	AB2 (3 Seiten)
11:57	Sammeln der Ergebnisse 8 min	Das Aufgabenblatt wird projiziert. LP schreibt Antworten der Schüler mit. - Die Stichprobe wird nicht zufällig ausgewählt, nur die Frauen aus einer Klinik - Die Stichprobe ist nur repräsentativ für alle Schwangeren. Nicht für alle Menschen	UG	Beamer+Tablet AB2 projizieren
12:05	Auflösung: Magnesium gegen Krämpfe	PowerPoint Präsentation zum Review	UG	Beamer+Tablet PPP
12:10	(Kurzes Blitzlicht?)	Jeder sagt kurz: würdet ihr Magnesium Tablette oder Banane nehmen und warum?		

Abbildung B.8.1: *Fragebogen zur Erhebung der Verwendungsgewohnheiten von Nahrungsergänzungsmitteln bei Schülerinnen und Schülern der 8. bis 10. Klasse.*



WWU MÜNSTER

Chemie Didaktik Münster

Nahrungsergänzungsmittel (NEM) sind Produkte, die bestimmte **Nährstoffe** in konzentrierter Form enthalten. Sie werden **ergänzend** zu der normalen Ernährung eingenommen. **NEM** können folgende Nährstoffe enthalten:

- **Vitamine**
  - z. B. Vitamin C, Vitamin D, Vitamin B, Folsäure, …
- **Mineralstoffe**
  - z. B. Magnesium, Calcium, Eisen, Selen, Iod, …
- **Fettsäuren**
  - z. B. Omega-3-Fettsäuren, ungesättigte Fettsäuren
- **Aminosäuren**
  - z. B. L-Carnitin, L-Tryptophan, L-Arginin, L-Ornithin, L-Glutamin, „BCAA“ (= Valin, Leucin, Isoleucin), …
- **Sportlernahrung**
  - z. B. Protein-Shakes, Energie-Riegel, Kreatin, Taurin, Phosphate, Q 10, …

Seite 1

WWU MÜNSTER

Chemie Didaktik Münster

1. Wo sind dir **NEM** bereits begegnet?

2. Was bewirken **NEM**?

3. Falls du NEM nimmst: Welche **NEM** hast du schon einmal eingenommen? Auf dem Info-Blatt findest du Nährstoffe, die in **NEM** enthalten sein können. Bitte trage in die Tabelle links den Produktnamen ein, rechts die Häufigkeit der Einnahme (nur ein Kreuz pro Zeile). Solltest du den genauen Namen nicht kennen, reicht eine ungefähre Bezeichnung (z. B.: „Vitamin E-Tablette“).

Produkt	**Häufigkeit der Einnahme**			
	❍ täglich	❍ mehrmals pro Woche	❍ einmal pro Woche	❍ nach Bedarf
	❍ täglich	❍ mehrmals pro Woche	❍ einmal pro Woche	❍ nach Bedarf
	❍ täglich	❍ mehrmals pro Woche	❍ einmal pro Woche	❍ nach Bedarf
	❍ täglich	❍ mehrmals pro Woche	❍ einmal pro Woche	❍ nach Bedarf
	❍ täglich	❍ mehrmals pro Woche	❍ einmal pro Woche	❍ nach Bedarf

4. Falls du NEM nimmst: Woher bekommst du/bekommt deine Familie die **NEM**?
(mehrere Kreuze möglich)

❑ Apotheke ❑ Drogeriemarkt ❑ Supermarkt
❑ Internethandel ❑ Weiß nicht
❑ Sonstige: _____

Seite 2

WWU MÜNSTER

Chemie Didaktik Münster

5. Welche Wirkung erhoffst du dir von den **NEM**, die du nimmst?

Produkt	**Wirkung**

6. Wo informierst du dich, welche **NEM** für dich sinnvoll sind? (mehrere Kreuze möglich)

❑ Bei meinem Arzt
❑ Bei meinen Eltern
❑ Bei meinen Freunden
❑ Bei meinem Trainer
❑ In der Werbung
❑ In Büchern/Zeitschriften
❑ In der Schule

☞ Im Unterrichtsfach: _____

❑ In den sozialen Medien (Youtube, Instagram, Facebook, …)

☞ Wo genau? _____

❑ Woanders: _____

7. Wie stark beeinflussen die folgenden Personen/Faktoren, ob du NEM einnimmst?
(Bitte benutze *nur die Kreise* und kreuze nicht zwischen den Kreisen an)

	Kein Einfluss ------|----------|------			**Starker Einfluss**	**Keine Angabe**
Arzt	❍	❍	❍	❍	❍
Eltern	❍	❍	❍	❍	❍
Freunde	❍	❍	❍	❍	❍
Werbung	❍	❍	❍	❍	❍
Trainer	❍	❍	❍	❍	❍
Soziale Medien (Youtube, Instagram, Facebook, …)	❍	❍	❍	❍	❍
Weitere Personen:					
_____	❍	❍	❍	❍	❍
_____	❍	❍	❍	❍	❍

WWU MÜNSTER

Chemie Didaktik Münster

8. Die folgende Auflistung enthält verschiedene Aussagen. Bitte kreuze an, wie sehr du der jeweiligen Aussage *zustimmst.* Beziehe dich dabei auf einen durchschnittlich gesunden Menschen in Deutschland.

	stimme überhaupt nicht zu	------|----------|------			**stimme voll und ganz zu**	**Weiß nicht**
NEM können die Gesundheit verbessern.	❍	❍	❍	❍		❍
NEM können die Fitness verbessern.	❍	❍	❍	❍		❍
NEM verhindern, dass es dem Körper an Nährstoffen mangelt.	❍	❍	❍	❍		❍
Je mehr NEM der Körper bekommt, desto besser ist es für ihn.	❍	❍	❍	❍		❍
NEM können dem Körper nicht schaden.	❍	❍	❍	❍		❍
NEM sind für mich Medikamente.	❍	❍	❍	❍		❍
Man kann den Bedarf an Vitaminen und Mineralstoffen allein durch eine gesunde Ernährung decken.	❍	❍	❍	❍		❍
Zusätzlich zur gesunden Ernährung sollte man auch NEM nehmen.	❍	❍	❍	❍		❍
Wenn man sich mal ungesund ernährt, lässt sich das durch NEM ausgleichen.	❍	❍	❍	❍		❍
Die meisten Deutschen haben einen Vitaminmangel.	❍	❍	❍	❍		❍
NEM sind für mich Lebensmittel.	❍	❍	❍	❍		❍

9. Bewerte auf der Skala: Aus deiner Sicht sind Nahrungsergänzungsmittel…

						Weiß nicht
ungesund	❍	❍	❍	❍	gesund	❍
gefährlich	❍	❍	❍	❍	ungefährlich	❍
teuer	❍	❍	❍	❍	günstig	❍
sinnlos	❍	❍	❍	❍	sinnvoll	❍

Seite 4

WWU MÜNSTER

Chemie Didaktik Münster

Geschlecht: ❍ weiblich ❍ männlich ❍ keine Angabe

Jahrgang: ❍ 7 ❍ 8 ❍ 9 ❍ 10 ❍ 11 ❍ 12

Alter: _____

Wie häufig treibst du welchen Sport? Trage links die Sportart ein und kreuze rechts die Häufigkeit an (Radfahren zur Schule brauchst du nicht aufschreiben).

Sportart	Häufigkeit
	❍ 0-1 h/Woche ❍ 1-2 h/Woche ❍ 2-4 h/Woche ❍ >4 h/Woche
	❍ 0-1 h/Woche ❍ 1-2 h/Woche ❍ 2-4 h/Woche ❍ >4 h/Woche
	❍ 0-1 h/Woche ❍ 1-2 h/Woche ❍ 2-4 h/Woche ❍ >4 h/Woche
	❍ 0-1 h/Woche ❍ 1-2 h/Woche ❍ 2-4 h/Woche ❍ >4 h/Woche
	❍ 0-1 h/Woche ❍ 1-2 h/Woche ❍ 2-4 h/Woche ❍ >4 h/Woche

Vielen Dank für deine Teilnahme!

Ch. Kralisch

Christopher Kralisch

Christopher Kralisch
Institut für Didaktik der Chemie
Fliednerstr. 21
48149 Münster

Tel.: 0251-83-39479
Mail: c.kralisch@wwu.de

Abbildungsverzeichnis

3.1 Das Modell ethischer Urteilskompetenz nach Reitschert et al. (2007). . 42
3.2 Das Modell epistemischer Kompetenz nach Kötter & Hammann (2018). 46
3.3 Gegenüberstellung der Modelle ethischer Urteilskompetenz (Reitschert et al. 2007) und epistemischer Kompetenz (Kötter & Hammann 2018). 47

4.1 Darstellung der fünf Phasen von choice^{2}reflect (Jungkamp 2021; Kralisch & Marohn o.D.). 65

5.1 Strukturmodell design-basierter Forschung, verändert nach Rohrbach-Lochner (2019, S. 31). 78

8.1 Übersicht über alle durchgeführten Erprobungen. Horizontale Achse bildet den zeitlichen Projektverlauf ab. QIA = Qualitative Inhaltsanalyse.108
8.2 Darstellung der Sitzpositionen der Schüler:innen in der Doppelstunde am 12.09.2017 zur Falsifizierbarkeit und Darstellung der verwendeten technischen Geräte zur Datenerhebung. 110
8.3 Darstellung der Sitzpositionen der Schüler:innen in der ersten (oben) und zweiten (unten) Doppelstunde am 21. bzw. 28.09.2020 und der verwendeten technischen Geräte zur Datenerhebung. 113

9.1 Arbeitsblätter 1 (links) und 2 (rechts) zur Erarbeitung des Prüfkriteriums Falsifizierbarkeit (Kralisch 2017). 122
9.2 Arbeitsblätter 3 (links) und 4 (rechts) zur Erprobung des Prüfkriteriums Falsifizierbarkeit (Kralisch 2017). 124
9.3 Prüfkarten zum Prüfkriterium Falsifizierbarkeit. 125
9.4 Code-Matrix-Browser zur Visualisierung der Codierungen für Arbeitsblatt 1. 128

9.5 Code-Matrix-Browser zur Visualisierung der Codierungen für Arbeitsblatt 2. 129

9.6 Code-Matrix-Browser zur Visualisierung der Codierungen in den Audioaufnahmen der Unterrichtsstunde. 130

9.7 Arbeitsblatt 1 zur Erarbeitung des Prüfkriteriums Korrelation und Kausalität. 137

9.8 Ausschnitt aus den Rohdaten der Befragung. Dargestellt werden die konstruierten Daten von acht Versuchspersonen. 138

9.9 Ergebnistabelle der Analyse von Rohdaten durch Setzen von Schnellfiltern in der App „Numbers“ mit bereits eingetragenen Ergebnissen. . 139

9.10 Arbeitsblatt 1 zur Erarbeitung des Prüfkriteriums Randomisierung. . 172

9.11 Prüfkarten zur Erarbeitung des Prüfkriteriums Randomisierung in der Differenzierungsstufe Stern 2 (Seiten 1 und 2). 173

9.12 Prüfkarten zur Erarbeitung des Prüfkriteriums Randomisierung in der Differenzierungsstufe Stern 2 (Seiten 3 und 4). 174

9.13 Schlagzeile der BILD-Zeitung am 04.05.2020 als Reaktion auf die Veröffentlichung der Heinsberg-Studie. 176

9.14 Arbeitsblätter 1 (links) und 2 (rechts) zur Erarbeitung des Prüfkriteriums Reproduzierbarkeit. 203

9.15 Arbeitsblätter 3a (links) und 3b (rechts) zur Erarbeitung des Prüfkriteriums Reproduzierbarkeit. 205

9.16 Arbeitsblätter 4 (oben) und 5 (unten) zur Erarbeitung des Prüfkriteriums Reproduzierbarkeit. 207

9.17 Code-Matrix-Browser zur Visualisierung der Codierungen für Arbeitsmaterial 3a des Prüfkriteriums Reproduzierbarkeit. 209

9.18 Code-Matrix-Browser zur Visualisierung der Codierungen für Arbeitsmaterial 3b des Prüfkriteriums Reproduzierbarkeit. 210

9.19 Code-Matrix-Browser zur Visualisierung der Codierungen für Arbeitsmaterial 6 des Prüfkriteriums Reproduzierbarkeit. 212

10.1 Darstellung der relativen Anteile der Antwortmöglichkeiten 0 bis 3 bei Item 8 des Fragebogens. Fettgedruckte Linie als Abgrenzung zwischen Zustimmung und Ablehnung. Dicht gepunktet = „Stimme voll und ganz zu“, leicht gepunktet = „Stimme eher zu“, leicht schraffiert = „Stimme eher nicht zu“, dicht schraffiert = „Stimme überhaupt nicht zu“.229

10.2 Darstellung der relativen Anteile der Antwortmöglichkeiten 1 bis 4 bei Item 9 des Fragebogens. Fettgedruckte Linie als Abgrenzung zwischen dem linken und rechten Begriff. Dicht gepunktet = Option 1, leicht gepunktet = Option 2, leicht schraffiert = Option 3, dicht schraffiert = Option 4. 230

10.3 Ausschnitt aus der Video-Datei „Gruppenchat“, die die Lernenden im Rahmen des Fachspezifischen Problemaufrisses verwenden. 243

11.1 Ausschnitt aus dem Prüfschema. Gezeigt wird die linke Seite mit aussagenlogischer Prüfung auf Falsifizierbarkeit. 267

11.2 Darstellung früherer Versionen der linken Seite des Prüfschemas. Erkennbar ist ein starker Wandel der Verwendung der Trichter-Symbolik.268

11.3 Ein Prüfschema zur Bewertung kontroverser Themen in den Phasen IV und V der Unterrichtskonzeption choice^{2}reflect , Version 1. 270

11.4 Ablaufplan der Seminarsitzungen im Juli 2021 zur Erprobung des Prüfschemas. 272

11.5 Schlagzeile und Einleitung des in der Erprobung verwendeten Artikels zum Thema Veganismus. Quelle des Artikels: Landswerk (2021). . 274

11.6 Aufgabenstellung der Primäraufgabe in der Erprobung des Prüfschemas. 275

11.7 Ausschnitt aus dem in MAXQDA deduktiv-induktiv erstellen Kategoriensystem für das Prüfkriterium Falsifizierbarkeit. 275

11.8 Code-Matrix-Browser der Kodierung von Fragebogen-Antworten zur Beantwortung von Subfrage 1. Jede Spalte stellt die Antworten einer Person dar. 279

11.9 Darstellung der Anteile zielführender und nicht-zielführender Antworten auf die Fragebogen-Items. 1 = Lerngruppe 1 (erster Durchlauf). 2 = Lerngruppe 2 (zweiter Durchlauf). Diagonal schraffiert = zielführende Antworten. Dicht gepunktet = nicht-zielführende Antworten. Leicht gepunktet = keine Antwort. Fettgedruckte Linie markiert Grenze zwischen zielführenden und nicht-zielführenden Antworten. 280

11.10 Code-Matrix-Browser der Kodierung von Gruppenvideos zur Beantwortung von Subfrage 2. Jede Spalte stellt eine Gruppe von Studierenden (zwei Studierende pro Gruppe) dar. 283

A.0.1 Tabellarische Übersicht der Definitionen und Konzeptualisierungen von Bewertungskompetenz in verschiedenen schulischen Lehrplänen, bezugnehmend auf das Kapitel 3.4.2. 348

B.1.1 Kategoriensystem für das Prüfkriterium Falsifizierbarkeit. 352

B.1.2 Vollständiges Kategoriensystem zur Codierung der zweiten Doppelstunde der Gesamterprobung. 354

B.1.3 Kategoriensystem zur Analyse der Fragebogen-Antworten in der Erprobung des Prüfschemas. 370

B.1.4 Kategoriensystem zur Analyse der Gruppen-Meetings in der Erprobung des Prüfschemas. 371

B.2.1 Auflistung aller in der zweiten Doppelstunde der Gesamterprobung identifizierten Impulse mit Angabe antizipierter sowie gegebener Antworten. * = Passung. ** = Anschlussfähigkeit. 373

B.2.2 Darstellung der identifizierten Lehrimpulse in der ersten Doppelstunde der Gesamterprobung am 21.09.2020. * = Passung. ** = Anschlussfähigkeit. 377

B.3.1 Codeline-Darstellung der Arbeitsphasen-Auswertung zum Prüfkriterium *Korrelation und Kausalität* für die Gruppen 1-5. 383

B.3.2 Codeline der Arbeitsphasen-Auswertung zum Prüfkriterium *Korrelation und Kausalität* in der Gesamtdarstellung aller Gruppen bezogen auf die Codes *Verständnisschwierigkeiten*, *Aha-Erlebnisse* und *Off-Task*. 384

B.3.3 Codeline-Darstellung der Arbeitsphasen-Auswertung zum Prüfkriterium *Randomisierung* für die Gruppen 1-5. 385
B.3.4 Codelines der Arbeitsphasen-Auswertung zum Prüfkriterium *Randomisierung* in der Gesamtdarstellung aller Gruppen bezogen auf die Codes *Verständnisschwierigkeiten*, *Aha-Erlebnisse* und *Off-Task*. 386
B.4.1 Auflistung von Einzelaussagen aus der Erprobung zum Prüfkriterium Falsifizierbarkeit, die als relevant für die Datenanalyse angesehen werden. 387
B.4.2 Transkribierte Antworten der Wiederholungsaufgabe in der zweiten Doppelstunde der Gesamterprobung. 388
B.5.1 Erstellte Instagram-Posts zu a) den Inhaltsstoffen einer Banane und b) einem Magnesium-Präparat zur Verwendung im Fachspezifischen Problemaufriss und der Positionierung. 390
B.5.2 Arbeitsblatt 1 in der Unterrichtsstunde zu den Phasen I und II zur Problemfrage: Helfen Magnesium-Präparate gegen Muskelkrämpfe? . . . 391
B.5.3 Arbeitsblätter für die Unterrichtseinheit zur Falsifizierbarkeit. 392
B.5.4 Nach Falsifizierbarkeit zu bewertende Aussagen für die Unterrichtseinheit zur Falsifizierbarkeit. 396
B.5.5 AB 2 zur Erarbeitung von *Randomisierung*. Enthalten ist eine vereinfachte Zusammenfassung der Heinsberg-Studie nach Streeck et al. (2020)398
B.5.6 AB 0 zur Wiederholung der Begriffe *Randomisierung* und *repräsentativ*, eingesetzt in der zweiten Doppelstunde der Gesamterprobung. . . 399
B.5.7 Arbeitsblätter zur Erarbeitung des Prüfkriteriums *Korrelation und Kausalität*. 400
B.5.8 Arbeitsblätter zur Erarbeitung des Prüfkriteriums *Reproduzierbarkeit*. 403
B.5.9 Arbeitsblätter zur Anwendung der Prüfkriterien *Randomisierung* und *Korrelation und Kausalität* auf eine (Meta-)Studie, die den Effekt von Magnesium auf Wadenkrämpfe bei schwangeren Frauen untersucht hat.411
B.6.1 Prüfkarten zur Erarbeitung des Prüfkriteriums Falsifizierbarkeit. . . . 414
B.6.2 Prüfkarten zur Erarbeitung des Prüfkriteriums Randomisierung in der Differenzierungsstufe Basis. 416
B.6.3 Prüfkarten zur Erarbeitung des Prüfkriteriums Randomisierung in der Differenzierungsstufe Stern 1. 417

B.6.4 Prüfkarten zur Erarbeitung des Prüfkriteriums Randomisierung in der Differenzierungsstufe Stern 2. 418
B.6.5 Prüfkarten zur Erarbeitung des Prüfkriteriums Korrelation und Kausalität. 419
B.6.6 Prüfkarten zur Erarbeitung des Prüfkriteriums Korrelation und Kausalität. 420
B.7.1 Stundenverlaufsplan der Erprobungsstunde am 21.09.2020. Im ersten Teil der Doppelstunde wurden die Phasen I und II durchgeführt. Der zweite Teil beinhaltet die Erarbeitung des Prüfkriteriums Randomisierung. 422
B.7.2 Stundenverlaufsplan der Erprobungsstunde am 28.09.2020 zur Erarbeitung des Prüfkriteriums Korrelation und Kausalität. 424
B.7.3 Stundenverlaufsplan der Erprobungsstunde zur Falsifizierbarkeit. . . 427
B.7.4 Stundenverlaufsplan der dritten Stunde der Gesamterprobung. Inhalte der Stunde sind Anwendung und Reflexion der Wissenschaftskriterien. 428
B.8.1 Fragebogen zur Erhebung der Verwendungsgewohnheiten von Nahrungsergänzungsmitteln bei Schülerinnen und Schülern der 8. bis 10. Klasse. 430

Tabellenverzeichnis

3.1 Zusammenstellung aller im Kernlehrplan Chemie (KLP 2019) genannten konkretisierten Kompetenzerwartungen für den Kompetenzbereich Bewertung. IF = Inhaltsfeld. ÜK = übergeordnete Kompetenzerwartungen. 53

3.2 Zusammenstellung aller im Kernlehrplan Chemie (KLP 2019) genannten konkretisierten Kompetenzerwartungen für den Kompetenzbereich Bewertung der zweiten Progressionsstufe. IF = Inhaltsfeld. ÜK = übergeordnete Kompetenzerwartungen. 54

5.1 Vergleich der Charakteristika von Laborstudien und Design-Based Research nach Collins et al. (2004). 76

7.1 Übersicht aller in diesem Projekt durchlaufenen Mesozyklen innerhalb des Makrozyklus (Design-Experiment) sowie der zugehörigen Forschungsfragen. MiZ = Mikrozyklus. 93

8.1 Ablaufschema der Gesamterprobung im September 2020. Spaltenbreiten proportional zur Dauer der Unterrichtsstunden. 114

9.1 Auszug aus dem Kategoriensystem zur Analyse der drei Schwerpunkte in der Unterrichtseinheit zum Prüfkriterium *Korrelation und Kausalität*. 152

9.2 Auszug aus den mit *Aha-Erlebnis* codierten Segmenten aus der Unterrichtsstunde zum Prüfkriterium *Korrelation und Kausalität*. 156

9.3 Darstellung der mit *Aha-Erlebnis* codierten Segmente aus der Unterrichtsstunde zum Prüfkriterium *Randomisierung*. 188

9.4 Darstellung der als *besondere Aussagen* eingestuften Unterrichtssituationen in der Erprobung des Prüfkriteriums *Randomisierung*. 190

10.1 Mittelwerte (MW) und Standardabweichungen (St.abw) über alle Einschätzungen zu den Aussagen a. bis k. des Items 8. 4-stufige Skala: 0 = stimme überhaupt nicht zu, 1 = stimme eher nicht zu, 2 = stimme eher zu, 3 = stimme voll und ganz zu. 228
10.2 Mittelwerte (MW) und Standardabweichungen (St.abw) über alle Bewertungen der Begriffspaare a. bis d. des Items 9. 4-stufige endpunktverbalisierte Skala: 1 = linksstehender Begriff, 4 = rechtsstehender Begriff. 229
10.3 Darstellung aller als „unsicher“ eingestuften Antworten auf dem Arbeitsblatt zur Positionierung. Indikatoren für Unsicherheit sind fettgedruckt. Spalte * zeigt die von den Lernenden angekreuzte Option (Mg = Magnesium; Ban = Banane). 239
10.4 Darstellung der im Review von Garrison et al. (2020) beschriebenen exkludierten Studien mit Angabe des zugehörigen Prüfkriteriums. . . 249

11.1 Auflistung über die im Fragebogen verwendeten Aussagen zur Erklärung der Prüfkriterien. 273
11.2 Auflistung über die Reaktionen der Studierenden (erste Lerngruppe) auf das erstmalige Betrachten des Prüfschemas. 287

Danksagung

Von der Idee dieses Projekts über mannigfaltige theoretische Überlegungen bis zur praktischen Umsetzung haben mich viele Personen begleitet, unterstützt und ermutigt.

Allen voran meine Doktormutter Annette Marohn. Ich danke dir, dass du mir mit choice^{2}reflect das Tor zur Bewertungskompetenz im Chemieunterricht geöffnet hast und ich dadurch Teil der chemiedidaktischen Forschung werden durfte. Danke für die Unterstützung und Beratung in vielen ~~an~~netten Gesprächen, Tagungen und Momenten.

Ich danke Frau Prof.'in Dr. Katrin Sommer für die Übernahme des Zweitgutachtens sowie Frau Prof.'in Dr. Kirsten Schlüter und Frau Prof.'in Dr. Ulrike Weyland für Ihre Beteiligung in der Promotionskommission.

Die Zeit im Institut für Didaktik der Chemie verging wie im Flug –- Ich danke allen Mitgliedern des Instituts dafür, dass man sich stets auf euch und eine gelassene Arbeitsatmosphäre verlassen kann.

Danke an Marius für „Kaffee bereit". Natürlich nur dafür. Danke an Caro Banse und Theresa Flach für eure durchgängig überragende Bereitschaft, das Projekt voranzubringen und dabei Exzellenzleistungen abzuliefern. Vielen Dank an Joana für die Korrektur der Arbeit.

Ein großer Dank geht an Mama und Papa sowie an meine Schwester. Ich bin sehr froh, dass ihr immer für mich da seid. Es ist ein Geschenk, derart großen Rückhalt in der Familie finden zu dürfen.

Ich danke dir, Ina, dass du als meine Ehefrau für mich auch beste Freundin, Motivations-Coach und noch so viel mehr bist.

Bisher erschienene Bände der Reihe

Lernen in Naturwissenschaften

ISSN 2566-493X

1	Maria Egbers	Konzeptentwicklungs- und Gesprächsprozesse im Rahmen der Unterrichtskonzeption „choice2learn“ ISBN 978-3-8325-4510-9 53.00 EUR
2	Daniel Laumann	Magnetismus hoch 4 Fachliche Strukturierung und Entwicklung multipler Repräsentationen zum Magnetismus für die Hochschule ISBN 978-3-8325-4571-0 60.50 EUR
3	Friederike Rohrbach-Lochner	Design-Based Research zur Weiterentwicklung der chemiedidaktischen Lehrerausbildung zu Schülervorstellungen: Entwicklung und Evaluation eines an Forschendem Lernen orientierten Seminarkonzepts ISBN 978-3-8325-4944-2 53.50 EUR
4	Lisa Rott	Vorstellungsentwicklungen und gemeinsames Lernen im inklusiven Sachunterricht initiieren: Die Unterrichtskonzeption „choice2explore“ ISBN 978-3-8325-4817-9 83.00 EUR
5	Eva Julia Kolbeck	Schulung von Vermittlungsfähigkeiten Promovierender im Fach Chemie: Die Weiterbildung „How to communicate chemistry?“ ISBN 978-3-8325-4953-4 71.00 EUR
6	Björn Dellbrügge	Choice2interact Interaktiv Lernen mit Tablets im Chemieunterricht ISBN 978-3-8325-5119-3 79.00 EUR

7	Felicitas Jürgensmeier	Offene Experimentiersituationen ermöglichen und charakterisieren – eine Schülerlaboreinheit zu einem curricular innovativen Thema ISBN 978-3-8325-5184-1 89.00 EUR
8	Jan-Bernd Haas	chem.LEVEL – fachsprachlich sensibler Chemieunterricht auf Basis des Johnstone Dreiecks ISBN 978-3-8325-5345-6 89.00 EUR
9	Florian Jungkamp	Kontroversen mit Hilfe wissenschaftlicher Prüfkriterien bewerten: Die Unterrichtskonzeption choice^{2}reflect ISBN 978-3-8325-5362-3 69.50 EUR
10	Christopher Kralisch	Wissenschaftskriterien verstehen und anwenden Weiterentwicklung der Unterrichtskonzeption choice^{2}reflect zur Bewertung gesellschaftlicher Kontroversen ISBN 978-3-8325-5517-7 91.50 EUR
11	Fabian Gust	Das Self-assembLAB: Entwicklung, Erprobung und Optimierung eines curricular innovativen Schülerlabors zum Thema Self-assembly ISBN 978-3-8325-5517-7 85.50 EUR
12	Marius van den Boom	Die Unterrichtskonzeption *feil*: Fehlschlüsse identifizieren lernen ISBN 978-3-8325-5562-7 87.00 EUR

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

Verstehen und *Entwickeln* bilden eine fachdidaktische Einheit: das Verstehen von Lernprozessen hilft uns, innovative Lernangebote zu entwickeln und an Rahmenbedingungen anzupassen. Die Erprobung und Analyse dieser Angebote kann wiederum zu erweiterten Erkenntnissen und einem tieferen Verstehen führen. Diese Schriftenreihe rückt daher beide Aspekte in den Fokus – mit dem Ziel, das zu fördern, was uns als Naturwissenschaftsdidaktikern am Herzen liegt: das Lernen in Naturwissenschaften!

Im Rahmen der Unterrichtskonzeption choice^{2}reflect nach Jungkamp und Marohn erarbeiten Schülerinnen und Schüler wissenschaftliche Prüfkriterien. Diese sollen sie dabei unterstützen, Aussagen und Studien zu beurteilen und in kontrovers diskutierten Themen rationale Entscheidungen zu treffen. Band 9 dieser Reihe stellt die Konzeption vor und zeigt, dass die Prüfkriterien sinnvolle Werkzeuge zur Etablierung eines wissenschaftlichen Weltbildes bei Lernenden bilden können.
Im Fokus dieses Bandes steht die Weiterentwicklung der fünfphasigen Konzeption im Rahmen des Design-Based Research-Ansatzes: Weitere Prüfkriterien werden ausgearbeitet, im Unterricht erprobt und erforscht. So erfahren Lernende zum Beispiel was es bedeutet, zwischen Korrelationen und Kausalitäten zu unterscheiden, falsifizierbare Aussagen zu erkennen, Stichproben zu randomisieren oder Versuchsdokumentationen reproduzierbar zu gestalten.
Zudem werden die Unterrichtsmaterialien in einen weiteren kontroversen Kontext eingebettet: die Thematik Nahrungsergänzungsmittel. Mithilfe des entwickelten Materials explizieren Lernende zunächst ihren Standpunkt zur Kontroverse und werden nach Erarbeitung der wissenschaftlichen Prüfkriterien dazu angeregt, die eigene Position zu reflektieren.
In einem dritten Schritt des Projekts wird ein Prüfschema zur Anwendung der Wissenschaftskriterien entwickelt und empirisch untersucht. Anhand dieses Schemas können Lernende alltägliche Aussagen und empirische Studien auf Wissenschaftlichkeit überprüfen.

ISBN 978-3-8325-5517-7

Logos Verlag Berlin

ISSN 2566-493X