

Umgang mit Evidenzen angehender Lehrkräfte in den Naturwissenschaften

- Epistemologie in der Lehrkräfteausbildung -

Dissertation
zur Erlangung des Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

vorgelegt dem Department Chemie der
Fakultät für Naturwissenschaften der Universität Paderborn

von Pascal Pollmeier
aus Schloß Holte-Stukenbrock

Paderborn, November 2023

Vorsitzender:	Prof. Dr. Thomas Werner
1. Gutachterin:	Prof. Dr. Sabine Fechner
2. Gutachter:	Prof. Dr. Sebastian Habig
3. Prüfer:	Nachwuchsgruppenleiter Dr. Christoph Vogelsang

Tag der Disputation: 20.12.2023

„And let it not be forgotten that all of us, we the contributors to the text and you the readers, will continue to grow and develop epistemologically. We must ask ourselves, where will our thinking be in 10 years, 20 years, 30 years, or longer?“

Marlene Schommer-Aikins

Für meine Eltern

DANKSAGUNG

Zuerst möchte ich meiner Betreuerin Prof. Dr. Sabine Fechner für ihre stetige Unterstützung und das Vertrauen während meiner Promotionszeit danken. Herzlichen Dank für alle Herausforderungen, an denen ich wachsen durfte, die konstruktiven Diskussionen sowie die stets zugewandte und freundschaftliche Betreuung. Ebenso möchte ich mich für die optimalen Arbeitsbedingungen, die flexible Reaktion auf Schwierigkeiten im Zusammenhang mit der Corona-Pandemie und die fantastische Arbeitsatmosphäre in ihrem Arbeitskreis bedanken.

Einen besonderen Dank möchte ich auch Prof. Dr. Sebastian Habig für die Übernahme des Zweitgutachtens richten.

Dr. Christoph Vogelsang möchte ich für die Übernahme der Funktion des Drittprüfers sowie die stets kritisch-konstruktive Diskussion meiner Arbeit und die Unterstützung in der Datenerhebung in Kooperation mit der *PLAZ – Professional School of Education* herzlich danken.

Ein besonderer Dank gilt auch Prof. Dr. Thomas Werner für die Übernahme des Prüfungsvorsitzes meines Promotionsverfahrens.

Furthermore, I would like to thank the consortium of the Erasmus+ project *Research in Teacher Education* Dr. Ria Dolfing, Prof. Dr. Christian Bokhove, Dr. Sally Bamber and Prof. Dr. Eliza Rybska for the fruitful discussions on evidence-based practice and the critical reflection on our instruments. Last but not least, thank you for giving me insights into international collaboration.

Herzlich bedanken möchte ich mich auch bei meinen Arbeitskolleg*innen Dr. Perihan Akman, Lisa Schmitz, Dr. Thomas Witte, Julia Elsner, Hendrik Peeters, Lisa Wedekind, Jonas Ponath, Christoph Fröhleke und Salome Janke. Vielen Dank für die gemeinsame Zeit im Büro, auf Tagungen oder vertieft in Diskussionen. Besonders danken möchte ich Stefanie Michaelis für die stetige Unterstützung in allen Belangen sowie Sina Pankoke für die organisatorische Unterstützung.

Danken möchte ich außerdem den studentischen Hilfskräften, die zum Gelingen dieser Arbeit beigetragen haben. Allen voran Charlotte für die Unterstützung bei der Kodierung.

Nicht zuletzt möchte ich den Studierenden danken, welche in den herausfordernden Zeiten der Corona-Pandemie an meinen Studien teilgenommen haben.

Zuletzt möchte ich einen herzlichen Dank an meine Familie und Freunde richten. Mama, Papa – vielen Dank für eure unentwegte Unterstützung auf meinem Weg von der Realschule bis zur Promotion. Ohne euer Zutun wäre dieser Weg für mich sicher nicht möglich gewesen.

Ebenso danke ich von ganzem Herzen dir, Lena! Vielen Dank für deine Geduld und dein Vertrauen in mich.

INHALT

1. Einleitung	1
2. Evidenzbasierung	3
2.1 Der Evidenzbegriff.....	3
2.2 Argumentation als Evidenz	5
2.3 Evidence-based practice	9
2.4 Evidence-based teaching	13
3. Einfluss von Epistemologie auf evidenzbasierte Praxis	22
3.1 Begriffsabgrenzung	22
3.2 Definition Epistemologie.....	23
3.3 Conceptual Change und Epistemologie	35
3.3.1 Conceptual Change	35
3.3.2 Framework-theories	43
3.3.3 Umgang mit anomalen Daten	45
3.3.3.1 Schulische vs. Authentische wissenschaftliche Arbeit.....	53
3.4 Epistemologie und Nature of science	54
4. Zusammenfassung und Forschungsdesiderat.....	60
5. Zielstellung der Arbeit.....	62
6. ERSTER EMPIRISCHER TEIL – FÖRDERUNG DER KOMPETENZEN IN EVIDENZBASIERTER PRAXIS ANGEHENDER (CHEMIE-) LEHRKRÄFTE	63
6.1 Studiendesign der fächerübergreifenden Erhebung im Rahmen des Praxissemesters.	64
6.2 Stichprobe der fächerübergreifenden Erhebung.....	66
6.3 Studiendesign der Interventionsstudie zur Förderung von Kompetenzen angehender Chemielehrkräfte in evidenzbasierter Praxis.....	70
6.4 Stichprobe der Interventionsstudie zur Förderung von Kompetenzen angehender Chemielehrkräfte in evidenzbasierter Praxis.....	77
6.5 Instrumente	78
6.5.1 Evidence-informed decision making in education test (EIDM-ET)	78
6.5.1.1 Kurzfassung des EIDM-ET.....	84
6.5.2 Evidence-based practice scales.....	90
6.5.3 Interview	96
6.5.4 Zusätzlich erhobene Daten	97
6.6 Ergebnisse und Diskussion	97
6.6.1 Ergebnisse der fächerübergreifenden Befragung.....	97
6.6.1.1 Ergebnisse des gekürzten EIDM-ET.....	97
6.6.1.2 Ergebnisse der <i>EBP beliefs scale</i>	100

6.6.1.3 Ergebnisse der <i>EBP implementation scale</i>	102
6.6.2 Diskussion der Ergebnisse der fächerübergreifenden Befragung	104
6.6.3 Ergebnisse der Interventionsstudie zur Förderung angehender Chemielehrkräfte in evidenzbasierter Praxis	112
6.6.3.1 Ergebnisse des EIDM-ET	112
6.6.3.2 Ergebnisse der EBP beliefs scale.....	114
6.6.3.3 Ergebnisse der EBP implementation scale	114
6.6.3.4 Ergebnisse der Interviews	115
6.6.3.5 Ergebnisse der zweiten Phase der Interventionsstudie	119
6.6.4 Zusammenführung der Ergebnisse.....	124
6.6.5 Diskussion der Ergebnisse der Interventionsstudie zur Förderung angehender Chemielehrkräfte in evidenzbasierter Praxis	126
6.7 Zusammenfassende Diskussion des ersten empirischen Teils	131
7. ZWEITER EMPIRISCHER TEIL - UMGANG MIT ANOMALEN BEOBACHTUNGEN ANGEHENDER LEHRKRÄFTE DES NATURWISSENSCHAFTLICHEN SACHUNTERRICHTS	134
7.1 Studiendesign	134
7.1.1 Experimente mit potentiell anomalen Beobachtungen	136
7.1.2 Model-of-data	141
7.1.3 Digitale Umsetzung der Modellierungsstudie	146
7.2 Stichprobe der Modellierungsstudie.....	147
7.3 Ablauf der Analyse.....	148
7.4 Instrumente	149
7.4.1 Fachwissenstest.....	149
7.4.2 Prozessdaten (lautes Denken, Notizen)	152
7.5 Datenauswertung	153
7.5.1 Erstellung des Kodiermanuals	153
7.5.2 Güte der Datenauswertung.....	157
7.6 Ergebnisse und Diskussion	161
7.6.1 Analyse der inhaltlichen Aspekte	161
7.6.2 Beschreibung anomaler Beobachtungen	168
7.6.3 Analyse der gebildeten Verknüpfungen.....	172
7.6.4 Ergebnisse der Auswertung des Fachwissenstests	178
7.6.5 Ergebnisse zur Validität der erstellen Model-of-data	180
7.7 Diskussion der Ergebnisse der Modellierungsstudie.....	184
8. Zusammenführung und Fazit.....	198
9. Implikationen.....	202

9.1	Implikationen für die Forschung	202
9.2	Implikationen für die Lehrkräftebildung	203
10	Verzeichnisse	205
10.1	Literaturverzeichnis	205
10.2	Abbildungsverzeichnis	222
10.3	Tabellenverzeichnis.....	222
11.	Anhang	225

1. Einleitung

Ein zentrales Ziel des Chemieunterrichts liegt in einer vertieften naturwissenschaftlichen Grundbildung (MSB NRW, 2019). Dabei beinhaltet diese, dass Schüler*innen „eine aktive Teilhabe an gesellschaftlicher Kommunikation und Meinungsbildung über technische Entwicklungen und naturwissenschaftliche Forschung“ (MSB NRW, 2019, S. 8) ermöglicht werden soll. Dazu gehört eine Kenntnis über wissenschaftliche Erkenntnismethoden wie dem chemischen Experiment und den damit verbundenen Beobachtungs- und Auswertungsprozessen. Lebensweltliche Phänomene sollen durch die Nutzung geeigneter Modelle erklärt werden, um daraus begründete Beurteilungen und Entscheidungen abzuleiten.

Der Kernlehrplan Chemie für die Sekundarstufe II stellt dabei besonders „aktuelle [...] und zukünftige[...] Umwelt-, Verbraucher-, Ressourcen- oder Alltagsfragen“ (MSB NRW, 2022, S. 9) in den Fokus der Diskussionen. Dabei können zahlreiche aktuelle Themen z.B. bzgl. der Energieversorgung oder Elektromobilität für den Chemieunterricht beschrieben werden. Im Kontext einer Bildung für nachhaltige Entwicklung können ebenso der Klimawandel, die Gewässerverschmutzung sowie mobile Energiequellen als relevante Themen identifiziert werden (MSB NRW, 2019). Dabei stellen diese Themen gesellschaftlich relevante Aspekte in den Vordergrund.

In der gesellschaftlichen Realität kann aktuell jedoch eine Abkehr von wissenschaftlichen Argumenten und Methoden beobachtet werden. Kienhues et al. (2020, S. 145) beschreiben in diesem Zusammenhang: „the more detrimental effect of post-truthism may be the public loss of confidence in sciences as a fact-based, objective discourse“. Beispiele stellen Diskussionen um den Klimawandel, Impfkampagnen oder der Homöopathie dar (Kienhues et al., 2020). Ein Verlust des Vertrauens in Wissenschaft kann dabei zu einer abnehmenden Bedeutung wissenschaftlicher Erkenntnisse für Diskussionen führen.

Für den Klimawandel formulieren E. U. Weber und Stern (2011) unterschiedliche Begründungen für Unterschiede in der gesellschaftlichen und wissenschaftlichen Argumentation, welche sich ggf. auf weitere gesellschaftsrelevante Themen übertragen lassen könnten. Neben der großen Komplexität der konzeptuellen Wissensaspekte sowie dem Einfluss politischer Kampagnen im Sinne des *motivated reasonings* (Kienhues et al., 2020) erkennen E. U. Weber und Stern (2011) einen Unterschied im Verständnis von Expert*innen und Lai*innen. Dabei können diese Unterschiede auf unterschiedliche Vorstellungen über die Natur von Wissen zurückgeführt werden (D. Kuhn & Weinstock, 2002). Für diese epistemologischen Überzeugungen konnten Einflüsse auf Lernprozesse wie z.B. das Problemlösen (Schraw et al., 1995) sowie die Akzeptanz anomaler Beobachtungen (Mason, 2002) beschrieben werden. Ebenso zeigen Kienhues et al. (2020, S. 146) „gateways for post-truth agendas“ auf, welche auf epistemologische Überzeugungen zurückgeführt werden können. Demnach stellen die Subjektivität wissenschaftlicher Interpretationsprozesse, die Einbettung von Wissenschaft in soziale Prozesse sowie die Grenzen von wissenschaftlicher Erkenntnis mögliche Ansatzpunkte für die Abkehr von wissenschaftlichen Argumentationen dar. Diese *gateways* können dabei als anschlussfähig an epistemologische Überzeugungen (Hofer & Pintrich, 1997) wie auch Vorstellungen aus dem Feld *Nature of Science* (N. G. Lederman, 2007) beschrieben werden.

Eine Möglichkeit die Grenzen und Vorläufigkeit von wissenschaftlicher Erkenntnis produktiv in gesellschaftliche Diskussionen einzubringen, kann im Konzept der evidenzbasierten Praxis erkannt werden. Dieses Konzept fokussiert auf den Einbezug von wissenschaftlichen Evidenzen sowie der Erfahrung von Expert*innen verschiedener Bereiche (Sackett et al., 1996). Dabei wird ebenfalls

angenommen, dass nicht für alle Inhaltsbereiche hochqualitative Evidenzen vorhanden sind und somit der Erfahrungen von Expert*innen eine wichtige Bedeutung zukommt. Im Sinne gesellschaftlicher Debatten könnte so die Bedeutung wissenschaftlicher Erkenntnis deutlich werden, jedoch auch eine Offenheit für weitere Informationsquellen existieren.

Das Konzept evidenzbasierter Praxis wurde für verschiedene Bereiche diskutiert (Bauer & Kollar, 2023; Breckon & Dodson, 2016; Davies, 1999; Sackett et al., 1996). Seinen Ursprung nahm das Konzept im Bereich evidenzbasierter Medizin (Sackett et al., 1996). Zuletzt wurden erste Ansätze evidenzbasierter Praxis im Bildungsbereich diskutiert (Bromme et al., 2014). Dabei kann festgehalten werden, dass das Konzept (noch) keinen umfassenden Einzug in die Lehrkräftebildung gehalten hat.

Die Etablierung einer evidenzbasierten Praxis im Unterricht, könnte sich auch auf gesellschaftliche Debatten übertragen lassen. Dadurch könnte der Einbezug von Evidenzen in gesellschaftliche Diskurse erreicht werden. Das Ziel der Förderung einer naturwissenschaftlichen Grundbildung im Sinne der Teilnahme an gesellschaftlichen Diskursen der Schüler*innen macht entsprechend auch einen Einbezug epistemologischer Überzeugungen notwendig. In diesem Zusammenhang stellt sich jedoch die Frage, ob (angehende) Lehrkräfte notwendige epistemologische Überzeugungen bzw. Kompetenzen in evidenzbasierter Praxis besitzen.

Ziel dieser Arbeit ist es entsprechend, die Kompetenzen angehender Lehrkräfte in evidenzbasierter Praxis zu untersuchen und zu fördern. Dabei sollen diese auf Grundlage bestehender epistemologischer Überzeugungen der Proband*innen fokussiert werden. Es stellt sich zuerst die Frage, inwiefern bereits Kompetenzen in evidenzbasierter Praxis bei angehenden Lehrkräften unterschiedlicher Unterrichtsfächer vorhanden sind (Teilstudie 1). Dabei wird auch ein Einfluss erster schulpraktischer Erfahrungen im Rahmen des Praxissemesters untersucht. In Teilstudie 2 soll die Förderung angehender Chemielehrkräfte in evidenzbasierter Praxis untersucht werden. Dabei soll ebenfalls die Stabilität der Kompetenzen im Sinne einer längsschnittlichen Untersuchung in die zweite Ausbildungsphase (Referendariat) betrachtet werden. Zuletzt soll der Umgang mit anomalen Beobachtungen als widersprüchliche Evidenzen fokussiert werden (Teilstudie 3). Hier stellt sich die Frage, wie angehende Lehrkräfte mit anomalen Beobachtungen aus naturwissenschaftlichen Experimenten umgehen und welche epistemologischen Überzeugungen dem jeweiligen Umgang zu Grunde liegen.

2. Evidenzbasierung

Um Qualität in einem Bildungssystem zu sichern und zu entwickeln, müssen wissenschaftliche Grundlagen geschaffen werden, die eine verlässliche Beurteilung der Situation und der Perspektiven ermöglichen.

(BMBF, 2008, S. 7)

Diese Forderung stellt das Bundesministerium für Bildung und Forschung im Jahr 2008 im „Rahmenprogramm zur Förderung der empirischen Bildungsforschung“. Das Programm stellt einerseits die Notwendigkeit exzellenter Forschung zu Bildungsprozessen in den Fokus, zeigt aber gleichzeitig auch die notwendige Implementation der gewonnenen Erkenntnisse in das Bildungssystem und Schule auf. Damit wird die Rolle von Evidenzen zur Weiterentwicklung des Bildungssystems in den Fokus gestellt. Das obige Zitat stellt dabei einige Bedingungen an die beschriebenen wissenschaftlichen Grundlagen. Diese sollen **verlässlich** sein und zur **Beurteilung** einer Situation sowie dem **Aufzeigen möglicher Perspektiven** genutzt werden können. Bromme et al. (2014, S. 4) erkennen in diesem Programm „eine Entwicklung in Richtung einer daten- und evidenzgestützten Politik [...], die in anderen Bereichen (z.B. Gesundheit, Wirtschaft, Umwelt) bereits seit längerer Zeit begonnen hat.“. Gleichzeitig stellen die Autoren zentrale Problemstellen in der Evidenzbasierung des Bildungssystems in den Fokus. So muss die Aussagekraft jedweder Erkenntnisse vor dem Hintergrund der genutzten Methoden, Fragestellungen und Theorien betrachtet werden. Auch sei es nicht immer möglich, Fragestellungen (einfach) zu beantworten, weil noch nicht viel Forschung in diesem Bereich geschehen ist oder Erkenntnisse sich gegenseitig widersprechen (Bromme et al., 2014). Neben dem Nutzen von Evidenzen zur Weiterentwicklung des Bildungssystems, kann evidenzbasierte Praxis auch auf Unterrichtsebene konzeptualisiert werden.

International wird Evidenzbasierung als wichtiger Bestandteil der Lehrkräfteprofessionalisierung verstanden. Bauer und Prenzel (2012) erkennen mit den Umstrukturierungen der Lehrkräfteausbildung im Rahmen der Bologna-Reformen einen curricularen Trend zur Implementierung von Evidenzbasierung im Zusammenhang mit lebenslangem Lernen. Dies könne unter anderem auf das Verständnis der Rolle von Lehrkräften zurückgeführt werden, wie es durch die Europäische Kommission beschrieben wird:

As with any other modern profession, teachers also have a responsibility to extend the boundaries of professional knowledge through a commitment to reflective practice, through research, and through a systematic engagement in continuous professional development from the beginning to the end of their careers. Systems of education and training for teachers need to provide the necessary opportunities for this.

(Commission of the European Communities, 2007, S. 5)

In dem folgenden Kapitel soll eine multiperspektivische Beschreibung des Konstrukts Evidenz und der damit verbundenen evidenzbasierten Praxis erfolgen. Dabei soll der Blick auch auf Konzeptualisierungen aus anderen Bereichen eingegangen werden (vgl. Bromme et al., 2014).

2.1 Der Evidenzbegriff

Der Begriff „Evidenz“ stammt vom lateinischen Wort „evidentia“ und bedeutet nach der Dudenredaktion (o.D.) „unmittelbare und vollständige Einsichtigkeit, Deutlichkeit, Gewissheit“. Eine andere Definition des Begriffs stammt aus dem Cambridge Dictionary (o.D.) „evidence – one

or more reasons for believing that something is or is not true". Die beiden Definitionen zeigen einen uneinheitlichen Sprachgebrauch des Worts „Evidenzen“ auf. In diesem Zusammenhang erkennen Tippelt und Reich-Claassen (2010) einen Bruch zwischen der umgangssprachlichen und fachsprachlichen Verwendung des Begriffs „Evidenz“, wie er schon in den obigen Definitionen deutlich wird. Demnach würde der Begriff umgangssprachlich immer dann verwendet, wenn die „Gültigkeit und Offensichtlichkeit [von Schlussfolgerungen] so offensichtlich »auf der Hand« liegen, dass jegliche empirische Untermauerung obsolet erscheint“ (Tippelt & Reich-Claassen, 2010, S. 22). In der fachsprachlichen Verwendung steht die empirische Untermauerung entlang definierter Standards deutlich im Vordergrund. Dabei „meint ‚evidence‘ stets etwas Mittelbares, nämlich die durch Forschungsmethoden vermittelten empirischen Erkenntnisse, die wiederum selbst nicht unmittelbare Wahrheiten darstellen, sondern nur die interpretationsbedürftige Grundlage für bestimmte Schlussfolgerungen“ (Bellmann & Müller, 2011a, S. 11). Das hier dargestellte Verständnis des englischen Begriffs *evidence* entspricht dem deutschen, fachsprachlichen Gebrauch des Begriffs Evidenz. Folglich ist ein differenzierterer Blick auf Evidenzen nötig, um das Konstrukt in seiner Gänze zu erfassen.

Aus der Literatur können verschiedene Aspekte einer Definition von „Evidenz“ abgeleitet werden. Ein erster Aspekt ist die **Quelle von Evidenzen**. Evidenzen können durch Forschungseinrichtungen und Forschende erstellt werden (Breckon & Dodson, 2016). Wie bereits dargestellt, gibt es jedoch Bereiche, in denen noch nicht genügend Forschung vorhanden ist, oder sich bestehende Forschungsergebnisse gegenseitig widersprechen (Bromme et al., 2014). Vor allem in diesen Bereichen können Expert*innen und Praktiker*innen auf dem spezifischen Feld Evidenzen darstellen (Breckon & Dodson, 2016). Hier muss jedoch die **Qualität von Evidenzen** kritisch betrachtet werden. Während Forschende durch spezifische Verfahren und kontrollierte Umgebungen Evidenzen erzeugen, geben Expert*innen und Praktiker*innen vorhandene Erfahrungen weiter. Diese könnten, je nach Fragestellung, unsystematisch zustande gekommen sein. In diesem Fall sollten Evidenzen aus wissenschaftlichen Studien (mit kontrollierter Güte) bevorzugt werden (Canadian Health Services Research Foundation, 2005; Sackett et al., 1996). Vor allem das intentionale und systematische Erstellen von Evidenzen erhält in diesem Zusammenhang eine besondere Bedeutung. Liegen derartige Evidenzen vor, können die Erfahrungen von Expert*innen und Praktiker*innen dennoch als weitere Hinweise in eine Untersuchung eingehen. Hier wird gleichzeitig der **summative Charakter von Evidenzen** deutlich. Frei nach der Redensart ‚Eine Studie ist keine Studie‘, sollten vor allem Zusammenführungen von Evidenzen besonders berücksichtigt werden (Gorard et al., 2020). Damit erhalten vor allem Meta-Analysen, oder systematische Reviews eine besondere Bedeutung, da diese Evidenzen aus verschiedenen Studien vergleichend zusammentragen. Gleichzeitig bestünde bei diesen Arten von Studien (Meta-Analysen, Hyper-Analysen) die Gefahr, dass die Qualität und die Eigenheiten der eingeschlossenen Einzelstudien zu wenig beachtet würden. Dies könne zu fehlgeleiteten Entscheidungen führen (Gorard et al., 2020). Zuletzt muss die **Subjektivität von Evidenzen** beachtet werden (Gorard et al., 2020). Innerhalb eines Forschungsprozesses werden verschiedene Gütekriterien beachtet, welche z.B. zu einer hohen Objektivität der Evidenzen führen sollen (Döring & Bortz, 2016). Durch die Involvierung der Forschenden als Individuen mit eigenen Interessen, Überzeugungen und Zielsetzungen kann es jedoch subjektive Einflüsse geben. Dies könne auch dazu führen, dass z.B. (bildungs-)politische Entscheidungen weniger auf Evidenzen, als auf subjektiven Überzeugungen beruhen (Gorard et al., 2020). Es wird also deutlich, dass Evidenzen durch Forschungsmethoden und an wissenschaftlichen Qualitätsstandards orientierte Hinweise für Schlussfolgerungen darstellen. Zur Erstellung von Schlussfolgerungen, bedarf es der Interpretation verschiedener Evidenzen. Hier wird erneut die große Bedeutung der Subjektivität deutlich. Sowohl im Forschungsprozess, also der Generierung von Evidenzen, als

auch in der Interpretation von Evidenzen spielen subjektive Überzeugungen und Einstellungen eine wichtige Rolle.

Bromme et al. (2014) bringen einen weiteren Impuls ins Spiel, wenn sie Evidenzen als funktional beschreiben. Demnach könnten Daten die Funktion von Evidenzen erfüllen, wenn sie dafür genutzt werden Theorien oder Hypothesen zu stützen oder zu widerlegen. Die daraus resultierende Verknüpfung von Evidenzen mit Theorien und Hypothesen bezeichnen die Autoren als „*innerhalb der Wissenschaften selbstverständlich*“ (Bromme et al., 2014, S. 7). Auch aus diesem Verständnis wird deutlich, dass Evidenzen nicht unmittelbar vorhanden sind, sondern erst durch die Interpretation im Zusammenhang mit Theorien und Hypothesen eine Funktion erfüllen. Die Interpretation sei dabei „ein aufwändiger und komplexer Vorgang, der insbesondere Expertise und kritische Reflexion verlangt“ (Bromme et al., 2014, S. 8). Die kritische Interpretation von Evidenzen im Zusammenhang mit Theorien spiegelt dabei auch die Vorläufigkeit wissenschaftlicher Erkenntnis wider (Bromme et al., 2014).

Zusammenfassend könnte eine Definition des Konstrukts „Evidenz“ wie folgt lauten:

Evidence is the best available information regarding a particular topic, especially if it refers to results of various studies. Evidence can be gathered and interpreted by researchers and anybody else (e.g. practitioners, experts, ...) through intentional and systematic procedures. The focus and format of evidence can vary with regard to internal interests, values and purposes of the authors or the target group. (Dolfing et al., 2022, S. 7)¹

Wie gezeigt wurde, wird das Konstrukt „Evidenz“ innerhalb verschiedener Kontexte unterschiedlich definiert. Während ein eher einfaches Verständnis von Evidenzen in der umgangssprachlichen Nutzung des Begriffs erkannt werden kann, wird das Konstrukt in der wissenschaftlichen Community als komplex und vorläufig beschrieben. So kann Evidenz im wissenschaftlichen Verständnis nur im Zusammenhang mit einer Theorie/Hypothese verstanden werden. Ebenso zeigen sich verschiedene Voraussetzungen (Generierung durch anerkannte Forschungsmethoden, Einhaltung wissenschaftlicher Qualitätsstandards, etc.), welche Evidenzen erfüllen müssen. Deutlich wird vor allem die Bedeutung subjektiver Einflüsse auf Evidenzen. So ist vor allem die Interpretation von Daten im Zusammenhang mit einer Theorie/Hypothese anfällig für subjektive Meinungen, Werthaltungen, Überzeugungen, etc. (Voss, 2022). Diese Arbeit basiert auf dem wissenschaftlichen Verständnis von Evidenz, wie es in der oben dargestellten Definition aus dem Projekt „*Research in Teacher Education*“ ersichtlich wird.

2.2 Argumentation als Evidenz

Wie beschrieben, kommt der Subjektivität von Evidenzen eine besondere Bedeutung zu. Diese kann sich vor allem in der Interpretation von Daten zu Evidenzen äußern. Bromme et al. (2014) stellen Evidenzen in den Zusammenhang mit theoretischen Überlegungen und damit als Evidenz *für*, oder *gegen* bestimmte Hypothesen. Gesammelte Daten stellen entsprechend nicht per se Evidenzen dar. Es bedarf der Interpretation der Daten, damit diese die Funktion von Evidenzen erfüllen. Aus den Evidenzen können dann wiederum Schlussfolgerungen gezogen werden,

¹ Die vorliegende Definition hat der Autor dieser Arbeit im Zusammenhang mit dem durch Erasmus+ geförderten Projekt „*Research in Teacher Education*“ (Erasmus+ programme agreement number: 2019-1-NL01-KA203-060339) erarbeitet.

welche auch aus mehreren, widersprüchlichen Evidenzen abgeleitet werden können. Damit stellen Schlussfolgerungen die Vermittlung zwischen unterschiedlichen Evidenzen dar, welche anhand gewählter Kriterien in ihrer Bedeutsamkeit abgewogen werden. Das Aufstellen von Schlussfolgerungen und der Zusammenhang mit Evidenzen kann durch Argumentationsmodelle genauer untersucht werden.

Verschiedene Modellierungen des Argumentationsprozesses basieren auf unterschiedlichen philosophischen Positionen. Grundsätzlich kann zwischen einer mechanistischen und einer systemischen Position unterschieden werden (Ebert, 2001). Dabei beziehen sich diese Positionen nicht allein auf Argumentationsprozesse, sondern können als Weltanschauungen verstanden werden. Innerhalb der mechanistischen Position werden komplexe Probleme als Summe einzelner Phänomene verstanden (Ebert, 2001). Daraus folgt, dass diese Probleme bearbeitet werden können, indem die Bestandteile identifiziert und gelöst werden. Es wird somit ein linear-kausaler Zusammenhang zwischen den Bestandteilen und dem Problem als Ganzes angenommen. Diese Betrachtungsweise kann wissenschaftshistorisch auf die klassische Mechanik zurückverfolgt werden. So könne der Grundgedanke dieser Position z.B. in der Reduktion komplexer Bewegungen auf die drei Newtonschen Axiome erkannt werden. Komplexe Probleme werden auf einzelne Phänomene reduziert, für welche linear-kausale Beziehungen postuliert werden (Ebert, 2001). Diese Herangehensweise wird in der mechanistischen Position auch auf das Universum als Gesamtsystem übertragen. So formuliert Merchant (1989) fünf Grundannahmen der mechanistischen Sichtweise:

Erstens: Materie setzt sich aus einzelnen Atomen zusammen. Zweitens: Das Ganze entspricht der Summe seiner Teile – das Gesetz der Identität in der Mathematik. [...] Die dritte Annahme: Äußere Ursachen wirken auf tote, unbelebte Teile ein. [...] Für den Physiker Isaac Newton verharren Körper im Ruhezustand, oder sie bewegen sich gradlinig, wenn keine äußere Kraft auf sie einwirkt. Die vierte Annahme besagte, Veränderungen seien die Folge der Neuordnung von Teilen. Die fünfte schließlich ist das Konzept des Dualismus: Geist und Körper, Natur und menschliche Natur (= Kultur) sind grundsätzlich getrennt. (Merchant, 1989, S. 137)

In der umfassenden Veränderung vormals als sicher geltender linear-kausaler Verbindungen durch Albert Einsteins Relativitäts- und Quantentheorie, sieht Ebert (2001) Kritikpunkte für die mechanistische Weltanschauung. Spätestens durch die von Heisenberg postulierte Unschärfere-lation, seien linear-kausale Beziehungen im deterministischen System geschwächt. So könnten nur Wahrscheinlichkeiten für bestimmte Zusammenhänge betrachtet werden. Daraus folgt, dass die Summe der Einzelbestandteile nicht das Gesamtsystem widerspiegelt, sondern nur eine zu einem bestimmten Prozentsatz wahrscheinliche Variante des Systems. Diese Betrachtung führt zur systemischen Weltanschauung:

Der bis dahin gültige monokausale Determinismus wird durch einen probalistischen Determinismus abgelöst. Es wird nun angenommen, daß (sic) mehrere Faktoren summativ und aufgrund ihres spezifischen Strukturmusters die Wahrscheinlichkeit so lange erhöhen, bis bestimmte Wirkungsweisen und Phänomene, entstehen. Der Faktor Wahrscheinlichkeit weist darauf hin, daß (sic) bestimmte Phänomene auch bei ähnlichen Mustern, nicht entstehen müssen, sowie Aussagen über eine Wiederholbarkeit von bestimmten Phänomenen nicht getroffen werden können. (Ebert, 2001, S. 86)

Die systemische Betrachtung wird als holistischer Ansatz beschrieben, der auch Wechselwirkungen zwischen Aspekten betrachtet, welche in der mechanistischen Betrachtung durch den zu Grunde gelegten Reduktivismus vernachlässigt werden. Da beide Betrachtungsweisen auf

unterschiedlichen philosophischen Begründungen aufbauen, ist keine der beiden der anderen generell überlegen. Ebert (2001) weist darauf hin, dass in den Naturwissenschaften zumeist eine mechanistische Betrachtung favorisiert wird. Die Arbeiten auf dem Feld der Relativitätstheorie zeigen jedoch auch die Notwendigkeit systemischer Betrachtungsweisen auf.

Die hier dargestellten philosophischen Positionen könnten in Argumentationsmodellen identifiziert werden. Im Zusammenhang mit der datenbasierten Erstellung von Schlussfolgerungen kann das Modell von Toulmin (2003) hervorgehoben werden (Abb. 1). In seinem Modell stellt Toulmin (2003) die Beziehung zwischen Daten (*data*) und Schlussfolgerungen (*claim*) dar.

Grundlegend wird der *claim* durch **Data** gestützt. *Data* bildet die Grundlage, auf welcher der *claim* begründet wird (Toulmin, 2003). *Data* stellen dabei noch nicht interpretierte Messwerte, Beobachtungen, etc. dar. *Data* können in eigenen empirischen Erhebungen gesammelt werden und müssen sich hier an den gängigen Qualitätsstandards wissenschaftlicher Arbeit orientieren. Die Begründung des *claims* mit *data* bedarf aber weiterer Annahmen:

Our task is no longer to strengthen the ground on which our argument is constructed, but is rather to show that, taking these data as a starting point, the step to the original claim or conclusion is an appropriate and legitimate one. (Toulmin, 2003, S. 91)

Annahmen, welche die Anwendung der *data* auf den *claim* unterstützen, werden im *Model of argument* als **warrant** bezeichnet (Toulmin, 2003). Ein *warrant* unterstützt somit die Verbindung von *data* und *claim* zueinander und fungiert als Brücke. Der *warrant* legitimiert die Nutzung der *data* für den spezifischen *claim*. Am Beispiel einer empirischen Forschungsstudie könnte z.B. die Ausprägung des Interesses bzgl. eines bestimmten Inhaltsfeldes untersucht werden. Selbstauskünfte können hier als *data* verstanden werden, welche bzgl. des *claims* interpretiert werden können. Die Annahme, dass sich das Interesse während der Befragung nicht von dem Interesse kurz nach der Befragung unterscheidet, da es sich um ein stabiles Konstrukt handelt, könnte als *warrant* angeführt werden. Dabei ist nach Toulmin (2003) die Unterscheidung von *data* und *warrant* weder trennscharf möglich noch in jeder Situation nötig. In ihrer Natur unterscheiden sich beide Bestandteile des Modells. Je nach Komplexität der Argumentation könnten die unterschiedlichen Funktionen klarer hervortreten. Toulmin (2003) zeigt in seinen Beispielen auf, dass ein *warrant* je nach Zusammenhang sehr trivial erscheinen kann². In komplexeren Argumentationen könnte die Natur des *warrant* deutlicher Einfluss nehmen.

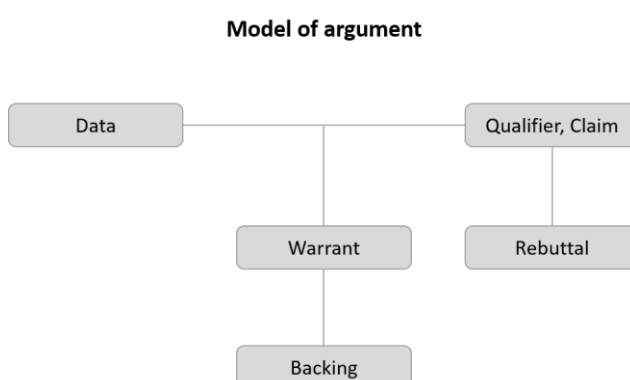


Abbildung 1 - *Model of argument* nach Toulmin (2003).

Die Frage nach dem Einfluss des *warrants* kann durch den **qualifier** verdeutlicht werden. Der *qualifier* kann die Aussagekraft der *data* inkl. des *warrant* einschränken, oder auch vollumfänglich freigeben. Der *qualifier* gibt so eine Art Grad der Aussagekraft an, ist dabei aber unabhängig vom *warrant* selbst (Toulmin, 2003). Bezogen auf das obige Beispiel könnte der *qualifier* einschränken, dass bspw. nur das situationale Interesse abgefragt wurde, welches

² Im seinem Beispiel beschreibt Toulmin (2003), dass eine Person rote Haare hat. Durch den *warrant* „Wenn etwas rot ist, kann es nicht gleichzeitig schwarz sein.“ kann der *claim* unterstützt werden, dass die Haare der Person schwarz sind.

ggf. nicht auf das individuelle Interesse übertragen werden kann. Damit wird die Aussagekraft der *data* bzgl. des *claims* eingeschränkt. Neben dem *qualifier* kann auch das **rebuttal** angegeben, ob ein *warrant* inkl. der zu Grunde gelegten *data* den jeweiligen *claim* unterstützen. Dabei zeigt das *rebuttal* Bedingungen auf, unter denen die Aussagekraft des *warrant* verworfen werden muss. Werden die im *rebuttal* beschriebenen Bedingungen erfüllt, ist die Aussagekraft der Argumentation in Bezug zum jeweiligen *claim* zu verwerfen (Toulmin, 2003). Im obigen Beispiel könnte ein *rebuttal* durch eine mögliche Konfundierung mit einem anderen Konstrukt, wie z.B. der Motivation erkannt werden. Sollte das verwendete Testinstrument nicht geeignet sein, um das Interesse zu messen, muss die gesamte Argumentation verworfen werden. Das *rebuttal* gibt damit Bedingungen an, unter welchen die Argumentation nicht zulässig ist. Selbst wenn alle Bedingungen zur Unterstützung des *claims* erfüllt sind, könnten generelle Zweifel den *claim* schwächen. Diese generellen Zweifel würden am *warrant* ansetzen. Der *warrant* beschreibt die Angemessenheit der *Daten* in Bezug zu einem spezifischen *claim*. Das bedeutet allerdings nicht, dass der *warrant* nicht generell in Zweifel gezogen werden könnte. In diesem Zusammenhang gibt Toulmin (2003, S. 96) das folgende Beispiel: „You presume that a man born in Bermuda can be taken to be a British subject,’ he may say, ‘but why do you think that?’“. Ein *warrant* wird nach Toulmin (2003) selbst durch verschiedene Annahmen gestützt, die die generelle Anwendbarkeit des *warrant* auf die *data* erlauben. Diese Annahmen bilden das **backing** für den *warrant*. Im obigen Beispiel könnte das *backing* die Aussage untermauern, dass sich das Interesse einer Person nicht innerhalb kürzester Zeit ändert. Dazu könnten Studien zur Langzeituntersuchung des Interesses angeführt werden.

Wie deutlich wurde, ist die Beziehung zwischen *data* und *claim* nicht einfach linear, sondern wird durch unterschiedliche weitere Aspekte beeinflusst. Im Sinne des funktionalen Evidenzverständnisses von Bromme et al. (2014) stellt der **claim** im *Model of argument* die Zieldimension dar. Der *claim* soll durch den Argumentationsprozess unterstützt werden. Somit ist die Argumentation darauf ausgerichtet den *claim* zu stützen, wodurch sie den Charakter von Evidenzen nach Bromme et al. (2014) erfüllen könnte.

Das *Model of argument* stellt die vielfältigen Bedingungen und Zusammenhänge zwischen einzelnen Elementen einer Argumentation dar. Toulmin (2003) macht durch seine Ausführungen auf die Komplexität des Prozesses aufmerksam. Einen *claim* durch *data* zu unterstützen, bedarf demnach vieler Bedingungen und Annahmen. Toulmin (2003) selbst schlägt in seiner Beschreibung des *Model of argument* den Bogen zum Syllogismus. Der Syllogismus als Form des deduktiven Schließens geht zurück auf die philosophische Logik von Aristoteles (Mittelsten Scheid & Hößle, 2008). Im Syllogismus führen zwei Prämissen zu einer Schlussfolgerung (Johnson-Laird, 1980). Dabei können in unterschiedlichen Konzeptualisierungen auch mehr Prämissen eingefügt werden (z.B. vier Prämissen im normativen Syllogismus, vgl. Mittelsten Scheid & Hößle, 2008). Unabhängig von der Anzahl der Prämissen, führen diese linear zur Schlussfolgerung, welche inhaltlich nichts enthält, was nicht durch die Prämissen bereits vorgegeben wurde. Die Schlussfolgerung folgt aus den Prämissen und ist entsprechend nur dann gültig, wenn die Prämissen gültig sind (Mittelsten Scheid & Hößle, 2008). Übertragen auf das *Model of argument* von Toulmin (2003) stellen vor allem der *warrant* und das *backing* Prämissen im Sinne des Syllogismus dar. Im Syllogismus verwendete Quantoren (Spektrum.de, o.D.) könnten im Modell von Toulmin (2003) im *qualifier* erkannt werden. Hier wird der Grad der Aussagekraft einer Schlussfolgerung beschrieben. Im Syllogismus treten Quantoren vor allem im Bereich der Prämissen auf. Diese Quantoren können einen Einfluss auf die Akzeptanz von Schlussfolgerungen haben und so zum ‚Atmosphäreneffekt‘ führen (Johnson-Laird, 1980). Auch für die Reihenfolge der Prämissen könnte ein Effekt auf die Schlussfolgerungen aufgezeigt werden (Johnson-Laird, 1980). Inwiefern beide Effekte auch im *Model of argument* eine Rolle spielen könnten, ist offen.

In Bezug auf die unterschiedlichen Weltanschauungen mit potentielltem Einfluss auf Argumentationsprozesse, kann das Modell von Toulmin (2003) nicht eindeutig einer Position zugeordnet werden. Während die Elementarisierung der Argumentation Hinweise auf eine mechanistische Sichtweise geben könnte, zeigt z.B. das Einbringen des *qualifiers* eindeutige Referenz zur systemischen Weltanschauung auf. Toulmin (2003) betont in seinen Ausführungen, dass ein *warrant* gemeinsam mit den *data* einen *claim* nur mit einer bestimmten Aussagekraft unterstützen kann. Hier drückt sich der Gedanke der Wahrscheinlichkeit aus, welcher für ein systemisches Verständnis grundlegend ist. Der Grad mechanistischer Anschauung hängt eng mit der Nutzung, bzw. dem spezifischen *claim* zusammen. Stellt der *claim* ein Einzelphänomen im Zusammenhang eines größeren Problems dar, kann die Nutzung ggf. als mechanistisch beschrieben werden. Soll der *claim* aber ein größeres Problem mit seinen komplexen Zusammenhängen lösen, stehen eher systemische Betrachtungen im Fokus. Somit kann die Nutzung des *Model of argument* über den Grad einer mechanistischen oder systemischen Betrachtung entscheiden.

Das *Model of argument* von Toulmin (2003) stellt nur ein Beispiel für die Modellierung des Argumentationsprozesses dar. Der Rückbezug auf *data* und weitere, notwendige Annahmen zur Erstellung von Schlussfolgerungen eignet sich im Sinne dieser Arbeit besonders gut, um den Umgang mit Evidenzen näher zu untersuchen.

2.3 Evidence-based practice

Das Konzept der Evidenzbasierung spielt in verschiedenen Professionen eine Rolle (Bauer & Prenzel, 2012; Breckon & Dodson, 2016; Davies, 1999). Die Grundannahme evidenzbasierter Praxis in den verschiedenen Professionen ist dabei, „dass der Einbezug wissenschaftlichen Wissens dazu beiträgt, professionelles Handeln und Entscheiden rational zu begründen und seine Erfolgswahrscheinlichkeiten zu erhöhen“ (Bauer & Kollar, 2023, S. 124). Die Ausprägung evidenzbasierter Praxis kann dabei zwischen den unterschiedlichen Professionen teils sehr unterschiedlich ausfallen. Während evidenzbasierte Praxis z.B. im Gesundheitssektor institutionell verankert ist, wird im Bildungsbereich häufig ein Gap zwischen Forschung und Unterrichtspraxis beschrieben (Broekkamp & van Hout-Wolters, 2007). Voss (2022) hebt dabei vor allem die Rolle persönlicher Überzeugungen hervor, welche teils durch eigene Unterrichtserfahrung aus Perspektive der Lernenden gebildet wurden. Diese Überzeugungen können den Zusammenhang zwischen Forschung und Unterrichtspraxis verdecken und somit hinderlich bei der Umsetzung evidenzbasierter Techniken sein.

Vor allem im Gesundheitssektor zeigt sich eine lange Forschungstradition und Entwicklung evidenzbasierter Ansätze. Ein möglicher Grund dafür könnte der folgende sein:

Ausgangspunkt des EBP-Ansatzes [evidence-based practice] ist es, dass viele traditionelle Maßnahmen der Gesundheitsversorgung – seien sie pharmakologisch, chirurgisch, physiotherapeutisch, sprachtherapeutisch etc. – in ihrer Wirksamkeit nicht überzeugend belegt sind. Dadurch besteht die Gefahr, dass sie wirkungslos sind, somit überflüssig (und auch ökonomisch unvertretbar) oder sogar schädlich. (Cholewa et al., 2015, S. 20)

Im Gesundheitssektor scheint eine Notwendigkeit der Überprüfung der Wirksamkeit von medizinischen Behandlungen zwingend nötig und u.a. lebenserhalten wichtig zu sein. Einerseits sind Nachteile für die Gesundheit der Patient*innen zu befürchten, auf der anderen Seite können ökonomische Verluste entstehen. Daher hat sich in den 1990er Jahren eine Bewegung zu *evidence-based medicine* gebildet (Dawes et al., 2005). Der Ansatz hat sich nach und nach auf weitere Bereiche des Gesundheitssektors neben dem ärztlichen Dienst ausgeweitet. So wurde z.B. auch im

Bereich der Pflege ein evidenzbasierter Ansatz populär. Davon ausgehend, plädieren Dawes et al. (2005) den Begriff *evidence-based medicine* zu *evidence-based practice* (EPB) zu erweitern. Die Entwicklungen im Bereich evidenzbasierter Praxis in der Gesundheitsversorgung bieten wichtige Impulse für eine Übertragung auf den Bildungsbereich. Im Folgenden sollen daher relevante Konzeptualisierungen vorgestellt werden.

Im Zuge der Fokussierung auf EBP in verschiedenen medizinischen Bereichen, sind verschiedene Definitionen entstanden, die jeweils die Eigenheiten der jeweiligen Bereiche abbilden. Eine viel zitierte Definition stammt von Sackett et al. (1996). David Sackett gilt als Pionier der evidenzbasierten Medizin, gründete er doch 1967 das *Department of Clinical Epidemiology and Biostatistics* an der McMaster Universität in Hamilton (Kanada) (McMaster University, o.D.) und war später Gründungsdirektor des *Centre for Evidence-Based Medicine* an der Oxford University (England) (Heneghan, 2015). In ihrer Definition fokussieren Sackett et al. (1996) verschiedene Aspekte:

Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research. By individual clinical expertise we mean the proficiency and judgment that individual clinicians acquire through clinical experience and clinical practice. (S. 71)

In Verbindung zu Kapitel 2.1 müssen die Eigenheiten von Evidenzen beachtet werden. Diese sollen umsichtig und gewissenhaft genutzt werden. Hier könnten Argumentationsmodelle wie das *Model of argument* von Toulmin (2003) ansetzen. Im nächsten Schritt fordern Sackett et al. (1996) die Vermittlung von Evidenzen und der persönlichen Erfahrung. Damit stehen Erfahrung und Evidenzen gleichberechtigt nebeneinander. Die Autoren heben vor allem den mitfühlenden Umgang mit den Patient*innen als wichtigen Bestandteil der individuellen Erfahrung hervor. Sackett et al. (1996) beschreiben das Zusammenspiel von Evidenz und Erfahrung als wichtig:

Good doctors use both individual clinical expertise and the best available external evidence, and neither alone is enough. Without clinical expertise, practice risks becoming tyrannized by evidence, for even excellent external evidence may be inapplicable to or inappropriate for an individual patient. Without current best evidence, practice risks becoming rapidly out of date, to the detriment of patients. (S. 72)

Die Forderungen der Befürworter evidenzbasierter Medizin trafen auf Kritik von Seiten der praktizierenden Mediziner*innen. Der Einbezug von wissenschaftlicher Evidenz in jede Entscheidung wurde als überfordernd und entmündigend angesehen (Dawes et al., 2005). Derlei Kritik rückt das Verhältnis zwischen Evidenzen und Erfahrungen in den Mittelpunkt der Diskussion. Auf der anderen Seite wurde ein Bedarf für hochwertige Evidenzen zur optimalen Versorgung der Patient*innen anerkannt. Dabei wird aber auch deutlich, dass die Einbindung hochwertiger Evidenzen in sich Probleme beinhalten kann. So könnten fehlerhafte Randomisierungen innerhalb berichteter Studien oder der *publication bias* zu Überschätzungen bei der Wirksamkeit von Therapien führen, welche sich negativ auf die Gesundheit der Patient*innen auswirken könnten (Dawes et al., 2005). Ebenfalls liegen nicht zu allen spezifischen Situationen hochwertige Evidenzen vor. Hier sollte nach den nächst-besten Evidenzen gesucht und die eigene Erfahrung eingebracht werden (Sackett et al., 1996). Die hier beschriebenen Herausforderungen stehen in engem Kontakt mit den beschriebenen Definitionsmerkmalen von Evidenzen (vgl. Kap. 2.1). So sollten immer die best-möglichen Informationen als Evidenzen angeführt werden, was jedoch nicht gleichzeitig bedeutet, dass zu jedem Bereich qualitativ hochwertige Evidenzen existieren. In Bezug auf die Quelle von Evidenzen

erhalten in diesem Zusammenhang gesammelte Erfahrungen eine wichtige Bedeutung (Breckon & Dodson, 2016).

Im Zuge der *Second International Conference of Evidence-Based Health Care* in 2003, entstand das sogenannte *silicy statement*. In diesem haben sich Forschende und Lehrende auf eine gemeinsame Definition von evidenzbasierter Praxis geeinigt. Ein Aspekt, der durch das *silicy statement* aufgeworfen wird, ist die Rolle der Patient*innen. So wird beschrieben: „These decisions should be made by those receiving care, informed by the tacit and explicit knowledge of those providing care, within the context of available resources“ (Dawes et al., 2005, S. 1). Dadurch wird eine weitere Facette evidenzbasierter Praxis eröffnet. Während Patient*innen oft nicht über das notwendige Fachwissen zur Einschätzung von Therapieoptionen verfügen, müssen sie dennoch Entscheidungen treffen, deren Auswirkungen gravierend sein können. Dadurch wird es zur Aufgabe der Mediziner*innen, wissenschaftliche Erkenntnisse zu kommunizieren bzw. das Zusammenspiel aus persönlicher Erfahrung und wissenschaftlicher Evidenz transparent darzustellen. So schließen sich Kompetenzanforderungen der Wissenschaftskommunikation an. Dabei stellen Dawes et al. (2005) die durch Mediziner*innen gezogenen Schlussfolgerungen als subjektiv dar. Die Schlussfolgerungen werden (im Sinne von Argumentationsmodellen) als subjektiv aus Informationen erstellt beschrieben. Damit spielt schon in der Argumentation die persönliche Erfahrung eine relevante Rolle. Bezogen auf das *Model of argument* könnte hier vor allem die Bewertung von *warrant*, *backing* und *rebuttal* eine entscheidende Rolle spielen (Toulmin, 2003).

Im Laufe der Entwicklung evidenzbasierter Praxis im Gesundheitssektor, sind auch institutionalisierte Rahmenbedingungen entstanden. So gibt es bspw. diverse Fachdatenbanken, in denen medizinische Evidenzen gefunden werden können (z.B. *Physiotherapy Evidence Database*, etc.) (Dawes et al., 2005). Ebenso haben sich Metadatenbanken entwickelt (z.B. *PubMed*), die teilweise wissenschaftliche Evidenzen sichten und im Sinne eines *systematic reviews* zusammenfassende Empfehlungen geben (z.B. *Cochrane*). Weiterhin wurde bspw. in der Bundesrepublik Deutschland das unabhängige *Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen* [IQWiG] gegründet. Das IQWiG erstellt evidenzbasierte Gutachten (im Sinne von *systematic reviews*) zu Arzneimitteln und Therapieformen. Dabei verpflichtet es sich explizit der evidenzbasierten Medizin (IQWiG, o.D.). Innerhalb des Gesundheitssektors könnten derlei Institute jedoch auch zu einer kritischen Betrachtung evidenzbasierter Medizin beitragen. So formulieren Sackett et al. (1996) die Gefahr, dass evidenzbasierte Medizin als Methode zur Kostenreduktion verstanden werden könnte. Leitlinien könnten die klinische Freiheit der Mediziner*innen verletzen und individuelle Kontexte außen vor lassen. So könnte durch den Verlust der kontextuellen Einbettung von medizinischen Entscheidungen auch Einfluss auf die Argumentationsprozesse genommen werden. Leitlinien könnten unreflektiert und ungeachtet entgegenstehender Überzeugungen umgesetzt werden. Dies könnte als weitere Entmündigung wahrgenommen werden, oder sich negativ auf die tiefenstrukturelle Begründung in Argumentationsprozessen auswirken. Diesen Befürchtungen begegnen die Autoren durch den Einbezug der persönlichen, klinischen Erfahrung der Mediziner*innen in evidenzbasierter Praxis.

Im Gesundheitssektor wird evidenzbasierte Praxis prozedural durch fünf Schritte modelliert (Cholewa et al., 2015; Dawes et al., 2005):

1. Ask Unsicherheiten in beantwortbare Fragen überführen
2. Acquire Systematische Suche nach den besten verfügbaren Evidenzen
3. Appraise Kritische Bewertung der Evidenzen bzgl. Validität, klinischer Relevanz und Anwendbarkeit

- | | | |
|----|--------|--|
| 4. | Apply | Anwendung der Ergebnisse in der Praxis |
| 5. | Assess | Evaluation der Performanz |

Dabei können die beschriebenen Schritte zugleich als Teilkompetenzen für die evidenzbasierte Praxis verstanden werden (Cholewa et al., 2015; Dawes et al., 2005). Ausgangspunkt ist die Erstellung einer präzisen Fragestellung (*Ask*). Dabei sollte die Fragestellung so präzise formuliert werden, dass relevante Evidenzen gefunden werden können (Cholewa et al., 2015). Im Gesundheitssektor wird dazu häufig auf das PICO-Schema zurückgegriffen (Cholewa et al., 2015). Dieses Schema fokussiert auf die Patient*innen („P“), eine mögliche Intervention („I“), eine vergleichbare Intervention („C“ für *comparison*) sowie das Ergebnis der Intervention („O“ für *outcome*). Das PICO-Schema erlaubt es eine möglichst konkrete Fragestellung zu erstellen, welche spezifisch auf die Situation angepasst ist (Cholewa et al., 2015). Im nächsten Schritt werden passende Evidenzen recherchiert (*Acquire*). Dabei werden bestehende, einschlägige Datenbanken genutzt. Entscheidend ist, dass passende Suchbegriffe gewählt werden. Diese sollten sich im Idealfall eng an der Fragestellung aus dem PICO-Schema orientieren (Cholewa et al., 2015). Anschließend müssen die gefundenen Evidenzen evaluiert werden (*Appraise*). Hierzu sollten forschungsmethodologische Qualitätskriterien überprüft werden. Die genutzten Methoden, abgeleitete Ergebnisse sowie die Validität und Reliabilität der Befunde sollen genau analysiert werden (Cholewa et al., 2015). Wurden hochwertige Evidenzen gefunden, sollte im nächsten Schritt die Passung auf den spezifischen Kontext überprüft werden (*Apply*). Hier gilt es zu prüfen, ob die institutionellen Rahmenbedingungen sowie die personellen Gegebenheiten eine positive Implementation der Evidenzen erlauben (Cholewa et al., 2015). Durch diesen Schritt wird die Rolle des spezifischen Kontexts explizit hervorgehoben. Ist dies der Fall, kann die Behandlung entsprechend der Evidenzen erfolgen. Im letzten Schritt wird dann überprüft, ob die Behandlung zum Erfolg geführt hat (*Assess*). Dazu müssen Evaluationstechniken eingesetzt werden. Die hier gewonnenen Erfahrungen könnten dem Forschungsprozess erneut zugeführt werden (Cholewa et al., 2015).

Dawes et al. (2005) weisen auf die Bedeutung von balancierten Kompetenzen innerhalb der verschiedenen Teilkompetenzbereiche hin. Demnach sei es wichtig, dass Akteur*innen im Gesundheitssektor in allen Teilkompetenzen gefördert werden. Sind diese Akteur*innen nur in einzelnen Prozessschritten kompetent, könnte dies den gesamten Prozess ins Stocken geraten lassen (Dawes et al., 2005). Weiterhin beschreiben die Autor*innen einen grundlegenden Prozessschritt, noch vor den beschriebenen Schritten. Eine Grundvoraussetzung sei, dass die Akteur*innen Unsicherheiten erkennen und eingestehen. Dies sei gleichzeitig der schwierigste Prozessschritt und sollte daher schon in die Ausbildung von Akteur*innen im Gesundheitssektor einfließen.

Die große Bedeutung evidenzbasierter Praxis innerhalb des Gesundheitssektors zeigt sich auch in der Vielzahl entwickelter Interventionen zur Förderung (angehender) Akteur*innen in eben dieser (für einen Überblick siehe z.B. Albarqouni et al., 2018; Larsen et al., 2019; Patellarou et al., 2020; Scurlock-Evans et al., 2014). Die verschiedenen Interventionen fokussieren dabei häufig auf einzelne medizinische Fachbereiche oder unterschiedliche Zeitpunkte in der Aus-/Fortbildung von Akteur*innen im Gesundheitssektor. Untersuchungen bzgl. der Kompetenz von Akteur*innen zeigen eine moderat ausgeprägte Kompetenz in evidenzbasierter Praxis auf (Ashktorab et al., 2015; McCluskey & Bishop, 2009). Durch eine Vielzahl unterschiedlicher Testinstrumente, bzw. der fortwährenden Adaption bestimmter Testinstrumente für verschiedene medizinischen Fachbereiche, ist ein Vergleich unterschiedlicher Stichproben nur bedingt möglich.

Eine weitere Folge der langen Forschungstradition im Bereich evidenzbasierter Medizin, ist die Vielzahl an entstandenen Leitlinien und Empfehlungen zur Behandlung unterschiedlicher Erkrankungen. Leitlinien können dabei als komprimierte Übersichtsarbeiten verstanden werden, die konkrete Handlungsempfehlungen geben (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V.). In Deutschland werden die Leitlinien durch die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. [AWMF] registriert und bzgl. ihrer Qualität eingestuft. Allein im Leitlinienregister³ der AWMF finden sich mehrere hundert Leitlinien, welche bei der Behandlung von Patient*innen Orientierung geben sollen. Diese Vielzahl an Leitlinien könnte schnell zu einer Überforderung führen oder praktisch nicht umsetzbar sein. Greenhalgh et al. (2014) zeigen mit einem Praxis-Beispiel eindrücklich die Herausforderungen auf. Die Autor*innen berichten von einer 24-Stunden Schicht in einem Akutkrankenhaus. In dieser Zeit sind 18 Patient*innen mit 44 Diagnosen versorgt worden. Zu den Diagnosen kämen 3679 Seiten nationaler Leitlinien in Frage, ein geschätzter Leseaufwand von 122 Stunden. Nach dieser Darstellung wird schnell klar, dass die praktische Umsetzung hier an ihre Grenzen kommt. Weiterhin kritisieren Greenhalgh et al. (2014), dass die Arbeit mit Leitlinien den Schwerpunkt der Behandlung weg vom Ziel dieser und hin zu den Mitteln verschiebe. So werden spezifische Einzelfälle in Gruppen subsummiert, welche durch Evidenzen oder Leitlinien vorgegeben werden. In Leitlinien notwendige Verallgemeinerungen könnten so spezifische Kontextmerkmale verdrängen. Einen letzten Kritikpunkt an Leitlinien (als Produkt evidenzbasierter Medizin) erkennen Greenhalgh et al. (2014) in der fehlenden Passung dieser mit der Realität. Während Leitlinien Empfehlungen für spezifische Erkrankungsmuster geben, liegen in der Realität häufig multiple Erkrankungen nebeneinander vor. Damit sind mehrere Variablen gleichzeitig bei der Wahl der passenden Therapie zu beachten. Durch die vielen Kombinationsmöglichkeiten unterschiedlicher Erkrankungen, und die mögliche Kontraindizierung einer Therapie durch weitere Erkrankungen scheint die Bildung von Leitlinien wenig effektiv. Die hier formulierte Kritik an Leitlinien schließt sich an die oben formulierte Kritik an evidenzbasierter Praxis insgesamt an. Ein zentraler Kritikpunkt ist die Kontextabhängigkeit medizinischer Entscheidungen, welche in einzelnen Studien, oder summarischen Leitlinien nicht abgebildet werden könne. Befürworter von EBP könnten hier die persönliche Erfahrung der Mediziner*innen hervorheben, welche konstitutiv in EBP verankert ist.

Insgesamt zeigt sich, dass die Idee der evidenzbasierten Medizin immer mehr Einfluss auf die Arbeit von Akteur*innen im Gesundheitssektor gewinnt. Im Laufe der Forschungstradition in diesem Bereich fand eine immer deutlichere Diversifizierung statt. In immer mehr medizinischen Fachbereichen wurden Spezifika evidenzbasierter Praxis ausgearbeitet. Die Formulierung von spezifischen Leitlinien soll zur Verbesserung der medizinischen Versorgung insgesamt beitragen. Somit wurde EBP in die Curricula aufgenommen und institutionell verankert. Unterschiedliche Ansätze der Förderung von Akteur*innen im Gesundheitssektor in EBP zeigen unterschiedliche Erfolge. Diese Erfolge können aufgrund der unterschiedlichen Messinstrumente nur bedingt miteinander verglichen werden. Die bestehende Literaturlage zeigt gleichzeitig den Bedarf nach der Förderung der Akteur*innen in evidenzbasierter Praxis auf.

2.4 Evidence-based teaching

Auf Basis der bestehenden Untersuchungen im Feld evidenzbasierter Medizin, fand auch eine Übertragung auf Lehr-Lern-Prozesse statt. Forschung im Bereich evidenzbasierter Praxis im

³ <https://register.awmf.org/de/start>

Bildungsbereich blickt auf eine 20-jährige Forschungstradition zurück, in welcher Bauer und Kollar (2023) stellenweisen Konsens erkennen.

Ausgangspunkt der Forschung zu evidenzbasierter Praxis im Bildungsbereich stellt die dargestellte Forschung im Zusammenhang mit evidence-based medicine dar (vgl. Kap. 2.3). Die Ausdifferenzierung des Konzepts evidenzbasierter Praxis in den einzelnen medizinischen Fachbereichen hat zu einer Öffnung des Konzepts beigetragen, die auch eine Übertragung auf andere Professionen, z.B. den Bildungsbereich, erlauben könnte. Bevor auf spezifische Konzepte einer Evidenzbasierung im Bildungsbereich eingegangen werden soll, wird ein Vergleich mit dem Feld evidenzbasierter Medizin angestrebt.

Ein Vergleich zwischen Bildungs- und Gesundheitsbereich kann grundlegende Gemeinsamkeiten und Unterschiede der Bereiche aufdecken. Davies (1999) vergleicht in seinem Artikel die Entwicklungen in den beiden Bereichen *evidence-based medicine* und *evidence-based education* auf Grundlage der Ähnlichkeiten und Unterschiede der beiden Bereiche miteinander. Albrecht (2016) gibt in ihrer Arbeit einen umfassenden Überblick über Ähnlichkeiten und Unterschiede in der Arbeit, bzw. dem Belastungserleben von Lehrkräften und Mediziner*innen.

Tabelle 1 – Gemeinsamkeiten und Unterschiede von Gesundheitssektor und Bildungsbereich

Gemeinsamkeiten	Unterschiede
Herausforderungen bzgl. der Komplexität, Kontextabhängigkeit, Messung und Kausalität von beobachteten Effekten (Davies, 1999)	Unterschiedliche Forschungsmöglichkeiten in den Bereichen (Im Gesundheitssektor sind auch (quasi) experimentelle Designs, sowie doppelt verblindete randomisierte Kontrollstudien möglich Im Bildungsbereich werden eher Vergleichsstudien geführt.) (Davies, 1999)
Unterschied im Verhalten von Personen im Krankenhaus/in der Schule und im Alltag (Davies, 1999)	Umfassende Datenbanken (inkl. erstellter systematischer Reviews) wie Cochrane fehlen für den Bildungsbereich (Davies, 1999).
Einsatz von Evidenzen erfordert Anpassung auf spezifischen Kontext (Davies, 1999)	Schlechtere finanzielle Förderung von Forschung im Bildungsbereich im Vergleich zum Gesundheitssektor (Davies, 1999).
Eingebundenheit der Disziplinen in naturwissenschaftliche und kultur-/gesellschaftswissenschaftliche Bereiche (Davies, 1999)	Unterschiedliches ‚epistemisches Klima‘ in Medizin- und Lehramtsstudium und daraus folgende Unterschiede in Forschungs- & Praxisorientierung im Rahmen der Ausbildung (Rochnia et al., 2020)
Nutzung qualitativer und quantitativer Verfahren (Davies, 1999)	Unterschiede in Arbeitsformen (Mediziner*innen arbeiten meist in interdisziplinären Teams, Lehrkräfte eher allein) (Albrecht, 2016)
Nutzung von systematischen Reviews, oder Meta-Analysen (Davies, 1999)	Fehlendes Berufsgeheimnis für Lehrkräfte (Albrecht, 2016; Rothland, 2013)
Großer Einfluss auf Patient*innen/Schüler*innen, mit möglichen weitreichenden Konsequenzen, asymmetrisches Verhältnis zu Patient*innen/Schüler*innen (Albrecht, 2016; Davies, 1999)	Erhöhte Interdisziplinarität bei Lehrkräften (durch zweites Unterrichtsfach) (Rochnia et al., 2020)
Ethische Problemstellungen (Albrecht, 2016; Davies, 1999)	
Normbildung durch hergebrachte Praxis und Autoritäten (Davies, 1999) (vgl. „normative Kraft des Faktischen“, Jellinek & Jellinek, 1921)	
Fehlende Rückmeldung über langfristige Erfolge bei gleichzeitiger schneller Rückmeldung über Misserfolge (Albrecht, 2016; Davies, 1999)	

Wie in Tabelle 1 dargestellt, können einige Gemeinsamkeiten zwischen Gesundheitssystem und Bildungsbereich erkannt werden. Dabei sticht vor allem die Komplexität der jeweiligen Kontexte

hervor (Berliner, 2002; Davies, 1999; Rothland, 2013). Sowohl im Bildungsbereich als auch im Gesundheitssektor müssen viele Variablen für die Auswahl von Methoden oder Therapien beachtet werden. Berliner (2002) zeigt die Kontextabhängigkeit vor allem für den Bildungsbereich als Herausforderung auf. Auch Rothland (2013, S. 29) beschreibt die Vielzahl von Anforderungen an Lehrkräfte (Unterrichten, Erziehen, Beurteilen, Beraten, individuelle Kompetenzentwicklung, Schulentwicklung), welche zu einer großen „Komplexität der Berufsausübung“ führen. Im ersten Schritt müssen entsprechend Diagnosen gestellt werden, um relevante Einflussvariablen auf einen Zusammenhang zu erkennen. Für die notwendige Diagnostik wurden sowohl im Gesundheitssystem (PICO-Schema), als auch im Bildungsbereich (diagnostische Tests) unterschiedliche Methoden entwickelt. Die Übertragung von Fragestellungen in Messungen zur Untersuchung kausaler Zusammenhänge stellt in beiden Bereichen eine Herausforderung dar. Dabei unterscheiden sich jedoch die forschungsmethodologischen Zugänge und Möglichkeiten (Bromme et al., 2014; Davies, 1999). Hier schlägt sich erneut die Kontextabhängigkeit nieder. Die Kontextabhängigkeit zeigt sich in der Bildungsforschung in der Vielzahl und Komplexität vorhandener Interaktionen im Klassenraum. So seien die Untersuchungsgegenstände innerhalb empirischer Bildungsforschung oft Interaktionen der 10.-15. Ordnung (Berliner, 2002). Dies meint, dass eine Vielzahl beobachtbarer Interaktionen aus Zusammenhängen entspringen, welche mit dem eigentlich Untersuchungsgegenstand nichts/wenig zu tun haben, z.B. sozioökonomischer Status, Intelligenzquotient, Lernmotivation, etc. Dadurch wird die kausale Begründung von beobachtetem Verhalten erschwert. Weiterhin existieren im Gesundheitssektor anerkannte Evidenzhierarchien, welche auf der Kategorisierung von Evidenzen bzgl. ihrer Qualität basieren. Wenngleich die Qualität von Evidenzen innerhalb eines Kontinuums beschrieben wird, könnte die Kategorisierung von Evidenzen und den daraus abgeleiteten Leitlinien zur Orientierung beitragen (Guyatt et al., 2008). Aufgrund dieser Überlegungen sind unter Zuhilfenahme unterschiedlicher Methoden (Zeng et al., 2015) eine Vielzahl von Modellierungen der Qualität von Evidenzen entstanden (z.B. Borgetto et al., 2020; Guyatt et al., 2008). Nationale (z.B. AWMF e.V.) und internationale Organisationen (z.B. Cochrane Collaboration) orientieren sich in der Formulierung von Leitlinien an der Qualität von Evidenzen (zumeist in Anlehnung an das GRADE-System, Guyatt et al., 2008) und graduieren so die formulierten Leitlinien bzgl. ihrer Aussagekraft. An der Spitze formulierter Evidenzhierarchien stehen zumeist systematische Reviews und Metaanalysen (Wilkes & Stark, 2022). Die Übertragung medizinischer Evidenzhierarchien auf den Bildungsbereich wird kritisch diskutiert:

Im Bildungsbereich steht die Forschung außerdem vor einigen fast unüberwindbar scheinenden Problemen. Zum Beispiel sind eine konsequente Randomisierung (z. B. zufällige Zuordnung von Schülerinnen und Schüler zu Klassenverbänden und Lehrkräften) oder die Kontrolle von potentiell relevanten Einflussfaktoren über längere Zeiträume (z. B. in den Kontexten Elternhaus, Freundeskreis, Mediennutzung, Klasse, Schule) zu gewährleisten. Im Ergebnis führt das dazu, dass die „Spitze“ der Evidenzhierarchie im Bildungsbereich sehr viel schwächer ausgeprägt oder besetzt ist als der „Rumpf“. (Bromme et al., 2014, S. 13–14)

Unterschiede in den forschungsmethodologischen Zugängen und Möglichkeiten der empirischen Bildungsforschung im Vergleich zur medizinischen Forschung kann die Qualität von Evidenzen (im Sinne von Evidenzhierarchien) verringern. So werden in der Bildungsforschung häufiger Studien geführt, welche nach medizinischen Evidenzhierarchien eher am unteren Ende des Kontinuums verortet werden würden (vgl. Schwerpunkt auf qualitative Forschung nach Berliner, 2002). Trotz formulierter Evidenzhierarchien sollte eine kritische Reflexion der Studien erfolgen. So sollten auch Metaanalysen (an der Spitze bestehender Evidenzhierarchien) aufgrund ihrer

methodologischen Umsetzung oder spezifischen Zielstellungen kritisch reflektiert werden (Beelmann, 2014; Pant, 2014).

Hinzu kommen auch institutionelle bzw. infrastrukturelle Unterschiede der beiden Bereiche. So heben Bromme et al. (2014) die Anstrengungen hervor, welche nötig sind, um Evidenzen zur Grundlage von Entscheidungen zu machen. Die im Medizinsektor etablierten Systeme (internationale Netzwerke zur Bewertung der Qualität von Evidenzen, systematische Leitlinienerstellung, Plattformen zum vereinfachten Konsum von Evidenzen, etc.) sind aktuell im Bildungsbereich wenig existent (Bromme et al., 2014; Davies, 1999). Zwar erkennen Bromme et al. (2014) erste Ansätze von vergleichbaren Netzwerken (z.B. *Campbell Collaboration*), jedoch bestünde weiterhin ein erhöhter Bedarf nach systematischen Übersichtsarbeiten. Gepaart mit der Unterfinanzierung empirischer Bildungsforschung im Vergleich zu medizinischer Forschung entstehe so der Unterschied bzgl. bereitstehender Evidenzen (Bromme et al., 2014; Davies, 1999). Ebenso gäbe es institutionelle Eingriffe im Bildungsbereich, welche zu einer geringen Halbwertszeit von Evidenzen führen (Berliner, 2002). Während Berliner (2002) in diesem Zusammenhang vor allem auf den Wandel sozialer Systeme und Normen fokussiert, stellt Spitzer (2010) vor allem politische Reformen ins Zentrum der Betrachtung. Im Gegensatz zur Medizin handle es sich bei Reformen im Bildungsbereich zumeist um inhaltliche Reformen, welche durch „Gruppeninteressen und ideologische Vorurteile“ (Spitzer, 2010, S. 7) geleitet werden. Nur selten stellen empirische Untersuchungen die Basis für Reformen da. Dabei stelle auch das föderale Organisationssystem der Bildung in Deutschland ein Problem für Durchlässigkeit und Chancengleichheit im Schulsystem sowie gleichzeitig für die Generalisierbarkeit empirischer Befunde dar.

Wie gezeigt wurde, stellen sich in der empirischen Bildungsforschung und der medizinischen Forschung ähnliche Herausforderungen dar, welche in ihrer spezifischen Ausprägung zwischen den Bereichen variieren. Die Orientierung von *evidence-based teaching* and *evidence-based medicine* wurde und wird kritisch diskutiert (für einen Überblick siehe J. Schrader, 2014). Neben Argumenten für eine Vergleichbarkeit der Bereiche, gibt es immer wieder auch Argumente, die eine größere Distanz zwischen den Bereichen verdeutlichen (Albrecht, 2016; Davies, 1999; Rochnia et al., 2020). Die eigene Positionierung in diesem Feld „hängt nicht zuletzt davon ab, welche Erwartungen man an die Qualität des Forschungswissens stellt, welche Hoffnungen oder Befürchtungen man mit einer wissenschaftlichen und linearen Steuerung in Politik und Praxis verbindet [...]“ (J. Schrader, 2014, S. 214). Auffällig in den durch J. Schrader (2014, S. 210) dargestellten Entwicklungen und Diskussionen bzgl. evidenzbasierter Praxis im Bildungsbereich ist, dass die Bildungswissenschaft (und damit relevante Perspektiven auf evidenzbasierte Praxis) als „eine kleine, aber wachsende Fraktion vornehmlich innerhalb der Erziehungswissenschaft, der Psychologie, der Soziologie und der Ökonomie“ beschrieben wird. Eine (fach-)didaktische Perspektive auf evidenzbasierte Praxis scheint in den bestehenden Diskursen keine Rolle gespielt zu haben. Dies könnte einerseits auf Probleme der Translation von Evidenzen zwischen Forschung und Praxis zurückzuführen sein, andererseits aber ebenfalls ein Potential darstellen, um evidenzbasierte Praxis im Bildungsbereich zu etablieren. Weitere Kritik am Konstrukt *evidence-based teaching* schließt an Kritik am Konstrukt *evidence-based medicine* an (siehe z.B. Bellmann & Müller, 2011b; Kvernbekk, 2011).

Wie dargestellt wurde, besteht eine Forschungstradition zum Abgleich evidenzbasierter Medizin und Bildungspraxis. In ihrer Zuspitzung auf eine reale Unterrichtssituation machen Bauer et al. (2015) deutlich, warum Evidenzbasierung auch im Bildungsbereich eine relevante Rolle spielen sollte. Am Beispiel der Frage nach den passenden Hausaufgaben für eine Unterrichtsstunde machen die Autoren deutlich, inwiefern bildungswissenschaftliche Evidenzen bei der

Entscheidungsfindung unterstützen können. Dabei heben sie noch einmal hervor, dass Evidenzbasierung nicht die Entscheidungsfindung ersetzt, sondern dabei unterstützen kann (vgl. Kap. 2.3). Auch Bauer und Kollar (2023) heben die Bedeutung evidenzbasierter Praxis für den Unterricht hervor. Im Sinne eines Querschnittsthemas beschreiben die Autoren die Rolle des bildungswissenschaftlichen Wissens im Zusammenhang evidenzbasierter Entscheidungen im Unterricht. Im Sinne einer Zusammenfassung der Forschung in diesem Bereich fokussieren Bauer und Kollar (2023) fünf Leitfragen zur weiteren Untersuchung. Die Leitfragen beschäftigen sich mit der Rolle und Güte von Erfahrungs- und Kontextwissen im Prozess der Entscheidungsfindung. Ebenso werden weitere Interventionsstudien gefordert, „die prüfen, unter welchen Bedingungen Lehramtsstudierende und Lehrkräfte wissenschaftliches Wissen tatsächlich rezipieren und nutzen können“ (Bauer & Kollar, 2023, S. 15). Dabei stellen die Autoren diese Leitfragen vor dem Hintergrund eines moderaten/liberalen Verständnisses evidenzbasierter Praxis (vgl. auch Stark, 2017). Dieses Verständnis ist zu unterscheiden von „strengerer Auslegung evidenzbasierter Praxis, die im Sinne eines *What Works*-Ansatzes primär die Verwendung von Maßnahmen und Methoden fordern, deren Wirksamkeit auf Basis randomisierter Kontrollgruppenstudien [...] nachweisbar ist [...]“ (Bauer & Kollar, 2023, S. 127). Stark (2017) bestärkt die Bedeutung individuellen Erfahrungswissens im Zusammenhang mit bestehenden Evidenzen. In einem moderaten Verständnis evidenzbasierter Praxis obliegt die Entscheidung der Lehrkraft auf Grundlage aller zur Verfügung stehenden Evidenzen. Auch dieser Arbeit liegt ein moderates Verständnis evidenzbasierter Praxis zugrunde. Weiterhin stellen die Autoren heraus, dass ein „Kulturwechsel“ (Bauer & Kollar, 2023, S. 17) der Bildungspraxis nötig sei. Durch diesen sollen Lehrkräfte dazu angeregt werden, wissenschaftliche Evidenzen (auch der Fachdidaktik) bei unterrichtlichen Fragestellungen zu nutzen und als relevant anzuerkennen.

Ein genereller Ansatz zur Förderung von Kompetenzen der Evidenzbasierung kann im SSCO-Ansatz erkannt werden (Rochnia et al., 2020; Schaeper & Weiß, 2016). Dieser Ansatz beschreibt die Integration evidenzbasierter Praxis in das Studium. Entlang der Dimensionen *Structure*, *Support*, *Challenge* und *Orientation* sollten Forschung und Forschungsergebnisse an möglichst vielen Stellen des Studiums präsent werden, durch Dozierende vermittelt und in der eigenen Tätigkeit erfahrbar werden. Ebenso sollte eine Einführung in Forschungsmethodik erfolgen (Rochnia et al., 2020). Dieser Ansatz bedarf der Verortung in Prüfungsordnungen und einer durchgängigen Anwendung innerhalb des Studiums. Hier wird erneut die Rolle von Bildungspolitik und -administration deutlich (Bromme et al., 2014). Wenglein et al. (2015) bemängeln die kleine Zahl vorhandener Interventionsstudien, welche die Förderung von Kompetenzen evidenzbasierter Praxis im Bildungsbereich untersuchen. Innerhalb ihrer eigenen Interventionsstudie stellen die Autor*innen die Wirksamkeit eines Heuristik-Trainings auf die Argumentationsqualität dar. Generell fokussieren bestehende Interventionen zumeist auf bildungswissenschaftliche Evidenzen (Bauer & Kollar, 2023).

Mit Blick auf den Chemieunterricht konnten Rahman und Lewis (2020) in ihrer um den *publication-bias* bereinigten Meta-Analyse die Effektivität evidenzbasierter Praxis für die Lernendenleistung aufzeigen. Dazu haben die Autoren „evidence-based instructional practices“ (Rahman & Lewis, 2020, S. 765) bzgl. ihrer Effektstärken im Rahmen von quasi-experimentellen und experimentellen Studien verglichen. Dabei fokussieren die Autoren auf Methoden wie *problem-based learning* oder auch *process oriented guided inquiry learning*. Der Fokus dieser Methoden liegt im Umgang mit Evidenzen im Unterricht. Nach Rochnia et al. (2020, S. 126) könnte hier von „forschungsorientiertem Denken“ gesprochen werden. Dieses beinhaltet nicht das eigene Erstellen von Evidenzen. Mit einer durchschnittlichen Effektgrößen-Range von $d=0.29-0.62$ ($N=94$) kann eine Effektivität aufgezeigt werden. Gleichzeitig zeigen die Autoren Herausforderungen bei der genaueren Analyse

und dem Vergleich unterschiedlicher Studien auf. So könnten Effekte teils aufgrund der Konfundierung mit anderen Variablen nicht zweifelsfrei interpretiert und betrachtet werden.

In Folge der beschriebenen Diskurse wurden Operationalisierungen evidenzbasierter Praxis entwickelt, die auf den Bildungsbereich fokussieren. So teilt Davies (1999) evidenzbasierte Praxis generell in zwei unterschiedliche Praxen ein: *using evidence* und *establishing evidence*. Dabei sind diese nicht zwangsläufig gegensätzlich, beschreiben aber zwei unterschiedliche Bezüge zu Evidenzen. So könnten Lehrkräfte Evidenzen für ihren Unterricht oder dessen Planung nutzen, müssten aber nicht selbst im Forschungsprozess aktiv sein (*using evidence*). Auf der anderen Seite könnten Lehrkräfte durch eigene, kleine Forschungsprojekte selbst Evidenzen erzeugen (*establishing evidence*). Auch wäre denkbar, dass die beiden Praxen von unterschiedlichen Personen praktiziert werden. Während Lehrkräfte Evidenzen eher nutzen, sind Forschende eher mit der Erstellung von Evidenzen befasst (wenngleich dies die Nutzung von Evidenzen einschließt). Tack und Vanderlinde (2016) differenzieren evidenzbasierte Praxis in unterschiedliche Dimensionen. Demnach gäbe es eine affektive Dimension, welche die Einstellung gegenüber Evidenzen, aber auch der eigenen Rolle als forschende Lehrkraft beinhaltet. Eine kognitive Dimension beinhaltet die selbstwahrgenommenen Fähigkeiten im Umgang mit Evidenzen als Konsument*in und Produzent*in. In der Verhaltensdimension wird das eigene Engagement im Umgang mit Evidenzen beschrieben. Auch hier kann es sich um das Konsumieren oder Generieren von Evidenzen handeln. Diese Unterscheidung kann mit der Operationalisierung von Davies (1999) verglichen werden. Beide Operationalisierungen stehen im Einklang mit dem Modell von Dawes et al. (2005). Ebenso erweitern beide Operationalisierungen das ursprüngliche Konzept: Davies (1999) ergänzt das ursprüngliche Modell um den Bereich *establishing evidence*, Tack und Vanderlinde (2016) zeigen eine affektive Dimension auf.

Die beschriebenen Konzepte können im Modell von Breckon und Dodson (2016) subsummiert werden. Breckon und Dodson (2016) haben ein systematisches Review von 36 systematischen Reviews geführt. Diese entspringen hauptsächlich dem Gesundheitssystem, wurden jedoch im zweiten Schritt durch ein *Scoping Review* in angrenzenden Fachbereichen (z.B. Sozialwissenschaften, etc.) ergänzt. Ziel der Autor*innen war es Mechanismen der Evidenzbasierung für die Wissenschaftskommunikation mit Politiker*innen zu synthetisieren. Die Mechanismen (vgl. Tab. 2) zeigen dabei Möglichkeiten auf, Evidenzen in Entscheidungsprozesse zu integrieren. Wenngleich der Fokus der Autor*innen auf dem Einsatz von Evidenzen in politischen Entscheidungsprozessen liegt, scheint eine Übertragung auf andere Bereiche möglich zu sein. Dazu schreiben Breckon und Dodson (2016, S. 28) selbst: „Too many of the reviews covered here are from the health sector. We need more primary studies from teaching, policing, social work and other areas of social policy“. Die Autor*innen fokussieren auf Prozesse zwischen Forschenden und Politiker*innen. Damit werden Akteur*innen unterschiedlicher Systeme adressiert. Bezogen auf den Bildungsbereich könnte hier die Vermittlung zwischen Forschenden und Lehrkräften synonym verwendet werden. Im Sinne einer forschenden Grundhaltung, welche durch die Standards der Lehrerbildung in den Fächern, wie auch den Bildungswissenschaften beschrieben wird, können Lehrkräfte in beiden Systemen verortet werden (KMK, 2019a, 2019b). Die Tätigkeit von Lehrkräften umfasst neben dem Unterrichten folglich auch die Rezeption und Implementation von Forschungsergebnissen. Damit könnten die evidence-use Mechanismen auch für Lehrkräfte interpretiert werden, ohne dass es einer weiteren Personengruppe bedarf. Vielmehr können diese auf verschiedene Zielgruppen übertragen werden. Für den Bildungsbereich ergibt sich dadurch die Möglichkeit die Mechanismen für Unterrichtssituationen zu adaptieren. Eine Passung mit den zuvor vorgestellten Konzepten unterstreicht dabei die Anwendbarkeit auch auf den Bildungsbereich. Im Folgenden sollen die Mechanismen dargestellt und in die bestehende Literatur eingebettet werden.

Tabelle 2 – Evidence-use Mechanismen nach Breckon und Dodson (2016, S. 6)

1	Awareness	Building awareness and positive attitudes towards evidence use
2	Agree	Building mutual understanding and agreement on policy-relevant questions and the kind of evidence needed to answer them
3	Access and Communication	Providing communication of, and access to, evidence
4	Interact	Facilitating interactions between decision-makers and researchers
5	Skills	Supporting decision-makers to develop skills accessing and making sense of evidence
6	Structures and Processes	Influencing decision-making structures and processes

Der erste Mechanismus (*Awareness*) befasst sich mit der Einstellung gegenüber Evidenzen. Innerhalb des systematischen Reviews konnten Breckon und Dodson (2016) keine Reviews finden, welche lediglich auf Awareness-Interventionen fokussiert haben. Dies kann dadurch begründet werden, dass die Mechanismen nicht in Gänze trennscharf abzugrenzen sind, bzw. sich die Förderung eines Mechanismus auch auf andere Mechanismen auswirken kann (vgl. auch Konfundierung innerhalb *evidence-based instructional practices*, Rahman & Lewis, 2020). Im Rahmen des *Scoping Reviews* konnten die Autor*innen jedoch die Notwendigkeit von Awareness-Building aufzeigen. So gehe es vor allem darum, evidenzbasierte Praxis als eine Norm zu etablieren. Dazu muss ein Wert in evidenzbasierter Praxis erkannt werden, welche zu einem intrinsischen Interesse führen kann, Evidenzen in Entscheidungsprozesse einzubinden. Gleichzeitig weisen die Autor*innen darauf hin, dass die Abwesenheit von Evidenzen bzgl. der Bedeutung dieses Mechanismus nicht Evidenzen für die Unwichtigkeit des Mechanismus sein müssen (Dies ist jedoch im Sinne eines logischen Fehlschlusses – *argumentum ad ignorantiam* – kein Beweis für die Notwendigkeit dieses Mechanismus, vgl. Halbedl, 2011). Basierend auf der Erwartungs-mal-Wert-Theorie ist eine notwendige, wertschätzende Einstellung gegenüber evidenzbasierter Praxis begründbar. Wird ein Wert in evidenzbasierter Praxis erkannt, z.B. dass diese zu guten Ergebnissen führt, könnte diese häufiger als Methode ausgewählt werden (Kuhl, 1983). Somit ist der evidence-use Mechanismus *Awareness* auch aus einer theoretischen Perspektive begründbar.

Der zweite Mechanismus (*Agree*) fokussiert auf die Formulierung relevanter Fragen und geeignete Evidenzen, um diese zu beantworten (Breckon & Dodson, 2016). Im Rahmen des systematischen Reviews konnten erneut keine Reviews gefunden werden, die eine alleinige Wirkung von *Agree*-Interventionen aufzeigen. In zwei der gefundenen Reviews wurden *Agree*-Anteile erkannt, jedoch in Kombination mit Techniken, welche im Rahmen dieser Konzeption anderen Mechanismen zugeordnet werden müssen. Im *Scoping Review* konnten einige Interventionen gefunden werden. Die Autor*innen heben vor allem *Journal Clubs* und *Delphi Panels* hervor. Durch diese Methoden könnten relevante Fragen gebildet, oder nach Antworten auf diese gesucht werden. Dabei weisen die Autor*innen jedoch darauf hin, dass durch diese Methoden auch andere Mechanismen angesprochen werden. So kann z.B. in *Journal Clubs* der *Awareness*-Mechanismus angesprochen

werden, indem evidenzbasierte Praxis als Norm innerhalb der Gruppe etabliert wird. Inhaltlich kann eine Nähe zur Teilkompetenz „Ask“ in der Modellierung von Dawes et al. (2005) erkannt werden.

Der dritte Mechanismus (*Access and Communication*) richtet sich vor allem an Forschende, die Evidenzen erstellen (Breckon & Dodson, 2016). Sowohl durch das systematische Review als auch das *Scoping Review* geben die Autor*innen Empfehlungen, wie Evidenzen vermarktet werden sollten. Dabei spielt die Ausrichtung an der Zielgruppe eine zentrale Rolle. Evidenzen müssten in ihrer Präsentation immer an die jeweiligen Bedürfnisse der Zielgruppe angepasst werden. Dies gilt sowohl hinsichtlich des Designs der Dokumente als auch in Bezug zu den genutzten Kommunikationskanälen. So sollte Social Media in die Kommunikation der Evidenzen einbezogen werden. Damit spielen Breckon und Dodson (2016) auf die Teilkompetenz „Acquire“ nach Dawes et al. (2005) an. Zwar beschäftigt sich dieser Mechanismus nicht mit Recherche-Techniken oder dem effektiven Suchen nach Evidenzen, jedoch scheint hier nur ein Perspektivunterschied zu bestehen. Während Dawes et al. (2005) sich damit auseinandersetzen, wie Evidenzen zu einem bestimmten Thema gefunden werden können, fokussieren Breckon und Dodson (2016) auf die Gestaltung von Evidenzen durch die Forschenden, damit diese gefunden werden könnten. Damit legen sie einen Fokus auf die Produzierenden von Evidenzen (im Sinne von *establishing evidence*, vgl. Davies, 1999), während Dawes et al. (2005) auf die Konsumenten von Evidenzen fokussieren (im Sinne von *using evidence*, vgl. Davies, 1999).

Ähnlich wie bei den bisher vorgestellten Mechanismen, begründet sich der nächste Mechanismus (*Interact*) ebenfalls vor allem durch das *Scoping Review* (Breckon & Dodson, 2016). Im Rahmen des systematischen Reviews konnte wiederum keine alleinige Wirksamkeit erkannt werden. Vorhandene Reviews basieren wiederum auf Studien, welche eine Kombination unterschiedlicher Mechanismen zur Förderung in evidenzbasierter Praxis implementieren. Im Rahmen des *Scoping Reviews* konnte vor allem die soziale Beeinflussung als relevant erkannt werden. Hier ist vor allem der Einfluss von „evidence champions, opinion leaders, messengers, role models and ‚change agents‘“ (Breckon & Dodson, 2016, S. 20) gemeint. Diese könnten durch ihren Einfluss nachhaltige Verhaltensänderungen einzelner Peers, oder gesamter Systeme induzieren. Dieser Mechanismus kann mit der Teilkompetenz „Apply“ nach Dawes et al. (2005) in Verbindung gebracht werden. Gerade in der Anwendung von Evidenzen im spezifischen Kontext könnte die Interaktion (auch im Sinne von Ermutigung) hilfreich sein. Im Sinne multiprofessioneller Teams kann die Interaktion von Personen unterschiedlicher Professionen gewinnbringend eingesetzt werden (MSB NRW, 2021b).

Der fünfte evidence-use Mechanismus („Skills“) fokussiert auf den sinnerschließenden Umgang mit Evidenzen (Breckon & Dodson, 2016). Wenngleich Design und Marketing von Evidenzen einen großen Einfluss auf den Umgang mit diesen haben können (vgl. Mechanismus „Access and Communication“), müssen bestimmte Fähigkeiten und Wissen erworben werden, um Evidenzen zu nutzen. Dazu zählt auch, dass zentrale wissenschaftliche Konzepte, Messgrößen, Qualitätskriterien, etc. verstanden werden müssen. Andernfalls könnten Evidenzen nicht verstanden, oder aber deren Qualität nicht eingeschätzt werden. Neben der zielgruppenspezifischen Aufarbeitung von Evidenzen muss in der Zielgruppe also auch ein bestimmtes Verständnis von Evidenzen und deren Charakteristika vorliegen. Dieser evidence-use Mechanismus zeigt eine Nähe zur Teilkompetenz „Appraise“ auf (Dawes et al., 2005). Vorhandenes Wissen über Evidenzen soll eingesetzt werden, um die Qualität vorliegender Evidenzen kritisch zu prüfen. Dabei lassen Breckon und Dodson (2016) offen, inwiefern auch eigene Forschungspraxis notwendig ist, um das notwendige Wissen über Evidenzen zu erarbeiten. Im Sinne der Unterscheidung von *using evidence* und *establishing*

evidence nach Davies (1999) könnten beide Zugänge zu einem fundierten Wissen über Evidenzen beitragen. Im Sinne von *establishing evidence*, könnten erworbene Kompetenzen und Wissensanteile auch auf die Teilkompetenz „Assess“ (Dawes et al., 2005) auswirken. Erworbene Kompetenzen in der Erstellung von Evidenzen (vor allem Aspekte der Evaluation und Argumentation) könnten bei der Bewertung der eigenen Implementation von Evidenzen unterstützend wirken. Kenntnisse über die Auswahl geeigneter Untersuchungsvariablen, grundlegender Qualitätskriterien, etc. könnten derlei Evaluationsprozesse strukturieren und im Sinne von Orientierungswissen erleichtern. Breckon und Dodson (2016) beschreiben vor allem Interventionen, welche auf den Umgang mit Evidenzen fokussieren und weniger solche, die die eigene Erstellung von Evidenzen beinhalten. Im Rahmen des *Scoping Reviews* verweisen die Autor*innen vor allem auf Konzepte der Erwachsenenbildung, welche in Interventionen gewinnbringend eingesetzt werden könnten.

Der letzte evidence-use Mechanismus („*Structures and Processes*“) nimmt eine systemische Perspektive ein (Breckon & Dodson, 2016). Wenngleich das systematische Review erneut keinen alleinigen Einfluss des Mechanismus auf die Nutzung von Evidenzen zur Entscheidungsfindung aufzeigen kann, gibt es auch im *Scoping Review* vielversprechende Hinweise auf einen Einfluss. So berichten die Autor*innen über positive Einflüsse von unterschiedlichen Anstößen zur Arbeit mit Evidenzen. Seien es institutionelle Vorgaben zur Nutzung von Evidenzen in politischen Entscheidungen oder vereinfachte Zugänge zu Evidenzen im privaten Bereich (z.B. das Einrichten einer Datenbank als Startseite des Internetbrowsers). Ebenso plädieren die Autor*innen für die Einrichtung von Evidenz-Zentren als Organisationen, welche bei der Implementation von Evidenzen in Entscheidungsprozesse im Sinne von *stake holdern* unterstützend wirken können (vgl. bestehende Strukturen im Bereich *evidence-based medicine*). Bezüglich der Teilkompetenzen nach Dawes et al. (2005) kann hier eine Nähe zu „*Apply*“ festgestellt werden. Die letztliche Nutzung und Anwendung von Evidenzen könnte systemisch durch verschiedene Prozesse und Strukturen beeinflusst/gefördert werden.

Breckon und Dodson (2016) konnten durch ihr umfassendes Review einer Vielzahl von Interventionen evidence-use Mechanismen postulieren, welche weitere Operationalisierungen evidenzbasierter Praxis subsumieren. Wie die Autor*innen beschrieben haben, können und sollten die *evidence-use* Mechanismen nicht getrennt voneinander betrachtet werden. Die Kombination der Mechanismen kann zu Synergieeffekten führen, die zur Förderung von Kompetenzen einer evidenzbasierten Praxis führen könnte. Die Autor*innen fokussieren in ihrem Review auf „*evidence informed decision-making*“ (Breckon & Dodson, 2016, S. 2). Wie durch die Darstellungen zum Evidenzbegriff (Kap. 2.1) sowie den Argumentationsprozessen (Kap. 2.2) gezeigt worden ist, beinhaltet evidenzbasierte Praxis Entscheidungen. Evidenzen als Argument für oder gegen eine These (Bromme et al., 2014) führen zwangsläufig zu Momenten der Entscheidung: Widerlegen die Evidenzen die These oder unterstützen sie diese? Damit inkludiert evidenzbasierte Praxis auch Entscheidungsfindungsprozesse. Diese Prozesse können bei Lehrkräften z.B. für die Unterrichtsplanung beschrieben werden.

Wie gezeigt werden konnte, wurde und wird das Konzept evidenzbasierter Praxis auch für den Bildungsbereich diskutiert. Ausgangspunkt dazu bildet die Konzeptualisierung von *evidence-based medicine*, wenngleich eine identische Übertragung auf den Bildungsbereich aufgrund dessen Spezifika nicht möglich erscheint. Weiterhin scheinen (fach-)didaktische Überlegungen bisher wenig Berücksichtigung in den Diskursen gefunden zu haben. Hinweise eines unterschiedlichen Verständnisses von Evidenzen in einzelnen wissenschaftlichen Disziplinen (Bellmann & Müller, 2011a) könnten auch auf Unterschiede im Verständnis evidenzbasierter Praxis oder von Evidenzen als Konstrukt hindeuten.

3. Einfluss von Epistemologie auf evidenzbasierte Praxis

Wie gezeigt werden konnte, stellt die Evidenzbasierung ein relevantes Konzept in verschiedenen Professionen dar. Vor dem Hintergrund bestehender Argumentationsmodelle, lassen sich die Funktion und Generierung von Evidenzen durch die Interpretation von Daten erklären. Dabei basieren diese Prozesse jedoch auf Überzeugungen, welche sich auch als hinderlich bei der Arbeit mit Evidenzen erweisen können (Voss, 2022). Die hier beschriebenen Überzeugungen stellen epistemologische Überzeugungen der Individuen dar. Diese nehmen z.B. bei der Unterscheidung von Evidenzen und Erfahrungen eine zentrale Rolle im Prozess der Evidenzbasierung ein. Hier soll das Konstrukt Epistemologie und dessen Einfluss auf die Verarbeitung von Daten, bzw. deren Interpretation als Evidenzen dargelegt werden.

Die Epistemologie (auch Wissenschafts-/Erkenntnistheorie) kann als ein Teilbereich der Philosophie beschrieben werden, welcher sich vor allem mit der Natur und Begründung von Wissen beschäftigt (Hofer & Pintrich, 1997).

3.1 Begriffsabgrenzung

Bevor jedoch eine tiefergreifende Beschreibung des Konstrukts möglich ist, müssen in der Literatur unterschiedlich verwendete Begriffe voneinander abgegrenzt werden. Die Epistemologie kann als Wissen, oder auch Theorie (griech.: logos) über das Wissen (griech.: Episteme) beschrieben werden (Kitchener, 2002). Damit ist die Epistemologie ein Meta-Wissen über das Wissen selbst. Wichtig ist allerdings die Unterscheidung zwischen Wissen/Episteme und Kognitionen. Kognitionen beschreiben alle Wahrnehmungen eines Individuums von der Außenwelt. Dabei unterscheiden sich Kognitionen und Episteme insofern, dass Episteme ‚wahr‘ ist und vor dem Hintergrund von Evidenzen begründet werden kann (Kitchener, 2002). Die gleiche Unterscheidung kann auch auf der Meta-Ebene getroffen werden: Während Epistemologie das Meta-Wissen über die Episteme beschreibt, beschreiben Meta-Kognitionen Überzeugungen über eine Entität, welche nicht zwangsläufig wahr und evidenzbasiert begründbar sein müssen. Die damit verbundene Unterscheidung von Meta-Wissen und Meta-Kognitionen ist zentral für diese Arbeit. Naive, epistemologische Überzeugungen von Schüler*innen können entsprechend als Meta-Kognitionen beschrieben werden. Handelt es sich hingegen um wissenschaftlich anerkannte Aspekte der Epistemologie, können diese als Meta-Wissen über die Episteme beschrieben werden.

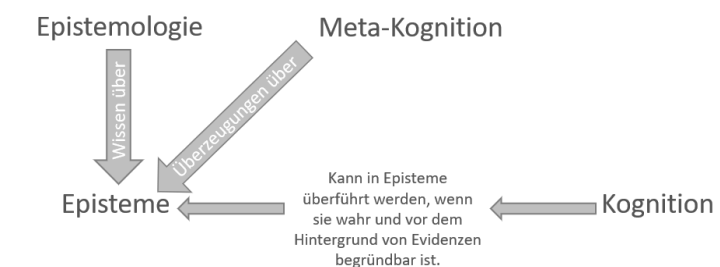


Abbildung 2 - Übersicht Begriffsklärung Epistemologie

Weiterhin wird in englischsprachigen Texten zwischen *epistemic beliefs* und *epistemological beliefs* unterschieden (z.B. Mason & Bromme, 2010; Murphy et al., 2012). Eine von Kitchener (2002) vorgeschlagene Unterscheidung beschreibt *epistemological beliefs* als das Wissen über die Epistemologie (also als Meta-Wissen über die Epistemologie), während *epistemic beliefs* das Meta-Wissen über das Wissen beschreiben. Andere Autor*innen nutzen den Begriff *epistemological beliefs* zur Beschreibung von „individual representations about knowledge and knowing“ (Mason & Bromme, 2010, S. 1).

In der vorliegenden Arbeit wird der Begriff ‚epistemologische Überzeugungen‘ als Meta-Kognitionen eines Individuums über das Wissen verstanden. Damit wird einerseits der vorläufige Charakter von individuellen Meta-Kognitionen im Vergleich zu Meta-Wissen beschrieben und andererseits auf eine mögliche Verwirrung durch die Nutzung unterschiedlicher Begriffe verzichtet. Weiterhin kann das Konstrukt der epistemologischen Überzeugungen damit auch sprachlich von *Nature of science*-Vorstellungen unterschieden werden (dazu Kapitel 3.4) (Neumann & Kremer, 2013).

3.2 Definition Epistemologie

Das Forschungsfeld zu Epistemologie kann in zwei grundsätzliche Strömungen unterschieden werden: Entwicklungspsychologische Ansätze und kognitionspsychologische Ansätze. Dabei sind vor allem in den 1990er-2010er Jahren viele Publikationen entstanden. Vor allem das umfassende Review von Hofer und Pintrich (1997) findet bis heute große Verwendung.

Entwicklungspsychologische Ansätze

Entwicklungspsychologische Ansätze fokussieren vor allem auf die Entwicklung epistemologischer Überzeugungen im Laufe der eigenen Lernbiographie und mögliche Entwicklungsstufen, welche durchlaufen werden. Die Grundlage aller entwicklungspsychologischen Ansätze bildet das *Scheme of Intellectual and Ethical Development* von William Perry aus 1970. Als einer der ersten Forschenden setzt sich Perry mit den epistemologischen Überzeugungen von Studierenden auseinander. Das daraus resultierende Schema beinhaltet neun Positionen, welche einen schrittweisen Aufbau, bzw. eine Neustrukturierung epistemologischer Überzeugungen beschreiben. Knefelkamp und Slepitz (1976) fassen diese neun Positionen in vier Kategorien zusammen:

i.	Dualism	Dualistisches, dichotomes Verständnis von richtig/falsch
ii.	Multiplicity	Anerkennung der Möglichkeiten falsche Entscheidungen zu treffen, bzw. unterschiedliche Entscheidungen zu treffen, welche individuelle Berechtigung haben
iii.	Relativism	Wahrnehmung des Selbst als aktiven Part im Prozess
iv.	Commitment with relativism	Wahrnehmung unterschiedlicher Einflussfaktoren auf die eigene Entscheidung, aktive Teilnahme am Entscheidungsprozess

Vor allem der Übergang zwischen Kategorie zwei zu drei zeigt eine größere Veränderung in den epistemologischen Überzeugungen der Studierenden auf. Während in den Kategorien eins und zwei noch von einem einfachen Verhältnis zwischen richtig und falsch ausgegangen wird, wird dieses in den weiteren Kategorien differenzierter betrachtet. Eine ‚Wahrheit‘ wird nicht mehr als ‚natürlich auffindbar‘ verstanden, sondern als Produkt eines Aushandlungsprozesses, in welchem das Individuum eine zentrale Rolle einnimmt. Die hier beschriebene Abkehr von der Überzeugung, dass Wissen richtig oder falsch sein kann, markiert einen Kernaspekt in der Erweiterung der epistemologischen Überzeugungen. Veränderungen in den Kategorien drei bis vier zeigen einen weniger deutlichen Zusammenhang zu epistemologischen Überzeugungen auf (Hofer & Pintrich,

1997). Hier wird der Fokus auch auf mögliche Einflussfaktoren für persönliche Entscheidungen gelenkt.

Perrys Arbeit musste sich einiger Kritik stellen. So wurden in seiner Studie nur zwei Frauen in die Analyse und Modellierung mit einbezogen. Auf Grundlage dieser Kritik leiteten Mary Field Belenky und Kolleg*innen eine eigene Studie ab, welche ausschließlich Frauen einbezog (Hofer & Pintrich, 1997). Das resultierende Modell (*Women's ways of knowing*) weist eine deutliche Nähe zu Perrys Schema auf. Als Abweichung wird die Kategorie *silence* vorgestellt, welche auf eine sehr passive, untergeordnete Rolle des Selbst abhebt. Die weiteren Kategorien sind vergleichbar mit Perrys Schema, sodass keine nennenswerten Gender-Effekte bei der Untersuchung auftraten. Das methodische Vorgehen und vor allem die Konzentration auf eine ausschließlich weibliche Stichprobe wurden teils wiederum deutlich kritisiert (Hofer & Pintrich, 1997). Marcia Baxter Magolda führte schließlich eine Langzeitstudie (5 Jahre, $N=101$) mit Frauen und Männern durch. Auch in dieser Studie konnten keine Gender-Effekte beobachtet werden, jedoch zumindest unterschiedliche Begründungsmuster der epistemologischen Überzeugungen erkannt werden (Hofer & Pintrich, 1997). Diese Begründungsmuster sind jedoch nicht spezifisch für eines der Geschlechter. Mit ihrer Studie und dem daraus entwickelten *Epistemological Reflection Mode* schließt Baxter Magolda einerseits die Forschungslücke bzgl. unterschiedlicher epistemologischer Überzeugungen bei Männern und Frauen, andererseits lässt sich ihr entwickeltes Modell als anschlussfähig bzgl. der vorherigen Konzeptualisierungen beschreiben.

Einen völlig anderen Schwerpunkt setzten King und Kitchener mit ihrem *Reflective Judgement Model*. Hier werden die Proband*innen (im Alter zwischen High-School- und Erwachsenenalter) mit komplexen Problemsituationen konfrontiert, auf die keine einfachen Antworten gefunden werden können. Zur Bearbeitung einer solchen Problemsituation müssen die Proband*innen sich mit ihren epistemologischen Überzeugungen auseinandersetzen und vor allem Unsicherheiten akzeptieren (King & Kitchener, 2002). Das Modell sieht sieben unterschiedliche Ebenen vor, welche wiederum in drei Kategorien zusammengefasst werden können. Dabei ist es möglich, dass sich Proband*innen in mehreren Kategorien, oder auf mehreren Ebenen gleichzeitig befinden. Dies sei vor allem abhängig vom Kontext und der Situation. Weiterhin erkennen die Autor*innen eine funktionale und eine optimale Ebene der epistemologischen Überzeugungen. Während die funktionale Ebene vor allem in akut-herausfordernden Situationen (wie im durch die Autor*innen durchgeführten Interview) aktiviert wird, kann mehr Zeit und Ruhe bei der Beschäftigung mit einem Problem zu einer Verarbeitung auf optimaler Ebene führen (Hofer & Pintrich, 1997). King und Kitchener (2002) heben in ihrer Arbeit explizit den meta-kognitiven Charakter der epistemologischen Überzeugungen hervor. Es geht explizit um die Entwicklung der eigenen Überzeugungen hin zu anerkanntem epistemologischem Wissen.

Zusammenfassend fokussieren die hier vorgestellten Modelle auf ähnliche Entwicklungsmuster epistemologischer Vorstellungen. Hofer und Pintrich (1997) erkennen drei grundlegende Entwicklungsstufen epistemologischer Überzeugungen in allen Modellen:

- i. Wissen wird dichotom als richtig/falsch verstanden
- ii. Wissen wird als durch Unsicherheiten geprägtes Konstrukt wahrgenommen
- iii. Das Selbst als aktiver Gestalter von Bedeutung in relativistischen Kontexten

Die Ausprägung der unterschiedlichen Entwicklungsstufen variiert in den unterschiedlichen Modellen. Auch die mögliche Integration von Unterstufen und die Übergänge zwischen den Stufen

werden unterschiedlich modelliert. Als zentrale Erweiterung der epistemologischen Überzeugungen treten in allen dargestellten Modellen die Unsicherheit von Wissen, wie auch die Abkehr von einem dichotomen richtig-falsch-Verständnis von Wissen auf. Vor allem vor dem Hintergrund schulischen Unterrichts, könnten diese Überzeugungen besonders relevant sein. In einem instruktivistischen Verständnis von Unterricht könnte die Lehrkraft als „Meister“ angesehen werden, welcher epistemologische Autorität im Klassenraum besitzt (Crawford, 2008). Durch die Bestätigung, oder auch Korrektur der Lehrkraft könnte ein dichotomes Verständnis der Richtigkeit von Wissen im Unterricht gefestigt werden.

Bezüglich der Vergleichbarkeit der unterschiedlichen Modelle muss einschränkend erwähnt werden, dass die Definitionen von Epistemologie nicht deckungsgleich sind. Perry begründet sein Verständnis von Epistemologie auf den Bereichen *Nature of knowledge and truth*. Belenky hingegen fokussiert auf *Source of knowledge* (Hofer & Pintrich, 1997), Baxter Magolda begründet ihr Verständnis von Epistemologie eher im Bereich *Nature of learning*. King und Kitchener (2002) legen in ihrem *Reflective Judgement Model* einen deutlich abweichenden Fokus auf die Lösung von komplexen Problemen. Diese unterschiedlichen Schwerpunktsetzungen müssen – trotz ihrer inhaltlichen Nähe zueinander – bei der Interpretation berücksichtigt werden. Die Modellierung der unterschiedlichen Entwicklungsstufen zeigt jedoch eine deutliche Nähe auf. D. Kuhn und Weinstock (2002) zeigen in ihrer Arbeit vor allem die Zusammenhänge und Problemstellen der verschiedenen Stufenmodelle auf. So stellt sich vor allem die Frage, was die einzelnen Ebenen essentiell ausmacht. Welche epistemologischen Überzeugungen werden auf spezifischen Ebenen, welche über das gesamte Modell hinweg erweitert? Mit ihrer Frage: „What is developing?“ deuten D. Kuhn und Weinstock (2002, S. 122) auf eine sich ergebende Frage hin: Was beinhaltet Epistemologie? Während die entwicklungspsychologischen Ansätze elaborierte Stufenmodelle zum Aufbau adäquater epistemologischer Überzeugungen entwickelt hat, fehlt eine einheitliche Definition, was Epistemologie umfasst.

Kognitionspsychologische Ansätze

Mit der Frage nach der Struktur epistemologischer Überzeugungen setzen sich vor allem kognitionspsychologische Ansätze auseinander (Neumann & Kremer, 2013).

Motiviert durch die Vielzahl an Publikationen im Bereich Epistemologie zwischen den später 1960er und den frühen 1980er Jahren sowie die zunehmende gegenseitige Unkenntnis der unterschiedlichen Forschungsrichtungen, erstellte Marlene Schommer ein System epistemologischer Überzeugungen (Schommer, 1990; Schommer-Aikins, 2002). Schommer et al. (1992) konnten letztlich vier verschiedene Bestandteile epistemologischer Überzeugungen in einer Studie aufzeigen: (a) *Structure of knowledge*, (b) *Stability of knowledge*, (c) *Speed of learning* und (d) *Control of learning* (Schommer-Aikins, 2002, S. 105). Den Einbezug des Lernens in die Epistemologie rechtfertigt Schommer-Aikins (2002) durch einen möglichen entwicklungspsychologischen Zusammenhang. So führe ein „developmental trend“ (Schommer-Aikins, 2002, S. 109) von Überzeugungen über das Lernen zu Überzeugungen über das Wissen insgesamt. Für eine genauere Aufklärung dieses möglichen Zusammenhangs seien jedoch weitere Studien nötig. Es zeige sich aber immer wieder, wie deutlich epistemologische Überzeugungen mit Überzeugungen über das Lernen zusammenhängen. Hofer und Pintrich (1997) kritisieren, dass Überzeugungen bzgl. des Lernens nicht im Kern epistemologischer Natur sind. Sie unterscheiden deutlich zwischen Überzeugungen über das Wissen und Überzeugungen über den Wissenserwerb. Zwar gäbe es eine Nähe der Konstrukte, diese bestünde aber auch zu anderen Aspekten (z.B. Überzeugungen über Intelligenz & Talent, etc.).

Insgesamt versteht Schommer-Aikins (2002) epistemologische Überzeugungen als ein System aus verschiedenen Überzeugungen, welche sich jedoch nicht synchron entwickeln müssen. Die Entwicklung der verschiedenen Überzeugungen sei nicht durch ein Kontinuum, oder ein dichotomes Verständnis abzubilden. Schommer-Aikins plädiert für ein Verständnis von epistemologischen Überzeugungen als Häufigkeitsverteilung. So könne es zum Beispiel bei Überzeugungen zur Stabilität von Wissen möglich sein einige Wissensaspekte als feststehend zu verstehen, während in anderen Bereichen die Vorläufigkeit von Wissen akzeptiert wird. Ein möglicher Einflussfaktor hierauf könnten auch die persönlichen Erfahrungen sein. Diesen spricht Schommer-Aikins (2002) eine wichtige Bedeutung zu, was wiederum auf den „developmental trend“ (Schommer-Aikins, 2002, S. 109) in Bezug auf Überzeugungen über Lernen und Wissen eingeht. Erfahrungen, welche in Lernprozessen gesammelt werden, könnten so zu epistemologischen Überzeugungen transformiert werden. Dabei müssen diese Erfahrungen jedoch nicht immer mit hochwertigen epistemologischen Überzeugungen einhergehen (vgl. epistemologische Autorität von Lehrkräften im Unterricht).

Ein zentrales Konzept im Verständnis epistemologischer Überzeugungen von Schommer-Aikins (2002) ist die Balance. Gemeint ist vor allem eine Balance bei der Anzahl, der Generalisierbarkeit und Unabhängigkeit epistemologischer Überzeugungen. Vor dem Hintergrund einer Weiterentwicklung epistemologischer Überzeugungen könnten sich Grundannahmen verändern. So könnten zum Beispiel weitere epistemologische Überzeugungen im Verlauf der eigenen Lernendenbiographie hinzutreten, oder wegfallen. Weiterhin könnten, vor allem im Kindesalter, generalisierbare Überzeugungen entwickelt werden, welche sich später auf einzelne Bereiche übertragen lassen, in diesen jedoch auch spezifisch ausgebildet werden können (Domänenspezifität epistemologischer Überzeugungen). Somit könnten generalisierende epistemologische Überzeugungen bei der Erschließung neuer Inhaltsbereiche hilfreich sein, sich aber später davon ausgehend domänenspezifisch weiterentwickeln.

Diese Balance ist auch im Bereich einzelner epistemologischer Überzeugungen interpretierbar. So muss zum Beispiel in Bezug auf die Stabilität von Wissen in spezifischen Lernszenarien ein gewisser Anteil konzeptuellen Wissens als gesichert angenommen werden, damit ausgehend von diesem Wissen weitergedacht werden kann. Zwar widerspricht diese Annahme eigentlich der Unsicherheit/Vorläufigkeit des Wissens, jedoch braucht es eine gewisse Balance, um weiteres Wissen erschließen zu können: „I am not saying the solid core dominates the learners' thinking. Rather, it may be a small core that allows one to build on and around that one entertains new ideas and modifies old ideas.“ (Schommer-Aikins, 2002, S. 114). Hier wird erneut deutlich, warum epistemologische Überzeugungen nach Schommer-Aikins nicht als Kontinuum, oder dichotomes Verhältnis verstanden werden dürfen. Auch Lernende mit adäquateren epistemologischen Überzeugungen können derlei Kernannahmen halten, ohne dass sich ihre epistemologischen Überzeugungen dadurch als naiv bezeichnen lassen würden. Das Konzept der Balance wird von Schommer-Aikins seit den 2000er Jahren immer wieder hervorgehoben, „but not at the price of unbridled relativism“ (Schommer-Aikins, 2002, S. 117). Es geht also nicht darum jedwede Sicherheit in der Konzeption eines Systems epistemologischer Überzeugungen zu verneinen, aber eine Offenheit für den vorläufigen Charakter eines solchen Systems zu erhalten. Dies schließt an das Konzept des lebenslangen Lernens im Bereich der Epistemologie an, welches ebenfalls durch Schommer-Aikins (2002) formuliert wurde. Hieraus resultiert auch die Frage, ob eine Ziel-Definition epistemologischer Überzeugungen endgültig geklärt werden kann (vgl. Unsicherheit/Vorläufigkeit von Wissen). Die Konstruktion eines Systems epistemologischer Überzeugungen muss also stets auch vor dem Hintergrund der eigenen epistemologischen Überzeugungen reflektiert werden. Damit hebt

Schommer-Aikins auf eine weitere Meta-Ebene ab, welche Überzeugungen über die Epistemologie beinhaltet.

Einen aus Schommer-Aikins' (2002) Perspektive deutlich philosophischer geprägte Konzeptualisierung epistemologischer Überzeugungen stammt von Hofer und Pintrich (1997). In ihrem umfassenden Review zeigen Hofer und Pintrich (1997) neun Herausforderungen im Bereich epistemologischer Überzeugungen auf (siehe Tab. 3).

Tabelle 3 – Herausforderungen im Bereich epistemologischer Überzeugungen nach Hofer und Pintrich (1997)

Herausforderungen im Bereich epistemologischer Überzeugungen
Umfassende Definition von Epistemologie
Elemente/Dimensionen epistemologischer Überzeugungen
Unklare Entwicklung epistemologischer Überzeugungen (vgl. Entwicklungsmodelle)
Unklare Verbindung zu weiteren Aspekten kognitiver Entwicklung (z.B. Alter, Bildungsbiographie)
Unklarheit über Prozesse der Konstruktion und Veränderung epistemologischer Überzeugungen
Unklarheit bzgl. Domänenspezifität epistemologischer Überzeugungen
Unklare Verbindung zu individueller Motivation, Lernen, etc.
Unklare Rolle von Gender, Ethnie, etc.
Methodologischer Herausforderungen

Im Bereich der Definition des Konstrukts Epistemologie weisen die Autor*innen auf unterschiedliche Grenzsetzungen des Konzepts hin. Wie oben beschrieben, inkludiert Schommer-Aikins (2002) Überzeugungen über das Lernen in der Epistemologie. Aus dem Vergleich der oben aufgeführten Modellierungen fokussieren Hofer und Pintrich (1997, S. 112) auf die zwei Bereiche „Nature of knowledge“ und „Nature or process of knowing“. Dennoch anerkennen die Autor*innen eine Nähe und Verzahnung mit anderen Konzepten wie Intelligenz, Lernen, Unterrichten, etc. Gleiches gilt für generelle Denk- und Begründungsprozesse. Epistemologische Überzeugungen seien eher als übergeordnete Struktur auf einer höheren (Meta-)Ebene zu verstehen, welche auch generelle Denk- und Begründungsprozesse beeinflussen. Hofer und Pintrich (1997) beschreiben epistemologische Überzeugungen als theorie-ähnliche Konstrukte. Demnach besitzen epistemologische Überzeugungen eine gewissen Kohärenz zwischen ihren Bestandteilen, welche dennoch ontologisch voneinander abgegrenzt werden können. Gopnik und Wellman (1992) beschrieben Theorien weiterhin als abstrakte Erklärungsnetze, welche in weiter entwickelten Formen auch widersprüchliche Informationen adressieren können. Inwiefern auch epistemologische Überzeugungen als Erklärungsnetze nutzbar sind, ist für Hofer und Pintrich (1997) offen. Die Betrachtung epistemologischer Überzeugungen als persönliche Theorie stellt den möglichen Zusammenhang mit weiteren Bereichen (z.B. Lernen, Intelligenz, etc.) dar. Epistemologische Überzeugungen können als Theorie Einfluss auf diese Bereiche nehmen, oder durch diese beeinflusst werden. Annahmen über das Lernen müssen sich in die epistemologischen Überzeugungen einbetten lassen. Andernfalls bedarf es einer Theorieanpassung. Hier wird die gegenseitige Einflussnahme der verschiedenen Konstrukte deutlich.

Die Betrachtung epistemologischer Überzeugungen als persönliche Theorien erlaubt gleichzeitig die weitere Ausdifferenzierung des Konzepts. Hofer und Pintrich (1997) extrahieren in ihrem Review je zwei Dimensionen für die beiden Bereiche, welche in Abbildung 3 dargestellt sind.

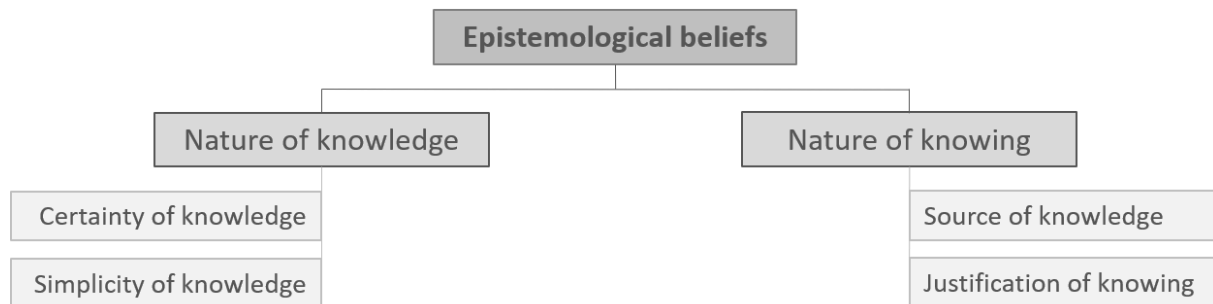


Abbildung 3 - Dimensionen epistemologischer Überzeugungen nach Hofer und Pintrich (1997, S. 119)

Im Bereich *Nature of knowledge* fokussieren die Autor*innen auf das Verständnis dessen, was Wissen ist. Die Dimension „Certainty of knowledge“ (Hofer & Pintrich, 1997, S. 120) stellt das Spannungsfeld zwischen einem Verständnis von Wissen als sicher/beständig und unsicher/flexibel in den Fokus. Im Sinne der vorgestellten entwicklungspsychologischen Modelle epistemologischer Überzeugungen kann die Entwicklung von einem starren zu einem fluiden Wissensverständnis erfolgen. Dabei stellt die Offenheit für Neuinterpretationen und Anpassungen der Theorie eine mögliche Zieldimension dar (vgl. Modell von King & Kitchener). Die Dimension „Simplicity of knowledge“ (Hofer & Pintrich, 1997, S. 120) fokussiert auf die Komplexität von Wissen. Dabei wird Wissen als Kontinuum zwischen einem Verständnis von Wissen als Akkumulation einzelner Fakten und als hochgradig kontextabhängiges, relatives Konzept verstanden. In Rückbezug auf Schommer-Aikins (2002) spielt dabei die Balance des Verständnisses eine wichtige Rolle. Wenngleich eine adäquate Zielvorstellung von Wissen durch Unsicherheit und kontextuelle Einflüsse beschrieben wird, kann das Anerkennen bestimmter, faktischer Wissensinhalte für die Bearbeitung einer Fragestellung entscheidend sein. Dabei ist vor allem das eigene Bewusstsein für das unterschiedliche Verständnis von Wissen wichtig.

Im Bereich *Nature of knowing* lassen Hofer und Pintrich (1997) eine Nähe zum Konzept des Lernens erkennen, da sie hier auch auf prozedurale Aspekte der Epistemologie eingehen. Die erste postulierte Dimension in diesem Bereich ist „Source of knowledge“ (Hofer & Pintrich, 1997, S. 120). Die Quelle von Wissen wurde in allen oben dargestellten Modellen epistemologischer Überzeugungen aufgegriffen und kann auch in der Definition des Konstruktes „Evidenz“ wiedergefunden werden. Innerhalb dieser Dimension wird die Quelle von Wissen kritisch reflektiert. Die Veränderung der Rolle des Selbst im Prozess der Wissensgenerierung wird als entscheidender Entwicklungsschritt modelliert. Werden zu Beginn Autoritäten als Quelle von Wissen verstanden, erhält das Selbst nach und nach größeren Einfluss im Prozess. Adäquatere Überzeugungen (im Sinne entwicklungspsychologischer Konzepte) stellen das Selbst als aktiven, gestaltenden Part im Wissensgenerierungsprozess in den Mittelpunkt. In der Zusammenarbeit mit anderen wird Wissen erschlossen. Autoritäten spielen in einem erweiterten Verständnis eine sehr untergeordnete Rolle. Grundlage der Wissensgenerierung bieten Evidenzen, welche in der Dimension „Justification of knowledge“ (Hofer & Pintrich, 1997, S. 120) kritisch betrachtet werden. Verschiedene Schlussfolgerungen werden auf ihrer evidenzbasierten Begründung basierend akzeptiert. Es bedarf also der Begründung von Schlussfolgerungen wie diese z.B. im *Model of argument* nach Toulmin (2003) modelliert ist. Auch die Rolle von Autoritäten und Experten wird hier bewertet, jedoch mit weiteren Evidenzen, eigenen Fragestellungen und Überzeugungen abgeglichen. Insofern

spielen Autoritäten in adäquateren epistemologischen Überzeugungen weiterhin eine Rolle. Jedoch werden diese im Vergleich zu naiven Überzeugungen nun bzgl. ihrer Glaubwürdigkeit bewertet und als eine weitere Quelle von Evidenzen verstanden. Der individuelle Aushandlungsprozess verschiedener Evidenzen führt dann zu Wissen. Hier sind viele Parallelen zum Evidenzbegriff und dem *Model of argument* (Toulmin, 2003) zu erkennen. Es wird ersichtlich, dass jegliche Information als *data* akzeptiert werden kann. Um jedoch als Evidenz eine Schlussfolgerung unterstützen zu können, müssen weitere Bedingungen (vgl. *warrant*, *backing*, *rebuttal* und *qualifier*) erfüllt sein. Dabei spielen für die Bewertung der einzelnen Bedingungen die epistemologischen Überzeugungen eine entscheidende Rolle. So kann ein *warrant* z.B. aufgrund seiner Herkunft von einer Autorität in Zweifel gezogen werden, da Überzeugungen vorliegen, welche Autoritäten als einziges Argument für die Glaubwürdigkeit von Aussagen ablehnen. Ebenso könnten einfache Lösungen auf komplexe Probleme aufgrund von Überzeugungen zum komplexen, vernetzten Wissen in Zweifel gezogen werden. Es zeigt sich also, dass insbesondere die Begründungsschritte innerhalb des *Model of argument* und zwischen den konstituierenden Bedingungen entscheidend von den epistemologischen Überzeugungen abhängen können.

Die Unterscheidung von ‚adäquaten‘ und ‚inadäquaten‘ epistemologischen Überzeugungen scheint auf Grundlage der obigen Dimensionen gut möglich und anschlussfähig an entwicklungspsychologische Modelle zu sein. Dabei stellen die Dimensionen vier wichtige Merkmale ‚adäquater‘ Überzeugungen dar, ohne dass diese normativ als Ideal-Überzeugungen gesetzt werden könnten. Die Formulierung von Ideal-Überzeugungen bei gleichzeitiger Berücksichtigung wichtiger epistemologischer Überzeugungen (z.B. Unsicherheit/Vorläufigkeit von Wissen) ist kritisch zu betrachten (Hofer & Pintrich, 1997; Schommer-Aikins, 2002). Die Beschreibung verschiedener Entwicklungsstufen basiert auf der Annahme, dass „höherwertige“ epistemologische Überzeugungen in jeder Situation anwendbar sind. Es gäbe somit einen Endpunkt jeder Entwicklung von Überzeugungen. Dies würde aber die hochgradige Kontextabhängigkeit von Wissen völlig vernachlässigen. Hofer und Pintrich (1997) mahnen ein solch einfaches Verständnis an. Schon Theorien von Piaget seien für die Annahme eines Endpunktes von Entwicklung (mit einem Idealtyp basierend auf westlich-kulturellen Aspekten) kritisiert worden. Epistemologische Überzeugungen sind somit vor dem Hintergrund des relativistischen Kontexts zu betrachten und zu bewerten. Damit scheinen entwicklungspsychologische Modelle, die den Aufbau epistemologischer Überzeugungen über vertikale Ebenen beschreiben, dem komplexen Konstrukt epistemologischer Überzeugungen nicht gerecht zu werden. Hofer und Pintrich (1997) folgen dem Verständnis von Schommer-Aikins, welche epistemologische Überzeugungen nicht durch vertikale Ebenenübergänge charakterisiert, sondern als kontextabhängige Häufigkeitsverteilung. In verschiedenen Situationen können somit verschiedene epistemologische Überzeugungen aktiviert werden, die nur auf Grundlage der jeweiligen Situation bzgl. ihrer Adäquatheit bewertet werden können. Damit scheinen postulierte Entwicklungsstufen nicht zur Beschreibung epistemologischer Überzeugungen geeignet zu sein. Dies begründet sich dabei weniger in der ‚Adäquatheit‘ der Überzeugungen, sondern eher in der Balance zwischen adäquaten und weniger adäquaten Überzeugungen, welche je nach Kontexte dominieren können. Somit könnten epistemologische Überzeugungen, welche Wissen als vorläufig, komplex und in seiner Begründung evidenzbasiert und ungeachtet potentieller Autoritäten beschreiben als sehr hochwertig beschrieben werden.

Auf Grundlage der Kontextabhängigkeit epistemologischer Überzeugungen stellt sich die Frage nach Einflussfaktoren, welche eher adäquate Überzeugungen aktivieren könnten. Inwieweit die Entwicklung epistemologischer Überzeugungen mit dem Alter, oder auch der individuellen Bildungsbiographie zusammenhängt, wurde mehrfach untersucht. Hofer und Pintrich (1997) leiten aus bestehender Literatur ab, dass es einen geringen Zusammenhang zwischen der Entwicklung

kognitiver Fähigkeiten und epistemologischer Überzeugungen geben könnte. So könnten gewisse Überzeugungen nur bei entsprechender kognitiver Entwicklung möglich werden. Komplexe Begründungsfähigkeiten könnten eine notwendige Voraussetzung für relativistische epistemologische Überzeugungen sein. Hofer und Pintrich (1997) sprechen daher von einer deutlich positiven Beziehung zwischen Alter, Bildungsstand und epistemologischer Entwicklung. Forschung im Bereich epistemologischer Überzeugungen von Kindern und Jugendlichen zeigt auf, dass auch im jungen Alter epistemologische Überzeugungen vorhanden sind, die wiederum Einfluss auf bestimmte Handlungen nehmen können (Elder, 2002; D. Kuhn et al., 2000). Dabei können vornehmlich naive Überzeugungen gefunden werden, welche sich vor allem im Bereich *Source of knowledge* und *Certainty of knowledge* mit der Zeit hin zu adäquateren Überzeugungen entwickeln können (Conley et al., 2004). Studienergebnisse zeigen weiterhin erste Indizien für einen Einfluss der Instruktionsart auf (Conley et al., 2004; Hofer, 2004).

Die durch Hofer und Pintrich (1997) beschriebene Forschungslücke bzgl. epistemologischer Überzeugungen von jungen Schüler*innen wurde bis heute weiterbearbeitet. Studien in der Primar- und Sekundarstufe zeigen relevante epistemologische Überzeugungen auch für jüngere Schüler*innen auf (Belland et al., 2019; Kizilgunes et al., 2009; F. Lin, 2018; Schommer-Aikins et al., 2005). So konnte F. Lin (2018) für Proband*innen an einer Grundschule in Hong-Kong eine fehlende Unterscheidung zwischen Evidenzen und Theorien und lediglich rudimentäre Vorstellung der Veränderbarkeit von Theorien aufzeigen. Belland et al. (2019) konnten Indizien für einen positiven Einfluss problemorientierter Arbeitsweisen auf die Förderung epistemologischer Überzeugungen für Middle-School Schüler*innen darlegen. Auch der Einfluss affektiver Variablen wie Motivation konnten nachgewiesen werden (Kizilgunes et al., 2009). Berücksichtigt werden muss jedoch, dass in einigen Studien keine klare Trennung zwischen den Konstrukten Epistemologie und *Nature of science* (z.B. F. Lin, 2018), bzw. weiteren verwandten Konstrukten geschieht (vgl. Kap. 3.4). Ebenso erfolgen teils Untersuchungen der Wirksamkeit bestimmter Interventionen, ohne eine umfassende Definition des Konstrukts epistemologischer Überzeugungen darzulegen, bzw. eine Verortung im Diskurs um diese.

Nicht nur Einflussvariablen auf die Entwicklung epistemologischer Überzeugungen sollten genau betrachtet werden. Auch der eigentliche Aufbau oder auch die Veränderung epistemologischer Überzeugungen ist auf der Prozessebene bisher unklar. Aus Perspektive der Arbeiten von Piaget könnten Assimilation und Akkommodation eine wichtige Rolle spielen. Dabei meint Assimilation die Interpretation neuer Kognitionen vor dem Hintergrund bestehender Schemata. Ist eine Interpretation durch bestehende Schemata nicht möglich, kann durch Akkommodation das bestehende Schema erweitert, oder auch ein völlig neues Schema erworben werden (Hannover et al., 2014). Hofer und Pintrich (1997) beschreiben auch die Rolle eines Ungleichgewichts, bzw. einer kognitiven Dissonanz, wie diese im Bereich der *Conceptual Change* Forschung thematisiert wird (näheres dazu in Kap. 3.2). So könnten z.B. anomale Beobachtungen zu gefestigten Konzepten wie dem Siedepunkt von Wasser epistemologische Überzeugungen zur Stabilität und Unveränderbarkeit von Wissen in Frage stellen. Ebenso könnte der Aufbau epistemologischer Überzeugungen auch durch einen Enkulturationsprozess der jeweiligen Disziplin verstanden werden. So könnten epistemologische Überzeugungen im Sinne von *cognitive apprenticeship* erworben, also durch Lehrkräfte an Schüler*innen weitergegeben werden.

Einen Einfluss auf den Erwerb epistemologischer Überzeugungen könnten auch unterschiedliche wissenschaftliche Disziplinen oder Domänen darstellen (Hofer & Pintrich, 1997). Die Betrachtung derartiger Effekte wird jedoch durch Unschärfen in der Definition und Grenzsetzung von Domänen/Disziplinen erschwert. Auch werden die Begriffe Domäne und Disziplin in der bisherigen

Forschung teils synonym verwendet, bzw. nicht deutlich voneinander unterschieden. Aus einer theoretischen Perspektive sehen Hofer und Pintrich (1997) durchaus eine Begründung für die Unterscheidung domänenspezifischer und genereller epistemologischer Überzeugungen. Das Verständnis derartiger Überzeugungen als theorieähnlich erlaubt ein Netz epistemologischer Überzeugungen welches einerseits generalisierend verstanden werden könnte, sich aber in spezifischen Kontexten unterschiedlich ausdrücke. Bestehende Studien zu dieser Fragestellung stellen zumeist deutlich unterschiedliche Domänen dar (z.B. Urhahne, 2006; Schommer-Aikins et al., 2003; Urhahne & Kremer, 2023; Buehl & Alexander, 2005). In einer aktuellen Untersuchung von Urhahne und Kremer (2023) wurden bspw. epistemologische Überzeugungen innerhalb verschiedener Disziplinen miteinander verglichen. Dabei fand die Unterscheidung der Disziplinen anhand des „hard-soft“-Kontinuums nach Biglan (1973) statt, sodass die Disziplinen *Natural sciences*, *Mathematical sciences*, *Social sciences* und *Linguistic sciences* unterschieden wurden (Urhahne & Kremer, 2023). Biglan (1973) argumentiert diese Unterscheidung durch geteilte Paradigmen im Sinne T. S. Kuhns (1996) der einzelnen Disziplinen. Signifikante Mittelwertsunterschiede zwischen den einzelnen Disziplinen weisen hier auf unterschiedliche Überzeugungen der Schüler*innen in unterschiedlichen Disziplinen hin (Urhahne & Kremer, 2023). Gleichzeitig erkennen die Autor*innen Gemeinsamkeiten zwischen eher empirisch orientierten Disziplinen (*Natural sciences* & *Social sciences*) und stark reglementierten Disziplinen (*Mathematical sciences* & *Linguistic sciences*). Während Wissen in den erstgenannten Disziplinen als weniger statisch wahrgenommen wird, ist für die letztgenannten Disziplinen ein Fokus auf die Sicherheit und die Rolle von Autoritäten bei der Generierung von Wissen vorherrschend. Urhahne und Kremer (2023) argumentieren entsprechend eine Unterscheidung zwischen empirischen und theoretisch/hermeneutischen Disziplinen in das Konzept von Biglan (1973) mit aufzunehmen.

Muis et al. (2006) haben mit der *Theory of integrated Domains in Epistemology* (TIDE) die Beziehung zwischen generellen und spezifischen epistemologischen Überzeugungen modelliert (vgl. Abb. 4). Dabei gehen die Autor*innen davon aus, dass generelle epistemologische Überzeugungen bei Kindern ab der Geburt aufgebaut werden, während akademische epistemologische Überzeugungen mit dem Eintritt in Bildungsprozesse entwickelt werden. Akademische epistemologische Überzeugungen stellen dabei auch einen Übergang zu domänenspezifischen Überzeugungen dar. Diese Art von Überzeugungen werden im Rahmen instruktionaler Kontexte z.B. in der Schule aufgebaut. Die domänenspezifischen, wie auch akademischen Überzeugungen sind dabei in einen akademischen Kontext eingebettet. Hier heben die Autor*innen erneut auf den Aufbau derlei Überzeugungen durch Bildungsprozesse ab. Insgesamt ist der Aufbau epistemologischer Überzeugungen in den sozio-kulturellen Kontext eingebunden. Durch die Einbettung in verschiedene Kontexte ist eine Nähe zum Verständnis nach Schommer-Aikins (2002) erkennbar.

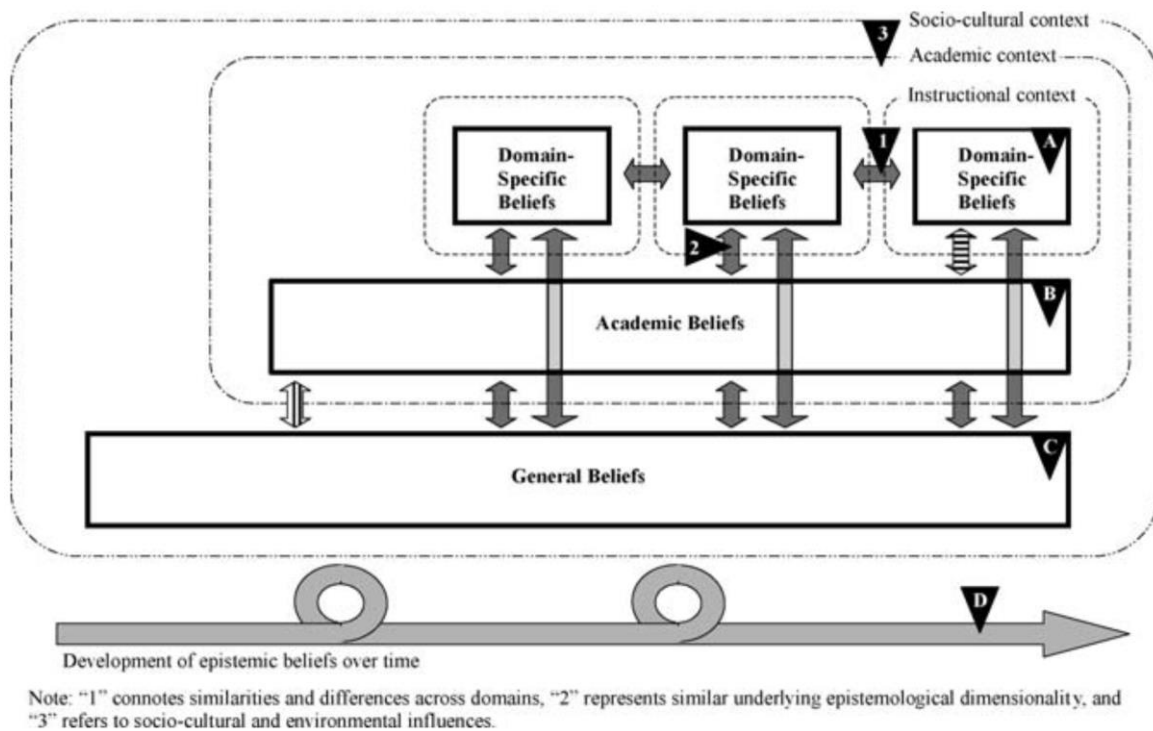


Abbildung 4 - *Theory of integrated Domains in Epistemology* (TIDE) nach Muis et al. (2006, S. 30)

In Bezug auf die Begriffsunterscheidung zwischen Disziplin und Domäne werden disziplinübergreifende Überzeugungen (z.B. über die Disziplin „Naturwissenschaften“) und domänenspezifischen (verstanden als Inhaltsbereiche einer Disziplin) Überzeugungen (Hofer & Pintrich, 1997) im Rahmen dieser Arbeit unterschieden. Je nach der Größe ontologisch voneinander unterschiedener Domänen könnten sich hier Unterschiede ergeben. Konkret könnte sich diese in übergreifenden Überzeugungen im Bereich Naturwissenschaften, aber auch im Bereich Chemie äußern. So könnten übergeordnete Überzeugungen im Bereich Chemie für unterschiedliche Fachgebiete tragfähig sein (z.B. Anorganische Chemie und Organische Chemie). Innerhalb der einzelnen Fachgebiete könnten dann wiederum spezifische epistemologische Überzeugungen auftreten. Es bleibt also die Frage nach der Auflösung, mit welcher die Überzeugungen unterschieden werden sollen. Die unterschiedliche Auflösung unterstützt ein Verständnis von generalisierbaren und spezifischen Überzeugungen. Das Verhältnis dieser untereinander ist nicht zuletzt durch die herausfordernde Grenzsetzung nicht umfassend geklärt. Festgehalten werden kann jedoch, dass epistemologische Überzeugungen sich zumindest in unterschiedlichen Disziplinen unterscheiden (Hofer, 2006; Muis et al., 2006; Urhahne & Kremer, 2023). Untersuchungen für die feingliedrige Unterscheidung zwischen einzelnen Domänen stehen größtenteils noch aus.

Eine weitere von Hofer und Pintrich (1997) beschriebene Herausforderung ist die Klärung der Beziehung von epistemologischen Überzeugungen gegenüber Motivation sowie kognitive Prozesse. Studienergebnisse legen einen Einfluss epistemologischer Überzeugungen auf die beiden genannten Konstrukte nahe (z.B. Buehl & Alexander, 2005; Chen & Pajares, 2010; Kizilgunes et al., 2009; Urhahne, 2006; Urhahne & Hopf, 2004). So können naive epistemologische Überzeugungen Einfluss auf die Wahl von Lernstrategien (Urhahne, 2006; Urhahne & Hopf, 2004), die Motivation (Buehl & Alexander, 2005; Chen & Pajares, 2010; Urhahne & Hopf, 2004), das Selbstkonzept (Urhahne & Hopf, 2004), oder Selbstwirksamkeit (Kizilgunes et al., 2009) nehmen.

Auch die Rolle des Genders wurde im Zusammenhang mit epistemologischen Überzeugungen diskutiert. Neben Studien mit ausschließlich männlichen oder weiblichen Teilnehmenden wurden auch gemischte Stichproben genauer untersucht. Dabei sind die Ergebnisse jedoch uneindeutig: Während einige Studien keine Gender-Effekte aufzeigen konnten, konnten andere zumindest unterschiedliche Begründungsmuster aufdecken. So konnte Baxter Magolda (1990) in einer frühen Studie und mit einem binären Gender-Verständnis aufzeigen, dass Frauen eher geneigt waren auf Aussagen von Autoritäten oder Peers zu vertrauen, während Männer sich selbst aktiv in den Erkenntnisgewinnungsprozess einbrachten (*Source of knowledge*). Ebenso haben Frauen bei auftretenden Widersprüchen in dieser Studie häufiger den Gesamtkontext betrachtet als Männer, welche versucht haben, die Widersprüche durch Logik und Inhaltswissen zu lösen (vgl. mechanistische Weltanschauung). Andere Studien zeigen weniger absolutistische Überzeugungen für Mädchen als für Jungen auf (Alpaslan, 2019). Inwiefern diese Ergebnisse eine Verallgemeinerung zulassen oder die Unterschiede auf Kontexteinflüsse zurückgeführt werden können, ist fraglich. Dennoch können Gender-Einflüsse auf epistemologische Überzeugungen bis heute nicht klar dargelegt werden. Hofer und Pintrich (1997) fordern daher eine theoretische Begründung für angenommene Gender-Effekte. Gleiches gilt auch für kulturelle Einflüsse (für einen Überblick: K. Chan & Elliott, 2004). Odebiyi und Choi (2022) konnten bspw. kulturell geprägte Überzeugungen zur Rolle von Autoritäten im Prozess der Wissensgenerierung für eine nigerianische Stichprobe aufzeigen. Gleichzeitig konnten auch Ähnlichkeiten mit westlichen Kulturen in Bezug auf notwendige Anstrengungen zur Wissenserweiterung festgestellt werden (Odebiyi & Choi, 2022). Schon Hofer und Pintrich (1997) vermuteten einen kulturellen Einfluss auf epistemologische Überzeugungen. Als Beispiel nennen die Autoren den Zweck von Wissen. Wissen könnte einerseits mit dem Zweck der Theorieerstellung/-erweiterung, andererseits jedoch rein instrumentell zur Umsetzung bestimmter Praktiken eingesetzt werden (vgl. die Unterscheidung anwendungsbezogener und theorieentwickelnder Domänen bei Schommer-Aikins et al., 2003). In Anlehnung an die Frage nach einer Domänenspezifität von epistemologischen Überzeugungen müssen die kulturellen Hintergründe genauer betrachtet werden. Die unklare Grenzziehung zwischen Disziplinen, Domänen und möglichen Subdomänen wird durch den Einbezug kultureller Hintergründe weiter erschwert. So könnten Domänen auch den Charakter einer Fachkultur erhalten (Hofer & Pintrich, 1997). In diesem Zusammenhang tritt erneut die hochgradige Kontextabhängigkeit epistemologischer Überzeugungen zu Tage. Epistemologische Überzeugungen müssen immer auf Grundlage des individuellen Kontexts betrachtet werden (Muis et al., 2006; Schommer-Aikins, 2002). Jedoch wird eine Unterscheidung zwischen Disziplin, Domäne und Fachkultur auch auf der Ebene des individuellen Kontexts kaum trennscharf möglich sein.

Unklare Grenzsetzungen, definitorische Uneinheitlichkeit und die Rolle theorieähnlicher epistemologischer Überzeugungen mit Einfluss auf viele verschiedene lernrelevante Bereiche (Alvermann & Qian, 2000; Jacobson & Spiro, 1995; Mason, 2000, 2002; Neumann & Kremer, 2013; Schommer et al., 1992; Schraw et al., 1995; Sinatra & Pintrich, 2003; Vosniadou et al., 2001) erschweren auch die methodische Auseinandersetzung mit epistemologischen Überzeugungen. In der bisherigen Forschungspraxis wurden unterschiedliche Formate erprobt (für einen Überblick: S. W.-Y. Lee et al., 2021). Neben strukturierten Interviews (z.B. F. Lin, 2018; Schroeder et al., 2019) wurden auch unterschiedliche Fragebögen eingesetzt (z.B. Conley et al., 2004; Schommer, 1990). S. W.-Y. Lee et al. (2021) bemängeln vor allem fehlende Diskussionen der Validität entwickelter Fragebögen. Ebenso scheint die Vergleichbarkeit der Fragebögen mit Schwerpunkten auf unterschiedlichen Aspekten von Epistemologie oder auch verwandten Konzepten wie *Natur of science* schwer. Weiterhin formulieren S. W.-Y. Lee et al. (2021) eine Notwendigkeit für Instrumente, welche *practical epistemology* fokussieren: „The focus of practical epistemology is the performance

of students in the school science learning environment [...]“ (S. W.-Y. Lee et al., 2021, S. 899). Damit geht ein Bedürfnis nach kontextualisierten Fragebögen, bzw. Testverfahren einher (S. W.-Y. Lee et al., 2021), welcher den holistischen Charakter epistemologischer Überzeugungen und deren Auswirkung auf Lernprozesse verdeutlicht. Die methodologische Anlage von Studien geht mit einem Fokus auf verschiedene Aspekte epistemologischer Überzeugungen einher. S. W.-Y. Lee et al. (2021) erkennen in den Aspekten *Uncertainty of knowledge* und *Source of knowledge* eine Übereinstimmung innerhalb verschiedener Fragebögen. Weitere Schwerpunkte können dann z.B. im Konstrukt *Nature of science* darstellen. Diesen Anforderungen müssen Forschende sich bei der Auseinandersetzung bewusst sein. So kann eine Methodenvielfalt in verschiedenen Studien zu einem holistischen Verständnis epistemologischer Überzeugungen beitragen.

Hofer und Pintrich (1997) stellen in ihrem bis heute viel beachteten Review grundlegende Richtungen der Forschung um epistemologische Überzeugungen vor. Die vorgestellte Theorie epistemologischer Überzeugungen (vgl. Abb. 3) mit ihren zwei Bereichen und vier Dimensionen ermöglicht eine Strukturierung der Forschungsrichtungen, welchen vielen Folgearbeiten und Testinstrumenten zu Grund gelegt wurde. Durch die Genese des Modells aus relevanten Beiträgen im Forschungsfeld ist dieses anschlussfähig an bestehende Forschung. Die Betrachtung der Entwicklung epistemologischer Überzeugungen als linear und vertikal scheint vor dem Hintergrund der obigen Diskussion zu kurz zu fassen. Schommer-Aikins' (2002) Konzept der Balance im Sinne einer kontextabhängigen Häufigkeitsverteilung aktivierter epistemologischer Überzeugungen, kann an vielen Stellen rund um das Verständnis epistemologischer Überzeugungen angesetzt werden. So können verschiedene Kontexte (Muis et al., 2006), aber auch kulturelle Einflüsse (K. Chan & Elliott, 2004) durch das Konzept der Balance verstanden werden. Folglich schließen sich adäquatere epistemologische Überzeugungen, bei gleichzeitiger Aktivierung eher naiver Überzeugungen nicht einander aus. In verschiedenen (sozio-kulturellen) Kontexten, könnten verschiedene Überzeugungen aktiviert werden. Über das Verständnis von epistemologischen Überzeugungen als kontextabhängig können auch weitere Aspekte einbezogen werden. So wird im spezifischen Kontext auf eine Disziplin/Domäne, wie auch eine spezifische Situation (im Sinne eines Lehr-Lern-Szenarios, etc.) fokussiert. Durch die Spezifizierung des Kontexts können viele Einflussvariablen auf die Entwicklung epistemologischer Überzeugungen in die Betrachtung eingebunden werden. Die *Theory of integrated Domains in Epistemology* (vgl. Abb. 4) kann dabei durch die Unterscheidung verschiedener Arten epistemologischer Überzeugungen, welche wiederum unterschiedlichen Prozessen (Alltagserfahrung, Schulbildung, etc.) entspringen, den direkten Einfluss verschiedener Kontexte auf aktivierte Überzeugungen ein Stück weit erklärbar machen.

Eine essentielle Herausforderung bei der Untersuchung epistemologischer Überzeugungen bleibt die Meta-Ebene in Bezug auf die Rolle des Forschenden dieses Prozesses. Forschung bzgl. epistemologischer Überzeugungen muss vor einem Bewusstsein der eigenen epistemologischen Überzeugungen analysiert werden (Schommer-Aikins, 2002). Damit müssen epistemologische Überzeugungen auf der Meta-Ebene durch Konstituenten des Konstrukts selbst reflektiert werden (Schommer-Aikins, 2002). Hier besteht die Gefahr, in einen Relativismus abzurutschen, der jegliche Vergleichbarkeit wissenschaftlicher Studien unmöglich zu machen scheint. Das Konzept der Balance kann auch in diesem Fall angewendet werden. Bromme und Kienhues (2014) fokussieren in diesem Zusammenhang auf den wissenschaftlichen Konsens, welcher relativ abgesichertes Wissen als Produkt hat. Bei diesem Wissen „muss man praktisch den Vorbehalt der Vorläufigkeit und Revidierbarkeit nicht mehr berücksichtigen, kann sich also auf dieses Wissen verlassen, auch wenn dieser Vorbehalt prinzipiell, d.h. in einem wissenschaftslogischen Sinne, immer gültig bleibt.“ (Bromme & Kienhues, 2014, S. 59). Es braucht also die Balance zwischen konsensuell abgesichertem Wissen und vorläufigem, sich entwickelndem Wissen. Ein solches Verständnis von Wissen

wird in den entwicklungspsychologischen Modellen innerhalb der obersten Ebenen/Stufen beschrieben. Hier wird ein relativistisches Verständnis von Wissen mit dem Selbst als aktivem*r Wissensgestalter*in postuliert (Hofer & Pintrich, 1997). Tritt ein Verständnis von Wissen als gesichert ohne die Berücksichtigung von relativistischen Kontexten auf, würde dies eher in den unteren Ebenen/Stufen verortet. Das Konzept der Balance, oder auch die durch Schommer-Aikins (2002) postulierte Häufigkeitsverteilung erlaubt auf den ersten Blick naive epistemologische Vorstellungen auf höheren Ebenen entwicklungspsychologischer Modelle zu halten. Wichtig ist dabei aber ein Verständnis und ein Bewusstsein für die eigenen Überzeugungen vor dem Hintergrund der jeweiligen Situation.

Im Rahmen dieser Arbeit werden epistemologische Überzeugungen als relativistisch im Sinne einer kontextabhängigen Häufigkeitsverteilung (vgl. Schommer-Aikins, 2002) verstanden. Dieses Verständnis lässt eine Unterscheidung bzgl. der Qualität epistemologischer Überzeugungen zu, wenngleich diese nicht in Form von Stufen wie in entwicklungspsychologischen Modellen unterschieden werden kann. Im Rahmen der Häufigkeitsverteilung könnten sich jedoch Schwerpunkte ausmachen lassen, welche eine qualitative Unterscheidung von individuellen Überzeugungen möglich machen könnten. So könnte im Sinne einer Mittelwertbetrachtung untersucht werden, inwiefern adäquatere oder weniger adäquate Überzeugungen aktiviert werden. Dies erlaubt einerseits die Kategorisierung der Qualität individueller Überzeugungen, andererseits kann das Auftreten unterschiedlichster Überzeugungen erklärbar werden. Als Grundlage eines Verständnisses epistemologischer Überzeugungen wird das Modell von Hofer und Pintrich (1997) angelegt. Dieses Modell zeigt Betrachtungsaspekte für epistemologische Überzeugungen auf. Dieses Verständnis ermöglicht einerseits die Offenheit für vielfältige Einflussfaktoren auf aktivierte epistemologische Überzeugungen (siehe oben) und bietet andererseits einen Orientierungsrahmen, um nicht dem beschriebenen Relativismus zu erliegen (vgl. Schommer-Aikins, 2002). Epistemologie wird hier als ein theorieähnliches, holistisches Konzept verstanden, welches Einfluss auf viele lernrelevante Bereiche nehmen kann, von diesen aber unterscheidbar ist. Die enge Vernetzung mit anderen Konzepten (wie z.B. *Nature of science*, siehe Kap. 3.4) zeigt damit die umfassende Bedeutung der Epistemologie auf.

3.3 Conceptual Change und Epistemologie

Im Bereich der Forschung um das Konstrukt *Conceptual Change* kann auf eine lange Forschungstradition zurückgeblickt werden. Ausgehend von einer umfassenden Arbeit von Posner et al. (1982) wurde das Konstrukt mit der Zeit immer weiter überarbeitet und empirisch erschlossen. Die Forschungslinien der Epistemologie (insbesondere Vorläufigkeit von Erkenntnissen und kognitiver Konflikt) und des *Conceptual Change* überschneiden sich an einigen Stellen. So zeigen Hofer und Pintrich (1997), Bendixen (2002) oder auch Liptow (2022) eine mögliche Entwicklung epistemologischer Überzeugungen durch *Conceptual Change* Prozesse auf. In diesem Kapitel soll das Konstrukt des *Conceptual Change*, dessen Bedingungen, philosophischen Grundlagen wie auch die Verbindung zur Epistemologie dargestellt werden.

3.3.1 Conceptual Change

Die *Conceptual Change* Theorie versucht die Veränderung von Vorstellungen von Lernen zu beschreiben. Dabei gibt es trotz langer Forschungstradition bisher keinen Konsens, wie ein solcher Prozess ablaufen kann (J.-W. Lin et al., 2016). Geteilt wird das theorieverändernde/-erweiternde Verständnis von *Conceptual Change* Prozessen, die zumeist an Vorstellungen von Lernen anknüpfen. Dabei können jedoch Unterschiede im Verständnis von Präkonzepten,

Einflussfaktoren auf den Prozess wie auch möglichen Instruktionsmethoden erkannt werden (Pacaci et al., 2023).

Allem voran stellt sich die Frage, wie das Konstrukt *Vorstellungen* beschrieben werden kann. Gropengießer und Marohn (2018, S. 51) beschreiben Vorstellungen als „subjektive gedankliche Konstruktionen“. Diese Vorstellungen können naturwissenschaftlichen Konzepten gegenüberstehen. Ausgehend von (häufig) naiven oder wissenschaftlich inadäquaten Schülervorstellungen ist das Ziel, die Vorstellungen im Sinne einer wissenschaftlich anerkannten Theorie zu transformieren (für einen Überblick: Gropengießer & Marohn, 2018). Die Forschung im Bereich der *Conceptual Change* Theorie befasste sich zuerst mit diesem Transformationsprozess, also der Frage, wie Schülervorstellungen in wissenschaftlich anerkannte Theorien überführt werden können. Dabei ist der Aspekt des Transformierens, des Erweiterns oder auch der Änderung der Vorstellungen in verschiedenen Konzeptualisierungen diskutiert worden. Die unterschiedlichen Begrifflichkeiten sind (wie gezeigt werden soll) voneinander abgrenzbar, werden hier aber unter dem Begriff *Conceptual Change* subsummiert. Neuere Modellierungen des *Conceptual Change* Prozesses betrachten Lernlinien für unterschiedliche Inhaltsbereiche.

Posner et al. (1982) postulieren einen *Conceptual Change* als einen radikalen Wechsel von einer Theorie zu einer anderen. In diesem Sinne folgen sie der durch Piaget beschriebenen Veränderung in linearen, größeren Schritten. Dabei beziehen sich Posner et al. (1982) auf das Konzept der Assimilation und Akkommodation nach Piaget. Wie in Kapitel 3.2 beschrieben, meint die Assimilation die Verarbeitung neuer Kognitionen mit Hilfe von bestehenden Schemata, wohingegen die Akkommodation den Aufbau neuer Schemata erfordert. Wissenschaftsphilosophisch steht dieses Konzept vor allem in der Tradition von T. S. Kuhn's (1996) Theorie der wissenschaftlichen Revolution, welche die Vorstellungsveränderung in der Wissenschaft beschreibt. In seiner Theorie wissenschaftlicher Revolution geht Kuhn davon aus, dass die Wissenschaft durch bestimmte Paradigmen geprägt ist, welche als Erklärungsrahmen für auftretende Phänomene dienen. Bezogen auf die Fachwissenschaft Chemie könnte beispielsweise das Bohr'sche Atommodell ein Paradigma darstellen, das zur Erklärung auftretender Phänomene (hier z.B. das Auftreten von Spektrallinien) genutzt wird. Treten neue Phänomene auf, welche durch die bestehenden Paradigmen nicht mehr erklärbar sind, gerät die Wissenschaft in eine Krise. In dieser Krise müssen neue Paradigmen entwickelt werden, die die Erklärung der neuen Phänomene ermöglichen. Der Wechsel zu dem neuen Paradigma wird durch T. S. Kuhn (1996) als wissenschaftliche Revolution bezeichnet und könnte nach Posner et al. (1982) (in deren Anwendung auf Lernprozesse) als Akkommodationsprozess verstanden werden. Sowohl in der wissenschaftsphilosophischen Basierung von T. S. Kuhn (1996), als auch in der Anwendung auf Lernprozesse durch Posner et al. (1982) wird der Wechsel zwischen den Paradigmen/Konzepten als radikal beschrieben. Ein bestehendes Schema wird zugunsten des neu zu entwickelnden Schemas in Gänze gewechselt.

Wissenschaftsphilosophisch lässt sich dieses Verständnis von der Konzeptualisierung von Lakatos (1978) unterscheiden. Seine Kritik am (simplen) Falsifikationismus nach Popper (2005) wird in Bezug auf ein Verständnis der Genese wissenschaftlicher Erkenntnis durch *trial-and-error* Prozesse begründet. Während Popper (2005) durch wissenschaftliche Erkenntnisse ganze Theorien in Zweifel zieht, bzw. diese falsifiziert, fokussiert Lakatos (1978) auf die Veränderungen und Bearbeitung von Theorien durch neue Erkenntnisse. Die wissenschaftliche Praxis wird daher in größeren Zusammenhängen verstanden. In diesem Verständnis von Wissenschaft werden Theorien als Zusammensetzung von „hard core“ (Lakatos, 1978, S. 4) Annahmen und einem „protective belt“ (Lakatos, 1978, S. 4) beschrieben. Dabei enthält der *protective belt* vor allem Hypothesen im Bereich der Theorie oder leichter abwandelbare Vorstellungen, welche sich spezifisch auf Einzelfälle

anwenden lassen. Kommt es zu Anomalien und daraus resultierender Unzufriedenheit, werden Veränderungen am *protective belt* vorgenommen. Erst wenn die Anomalien das bestehende Konzept (hier: „scientific research programmes“ Lakatos, 1978, S. 4) weitreichend in Frage stellen, werden Veränderungen am *hard core* vorgenommen. Bezogen auf das obige Beispiel des Bohrschen Atommodells, könnten bei einzelnen Anomalien Ausnahmen für das Modell argumentiert werden. So müsste das Wasserstoffatom nach Bohrs Definition scheibenförmig, also zweidimensionaler Struktur sein, was jedoch im Widerspruch zur kinetischen Gastheorie steht (Lichtenegger, 2015). Hier könnte eine Einschränkung an Bohrs Theorie beschrieben werden. Die Abweichung zwischen der Messung und Berechnung von Spektrallinien auf Grundlage Bohrs Postulate von Mehrelektronensystemen hingegen, könnte zur notwendigen Veränderung von *hard core* Annahmen führen. Es würden deutliche Eingriffe in die Theorie nötig werden, welche auf die *hard core*-Annahmen zurückzuführen sind. Nach Lakatos (1978) kann eine Transformation von Theorien nicht nur durch radikale Wechsel geschehen, sondern auch durch graduelle Veränderungen. Solche Veränderungen am *protective belt* könnten den Übergang zwischen den von Posner et al. (1982) beschriebenen Assimilations- und Akkommodationsprozessen darstellen. Der durch Lakatos (1978) beschriebene *hard core* kann mit dem Verständnis eines „solid core“ nach Schommer-Aikins (2002, S. 114) verglichen werden (vgl. Kapitel 3.2).

Als Ausgangspunkt der *Conceptual Change* Forschung wird in der Arbeit von Posner et al. (1982) der Wortursprung der *Conceptual Change* Theorie deutlich. Ursprünglich als radikaler Wechsel (*Change*) verstanden, wurden in späteren Konzeptualisierungen andere Begriffe (z.B. *Conceptual growth*; Gropengießer & Marohn, 2018) aufgrund anderer Schwerpunktsetzungen im Prozess gewählt. Damit verbunden sind oft auch unterschiedliche wissenschaftsphilosophische Begründungen.

In ihrer Ursprungskonzeption fokussieren Posner et al. (1982) auf Akkommodationsprozesse. Die Autoren beschreiben unterschiedliche Bedingungen, welche zum Auslösen dieser Prozesse notwendig sind:

- i. Unzufriedenheit mit dem bestehenden Konzept
- ii. Verständlichkeit des neuen Konzepts
- iii. Plausibilität des neuen Konzepts
- iv. Fruchtbarkeit des neuen Konzepts in Bezug auf weitere Phänomene

Um einen Konzeptwechsel herbeizuführen, beschreiben Posner et al. (1982) die Notwendigkeit von (i) Unzufriedenheit mit dem bestehenden Konzept. Besteht die Unzufriedenheit mit dem Konzept an mehreren Stellen, bzw. in Zusammenhang mit mehreren (aus Perspektive des initialen Konzepts) unerklärbaren Phänomenen, ist die Wahrscheinlichkeit höher, dass ein *Conceptual Change* auftreten kann. Während kleinere Unzufriedenheiten ggf. mit kleineren Modifikationen am bestehenden Konzept gelöst werden können, verlangen größere sowie multiple Unzufriedenheiten nach Posner et al. (1982) einen radikalen Konzeptwechsel. Dabei ist jedoch die (ii) Verständlichkeit des neuen Konzepts für die Lernenden entscheidend. Es muss also erkennbar sein, wie das neue Phänomen durch das neue Konzept strukturiert werden kann. Ebenso muss dem neuen Konzept eine gewisse (iii) Plausibilität inhärent sein. Phänomene, die mit dem initialen Konzept erklärt werden konnten, müssen also auch durch das neue Konzept erklärbar sein. Auch sollten keine Widersprüche zu anderen bestehenden Konzepten vorliegen. Dieser Punkt hängt gleichzeitig mit der (iv) Fruchtbarkeit des Konzepts in Bezug auf andere Phänomene zusammen. Das

neue Konzept sollte den Anschein erwecken, dass auch zukünftige Phänomene durch dieses erklärbar sein werden.

Neben diesen Bedingungen, welche direkt an den Unzufriedenheiten, bzw. dem Konzept ansetzen, beschreiben Posner et al. (1982, S. 213) auch einen wichtigen Einfluss der „conceptual ecology“. In diese konzeptuelle Umwelt beziehen die Autoren verschiedene Aspekte ein, welche einen Einfluss auf die Auswahl des neuen Konzeptes haben. Eine wichtige Rolle schreiben sie hier den auftretenden Anomalien zu. Die zur Unzufriedenheit führenden Anomalien haben Einfluss auf die Wahl des neuen Konzepts. Auch Analogien oder Metaphern können bei der Wahl des neuen Konzepts helfen.

Kognitive Konflikte

Zur Initiierung eines *Conceptual Change* Prozesses eignet sich nach Posner et al. (1982) vor allem ein kognitiver Konflikt. Ein kognitiver Konflikt meint (aus Perspektive Piagets Äquilibration) das Ungleichgewicht zwischen Präkonzepten und beobachteten Phänomenen oder zwei kognitiven Strukturen (Kang et al., 2004). Vor allem der erste beschriebene Konflikt fokussiert dabei auf „discrepant events“ (Kang et al., 2004, S. 73), welche das Auftreten anomaler, diskrepanter Daten/Beobachtungen beschreiben. Appleton (1993, S. 1) hebt in seiner Beschreibung eines kognitiven Konflikts besonders die herausfordernde Situation hervor, wenn er einen kognitiven Konflikt als „puzzling situation which was counter-intuitive“ beschreibt. Wissenschaftsphilosophisch steht das hier formulierte Verständnis eines kognitiven Konflikts in einer Linie mit der durch T. S. Kuhn (1996) beschriebenen Krise als Notwendigkeit einer wissenschaftlichen Revolution.

Das Konzept des kognitiven Konflikts wurde durch Limón (2001) deutlich kritisiert. Ein zentraler Kritikpunkt ist die (fehlende) Bedeutsamkeit kognitiver Konflikte für Lernende. Die hohe Individualität der Bedeutsamkeit von Anomalien (im Sinne diskrepanter Daten/Beobachtungen) kann zu sehr unterschiedlichen Erfolgen bei deren Einsatz im Klassenraum führen (siehe auch Kap. 3.3.3). Gleichzeitig wird kritisiert, dass die bestehenden Lernendenvorstellungen nicht entsprechend wertgeschätzt, bzw. deren Potential als Schritt hin zu einer wissenschaftlichen Vorstellung nicht erkannt wird. In diesem Zusammenhang postulieren Li et al. (2006, S. 405) die Strategie der „cognitive perturbation“, welche das Ziel hat, Lernendenvorstellungen durch widersprechende Informationen, bzw. Beobachtungen Schritt für Schritt in wissenschaftliche Vorstellungen zu überführen. In einer Untersuchung von Dega et al. (2013) konnte eine größere Effektivität dieser Strategie im Vergleich zur klassischen Induzierung eines kognitiven Konflikts für die Physik beschrieben werden. Bucat (2015) kritisiert, dass Anomalien bei unterschiedlichen Lernenden zu völlig unterschiedlichen kognitiven Konflikten führen könnten. Intendierte kognitive Konflikte werden zumeist vorab spezifisch ausgeplant und vernachlässigen dabei die Individualität der Lernenden, wie auch die nicht lineare und deterministische Veränderungen von Lernendenvorstellungen zu wissenschaftlichen Vorstellungen (Li et al., 2006). Limón (2001) zeigt weiterhin die Motivation und das Interesse am Thema, aktiviertes Vorwissen, epistemologische Überzeugungen und adäquate Begründungsmechanismen als entscheidend für einen durch einen kognitiven Konflikt induzierten *Conceptual Change* Prozess auf (Kritik an *cold cognition*). Die Fülle der Anforderungen und relevanten Konstrukte deutet auf die Komplexität der Situation hin. All diese Variablen nehmen Einfluss auf die Bedeutsamkeit des kognitiven Konflikts bei den Lernenden. Weiterhin kritisiert Limón (2001), dass bestehende Forschung zumeist auf das Induzieren eines kognitiven Konflikts durch Anomalien fokussiert. So seien auch Analogien (Duit et al., 2001), Metaphern (Amin, 2009) oder Gruppendiskussionen (Eryilmaz, 2002) geeignete Methoden, um einen kognitiven Konflikt auszulösen. Aktuell werden verschiedene Methoden und Instruktionsarten als effektiv zur Auslösung von *Conceptual Change* Prozessen in der Literatur beschrieben (Pacaci et al., 2023). Diskutiert

werden im Zusammenhang mit den spezifischen Methoden vor allem auch die affektiven Einflussfaktoren, welche einen Einfluss auf verschiedene Methoden zeigen. Für den Chemieunterricht konnten z.B. Baddock und Bucat (2008) Herausforderungen bei der Wahrnehmung eines kognitiven Konflikts aufzeigen. In einem Experiment stellen die Autor*innen starke und schwache Säuren und die jeweilige Färbung eines Indikators gegenüber. Innerhalb der Studie hat sich der kognitive Konflikt als nicht effektiv erwiesen. Baddock und Bucat (2008) postulieren, dass das Experiment nur die ersten drei der durch Posner et al. (1982) Bedingungen für einen *Conceptual Change* (Unzufriedenheit mit dem bestehenden Konzept, Verständlichkeit des neuen Konzepts & Plausibilität des neuen Konzepts) erfüllt habe. Vor allem die Fruchtbarkeit des neuen Konzepts für weitere auftretende Phänomene könne durch dieses Experiment nicht erreicht werden. Gleichzeitig geben Baddock und Bucat (2008) dezidierte Anweisungen, welche Instruktionen additiv zu dem Experiment gegeben werden sollten, damit ein suffizienter kognitiver Konflikt bei den Lernenden entsteht. In Anlehnung an Johnstone (1997) erkennen Baddock und Bucat (2008) Herausforderungen bei der Unterscheidung zwischen *signal* und *noise*. Demzufolge sei es für die Lernenden nicht leicht, zwischen den relevanten und weniger relevanten Beobachtungs- und Konzeptaspekten zu unterscheiden. Die Instruktionen sollen entsprechend einen sortierenden Charakter besitzen, damit der eigentliche kognitive Konflikt besser erkannt werden kann. Die hier im Zusammenhang mit einem Demonstrationsexperiment beschriebenen Herausforderungen könnten sich bei der eigenständigen Durchführung der Experimente weiter verschärfen. Bucat (2015) gibt Beispiele für Situationen die kognitive Konflikte im Chemieunterricht auslösen könnten. So könnte z.B. der unterschiedliche pH-Wert von starken und schwachen Säuren gleicher Konzentration, oder etwa die unterschiedliche pH-Wert-Veränderung von Wasser im Vergleich zu einer Pufferlösung bei Zugabe einer Base zu einem kognitiven Konflikt führen. Dabei kommt es nach Bucat (2015) auf das Unterrichtsetting an, ob ein kognitiver Konflikt ausgelöst werden kann. Ebenso sollte die Auflösung des Konflikts nur wenige Auswertungsschritte umfassen.

Conceptual Change Prozesse generell sind für viele verschiedene Fachbereiche der Chemie untersucht worden, z.B. das Gas-Konzept (Çetin et al., 2009), die Reaktionskinetik (Taştan Kırık & Boz, 2012) und Thermodynamik (Afadil et al., 2022). Dabei sind – ähnlich der generellen Forschungslinie zu *Conceptual Change* – unterschiedliche Instruktionsweisen und Methoden in die Untersuchungen eingeflossen.

Strike und Posner (1992) haben auf Grundlage bestehender Kritik (z.B. radikaler Konzeptwechsel, *cold cognition*, etc.) ihr *Conceptual Change* Modell überarbeitet. So werden im überarbeiteten Modell weitere Faktoren in die konzeptuelle Umwelt aufgenommen. Auch die Interaktion bestehender und sich entwickelnder Konzepte mit der konzeptuellen Umwelt wird in den Blick genommen. Die Notwendigkeit eines radikalen Wechsels zwischen Konzepten wird ebenfalls in Teilen revidiert. Dennoch kann das durch Posner et al. (1982) entwickelte *Conceptual Change* Verständnis – auch als „Classical Approach“ (Vosniadou & Mason, 2012, S. 222) bezeichnet – als einflussreich für das Forschungsfeld beschrieben werden (Potvin et al., 2020). Ihre Übertragung wissenschaftsphilosophischer Theorien auf den Lernprozess hat den Ausgangspunkt für eine Vielzahl unterschiedlicher Modelle gebildet. Neben dem klassischen Ansatz von Posner et al. (1982; 1992) sind viele weitere Modellierungen des *Conceptual Change* Prozesses entstanden (für einen Überblick: Potvin et al., 2020). Im Folgenden sollen ausgewählte und verbreitete Modellierungen knapp vorgestellt werden.

Einen anderen Ansatz zur Beschreibung von *Conceptual Change* Prozessen wählt Chi (2008). Die Autorin fokussiert auf die unzutreffende Zuordnung von Konzepten in ontologische Kategorien. Dabei geht die Autorin davon aus, dass Individuen drei verschiedene ontologische Kategorien

ausprägen, in welche Konzepte gespeichert werden können: „Entities, processes, mental states“ (Chi, 2008, S. 64). Diese ontologischen Kategorien werden dann durch viele Unterkategorien für verschiedene Inhalte differenziert. Durch die Zuordnung eines Konzepts in eine der Kategorien könne so Wissen aus den jeweiligen Kategorien zur Erklärung (auch neuer) Konzepte genutzt werden. Lernendenvorstellungen entstünden vor allem durch die unzutreffende Zuordnung der Konzepte zu den Kategorien (Vosniadou & Mason, 2012). Das Aufzeigen der unterschiedlichen Kategorien in Verbindung mit einer Bewusstmachung der fehlerhaften Zuordnung könne somit zu einer Vorstellungsveränderung führen (Chi, 2008). In Bezug auf die vorliegende Studie könnte die Betrachtung von Zuordnungen zu epistemologischen Kategorien relevant werden. So könnten epistemologische Kategorien wie „Wissen“ und „Vermutungen“ relevanten Einfluss auf die Verarbeitung von Informationen und deren Nutzung als Evidenzen haben. Auch die Zuordnung von Beobachtungen als „richtig“ oder „falsch“ (im Sinne anomaler Daten) könnte großen Einfluss auf die Verarbeitung von Informationen und somit einen möglichen kognitiven Konflikt nehmen. Ebenso könnte die ontologische Kategorisierung von Entitäten innerhalb von Experimenten zu Verständnis von kognitiven Konflikten beitragen, oder dieses erschweren.

diSessa (2002) beschreibt die Konstruktion/Veränderung von Konzepten durch sogenannte „*p-prims*“ (diSessa, 2002, S. 38). *P-prims* („phenomenological primitives“ diSessa, 2002, S. 38) beschreiben kleine Wissensaspekte, welche selbst keine weitere Erläuterung benötigen. Dem Verständnis des Aufbaus der Materie durch Atome ähnlich, beschreibt diSessa (2002) den Aufbau von Konzepten durch die Vernetzung von *p-prims*. Dabei ist vor allem die Vernetzung ein entscheidender Faktor. Während Noviz*innen *p-prims* häufig nicht, oder wenig miteinander vernetzen, entstehen durch die Vernetzung der *p-prims* bei Expert*innen „*coordination classes*“ (diSessa, 2002, S. 43). *Coordination classes* können als komplexe Systeme verstanden werden, welche *p-prims* enthalten. Die Vernetzung der *p-prims* untereinander macht dabei die Natur der *coordination classes* aus. Wird ein Konzept entwickelt, werden neue *p-prims* aufgenommen, welche durch Vernetzung zu *coordination classes* geformt werden können (Gropengießer & Marohn, 2018; Vosniadou & Mason, 2012). Vor allem im Zusammenhang mit der Modellierungsstudie innerhalb dieser Arbeit (vgl. Kap. 8), spielt die Vernetzung und Umstrukturierung von *p-prims* eine wichtige Rolle. Die Offenheit zur Umstrukturierung bestehender Strukturen (und das damit verbundene Anerkennen der Unsicherheit von Wissen) könnten großen Einfluss auf die Akzeptanz von anomalen Daten sowie das Auslösen eines möglichen kognitiven Konflikts nehmen. Hier zeigt sich erneut der mögliche direkte Einfluss epistemologischer Überzeugungen auf die Verarbeitung von Beobachtungen und Daten.

Besonders hervorzuheben ist das Konzept der *learning progressions* (auch: *learning pathways*, *conceptual pathways*). Diese beschreiben die Entwicklung von Wissen zu einem Basiskonzept (z.B. Energie) ausgehend von Präkonzepten, bis hin zu wissenschaftlich anerkannten Konzepten (Smith & Wiser, 2015). Dabei können *learning progressions* zur Kohärenz auf mehreren Ebenen beitragen: Im Sinne einer Entwicklungskohärenz können *learning progressions* den Weg von intuitiven zu wissenschaftlichen Vorstellungen beschreiben. Dabei plädieren Jin et al. (2019) dafür, konzeptuelle und prozessuale Kompetenzentwicklungen in Verbindung miteinander zu betrachten. So sollten *learning progressions* im Sinne der Kohärenz konzeptuelle und prozessuale Kompetenzen gleichermaßen abbilden. Ausgehend von der Entwicklungskohärenz, können *learning progressions* zu horizontaler Kohärenz beitragen. Damit ist gemeint, dass entwickelte *learning progressions* durch wissenschaftliche Studien auf die kausalen Zusammenhänge der Konzeptualisierung und entstehende Kompetenzzuwächse untersucht werden, sodass eine empirische Untermauerung der jeweiligen *learning progression* entsteht. Können *learning progressions* der wissenschaftlichen Prüfung entlang diverser Qualitätskriterien standhalten, sollten diese zur Grundlage von

bestehenden Curricula werden (Jin et al., 2019). Ein Beispiel hierfür kann in den spiralcurricular aufgebauten Kernlehrplänen im Fach Chemie erkannt werden (siehe auch: *learning progressions* für die Sekundarstufe I des Fachs Chemie bei Celik, 2022 und K. Weber, 2018). Ein Kompetenzaufbau im Bereich der verschiedenen Basiskonzepte folgt dabei der Idee von *learning progressions* (MSB NRW, 2019, 2022). In Bezug auf vertikale Kohärenz beschreiben Jin et al. (2019) einige Herausforderungen. Vertikale Kohärenz stellt auf die Messung des Kompetenzzuwachs' ab. Dabei ergeben sich einige Probleme, die aus der Natur langfristig angelegter *learning progressions* entspringen (Für eine Übersicht der unterschiedlichen Arten von Kohärenz und den resultierenden Herausforderungen siehe Jin et al., 2019). *Learning progressions* können dabei nicht nur auf konzeptuellen oder prozessualen Kompetenzen angewendet werden. Auch der Kompetenzerwerb meta-konzeptueller Strukturen wie der Epistemologie könnte durch *learning progressions* beschrieben werden (D. Kuhn et al., 2000). Im Sinne der oben geführten Diskussionen ist jedoch fraglich, inwiefern ein Entwicklungsendpunkt der *learning progression* beschrieben werden kann, oder sollte (im Sinne der Vorläufigkeit von Wissen). Ebenso sollten meta-konzeptuelle Konstrukte als Einflussfaktor oder Konstituente von *learning progressions* verstanden werden (Adadan et al., 2010).

Learning progressions sollten dabei nicht als Methode per se verstanden werden. Vielmehr handelt es sich um eine übergeordnete Konzeptualisierung des Erkenntniswegs von Lernenden, in welche weitere Methoden wie z.B. *Conceptual Change* Prozesse eingebettet werden können (Smith & Wiser, 2015). Dennoch zeigen *learning progressions* einen langangelegten Kompetenzaufbau auf, in welchem Vorstellungen von Lernenden immer wieder erweitert, bzw. verändert werden. Während dabei der radikale Konzeptwechsel nach Posner et al. (1982) eine Methode darstellen kann, wird Lernen insgesamt eher im Sinne der obigen Formulierungen (z.B. *conceptual growth*) beschrieben.

Die hier dargestellte Auswahl von *Conceptual Change* Konzeptualisierungen ließe sich weiter ergänzen (z.B. Niaz, 1998; Berland & Russ, 2017; etc. – Für einen Überblick: Potvin et al., 2020; Vosniadou & Mason, 2012). Deutlich wird jedoch, dass unterschiedliche Konzeptualisierungen teils unterschiedliche wissenschaftsphilosophische Schwerpunkte haben. Ein genereller Trend hin zu Konzepten gradueller Bearbeitung bestehender Vorstellungen kann in den Arbeiten von Lakatos (1978) begründet werden.

Die hier beschriebenen Modellierungen des *Conceptual Change* Prozesses wurden über die Zeit viel diskutiert und empirisch untersucht. Amin et al. (2014) erkennen drei verschiedene Phasen der *Conceptual Change* Forschung. Innerhalb der ersten Phase wurde die domänenspezifische Ausprägung von *Conceptual Change* Prozessen sowie die Bedeutung der Argumentation innerhalb der Naturwissenschaft fokussiert. In diese Phase fällt auch die Konzeption relevanter Modellierungen des *Conceptual Change* Prozesses, etwa von Posner et al. (1982), Vosniadou und Brewer (1987) und diSessa (2002). Innerhalb der zweiten Forschungsphase wurde ein Fokus auch auf ontologische und epistemologische Überzeugungen und deren Einfluss auf den *Conceptual Change* gelegt. Innerhalb der dritten Phase folgt eine systemische Betrachtung. *Conceptual Change* Prozesse werden dabei als voraussetzungsvoll und kontextabhängig betrachtet. In diese Phase fällt auch die Betrachtung von *learning progressions* (Amin et al., 2014).

Potvin et al. (2020) zeigen in ihrem umfassenden Review von Artikeln ($N=245$) aus fünf bedeutenden Zeitschriften naturwissenschaftsdidaktischer Forschung die Bedeutung von insgesamt 86 unterschiedlichen Modellierungen des *Conceptual Change* Prozesses dar. Die Autor*innen ordnen die unterschiedlichen Modellierungen dabei hierarchisch entsprechend den im Korpus zu findenden empirischen und theoretischen Argumenten. Dabei sind die hier beschriebenen

Modellierungen unter den ersten 6 Plätzen der hierarchischen Liste zu finden. Bezogen auf die Anzahl empirischer Untermauerungen liegen die Konzeption von Posner et al. (1982) und Vosniadou (1994) deutlich vorn. Gleichzeitig zeigt sich insbesondere für die Konzeption von Posner et al. (1982) eine abnehmende Bedeutung im Sinne der Nutzung innerhalb aktueller Publikationen. Dieser Trend lässt sich jedoch auch insgesamt interpretieren. Einen Schwerpunkt der Anzahl von Publikationen zu den *Conceptual Change* Modellen liegt in den 2000er-Jahren. Bis 2018 ist insgesamt eine deutlich Abnahme der Publikationszahlen erkennbar (Potvin et al., 2020).

Dieser Arbeit wird eine Kombination der Modelle von diSessa (2002) und Chi (2008) zu Grunde gelegt. Dabei können *p-prims* auch als aufgenommene Beobachtungen verstanden werden, welche ggf. ontologischen Kategorien zugeordnet werden können.

Posner et al. (1982) beschreiben einen Einfluss epistemologischer Überzeugungen auf *Conceptual Change* Prozesse. So könnten Überzeugungen hinsichtlich der Qualität guter, wissenschaftlicher Erklärungen in der jeweiligen Domäne wie auch allgemeine epistemologische Überzeugungen einen Einfluss auf die Konzeptwahl haben. Auch metaphysische Überzeugungen und Konzepte über Wissenschaft könnten die Wahl lenken. Als Beispiel führen Posner et al. (1982) Einsteins Überlegungen an. So könnten zum Beispiel Überzeugungen über die Symmetrie und Unzufälligkeit des Universums zu bestimmten epistemologischen Überzeugungen führen, welche Einfluss auf die Wahl von Konzepten haben würden. Als weiteren Einflussfaktor der Umwelt auf die Wahl eines Konzepts beschreiben die Autoren Wissen in anderen Bereichen sowie konkurrierende Konzepte. Diese „conceptual ecology“ (Posner et al., 1982, S. 214) kann einen deutlichen Einfluss auf die Wahl eines Konzepts haben. In Bezug auf Forschung im Bereich der Epistemologie könnte diese Umwelt als Kontext beschrieben werden, welcher im individuellen Fall unterschiedlich auf den Prozess einwirken kann. Je nachdem welche Phänomene aus Perspektive welches Konzepts betrachtet werden, könnten unterschiedliche (epistemologische) Überzeugungen aktiviert werden. Das Konzept einer konzeptuellen Umwelt nach Posner et al. (1982) kann damit als anschlussfähig bzgl. der Kontextabhängigkeit epistemologischer Überzeugungen (Schommer-Aikins, 2002) beschrieben werden. Posner et al. (1982) weisen weiterhin auf die besonders relevante Rolle epistemologischer Überzeugungen im *Conceptual Change* Prozess hin.

Neuere Forschungsbereiche fokussieren vor allem auf die Verknüpfung von *Conceptual Change* Prozessen zu den epistemologischen Überzeugungen von Lernenden (Hammer, 2017; Hofer, 2017; Jiménez-Aleixandre & Brocos, 2017; Sandoval, 2017). Dabei wird grundsätzlich ein Einfluss epistemologischen Denkens und Argumentierens auf Lernprozesse und damit *Conceptual Change* Prozesse formuliert: “How individuals think and reason epistemically can influence how they learn, how they resolve competing truth claims, what they count as worthy evidence, and the degree to which they place trust in authority, expertise, and their own experiences.” (Hofer, 2017, S. 229). Epistemologischen Überzeugungen wird vor allem ein Einfluss auf die Verarbeitung von Daten und Informationen zugestanden, welcher sich somit auf den *Conceptual Change* Prozess auswirken können. In diesem Zusammenhang wird auch deutlich, dass die letztliche Modellierung des *Conceptual Change* Prozesses (z.B. verstanden als radikaler Wechsel, oder schrittweise Erweiterung von Vorstellungen) weniger entscheidend ist. Die epistemologischen Überzeugungen können sich in diesem Verständnis auf jegliche Modellierungen auswirken. Dabei schlagen Sinatra und Chinn (2012) die Erweiterung epistemologischer Überzeugungen durch *Conceptual Change* Prozesse vor. So könnten epistemologische Überzeugungen Gegenstand von *Conceptual Change* Prozessen werden. *Conceptual Change* Prozesse könnten folglich konzeptuelle, aber auch metakonzeptuelle Entitäten umfassen. Dabei sei beiden Ansätzen die notwendige Identifizierung von Verständnishürden inhärent (Sinatra & Mason, 2013). Je nachdem wie stabil die zu ändernden

Überzeugungen sind, könnten verschiedene methodische Initiierungen eines *Conceptual Change* Prozesses bevorzugt werden (Sinatra & Mason, 2013). Nach Jiménez-Aleixandre und Brocos (2017) könnte der spezifische Kontext, innerhalb dessen *Conceptual Change* Prozesse ablaufen, ebenfalls einen Einfluss auf aktivierte epistemologische Überzeugungen haben. Entlang der durch Posner et al. (1982) formulierten Bedingungen eines *Conceptual Change* Prozesses, machen Jiménez-Aleixandre und Brocos (2017) Unterschiede zwischen der Entwicklung von Erklärungen und Modellen und der Bearbeitung von *socio-scientific issues* (SSI) deutlich. Ähnliche Ansätze sind auch bei Hofer (2017) erkennbar, welche zwischen dem Verständnis einer Theorie und deren Akzeptanz im Kontext der Evolutionstheorie unterscheidet. Demnach könnten Theorien verstanden werden, ohne dass diese auch akzeptiert werden. Hier wird das Zusammenspiel kognitiver und metakognitiver Überzeugungen deutlich. Die Kontextabhängigkeit dieser *Conceptual Change* Prozesse spiegelt hier die Kontextabhängigkeit epistemologischer Überzeugungen generell wider (s.o.). Viele der beschriebenen Studien fokussieren nicht allein auf epistemologische Überzeugungen. Im Sinne von *warm cognition* (als Reaktion auf Kritik von Pintrich et al., 1993) werden vermehrt auch affektive Variablen in die Betrachtung einbezogen. Im Zuge der zunehmenden Bedeutung des Forschungsbereichs *Nature of science* (vgl. Kap. 3.4) findet jedoch immer seltener eine Abgrenzung zu eben diesem Konstrukt statt. Folglich müssen, ähnlich wie bereits dargelegt, die zu Grunde gelegten Definitionen von Epistemologie, bzw. deren Umfang kritisch reflektiert werden.

Wie exemplarisch aufgezeigt wurde, können epistemologische Überzeugungen deutlichen Einfluss auf unterschiedliche Aspekte von *Conceptual Change* Prozessen nehmen. Sei es die Verarbeitung von Informationen, die Abschätzung von Expertise und Autorität, oder Fragen bzgl. der Akzeptanz einer Theorie. Basierend auf dem Gedanken der konzeptuellen Umwelt nach Posner et al. (1982) werden epistemologische Überzeugungen als ein ausschlaggebender Faktor für *Conceptual Change* Prozesse beschrieben. Dabei wird vor allem auf den Bereich *Nature of knowing* im Sinne der Modellierung von Epistemologie nach Hofer und Pintrich (1997) eingegangen, in dem die Bewertung der Glaubwürdigkeit von Kognitionen und die Begründbarkeit/Argumentation von Konzepten fokussiert wird. Gleichzeitig scheint dieser Einfluss epistemologischer Überzeugungen unabhängig von der spezifischen Modellierung des *Conceptual Change* Prozesses relevant zu sein. Dies ist anschlussfähig an das Verständnis von epistemologischen Überzeugungen als Bestandteil der Metakognition.

3.3.2 Framework-theories

Einen möglichen Ansatz zur genaueren Untersuchung des Einflusses epistemologischer Überzeugungen auf den *Conceptual Change* Prozess kann die *framework theory* nach Vosniadou (1994, 2014) darstellen.

Vosniadous Verständnis eines *Conceptual Change* basiert auf der Verarbeitung neuer Beobachtungen und Informationen in mentalen Modellen (Vosniadou & Brewer, 1992). Mentale Modelle können als interne Repräsentationen der wahrgenommenen Außenwelt beschrieben werden (für einen Überblick: Nitz & Fechner, 2018). Sie stellen so einen Erklärungsrahmen für Phänomene dar, ohne zwangsläufig mit einer wissenschaftlich adäquaten Erklärung des Phänomens übereinzustimmen (Vosniadou & Brewer, 1992). In einem ersten Zugang zu neuen Phänomenen, können aufgenommene Kognitionen mithilfe eines mentalen Modells sortiert werden (Jones et al., 2011). Um anschließend die Passung der neuen Kognitionen mit dem (vorhandenen) mentalen Modell zu überprüfen, können Plausibilitätsprüfungen und mentale Simulationen durchgeführt werden (Chinn & Brewer, 2001; Gentner, 2001). Damit die aufgenommenen Informationen als ‚Wissen‘ in das mentale Modell aufgenommen werden, müssen diese aus Sicht der Lernenden evidenzbasiert

begründbar und ‚wahr‘ sein (Kitchener, 2002). Die notwendige Wahrheit der Informationen bezieht sich hier auf die Existenz in der realen Welt. Damit unterscheidet Kitchener (2002) wahre Informationen z.B. von Informationen aus Paranoia. Vosniadou (1994) postuliert, dass mentale Modelle aus dem Langzeitgedächtnis (re-)aktiviert oder situativ gebildet werden können, vor allem bei hoher Erklärungsgüte und -reichweite des Modells. Die essentiellen Aspekte Vosniadous Verständnisses mentaler Modelle als dynamische, veränderbare interne Repräsentationen von Phänomenen der Außenwelt, inkl. der Möglichkeit aufgenommene Informationen vor dem Hintergrund dieser zu reflektieren, stellen die Grundlage für den Umgang mit Evidenzen im Sinne evidenzbasierter Praxis dar.

Mentale Modelle werden von Vosniadou (1994, S. 48) in größere „framework theories“ eingebettet. Diese *framework theories* beinhalten konzeptuelle Vorstellungen und epistemologische (sowie ontologische) Überzeugungen, die durch eine Vielzahl an Alltagserfahrungen bestätigt wurden (vgl. Verständnis von Epistemologie als *theory in action* bei D. Kuhn & Weinstock, 2002). Sie stellen daher zumeist ein sehr kohärentes Erklärungssystem dar. Als Beispiel führt Vosniadou ihre Forschung zu Vorstellungen einer flachen Erde bei Kindern an. Die wissenschaftliche Vorstellung einer Erde als Kugel wird durch viele ontologische Kategorien (wie zum Beispiel die Festlegung von Richtungen als ‚unten‘ und ‚oben‘) behindert. Damit wird es für Kinder sehr schwer ihre Vorstellung einer flachen Erde zu der Vorstellung einer Erde als Kugel zu verändern (Vosniadou, 1994; Vosniadou & Brewer, 1992). Derlei grundlegende Überzeugungen sind Bestandteil von *framework theories*. Ebenso können Vorstellungen über die Quelle von Wissen, z.B. durch Expert*innen und Autoritäten, zur Annahme oder Ablehnung von Informationen führen. Enthaltene epistemologische und ontologische Überzeugungen bilden die Grundlage von Wissen, bzw. das Verständnis des Wissensaufbaus. Änderungen im Bereich meta-kognitiver Überzeugungen sind daher sehr schwer herbeizuführen und nehmen großen Einfluss auf (konzeptuell-geprägte) mentale Modelle. Hier spiegelt sich das Verständnis epistemologischer Überzeugungen als theorieähnlich und damit einflussreich auf viele verschiedene Wissensbereiche wider (Hofer & Pintrich, 1997).

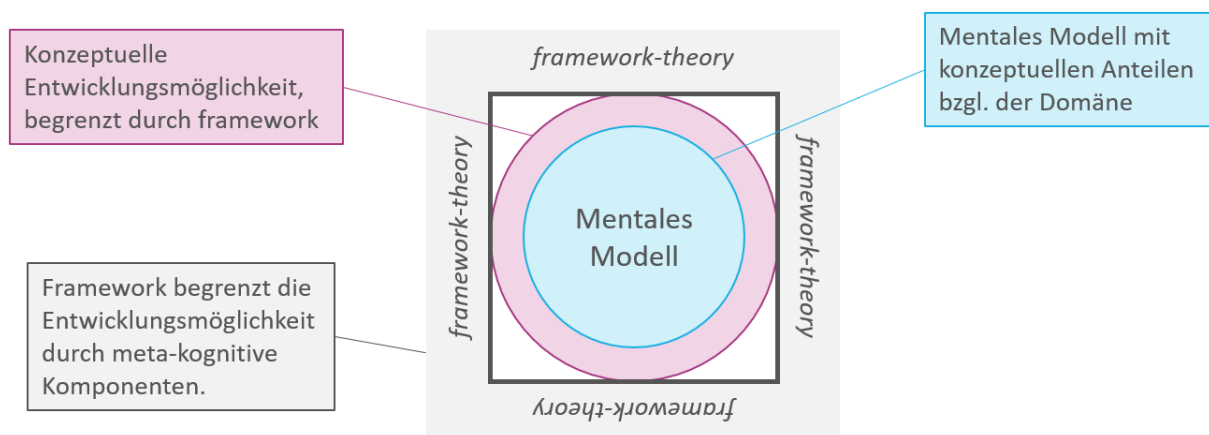


Abbildung 5 - Einbettung mentaler Modelle in *framework theories* nach Vosniadou (1994)

Abbildung 5 stellt den Zusammenhang zwischen mentalen Modellen und *framework theories* graphisch dar. Das mentale Modell umfasst dabei konzeptuelle Aspekte zu einem spezifischen Phänomen. Im Sinne eines *Conceptual Change* Prozesses können diese mentalen Modelle erweitert werden, um auch neue Phänomene erklären zu können. Die konzeptuellen Entwicklungsmöglichkeiten können dabei jedoch durch *framework theories* begrenzt werden. Dadurch haben *framework theories* Einfluss auf einen möglichen *Conceptual Change*. Vosniadou (1994) stellt daher die

Notwendigkeit eines *Conceptual Change* heraus, welcher auch die Überarbeitung ontologischer und epistemologischer Überzeugungen nötig macht (vgl. auch Sinatra & Chinn, 2012). Durch die hohe Stabilität der *framework theories* sei dieser Prozess jedoch nur schwer zu initiieren. Ein möglicher Ansatz könnte es sein, den hypothesenartigen Charakter epistemologischer und ontologischer Überzeugungen zu fokussieren (Vosniadou & Ioannides, 1998). So könnte die hohe Stabilität der *framework theories* ggf. verändert werden.

Das Modell der *framework theories* wurde empirischen Testungen unterzogen. So konnten bspw. Brown et al. (2018) die Vorstellungsveränderung bzgl. Werkstoffmechanik entlang Vosniadous Modell erläutern. Auch Untersuchungen zum Tag-Nacht-Rhythmus und der damit verbundenen Stellung der Planeten zueinander wurde in der Gruppe um Vosniadou vielfach aus Perspektive der *framework theories* untersucht (z.B. Vosniadou & Skopeliti, 2014, in Ansätzen: Vosniadou & Brewer, 1994).

In knapp 40 Jahren der Forschung im Bereich *Conceptual Change* sind viele verschiedene Konzeptualisierungen entstanden. Es zeigt sich ein zunehmendes Verständnis eines graduellen *Conceptual Change* (z.B. Vosniadou & Ioannides, 1998), welcher auch durch epistemologische und ontologische Überzeugungen beeinflusst wird (z.B. Alvermann & Qian, 2000; Qian & Alvermann, 1995). Basierend auf philosophischen Konzepten zur Vorstellungsveränderung können unterschiedliche Operationalisierungen für den Unterricht erkannt werden. Dabei spielt vor allem die Induzierung eines kognitiven Konflikts eine wichtige Rolle (z.B. Bucat, 2015; Posner et al., 1982). Epistemologische und ontologische Überzeugungen können als Bestandteil der *framework theories* verstanden werden und haben so Einfluss auf den *Conceptual Change* (Vosniadou, 1994; Vosniadou & Ioannides, 1998). Neben diesen hauptsächlich theoretischen Überlegungen werden vermehrt auch unterrichtspraktische Einflüsse betrachtet. So kritisieren Pintrich et al. (1993) rein kognitive Modelle für die fehlende Betrachtung von Motivation und der Situation im Klassenraum. Beides seien wichtige Einflussfaktoren für die kognitiven Prozesse von Lernende und mögliche *Conceptual Change* Prozesse.

3.3.3 Umgang mit anomalen Daten

Die Initiierung eines kognitiven Konflikts stellt nach wie vor eine wichtige Strategie zur Auslösung eines *Conceptual Change* dar (Gropengießer & Marohn, 2018). Um einen solchen Konflikt auszulösen, müssen Präkonzepte mit beobachteten Phänomenen in Kontrast stehen (vgl. Kapitel 3.2.1). In diesem Zuge erhält die Präsentation von anomalen Daten (oder auch: „discrepant events“, Kang et al., 2004, S. 73) eine besondere Bedeutung (C. K. K. Chan, 2001). In Rückbezug auf Appleton (1993), werden anomale Daten (im Sinne des *Model of argument* nach Toulmin, 2003) hier als Beobachtungen oder Informationen insgesamt betrachtet, welche den ursprünglichen Annahmen entgegenstehen. Anomale Daten werden im Kontext dieser Arbeit auch als widersprüchliche Beobachtungen verstanden. Dabei ist die Frage ob Beobachtungen als anomal betrachtet werden nur vor den individuellen Vorstellungen der Lernenden zu bewerten (vgl. Bucat, 2015).

Chinn und Brewer (1993b) befassen sich in ihrer Forschung mit dem Umgang der Lernenden mit anomalen Daten. Noch heute, knapp 30 Jahre nach der Studie von Chinn und Brewer (1993b), stellt deren Arbeit einen wichtigen Orientierungspunkt für das Forschungsfeld dar (Potvin, 2023). Ebenso haben sie Charakteristika ausgearbeitet, welche einen Einfluss auf die unterschiedlichen Umgangsarten haben können.

Ein entscheidendes Charakteristikum anomaler Daten ist ihre **Glaubwürdigkeit** (Chinn & Brewer, 1993b). Stammen die Daten aus unglaublichen Quellen oder ist der Prozess der Datengewinnung unklar, ist es unwahrscheinlicher, dass durch sie ein kognitiver Konflikt ausgelöst werden kann. Im Idealfall erheben die Lernenden die Daten selbst, da sie so im Sinne einer Primärerfahrung direkte Kontrolle über die Beobachtungssituation und den Experimentierprozess haben. Weiterhin sollten die anomalen Daten möglichst eindeutig sein. Je klarer die Daten dem bestehenden mentalen Modell widersprechen, desto deutlicher werden sie als anomal wahrgenommen (Chinn & Brewer, 1993b). Für die Beschreibung der **Eindeutigkeit** ist vor allem eine Vergleichbarkeit zum bestehenden mentalen Modell wichtig. Konkurrierende Theorien verwenden häufig ein unterschiedliches Vokabular, wodurch die Vergleichbarkeit und damit die Bewertung der eindeutigen Abgrenzung nicht einfach möglich sind. Wird der Aufenthaltsort von Elektronen nach Niels Bohr als ‚Schale‘ beschrieben, im Molekül-Orbital-Modell jedoch als ‚Orbital‘, kann das unterschiedliche Vokabular bereits dazu führen, dass Daten nicht als anomal wahrgenommen werden. Es kann eine Unterscheidung der beiden Datensätze aufgrund der Betrachtung aus Perspektive unterschiedlicher Theorien erfolgen. Zuletzt heben Chinn und Brewer (1993b) die Bedeutung **multipler Anomalien** hervor. Während einzelne Anomalien ggf. in das vorhandene mentale Modell assimiliert werden können, oder nur Veränderungen am *protective belt* nötig machen (vgl. Lakatos, 1978), könnten multiple Anomalien größere Veränderungen erfordern. Chinn und Brewer (1993b, S. 29) schreiben dazu: „[...] students began to run out of counterexplanations for the anomalous data.“. Folglich müssen sie die Anomalien als abweichend von der Ausgangstheorie wahrnehmen, es gibt keine Gelegenheit mehr, diese mit der Theorie in Einklang zu bringen.

Zusammenfassend sind also multiple, möglichst eindeutige und zugleich glaubwürdige anomale Daten besonders geeignet, um einen *Conceptual Change* Prozess auszulösen. Entscheidend ist jedoch auch der Bezugsrahmen: Daten können nur dann als anomal betrachtet werden, wenn sie eine bestehende Theorie oder ein mentales Modell diskreditieren. Es wird also ein Abgrenzungsbeispiel für anomale Daten benötigt, wodurch anomale Daten individuell deutlich unterschiedlich aufgegriffen werden können. So können individuelle mentale Modelle zu unterschiedlichen Umgangsarten mit anomalen Daten führen. Gleichzeitig wird die große Bedeutung epistemologischer sowie ontologischer Überzeugungen deutlich. Beispielsweise könnte die Einschätzung der Glaubwürdigkeit von Daten deutlich mit den epistemologischen Überzeugungen der Dimension *Nature of knowing* nach Hofer und Pintrich (1997) zusammenhängen.

Chinn und Brewer (1993b) postulieren sieben unterschiedliche Umgangsarten von Lernenden mit anomalen Daten, welche nach einer empirischen Testung um eine weitere Umgangsart ergänzt wurden (Chinn & Brewer, 1998). Die unterschiedlichen Umgangsarten sind hierarchisch sortiert, basierend auf ihrem Grad der Theoriebearbeitung, bzw. des Protektionismus bezüglich des bestehenden Modells (vgl. Abb. 6). Protektionismus meint dabei das Verteidigen und Festhalten an der initialen Theorie.

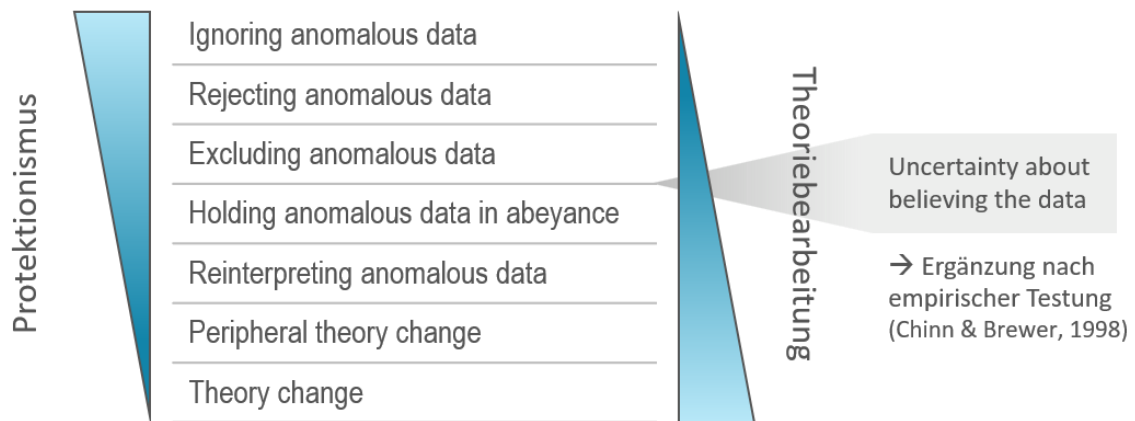


Abbildung 6 - Umgangsarten mit anomalen Daten nach Chinn und Brewer (1993b, 1998)

Während die oberen Umgangsarten vor allem auf das Beibehalten des bestehenden mentalen Modells fokussieren, nimmt der Grad der Theoriebearbeitung hin zu den unteren Umgangsarten zu (vgl. Abb. 6). Der *theory change* (vergleichbar mit dem Verständnis eines radikalen *Conceptual Change*) beschreibt dann die komplette Veränderung des bestehenden mentalen Modells. Im Folgenden sollen die Charakteristika der unterschiedlichen Umgangsarten dargestellt werden.

Ignoring anomalous data

Die Ignoranz anomaler Daten kann als die extremste Umgangsart beschrieben werden. Die anomalen Daten werden ignoriert, sodass keine Veränderung an dem bestehenden mentalen Modell nötig wird (Chinn & Brewer, 1993b). Es finden weder Assimilations- noch Akkommodationsprozesse statt. Die Daten werden nicht mit dem mentalen Modell in Zusammenhang gebracht. Beispiele für diese Umgangsart erkennen Chinn und Brewer (1993b) sowohl in der Wissenschaftsgeschichte als auch in psychologischen Studien. Anomale Daten werden hier nicht akzeptiert, bzw. ihnen wird keine Bedeutung zugemessen.

Rejecting anomalous data

Diese Umgangsart ähnelt dem Ignorieren anomaler Daten insofern, dass keine Theoriebearbeitung stattfindet. Auch in diesem Fall wird das ursprüngliche mentale Modell beibehalten (Chinn & Brewer, 1993b). Allerdings wird für die Ablehnung der anomalen Daten ein Grund angegeben. Dieser generiert sich jedoch nicht aus der Evaluation der Daten, sondern basiert auf drei möglichen Begründungen. Eine mögliche Begründung können (1) methodologische Fehler darstellen. Dabei erfolgt nicht zwingend eine wissenschaftliche Diskussion über die Güte von Methoden. Vielmehr werden Methoden generell in Zweifel gezogen, sodass resultierende Daten unglaubwürdig werden. Eine weitere Begründung zum Ablehnen der anomalen Daten kann durch das (2) zufällige Auftreten der Anomalien erfolgen. Dabei kann das zufällige Auftreten der Anomalien auch mit methodologischen Schwächen verknüpft werden. Die extremste Art der Ablehnung anomaler Daten kann in der Bezeichnung dieser als ‚Witz‘ erkannt werden. Hier werden Anomalien nicht auf Grundlage wissenschaftlicher Argumente abgelehnt, sondern als ‚schlechter Witz‘ herabgewürdigt. Durch diese Herabwürdigung scheint eine Auseinandersetzung mit den Anomalien nicht weiter relevant zu sein. Wenngleich der Grad der Theoriebearbeitung und des Protektionismus bzgl. des mentalen Modells vergleichbar mit dem Ignorieren anomaler Daten ist, lässt die Ablehnung anomaler Daten eine größere Eingebundenheit der Individuen in den Prozess erkennen. Die Begründung der Ablehnung anomaler Daten misst diesen gleichzeitig eine Bedeutung bei. Die Daten

werden also wahrgenommen und es findet eine (wenn auch unwissenschaftliche) Auseinandersetzung mit ihnen statt.

Excluding anomalous data

Das Exkludieren anomaler Daten verlangt keine Bewertung der Validität, bzw. Glaubwürdigkeit dieser (Chinn & Brewer, 1993b). Die Daten werden dabei als nicht zugehörig zur Theorie beschrieben. Theorien können als abstrakte Erklärungsnetze verstanden werden (Gopnik & Wellman, 1992). Diese Erklärungsnetze haben jedoch Grenzen sodass nicht alle Daten durch eine Theorie erklärbar sein können. Beim Exkludieren anomaler Daten, werden diese als außerhalb der Erklärungsreichweite der Theorie beschrieben. Die anomalen Daten müssen so nicht mit der Theorie/dem mentalen Modell in Einklang gebracht werden (Chinn & Brewer, 1993b). Diese Begründung bedarf des Versuchs einer Assimilation der anomalen Daten, welche jedoch scheitert. So können die Daten anschließend ausgeschlossen werden, ohne dass weitere Begründungen (auch bzgl. der Güte der Daten) notwendig werden. Diese Umgangsart stellt die einzige innerhalb der durch Chinn und Brewer (1993b) ursprünglich postulierten Umgangsarten dar, welche keinerlei Bewertung der anomalen Daten bedarf. Die anomalen Daten werden weder akzeptiert noch abgelehnt. Sie werden in ihrer Glaubwürdigkeit nicht bewertet.

Uncertainty about believing the data

Nach einer empirischen Testung der postulierten Taxonomie konnten Chinn und Brewer (1998) eine weitere Umgangsart ergänzen. Diese Umgangsart beschreibt die Unsicherheit bei der Bewertung der Glaubwürdigkeit der anomalen Daten. Dabei wird weder eine ablehnende noch eine akzeptierende Haltung gegenüber den anomalen Daten beobachtet. Die Proband*innen innerhalb der Studie führen vor allem methodologische Unklarheiten an. So wird z.B. fehlende Replikation, oder auch das Fehlen eigener Beobachtungen als Grund für die Unsicherheit angeführt. Die Unsicherheit bzgl. der Glaubwürdigkeit der Daten stellt neben dem Exkludieren der Daten eine weitere Umgangsart dar, in welcher keine Bewertung der anomalen Daten stattfindet. Letztlich findet bei dieser Umgangsart keine finale Entscheidung über die anomalen Daten statt. Entsprechend geschieht keine Bearbeitung der Theorie.

Holding anomalous data in abeyance

Ähnlich zur Unsicherheit bzgl. der Daten, ist die Umgangsart des ‚In-der-Schwebe-haltens‘. Hier werden die anomalen Daten nicht aus dem Erklärungsradius der Theorie ausgeschlossen, jedoch auch nicht assimiliert bzw. akkommodiert (Chinn & Brewer, 1993b). Der Theorie/dem mentalen Modell wird die Fähigkeit zugesprochen, die anomalen Daten in Zukunft erklären zu können. Diese Sichtweise beinhaltet die epistemologische Überzeugung, dass sich eine Theorie/Wissen ändern kann. Durch eine mögliche Erweiterung der Theorie wird es möglich die anomalen Daten zu erklären. Ähnlich wie bei der Unsicherheit bzgl. der anomalen Daten wird auch hier keine finale Entscheidung über die Daten getroffen. Sie werden weder in das bestehende mentale Modell assimiliert und aufgenommen, noch findet eine Ablehnung statt. Der Versuch einer Assimilation scheitert, zeigt jedoch auf, dass sich die Daten in Reichweite der Theorie befinden. Nach der *Conceptual Change* Theorie entsteht bei scheiternder Assimilation ein kognitiver Konflikt, welcher zu einem *Conceptual Change*, also zur Akkommodation führen kann (Posner et al., 1982). In diesem Fall wird eine spätere Assimilation der anomalen Daten jedoch für möglich gehalten, sodass keine Akkommodation ausgelöst wird.

Reinterpreting anomalous data

Bei der Reinterpretation anomaler Daten wird der Assimilationsprozess umgekehrt. Wird nach Definition der Assimilation die Theorie an die Daten angepasst, um ihre Erklärungsreichweite zu vergrößern, werden in diesem Fall die Daten an die bestehende Theorie angepasst (Chinn & Brewer, 1993b). Die Daten werden so reinterpretiert, dass die Theorie diese erklären kann und nicht verändert werden muss. Die Daten werden bei dieser Umgangsart als anomal akzeptiert und daraus wird eine Notwendigkeit zur Bearbeitung abgeleitet. Chinn und Brewer (1993b) beschreiben teils extreme Arten der Reinterpretation, bei welchen klar anomale Daten als Argument für das Bestehen der Theorie angeführt werden („Die Ausnahme bestätigt die Regel“).

Peripheral theory change

Die periphere Veränderung der Theorie nimmt erste Veränderungen an der bestehenden Theorie vor und unterscheidet sich damit grundlegend von allen zuvor beschriebenen Umgangsarten (Chinn & Brewer, 1993b). Die Veränderungen der Theorie/des mentalen Modells geschehen hier vor allem im *protective belt* (vgl. Lakatos, 1978). Es werden kleine Veränderungen vorgenommen, welche die Erklärung der anomalen Daten ermöglichen, die Kernannahmen der Theorie jedoch nicht in Zweifel ziehen. So kann die Ausgangstheorie bestehen bleiben und wird nur leicht verändert. Chinn und Brewer (1993b) beschreiben dabei vor allem die Bildung von Subgruppen als häufig genutzte Strategie. Hier werden anomale Daten als eine Subgruppe auftretender Daten verstanden. Solange die Subgruppen jedoch einer gleichen Hauptgruppe angehören, bedarf es keiner weitreichender Veränderungen der Theorie.

Theory change

Umfassende Veränderungen der Kernannahmen der Theorie finden im *theory change* statt. Diese Umgangsart ist vergleichbar mit einem radikalen *Conceptual Change* (Posner et al., 1982), oder einer dateninduzierten Krise der Wissenschaft (T. S. Kuhn, 1996). Die Theorie wird in ihren Kernannahmen verändert, bzw. kann durch eine neue Theorie abgelöst werden. Hier treten Akkommodationsprozesse zum Vorschein, die die anomalen Daten als Ausgangspunkt eines radikalen Wandels nutzen. Entscheidend ist, dass eine Theorie nicht durch eine neue Theorie ersetzt werden muss. Hier unterscheidet sich die Umgangsart vom radikalen Verständnis von T. S. Kuhn (1996) und Posner et al. (1982). Es kann auch zu grundlegenden Veränderungen der Kernannahmen kommen. Nach Lakatos (1978) werden Veränderungen am *hard core* vorgenommen. Damit bildet der *theory change* den Endpunkt der Taxonomie von Chinn und Brewer (1993b) mit umfassenden Veränderungen der Theorie.

Die durch Chinn und Brewer (1993b, 1998) dargestellten Umgangsarten mit anomalen Daten reichen von simplen, protektionistischen bis hin zu komplexen, theoriebearbeitenden Handlungsweisen. Dabei heben die Autoren hervor, dass es für jede Umgangsart Beispiele aus psychologischen Studien mit Schüler*innen, jedoch ebenfalls aus der Wissenschaftsgeschichte gibt. Auch Wissenschaftler*innen zeigen die unterschiedlichen Umgangsarten in ihrem Umgang mit anomalen Daten. An einigen Stellen wurde ebenfalls deutlich, dass epistemologische Überzeugungen einen Einfluss auf den Umgang mit anomalen Daten haben können (vgl. z.B. Holding anomalous data in abeyance). Sofern Wissenschaftler*innen erweiterte epistemologische Überzeugungen unterstellt werden können, mag es verwundern, dass auch hier alle (zum Teil naiven) Umgangsarten vorgefunden werden können. Auch hier könnte das Konzept der Balance von Schommer-Aikins (2002) angewendet werden. Im Sinne einer Häufigkeitsverteilung könnten auch

Wissenschaftler*innen naive Umgangsarten mit anomalen Daten praktizieren. Je nach dem spezifischen Kontext könnten so unterschiedliche Umgangsarten initiiert werden.

Neben dem spezifischen Kontext können weitere Einflussfaktoren auf den Umgang mit anomalen Daten formuliert werden. Limón (2001) beschreibt den wichtigen Einfluss des **Vorwissens** für *Conceptual Change* Prozesse. Das bestehende Vorwissen, oder auch naive Vorstellungen (vgl. Schülervorstellungen), können in mentalen Modellen organisiert sein (Gropengießer & Marohn, 2018). Diese mentalen Modelle können zur Erklärung von Phänomenen genutzt werden. Je weitreichender ein mentales Modell neue Phänomene erklären kann und je besser es dabei mit mentalen Modellen aus anderen Domänen/Disziplinen vernetzt ist, umso stabiler sind die beinhalteten Vorstellungen. Chinn und Brewer (1993b, S. 15) schreiben dazu: „More specifically, a deeply entrenched belief is one that (a) has a great deal of evidentiary support and (b) participates in a broad range of explanations in various domains“. Das hier beschriebene „entrenchement“ (Chinn & Brewer, 1993b, S. 15) einer Theorie basiert auch auf eigenen Wert-/Zielvorstellungen. Es ist wahrscheinlicher, dass anomale Daten abgelehnt werden, wenn diese der eigenen sozialen Praxis widersprechen, oder die eigene Reputation in Gefahr bringen. So könnten auch Wissenschaftler*innen anomale Daten eher ablehnen, wenn sie dadurch anerkannte Theorien beibehalten können, welche innerhalb der eigenen Peergroup selbstverständlich genutzt werden (Chinn & Brewer, 1993b). Stark vernetztes, stabiles Vorwissen könnte so einen Einfluss auf den Umgang mit anomalen Daten nehmen. Zu Vorwissen können auch epistemologische und ontologische Überzeugungen (z.B. im Bereich der *framework theories*) gezählt werden. Chinn und Brewer (1993b) heben die Bedeutung ontologischer Überzeugungen für den Umgang mit anomalen Daten hervor. So könne die Kategorisierung unterschiedlicher Entitäten eine notwendige Neukategorisierung behindern (vgl. *Conceptual Change* durch ontologische Kategorisierung, Chi, 2008). Als Beispiel könnte die Neudefinition von Raum und Zeit durch Albert Einstein angeführt werden. Die Neudefinition von Raum und Zeit als relativ und variabel hat die ontologische Kategorisierung der beiden Größen grundlegend verändert. Durch die Neudefinition sind viele neue Zusammenhänge aufkommen, welche zuvor nicht denkbar waren. Somit können ontologische Kategorisierungen entscheidenden Einfluss auf den Umgang mit anomalen Daten und einen möglichen *Conceptual Change* haben.

Im Zusammenhang mit ontologischen Überzeugungen spielen auch epistemologische Überzeugungen eine wichtige Rolle (Chinn & Brewer, 1993b). Samarapungavan (1992) konnte zeigen, dass selbst Kinder gewisse epistemologische Überzeugungen halten, welche Einfluss auf den Umgang mit Theorien und Daten haben können. Chinn und Brewer (1993b) erkennen eine besondere Bedeutung des Verständnisses des Zusammenhangs zwischen Theorien und Evidenzen (vgl. Toulmin, 2003). Erweiterte epistemologische Überzeugungen könnten so zu einem adäquaten Umgang mit anomalen Daten führen, welcher diese in die Theoriebearbeitung einbindet (vgl. Kap. 3.2). In Rückbezug auf Hofer und Pintrich (1997) können Überzeugungen aus den Dimensionen *Nature of knowledge* und *Nature of knowing* als relevant erkannt werden. Im Bereich von *Nature of knowledge* können Überzeugungen bezüglich der Vorläufigkeit von Wissen entscheidenden Einfluss auf den Umgang mit anomalen Daten nehmen. Ein Verständnis von Vorläufigkeit ermöglicht überhaupt erst die Theoriebearbeitung. Im Bereich *Nature of knowing* heben Chinn und Brewer (1993b) besonders den prozesshaften Charakter wissenschaftlicher Theoriebildung hervor. Die aktive Gestaltung von Wissen, eine notwendige Debattenkultur und multiperspektivische Betrachtungen von Phänomenen seien entscheidende epistemologische Überzeugungen, welche Einfluss auf den Umgang mit anomalen Daten nehmen könnten.

Ebenfalls zum Einfluss des Vorwissens zählen Chinn und Brewer (1993b) das Hintergrundwissen. Das Hintergrundwissen unterscheidet sich insofern vom Vorwissen im Sinne einer Theorie, dass es nicht zur jeweiligen Disziplin/Domäne gehört. Es beinhaltet z.B. das Wissen über methodologische Zugänge oder essentielle Verfahren wie die Aufnahme elektromagnetischer Spektren oder der Temperatur. Das Wissen über derlei Methoden kann die Akzeptanz anomaler Daten sowohl fördern als auch erschweren. Basieren anomale Daten auf bekannten Methoden oder akzeptierten Wissensaspekten anderer Domänen, könnten diese eher akzeptiert werden. Die Glaubwürdigkeit der Daten kann so durch die Vernetzung zum Hintergrundwissen gestärkt werden. Ebenso kann sich Hintergrundwissen auch negativ auf die Akzeptanz von anomalen Daten auswirken. So könnten alternative Erklärungen für anomale Daten aus dem Hintergrundwissen abgeleitet werden und damit zur Reinterpretation oder Ablehnung der anomalen Daten führen. Das Hintergrundwissen kann so vielfältigen Einfluss auf den Umgang mit anomalen Daten nehmen.

Neben dem Vorwissen und den zuvor beschriebenen Charakteristika anomaler Daten, spielen auch **Charakteristika der neuen Theorie** eine wichtige Rolle. Die von Chinn und Brewer (1993b) beschriebenen Charakteristika der neuen Theorie decken sich zu großen Teilen mit den durch Posner et al. (1982) postulierten Bedingungen für einen *Conceptual Change* (vgl. Kap. 3.3.1). Basierend auf dem Modell wissenschaftlicher Revolution von T. S. Kuhn (1996), muss eine alternative, plausible Theorie vorhanden sein, damit eine vorhandene Theorie verworfen wird. Im Sinne der vorherigen Ausführungen kann hier auch ein graduelles Verständnis der Theorieveränderung angewendet werden. Gibt es keine plausible, alternative Erklärung für auftretende Daten, wird die vorhandene Erklärung beibehalten. Dies gilt nach Chinn und Brewer (1993b, S. 22) sogar dann, wenn die vorhandene Theorie nicht von hoher Qualität ist: „Apparently, a bad theory was better than no theory at all [...]“. Entscheidend sei vor allem, dass ein neuer Mechanismus zur Erklärung vorhanden ist, welcher als Kern einer (neuen) Theorie verstanden werden könnte (Chinn & Brewer, 1993b).

Aus der Komplexität dieser Mechanismen, erwachsen einige Herausforderungen für den Unterricht in der Schule. Chinn und Brewer (1993b) beschreiben Schwierigkeiten beim Entdecken neuer Theorien durch die Schüler*innen. Die Ableitung der Theorien aus vorhandenen Daten kann sich als sehr komplex herausstellen, sodass es sogar Beispiele aus der Wissenschaftsgeschichte gibt, in welchen dieser Prozess nicht funktioniert hat. Die Autoren schlagen daher vor, die anerkannten Theorien in Gänze, oder in Teilen im Unterricht zu präsentieren. Hier ist eine Nähe zu *refutation texts* erkennbar, welche eine Lernendenvorstellung explizit durch wissenschaftlich anerkannte Theorien widerlegt werden (für einen Überblick: Tippet, 2010). Vor allem muss die anerkannte, neue Theorie die auftretenden Anomalien besser erklären können als die Ausgangstheorie (vgl. Posner et al., 1982). In diesem Fall kann die Präsentation/Erarbeitung der anerkannten Theorie schneller zur Akzeptanz dieser führen. Hier heben Chinn und Brewer (1993b) auf eine nötige, hohe Qualität der anerkannten Theorie ab. Verbunden mit einer guten Verständlichkeit der Theorie können so wichtige Grundsteine für eine effektive Bearbeitung des mentalen Modells erkannt werden.

Einen weiteren wichtigen Einflussfaktor auf den Umgang mit anomalen Daten erkennen Chinn und Brewer (1993b) in der **Verarbeitungstiefe**. Eine tiefgreifende Verarbeitung der anomalen Daten könne nach den Autoren durch die individuelle Eingebundenheit erreicht werden. Dazu würde sich z.B. die Kontextualisierung im Sinne eines Einbezugs von Alltagsbeispielen eignen (vgl. Chemie im Kontext, z.B. Parchmann & Kuhn, 2018; Parchmann et al., 2001). Auch eine Forderung nach der Begründung von Argumenten, könne zu einer entsprechenden Verarbeitungstiefe beitragen.

Potvin (2023) stellt in ihrem umfassenden Review empirische Studien auf Grundlage des Modells von Chinn und Brewer (1993b, 1998) dar. Die beschriebenen Umgangsarten mit anomalen Daten konnten in anderen Studien repliziert werden, wenngleich die Verteilung auf die Umgangsarten variiert. Eine Erklärung dafür sehen Kang et al. (2004) in den spezifischen Themenbereichen. So könnten unterschiedliche Umgangsarten auch durch Charakteristika des jeweiligen Themas beeinflusst werden. Potvin (2023) beschreibt ebenso Ähnlichkeiten in der Konzeption von Chinn und Brewer (1993b), C. K. K. Chan et al. (1997) und Kang et al. (2004). Als generelles Modell des Umgangs mit anomalen Daten stützt Potvin (2023) ihre Arbeit von dem *cognitive conflict process model* (CCPM) nach G. Lee et al. (2003). Innerhalb dieses Modells wird der Umgang mit anomalen Daten in drei Phasen unterteilt. In der ersten Phase sind die Lernenden in ihren Vorstellungen verhaftet, bis anomale Daten präsentiert werden. In der zweiten Phase entsteht ein Konflikt durch die anomalen Daten. Hier stellt sich die Frage, inwiefern die anomalen Daten wahrgenommen werden, und ggf. kognitive und affektive Reaktionen auslösen. Die dritte Phase stellt die Lösung des Konflikts durch Entscheidungsprozesse dar (Potvin, 2023). Auch dieses Modell wurde im Zusammenhang und im Einklang mit der durch Chinn und Brewer (1993b) postulierten und empirisch überprüften (1998) Taxonomie formuliert.

Insgesamt stellen Chinn und Brewer (1993b, 1998) eine umfassende Rahmentheorie zum Umgang mit anomalen Daten dar. Neben den unterschiedlichen Umgangsarten werden ebenfalls Einflussfaktoren dargestellt, welche im Unterricht variiert, werden könnten. Dabei sollte Wissenschaft vor allem als Bearbeiten und Erweitern von Theorien auf Grundlage von (anomalen) Daten verstanden werden (Chinn & Brewer, 1993b). Durch den sehr grundlegenden Charakter der vorgestellten Rahmentheorie, halten Chinn und Brewer (1993b) die Anwendung auch außerhalb des naturwissenschaftlichen Unterrichts für denkbar. Damit würden die Umgangsarten mit anomalen Daten zu den domänenübergreifenden Aspekten gehören, welche mit akademischen epistemologischen Überzeugungen verknüpft sind (vgl. Kap. 3.2). Im Unterricht muss der Einbezug anomaler Daten entlang der dargestellten Einflussfaktoren geplant werden, um der hohen Resistenz gegenüber anomalen Daten begegnen zu können.

Um die möglichen Ursachen für die hohe Resistenz gegenüber anomalen Daten genauer zu verstehen, unterteilen Chinn und Malhotra (2002a) den Auswertungsprozess von anomalen Daten in vier psychologische Prozesse: Beobachtung, Interpretation, Generalisierung und Retention, also das Beibehalten der neuen Idee auch über einen längeren Zeitraum. Für die Beobachtung heben Chinn und Malhotra (2002a, S. 328) die Theoriegeladenheit hervor: „When sensory stimuli are ambiguous, prior expectations exert a powerful influence on what is perceived“. Die Ausgangstheorie, oder auch bestehende Schülervorstellungen, können somit einen großen Einfluss auf die Wahrnehmung von anomalen Daten haben. Im Prozess der Interpretation können vor allem zwei Probleme auftreten: „rational nonalignment“, oder „irrational nonalignment“ (Chinn & Malhotra, 2002a, S. 328). Beim *rational nonalignment* handelt es sich um einen fehlenden Abgleich von Beobachtung und Interpretation, welcher zur Reinterpretation oder Verwerfung von Daten führen kann. Beim *irrational nonalignment* werden valide Argumentationslinien gebrochen, um die bestehende Theorie beizubehalten. Dies führt zu Umgangsarten wie dem Ignorieren von Daten, oder dem Verwerfen von Daten ohne plausiblen Grund. Der Prozess der Generalisierung fokussiert auf Verallgemeinerungen zwischen den beobachteten Phänomenen und anderen ontologischen Kategorien. Als Beispiel führen Chinn und Malhotra (2002a) ein physikalisches Experiment an, in welchem zwei unterschiedlich schwere Steine gleichzeitig fallen gelassen werden und entsprechend gleichzeitig am dem Boden einschlagen. Im Prozess der Generalisierung könnte nun eine Verallgemeinerung dieses Zusammenhangs geschehen, welche nicht mehr auf zwei Steinen, sondern (allgemeiner) auf zwei unterschiedlich schweren Gegenständen beruht. Die Generalisierung

würde entsprechend das Phänomen vom Einzelfall lösen und die Entwicklung einer umfassenden Theorie ermöglichen. Eine Herausforderung im Prozess der Retention ist das Beibehalten der neuen Erklärung/Idee über einen längeren Zeitraum. Es bestünde die Gefahr, dass die neue Erklärung für das Auftreten anomaler Daten nach wenigen Tagen nicht mehr genutzt wird, sondern die Person in ihr ursprüngliches Erklärungsschema zurückfällt (Chinn & Malhotra, 2002a). Als wichtigste Ursache für eine hohe Resistenz gegenüber anomalen Daten erkennen Chinn und Malhotra (2002a) den Prozess der Beobachtung. In ihrer durchgeführten Studie scheiterten einige Proband*innen mit inadäquaten Erwartungen an der Aufnahme objektiver Beobachtungen. Ihre Beobachtung war so sehr durch das bestehende mentale Modell beeinflusst, dass die eigentlichen Anomalien in der Beobachtung nicht mehr aufgefallen sind. Vor allem beim Umgang mit anomalen Daten spielen vorhandene Erklärungsschemata eine wichtige Rolle (Rock, 1985). Neben der Beeinflussung während der Aufnahme von Beobachtungen, können vorhandene Schemata auch die Bewertung von Anomalien als ‚noch angemessen‘ und ‚unterschiedlich von der Theorie‘ ermöglichen. Chinn und Malhotra (2002a) heben zusätzlich hervor, dass die adressierten Schemata einen großen Einfluss auf die Resistenz gegenüber anomalen Daten haben könnten. Damit nehmen die Autoren Bezug auf das *entrenchement* (Chinn & Brewer, 1993b) einer Theorie. Insgesamt erkennen die Autoren die Notwendigkeit von erweiterten epistemologischen Überzeugungen bzgl. der Beobachtung. Überzeugungen darüber, wie Beobachtungen getätigt werden sollten, gepaart mit der Offenheit für Veränderungen bestehender Wissensaspekte, könnten die Resistenz gegenüber anomalen Daten verringern (Chinn & Malhotra, 2002a). Damit wird der Kreis zu epistemologischen Überzeugungen und der Modellierung des *Conceptual Change* geschlossen. Die Aufnahme anomaler Daten als Auslöser kognitiver Konflikte (als Initiator des *Conceptual Change* Prozesses) wird durch epistemologische Überzeugungen beeinflusst. Somit zeigt sich in den dargestellten Überlegungen ein weiterer spezifischer Mechanismus, über welchen epistemologische Überzeugungen Einfluss auf einen *Conceptual Change* nehmen können (vgl. Kap. 3.2).

Während im Bereich bestehender Forschung ein Fokus auf anomale Daten insgesamt vorherrscht, liegt der Fokus dieser Arbeit auf anomalen Beobachtungen.

3.3.3.1 Schulische vs. Authentische wissenschaftliche Arbeit

Während Chinn und Malhotra (2002a) auf den Umgang mit anomalen Daten von Kindern fokussieren, schließt die von Chinn und Brewer (1993b) postulierte Taxonomie auch Forschende mit ein. In der Begründung der Taxonomie stellen die Autor*innen immer wieder Beispiele aus der Wissenschaftsgeschichte und psychologischen Studien mit Schüler*innen vor. Chinn und Malhotra (2002b) stellen klare Unterschiede zwischen authentisch-wissenschaftlicher und schulischer Praxis heraus, welche sich wiederum auf den Umgang mit anomalen Daten, aber auch den Aufbau epistemologischer Überzeugungen auswirken könnten.

Insgesamt tritt an vielen Stellen ein Unterschied in der Komplexität schulischer und authentisch-wissenschaftlicher Erkenntnisprozesse auf. Schulische Prozesse sind häufig inhaltlich weniger anspruchsvoll, wodurch vereinzelte Prozessschritte authentisch-wissenschaftlicher Erforschung ausblieben. Hier sei z.B. auf die seltener notwendige Datentransformation in schulischen Prozessen verwiesen, welche sich direkt aus der Komplexität der Experimente ableitet. Mit dem inhaltlichen Anspruchsniveau steigen teilweise auch die Anforderungen an den Erkenntnisprozess. Andererseits unterscheiden Chinn und Malhotra (2002b) nicht aktiv bzgl. des Umgangs mit anomalen Daten. Zwar steht die Reflexion der eigenen und fremden Praxis in authentisch-wissenschaftlicher Praxis deutlich mehr im Vordergrund als in schulischer Praxis, dennoch gehen die Autor*innen nicht auf Unterschiede im Umgang mit anomalen Daten im Vergleich zu nicht anomal wahrgenommenen Daten ein. Dennoch sind durch die dargestellten Unterschiede verschiedene

Umgangsarten erwartbar. Nutzen Forschende z.B. hochkomplexe instrumentelle Zugänge, um Messdaten zu erheben, geht damit ggf. auch eine höhere, wahrgenommene Fehleranfälligkeit einher, welche entsprechend reflektiert wird. So könnten anomale Daten in Zweifel gezogen werden, da Fehler im Messprozess befürchtet werden könnten. In schulischer Praxis stellen sich die Messmethoden häufig deutlich einfacher dar. Auch könnte in der eigenen, direkten Beobachtung (durch die Sinnesorgane) weniger Fehleranfälligkeit durch die Schüler*innen vermutet werden. Somit könnten sich aus den Unterschieden zwischen schulischer und authentisch-wissenschaftlicher Praxis durchaus Unterschiede im Umgang mit anomalen Daten ableiten lassen.

Chinn und Brewer (1993b) postulieren hingegen eine Taxonomie, welche Forschenden und Schüler*innen gleichermaßen gerecht werden soll. Auch die Übertragung auf andere Domänen/Disziplinen wird als möglich beschrieben. Um dies auch vor dem Hintergrund der beschriebenen Unterschiede verständlich zu machen, muss das Bild von Unterricht, welches durch Chinn und Malhotra (2002b) gezeichnet wird, kritisch hinterfragt werden. Die Autor*innen beschreiben einen instruktiven, geschlossenen und durch die Lehrkraft gelenkten Unterricht. Dieses Bild von schulischem, naturwissenschaftlichen Unterricht scheint die Abkehr von inputorientierten Curricula innerhalb der aktuellen kompetenzorientierten Rahmenbedingungen außer Acht zu lassen. Viele der hier beschriebenen Bestandteile authentisch-wissenschaftlicher Erforschung lassen sich in den Kompetenzbereichen (vor allem: Erkenntnisgewinnung) des aktuellen Kernlehrplans wieder finden (MSB NRW, 2019, 2022). Weiterhin obliegt die unterrichtliche Ausgestaltung im Klassenraum zu großen Teilen der jeweiligen Lehrkraft. Naturwissenschaftlicher Unterricht kann daher an der gleichen Schule bei zwei unterschiedlichen Lehrkräften komplett anders gestaltet werden. Entsprechend müssen die harten Kontraste, die Chinn und Malhotra (2002b) zwischen schulischer und authentisch-wissenschaftlicher Praxis beschreiben, als überzeichnet verstanden werden. An einigen Stellen sind die Unterschiede nicht von der Hand zu weisen (z.B. die ausbleibende Literaturanalyse zu spezifischen Themen), an anderen Stellen könnte es aber methodische Zugänge geben, welche auch in der Schule authentisch-wissenschaftliche Erforschung simulieren/ermöglichen können. So können Schüler*innen z.B. durch die Nutzung von Experimentierboxen in offen-konstruktiven Unterrichtssituationen selbst Kontrolle über die Planung eines Experiments übernehmen und so vor ähnliche Herausforderungen gestellt werden, wie Forschende. Entsprechend könnten ähnliche Umgangsarten mit anomalen Daten für Schüler*innen und Forschende möglich werden. In Bezug auf epistemologische Unterschiede zwischen authentisch-wissenschaftlicher und schulischer Praxis muss einschränkend erwähnt werden, dass Chinn und Malhotra (2002b) in ihrer Arbeit eher das Konstrukt *Scientific Inquiry*, oder auch *Nature of Scientific Inquiry* adressieren.

3.4 Epistemologie und Nature of science

Die Forderung von Chinn und Malhotra (2002b) nach authentisch-wissenschaftlichen Prozessen in der Schule schließt an Forschung im Bereich von *Nature of science* [NOS] und *Scientific Inquiry* an. Trotz einer knapp 100-jährigen Forschungstradition (hauptsächlich) im anglo-amerikanischen Bereich ist es bisher nicht gelungen, eine allgemein anerkannte Definition des Konstrukts NOS zu entwickeln (N. G. Lederman, 2007). Gleichzeitig stellt sich – ähnlich wie beim Konstrukt Epistemologie – die Frage, inwiefern eine einheitliche Definition möglich bzw. erstrebenswert ist. Dennoch hat NOS Einzug in verschiedene Curricula erhalten (z.B. National Research Council [NRC], 1996).

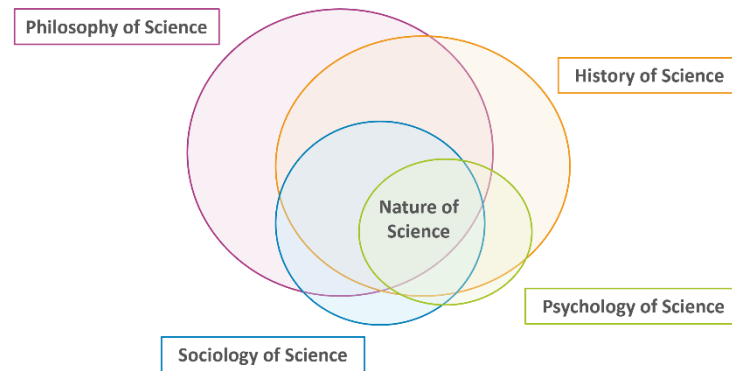


Abbildung 7 - Bestandteile von *Nature of science* nach McComas und Olson (1998/2002, S. 50). Die Größe der Kreise gibt den relativen Anteil am Konstrukt NOS an.

McComas und Olson (1998/2002) postulieren 4 Bezugsfelder, welche in ihrer Überlappung das Konstrukt NOS ergeben (Abb. 7). Diese entspringen einem Review verschiedener, relevanter Curricula. Den größten Anteil schreiben die Autor*innen dem Bereich *Philosophy of Science* zu. Innerhalb dieses Bereichs werden die unterschiedlichen Dimensionen von Epistemologie nach Hofer und Pintrich (1997) aufgeführt. Die

Epistemologie als Bestandteil von *Philosophy of Science* zeigt hier folglich einen deutlichen Einfluss auf das Konstrukt NOS auf. Ebenso zählen aber auch *History of Science*, *Psychology of Science* und *Sociology of Science* zu den Bezugsfeldern von NOS hinzu. Im Bereich von *History of Science* wird die Verknüpfung von Wissenschafts- und Menschheitsgeschichte fokussiert. Technologischer Fortschritt wird als Indikator für den Einfluss der Wissenschaft auf die Gesellschaft/Geschichte beschrieben. Im Feld der *Psychology of Science* werden charakteristische Persönlichkeitszüge von Forschenden thematisiert. So seien Forschende kreativ, offen für neue Eindrücke und müssen sich ihrer Subjektivität bewusst sein. Den letzten Bereich bildet die *Sociology of Science*. Hier wird ein Augenmerk vor allem auf ethische Entscheidungen und die kollaborative Forschungsprozesse gelegt.

Während McComas und Olson (1998/2002) vor allem auf Bezugsfelder fokussieren, welche sich in ihren Schnittmengen und darüber hinaus zu dem Konstrukt NOS zusammenfinden, postulieren andere Konzeptualisierung vor allem Bestandteile von NOS. So entstanden verschiedene, sog. ‚Aspektlisten‘, welche NOS beschreiben sollten. Je nach Forschungsgruppe differieren diese Aspektlisten in ihrer Aufteilung und ihrem Fokus. Eine viel beachtete Aspektliste stammt von N. G. Lederman (2006). Die sog. *Lederman seven* könnten daher viel Beachtung in der wissenschaftlichen Community gefunden haben, da sie die Grundlage für die VNOS-Fragebögen bilden (N. G. Lederman et al., 2002), die als Testinstrument häufig eingesetzt wurden/werden. Für die Auswahl der jeweiligen Aspekte von NOS stellt N. G. Lederman (2006, S. 304) drei Fragen:

- i. Is knowledge of the aspect of NOS accessible to students (can they learn and understand)?
- ii. Is there general consensus about the aspect of NOS?
- iii. Is it useful for all citizens to understand the aspect of NOS?

Mithilfe dieser Ausgangsfragen stellt N. G. Lederman (2006) die Praktikabilität der NOS-Aspekte in den Vordergrund. Es sei durchaus möglich, sehr lange Listen von relevanten NOS-Aspekten aufzustellen. Dies sei aber eine rein akademische Betrachtung, die die praktische Umsetzung in der Schule und die Relevanz für alle Schüler*innen außer Acht ließe. Gleichzeitig spiegle der fehlenden Konsens darüber, was NOS ist, grundlegende Überzeugungen aus dem Feld NOS wider (N. G. Lederman, 2007). So stellt NOS-Vorstellungen zur Subjektivität und der Vorläufigkeit von Wissen in das Zentrum der Überlegungen. Diese beeinflussen dabei gleichzeitig auch die Diskussionsprozesse über das Konstrukt. Somit scheint ein Konsens darüber was NOS genau beschreibt aus NOS-

Perspektive nicht möglich zu sein. Hier kann eine Verbindung zur Dimension *Certainty of knowledge* bei Hofer und Pintrich (1997) gezogen werden. Ein Verständnis der Vorläufigkeit von Wissen wirkt sich auch auf epistemologische Überzeugungen aus. Somit beeinflusst das Konstrukt sich in einer Art Zirkelschluss selbst. Gleiches kann in den Aussagen von N. G. Lederman (2007) erkannt werden. Eine mögliche Aspektliste von NOS-Bestandteilen kann nur als „solid core“ (Schommer-Aikins, 2002, S. 114) verstanden werden, welcher für Veränderungen im Sinne der Aspekte offen bleiben muss.

Tabelle 4 zeigt die um N. G. Lederman (2006) entstandene Aspektliste. Auf den ersten Blick ergeben sich hier deutliche Überschneidungspunkte mit dem Konstrukt der Epistemologie. Vor allem die Aspekte sechs und sieben zeigen einen direkten Bezug zur Epistemologie auf und sind in den Dimensionen von Hofer und Pintrich (1997) zu verorten. Die Nutzung empirischer Forschungsmethoden hat einen direkten Bezug zu den Dimensionen *Justification of knowing* und *Source of knowledge* nach Hofer und Pintrich (1997). Das Aspekt der Vorläufigkeit wissenschaftlicher Erkenntnis kann in der Dimension *Certainty of knowledge* wiedererkannt werden.

Tabelle 4 – Aspekte von *Nature of science* nach N. G. Lederman (2006)

Aspekt	Erläuterung
1 Beobachtung vs. Auswertung	Beobachtungen sind direkt durch die Sinne aufnehmbar, wohingegen Auswertungsprozesse Schlussfolgerungen oder Erklärungen aus/für die Beobachtungen erzeugen sollen.
2 Gesetz vs. Theorie	Wissenschaftliche Gesetze und Theorien sind (ontologisch) unterschiedlich und nicht ineinander überführbar. Gesetze beschreiben die Beziehung beobachtbarer Phänomene untereinander, Theorien sind abgeleitete Erklärungen für Phänomene.
3 Kreativität	Die Konstruktion wissenschaftlicher Theorien benötigt Kreativität, Vorstellungskraft und das aktive Gestalten der Forschenden.
4 Subjektivität	Wissenschaft ist geprägt durch die Subjektivität der Forschenden. Diese entsteht durch individuelle Überzeugungen, Erfahrungen, Vorwissensbestandteile und die Betrachtung aus einer spezifischen theoretischen Perspektive.
5 Kulturelle und soziale Eingebundenheit	Wissenschaft wird durch gesellschaftliche und kulturelle Gegebenheiten beeinflusst. Die Aufnahme und Genese wissenschaftlicher Erkenntnis können in verschiedenen kulturellen Zusammenhängen stark differieren.
6 Nutzung empirischer Forschungsmethoden	Wissenschaft basiert auf empirischer Evidenz und sollte, soweit es geht an dieser ausgerichtet sein.
7 Vorläufigkeit wissenschaftlicher Erkenntnis	Theorien, Gesetze und wissenschaftliche Erkenntnisse können nie bestätigt werden. Sie müssen immer als vorläufig und veränderbar verstanden werden.

Wenngleich eine Nähe zwischen den Konstrukten NOS und Epistemologie vorliegt, erkennen Neumann und Kremer (2013) auch Unterscheidungsaspekte. Einen ersten Unterscheidungsaspekt erkennen die Autorinnen in der **Disziplinspezifität**. Während epistemologische Überzeugungen im ersten Zugriff nicht auf einzelne Disziplinen fokussieren, befasst sich Forschung im Bereich NOS rein mit den Naturwissenschaften und deren experimentellen Arbeitsweise. Wie bereits dargestellt, ist die Disziplinspezifität im Bereich epistemologischer Überzeugungen nicht endgültig geklärt (vgl. Kap. 3.2). Fachbereiche außerhalb der Naturwissenschaften finden jedoch in der Betrachtung von NOS keine Beachtung. Im Feld von NOS wird über eine mögliche Disziplinspezifität innerhalb der Naturwissenschaften diskutiert (N. G. Lederman, 2007). Dadurch könnte der Geltungsbereich einer möglichen *Nature of science in chemistry* deutlich verkleinert werden. Weiterhin unterscheiden Neumann und Kremer (2013) NOS und Epistemologie auf Grundlage ihrer **Inhalte** voneinander.

In Bezug auf das von Hofer und Pintrich (1997) postulierte Modell epistemologischer Überzeugungen, erkennen Neumann und Kremer (2013) eine Überlappung der beiden Konstrukte in den Bereichen *Certainty of knowledge* (Aspekt 7) und *Justification of knowledge* (Aspekt 6). Dabei besteht diese Überlappung mit Aspekten der *Lederman seven*. Hierbei muss berücksichtigt werden, dass die *Lederman seven* zwar einen guten Ausgangspunkt zum Vergleich der Konstrukte darstellen können, jedoch selbst keinen allgemeinen Konsens im Forschungsfeld widerspiegeln. Auch hier erkennen Neumann und Kremer (2013) einen wichtigen Unterschied: Während im Bereich epistemologischer Überzeugungen durch das Modell von Hofer und Pintrich (1997) ein gewisser Konsens erreicht werden konnte, kann dieser innerhalb der NOS-Forschung nur sehr basal erkannt werden. Hier weisen die Autorinnen erneut darauf hin, dass ein NOS-Konsens evtl. gar nicht zulässig wäre, „da es eben die Diskursivität über die inhaltlichen Aspekte ist, die dieses Thema auszeichnet“ (Neumann & Kremer, 2013, S. 221).

Abseits der dargestellten Überschneidungspunkte der Konstrukte gibt es jedoch einige Unterscheidungspunkte. So fokussiere epistemologische Forschung auf die Komplexität und Quelle von Wissen, NOS-Forschung hingegen auch auf die soziale Eingebundenheit und kulturelle Einflüsse auf Forschung (Neumann & Kremer, 2013). Verbunden mit den Inhalten der beiden Konstrukte ist auch die **Perspektive**, aus welcher diese Inhalte betrachtet werden. Während es im Feld von NOS nicht um die individuelle Einschätzung des Wissens geht, sondern um eine wissenschaftsphilosophische Betrachtung, findet im Feld der Epistemologie eine individuelle Einschätzung bspw. im Bereich der Quelle von Wissen statt.

In Anlehnung an die Unterscheidung von Meta-Wissen und Meta-Kognition in Kapitel 3.1, kann NOS als **Wissensbereich**, Epistemologie hingegen als individuelle Überzeugung beschrieben werden. Verschiedene Konzeptualisierungen von NOS legen ein Verständnis als inhaltliches Konzept (vergleichbar mit anderen Konzepten wie dem Erhalt der Masse, etc.) nahe (Neumann & Kremer, 2013), welches folglich an normativen Maßstäben gemessen werden kann. In Zusammenhang mit Forschung im Bereich *Scientific inquiry* entstand aus diesem Verständnis von NOS die Forderung nach expliziter Thematisierung des Konstrukts im Unterricht (J. Lederman et al., 2019; N. G. Lederman, 2006, 2007; Schwartz et al., 2004; Vorholzer, 2016; Vorholzer & Aufschnaiter, 2019). Epistemologie kann von einem inhaltlichen Konzept unterschieden werden. Hier handelt es sich um individuelle Überzeugungen bzgl. der verschiedenen Dimensionen von Epistemologie (Neumann & Kremer, 2013). Hier spiegelt sich zugleich die Perspektive auf die beiden Konstrukte wider. Der Charakter epistemologischer Überzeugungen als individuell zeigt sich auch in weiteren Bezeichnungen des Konstrukts. Pintrich (2002) zeigt viele unterschiedliche Bezeichnungen auf, welche gemein haben, dass sie auf individuelle Überzeugungen fokussieren. Der Titel „Personal

Epistemology“ für das Sammelwerk von Hofer und Pintrich (2002, S. I) deutet auf die persönliche Bedeutung epistemologischer Überzeugungen hin. Hier ist also ein deutlicher Unterscheidungspunkt zwischen NOS und Epistemologie erkennbar.

Damit einher geht der **normative Charakter** von NOS-Aspekten im Kontrast zum deskriptiven Charakter epistemologischer Überzeugungen (Neumann & Kremer, 2013). Im Bereich NOS ist es durch das Verständnis von einem konzeptuellen Wissensbereich möglich normativ adäquate Vorstellungen festzulegen. Dieser erlaubt auch die Betrachtung von Schülervorstellungen zu NOS und Möglichkeiten, diese in adäquate Vorstellungen zu überführen. Epistemologische Überzeugungen hingegen werden bzgl. verschiedener Ausprägungsarten (vor allem in Verbindung zu entwicklungspsychologischen Modellen) beschrieben. Die Formulierung adäquater Ziel-Überzeugungen ist jedoch aufgrund konstruktinhärenter Grundsätze nicht möglich (vgl. Kap. 3.2). Damit bleibt es hier bei einer Beschreibung der individuellen epistemologischen Überzeugungen. Um diesen Unterschied zwischen den beiden Konstrukten zu berücksichtigen, schlagen Neumann und Kremer (2013) eine einheitliche sprachliche Regelung vor:

Eine mögliche Sprachregelung könnte sein, die Worte *Ansichten* und Überzeugungen (*beliefs*) im Sinne epistemologischer Überzeugungen zu nutzen. Im Abgrenzung [sic] dazu könnten inadäquate Vorstellungen (*views, ideas* oder *beliefs*) über *nature of science* als vorunterrichtliche „Alltagsvorstellungen“ bezeichnet werden – analog den Alltagsvorstellungen zu naturwissenschaftlichen Konzepten wie Evolution, Materie oder Kraft. (S. 224)

Diese sprachliche Regelung subsummiert die Unterscheidungspunkte der beiden Konstrukte. Dennoch halten Neumann und Kremer (2013) fest, dass eine trennscharfe Unterscheidung der Konstrukte nicht möglich ist. Dies kann auch durch die Abhängigkeit der Konstrukte voneinander begründet werden. So zeigen die Autorinnen auf, dass ein im Sinne des *Conceptual Change* modelliertes Lernen von NOS-Aspekten durch epistemologische Überzeugungen beeinflusst werden könnte (vgl. *framework theories*). Epistemologischen Überzeugungen, wie auch NOS-Aspekten kommt somit eine Doppelrolle zu (Duit, 1999).

Eine Unterscheidung der beiden Konstrukte kann vor allem auf Grundlage ihrer jeweiligen Forschungstradition erkannt werden. Während epistemologische Überzeugungen aus Perspektive entwicklungs- und kognitionspsychologischer Forschung untersucht wurden, stehen im Bereich NOS vor allem naturwissenschaftsdidaktische Aspekte im Vordergrund (Neumann & Kremer, 2013). Zwar wird die Entwicklung von Vorstellungen über NOS in der Forschung ebenfalls betrachtet (z.B. J. S. Lederman et al., 2021), jedoch existieren keine allgemein anerkannten Modelle mit unterschiedlichen Entwicklungsstufen. Auch scheint Forschung im Bereich NOS– nicht zuletzt durch die Verknüpfung zu vielen unterschiedlichen Bezugsdisziplinen (z.B. History of science, scientific inquiry, etc.) – weiter gefasst zu werden. Auch Vorstellungen über das Wissenschaftssystem mit sozialen und kulturellen Einflüssen werden hier genauer untersucht (N. G. Lederman, 2006). Weitere Konzeptualisierungen von NOS versuchen in holistischen Betrachtungen eine Abkehr von Konsens-Listen zu erreichen, welche als verkürzt und deklarativ kritisiert werden (z.B. *Natur of whole science*, Allchin, 2011; *family resemblance approach*, Kaya et al., 2019). Damit scheint der Umfang von NOS sehr groß und nicht klar umrissen zu sein. Ebenso spielen an vielen Stellen innerhalb der jeweiligen Konzeptualisierungen epistemologische Überzeugungen eine wichtige Rolle. N. G. Lederman (2007) beschreibt Epistemologie als eine Bezugsdisziplin für NOS. Im Sinne von *framework theories* könnten epistemologische Überzeugungen entsprechend an vielen Stellen von NOS eine wichtige Einflussvariable darstellen (Vosniadou, 1994, 2014).

Diese Arbeit fokussiert auf das Konstrukt Epistemologie als Einflussvariable auf konzeptuelles Lernen und die Bearbeitung mentaler Modelle. NOS wird dabei als Inhaltsbereich aufgefasst, welche ggf. durch einen *Conceptual Change* Prozess bearbeitet werden kann. Damit können epistemologische Überzeugungen auch Einfluss auf das Lernen über das Konzept *Nature of science* nehmen. Somit kann Epistemologie als Meta-Konzept verstanden werden, während NOS ein inhaltliches Konzept darstellt.

4. Zusammenfassung und Forschungsdesiderat

Vor dem Hintergrund bestehender Forschung besitzt das Konstrukt evidenzbasierter Praxis, wie gezeigt werden konnte, eine große Bedeutung für Bildungsprozesse. Dabei kann eine enge Verknüpfung zu den Konstrukten Epistemologie sowie *Conceptual Change* beschrieben werden (vgl. dazu Bauer & Kollar, 2023). So können epistemologische Überzeugungen grundlegenden Einfluss auf den Umgang mit Evidenzen (Chinn & Malhotra, 2002a; Hofer, 2017; Jiménez-Aleixandre & Brocos, 2017; D. Kuhn & Weinstock, 2002; Neumann & Kremer, 2013; Schommer-Aikins, 2002; Vosniadou, 1994) angehender Lehrkräfte und somit die Evidenzbasierung naturwissenschaftlichen Unterrichts nehmen. Dabei stellt das Konstrukt der evidenzbasierten Praxis die Vermittlung zwischen wissenschaftlichen Evidenzen, dem spezifischen Kontext der Situation und Erfahrungswissen in den Fokus.

Bezogen auf evidenzbasierte Praxis im Unterricht wird deutlich, wie vielfältig die Anforderungen an Lehrkräfte sind (Rothland, 2013). Diese sollen bei unterrichtlichen Entscheidungen zwischen wissenschaftlichen Evidenzen, dem spezifischen Kontext und dem persönlichen Erfahrungswissen vermitteln. Dabei sollten sie sich jederzeit ihrer epistemologischen Überzeugungen bewusst sein, welche Einfluss auf die Entscheidungen haben können. Hier kommt evidenzbasierter Praxis eine Doppelrolle zu: einerseits sollten Lehrkräfte ihren Unterricht evidenzbasiert gestalten, auf der anderen Seite sollten die Lernenden befähigt werden, selbst evidenzbasierte Entscheidungen zu treffen. Im Rahmen der vorliegenden Arbeit wird vor allem die evidenzbasierte Gestaltung von Unterricht durch angehende Lehrkräfte fokussiert. Somit ist fraglich, inwiefern angehende (Chemie-) Lehrkräfte über Kompetenzen in evidenzbasierter Praxis verfügen.

Ein weiterer Fokus innerhalb der vorliegenden Arbeit wird auf widersprechende Evidenzen im Sinne anomaler Beobachtungen innerhalb naturwissenschaftlicher Experimente gelegt. Wie beschrieben wurde, können epistemologische Überzeugungen Einfluss auf den Umgang mit anomalen Beobachtungen (Chinn & Brewer, 1993b, 1998) und ggf. resultierende *Conceptual Change* Prozesse (Chinn & Brewer, 2001; Gentner, 2001; Posner et al., 1982; Vosniadou, 1994; Vosniadou & Brewer, 1992; Vosniadou & Mason, 2012) nehmen. Dabei könnten angehende Lehrkräfte im Umgang mit anomalen Beobachtungen die Bedeutung epistemologischer Überzeugungen zur Verarbeitung von Evidenzen bewusst wahrnehmen, um diese anschließend im Unterricht vermitteln zu können. Anomale Beobachtungen können hier als Evidenzen, also Argument für oder gegen eine Aussage eingesetzt werden (Bromme et al., 2014; Toulmin, 2003). Somit kann der Umgang mit anomalen Beobachtungen als Bestandteil evidenzbasierter Praxis verstanden werden.

Für das Fach Chemie ergeben sich durch die experimentelle Natur der Disziplin Möglichkeiten, durch Experimente selbst Evidenzen im Unterricht zu erzeugen. Zwar gilt dies prinzipiell auch für andere/nicht-experimentelle Unterrichtsfächer (Haack, 2007), jedoch scheint der experimentelle Zugang im Fach Chemie Evidenzen in den Mittelpunkt von Unterricht zu stellen (MSB NRW, 2019, 2022). So ist es Aufgabe des Fachs Chemie, beobachtbare Phänomene auch auf submikroskopischer Ebene zu erklären (MSB NRW, 2022). Dazu ist es nötig, Beobachtungen und Messdaten in der Funktion von Evidenzen zur Modellbildung zu nutzen. Die Prüfung der Passung der aufgenommenen Beobachtungen aus chemischen Experimenten mit bestehenden mentalen Modellen, wie sie im *Conceptual Change* Prozess beschrieben wird, erhält so eine exponierte Rolle im Unterricht.

Die beschriebenen Anforderungen an Lehrkräfte machen einen Bedarf nach Kompetenzen in evidenzbasierter Praxis deutlich. Dawes et al. (2005) hebt dabei den großen Umfang derlei Kompetenzen hervor. Dabei müssten zur Implementation evidenzbasierter Praxis Kompetenzen in allen Bereichen dieser vorliegen. Bestehende Untersuchungen im Bereich evidenzbasierter Medizin lassen dabei

Förderbedarfe vermuten (Ashktorab et al., 2015; McCluskey & Bishop, 2009). Auch erste Untersuchungen zur Kompetenz (angehender) Lehrkräfte weisen auf einen Förderbedarf hin (Bauer & Kollar, 2023).

Innerhalb einer durch Hetmanek et al. (2015) durchgeführten Interviewstudie wird deutlich, dass Lehrkräfte Evidenzen nur selten für unterrichtliche Entscheidungen zu Rate ziehen. Eine Relevanz von (bildungswissenschaftlicher) Evidenz beschreiben die befragten Lehrkräfte vor allem in Bereichen problematischer Lernausgangslagen wie z.B. ADHS. Theoretisch begründete Annahmen vermuten keinen Unterschied zwischen (angehenden) Lehrkräften unterschiedlicher Unterrichtsfächer bzgl. der Ausprägung von Kompetenzen bzgl. evidenzbasierter Praxis (z.B. Haack, 2007; Schwandt, 2009). Stelter und Miethe (2019) erkennen auf Grundlage des Vergleichs unterschiedlicher Modulhandbücher Unterschiede in der Thematisierung forschungsmethodologischer Themen zwischen Lehramtsstudierenden und Studierenden aus erziehungs- und bildungswissenschaftlichen Studiengängen. Dabei werden in Lehramtsstudiengängen seltener forschungsmethodologische Themen adressiert als in erziehungs- und bildungswissenschaftlichen Studiengängen. Die Abstraktheit bestehender Modulbeschreibungen lässt dabei keinen detaillierten Rückschluss auf die thematisierten Methoden etc. zu (Stelter & Miethe, 2019).

Auf Grundlage der beschriebenen Kompetenzen angehender Lehrkräfte in evidenzbasierter Praxis zeigen Bauer und Kollar (2023) einen Bedarf nach Interventionen zur Stärkung der Kompetenz in evidenzbasierter Praxis auf. Bezogen auf das bildungswissenschaftliche Wissen von Lehrkräften aller Unterrichtsfächer zeigen die Autoren einige Beispiele auf. Dabei wird jedoch zumeist auf übergreifende, bildungswissenschaftliche Konstrukte fokussiert (Bauer & Kollar, 2023). Chemiedidaktische Interventionen scheinen bisher wenig im Fokus von Forschung gestanden zu haben. Im Zusammenhang mit der Forschungstradition im Feld *Nature of science* könnten Teilbereiche evidenzbasierter Praxis in Untersuchungen eingeflossen sein. Dabei kann vor allem die fehlende Differenzierung evidenzbasierter Praxis vor dem Hintergrund epistemologischer Überzeugungen sowie der holistische Fokus bestehender Untersuchungen kritisiert werden. Untersuchungen im Forschungsbereich *Nature of science* fokussieren zumeist auf Naturwissenschaften insgesamt und weniger auf einzelne Disziplinen, wie die Chemie.

Die Rolle des Fachs Chemie als empirisch-experimentelles Unterrichtsfach könnte spezifische Eigenheiten evidenzbasierter Praxis, wie z.B. die Deutung anomaler Beobachtungen aus naturwissenschaftlichen Experimenten, aufwerfen. Im Sinne der Kontextabhängigkeit von epistemologischen Überzeugungen und Argumentationsprozessen (Schommer-Aikins, 2002) sowie der Domänenspezifität dieser (Muis et al., 2006), lässt sich diese Annahme theoretisch stützen. Auch Bauer und Kollar (2023) stellen die Fachdidaktik als wichtige Bezugswissenschaft für unterrichtliche Entscheidungsprozesse heraus. Interventionen mit dem spezifischen Fokus auf evidenzbasierte Praxis vor dem Hintergrund epistemologischer Überzeugungen in fachspezifischen Domänen, wie z.B. dem Fach Chemie bilden somit ein relevantes Forschungsdesiderat, um evidenzbasierte Praxis im Chemieunterricht zu fördern.

Die Frage nach domänenspezifischen Ausprägungen evidenzbasierter Praxis steht in einer Linie mit ähnlichen Überlegungen in den Bereichen epistemologischer Überzeugungen und Vorstellungen zu *Nature of science*. Inwiefern die oben beschriebenen Charakteristika des empirisch-experimentellen Unterrichtsfachs Chemie einen Einfluss auf die Kompetenz von Lehrkräften in evidenzbasierter Praxis oder auch deren generelle Nutzung haben könnte, ist ungeklärt. Im Rahmen einer Untersuchung der Kompetenz in evidenzbasierter Praxis von Lehrkräften unterschiedlicher Fächer(-kombinationen) könnten gleichzeitig verschiedene Grenzsetzungen der einzelnen Domänen (im Sinne von einzelnen Fächern, oder Fächerbündeln ähnlicher Wissensbereiche) differenziert werden.

5. Zielstellung der Arbeit

Das Ziel dieser Arbeit ist es folglich, den Umgang mit Evidenzen angehender Lehrkräfte in chemiebezogenen Studiengängen genauer zu untersuchen. In Abgrenzung zu übergreifenden, bildungswissenschaftlichen Untersuchungen, sollen die Eigenheiten des Unterrichtsfachs Chemie aus einer fachdidaktischen Perspektive berücksichtigt werden.

Dazu sollen in einem ersten Schritt (erster empirischer Teil) bestehende Kompetenzen angehender Lehrkräfte in evidenzbasierter Praxis untersucht werden. Dabei sollen sowohl der Einfluss von Praxiserfahrungen im Praxissemesters sowie mögliche Unterschiede auf Grundlage der belegten Unterrichtsfächer betrachtet werden. Daran anschließend wird eine Intervention zur Förderung angehender Chemielehrkräfte in evidenzbasierter Praxis entwickelt und erprobt. In einem längsschnittlichen Untersuchungsdesign sollen die angehenden Chemielehrkräfte anschließend bis in den Vorbereitungsdienst in den Schulen begleitet werden. So kann erneut ein Einfluss der Schulpraxis im Sinne eines konkretisierten Kontexts untersucht werden.

Der zweite empirische Teil nimmt den Umgang mit anomalen Beobachtungen angehender Lehrkräfte im naturwissenschaftlichen Sachunterricht in den Blick. Dazu werden die Umgangsarten mit anomalen Daten nach Chinn und Brewer (1993b, 1998) zu Grunde gelegt und zu Grunde liegende epistemologische Überzeugungen analysiert. Ziel dieses Teils soll es sein, die evidenzbasierte Argumentation im Zusammenhang mit chemischen Beobachtungen sowie damit verbundene Herausforderungen zu beschreiben. Dabei wird auf die epistemologischen Begründungen zum spezifischen Umgang mit anomalen Beobachtungen fokussiert.

Durch die drei hier dargelegten Untersuchungen soll ein Beitrag zur Förderung evidenzbasierter Praxis vor dem Hintergrund epistemologischer Überzeugungen angehender Lehrkräfte in chemiebezogenen Studiengängen geleistet werden. Im Folgenden werden die drei empirischen Untersuchungen dargestellt und diskutiert. Es folgt eine abschließende, übergreifende Diskussion der Ergebnisse.

6. ERSTER EMPIRISCHER TEIL – FÖRDERUNG DER KOMPETENZEN IN EVIDENZBASIERTER PRAXIS ANGEHENDER (CHEMIE-) LEHRKRÄFTE

Wie gezeigt werden konnte, hält das Konzept der Evidenzbasierung vor dem Hintergrund epistemologischer Überzeugungen Eingang in die Forschung um Bildungsprozesse. Dabei spielt die Verknüpfung persönlicher Erfahrungen mit den besten verfügbaren Evidenzen eine entscheidende Rolle (vgl. Kap. 2.3 & 2.4). Trotz bestehender Studien vor allem im Bereich der Bildungswissenschaften, besteht ein Bedarf an Untersuchungen zur Ausprägung und möglichen Förderungsansätzen der Kompetenzen evidenzbasierter Praxis bei angehenden Lehrkräften (Bauer & Kollar, 2023). Im vorliegenden, ersten empirischen Teil dieser Arbeit sollen die Ausprägung von Kompetenzen in evidenzbasierter Praxis sowie Fördermöglichkeiten genauer untersucht werden. Dazu werden zwei Teilstudien berichtet, die sich in Bezug auf die Kontrolliertheit des jeweiligen Designs unterscheiden. Im Rahmen einer nicht-experimentellen Feldstudie wird der Einfluss des Praxissemesters auf die Entwicklung von Kompetenzen in evidenzbasierter Praxis sowie die Ausprägung dieser über verschiedene Unterrichtsfächer und Schulformen hinweg untersucht. Innerhalb einer Interventionsstudie erfolgt ein Fokus auf die gezielte Förderung der Kompetenzen evidenzbasierter Praxis angehender Chemielehrkräfte.

Auf Grundlage von Diskussionen in den Bereichen Epistemologie und *Nature of science* stellt sich die Frage, inwiefern eine fachspezifische Perspektive auf das Konstrukt evidenzbasierten Unterrichtens sinnvoll ist. Bauer und Kollar (2023) zeigen auf, dass fachdidaktische Ansätze für die Förderung von Kompetenzen evidenzbasierter Praxis im Bildungsbereich bisher eine sehr kleine Rolle gespielt haben. Vorhandene Studien zielen meist auf die Förderung im Bereich bildungswissenschaftlicher Themen ab. Untersuchungen zu Unterschieden zwischen unterschiedlichen Disziplinen deuten auf mögliche Unterschiede hin (Urhahne & Kremer, 2023). Dabei erfolgt die Unterteilung unterschiedlicher Disziplinen in größere Bündelungen entlang der vorgeschlagenen Unterscheidung im Rahmen des *hard-soft*-Kontinuums nach Biglan (1973). Es stellt sich folglich die Frage, inwiefern eine fachspezifische Betrachtung von Kompetenzen in evidenzbasierter Praxis nötig ist. Fachspezifische Besonderheiten des Konstrukts evidenzbasierter Praxis könnten sich auch in bestehenden Kompetenzen angehender Lehrkräfte widerspiegeln. Gleichzeitig könnten sich unterschiedliche Fachperspektiven als konkrete Anwendungskontexte erweisen, innerhalb welcher Kompetenzen evidenzbasierter Praxis erworben werden können. Hier könnte die Bedeutung des Kontextes für die Aktivierung epistemologischer Überzeugungen, wie auch die *conceptual ecology* im Zusammenhang mit *Conceptual Change* Prozessen angeführt werden. Die Sozialisation innerhalb der spezifischen Fachdisziplin könnte Einfluss auf den Umgang mit Evidenzen angehender Lehrkräfte nehmen. Entsprechend stellt sich folgende Forschungsfrage:

- I.a Inwiefern zeigen sich Unterschiede im Umgang mit Evidenzen bzw. den Einstellungen gegenüber Evidenzen bei Praxissemesterstudierenden verschiedener Fächergruppen im Praxissemester?

Aus der Unterscheidung zwischen Forschungsorientierung im Sinne des Umgangs mit Wissenschaft im Rahmen des Studiums und Praxisorientierung im Sinne der Verknüpfung von Theorie und Praxis (Rochnia et al., 2020), könnte auch praktische Unterrichtserfahrung einen Einfluss auf Kompetenzen evidenzbasierter Praxis angehender Lehrkräfte haben. Das Praxissemester im Master of Education (M.Ed.) Studiengang stellt ein zentrales Element der Theorie-Praxis-Verknüpfung dar. Dabei spielt die Theorie-Praxis-Verknüpfung in Sinne des forschenden Lernens eine entscheidende Rolle für das Praxissemester. So sollen Theorien einerseits dazu genutzt werden

Praxisbeobachtungen zu interpretieren, während andererseits eigene Forschungsprozesse umgesetzt werden sollen, um Fragestellungen aus der Praxis zu beantworten (Herzig & Wiethoff, 2019). Es ließe sich daher vermuten, dass das Praxissemester als Lerngelegenheit auch einen Einfluss auf die Kompetenzen in evidenzbasierter Praxis haben könnte. Daraus leitet sich die zweite Forschungsfrage dieser Teilstudie ab:

I.b Inwiefern nimmt das Praxissemester Einfluss auf die Kompetenzen im Umgang mit Evidenzen von Praxissemesterstudierenden sowie deren Einstellung gegenüber Evidenzen?

Im Rahmen der zweiten Teilstudie soll der Einfluss einer Intervention untersucht werden, welche entlang der durch Breckon und Dodson (2016) postulierten *evidence-use* Mechanismen geplant wurde⁴. Während diese Mechanismen ursprünglich zur Einflussnahme auf Entscheidungsprozesse von Politiker*innen entwickelt wurden, erfolgt im Rahmen der Intervention eine Adaption der Mechanismen im Fach Chemie. Bereits Bauer und Kollar (2023) haben auf die Notwendigkeit fachdidaktischer Perspektiven auf evidenzbasierte Praxis hingewiesen. Die *evidence-use* Mechanismen werden als Grundlage zur Förderung der Kompetenzen evidenzbasierter Praxis angehender Chemielehrkräfte angelegt. Es folgt die Frage:

II.a Inwiefern kann der Umgang mit Evidenzen von angehenden Lehrkräften durch eine an *evidence-use* Mechanismen orientierte Intervention im Lehramt Chemie gefördert werden?

Im Folgenden werden die Studiendesigns und Stichproben der beiden Teilstudien sowie die in beiden Studien eingesetzten Instrumente vorgestellt.

6.1 Studiendesign der fächerübergreifenden Erhebung im Rahmen des Praxissemesters

Die Datenerhebung der fächerübergreifenden Befragung wurde in eine bestehende Evaluation zum Praxissemester an der Universität Paderborn implementiert. Innerhalb der Evaluation stehen organisatorische Aspekte neben der Erhebung von Selbsteinschätzungen, etc. im Fokus der Untersuchung. Diese Evaluation wird seit dem Sommersemester 2015 durch die *PLAZ – Professional School of Education* an der Universität Paderborn durchgeführt und weiterentwickelt. Ziel der Umfrage ist es neben Erkenntnissen zu Organisation sowie Hürden im Praxissemester auch aktuelle Themen des Bildungsdiskurses aufzugreifen. Entsprechend wird die Evaluation immer wieder durch Aspekte, z.B. zur Digitalisierung oder Covid-Situation in Schulen, ergänzt. Die Befragung ist als Online-Fragebogen vor und nach dem schulpraktischen Teil des Praxissemester konzipiert (für weitergehende Informationen: <https://plaz.uni-paderborn.de/bildungsforschung/evaluation>). Die Evaluation wird dabei im Prä-Post-Design durchgeführt, sodass schon vor dem Praxissemester Erwartungen abgefragt werden können. Die Testinstrumente dieser Studie wurden in

⁴ Die vorliegende Teilstudie ist im Rahmen des Erasmus+ Projekts „Research in Teacher Education (RiTE)“ (Das Projekt "Research in Teacher Education" wird durch das Erasmus+ Programm unter der Nummer 2019-1-NL01-KA203-060339 vom 01.09.2019-31.12.2022 gefördert. <https://www.rite-project.eu/>) entstanden. Die vorgestellten Materialien und Daten beziehen sich auf die an der Universität Paderborn durchgeführten Studienanteile. Innerhalb des Projekts war die Universität Paderborn zuständig für die Entwicklung von Studiendesign und Instrumenten sowie die Datenauswertung und Verschriftlichung der Ergebnisse. Das Arbeitspaket der Universität Paderborn wurde durch den Autor dieser Arbeit bearbeitet.

die bestehende Erhebung aufgenommen. Die Positionierung der Instrumente direkt zu Beginn der Umfrage lässt eine inhaltliche Entkopplung zu.

Zum tieferen Verständnis der Studie muss der Erhebungskontext rund um das Praxissemester näher beschrieben werden. Das Praxissemester ist ein verpflichtendes Praxiselement der Lehramtsstudiengänge (M.Ed.) in Nordrhein-Westfalen. Dabei stellt das Praxissemester nach zwei verpflichtenden Praxiselementen im Rahmen des Bachelorstudiums (Eignungs- und Orientierungspraktikum & Berufsfeldpraktikum) die längste Praxisphase innerhalb der ersten Phase der Lehrkräftebildung dar (LABG, 2009/01.06.2023).

Innerhalb des ca. fünf-monatigen Praxissemesters steht das eigenständige Unterrichten der angehenden Lehrkräfte deutlicher im Vordergrund, als dies innerhalb der Praxisphasen im Bachelorstudium der Fall ist. So ist das eigene Unterrichten von 50 Schulstunden obligatorisch. Das Praxissemester ist insgesamt in ein vollständiges Studienjahr eingebettet. So finden im Semester vor dem Praxiselement vorbereitende Veranstaltungen an der Universität statt. Das Praxissemester kann in drei Lernorte gegliedert werden: An den Schulen verbringen die Studierenden i.d.R. vier Tage in der Woche. Hier sammeln die Studierenden, begleitet und angeleitet durch (schulische) Mentor*innen, Erfahrung im eigenverantwortlichen Unterrichten. Ebenso finden Hospitationen im Unterricht des Fachkollegiums statt. An den Zentren für schulpraktische Lehrerausbildung (ZfSL) werden Begleitveranstaltungen zu Beginn und gegen Ende des Praxissemesters angeboten. Ebenso findet eine verpflichtende Unterrichtsberatung durch eine Fachleitung im eigenen Unterricht statt. Dabei unterscheidet sich diese Unterrichtsberatung von Unterrichtsbesuchen innerhalb des Vorbereitungsdienstes darin, dass eine Leistungsbeurteilung des gehaltenen Unterrichts ausbleibt. Am Lernort Universität verbringen die Studierenden einen Tag in der Woche. An diesem sog. Studientag finden Begleitveranstaltungen in beiden Unterrichtsfächern sowie im Fach Bildungswissenschaften statt.

Weiterhin ist das wahlweise in einem Unterrichtsfach oder dem Fach Bildungswissenschaften zu belegende Begleitforschungsseminar Bestandteil des Lernorts Universität (Herzig & Wiethoff, 2019). Im Rahmen des Begleitforschungsseminars müssen die Studierenden ein Studienprojekt mit Bezug zur Schulpraxis bearbeiten. Dem Seminar liegt ein Verständnis forschenden Lernens zu Grunde, welches sich in die Bereiche des ‚forschend Lernen‘ sowie des ‚Forschen lernen‘ unterteilen lässt (Herzig & Wiethoff, 2019). Dies meint einerseits die Reflexion eigener Vorstellungen und Praxiserfahrungen vor dem Hintergrund bestehender wissenschaftlicher Theorien (forschend lernen) sowie die wissenschaftliche Bearbeitung von Fragestellungen aus der Praxis. Dabei steht im Begleitforschungsseminar die eigene Durchführung von Forschungsprozessen auf Grundlage unterrichtlicher Fragestellungen im Fokus. So sollen die Studierenden „forschungsmethodologische Kompetenzen [...] [aufbauen], die sie in die Lage versetzen sollen, einen eigenen wissenschaftlichen Erkenntnisgewinn zu erzielen“ (Herzig & Wiethoff, 2019, S. 17). Bestehende Untersuchungen weisen auf eine Wirksamkeit des Begleitforschungsseminars zum Aufbau forschungsmethodologischer Kompetenzen z.B. zur Formulierung von Forschungsfragen hin, zeigen jedoch gleichzeitig zentrale Herausforderungen des Seminars auf (Pollmeier & Fechner, 2021b). So wünschen sich die Studierenden einen deutlicheren Fokus auf die Schulpraxis und nehmen das Studienprojekt teilweise als Belastung wahr. Das Begleitforschungsseminar könnte insgesamt aufgrund seiner Ausrichtung ein besonderes Potential zur Förderung von Kompetenzen in evidenzbasierter Praxis aufzeigen.

6.2 Stichprobe der fächerübergreifenden Erhebung

Die Stichprobe der fächerübergreifenden Erhebung besteht insgesamt aus $N=411$ Proband*innen (davon $N_{\text{weiblich}}=304$, $N_{\text{männlich}}=104$, $N_{\text{divers}}=3$, $N_{\text{ohne Angabe}}=1$), die im Sommersemester 2022 an der Universität Paderborn das Praxissemester absolviert haben. Die Proband*innen sind im Durchschnitt $M=25.55$ Jahre alt ($SD=3.79$, $N=388$). Zur Einschätzung des Leistungsniveaus der Proband*innen wurde sowohl der Mittelwert der Abiturnoten $M=2.48$ ($SD=0.52$) als auch der Abschlussnote des Bachelorstudiums $M=2.16$ ($SD=0.39$) gebildet. Tabelle 5 gibt einen Überblick über die belegten Studiengänge sowie das Geschlecht der Proband*innen.

Tabelle 5 – Aufteilung der Stichprobe auf unterschiedliche Lehramtsstudiengänge inkl. Unterscheidung nach Geschlecht

Studiengang	N (w/m/d)
Lehramt an Grundschulen	104 (93/11/0)
Lehramt für sonderpädagogische Förderung	40 (32/6/2)
Lehramt an Haupt-, Real-, Sekundar- und Gesamtschulen	60 (49/16/0/+1 n.n.)
Lehramt an Gymnasien und Gesamtschulen	147 (99/47/1)
Lehramt an Berufskollegs	54 (31/23/0)

Die Stichprobe stellt die komplette Kohorte eines Praxissemester-Durchgangs am Standort Paderborn dar. Dabei verteilen sich die Proband*innen über das komplette Spektrum der Unterrichtsfächer (siehe Tab. 6). Neben der Verteilung auf Unterrichtsfächer ist aus Tabelle 6 ebenfalls die Sortierung der Unterrichtsfächer in Bündel zu erkennen. Für die spätere Auswertung und mögliche Gruppenvergleiche werden die Unterrichtsfächer in Bündel zusammengefasst. Innerhalb dieser Bündel sind verwandte Fächer in Anlehnung an die APO-GOST, 1998/23.03.2022 (1998) (§7) zusammengefasst. Neben den „klassischen“ Unterrichtsfächern beschreiben die *Lernbereiche* inhaltliche Ausrichtungen innerhalb des Lehramts an Grundschule. *Berufliche Fachrichtungen* stellen Ausrichtungen innerhalb des Lehramts an Berufskollegs dar, *Förderschwerpunkte* können im Lehramt für sonderpädagogische Förderung (angegliedert an die Lehramtsausbildung im Bereich der Primarstufe) belegt werden. Einzig der *Lernbereich* „Natur und Gesellschaftswissenschaften“ stellt ein eigenständiges Bündel dar. Dieser *Lernbereich* verbindet naturwissenschaftliche mit gesellschaftswissenschaftlichen Inhalten. Es ist entsprechend nicht möglich dieses Unterrichtsfach einem der bestehenden Bündel zuzuordnen. Folglich wurde der *Lernbereich* als eigenes Bündel konzipiert.

Die dargestellten Belegungszahlen für die einzelnen Unterrichtsfächer weichen aufgrund der Anzahl möglicher Unterrichtsfächer von der Stichprobengröße ($N=411$) ab. Tabelle 6 gibt entsprechend eine Übersicht über die Belegungszahlen für die einzelnen Unterrichtsfächer sortiert in Fächerbündel. Es müssen jeweils mindestens zwei, für das Lehramt an Grundschulen und für sonderpädagogische Förderung drei Unterrichtsfächer (oder je ein Förderschwerpunkt), gewählt werden. Auch innerhalb der Studiengänge für die Sekundarstufe kann ein drittes Unterrichtsfach im Erweiterungsstudium belegt werden. Um die Fälle jeweils nur einmal in die Analyse einzubeziehen, wurden diese auf Grundlage der jeweiligen Unterrichtsbündel und der Kombinationen aus

diesen kategorisiert. So wurden die belegten Unterrichtsfächer in einem ersten Schritt den unterschiedlichen Unterrichtsfächer-Bündeln zugeordnet.

Tabelle 6 – Belegungen der Unterrichtsfächer innerhalb der Unterrichtsfächer-Bündel für die Stichprobe

Sprachlich-literarisch-künstlerisches Bündel	Mathematisch-naturwissenschaftlich-technisches Bündel	Gesellschaftswissenschaftliches Bündel	Religiös-philosophisches Bündel	Sachunterricht
N=377	N=252	N=196	N=94	N=74
Lernbereich: sprachliche Grundbildung (N=118)	Lernbereich: mathematische Grundbildung (N=122)	Geschichte (N=52)	Philosophie/praktische Philosophie (N=37)	Lernbereich: Natur- und Gesellschaftswissenschaften (Sachunterricht) (N=72)
Deutsch (N=109)	Sport (N=42)	Förderschwerpunkt: Lernen (N=38)	Religionslehre, katholisch (N=35)	
Englisch (N=96)	Mathematik (N=37)	Förderschwerpunkt: Emotionale und soziale Entwicklung (N=38)	Religionslehre, evangelisch (N=22)	
Kunst (N=20)	Hauswirtschaft (N=15)	Pädagogik (N=34)		
Musik (N=10)	Ernährungslehre (N=8)	Berufliche Fachrichtung: Wirtschaftswissenschaften (N=19)		
Spanisch (N=10)	Chemie (N=7)	Wirtschaftswissenschaften (N=12)		
Französisch (N=9)	Physik (N=6)	Wirtschaftspädagogik (N=3)		
Textilgestaltung (N=5)	Berufliche Fachrichtung: ernährungs- und Hauswirtschaftswissenschaften (N=6)			
	Berufliche Fachrichtung: Elektrotechnik (N=4)			
	Berufliche Fachrichtung: Maschinenbautechnik (N=3)			
	Informatik (N=2)			

Anschließend wurde die Kombination der jeweiligen Bündel für die Proband*innen bestimmt. So konnten Kombinationen aus bis zu drei Unterrichtsfächern beschrieben werden. Innerhalb der Bündel-Kombination gehen die Fälle dann jeweils nur einmal in die Analyse ein. Dabei wurden insgesamt 15 Proband*innen aufgrund fehlender Angaben bzw. der Angabe eines sonstigen, nicht zuordenbaren Unterrichtsfachs keiner Bündelkombination zugeteilt. Entsprechend stellt die

Zuordnung zu Fächerbündeln nur einen Umformungsschritt für die Kategorisierung der Proband*innen dar. Während Tabelle 6 die Anzahl der Belegungen für die einzelnen Unterrichtsfächer sortiert in Fächerbündel darstellt, kann in Tabelle 7 die Verteilung der Fälle auf unterschiedliche Unterrichtsbündelkombinationen erkannt werden.

Tabelle 7 – Anzahl der Fälle pro Fächerbündelkombination

Fächerbündel 1	Fächerbündel 2	Fächerbündel 3	<i>N</i> _{Fächerbündelkombination}
Sprachlich-literarisch-künstlerisch	Mathematisch-naturwissenschaftlich-technisch	-	73
Sprachlich-literarisch-künstlerisch	Gesellschaftswissenschaftlich-technisch	-	75
Sprachlich-literarisch-künstlerisch	Religiös-philosophisch	-	50
Mathematisch-naturwissenschaftlich-technisch	Gesellschaftswissenschaftlich	-	18
Mathematisch-naturwissenschaftlich-technisch	Religiös-philosophisch	-	6
Mathematisch-naturwissenschaftlich-technisch	Sachunterricht	-	1
Gesellschaftswissenschaftlich	Religiös-philosophisch	-	16
Sprachlich-literarisch-künstlerisch	Sprachlich-literarisch-künstlerisch	-	31
Mathematisch-naturwissenschaftlich-technisch	Mathematisch-naturwissenschaftlich-technisch	-	22
Gesellschaftswissenschaftlich	Gesellschaftswissenschaftlich	-	7
Sprachlich-literarisch-künstlerisch	Mathematisch-naturwissenschaftlich-technisch	Gesellschaftswissenschaftlich	4
Sprachlich-literarisch-künstlerisch	Mathematisch-naturwissenschaftlich-technisch	Religiös-philosophisch	12
Sprachlich-literarisch-künstlerisch	Mathematisch-naturwissenschaftlich-technisch	Sachunterricht	55
Sprachlich-literarisch-künstlerisch	Gesellschaftswissenschaftlich	Religiös-philosophisch	5
Sprachlich-literarisch-künstlerisch	Gesellschaftswissenschaftlich	Sachunterricht	7
Mathematisch-naturwissenschaftlich-technisch	Gesellschaftswissenschaftlich	Religiös-philosophisch	5
Mathematisch-naturwissenschaftlich-technisch	Gesellschaftswissenschaftlich	Sachunterricht	9

Weiterhin können die Proband*innen auf Grundlage der gewählten Begleitforschungsseminare kategorisiert werden. Tabelle 8 gibt einen Überblick über die belegten Begleitforschungsseminare sowie deren Zuordnung zu den Bündeln. Im Vergleich zur Einteilung auf Grundlage der studierten Unterrichtsfächer, kann im gesellschaftswissenschaftlichen Bündel ein weiteres Fach erkannt

werden. So kann das Begleitforschungsseminar auch im Anteilsfach Bildungswissenschaften belegt werden, welches jedoch nicht als Unterrichtsfach studiert werden kann.

Tabelle 8 – Aufteilung der Stichprobe auf die unterschiedlichen Begleitforschungsseminare

Sprachlich-literarisch-künstlerisches Bündel	Mathematisch-naturwissenschaftlich-technisches Bündel	Gesellschaftswissenschaftliches Bündel	Religiös-philosophisches Bündel	Sachunterricht
N=70	N=70	N=142	N=46	N=18
Lernbereich: sprachliche Grundbildung (N=14)	Lernbereich: mathematische Grundbildung (N=28)	Geschichte (N=24)	Philosophie/praktische Philosophie (N=14)	Lernbereich: Natur- und Gesellschaftswissenschaften (Sachunterricht) (N=18)
Deutsch (N=13)	Sport (N=11)	Förderschwerpunkt: Lernen (N=2)	Religionslehre, katholisch (N=17)	
Englisch (N=26)	Mathematik (N=6)	Förderschwerpunkt: Emotionale und soziale Entwicklung (N=7)	Religionslehre, evangelisch (N=15)	
Kunst (N=4)	Hauswirtschaft (N=9)	Pädagogik (N=28)		
Musik (N=7)	Ernährungslehre (N=0)	Berufliche Fachrichtung: Wirtschaftswissenschaften (N=4)		
Spanisch (N=3)	Chemie (N=6)	Wirtschaftswissenschaften (N=1)		
Französisch (N=2)	Physik (N=2)	Wirtschaftspädagogik (N=3)		
Textilgestaltung (N=1)	Berufliche Fachrichtung: ernährungs- und Hauswirtschaftswissenschaften (N=4)	Bildungswissenschaften (N=73)		
	Berufliche Fachrichtung: Elektrotechnik (N=2)			
	Berufliche Fachrichtung: Maschinenbautechnik (N=1)			
	Informatik (N=1)			

6.3 Studiendesign der Interventionsstudie zur Förderung von Kompetenzen angehender Chemielehrkräfte in evidenzbasierter Praxis

Im Rahmen der Interventionsstudie soll die Förderung von Kompetenzen in evidenzbasierter Praxis angehender Chemielehrkräfte durch eine an *evidence-use* Mechanismen orientierte Intervention untersucht werden. Um Einblicke in den Einfluss der Intervention auf die Kompetenzen im Umgang mit Evidenzen der Studierenden zu erhalten, wurde ein längsschnittliches Mixed-Methods Design eingesetzt. Hierzu wurden Studierende im M.Ed.-Studiengang mit dem Unterrichtsfach Chemie im Rahmen der Seminare „Chemische Bildung und chemiedidaktische Konzeptionen“ sowie „Begleitforschungsseminar Praxissemester Chemie“ vor und nach der Intervention zu ihren Einstellungen gegenüber Evidenzen sowie der Kompetenz im Umgang mit diesen befragt. Ein kleiner Teil der Stichprobe wurde nach Abschluss des Studiums bis in den Vorbereitungsdienst begleitet.

Die Implementation der Studie innerhalb von zwei Seminaren erfolgte zur Stichprobenvergrößerung. Aufgrund der geringen Studierendenzahlen im M.Ed. Chemie würde die alleinige Implementation der Intervention in eines der Seminare nur eine sehr kleine Stichprobengröße zulassen. Entsprechend wurden Seminare mit einem ähnlichen Potential zur Implementation der Intervention ausgewählt. Bei beiden Seminaren handelt es sich um Seminare im bzw. nach dem Praxissemester. Somit können unterrichtspraktische Erfahrungen eingebracht werden, welche essentieller Bestandteil einer evidenzbasierten Praxis darstellen. Ebenso fokussieren beide Seminare auf Projektarbeiten mit Bezug zu wissenschaftlichen Evidenzen mit einer hohen Selbstständigkeit in der Planung und Durchführung eines individuellen Projekts. So können die Recherche und der Einbezug von Evidenzen in die individuellen Projekte als Vergleichsmerkmal der Seminare herangezogen werden. Im Folgenden werden die Charakteristika der beiden Seminar genauer vorgestellt.

Seminar „Chemische Bildung und chemiedidaktische Konzeptionen“

Ein Teil der Studierenden stammt aus dem Seminar „Chemische Bildung und chemiedidaktische Konzeptionen“. In diesem Seminar entwickeln die Studierenden in Gruppenarbeit einen Projekttag für Schüler*innen. Der Projekttag soll thematisch ein Thema der Nachhaltigkeit aufgreifen. Die Themenwahl ist dabei freigestellt. Oft gewählte Themenbereiche sind z.B. Elektromobilität (Lithium-Ionen-Akkumulator vs. Wasserstoff-Brennstoff-Zelle) oder auch Alternativen zu herkömmlichen Kunststoffen (biologisch-abbaubare Kunststoffe). Neben der fachlichen Klärung der Themen recherchieren die Studierenden ebenso mögliche Schülervorstellungen. Zusätzlich zur Recherche der Schülervorstellungen, werden diese in Interviews mit Schüler*innen aus den jeweiligen Schulklassen erhoben. Im Sinne der didaktischen Rekonstruktion (Kattmann, 2007) soll vor dem Hintergrund der fachlichen Klärung der Themen sowie der vorhandenen Schülervorstellungen die didaktische Strukturierung des Projekttagges erfolgen. Das Seminar ist nach dem Praxissemester vorgesehen, sodass die Studierenden auf erste Unterrichtserfahrung zurückgreifen können. Innerhalb des Seminars besteht ein Fokus auf Evidenzen, welche aus der Literatur abgeleitet werden können oder selbst generiert werden müssen. Das Konzept der didaktischen Rekonstruktion zeigt hier relevante Bereiche auf, in welchen Evidenzen zur Unterrichtsplanung notwendig sind. So kann das Seminar einen Beitrag zur Theorie-Praxis-Verknüpfung leisten (Pollmeier & Fechner, 2021b). Insofern ergibt sich hier ein Potential zur Förderung von Kompetenzen evidenzbasierter Praxis. Im Rahmen der Intervention sollen Aspekte expliziert werden, welche auf den Umgang mit Evidenzen abzielen. So könnte der bereits bestehende Fokus auf die evidenzbasierte Entwicklung eines Projekttagges für die Schüler*innen ausgeweitet werden. Dabei könnte vor allem die

Explication evidenzbasierter Tätigkeiten sowie die Reflexion der Prozesse einen Beitrag zum Aufbau von Kompetenzen in evidenzbasierter Praxis bei den Proband*innen leisten.

Seminar „Begleitforschungsseminar Praxissemester Chemie“

Ein zweiter Teil der Studierenden wurde aus dem Seminar „Begleitforschungsseminar Praxissemester Chemie“ rekrutiert. Im Rahmen des Begleitforschungsseminars sollen die Studierenden „im Sinne des Forschenden Lernens einer begrenzten spezifischen unterrichtlichen oder außerunterrichtlichen Fragestellung nachgehen, indem sie ihr Erkenntnisinteresse formulieren, Theoriebezüge aufarbeiten, [...], Daten erheben und in Bezug auf die Forschungsfrage auswerten, interpretieren und dokumentieren.“ (PLAZ, 2017, S. 6). Am Ende des Semesters muss das selbst durchgeführte Forschungsprojekt auf einem Forschungsposter dargestellt und verteidigt werden. Dieses Seminar ist verpflichtender Bestandteil des Praxissemesters im M.Ed.-Studiengang. Das Seminar teilt sich in drei Sitzungen à 4 Stunden im Laufe des Semesters auf. Auch hier ist ein Bezug zum Umgang mit Evidenzen zu erkennen. Die Studierenden müssen in dem Seminar Evidenzen zu ihren jeweiligen Fragestellungen selbst generieren. Dazu muss vorab eine Recherche bestehender Literatur erfolgen. Dabei spielt die Passung der Evidenzen in Bezug auf die gestellte Frage eine wichtige Rolle (vgl. *evidence-use Mechanismus Agree*, Breckon & Dodson, 2016). Auch für dieses Seminar sind direkte Anknüpfungspunkte zur Förderung der Kompetenz in evidenzbasierter Praxis zu erkennen. Der Forschungsprozess ist entsprechend bereits Bestandteil des Seminars und könnte durch Explication und Reflexion weiter fokussiert werden.

Auf Grundlage der Ähnlichkeiten werden die beiden vorgestellten Seminare im Rahmen dieser Studie gemeinsam betrachtet. Dies kann vor allem auf dem instruktional sehr ähnlichen Umgang mit Evidenzen innerhalb der Seminare begründet werden. Evidenzen wurden hier selbst recherchiert und generiert. Dadurch ergaben sich ähnliche Herausforderungen für die Studierenden der beiden Seminare. So mussten Qualitätskriterien zur Bewertung bestehender Evidenzen angelegt werden, während ebenfalls ein Fokus auf die Güte einer eigenen Datenerhebung gelegt werden sollte. Die Evidenzen mussten dann für die Gestaltung des eigenen Projekts verarbeitet werden. Hier ergab sich ein kleiner Unterschied in der Zielsetzung der Projekte. Während die Studierenden im Seminar „Chemische Bildung und chemiedidaktische Konzeptionen“ einen Projekttag für Schüler*innen entwickeln, sollte im Rahmen des „Begleitforschungsseminars Praxissemester Chemie“ eine selbst formulierte Forschungsfrage beantwortet werden. Entsprechend lag der Fokus in erstgenanntem Seminar eher auf der Unterrichtsplanung, während im Begleitforschungsseminar eher die wissenschaftlich orientierte Beantwortung einer Fragestellung fokussiert wurde. Auf Grundlage der dafür notwendigen Prozesse und des klaren Fokus der Begleitforschungsarbeiten auf den Chemieunterricht, können jedoch deutliche Ähnlichkeiten erkannt werden. Entsprechend erfolgt im Folgenden eine gemeinsame Betrachtung der beiden Seminare.

Abbildung 8 gibt einen Überblick über die zeitliche Gestaltung der Intervention und die gewählten Erhebungszeitpunkte. Ziel der Datenerhebung war es, Veränderungen in den Kompetenzen evidenzbasierter Praxis und der Einstellung gegenüber diesen zu untersuchen.

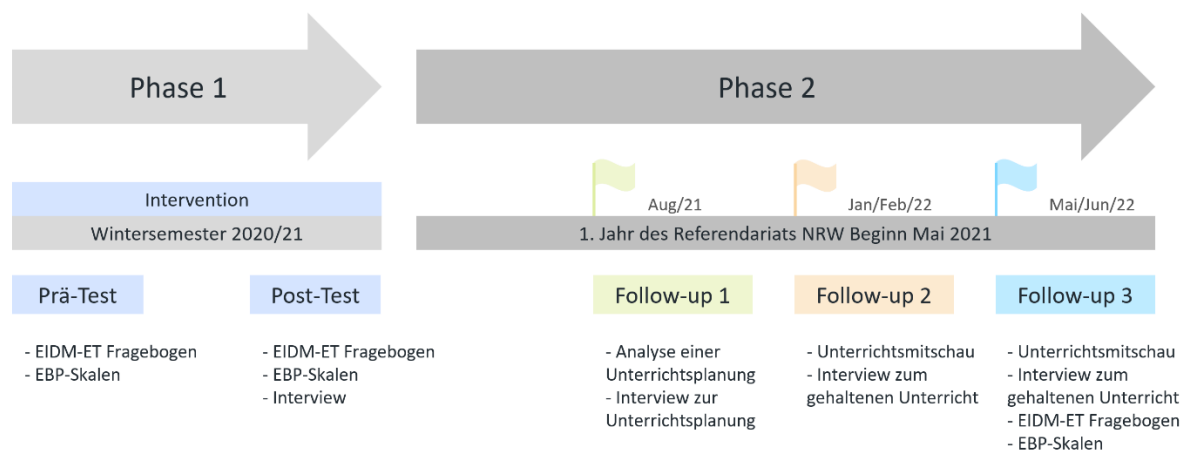


Abbildung 8 - Studiendesign der Interventionsstudie

In der ersten Phase fand die Intervention begleitet von einem Prä-Post-Test statt. Vor Beginn und nach Abschluss der Intervention wurden die gleichen Instrumente an den jeweiligen Testzeitpunkten eingesetzt. Um einen Einblick in die Kompetenzen evidenzbasierter Praxis der Proband*innen zu erhalten, wurde der *evidence-informed decision making in education test* (EIDM-ET) eingesetzt (vgl. Kap. 6.5.1). Dieser Fragebogen wurde sowohl im Prä- als auch im Posttest sowie dem letzten follow-up-Zeitpunkt eingesetzt. Zusätzlich wurden an diesen Zeitpunkt zwei *evidence-based practice scales* eingesetzt (vgl. Kap. 6.5.2), welche sich mit den Einstellungen zu evidenzbasierter Praxis, wie auch deren Implikation in der eigenen Praxis auseinandersetzen. Das zum Post-Zeitpunkt und im Rahmen des follow-up-Zeitpunktes eingesetzte Interview (vgl. Kap. 6.5.3) sollte vertiefte Einblicke in das Verständnis evidenzbasierter Praxis der Proband*innen ermöglichen. Als Grundlage für die Interviews im Rahmen der 2. Phase wurden Unterrichtsplanungen oder eine Unterrichtsmitschau eingesetzt (vgl. Kap. 6.5.4).

In der zweiten Phase wurde ein Teil der Gesamtstichprobe im ersten Jahr des Referendariats (Vorbereitungsdienst auf das Lehramt in Nordrhein-Westfalen) begleitet. Zu drei Zeitpunkten fand eine follow-up-Erhebung statt. Diese bestand jeweils aus einem Interview, welches entweder auf Grundlage einer geplanten und bereits durchgeführten Unterrichtsstunde oder einer Unterrichtsmitschau aufbaute. Die einzelnen Termine der Follow-up Interviews sollten dabei auf die Quartale 2-4 des ersten Jahrs des Referendariats verteilt werden. Bedingt durch Ferienzeiten und die Umsetzungsmöglichkeiten an den einzelnen Schulen bzw. der Lehramtsanwärter*innen kam es hier zu Verschiebungen.

Einschränkungen durch die Corona-Pandemie

Aufgrund der Corona-Pandemie mussten Einschränkungen an verschiedenen Stellen der vorliegenden Studie in Kauf genommen werden.

So konnten die innerhalb des Seminars „Chemische Bildung und chemiedidaktische Konzeptionen“ geplanten Schulbesuche zur Durchführung der Projekte nicht wie geplant realisiert werden. Anstatt dessen wurden die Projekttag in digitale Webquests überführt, welche anschließend synchron oder asynchron mit den Schulklassen erprobt wurden.

Auch im Rahmen des „Begleitforschungsseminars Praxissemester Chemie“ kam es zu Einschränkungen durch die Corona-Pandemie. Drei Monate nach dem Start des Praxissemesters wurde es den Schüler*innen der Jahrgangsstufen 1-7 in Nordrhein-Westfalen freigestellt, am Präsenzunterricht teilzunehmen. Schüler*innen der Jahrgangsstufen 8-13 wurden ab dem 14.12.2022

vollständig im Distanzunterricht beschult (Richter, 2020). Zuvor wurden einzelne Schulen bei einer besonders hohen Fallzahl-Inzidenz innerhalb der Kommunen geschlossen. Somit war die Planbarkeit des Praxissemesters und damit auch der Begleitforschungsarbeit deutlich beeinträchtigt. Den Studierenden dieses Durchgangs wurde daher freigestellt, auch Dokumentenanalysen etc. durchzuführen, sodass sie nicht auf das Gelingen einer Erhebung mit Schüler*innen angewiesen waren.

Einschränkungen lagen auch für die Datenerhebung der vorliegenden Studie vor. So konnten einige Besuche zur Unterrichtsmitschau der Phase 2 nicht realisiert werden, da an den jeweiligen Schulen Betretungsverbote für Externe gegolten haben. In diesen Fällen wurde das Interview erneut auf Basis einer Unterrichtsplanung digital durchgeführt.

Auch die Intervention wurde im Verlauf immer wieder an die sich ändernden Rahmenbedingungen der Corona-Pandemie angepasst. Hier gab es Auswirkungen vor allem auf die jeweiligen Durchführungs- und Prüfungsformen der Projekte. Wenngleich das Seminar komplett digital stattgefunden hat, waren in beiden Seminaren Präsenztermine zur Datenerhebung oder Erprobung chemischer Experimente angedacht. Aufgrund der dynamischen Situation mussten die Studierenden ihre Projekte teilweise umplanen oder hatten mit organisatorischen Hürden zu kämpfen. In Bezug auf die sich verändernde Situation mussten Anpassung in den Prozessen der Proband*innen vorgenommen werden. Hier kann eine Nähe zum Mechanismus *Structure and Processes* erkannt werden (Breckon & Dodson, 2016). Die Strukturen innerhalb derer die Proband*innen gearbeitet haben, unterlagen stetigen Veränderungen. So mussten auch Prozesse angepasst werden. Dies könnte Einfluss auf den Umgang mit Evidenzen nehmen. Ebenso kann ein Einfluss evidenzbasierter Diskussionen um COVID-19 innerhalb der Gesellschaft auf die Kompetenzen im Umgang mit Evidenzen der angehenden Lehrkräfte nicht ausgeschlossen werden. Auch im Rahmen der Datenerhebung kam es zu Einschränkungen durch die COVID-19 Pandemie.

Gestaltung der Intervention

Die Intervention wurde entlang der *evidence-use* Mechanismen nach Breckon und Dodson (2016) geplant. Dazu wurden als Schwerpunkte die vier Mechanismen (2) *Agree*, (3) *Access and Communication*, (5) *Skills* und (6) *Structures and processes* ausgewählt (Pollmeier & Fechner, 2021a).

Im Mechanismus *Agree* wird z.B. auf die Frage „Warum sind Evidenzen wichtig?“ fokussiert. Die Studierenden sollen hier Argumente aus ihrer vorherigen Arbeit mit Evidenzen zusammentragen, welche auf die Bedeutung der Evidenzbasierung eingehen. Dabei wird deutlich, dass ggf. weitere Mechanismen durch diese Frage angesprochen werden könnten. So könnten auch Aspekte des Mechanismus *Awareness* angesprochen werden. Dieser Mechanismus hebt vor allem auf eine wertschätzende Haltung gegenüber Evidenzen ab. Hier wird der interdependente Charakter der durch Breckon und Dodson (2016) postulierten Mechanismen deutlich (vgl. Kap. 2.4). Wenngleich auf diese Weise anteilig auch die weiteren Mechanismen thematisiert werden, bilden die oben benannten Mechanismen die Schwerpunkte für die inhaltliche Arbeit innerhalb der Intervention.

Zeitlicher Ablauf der Intervention

Beide beschriebenen Seminare unterscheiden sich in ihrer zeitlichen Gestaltung (vgl. Abb. 9). Bzgl. der genutzten Materialien und Aufgaben in der Intervention sind jedoch keine Unterschiede zu erkennen. Durch die unterschiedlichen Seminarkontexte können Unterschiede in der Diskussion der Inhalte nicht ausgeschlossen werden. Dabei können aber vor allem deutliche Ähnlichkeiten in der Herangehensweise der beiden Seminare erkannt werden (s.o.).

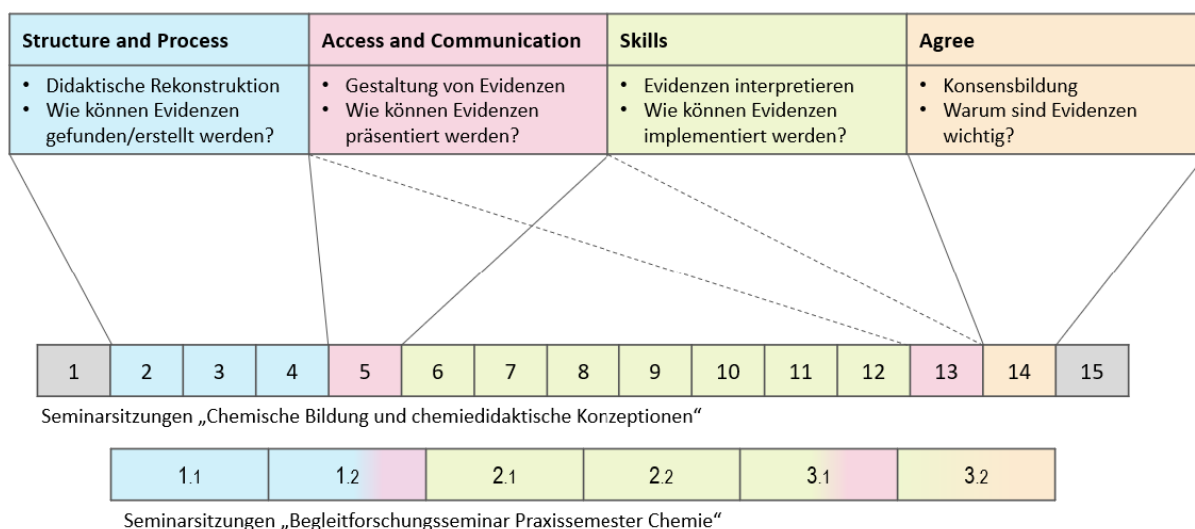


Abbildung 9 - Vergleich der Zeitgestaltung der Intervention für beide Seminare. Das „Begleitforschungsseminar Praxissemester Chemie“ findet an nur drei Terminen im Semester statt. Die Sitzungen umfassen je 3,5 Stunden.

Structure and Process

Zu Beginn der Intervention wurden Aspekte des Mechanismus *Structure and Process* fokussiert. Dieser Mechanismus fokussiert auf Strukturen, die bei der Nutzung von Evidenzen nützlich wirken können. So stellt z.B. das Auffinden von Evidenzen innerhalb von Datenbanken einen wichtigen Aspekt dar. Das Ziel dieser Phase ist es, dass die Proband*innen unterschiedliche Arten von Evidenzen und deren Einfluss auf Entscheidungsprozesse kennenlernen.

Dazu sollten die Studierenden als Einstieg eine eigene Definition für Evidenzen erarbeiten (Aufgabe SP1⁵, siehe Anhang). Anschließend wurden die Studierenden mit einer Studie zur Wirksamkeit von Chloroquin bei einer COVID-Infektion konfrontiert (Gautret et al., 2020). Im Frühjahr 2020 machte der damalige US-Präsident Donald Trump Schlagzeilen, als er das Malaria-Medikament Hydroxychloroquin zur Prophylaxe einer COVID-Infektion einnahm (Deutsches Ärzteblatt, 2020). Der Einsatz des Medikaments wurde in der Öffentlichkeit und der Fachwelt diskutiert. Die Studienlage war nicht eindeutig (Feldt et al., 2020), was zur Verunsicherung bzgl. des Einsatzes des Medikaments führte. Dieser aktuelle Kontext wurde für die Intervention aufgegriffen. Die Studierenden sollten zu Beginn überlegen, welche Evidenzen vorliegen müssten, um die Wirksamkeit eines Medikaments bewerten zu können (SP2). Anschließend wurde ein Ausschnitt aus dem NDR-Podcast⁶ mit Prof. Dr. Christian Drosten angehört. Dieser hat auf methodische Schwächen der Studie von Gautret et al. (2020) hingewiesen. Ebenso wurden in dem Podcast relevante Evidenzen zur Bewertung der Wirksamkeit dargelegt. An diesem Beispiel kann der Einfluss von Evidenzen auf (institutionelle) Entscheidungsfindungsprozesse diskutiert werden. Dabei werden hier vor allem die Prozesse der Entscheidungsfindung vor dem Hintergrund bestehender institutioneller Strukturen thematisiert. Nachdem anschließend unterschiedliche Typen von Evidenzen verglichen wurden (z.B. anekdotische Evidenz, Fallberichte, Kohorten-Studien, etc.), sollten die Studierenden Evidenz-Typen für naturwissenschaftsdidaktische Forschung ableiten (SP3). Somit sollte ein Transfer der Strukturen und Prozesse aus dem Beispiel auf den Bildungsbereich stattfinden. Dabei sollten die Evidenz-Typen bzgl. ihrer Stärken und Schwächen bewertet werden. Die Verknüpfung zu

⁵ Die Aufgaben der jeweiligen *evidence-use* Mechanismen finden sich im Anhang. Die Aufgaben werden wie folgt abgekürzt: *Structure and Processes*: SP, *Access and Communication*: AC, *Skills*: S, *Agree*: AG

⁶ <https://www.ndr.de/nachrichten/info/17-Malaria-Medikament-vorerst-kein-Hoffnungstraeger,audio655730.html> (20.06.2022)

medizinischen Evidenzen sollte das ggf. unterschiedliche Verständnis, bzw. die unterschiedlich wahrgenommene Relevanz von Evidenzen in Medizin und Bildung aufgreifen. Wenngleich sich die letztendlichen Studienarten der beiden Bereichen unterscheiden, können doch grundlegende Ähnlichkeiten und eine ähnliche Relevanz von Evidenzen abgeleitet werden. Dies gilt teils auch für die dahinterliegenden Strukturen und Prozesse, welche teils institutionell vorgegeben sind. Nach diesen Impulsen sollten die Studierenden ihr weiteres Vorgehen (zur Erarbeitung der Studienprojekte, bzw. Projekttag für Schüler*innen) planen. Dabei sollte besonders darauf eingegangen werden, welche Evidenzen für das eigene Projekt notwendig sind. Damit werden die eigenen Strukturen und Prozesse in den Blick genommen.

Im nächsten Schritt wurde vor allem die Recherche von Evidenzen fokussiert. Entsprechend des Mechanismus *Structure and Processes* sollen Strukturen erkannt werden, die zur Suche von Evidenzen helfen. Dazu wurde zuerst die Didaktische Rekonstruktion eingeführt (Kattmann, 2007). Die Studierenden sollten so im Sinne der Strukturen des eigenen Projektes Bereiche ausmachen, in denen die Recherche von Evidenzen für ihr Projekt relevant ist (SP4). Anschließend wurden verschiedene Datenbanken (z.B. PsychNET®, Web of Science™, ERIC, etc.) und diverse Suchstrategien (z.B. Verwendung einer Suchmatrix, Boolean Operatoren, Trunkierungen, etc.) vorgestellt. Um bei der Recherche schnell die Qualität eines Artikels, bzw. der Evidenzen abschätzen zu können, wurden einige Qualitätskriterien eingeführt (z.B. Signifikanz, Effektstärke, Gütekriterien, etc.). Die Studierenden sollten so recherchierte Artikel möglichst schnell bzgl. ihrer Qualität einschätzen können, um ihre Suche zeiteffizient voranzutreiben. Neben den Strukturen im Sinne von Datenbanken konnten so auch eigene Prozesse bei der Suche nach Evidenzen reflektiert werden. Anschließend sollten die Studierenden, unter Berücksichtigung der dargestellten Inhalte, ihr Literaturrecherche beginnen (SP5). Erneut stand damit das eigene Projekt im Mittelpunkt der Intervention.

Im folgenden Abschnitt der Intervention wurde das eigene Generieren von Evidenzen in den Mittelpunkt gestellt. Dazu wurden Aspekte der Stichprobenziehung sowie Beispiele unterschiedlicher Untersuchungsdesigns und -methoden dargestellt. Die zentrale Aufgabe dieser Einheit war die Erstellung eines Erhebungsinstruments für das eigene Projekt (SP6). Vor dem Hintergrund der Strukturen des Seminars müssen hier Einschränkungen im Umfang der entwickelten Instrumente berücksichtigt werden. So werden für das Seminar insgesamt 90 Arbeitsstunden angesetzt, welche zu einer vollumfassenden Beantwortung der Forschungsfrage ggf. nicht ausreichen könnten. Somit könnten sich Unterschiede zwischen den Charakteristika der hier generierten Evidenzen sowie der Definition von Evidenzen in Kap. 2.1 ergeben. Insbesondere könnte in Frage gestellt werden, ob die generierten Evidenzen die best-möglichen Evidenzen zur Beantwortung der jeweiligen Fragestellung darstellen. Eine Berücksichtigung der Rahmenbedingungen des Praxissemesters zur Bewertung der Qualität der generierten Evidenzen scheint entsprechend nötig zu sein. Auch bei den Stichproben handelt es sich zumeist um vorgefundene Lerngruppen innerhalb des individuellen Praxissemesters. Nachdem die Studierenden zuvor den aktuellen Forschungsstand mittels Literaturrecherche aufgearbeitet haben, sollten nun Evidenzen an Stellen erstellt werden, an denen es bisher wenig/unzureichende Hinweise aus der Literatur gab. Dazu kann auf bestehende Testinstrumente zurückgegriffen werden, sofern diese für den eigenen Kontext angemessen erscheinen. Ebenso ist die Adaption, oder völlige Neugestaltung von Instrumenten möglich gewesen. Die erstellten Instrumente wurden anschließend auf der zur Verfügung stehenden Lernplattform hochgeladen und durch Kommiliton*innen kommentiert. Diesem Online-Peerfeedback wurde ein synchroner Austausch in Präsenz nachgeschaltet. Hier konnten vor allem die Aspekte geklärt werden, welche aus dem schriftlichen Feedback noch unklar geblieben sind. Hier können ebenfalls Aspekte von Entscheidungsfindungsprozessen thematisiert werden. Ziel ist es, dass die Proband*innen einen Bedarf an weiteren Evidenzen im Prozess erkennen. Dazu müssen die vorhandenen Evidenzen

gefunden und bewertet werden. Anschließend werden Prozesse der Kooperation und des Peer-Feedbacks eingeführt. Diese Prozesse können ebenfalls als essentielle Aspekte des Umgangs mit Evidenzen gewertet werden. Bezogen auf die *evidence-use* Mechanismen sind hier auch Anteile des Mechanismus *Skills* erkennbar. Vor allem für die Erstellung der Testinstrumente sind auch Teilkompetenzen des Mechanismus *Skills* nötig.

Access and Communication

Bevor die Proband*innen in eine Phase eigenverantwortlichen Arbeitens an ihren Projekten starten, wurde noch die Vermarktung von Evidenzen innerhalb des *evidence-use* Mechanismus *Access and Communication* thematisiert. Der Mechanismus *Access and Communication* fokussiert vor allem die Gestaltung von Evidenzen in Abhängigkeit der jeweiligen Zielgruppe. Damit wird die Gestaltung von Evidenzen auch in Bezug zur Zugänglichkeit dieser gesetzt (Breckon & Dodson, 2016). Dazu wurden zwei Studien zum Thema Nachhaltigkeit in einer gekürzten Version eingesetzt. Die Proband*innen sollten auf Grundlage der genutzten Gestaltungsmittel sowie Sprache auf die mögliche Zielgruppe des Artikels schließen (AC1). Während ein Artikel sich an Forschende im Bereich richtet (Gräsel et al., 2012), adressiert der zweite Artikel eine breitere Leserschaft ohne notwendiges Vorwissen (Gossen et al., 2018). Die Artikel unterscheiden sich bereits in ihrem Layout und verwendeten Abbildungen deutlich voneinander. Im Anschluss sollte die Qualität verschiedener Grafiken bewertet werden (AC2). Das Ziel war es, dass die Proband*innen Merkmale für hochwertige Grafiken erarbeiten und in einer Checkliste festhalten, damit diese anschließend in der Erstellung eigener Grafiken berücksichtigt werden können. Dazu wurden verschiedene Grafiken zur Verfügung gestellt, in welchen die gleichen Daten berichtet wurden. Verschiedene Darstellungsarten und unterschiedliche Achseneinteilungen etc. lassen deutliche Unterschiede in den interpretierbaren Aussagen erkennen. Am Ende der Intervention werden die erstellten Checklisten genutzt, um Feedback zu den erstellten Materialien und Ergebnissen der anderen Gruppen zu geben.

Skills

Nach den dargestellten Inhalten und Aufgaben folgte eine Phase der eigenverantwortlichen Projektarbeit. Die Proband*innen konnten weiterhin Literatur recherchieren, Erhebungsinstrumente planen, oder frei an ihren Projekten arbeiten. Je nach Bedarf konnten die Proband*innen Beratungsangebote durch die Betreuenden in Anspruch nehmen. Innerhalb der Selbstlernphase sollten die Proband*innen ihr WebQuest, bzw. Forschungsprojekt erarbeiten. Mit sieben Sitzungen im Seminar „Chemische Bildung und chemiedidaktische Konzeptionen“ sowie zweieinhalb Sitzungen im Seminar „Begleitforschungsseminar Praxissemester Chemie“, hat die Selbstlernphase einen Großteil der Seminarzeit ausgemacht. Bezogen auf die *evidence-use* Mechanismen, welche hier zur Förderung von angehenden Lehrkräften im Umgang mit Evidenzen adaptiert werden, liegt der Schwerpunkt dieser Phase auf dem Mechanismus *Skills*. Die Selbstlernphase fällt zu großen Teilen in den erneuten Corona-Lockdown im Winter 2020 (CoronaSchVO, 2020/ab dem 18. Dezember 2020 gültige Fassung).

Die erstellten Materialien werden wie unter dem Mechanismus *Access and Communication* beschrieben anhand der entwickelten Checklisten bewertet.

Agree

Die Intervention wurde mit einem Rückblick auf die eigene Arbeit abgeschlossen. Dazu sollten die Studierenden ihre eigene Projektarbeit reflektieren. Sie sollten Stellen im Prozess ausmachen, in welchen sie selbst Evidenzen genutzt haben (AG1). Dabei sollte einerseits das Recherchieren der Evidenzen, andererseits der eigene Umgang mit Evidenzen reflektiert werden. Der Fokus hier lag

nun vor allem auf dem Mechanismus *Agree*. Hier soll ein gemeinsamer Konsens über die Bedeutung von Evidenzen in Entscheidungsfindungsprozessen erstellt werden. Es geht dabei auch darum relevante Fragen zu erkennen und die notwendigen Evidenzen zur Beantwortung dieser Fragen zu benennen. Die Studierenden sollten Ansätze finden, wie Schüler*innen auf eine wissenschaftsfundierte Argumentation bzgl. gesellschaftlicher Diskussionen über naturwissenschaftliche Aspekte vorbereitet werden können. Zu diesem Zweck wurde der wissenschaftliche Konsens thematisiert. Mittels eines Zitats von C. Schrader (o.D.) wurde der wissenschaftliche Konsens eingeführt. Der Fokus hier lag auf der Unterscheidung des wissenschaftlichen Konsens' von der Demokratie. Dabei steht vor allem die Vorläufigkeit von Wissen und die Bedeutung von Evidenzen im Gegensatz zu Meinungen und Mehrheitsverhältnissen innerhalb der Demokratie im Vordergrund. Die notwendige Güte von Methoden und Design innerhalb von Studien wurde mit einem Videoausschnitt (Nyguen-Kim, Mai Thi [maiLab]), wiederum am Beispiel einer Corona-Therapie, vertieft (AG2). In einer anschließenden Diskussion wurde der Begriff des wissenschaftlichen Konsenses kritisch reflektiert. In Anwendung auf die eigenen Projekte der Studierenden konnten Beispiele ergänzt werden. Hier ist ein direkter Bezug zu den individuellen epistemologischen Überzeugungen der Proband*innen erkennbar.

6.4 Stichprobe der Interventionsstudie zur Förderung von Kompetenzen angehender Chemielehrkräfte in evidenzbasierter Praxis

Die Stichprobe der Interventionsstudie besteht aus $N=14$ ($\approx 60\%$ weiblich) Studierenden im Lehramtsstudiengang Chemie (M.Ed.). Die kleine Größe der Stichprobe lässt sich hier auf die geringen Studierendenzahlen im M.Ed. Chemie zurückführen. Zur Stichprobenvergrößerung wurde bereits auf Studierende aus zwei Seminaren zurückgegriffen. Dennoch ist die Stichprobe, vor allem im Zusammenhang mit statistischen Auswertungsverfahren, als klein zu betrachten. Das Alter der Studierenden liegt zwischen 23 bis 33 Jahren ($M=25.93$ Jahre, $SD=3.32$ Jahre). Alle Studierenden gaben an vor der Teilnahme an der Studie an keinem (Fortbildungs-) Angebot zu evidenzbasierter Praxis teilgenommen zu haben. Bis auf eine*n Studierende*n haben alle Proband*innen das Praxissemester bereits in Gänze absolviert ($N=8$) oder waren zum Zeitpunkt der Datenerhebung im Praxissemester ($N=5$). Ein Studierender gab an, das Praxissemester noch nicht abgeschlossen zu haben oder dieses aktuell zu belegen.

Die Teilnahme an der Studie erfolgte freiwillig. Im Gegenzug zur Teilnahme an der Studie, wurden Abgaben im Rahmen der qualifizierten Teilnahme-Leistungen innerhalb des Seminars erlassen. Die Rolle der Lehrenden wurde innerhalb des Seminars klar getrennt: Während die Hauptlehrende des Seminars als Dozentin auftrat, wurde die Rolle des Autors dieser Arbeit klar als Studienleiter formuliert. So verließ die Dozentin z.B. während der Datenerhebung innerhalb der Studie den Raum, um keinen Eindruck einer moralischen Verpflichtung zur Teilnahme an der Studie aufkommen zu lassen.

Eine Maßnahme zur Stichprobenvergrößerung wurde zur Überprüfung der Güte der Testinstrumente umgesetzt. Für eine Faktorenanalyse des Testinstruments *evidence-based practice scales* (vgl. Kapitel 7.3.2) sowie eine Untersuchung des EIDM-ET wurde eine zweite Kohorte ($N=13$; Alter: $M=22.15$, $SD=1.52$; $\approx 46\%$ weiblich) in die Analyse einbezogen. Die Vergrößerung der Stichprobe kann hier zur statistischen Absicherung der Faktorenanalyse beitragen. Bei diesen Proband*innen handelt es sich um Studierende im Lehramtsstudiengang Chemie (B.Ed.). Die Studierenden haben das Seminar „Lernvorgänge im Chemieunterricht“ belegt. Dieses Seminar stellt das letzte chemiedidaktische Seminar vor Ende des Bachelor-Studiums dar. Durch Schulpraktika im Laufe des Bachelor-Studiums haben alle Studierenden dieser Kohorte erste praktische Erfahrungen in der

Schule sammeln können, wenngleich diese Erfahrungen in ihrem Umfang nicht mit dem Praxissemester vergleichbar sind. Diese zweite Kohorte wurde ausschließlich zur Bewertung der Güte der Skalen sowie Erhöhung der Stichprobenvarianz innerhalb des Prätests unabhängig von der Intervention in die Analyse aufgenommen.

Ebenso kann die Gesamtstichprobe des Projekts *Research in Teacher Education* (RiTE) zum Vergleich hinzugezogen werden. Bei dieser Stichprobe handelt es sich um $N=116$ Studierende (44.8% männlich, 52.6% weiblich, 2.6% ohne Angabe) unterschiedlicher Fächer und Länder (Tab. 9). 11.2% der Studierenden gaben an, vor der Teilnahme an dem Projekt bereits an Workshops o.Ä. zum Thema EBP teilgenommen zu haben. Auch diese zusätzliche Stichprobe wird zur Diskussion der Güte der genutzten Instrumente hinzugezogen, nimmt jedoch nicht an der Intervention im Rahmen der vorgestellten Teilstudie teil.

Tabelle 9 – Verteilung der Gesamtstichprobe des Projekts RiTE auf Fachbereiche und Länder

Fachbereich	Land	N
Promotion Biologie	Polen	9
Bachelorstudium Biologie	Polen	13
Masterstudium Biologie	Polen	6
Master of Education Chemie	Deutschland	14
Postgraduate Certificate in Education Mathematik	Großbritannien	49
Promotion Science & Engineering	Niederlande	25

6.5 Instrumente

Innerhalb des ersten empirischen Teils werden ähnliche Instrumente genutzt. Somit werden hier die Instrumente der beiden Teilstudien gemeinsam vorgestellt und aneinander abgeglichen. Wenngleich die Ergebnisse der fächerübergreifenden Erhebung zuerst berichtet werden, geht die Entwicklung der Testinstrumente größtenteils auf die vorherige Durchführung der Interventionsstudie zurück.

6.5.1 Evidence-informed decision making in education test (EIDM-ET)

Der *Evidence-informed decision making in education test* (EIDM-ET) fragt grundlegende Kompetenzen in evidenzbasierter Praxis der Proband*innen ab. Es handelt es sich dabei um einen Fragebogen mit geschlossenen und offenen Items (siehe Anhang). Der Fragebogen stellt eine Adaption des Fresno-Tests (Ramos et al., 2003) durch den Autor dieser Arbeit dar. Diese Adaption ist im Rahmen des EU-Projekts RiTE entstanden.

Der *Fresno-Test* wurde im Bereich der evidenzbasierten Medizin entwickelt. In ihrem umfassenden Review kommen Shaneyfelt et al. (2006) zu der Einschätzung, dass der *Fresno* realistische EPB-Aufgaben beinhaltet. Zu dieser Einschätzung kommen die Autor*innen vor allem auf Grundlage der Items, die angewandtes Wissen und Fähigkeiten abfragen. Die Integration eines Kontexts für die unterschiedlichen Items kann auch im Zusammenhang mit der Kontextabhängigkeit epistemologischer Überzeugungen und deren Einfluss auf Argumentation (Toulmin, 2003) und den Umgang mit Evidenzen (Kvernbekk, 2011) gesehen werden. So besteht der *Fresno-Test* aus 12 offenen Items, die z.B. zur Formulierung einer Fragestellung auf Grundlage einer Situationsbeschreibung auffordern. Zwar sei dadurch die Implementation zeitaufwendiger als bei anderen Testverfahren wie dem *Berlin-Questionnaire* (Fritsche et al., 2002), dafür können jedoch realistische Aufgabenformate abgebildet werden, wodurch die Messung von Kompetenzen in evidenzbasierter Praxis eine höhere Aussagekraft für die Praxis besitzt. Shaneyfelt et al. (2006) stellen in

ihrem Review den *Fresno-Test* sowie den *Berlin-Questionnaire* als die beiden Testinstrumente heraus, welche das Konzept *evidence-based practice* (EBP) am umfangreichsten abbilden können. Als Beispiel eines realistischen Aufgabenformats kann das erste Item des *Fresno-Tests* beschrieben werden. Die erste Hälfte des *Fresno-Tests* basiert auf der Beschreibung zweier klinischer Situationen, zu welchen die Proband*innen Forschungsfragen bilden, ein Forschungsdesign entwickeln und relevante Suchstrategien für Evidenzen vorstellen sollen (Ramos et al., 2003). Die Nähe zum bereits beschriebenen PICO-Schema lässt die hohe Realitätsnähe in Bezug auf evidenzbasierte Medizin deutlich werden. Während der Original-Test für Allgemeinmediziner*innen entwickelt wurde (Ramos et al., 2003), gibt es Adaptionen z.B. auch für angrenzende medizinische Fachbereiche (Cholewa et al., 2015; Coppenrath et al., 2017; Laibhen-Parkes et al., 2018) sowie unterschiedliche Sprachen (Argimon-Pallàs et al., 2010; Cakmakkaya et al., 2021; Rodrigues Salerno et al., 2019).

Auf Grundlage der beschriebenen Ähnlichkeiten zwischen Gesundheits- und Bildungsbereich wurde der *Fresno-Test* in der vorliegenden Teilstudie für den Einsatz im Bereich evidenzbasierten Unterrichtens adaptiert. Ausgangspunkt für die Adaption des *Fresno-Tests* ist die deutsche Adaption „TÜBEP-ST (1.0) Test zur Überprüfung von Basiskenntnissen in evidenzbasierter Praxis für SprachtherapeutInnen“ (Cholewa et al., 2015). Cholewa et al. (2015, S. 24) schlagen vier Schritte zur Adaption des *Fresno Tests* vor:

- I. Übersetzung
- II. Entwicklung fachspezifischer Items
- III. Festlegung der Auswertungskriterien
- IV. Erstellung und Erprobung der Pilotversion

Die Übersetzung der Items ist bereits durch den TÜBEP-ST (1.0) geschehen. An einigen Stellen wurden die Items zur besseren Verständlichkeit leicht umformuliert. Gleiches gilt für die Struktur der Items, welche zur besseren Übersichtlichkeit im EIDM-ET in mehrere Einzelaufgaben unterteilt wurden. Die Veränderungen an den Items und der Teststruktur wurde teils schon durch Cholewa et al. (2015) vorgenommen. So wurden Testitems für die einfachere Implementation des Tests in Single-Choice-Items überführt. Dadurch kann auch der Kritik durch Shaneyfelt et al. (2006) begegnet werden, die hier deutliche Vorteile vor allem in Bezug zum benötigten Zeitaufwand des *Berlin-Questionnaire* sahen (Fritsche et al., 2002). Tabelle 10 gibt einen Überblick über die Teststruktur des *Fresno* im Vergleich zum TÜBEP-ST sowie der Adaption innerhalb der vorliegenden Studie. Neben einigen Ähnlichkeiten der Items zu forschungsmethodologischen Begriffen bzw. dem Forschungsprozess, können für den *Fresno-Test* fachmedizinische Items beschrieben werden, welche innerhalb der beschriebenen Adaptionen nicht vorkommen bzw. angepasst wurden.

Tabelle 10 – Vergleich der Items des *Fresno-Tests* (Ramos et al., 2003), des TÜBEP-ST (Cholewa et al., 2015) sowie des EIDM-ET

Item	Item-format	Fresno (Ramos et al., 2003)	TÜBEP-ST (Cholewa et al., 2015)	EIDM-ET (Adap- tion des Fresno- Test innerhalb der vorliegenden Studie)
Formulierung von Forschungsfragen (PICO-Schema)	Offenes Item	X	X	X (ohne PICO- Schema)
Quellen inkl. Vor- und Nachteile nen- nen	Offenes Item	X	X	X

Item	Item-format	Fresno (Ramos et al., 2003)	TÜBEP-ST (Cholewa et al., 2015)	EIDM-ET (Adap- tion des Fresno- Test innerhalb der vorliegenden Studie)
Formulierung von Suchbegriffen	Offenes Item		X	X
Suchstrategien	Offenes Item	X	X	X
Begriffsdefinitionen zu forschungs- methodologischen Begriffen	Multiple- choice		X	X
Studiendesigns	Offenes Item	X	X	X
Bewertung der Relevanz von Studien- ergebnissen	Offenes Item	X	X	X
Bewertung der Validität von Studien- ergebnissen	Offenes Item	X	X	X
Bewertung der Signifikanz und Bedeu- tung von Studienergebnissen	Offenes Item	X	X	X
Fachmedizinische Fragestellung zu Blutgasen	Multiple- choice	X		
Fachmedizinische Fragestellung Medi- kation koronarer Erkrankungen bei Di- abetiker*innen	Multiple- choice	X		
Frage nach Konfidenzintervall zu fach- medizinischer Frage (Venenthrombose bei Östrogeneinnahme)	Offenes Item	X		
Item zu geeignetem Studiendesign	Multiple- choice	X		
Item zu geeignetem Studiendesign	Multiple- Choice	X		

Zu tieferegreifenden Veränderungen kam es im zweiten vorgeschlagenen Schritt. Einige Items haben eine fachspezifische Überarbeitung erfahren. Die größten Veränderungen sind in Item 1 zu erkennen. Hier sollen die Studierenden auf Grundlage zweier beschriebener Situationen Forschungsfragen ableiten. Die Situationen wurden hier auf den Kontext Schule und Unterricht angepasst. Ein Beispiel stammt dabei explizit aus dem Chemieunterricht, während im anderen Beispiel die (zum Erhebungszeitpunkt) aktuellen Herausforderungen des Homeschooling, während der COVID-19 Pandemie aufgegriffen wurden. Beide Beispiele fokussieren dabei auf spezifisch schulische/unterrichtliche Problemstellungen. Zur Entwicklung möglichst authentischer Situationen wurde innerhalb des Forschungsprojekts RiTE auf die Expertise der verschiedenen beteiligten Institutionen zurückgegriffen. An jeder Institution sollten je zwei Situationen entwickelt werden, welche aus der jeweiligen Fachperspektive mögliche Forschungsfragen aufwerfen können. Die so entstandenen 10 Situationsbeschreibungen wurden anschließend durch alle Partnerinstitutionen bewertet und die zwei geeignetsten ausgewählt. Mit einer solchen fachspezifischen Adaption des Fresno-Tests gingen auch Veränderungen an den Auswertungskriterien einher. Weitere Items konnten aus dem Fresno-Test übernommen werden, während einige Aufgaben aufgrund

fehlender Passung zum Bildungsbereich gestrichen wurden. Dabei erfolgte eine Orientierung am TüBEP-ST nach Cholewa et al. (2015).

Zur Auswertung des Fresno-Test wird ein *grading rubric* eingesetzt (siehe Anhang). Der verwendete *grading rubric* stellt eine Adaption des original *grading rubrics* des *Fresno-Tests* dar (Ramos et al., 2003). Adaptionen fanden vornehmlich für Item 1 statt. Durch die Anpassung der Situationen zur Generierung von Forschungsfragen auf Zusammenhänge aus dem Bildungsbereich, mussten diese auch im *rubric* angepasst werden. Dabei wird für Item 1 bspw. auf die Integration mehrerer relevanter Beschreibungen der Situation in Form der Forschungsfrage eingegangen. Insgesamt können dafür 24 Punkte erreicht werden, je nachdem ob Beschreibungen oberflächlich, vertieft oder multiple angebracht wurden. Dabei müssen die genutzten Beschreibungen nicht zwangsläufig mit den Musterlösungen übereinstimmen. Je nach individueller Forschungsfrage, können die genutzten Beschreibungen variieren. Für die weiteren Items stehen ebenfalls unterschiedliche Antwortniveaus zur Verfügung. Die maximale Punktzahl für jedes Item liegt bei 24 Punkten, sodass insgesamt maximal 216 Punkte erreicht werden können. Bis auf die beschriebenen inhaltlichen Anpassungen, wurde *grading rubric* von Ramos et al. (2003) übernommen.

Die weitverbreitete Nutzung des *Fresno-Tests* kann auch mit der umfangreichen Untersuchung seiner Testgüte erklärt werden. Der *Fresno-Test* wurde in seiner Testgüte mehrfach untersucht. Die Durchführung von Experten*innen-Interviews sowie einer anschließenden Pilotierung des überarbeiteten Testinstruments mit einer Stichprobe aus EBP-Expert*innen und Personen ohne vertiefte Kenntnisse durch Ramos et al. (2003), können als Validierungsargumente gewertet werden. Die Autor*innen berichten eine interne Konsistenz mit einem Wert von $\alpha = .88$. Auch die Auswertung des Tests mithilfe des entwickelten *grading rubrics* lässt mit einer Inter-coder-Übereinstimmung von $\kappa = .98$ auf eine hohe Güte schließen. Auch weitere Untersuchungen geben Hinweise auf eine zufriedenstellende Testgüte des Fresno-Test (für einen Überblick: Lewis et al., 2011). Während der Fresno-Test, sowie diverse weitere Adaptionen des Tests, in seiner Testgüte bereits überprüft wurde, steht eine Überprüfung des EIDM-ET noch aus.

Der erste Entwurf des EIDM-ET wurde im Rahmen einer Pilotstudie ($N=15$) erprobt. Ziel dieser Pilotstudie war es, vor allem mögliche Herausforderungen im Bereich der Itemformulierung zu lokalisieren sowie die Durchführung des Fragebogens zu erproben. Neben einzelnen Konkretisierungen der Aufgabenstellungen wurde in der Nutzung der entwickelten Forschungsfragen für weitere Aufgabenteile eine Beschränkung erkannt. Sofern die Proband*innen keine Forschungsfragen zu den beiden Situationsbeschreibungen erstellen können, kann Item 3 nicht beantwortet werden. Innerhalb von Item 3 sollen die Proband*innen nach geeigneten Suchbegriffen zur Beantwortung einer der Forschungsfragen suchen. Um eine Antwort auch im Falle einer fehlenden eigenen Fragestellung zu ermöglichen, wurde eine inhaltlich abweichende Frage zur Verfügung gestellt. Diese können die Proband*innen nutzen, sofern keine eigene Fragestellung vorhanden ist. Wurden eigene Fragestellungen entwickelt, soll eine davon für Aufgabe 3 genutzt werden. Ein hier möglicherweise entstehender Verstoß gegen die lokale stochastische Unabhängigkeit der Items muss diskutiert werden (Döring & Bortz, 2016). Dabei kann ein Einblick in den *grading rubric* dieser Aufgabe weitere Einblicke liefern. Innerhalb des *grading rubrics* wird ersichtlich, dass die Erstellung der Suchbegriffe in Item 3 unabhängig von der erreichten Punktzahl in Item 1 ist. Werden passende Suchbegriffe auch für eine niedrig bewertete Forschungsfrage aus Item 1 formuliert, kann eine hohe Punktzahl in Item 3 erreicht werden. Sollte keine Frage formuliert worden sein, kann auf die Beispielfrage zurückgegriffen werden. Folglich ist die erreichte Punktezahl in Item 3 nicht von der erreichten Punktzahl aus Item 1 abhängig. Es kann lediglich einen inhaltlichen

Zusammenhang der Items geben. Dieser Zusammenhang kann in Bezug auf die Kontextabhängigkeit des Konstrukts nachvollzogen werden.

Weiterhin wurde die Validität des EIDM-ET im Nachgang der durchgeführten Studie im Rahmen einer unveröffentlichten Masterarbeit (Danzer, 2022) untersucht. Mithilfe von lautem Denken sowie eines leitfadengestützten Interviews sollte der Frage nachgegangen werden, inwiefern die im schriftlichen Fragebogen angegebenen Antworten die mentalen Denkprozesse widerspiegeln. Die Arbeit kommt insgesamt zu dem Schluss, dass der EIDM-ET als „eingeschränkt valide“ (Danzer, 2022, S. 64) bezeichnet werden kann. Abweichungen zwischen den Antworten im Fragebogen und den Ergebnissen des lauten Denkens sowie des Interviews werden z.B. für Aufgabe 1 beschrieben. Begründungen könnten hier in unbewussten Ausschluss bestimmter Aspekte zur Fragerstellung erkannt werden. Ebenso könnte das Fehlen essentieller Aspekte innerhalb der Forschungsfragen auch auf Schwierigkeiten in der schriftlichen Formulierung von Forschungsfragen hindeuten. Unterschiede zwischen den im EIDM-ET formulierten und verbalisierten Fragestellungen zeigen zwei Schwerpunkte auf. Während einige Abweichungen auf Umformulierungen z.B. in Form der Nutzung anderer Fragewörter zurückzuführen sind, stellen andere inhaltliche Ausdifferenzierungen dar.

Die von Danzer (2022) berichteten Unterschiede zwischen formulierten Fragestellungen im EIDM-ET und im lauten Denken müssen kritisch bzgl. der Validität des Testinstruments diskutiert werden. Sowohl das Analysieren relevanter Aspekte in mentalen Prozessen als auch die schriftliche Formulierung einer Forschungsfrage können als relevante Kompetenzen im Umgang mit Evidenzen verstanden werden. Somit würde eine unvollständige Forschungsfrage auf Schwächen im Bereich der Kompetenz der Formulierung von Forschungsfragen als Teilkompetenz des Umgangs mit Evidenzen hindeuten. Im Rahmen des schriftlichen EIDM-ET kann keine Unterscheidung der Problemlagen erfolgen. Folglich könnte die zusätzliche Erhebung von Prozessdaten wie lautem Denken vertiefte Einblicke in die mentalen Modelle liefern. Auf Grundlage der geringen Anzahl beschriebener inhaltlicher Ergänzungen im Rahmen des lauten Denkens, ist hier lediglich eine kleine Beschränkung der Validität des EIDM-ET zu erkennen.

Bei genauerer Betrachtung der beiden genutzten Szenarien für Aufgabe 1, macht Danzer (2022) auf ein unterschiedliches Anforderungsniveau der beiden Situationsbeschreibungen aufmerksam. So berichteten die Proband*innen in den Interviews, dass die Generierung einer Forschungsfrage für Situation 1 (Suche nach Möglichkeiten zum Einbezug der Lernenden in Online-Unterricht) deutlich schwerer gewesen sei. Dies wird über die Offenheit sowie einen nicht vorhandenen Fachbezug begründet. Eine Analyse der Itemschwierigkeit der beiden Situationen auf Grundlage der Prä-Test-Daten der Stichprobe aus dieser Teilstudie ($N=14$) unterstreicht dieses Ergebnis ($p_{\text{Situationsbeschreibung}_1}=72.73$; $p_{\text{Situationsbeschreibung}_2}=84.85$). Aufgrund des nur kleinen Unterschieds der Itemschwierigkeiten sind jedoch keine großen Unterschiede im Anforderungsniveau der beiden Situationsbeschreibungen zu erwarten.

Innerhalb von Aufgabe 4 des EIDM-ET beschreibt Danzer (2022) Abweichungen zwischen den verschriftlichten Antworten und verbalisierten Gedanken von zwei Proband*innen. Bei diesen Fällen gab es Probleme bzgl. der Verständlichkeit der Fragestellung. So haben die Proband*innen keine Strategien zur Ausweitung/Eingrenzung der Suche angegeben, sondern die Suchbegriffe spezifiziert, bzw. ergänzt. Für eine erneute Nutzung des Fragebogens könnte eine wie von Danzer (2022) beschriebene Konkretisierung der Aufgabenstellung (Nutzung der Formulierung „Strategien, um die Suche einzugrenzen/auszuweiten“, anstelle von „Wegen die Suche einzugrenzen/auszuweiten“) derlei Verständnisprobleme verhindern. Für die vorliegende Version des EIDM-ET sind auf Grundlage der Arbeit von Danzer (2022) folglich leichte Abstriche der Validität von Aufgabe 4 zu

verzeichnen, wenngleich die Missinterpretation der Aufgabenstellung aufgrund der Ergebnisse der Pilotstudie als eher selten zu beschreiben ist.

Für die Aufgaben 5-8 beschreibt Danzer (2022) vereinzelte Verständnisprobleme, jedoch hauptsächlich eine gute Übereinstimmung der Bearbeitung des EIDM-ET und der verbalisierten Gedanken im lauten Denken. So kann die Validität dieser Aufgaben durch die Untersuchung gestützt werden. Für Aufgabe 9 hingegen sind größere Verständnisprobleme aufgetreten. Neben Verständnisproblemen der Aufgabenstellung, könnte hier auch ein konzeptionelles Verständnisproblem evidenzbasierten Unterrichtens erkannt werden. Die konzeptionelle Trennung von Forschung und Unterrichten in zwei ontologisch unabhängige Kategorien, könnte eine Verknüpfung beider Bereiche im Sinne der Aufgabenstellung erschweren/verhindern. Im Sinne dieses Erklärungsansatzes, könnte Aufgabe 9 als valide beschrieben werden, da die Proband*innen die geforderte Verknüpfung von Forschung und Unterrichten aufgrund konzeptioneller Hürden nicht herstellen könnten. Verständnisprobleme der Itemformulierung bzw. Begriffsunklarheiten können jedoch nicht ausgeschlossen werden. Im Sinne des argumentbasierten Validitätsverständnisses nach Kane (2016) scheint das fehlende konzeptionelle Verständnis evidenzbasierten Unterrichtens vor dem Hintergrund der eher geringen Punktzahlen in den übrigen Aufgaben des EIDM-ET eher zuzutreffen.

Danzer (2022) hebt noch einmal die hohe Bedeutung des Fachwissens für den EIDM-ET hervor. So müssen z.B. unterschiedliche statistische Verfahren oder Zusammenhänge erkannt werden. Dabei sind im EIDM-ET Schwerpunkte erkennbar, welche ggf. nicht die ganze Breite des Konstruktes EBP abbilden (z.B. Auswahl von Studiendesigns, etc.). Dieser Fokus auf eine Auswahl bzw. Teilfacetten des Konstrukts EBP muss im Vergleich zum Fresno-Test als Einschränkung verstanden werden. Es kann entsprechend kein Anspruch erhoben werden, dass der EIDM-ET die Kompetenzen im Bereich evidenzbasierter Praxis im Bildungsbereich im holistischen Sinne erhebt.

Die Reliabilität des EIDM-ET fällt für den Post-Zeitpunkt geringer aus als für den Prä-Zeitpunkt. Dabei sind in der vorliegenden Studie vor allem die niedrigen Proband*innen-Zahlen zu beachten. Im Rahmen des Erasmus+ Projekts *Research in Teacher Education* konnten mit einer größeren Stichprobe (siehe Kap. 6.4) für das gleiche Testinstrument akzeptable Werte für Cronbachs Alpha ermittelt werden. Auch eine Untersuchung der Stichprobe dieser Teilstudie, ergänzt um die zweite Kohorte des Lehramtsstudiums Chemie (B.Ed.) lässt eine solide Reliabilität erkennen.

Tabelle 11 – Reliabilität des EIDM-ET an Prä-, Post- und follow-up-Zeitpunkt für die Stichprobe dieser Studie, ergänzt um die zweite Kohorte des Lehramtsstudiums Chemie (B.Ed.) sowie die Gesamtstichprobe im Projekt *Research in Teacher Education*.

Stichproben	Itemanzahl	Cronbachs α Prä-Zeitpunkt [N]	Cronbachs α Post-Zeitpunkt [N]
Stichprobe der vorliegenden Studie	9	.511 [14]	.312 [14]
Stichprobe der vorliegenden Studie, ergänzt um die zweite Kohorte des Lehramtsstudiums Chemie (B.Ed.)	9	.595 [30]	-
Gesamtstichprobe des Projekts <i>Research in Teacher Education</i>	9	.674 [116]	.747 [116]

Eine Betrachtung der Einzelitems würde eine Verbesserung der Reliabilität des Tests bei Ausschluss einzelner Items ermöglichen. Dabei stellt sich jedoch keines der Items heraus, welches durch Ausschluss an beiden Messzeitpunkten einen positiven Einfluss auf die Reliabilität hätte. Auch vor dem Hintergrund theoretischer Überlegungen ergibt ein Ausschluss einzelner Items zur Steigerung der Reliabilität keinen Sinn. Der Ausschluss einzelner Items würde hier ggf. zu einem schmaleren Fokus auf einzelne Facetten evidenzbasierter Praxis führen. Der Wert der Operationalisierung des Konstruktes evidenzbasierter Praxis im vorliegenden EIDM-ET muss Schecker (2014) folgend einer unzufriedenstellenden Reliabilität gegenübergestellt werden. Ein strenges Anlegen bestimmter *alpha-thresholds* scheint auch vor dem Hintergrund messtheoretisch notwendiger Stichprobengrößen von 300-400 Proband*innen (Charter, 2003) für die vorliegende Studie nicht sinnvoll zu sein. Die gesteigerte Reliabilität des Testinstruments in der größeren Stichprobe der Projekts RiTE könnte als Indiz einer akzeptablen internen Konsistenz interpretiert werden. Für die vorliegende Stichprobe von 14 Proband*innen hingegen, muss die Reliabilität des Instruments am Post-Zeitpunkt auf Grundlage des Cronbachs α Werts als unzureichend bewertet werden.

Alle Datensätze zu Prä- und Postzeitpunkt wurden von zwei Kodierenden auf Grundlage des beschriebenen *grading rubrics* bewertet. Eine durchschnittliche Intercoderreliabilität für den Prä-Test von $K_{Prä} = .957$ sowie den Post-Test von $K_{Post} = .908$ ist vergleichbar mit berichteten Werten des Originalinstruments und kann als gut bezeichnet werden.

Im Rahmen der vorgestellten Verfahren konnte die Güte des EIDM-ET grundständig eingeschätzt werden. Auf Grundlage der Arbeit von Danzer (2022) konnten Indizien für die Validität des EIDM-ET gesammelt werden. Im Rahmen dieser Arbeit soll der EIDM-ET genutzt werden, um Einblicke in die Kompetenzen in EBP der Proband*innen zu erlangen. Nicht zuletzt aufgrund kleiner Stichprobengrößen dürfen die Ergebnisse jedoch nicht generalisierend bewertet werden. Vor allem Abstriche im Rahmen der Reliabilität sollten hier betrachtet werden.

6.5.1.1 Kurzfassung des EIDM-ET

Innerhalb der vorliegenden fächerübergreifenden Erhebung wurde eine Kurzfassung des vorgestellten EIDM-ET verwendet. Ziel der Verkürzung des Instruments war es, die benötigte Zeit für die Beantwortung zu reduzieren, damit das Instrument in die Evaluation im Rahmen des Praxissemesters integriert werden konnte. Ebenso sollten Schwächen der langen Fassung der EIDM-ET in der gekürzten Version des Instruments Beachtung finden.

Für die Auswahl der Items wurden zuerst die Itemschwierigkeiten auf Grundlage der in der Interventionsstudie erhobenen Daten errechnet. Zusätzlich wurden die Items inhaltlich kritisch betrachtet. Während alle Items als relevant für das Konstrukt evidenzbasierter Praxis angesehen werden können, bestand für die vorliegende Kurzfassung die Notwendigkeit eine Auswahl zu treffen. Entsprechend wurden die Items erneut bzgl. ihrer Relevanz bewertet und so in die Kurzfassung des Instruments integriert oder ausgeschlossen. Items 1-3 wurden in das gekürzte Instrument übernommen. Lediglich Item 1 wurde um Situation 2 gekürzt. Somit wurde noch eine Situationsbeschreibung genutzt, welche durch ihren fachunspezifischen Fokus auf Lernendenengagement im Rahmen des Online-Unterrichts besonders geeignet für eine fächerübergreifende Erhebung ist. Item 4 wurde vor allem auf Grundlage der häufigen Missverständnisse, Item 5 aufgrund geringerer Relevanz für das Konstrukt evidenzbasierter Praxis ausgeschlossen. Während Item 7 übernommen wurde, wurden Items 8 und 9 ausgeschlossen. Item 8 wurde aufgrund der hohen Itemschwierigkeit sowie auftretender Missverständnisse der Itemformulierung ausgeschlossen.

Für Item 9 waren die hohe Bearbeitungszeit sowie ebenfalls auftretende Missverständnisse der Itemformulierung ausschlaggebend für den Ausschluss.

Das gekürzte Testinstrument lässt nach wie vor eine Kontextorientierung der Items erkennen. Vor allem der inhaltliche Zusammenhang der ersten drei Items kann als Stärke aufgefasst werden (Shaneyfelt et al., 2006). Durch die umfassende Kürzung kann ebenso die Bearbeitungszeit verkürzt werden, was im Rahmen des Einsatzes des Instruments innerhalb der Evaluation des Praxissemesters notwendig ist. Tabelle 12 gibt einen Überblick über die Itemschwierigkeiten sowie die letztlich ausgewählten Items.

Weiterhin kann vermutet werden, dass die Proband*innen teils noch nicht mit dem Konstrukt evidenzbasierter Praxis in Kontakt gekommen sind. Dies könnte auch auf Begriffe zutreffen, welche im Rahmen des Konstrukts genutzt werden, z.B. Evidenz, Befunde, etc. Diese Verständnisprobleme, oder die Unkenntnis, was unter dem Konstrukt evidenzbasierter Praxis zu fassen ist, könnten zu Validitätsproblemen führen. Dadurch könnte eine unterschiedliche Beantwortung unterschiedlicher Aufgaben erklärbar sein. Mithilfe einer kurzen Einleitung in das Testinstrument sollte bspw. eine knappe Definition von Evidenzen gegeben werden.

Einleitungstext:

„Momentan spielt der Umgang mit Fakten vs. Meinungen auch im schulischen Umfeld eine große Rolle. Schülerinnen und Schüler müssen auch im Unterricht lernen, welche Informationen verlässlich sind und was im Bereich der Meinungsbildung anzusiedeln ist.

Im folgenden Teil des Fragebogens wollen wir daher Ihre Einstellung gegenüber Evidenzen (die besten verfügbaren, systematisch durch mehrere Studien gesammelten Informationen) erfragen. Auch möchten wir gern wissen, inwiefern Sie mit evidenz-basierter Praxis in Kontakt waren bzw. sie eingesetzt haben.“

Dieser Einleitungstext sollte zur Steigerung der Validität des Testinstruments dienen.

Tabelle 12 – Itemschwierigkeiten und -varianz für EIDM-ET sowie Auswahl der Items für das gekürzte Instrument

# Item	Item	Teilaufgaben	Itemschwierigkeit P_i	Item-varianz σ^2	Kommentar
1	<i>Formulieren Sie eine Forschungsfrage in Bezug auf die folgende Situation. Die Frage sollte Ihnen dabei helfen, Literatur zur Thematik zu finden und gute Artikel auszusuchen. (2 Situationen)</i>	Situation 1	72.73	.564	Nur Situation 1 wird in Kurzfassung übernommen, da Situation 2 sich spezifisch nur auf Chemieunterricht bezieht.
		Situation 2	84.85	.473	
2	<i>Wo könnten Sie eine mögliche Antwort auf Ihre Frage finden? Nennen Sie 4 Typen von Quellen und geben Sie jeweils Vor- und Nachteile an.</i>		57.58	.818	Übernahme in Kurzfassung aufgrund einfach Bearbeitung und mittlerer Itemschwierigkeit.
3	<i>Stellen Sie sich vor, Sie suchen nach Forschungsergebnissen zu Ihrer Frage (aus Aufgabe 1) in einer internationalen Datenbank. Welche wären Ihre besten 3 Suchbegriffe?</i>		78.79	.655	Übernahme in Kurzfassung aufgrund einfacher Bearbeitung und zusammenhängendem Charakter des Prozesses abgebildet durch Items 1-3.
4	<i>Denken Sie über Wege nach, um Ihre Suche einzugrenzen, bzw. auszuweiten. Geben Sie 2 Beispiele, wie Sie ihre Suche in einer Datenbank eingrenzen/ausweiten können.</i>		63.18	.455	Items ausgeschlossen, da es häufig missverstanden wurde und so viele fehlende Messwerte aufgetreten sind.
		Effektivität einer didak. Methode	18.18	.455	
5	<i>Studien müssen angemessen gestaltet werden, um passende Daten zu erheben. Welches Studiendesign eignet sich am besten für die folgenden Themen?</i>	Datensammlung bzgl. Präkonzepten der Lernenden	27.27	.467	Items (Inkl. Unteraufgaben) ausgeschlossen da vergleichsweise weniger relevant. Ebenso führte die strikte Anwendung des <i>grading rubrics</i> hier zu großen Itemschwierigkeiten.
		Leistung von Lernenden in verschiedenen Ländern	k.A.	k.A.	
6	<i>Was versteht man unter...?</i>	Randomisierung	72.73	.218	

7	Welche Charakteristika einer Studie sollten in einem Forschungsbericht dargelegt werden, damit Sie die Validität der Befunde akzeptieren? Geben Sie 5 Beispiele.	Reliabilität	63.64	.255	Übernahme aller Items in Kurzfassung aufgrund von mittleren Itemschwierigkeiten, thematischer Relevanz und kurzer Bearbeitungszeit.
		Kontrollgruppe	45.45	.273	
			54.56	1.164	Übernahme in Kurzfassung aufgrund von mittlerer Itemschwierigkeit und hoher thematischer Relevanz.
8	Nachdem Sie nun über Aspekte nachgedacht haben, welche zur Steigerung der Validität einer Studie beitragen, geht es nun um Signifikanz und Bedeutung von Studien. Geben Sie je zwei Charakteristika von Befunden an, damit sie als signifikant und bedeutungsvoll bezeichnet werden können.	Signifikanz	27.28	.273	
					Ausschluss beider Items aufgrund von hoher Itemschwierigkeit so häufigem Missverstehen der Aufgabenstellung.
9	Nun haben Sie sich über Validität, Signifikanz und Bedeutung von Studien Gedanken gemacht. In dieser Aufgabe sollen sie 4 Aspekte, inkl. kurzer Begründung angeben, mit denen Sie die Umsetzungsmöglichkeiten der Befunde im schulischen Alltag bewerten würden.	Bedeutung	18.18	.255	
			34.09	1.455	Ausschluss des Items aufgrund häufigem missverstehen der Aufgabenstellung sowie längerer Bearbeitungszeit.

Anmerkung. Die ausgegrauten Items wurden nicht in das gekürzte Testinstrument übernommen.

Wie aus Tabelle 12 ersichtlich, wurden letztlich fünf Items des EIDM-ET in die Kurzfassung des Instruments übernommen. Dabei sind Items mit Itemschwierigkeiten $45 < P_i < 78$ ausgewählt. Leichtere Items wurden aufgrund der teilweise geringeren thematischen Relevanz in Kombination mit dem Ziel eines möglichst kurzen Testinstruments ausgeschlossen. Items mit einer Itemschwierigkeit $P_i < 45$ wurden aufgrund der geringen Lösungswahrscheinlichkeit verworfen. Aus Teilstudie 2 kann abgeleitet werden, dass diese Items trotz Teilnahme an der Intervention schwer zu lösen sind. Es ist daher erwartbar, dass die Items für die Stichprobe der vorliegenden Teilstudie, welche keinerlei Intervention erhalten hat, noch schwerer sind. Andererseits könnte eine strikte Auslegung des *grading rubrics* an einzelnen Stellen die hohe Itemschwierigkeiten auf Grundlage der Daten aus der Interventionsstudie erklären (Danzer, 2022). Da in der vorliegenden Teilstudie ebenfalls auf den beschriebenen *grading rubric* zurückgegriffen wird, können die Itemschwierigkeiten dennoch zur Auswahl der Items für die gekürzte Fassung des Instruments berücksichtigt werden.

Weiterhin sind die durch Danzer (2022) herausgearbeiteten Validitätsargumente für die einzelnen Aufgaben bei der Auswahl der Items für die Kurzfassung beachtet worden. So wurde z.B. Item 4 aufgrund von deutlicheren Abweichungen zwischen den schriftlichen Antworten in EIDM-ET und den Verbalisierungen in den Interviews ausgeschlossen. Ähnliches gilt für Item 9. Für die Items 5-8 beschreibt Danzer (2022) vereinzelte Verständnisprobleme der Aufgabenstellungen. Für die Kurzfassung wurde die Auswahl dieser Items entsprechend hauptsächlich durch die inhaltliche Relevanz begründet. Wenngleich das Testinstrument in der Kurzfassung eine deutliche Verkürzung erfahren hat, können die Einzelitems auf Grundlage der Analysen von Danzer (2022) als valide betrachtet werden. Die Breite der abgefragten Kompetenzen im Bereich evidenzbasierter Praxis ist im vorliegenden Instrument jedoch geringer.

Für die Bepunktung der Test-Antworten wurde erneut der *grading rubric* genutzt (vgl. Interventionsstudie). Die maximale Punktzahl der Einzelitems liegt hier bei 24 Punkten. Eine Ausnahme davon bildet Item 1 für das nach der Verkürzung des EIDM-ET lediglich 12 Punkte erreicht werden können, da hier eine Situationsbeschreibung gestrichen wurde. Für Item 4 teilen sich die maximal erreichbaren Punkte auf 3 getrennte Teilfragen auf. Es wurden ca. 25% des Materials durch eine zweite Kodiererin kodiert. Die erreichte Interdecoder-Übereinstimmung zum Prä- und Post-Zeitpunkt ($\kappa_{\text{Prä}} = .69$; $\kappa_{\text{Post}} = .70$) erreicht nach Landis und Koch (1977) die zweit höchste Benchmark (*substantial*).

Die Untersuchung der internen Konsistenz der Kurzfassung des Tests schneidet im Vergleich zum vollständigen EIDM-ET schwächer ab ($\alpha_{\text{Prä}} = .296$, $N_{\text{Prä}}=364$; $\alpha_{\text{Post}} = .377$, $N_{\text{Post}}=362$). Ein möglicher Grund für diese Unterschiede könnte auch in der Befragungssituation liegen. Der gekürzte EIDM-ET wurde hier im Rahmen einer größeren Online-Evaluation eingesetzt. Im Rahmen der Interventionsstudie wurde der EIDM-ET zwar auch online eingesetzt, jedoch gab es die Möglichkeit Rückfragen zu stellen. Ebenso könnte die Befragung durch die Durchführenden in einen Projektkontext eingebettet werden. Einen anderen möglichen Grund für die mäßige interne Konsistenz könnte das Verständnis des Konstrukts „Evidenzen“ der Proband*innen darstellen. Wie gezeigt werden konnte, ist das Konstrukt ‚Evidenzen‘, oder auch ‚evidenzbasierte Praxis‘ sehr komplex und durch das Konzept der *Balance* geprägt (vgl. Kap. 3.2). Das sehr komplexe Konstrukt könnte ggf. zu wenig bekannt gewesen sein, als dass eine konsistente Beantwortung möglich gewesen wäre. Ebenso könnten Verständnisprobleme bei den genutzten Begriffen (auch in der Itemformulierung) aufgetreten sein. Auch der eingesetzte Einleitungstext konnte hier scheinbar nicht zu einem konsistenten Verständnis beitragen, bzw. wurde ggf. nicht entsprechend genutzt.

Eine weitere Begründung der geringen Konsistenz könnte in der Verkürzung des Testinstruments vermutet werden. Das Testinstrument scheint in seiner gekürzten Form nicht dazu geeignet zu sein, das Konstrukt in Gänze abzufragen. In Kombination mit möglicherweise bestehenden Verständnisproblemen der Aufgabenstellungen, könnten unterschiedliche Antwortverhalten für die unterschiedlichen Aufgabenstellungen resultieren. Gleichzeitig erschwert die geringe Itemanzahl (5 Items) des gekürzten Testinstruments das Erreichen hoher Werte für die interne Konsistenz, da die Itemanzahl in die Berechnung der internen Konsistenz eingeht (Schecker, 2014).

Zur tiefergehenden Analyse des geringen α -Werts wurden spezifische Gruppen bzgl. der Beantwortung des gekürzten Instruments verglichen. In einem ersten Schritt wurde das gekürzte Instrument für den Prä-Zeitpunkt der Stichprobe der Interventionsstudie simuliert. Der resultierende α -Wert von .214 ($N=14$) ist dabei geringer als der der vorliegenden Studie. Der geringere α -Wert könnte auf unterschiedliche Durchführungsbedingungen zurückgeführt werden. Wie berichtet, fand die Erhebung im Rahmen der Interventionsstudie in einem anderen Rahmen statt (Durchführung des Online-Fragebogens in Anwesenheit des Autors dieser Arbeit, persönliche-/fachliche Vertrautheit der Proband*innen, etc.). Damit könnte die Vergleichbarkeit dieser beiden Stichproben nicht gegeben sein. Anschließend wurden Teilstichproben der fächerübergreifenden Befragung untersucht. Die vermuteten Zusammenhänge zwischen dem Begriffsverständnis und der internen Konsistenz könnten durch die generelle Leistungsfähigkeit von Proband*innen erklärt werden. Somit könnte ein Einfluss der Abitur-, oder Bachelor-Note auf die Konsistenz der Beantwortung der Items vermutet werden. In den Analysen zeigt sich, dass eine Teilstichprobe mit einer Abiturnote < 1.5 einen α -Wert von .414 ($N=17$) und eine Teilstichprobe mit einer Bachelor-Note < 1.5 einen α -Wert von .489 ($N=20$) erreichen. Beide Werte liegen deutlich über dem berichteten α -Wert für die vorliegende Hauptstichprobe und könnten so auf einen Einfluss der Leistungsfähigkeit hindeuten. Dadurch könnte die Annahme, dass das komplexe Konstrukt ggf. nicht ausreichend verstanden wurde, um konsistent auf Items zu reagieren, unterstützt werden. Somit können die Ergebnisse eher auf Unterschiede in der Konsistenz des Antwortverhaltens durch die individuelle Leistungsfähigkeit im Sinne des Konstruktverständnisses gedeutet werden. Unterstrichen wird diese Vermutung durch die schwachen Korrelationen zwischen Abiturnote und EIDM-ET Punktzahl ($r = .124$, $p = .013$, $N=405$) sowie Bachelor-Note und EIDM-ET Punktzahl ($r = .115$, $p = .021$, $N=405$).

Wie gezeigt wurde, sollte vor allem das Verständnis des Konstrukts ‚Evidenzen‘ in der vorliegenden Studie kritisch betrachtet werden. Als Argument könnte angeführt werden, dass die Proband*innen innerhalb dieser Studie zuvor keinen Kontakt mit dem Konstrukt hatten und so kein einheitliches Verständnis vorliegt. Dem kann jedoch die Konkretheit der Itemformulierungen des EIDM-ET entgegengestellt werden. Der Einbezug von Situationsbeschreibungen sowie die darauffolgenden Itemformulierungen zur Literaturrecherche, etc. stellen konkrete Fragen dar, für welche das Verständnis des Konstruktes Evidenzen ggf. nicht vollumfänglich nötig ist. So könnte z.B. auch ohne ein präzises Verständnis des Konstruktes auf entscheidende Validitätsargumente von Studien eingegangen werden. Bessere α -Werte für einzelne Teilstichproben zeigen Indizien für den Einfluss der generellen Leistungsfähigkeit auf. Für den EIDM-ET in Gesamtlänge konnten sowohl zufriedenstellende Werte für die interne Konsistenz bestimmt als auch Validitätsargumente für die einzelnen Aufgaben (Danzer, 2022) abgeleitet werden. Für einen weiteren Einsatz des gekürzten Instruments sollte dieses überarbeitet werden. Andererseits könnte sich der Einsatz des EIDM-ET in voller Länge anbieten.

6.5.2 Evidence-based practice scales

Neben dem EIDM-ET wurden Skalen zu *evidence-based practice* (EBP) zur Datenerhebung eingesetzt. Es handelt sich dabei um die *evidence-based practice beliefs* sowie *evidence-based practice implementation scale* von Melnyk et al. (2008). Beide Skalen erheben Selbsteinschätzungen der Proband*innen. Die *EBP beliefs scale* fokussiert auf Werthaltung gegenüber EBP sowie der Einschätzung der eigenen Fähigkeiten in dieser. Dabei wird auch auf das Ergebnis evidenzbasierter Praxis im Sinne der Schüler*innen-Leistung abgehoben. Die *EBP implementation scale* legt einen Schwerpunkt auf die Durchführung von evidenzbasierter Praxis. Hier werden verschiedene Szenarien evidenzbasierter Praxis nach der Häufigkeit deren Durchführung durch die Proband*innen abgefragt. So können neben der Untersuchung individueller Kompetenzen durch den EIDM-ET auch Einstellungen und die tatsächliche Nutzung evidenzbasierter Praxis abgefragt werden. Während die *EPB beliefes scale* zum Prä- und Post-Zeitpunkt eingesetzt wurde, wurde die *EPB implementation scale* nur zum Post-Zeitpunkt eingesetzt. Dies wird durch den Fokus auf den Umgang mit Evidenzen im Praxissemester begründet. So wird angenommen, dass die Nutzung evidenzbasierter Praxis im Rahmen der Schulpraxis vor dem Praxissemester gering ausgeprägt ist, da es an Einsatzmöglichkeiten fehlt.

Beide Skalen wurden im Bereich evidenzbasierter Medizin entwickelt und eingesetzt. Eine Übertragung in den Bildungsbereich erscheint auf Grundlage der beschriebenen Ähnlichkeiten zum Gesundheitsbereich (vgl. Kap. 2.4) als vielversprechend. In einem ersten Schritt wurden die Skalen im Rahmen des Projekts *Research in Teacher Education* für den Bildungsbereich angepasst. In der englischen Version der Skalen sind dazu nur kleinere Veränderungen nötig geworden. So wurden, z.B. Begriffe wie „*patient*“ mit „*student*“ ersetzt. Tiefergehende Eingriffe in die Skalen sind häufig nicht nötig geworden. Die *implementation scale* wurde um zwei Items gekürzt, welche für den Bildungsbereich nicht passend erschienen (siehe Tab. 13). Item 15 fokussiert auf das medizinspezifische PICO-Schema, welche im Bildungsbereich nicht einfach übertragen werden kann. Item 17 spielt auf *Clearing Houses* für medizinische Leitlinien an. Wenngleich es auch im Bildungsbereich *Clearing Houses* gibt (z.B. Clearing House Unterricht der Technischen Universität München⁷), besitzen diese doch scheinbar eine andere Rolle als in der Medizin. Wie beschrieben werden Leitlinien institutionalisiert in der medizinischen Behandlung von Patient*innen eingesetzt. Dabei wird auch auf Standardverfahren in spezifischen Situationen abgehoben. Eine solche institutionalisierte Rolle von *Clearing Houses* bzw. Leitlinien ist im Bildungsbereich nicht zu beobachten. Ebenso scheint die Entwicklung von Leitlinien nicht der Zielstellung z.B. des *Clearing House* Unterricht zu entsprechen⁸. Auch die Popularität von *Clearing Houses* im Bildungsbereich kann als geringer vermutet werden. Aufgrund der sehr unterschiedlichen Rolle der *Clearing Houses* wurde das Item aus der Untersuchung ausgeschlossen. In einem zweiten Schritt wurden die bereits überarbeiteten Skalen auf Deutsch übersetzt. Die Proband*innen können dabei auf einer 5-stufigen Likert-Skala mit 1=stark negative *beliefs* gegenüber evidenzbasierter Praxis und 5=stark positive *beliefs* auf die Items der *EBP beliefs scale* reagieren. Die *EBP implementation scale* hingegen fragt auf einer 5-stufigen Skala ab, wie häufig eine Tätigkeit in den letzten 8 Wochen unternommen wurde.

⁷ <https://www.clearinghouse.edu.tum.de/>

⁸ <https://www.clearinghouse.edu.tum.de/projekt/>

Tabelle 13 – Englischsprachige Items der evidence-based practice beliefs/implementation scale

Itemzahl	Evidence-based practice implementation scale	Evidence-based practice beliefs scale
1	Shared the outcome data collected with colleagues.	I am sure that I can implement EBP in a time efficient way.
2	Shared evidence from a study/ies in the form of a report presentation to >2 colleagues.	I am sure that I can implement EBP.
3	Shared EBP guidelines with a colleague.	I believe that I can search for the best evidence to answer educational questions in a time efficient way.
4	Shared evidence from a research study with a multidisciplinary team member.	I am confident about my ability to implement EBP where I work.
5	Used an EBP guideline or systematic review to change educational practice where I work.	I believe that I can overcome barriers in implementing EBP.
6	Changed practice based on student outcome data.	I am sure about how to measure the outcomes of educational interventions .
7	Evaluated an educational initiative by collecting student outcome data.	I know how to implement EBP sufficiently enough to make practice change.
8	Evaluated the outcome of a practice change.	I am sure that I can access the best resources in order to implement EBP.
9	Promoted the use of EBP to my colleagues.	I am sure that implementing EBP will improve my courses .
10	Used evidence to change my educational practice.	I believe that critically appraising evidence is an important step in the EBP process.
11	Shared evidence from a research study with a student/parents .	I am clear about the steps of EBP.
12	Read and critically appraised an educational research study.	I am sure that evidence-based guidelines can improve education .
13	Informally discussed evidence from a research study with a colleague.	I believe that EBP results in the best education for the students .
14	Critically appraised evidence from a research study.	I believe the lessons I give are evidence-based.
15	Generated a PICO question about my clinical practice.	I believe EBP is difficult. (reversed score)
16	Collected data on student's problem.	I believe that EBP takes too much time. (reversed score)
17	Accessed the National Guideline Clearinghouse.	-
18	Accessed an educational database . Of systematic reviews.	-

Anmerkung. Markierte Wörter stellen Adaptionen der Originalinstrumente von Melnyk et al. (2008) dar.

Die Skalen wurden von Melnyk et al. (2008) als eindimensionale Skalen vorgestellt. Um eine mögliche Mehrdimensionalität der Skalen auszuschließen, wurde eine Faktorenanalyse durchgeführt. Für die Durchführung der Faktorenanalyse wurde wie beschrieben eine zweite Kohorte in die Analyse aufgenommen. Die Gesamtstichprobe für die Faktorenanalyse besteht folglich aus $N=27$ Proband*innen. Während die Daten der *EBP beliefs scale* sich für die Analyse mittels Faktorenanalyse (mit Abstrichen) eignen ($\text{Chi-Quadrat}(120)=289.966$; $p<.001$; $KMO=.551$), ergibt sich für die *EBP implementation scale* keine Eignung für eine Faktorenanalyse (nicht positiv definite Korrelationsmatrix).

Die Items der *EBP beliefs scale* wurden einer Hauptkomponentenanalyse zugeführt. Die Analyse zeigt fünf Hauptkomponenten mit einem Eigenwert >1 auf. Im Abgleich mit dem *Screeplot* sowie einer Analyse der Items (z.B. Querladungen von Items auf mehrere Faktoren, inhaltlicher Zusammenhang der Items, etc.) und der Itemanzahl pro Faktor werden drei Faktoren gebildet. Die inhaltliche Analyse der Faktoren zeigt unterschiedliche Schwerpunkte auf und machte den

Ausschluss einzelner Items nötig, um inhaltlich kohärente Faktoren zu erstellen. Während sich der erste Faktor mit dem grundsätzlichen *beliefs* des Einflusses von evidenzbasierter Praxis auf den Bildungsbereich auseinandersetzt, fokussieren die anderen Faktoren auf die *beliefs* bzgl. der Implementation evidenzbasierter Praxis in der Schule (Faktor 2) sowie die *beliefs* gegenüber eigener Fähigkeiten in evidenzbasierter Praxis (Faktor 3). Tabelle 14 gibt einen Überblick über die gebildeten Faktoren, die zugeordneten Items und die Reliabilität zu verschiedenen Zeitpunkten bzw. innerhalb der verschiedenen Substichproben.

Tabelle 14 – Faktoren der EBP beliefs scale, zugehörige Items sowie Reliabilität zu verschiedenen Zeitpunkten, bzw. innerhalb verschiedener Substichproben

Nr.	Faktor	Nr.	Item	Reliabilität der zusammengeführten Stichprobe (N=27) am Prä-Zeitpunkt	Reliabilität innerhalb Hauptstichprobe (N=14) am Prä-Zeitpunkt	Reliabilität innerhalb Hauptstichprobe (N=14) am Post-Zeitpunkt
1	Beliefs gegenüber dem generellen Einfluss von EBP auf Bildung	9	„Ich bin mir sicher, dass die Implementierung von EBP meinen Unterricht verbessert wird.“	.865	.801	.798
		12	„Ich bin mir sicher, dass evidenz-basierte Leitlinien Bildung verbessern können.“			
		13	„Ich glaube, dass EBP zu den besten Bildungsergebnissen von Lernenden führt.“			
2	Beliefs gegenüber der Implementation von EBP in der Schule	7	„Ich weiß, wie man EBP so weit umsetzt, dass sich die Praxis ändert.“	.873	.802	.335
		11	„Mir sind die EBP-Schritte klar.“			
		14	„Ich glaube, dass mein Unterricht evidenz-basiert ist.“			
3	Beliefs gegenüber persönlichen Fähigkeiten in EBP	2	Ich bin mir sicher, dass ich EBP implementieren kann.“	.769	.706	.756
		4	„Ich bin zuversichtlich, dass ich EBP dort implementieren kann, wo ich arbeite.“			
		5	„Ich glaube, dass ich bei der Implementierung von EBP-Hindernisse überwinden kann.“			
		8	„Ich bin mir sicher, dass ich auf die besten Ressourcen zurückgreifen kann, um EBP zu implementieren.“			

Für die weitere Analyse innerhalb dieser Teilstudie werden die gebildeten Faktoren der *EBP beliefs scale* betrachtet.

Die *EBP implementation scale* zeigt mit einer guten Reliabilität zum Prä-Zeitpunkt ($\alpha_{\text{Prä}} = .74$) und Post-Zeitpunkt ($\alpha_{\text{Post}} = .77$) akzeptable Kennwerte.

Im Rahmen der fächerübergreifenden Erhebung wurde ebenfalls eine Faktorenanalyse der Skala durchgeführt. In der *EBP beliefs scale* wurde Item 11 („Mir sind die EBP-Schritte klar.“) aus der Skala ausgeschlossen. Das Item zeigt insgesamt einen abweichenden Fokus zu den anderen Items der Skala auf. Weiterhin scheint das Item ohne den Zusammenhang zu einer Intervention, welche eben diese Schritte deutlich macht (vgl. Interventionsstudie), nur schwer beantwortbar zu sein. EBP-Schritte werden im Rahmen der Studiengänge bzw. des Praxissemesters nicht behandelt. Entsprechend wurde das Item aus der Skala ausgeschlossen.

Auf Grundlage der Daten des Prä-Zeitpunktes wurde erneut eine Faktorenanalyse der *EBP beliefs scale* durchgeführt. Diese zeigt eine ähnliche Struktur auf, wie sie im Rahmen der Interventionsstudie konstruiert werden konnte. Tabelle 16 zeigt die drei Faktoren. Im Vergleich zur Interventionsstudie konnten den Faktoren vereinzelt weitere Items zu sortiert werden. Die Items 15 und 16 wurden auf Grundlage der erfolgten Faktorenanalyse ausgeschlossen. Diese Items stellen zusammen einen weiteren Faktor dar, welcher jedoch aufgrund der geringen Itemanzahl sowie eines geringen Eigenwertes (1.03) nicht interpretiert wird. Ein möglicher Grund für die Abweichung der beiden Items von den restlichen Items der Skala könnte in der umgekehrten Polung erkannt werden. Während die meisten Items der Skala positiv formuliert sind, (hohe Zustimmung beschreibt hochwertige Überzeugungen), sind diese beiden Items negativ formuliert (hohe Zustimmung beschreibt naive Überzeugungen). Hieraus könnte ein unterschiedliches Antwortverhalten (trotz Umpolung in der Auswertung) entstehen, welches zum Ausschluss der Items innerhalb der Faktorenanalyse führen könnte.

Tabelle 15 – Faktoren der EBP beliefs scale, zugehörige Items sowie Reliabilität der Subskalen

Nr.	Faktor	Nr.	Item	Reliabilität (Cronbachs Alpha) am Prä-Zeitpunkt (N=411)	Reliabilität (Cronbachs Alpha) am Post-Zeit- punkt (N=411)
1	Beliefs gegenüber dem generellen Einfluss von EBP auf Bildung	9	„Ich bin mir sicher, dass die Implementierung von EBP meinen Unterricht verbessert wird.“	.725	.783
		10	„Ich glaube, dass die kritische Beurteilung von Evidenzen ein wichtiger Schritt im Prozess von evidenz-basierter Praxis ist.“		
		12	„Ich bin mir sicher, dass evidenz-basierte Leitlinien Bildung verbessern können.“		
		13	„Ich glaube, dass EBP zu den besten Bildungsergebnissen von Lernenden führt.“		
2	Beliefs gegenüber der Implementation von EBP in der Schule	6	„Ich bin mir sicher, wie Ergebnisse von unterrichtlichen Interventionen gemessen werden können.“	.655	.731
		7	„Ich weiß, wie man EBP so weit umsetzt, dass sich die Praxis ändert.“		
		8	„Ich bin mir sicher, dass ich auf die besten Ressourcen zurückgreifen kann, um EBP zu implementieren.“		
		14	„Ich glaube, dass mein Unterricht evidenz-basiert ist.“		

Nr.	Faktor	Nr.	Item	Reliabilität (Cronbachs Alpha) am Prä-Zeitpunkt (N=411)	Reliabilität (Cronbachs Alpha) am Post-Zeit- punkt (N=411)
3	Beliefs gegenüber per- sönlichen Fähigkeiten in EBP	1	„Ich glaube, dass ich zeiteffizient nach den besten Evidenzen außerhalb des Un- terrichts suchen kann, um unterrichtliche Fragen zu beantworte.“	.725	.822
		2	„Ich bin mir sicher, dass ich EBP imple- mentieren kann.“		
		3	„Ich bin mir sicher, dass ich evidenz-ba- sierte Praxis zeiteffizient implementieren kann.“		
		4	„Ich bin zuversichtlich, dass ich EBP dort implementieren kann, wo ich arbeite.“		
		5	„Ich glaube, dass ich bei der Implemen- tierung von EBP-Hindernisse überwinden kann.“		

Anmerkung. Graue Markierungen zeigen neu hinzugekommene Items im Vergleich zur Interventionsstudie für die einzelnen Faktoren auf.

Im Vergleich zu der Zuordnung aus der Interventionsstudie, sind weitere Items hinzugekommen. Diese Items lassen sich inhaltlich durch die Faktoren interpretieren, konnten jedoch durch uneindeutige Faktorladungen in Teilstudie 2 nicht zugeordnet werden. Item 8 stellt das einzige Item dar, welches einem anderen Faktor zugeordnet wurde. Im Rahmen der Interventionsstudie wurde das Item Faktor 3 zugeordnet, nun Faktor 2. Dies kann einerseits durch die deutliche Faktorladung des Items auf Faktor 3 begründet werden (.615, Abstand zunächst kleinerer Faktorladung ≈ 0.3). Auf der anderen Seite muss das Item inhaltlich analysiert werden. Es wäre hier denkbar, dass das Item zu beiden Faktoren passen könnte. Innerhalb der Interventionsstudie wurde das Item in Faktor 3 den persönlichen Fähigkeiten der Proband*innen zugeordnet. Der inhaltliche Fokus könnte hier als die eigene Fähigkeit zur Suche und Nutzung von Ressourcen beschrieben werden. Innerhalb der fächerübergreifenden Erhebung scheint eine Verschiebung des inhaltlichen Fokus vorzuliegen. Die Zuordnung zu Faktor 2 stellt die tatsächliche Implementation der unterrichtlichen Ressourcen im Unterricht in den Vordergrund. Der Fokus liegt entsprechend nicht auf den eigenen Fähigkeiten zur Suche der Ressourcen, sondern auf den Ressourcen selbst. Es stellt sich also die Frage, ob Ressourcen für die unterrichtliche Implementierung von evidenzbasierter Praxis zur Verfügung stehen. Damit erhält das Item weniger Bezug zu den Proband*innen persönlich, wird dafür aber eher auf die Ebene des Medien-/Materialangebots der Schule verortet. Diese beiden unterschiedlichen Zugänge zu dem Item könnten die unterschiedliche Verortung innerhalb der Faktoren in der Interventionsstudie und fächerübergreifenden Erhebung erklären.

Während eine Faktorenanalyse der *EBP implementation scale* innerhalb der Interventionsstudie nicht möglich war, eignen sich die Daten aus Teilstudie drei für die Durchführung dieser. Sowohl der Bartlett-Test ($\text{Chi-Quadrat}(120)=2839.69, p < .001$), als auch das Kaiser-Meyer-Olkin Kriterium ($\text{KMO} = .927$) weisen auf eine Eignung der Daten zur Faktorenanalyse hin. Die Faktorenanalyse zeigt zwei Faktoren mit einem Eigenwert > 1.0 auf. Zusammengenommen erklären diese Faktoren ca. 51% der Gesamtvarianz. Tabelle 16 zeigt die gebildeten Faktoren inkl. der zugeordneten Items. Vier Items wurden aufgrund von Querladungen auf beide Faktoren von der weiteren Analyse ausgeschlossen.

Tabelle 16 – Faktoren der EBP implementation scale, zugehörige Items sowie Reliabilität der Subskalen

Nr.	Faktor	Nr.	Item „Ich habe in den letzten 8 Wochen x-mal...“	Reliabilität (Cronbachs Alpha) am Post-Zeitpunkt (N=411)
1	Praxisorientierter Umgang mit Evidenzen	1	„... selbst gesammelte Daten mit Kolleg*innen geteilt.“	.695
		6	„... die Praxis aufgrund von Ergebnissen der Lernenden geändert.“	
		12	„... eine schulische Forschungsstudie gelesen und kritisch beurteilt.“	
		16	„... Daten über Probleme von Lernenden gesammelt.“	
		18	„... auf eine Bildungsdatenbank zugegriffen.“	
2	Forschungsorientierter Umgang mit Evidenzen	2	„... die Ergebnisse einer Studie in Form einer Präsentation mit >2 Kolleg*innen geteilt.“	.882
		3	„... evidenzbasierte Praxis-Leitlinien mit einer Kollegin/einem Kollegen geteilt.“	
		5	„... die Unterrichtspraxis durch den Gebrauch einer evidenzbasierten Praxis-Leitlinie oder ein systematisches Review dort verändert, wo ich arbeite.“	
		7	„... eine Bildungsinitiative durch Ergebnisse der Lernenden evaluiert.“	
		9	„... den Einsatz von evidenzbasierter Praxis bei Kolleg*innen beworben.“	
		11	„... die Ergebnisse einer Forschungsstudie mit Lernenden/Eltern geteilt.“	
		13	„... informell mit Kolleg*innen Evidenzen einer Forschungsstudie diskutiert.“	

Ein Vergleich der beiden gebildeten Faktoren zeigt Unterschiede zwischen diesen auf. Faktor 1 fokussiert auf den Umgang mit Evidenzen in einem praktischen, unterrichtlichen Kontext. Merkmale der Lernenden werden hier als Ausgangspunkt der eigenen Praxis beschrieben (vgl. Items 6 & 16). Andere Items innerhalb dieses Faktors beschreiben Aspekte, welche mit der Unterrichtsplanung in Verbindung gebracht werden können. So wird z.B. der Umgang mit Forschungsstudien und Bildungsdatenbanken beschrieben (Items 12 & 16). Beide Tätigkeiten könnten mit der Planung von Unterricht und der dafür notwendigen Recherche in Verbindung gebracht werden. Das letzte verbleibende Item dieses Faktors zielt auf den praxisbezogenen Austausch mit Kolleg*innen ab (Item 1), wie dieser z.B. durch Lehrkräfte-Konferenzen etc. teilweise institutionalisiert ist. Faktor 2 fokussiert hingegen eher auf einen forschungsorientierten Umgang mit Evidenzen. So steht z.B. der Austausch oder die Diskussion von Evidenzen mit verschiedenen Personengruppen im Fokus (z.B. Item 2, 9, 13 ...). Ebenso werden Aspekte der forschungsbasierten Modifikation bestehender Praxis (Item 5) und Evaluation von Evidenzen (Item 7) adressiert.

Die gebildeten Faktoren ließen sich auch vor dem Hintergrund der Unterscheidung zwischen *using evidence* und *establishing evidence* verstehen (Davies, 1999). Demnach könnten durch die Faktoren zwei unterschiedliche Tätigkeiten mit Evidenzen angesprochen werden. Während Faktor 1 eher auf die Unterrichtspraxis und damit die Perspektive der Lehrkräfte im Umgang mit Evidenzen fokussiert, kann die Perspektive von Faktor 2 eher als forschungsbasiert und aus Richtung Forschender beschrieben werden. Die Lehrkräftebildung und das Praxissemester im Besonderen zielt

auf die Verzahnung von beiden Perspektiven ab (Herzig & Wiethoff, 2019). Somit können beide Aktivitätssysteme als relevant für die Lehrkräftebildung beschrieben werden. Die Faktorenstruktur der *EBP implementation scale* kann somit sowohl theoretisch begründet als auch in bestehende Rahmungen zur Lehrkräftebildung verortet werden.

Die *EBP implementation scale* wird im Rahmen der fächerübergreifenden Erhebung lediglich zum Post und follow-up-Zeitpunkt eingesetzt. Auf eine Nutzung am Prä-Zeitpunkt wurde aufgrund der erwarteten sehr geringen Nutzung evidenzbasierter Tätigkeiten sowie dem enormen Umfang der gesamten Evaluationsbefragung zum Praxissemester verzichtet.

6.5.3 Interview

Um tiefergehende Einblicke in die Kompetenzen in evidenzbasierter Praxis der Proband*innen zu erhalten, wurden die quantitativen Instrumente durch ein leitfadengestütztes Interview ergänzt. Aufgrund von anhaltenden Kontaktbeschränkungen zur Bekämpfung der COVID-19-Pandemie konnten die Interviews nur als Online-Interviews mittels Videokonferenz (Zoom) durchgeführt werden. Die Interviews wurden mittels der Aufnahmefunktion des Tools mitgeschnitten und hatten eine durchschnittliche Länge von 33,6 min. Dabei konnten die Proband*innen sich entscheiden, das Kamerasignal auszustellen, da dies im Laufe der weiteren Analyse keinerlei Verwendung gefunden hat. Es wurden Transkripte für die Interviews erstellt, welche anschließend mittels der Software MAXQDA kodiert wurden. Die Daten wurden für die Analyse pseudonymisiert, sodass keine Rückführung auf die Person, jedoch ein *Matching* mit den weiteren Datensätzen der Proband*innen ermöglicht wurde.

Das Interview wurde an den *evidence-use* Mechanismen von Breckon und Dodson (2016) orientiert. Gleichzeitig sollte das Interview Raum bieten, um dem umfassenden und zusammenhängenden Charakter evidenzbasierter Praxis gerecht zu werden. Folglich wurde ein Interviewleitfaden erstellt, welcher als Orientierung innerhalb des Interviews genutzt wurde, jedoch z.B. in der Reihenfolge der Fragen variiert werden konnte. Zu Beginn des Interviews sollten die Proband*innen beschreiben, was sie unter „Evidenzen“ verstehen und welche Rolle diese in ihrem alltäglichen Leben, wie auch der Schule spielen. Gegen Ende des Interviews sollten die Proband*innen die eigene Vorbereitung auf den Umgang mit Evidenzen im Lehrberuf durch ihre Ausbildung bewerten. Innerhalb des Leitfadens (siehe Anhang) wurden jeder Frage weitere Impulsfragen hinzugefügt, um ein möglichst adaptives Interview zu ermöglichen.

Für die Auswertung der Interviews wurde ein an den *evidence-use* Mechanismen von Breckon und Dodson (2016) orientiertes Kodiermanual erstellt. Als Kodiereinheit wurde eine Mindestgröße von einem Satz festgelegt. Dieser kann dabei auch durch mehrere Codes beschrieben werden. Während die Oberkategorien des Kodiermanuals durch die *evidence-use* Mechanismen deduktiv gebildet wurden, wurden Unterkategorien hauptsächlich induktiv ergänzt. Ankerbeispiele dienen dem besseren Verständnis der einzelnen Kategorien. Hier schlägt sich ein explorativer Ansatz der Teilstudie nieder. Das Kodiermanual, welches durch den Autor dieser Arbeit entwickelt wurde, wurde innerhalb des Projektkonsortiums (*RiTE*) diskutiert und adaptiert. Durch die unterschiedlichen nationalen Kontexte innerhalb des *RiTE*-Projekts wurden induktive Ergänzungen auch im Laufe der weiteren Analyse ermöglicht. Dabei sind einige dieser Ergänzungen spezifisch für die beteiligten Länder. Das im Anhang befindliche Kodiermanual zeigt die für die Analyse am Standort Paderborn verwendeten Kategorien auf. Kategorien, welche lediglich Spezifika einzelner Ländern im *RiTE*-Projekt abbilden, sind nicht enthalten.

Um die Güte des Kodiermanuals und der Analyse der Interviews genauer zu untersuchen, wurde für die Interviews am Standort Paderborn eine Inter-coder-Übereinstimmung bestimmt. Dazu wurden 4 Interviews ($\approx 28.6\%$ des Datenmaterials) von einer zweiten Person kodiert, welche in die Projektarbeit am Standort einbezogen war und somit einen Einblick in das *RiTE*-Projekt, wie auch das Konstrukt evidenzbasierter Praxis hat. Eine bestimmte Inter-coder-Übereinstimmung von $\kappa = .77$ verpasst nach Landis und Koch (1977) nur knapp die Benchmark „almost perfect“.

Für Phase 2 der Studie, also die Begleitung der Proband*innen im Referendariat im Sinne einer follow-up-Erhebung, wurde der Interview-Leitfaden sowie das Kodiermanual angepasst. Wie beschrieben, wurden hier Unterrichtsentwürfe oder auch Unterrichtsmitschauen als Grundlage der Interviews verwendet. Entsprechend wurden materialbezogene Fragen ergänzt. Das Material sollte genutzt werden, um an spezifischen Beispielen die Gedanken bzgl. evidenzbasierter Praxis der Proband*innen aufzuzeigen. Das Kodiermanual konnte größtenteils übernommen werden. Lediglich eine Kategorie wurde leicht erweitert: Während in Phase 1 eine Bewertung der eigenen Ausbildung auf den Umgang mit Evidenzen im Lehrberuf erfolgen sollte, schließt die Kategorie innerhalb von Phase 2 auch weitere Professionalisierungselemente (z.B. Seminare der ZfsLs) mit ein. Aufgrund des hohen Deckungsgrades der Kodiermanuals aus beiden Phasen wurde keine erneute Inter-coder-Übereinstimmung ermittelt.

6.5.4 Zusätzlich erhobene Daten

Neben der Datenerhebung mithilfe der beschriebenen Instrumente, konnten zusätzliche Daten gesammelt werden. Dabei handelt es sich um Unterrichtsentwürfe oder Unterrichtsmitschauen, welche Grundlage für die Interviews waren. Dazu wurden die Entwürfe, aber auch die Unterrichtsmitschauen mithilfe von Checklisten genauer untersucht. Ziel war es, das Material vor dem Interview möglichst umfangreich zu erschließen, damit im späteren Interview auf wichtige Aspekte eingegangen werden kann. Die dazu entwickelten Checklisten beinhalten essenzielle Planungsschritte oder deren Umsetzung. Dabei wird auf operationalisierte Aspekte der *evidence-use* Mechanismen (Breckon & Dodson, 2016) abgehoben. Die tiefergehende Analyse der gesammelten zusätzlichen Daten ist nicht Bestandteil dieser Teilstudie. Gleichwohl werden entscheidende Aspekte der gesammelten Daten innerhalb der Interviews thematisiert. Entsprechend finden die Aspekte Berücksichtigung innerhalb der Datenanalyse. Die reine Analyse der Entwürfe bzw. Unterrichtsdurchführungen würde ebenfalls andere methodische Zugänge nötig machen (z.B. Videographien der Unterrichtsstunden), um vertiefte Einblicke in tatsächliche Umsetzung evidenzbasierter Praxis zu erhalten. Es war nicht das Ziel dieser Arbeit auf die Umsetzung evidenzbasierter Praxis im Klassenraum zu fokussieren. Gleichwohl bieten die beschriebenen Interviewgrundlagen Reflexionsanlässe für die Diskussion im Rahmen der Interviews.

6.6 Ergebnisse und Diskussion

6.6.1 Ergebnisse der fächerübergreifenden Befragung

6.6.1.1 Ergebnisse des gekürzten EIDM-ET

Deskriptive Ergebnisse

Innerhalb des gekürzten EIDM-ET konnten die Proband*innen maximal 108 Punkte erlangen. Tabelle 17 gibt einen Überblick über die erreichten Punkte der Proband*innen in Prä und Post-Test. Für fehlende Werte wurde innerhalb der Analysen ein paarweiser Fallausschluss vorgenommen. Fehlende Werte wurden im Datensatz immer dann mit dem Wert ‚99‘ markiert, wenn die

Proband*innen die Aufgabenstellung offensichtlich falsch verstanden haben. Wenn z.B. für Item 3 des gekürzten Instruments keine Strategien zur Eingrenzung der Suche angegeben wurden, sondern konkretisierte Begriffe, wurde der Wert ,99' eingetragen. Dabei könnten jedoch auch weitere Gründe für das Erreichen keiner Punkte vorliegen, welche nicht direkt auf fehlende Kompetenzen zurückzuführen sind (z.B. Testmotivation, etc.). Da diese jedoch zumeist nicht aus der Beantwortung der Items zu erkennen sind, konnte hier keine weitere Unterscheidung vorgenommen werden. Entsprechend können für einige Betrachtungen geringfügige Unterschiede der eingeschlossenen Stichprobe vorliegen.

Tabelle 17 – Deskriptive Statistik für den gekürzten EIDM-ET an Prä- und Post-Zeitpunkt

Testzeitpunkt	N	Minimum	Maximum	M	SD
Prä-Test	411	2.00	85.00	46.07 (42.66%)	16.92
Post-Test	411	0.00	87.00	44.28 (41.00%)	17.43

Die Mittelwerte für Prä- sowie Post-Test liegen unter 50% der maximal erreichbaren Punkte (vgl. Tab. 17). Eine Analyse der erreichten Punktzahlen für die einzelnen Items kann weitere Einblicke in die Leistungen der Proband*innen geben (vgl. Tab. 18).

Tabelle 18 – Deskriptive Statistik für einzelne Items des gekürzten EIDM-ET für Prä- und Post-Zeitpunkt

Items des EIDM-ET	N		M (SD)		Prozentual erreichte Punktzahl	
	Prä	Post	Prä	Post	Prä	Post
Item 1	399	393	7.53 (4.21)	7.19 (4.42)	62.75%	59.92%
Item 2	403	395	11.85 (8.53)	10.71 (8.13)	49.38%	44.63%
Item 3	383	381	17.40 (6.93)	17.13 (6.73)	75.50%	71.38%
Item 4_1	411	411	6.46 (3.16)	6.17 (3.36)	80.75%	77.13%
Item 4_2	411	411	6.01 (3.46)	6.29 (3.29)	75.13%	78.63%
Item 4_3	411	411	4.57 (3.96)	4.40 (3.98)	57.13%	55.00%
Item 5	406	405	8.90 (7.34)	9.25 (7.61)	37.08%	38.54%

Vergleiche

Ein Vergleich der erreichten Punkt der Proband*innen zum Prä- und Post-Zeitpunkt zeigt keine signifikanten Unterschiede auf ($t(410)=1.925$, $p=.055$). Dabei wird das Signifikanzniveau von $p=.05$ nur knapp verfehlt.

Auf Grundlage der im gekürzten EIDM-ET erreichten Punktzahlen wurden Gruppenvergleiche auf Grundlage unterschiedlicher Variablen bestimmt. Dabei sollen die Vergleiche Einblicke in die

Kontextabhängigkeit von EBP-Kompetenzen bieten. Die Vergleiche beziehen dabei die Bachelornote, das Geschlecht, die belegten Studiengänge, Unterrichtsfächer und Begleitforschungsseminare ein.

In einem ersten Schritt wurde ein Einfluss des Geschlechts auf die Leistung im Rahmen des gekürzten EIDM-ET untersucht. Für den Prä-Test konnte ein signifikanter Unterschied bestimmt werden ($F(2, 407)=3.78, p=.024$). Im Rahmen einer durchgeführten Post-Hoc Analyse können lediglich signifikante Unterschiede zwischen Proband*innen welche sich als divers identifizieren und Proband*innen welche sich als männlich oder weiblich identifizieren gefunden werden. Da die Gruppe diverser Proband*innen jedoch nur aus 3 Personen besteht, ist eine Interpretation hier nicht zulässig. Zwischen Proband*innen welche sich als männlich oder weiblich identifiziert haben, konnten keine Unterschiede erkannt werden. Ebenso bestehen keine signifikanten Unterschiede zum Post-Zeitpunkt ($F(2, 407)=1.17, p=.311$).

Unterschiede in der Punktzahl des gekürzten EIDM-ET wurden auch für die Bachelornote analysiert. Dabei ist sowohl für den Prä- ($r=.115, p=.021, N=405$) als auch den Post-Zeitpunkt ($r=.159, p=.001, N=405$) ein Zusammenhang zwischen der Bachelornote und der Punktzahl im gekürzten EIDM-ET zu erkennen. Da die Zunahme der Bachelornote mit einer schlechteren Note einhergeht, muss beschrieben werden, dass Proband*innen mit einer besseren Bachelornote eine niedrigere Punktzahl im gekürzten EIDM-ET zu Prä- und Post-Zeitpunkt erreichen und andersherum.

Auf Grundlage der belegten Studiengänge für das Lehramt an unterschiedlichen Schulformen (vgl. Kap. 7.3) konnten weder zum Prä-Zeitpunkt ($F(5, 405)=1.29, p=.266$), noch zum Post-Zeitpunkt ($F(5, 405)=.700, p=.624$) signifikante Unterschiede bestimmt werden. Ebenso konnten keine signifikanten Unterschiede zwischen den Unterrichtsfächern (sortiert in Fächerbündel) am Prä- ($F(16, 379)=1.43, p=.123$) und Post-Zeitpunkt ($F(16, 379)=1.00, p=.452$) erkannt werden.

Ein weiterer Gruppenvergleich wurde auf der Wahl der Begleitforschungsseminare im Rahmen des Praxissemesters aufgebaut. Für die in Fächerbündel sortierten Begleitforschungsseminare ergibt sich zum Prä-Zeitpunkt ein signifikanter Unterschied ($F(4, 341)=3.55, p=.007$). Entsprechend erhalten die Proband*innen innerhalb der Begleitforschungsseminare im religiös-philosophischen Bündel einen Mittelwert von $M=38.48$ Punkten ($SD=16.49$), während die Proband*innen der anderen Begleitforschungsseminare Mittelwerte zwischen $M=47.49$ - 48.72 Punkten erreichen. Im Rahmen von durchgeführten Post-Hoc Tests mit Bonferroni-Korrektur konnten signifikante Unterschiede für das religiös-philosophische Fächerbündel zu allen anderen Fächerbündeln mit Ausnahme des Bündels „Sachunterricht“ bestimmt werden.

Tabelle 19 gibt einen Überblick über die Unterschiedsmaße sowie die Effektstärke. Dabei wurde die Effektstärke nach Cohen durch ungepaarte t-Test für die Gruppen errechnet, für welche durch den Post-Hoc Test signifikante Unterschiede bestimmt wurden.

Tabelle 19 – Mittelwertsunterschiede im gekürzten EIDM-ET und Effektstärken für Fächerbündel

Fächerbündel 1	Fächerbündel 2	Mittelwerts- unterschied	<i>p</i>	Cohen's <i>d</i>
Religiös-philosophisches Bündel	Sprachlich-literarisch-künstlerisches Bündel	-9.91	.013	.598
	Mathematisch-naturwissenschaftlich-technisches Bündel	-10.01	.012	.603
	Gesellschaftswissenschaftliches Bündel	-9.01	.011	.565

Anmerkung. ANOVA mit Bonferroni Post-Hoc Test, $F(4, 341)=3.55$, $p=.007$, Die Effektstärken wurden aus ungepaarten t-Test für die einzelnen Gruppen errechnet.

Für die Unterschiede zwischen dem religiös-philosophischen Bündel zu den anderen Bündeln konnten mittlere Effektstärken ($d > 0.5$) bestimmt werden. Für den Post-Zeitpunkt ergeben sich keine Unterschiede bzgl. der gewählten Begleitforschungsseminare. Dabei deutet dies in Kombination mit dem ausbleibenden Haupteffekt der Zeit nicht auf einen größeren Lernzuwachs hin für die Proband*innen innerhalb der Begleitseminare des religiös-philosophischen Bündels hin. Der Mittelwert für diese Proband*innen am Post-Zeitpunkt liegt mit $M=40.20$ ($SD=15.17$) nur geringfügig höher.

6.6.1.2 Ergebnisse der *EBP beliefs scale*

Deskriptive Ergebnisse

Die *EBP beliefs scale* fragt Selbsteinschätzungen der Bedeutung evidenzbasierter Praxis sowie der eigenen Fähigkeiten in dieser ab. Dabei erfolgt die Selbsteinschätzung auf Grundlage einer 5-stufigen Likert-Skala mit 1=stark negative *beliefs* gegenüber evidenzbasierter Praxis bis 5=stark positive *beliefs*. stellt ein Skalenwert von 5.00 stark positive und ein Skalenwert von 0.00 stark negative *beliefs* gegenüber evidenzbasierter Praxis dar. Tabelle 20 gibt einen Überblick über die Ausprägungen der Selbsteinschätzungen der Proband*innen zu Prä- und Post-Zeitpunkt, getrennt für die bestimmten Faktoren der Skala (vgl. Kap. 6.5.2). Für alle Sub-Skalen liegt der Mittelwert zum Prä- und Post-Zeitpunkt leicht über dem Skalenmittelwert ($N=411$).

Tabelle 20 – Deskriptive Statistik für *EBP beliefs scale* zu Prä- und Post-Zeitpunkt

Testzeitpunkt und Faktoren	<i>M</i>	<i>SD</i>
Beliefs_EBP_Bildung - Beliefs gegenüber dem generellen Einfluss von EBP auf Bildung		
Prä-Zeitpunkt	3.66	.62
Post-Zeitpunkt	3.48	.70
Beliefs_EBP_Impl_Schule - Beliefs gegenüber der Implementation von EBP in der Schule		
Prä-Zeitpunkt	2.85	.64
Post-Zeitpunkt	3.03	.69
Beliefs_EBP_Fähigkeiten - Beliefs gegenüber persönlichen Fähigkeiten in EBP		
Prä-Zeitpunkt	3.22	.61
Post-Zeitpunkt	3.21	.71

Vergleiche

Eine Untersuchung von Unterschieden zwischen Prä- und Post-Zeitpunkt zeigt signifikante Unterschiede für *Beliefs_EBP_Bildung* ($t(410)=4.891$, $p<.001$, $d=.24$) und *Beliefs_EBP_Impl_Schule* ($t(410)=-4.527$, $p<.001$, $d=.22$) auf. Dabei können kleine Effekte ($d>0.2$) berichtet werden. Während für *Beliefs_EBP_Bildung* eine Abnahme des Mittelwerts zwischen Prä- und Post-Zeitpunkt erkennbar ist, steigt der Mittelwert für *Beliefs_EBP_Impl_Schule*. Für *Beliefs_EBP_Fähigkeiten* können keine signifikanten Unterschiede bestimmt werden.

Für eine Analyse bzgl. Unterschieden auf Grundlage des angegebenen Geschlechts können Unterschiede für die *Beliefs_EBP_Impl_Schule* ($F(2, 407)=6.840$, $p=.001$) und *Beliefs_EBP_Fähigkeiten* ($F(2, 407)=3.338$, $p=.036$) zum Prä-Zeitpunkt beschrieben werden. Erneut treten hier Unterschiede zwischen Proband*innen die sich als divers identifizieren mit männlichen sowie weiblichen Proband*innen auf. Diese Unterschiede basieren dabei erneut auf lediglich $N=3$ Proband*innen mit der Geschlechtsausprägung divers und sind folglich nicht interpretierbar.

Eine Untersuchung des Zusammenhangs der Bachelornote mit den Sub-Skalen der *EBP beliefs scale* zeigt signifikante Korrelationen zwischen der Bachelornote mit *Beliefs_EBP_Bildung_Prä* ($r=.251$, $p<.001$, $N=405$) sowie *Beliefs_EBP_Fähigkeiten_Post* ($r=.174$, $p<.001$, $N=405$) auf. Da bei einer Bachelornote die Zunahme der absoluten Zahl mit einer Verschlechterung einhergeht, muss hier beschrieben werden, dass eine bessere Bachelornote mit einem schlechteren Abschneiden auf den beiden Subskalen im Zusammenhang steht.

In Anlehnung an das Vorgehen zur Auswertung des gekürzten EIDM-ET wurden auch für die *EBP beliefs scale* Gruppenvergleiche bestimmt. Auf Grundlage einer Unterscheidung bzgl. der belegten Studiengänge kann für keinen Sub-Skala ein signifikanter Unterschied bestimmt werden (vgl. Tab. 21).

Tabelle 21 – Ergebnisse einer ANOVA zur Untersuchung von Unterschieden der Sub-Skalen der EBP beliefs scale für verschiedene Studiengänge

Testzeitpunkt	F(df)	p
Beliefs_EBP_Bildung - Beliefs gegenüber dem generellen Einfluss von EBP auf Bildung		
Prä-Zeitpunkt	$F(5, 405)=1.436$.210
Post-Zeitpunkt	$F(5, 405)=.286$.921
Beliefs_EBP_Impl_Schule - Beliefs gegenüber der Implementation von EBP in der Schule		
Prä-Zeitpunkt	$F(5, 405)=.447$.815
Post-Zeitpunkt	$F(5, 405)=1.169$.324
Beliefs_EBP_Fähigkeiten - Beliefs gegenüber persönlichen Fähigkeiten in EBP		
Prä-Zeitpunkt	$F(5, 405)=.933$.459
Post-Zeitpunkt	$F(5, 405)=1.339$.247

Auch für eine Untersuchung von Unterschieden der Sub-Skalen der *EBP beliefs scale* auf Grundlage der Fächerbündel-Kombinationen konnten keine signifikanten Unterschiede gefunden werden (vgl. Tab. 22).

Tabelle 22 – Ergebnisse einer ANOVA zur Untersuchung von Unterschieden der Faktoren der *EBP beliefs scale* für verschiedene Fächerbündel-Kombinationen

Testzeitpunkt	F(df)	p
Beliefs_EBP_Bildung - Beliefs gegenüber dem generellen Einfluss von EBP auf Bildung		
Prä-Zeitpunkt	F(16, 379)= .636	.855
Post-Zeitpunkt	F(16, 379)=1.481	.103
Beliefs_EBP_Impl_Schule - Beliefs gegenüber der Implementation von EBP in der Schule		
Prä-Zeitpunkt	F(16, 379)= .966	.494
Post-Zeitpunkt	F(16, 379)= .983	.475
Beliefs_EBP_Fähigkeiten - Beliefs gegenüber persönlichen Fähigkeiten in EBP		
Prä-Zeitpunkt	F(16, 379)= .791	.696
Post-Zeitpunkt	F(16, 379)=1.311	.187

Ebenso zeigen sich keine signifikanten Unterschiede für die Unterscheidung auf Grundlage der belegten Begleitforschungsseminare zum Prä- und Post-Zeitpunkt (vgl. Tab 23).

Tabelle 23 – Ergebnisse einer ANOVA zur Untersuchung von Unterschieden der Faktoren der *EBP beliefs scale* für verschiedene belegte Begleitforschungsseminare

Testzeitpunkt	F(df)	p
Beliefs_EBP_Bildung - Beliefs gegenüber dem generellen Einfluss von EBP auf Bildung		
Prä-Zeitpunkt	F(4, 341)= .146	.964
Post-Zeitpunkt	F(4, 341)= .945	.438
Beliefs_EBP_Impl_Schule - Beliefs gegenüber der Implementation von EBP in der Schule		
Prä-Zeitpunkt	F(4, 341)= .346	.847
Post-Zeitpunkt	F(4, 341)= .116	.977
Beliefs_EBP_Fähigkeiten - Beliefs gegenüber persönlichen Fähigkeiten in EBP		
Prä-Zeitpunkt	F(4, 341)= .403	.806
Post-Zeitpunkt	F(4, 341)= .931	.446

6.6.1.3 Ergebnisse der *EBP implementation scale*

Deskriptive Ergebnisse

Für die *EBP implementation scale* steht wiederum ein 5-stufiges Antwortformat zur Verfügung. Dabei werden Selbsteinschätzungen abgefragt, welche jedoch nicht auf *beliefs* der Proband*innen abzielen, sondern die Häufigkeit von evidenzbasierten Tätigkeiten abfragt. Die deskriptive Statistik (N=411) für die beiden Faktoren kann Tabelle 24 entnommen werden.

Tabelle 24 – Deskriptive Statistik für die Faktoren der *EBP implementation scale* für den Post-Zeitpunkt

Faktor der <i>EBP implementation scale</i>	<i>M</i>	<i>SD</i>
Implement_EBP_Praxis – Praxisorientierter Umgang mit Evidenzen	2.12	.74
Implement_EBP_Forschung – Forschungsorientierter Umgang mit Evidenzen	1.45	.61

Vergleiche

Da die *EBP implementation scale* nur am Post-Zeitpunkt erhoben wurde, können keine Vergleiche zwischen Prä- und Post-Zeitpunkt untersucht werden.

Eine Vergleich für das Geschlecht zeigt einen signifikanter Unterschied für *Implement_EBP_Forschung* ($F(2, 392)=5.073, p=.007$). Innerhalb der Post-Hoc Analyse erreicht lediglich der Unterschied zwischen Proband*innen männlichen und weiblichen Geschlechts das Signifikanzniveau. Dabei erreichen männliche Proband*innen einen signifikant höheren Mittelwert ($M=1.58, SD=.69$) als weibliche Proband*innen ($M=1.40, SD=.56$). ($p=.011, d=.35$). Die beiden Gruppen männlicher ($N=103$) und weiblicher Proband*innen ($N=304$) erlauben hier eine belastbare Interpretation. In diesem Fall ist kein Unterschied zu Studierenden mit diversem Geschlecht zu erkennen.

Weiterhin wurden Zusammenhänge des Skalenwerts auf der *EBP implementation scale* basierend auf der Bachelornote untersucht. Im Rahmen einer Korrelationsanalyse konnten keine Zusammenhänge zwischen den beiden Konstrukten gefunden werden.

Auch für die *EBP implementation scale* wurden Gruppenunterschiede berechnet. Für die Unterscheidung nach den unterschiedlichen Studiengängen für unterschiedliche Schulformen zeigen sich sowohl für *Implement_EBP_Praxis* ($F(5, 405)=3.384, p=.001$) als auch *Implement_EBP_Forschung* ($F(5, 405)=4.916, p<.001$) signifikante Unterschiede. Für *Implement_EBP_Praxis* kann ein signifikanter Unterschied zwischen den Studiengängen „Lehramt an Grundschulen“ und „Lehramt an Berufskollegs“ erkannt werden ($M_G=1.95, SD_G=.70, M_{BK}=2.33, SD_{BK}=.75, p=.044, d=.53$). Ein Unterschied zwischen den Studiengängen „Lehramt an Grundschulen“ und „Lehramt an Gymnasien und Gesamtschulen“ ($M_G=1.95, SD_G=.70, M_{GyGe}=2.22, SD_{GyGe}=.76, p=.053$) verpasst das Signifikanzniveau von $p=.05$ nur knapp.

Für *Implement_EBP_Forschung* wird der Unterschied zwischen „Lehramt an Grundschulen“ und „Lehramt an Gymnasien und Gesamtschulen“ signifikant ($M_G=1.29, SD_G=.45, M_{GyGe}=1.52, SD_{GyGe}=.60, p=.035, d=.43$). Ebenso zeigt sich ein signifikanter Unterschied zwischen „Lehramt an Grundschulen“ und „Lehramt an Berufskollegs“ ($M_G=1.29, SD_G=.45, M_{BK}=1.61, SD_{BK}=.82, p=.009, d=.59$). Weiterhin zeigt sich ein Unterschied zwischen „Lehramt für sonderpädagogische Förderung“ und „Lehramt für Haupt-, Real-, Sekundar- und Gesamtschulen“ ($M_{Sp}=1.20, SD_{Sp}=.54, M_{HRSGe}=1.55, SD_{HRSGe}=.59, p=.045, d=.62$). Einen Unterschied zwischen dem „Lehramt für sonderpädagogische Förderung“ besteht auch zum „Lehramt an Gymnasien und Gesamtschulen“ ($M_{GyGe}=1.52, SD_{GyGe}=.60, p=.038, d=.06$) sowie „Lehramt an Berufskollegs“ ($M_{BK}=1.61, SD_{BK}=.82, p=.007, d=.13$).

In den beschriebenen Gruppenvergleichen auf Grundlage der unterschiedlichen Studiengänge sind Unterschiede, teils mit mittelstarken Effekten zu erkennen. Die Studiengänge „Lehramt an Grundschulen“ sowie „Lehramt für sonderpädagogische Förderung“ zeigen dabei signifikant geringere Mittelwerte auf. Dabei müssen jedoch die eher kleinen Mittelwertsunterschiede für die einzelnen Analysen beachtet werden.

Da die vorgefunden Unterschiede auf Unterschiede zwischen den verschiedenen Schulstufen hinweisen, wurden diese ebenfalls auf Unterschiede untersucht. Dazu wurden folgende Schulstufen gebildet: Primarstufe (Lehramt an Grundschulen und Lehramt für sonderpädagogische Förderung), Sekundarstufe 1 (Lehramt an Haupt-, Real-, Sekundar- und Gesamtschulen), Sekundarstufe 2 (Lehramt an Gymnasien und Gesamtschulen) und berufliche Bildung (Lehramt an Berufskollegs). Sowohl für *Implement_EBP_Praxis* ($F(3, 407)=5.609, p=.001$) als auch für *Implement_EBP_Forschung* ($F(3, 407)=7.506, p<.001$) wurden signifikante Unterschiede gefunden. Signifikante Unterschiede für *Implement_EBP_Praxis* konnte mittels Post-Hoc Analyse zwischen der Primarstufe ($M_{\text{Primar}}=1.93, SD_{\text{Primar}}=.70$) und der Sekundarstufe 2 ($M_{\text{Sek_II}}=2.22, SD_{\text{Sek_II}}=.76, p=.005, d=.39$) sowie der beruflichen Bildung ($M_{\text{BK}}=2.33, SD_{\text{BK}}=.75, p=.004, d=.56$) gefunden werden. Für *Implement_EBP_Forschung* konnten signifikante Unterschiede zwischen der Primarstufe ($M_{\text{Primar}}=1.26, SD_{\text{Primar}}=.48$) und der Sekundarstufe 1 ($M_{\text{Sek_I}}=1.55, SD_{\text{GyGe}}=.59, p=.006, d=.56$), der Sekundarstufe 2 ($M_{\text{Sek_II}}=1.52, SD_{\text{Sek_II}}=.60, p=.001, d=.48$) sowie der beruflichen Bildung ($M_{\text{BK}}=1.61, SD_{\text{BK}}=.82, p=.001, d=.59$) aufgezeigt werden. Die jeweiligen Effektstärken erreichen sich wiederum aus ungepaarten t-Tests für die einzelnen Gruppen.

Eine Untersuchung auf Unterschiede zwischen den unterschiedlichen Fächerbündel-Kombinationen zeigt einen signifikanten Unterschied für Faktor 2 der *EBP implementation scale* auf ($F(15, 379)=1.839, p=.028$). Für die Durchführung einer Post-Hoc Analyse muss die Fächerbündel-Kombination „Mathematisch-naturwissenschaftlich-technisch & Sachunterricht“ aus der Analyse ausgeschlossen werden. Innerhalb der Stichprobe gibt es nur eine*n Probanden*in dieser Fächerbündel-Kombination. Entsprechend kann dieses Fächerbündel-Kombination innerhalb der Analyse nicht berücksichtigt werden. Im Rahmen der Post-Hoc Analyse erreicht keine Fächerbündel-Kombination ein Signifikanzniveau von $p=.05$.

Eine Untersuchung von Unterschieden auf Basis des belegten Begleitforschungsseminars zeigt weder für *Implement_EBP_Praxis* ($F(4, 341)=.617, p=.651$), noch für *Implement_EBP_Forschung* ($F(4, 341)=.583, p=.675$) signifikante Unterschiede auf.

6.6.2 Diskussion der Ergebnisse der fächerübergreifenden Befragung

Im Rahmen der vorliegenden Studie konnten Einblicke in die Kompetenzen in evidenzbasierter Praxis sowie Einstellungen gegenüber dieser von angehenden Lehrkräften unterschiedlicher Fächer gewonnen werden.

Mit Blick auf den gekürzten EIDM-ET muss festgestellt werden, dass die Kompetenzen der Proband*innen eher schwach ausgeprägt sind (vgl. Kap. 6.6.1.1). Die Proband*innen erreichen im Prä-Test nur unter 50% der maximalen Gesamtpunktzahl. Dieser Befund konnte auf Grundlage bestehender Literatur so erwartet werden (Bauer & Kollar, 2023; Hetmanek et al., 2015). Bauer und Kollar (2023) fassen mehrere mögliche Begründungsansätze für diese Befunde zusammen. So könnten unter anderem Überzeugungen zur Nützlichkeit von Evidenzen sowie bestehender Zeitmangel Gründe für gering ausgeprägte Kompetenzen darstellen. Im Zusammenhang der vorliegenden Studie müssen hier Prä- und Post-Zeitpunkt getrennt voneinander betrachtet werden.

Die Kompetenzen der Proband*innen zum Prä-Zeitpunkt können auf die bereits absolvierten Bestandteile des Lehramtsstudiums zurückgeführt werden. Im Rahmen dieses scheint das explizite Thematisieren von Evidenzen bzw. die Förderung im Umgang mit Evidenzen eine eher untergeordnete Rolle zu spielen, sodass die Proband*innen nur grundlegende Kompetenzen im Umgang mit Evidenzen aufbauen konnten. Dies könnte z.B. an einer fehlenden Thematisierung forschungsmethodologischen Wissens als Bestandteil evidenzbasierter Praxis im Rahmen des

Lehramtsstudiums begründet werden (Stelter & Miethe, 2019). Eine besondere Rolle für die Einsicht in das forschungsmethodologische Wissen der Proband*innen nehmen Item 4 und 5 ein. Item 4 fragt grundlegende forschungsmethodologische Begriffe wie ‚Kontrollgruppe‘ in einem *Multiple-choice* Format ab. Die Proband*innen schneiden für dieses Item überdurchschnittlich ab. Item 5 fokussiert auf die Validität von Befunden und fällt unterdurchschnittlich aus. Die Proband*innen sollen hier Merkmale von Befunden angeben, die zur Bewertung der Validität notwendig sind. Auf Grundlage der beiden beschriebenen Items kann das forschungsmethodologische Wissen am Prä-Zeitpunkt als eher schwach ausgeprägt beschrieben werden. Zwar scheinen grundlegende Definitionen von forschungsmethodologischen Begriffen bekannt zu sein (Item 4), die tiefergehende Anwendung des Gütekriteriums Validität auf Befunde aus Studien scheint jedoch weniger gut möglich zu sein (Item 5). Im Vergleich könnte Item 5 als anspruchsvoller für die Proband*innen beschrieben werden, da hier Bewertungsaspekte mit einfließen. Eine Begründung für die eher schwach ausgeprägten Kompetenzen könnte in dem von Stelter und Miethe (2019) beschriebenen Mangel an der Thematisierung von Forschungsmethoden im Lehramtsstudium erkannt werden.

Im Vergleich zu den dargestellten Items fokussieren Items 1-3 eher auf der Generierung einer Fragestellung und eines Prozesses zur Recherche nach passenden Evidenzen zu dieser. Damit steht hier weniger das forschungsmethodologische Wissen der Proband*innen als mehr ihre Kompetenzen im Forschungsprozess bzw. einem Ausschnitt davon im Vordergrund. Vor allem in Item 3 wird dabei der Umgang mit den Evidenzen fokussiert, indem diese auf Grundlage des jeweiligen Mediums bewertet werden. Vor allem Item 2 zeigt Schwächen in der Bewertung von Quellentypen auf. Auch hier könnte eine mangelnde Thematisierung von Quellen sowie deren Bewertung im Lehramtsstudium vermutet werden (Stelter & Miethe, 2019). Damit werden auch Schwächen im praktischen Umgang mit Evidenzen bzw. der Suche nach diesen bei den Proband*innen deutlich.

Im Rahmen des Post-Test können nun auch Kompetenzzuwächse im Kontext des Praxissemesters in die erreichte Punktzahl eingehen. Wie beschrieben, stellt vor allem das Begleitforschungsseminar eine Lerngelegenheit für den Umgang mit Evidenzen, wie auch forschungsmethodologische Aspekte im Praxissemester dar. Die Durchführung eines eigenen Studienprojekts kann dabei an die Items des gekürzten EIDM-ET angegliedert werden. Der durch Items 1-3 beschriebene Prozess der Findung einer Forschungsfrage, wie auch der Recherche nach Literatur und deren Bewertung auf Grundlage der jeweiligen Medien, stellt einen essentiellen Teil der Durchführung eines Studienprojekts dar. Je nach individueller Durchführung des Studienprojekts können dabei unterschiedliche Forschungsmethoden (vgl. Item 4) sowie Validitätsaspekte (vgl. Item 5) eine wichtige Rolle spielen. Ein vermuteter Kompetenzzuwachs ist auf Grundlage des Mittelwerts für den Post-Test nicht zu erkennen. Die Durchführung eines eigenen Studienprojekts hätte vor allem bei diesen Items, welche den Forschungsprozess und Umgang mit Evidenzen fokussieren eine Steigerung erwarten lassen. Durch die Betrachtung von Item 5 könnten die Proband*innen ihre Kompetenzen in der Anwendung von Validitätsargumenten für Studienbefunde leicht gestärkt haben.

Vor dem Hintergrund der beschriebenen Ergebnisse zum Post-Zeitpunkt sowie deren Vergleich zum Prä-Zeitpunkt muss das Praxissemester als Lerngelegenheit für den Aufbau von Kompetenzen in evidenzbasierter Praxis als nicht effektiv beschrieben werden. Die Ergebnisse für *Beliefs_EBP_Impl_Schule* lassen eine leichte Wirksamkeit der Intervention erkennen. Trotz des fehlenden Kompetenzzuwachses im Rahmen des gekürzten EIDM-ET schätzen sich die Proband*innen entsprechend als kompetenter im Umgang mit Evidenzen im Unterricht ein. Dies könnte als Hinweis verstanden werden, dass innerhalb des Praxissemesters durchaus Lerngelegenheiten zum

Umgang mit Evidenzen vorhanden sind. Diese haben scheinbar einen größeren Einfluss auf die Selbsteinschätzung der Proband*innen als auf den tatsächlichen Kompetenzzuwachs.

Wie beschrieben, stellt das Begleitforschungsseminar eine besondere Lerngelegenheit für den Aufbau von Kompetenzen in evidenzbasierter Praxis innerhalb des Praxissemesters dar. Unterschiede für die erreichte Punktzahl im EIDM-ET zum Prä-Zeitpunkt konnten so auch für die Wahl des Begleitforschungsseminars festgestellt werden. Dabei zeigen sich Unterschiede zwischen den Begleitforschungsseminaren innerhalb des religiös-philosophischen Fächerbündels zu allen weiteren Fächerbündeln außer dem Bündel Sachunterricht. Dabei schneiden Studierende innerhalb des religiös-philosophischen Bündels zum Prä-Zeitpunkt signifikant schlechter im gekürzten EIDM-ET ab als Studierende der anderen Fächerbündel. Eine mögliche Begründung könnte hier im Evidenzverständnis vermutet werden, welches jedoch im Rahmen der vorliegenden Studie nicht abgefragt wurde. In Zusammenhang mit der Untersuchung epistemologischer Überzeugungen von Urhahne und Kremer (2023) könnten die Begleitforschungsseminare des religiös-philosophischen Fächerbündels als *soft-sciences* (Biglan, 1973) betrachtet werden. Für diese konnten die Autor*innen epistemologische Überzeugungen in der Dimension *Source of knowledge* ausmachen, welche eher auf das Vertrauen in Autoritäten fokussieren. Ebenso kann in der Dimension *Certainty of knowledge* ein Schwerpunkt auf die Stabilität von Wissen erkannt werden. Diese Überzeugungen könnten einen direkten Einfluss auf die Beantwortung der Items des gekürzten EIDM-ET haben. So könnte z.B. ein Vertrauen auf Autoritäten einen Einfluss auf die Angabe von Validitätsargumenten für Befunde haben (vgl. Item 5). Vertrauen die Proband*innen auf Autoritäten, könnten weitere Validitätsargumente aus deren Perspektive nicht notwendig sein.

In Unterscheidung zu den Ergebnissen von Urhahne und Kremer (2023) können jedoch keine Unterschiede zwischen anderen Fächerbündeln hergestellt werden. Dabei muss erneut hervorgehoben werden, dass die Autor*innen die Proband*innen auf Grundlage des erst-gewählten Unterrichtsfachs gebündelt haben, während hier eine Bündelung auf Grundlage des gewählten Begleitforschungsseminars berichtet wurde. Während auf Grundlage der belegten Fächerbündel-Kombinationen keine Unterschiede zu erkennen sind, treten durch die individuelle Wahl des Begleitforschungsseminars die beschriebenen Unterschiede auf. Dabei kann die Wahl der jeweiligen Begleitforschungsseminare auf verschiedenen Begründungen basieren. Neben der Begründung durch verschiedene Rahmenbedingungen der Seminare z.B. bzgl. der zu erbringenden Prüfungsleistung könnten auch Aspekte epistemologischer Überzeugungen bzw. evidenzbasierter Praxis eine Rolle spielen. Sofern akademische Überzeugungen der Proband*innen mit domänenspezifischen Überzeugungen für ein Begleitforschungsseminar übereinstimmen (Muis et al., 2006), könnten die Proband*innen dieses Seminar eher wählen. Im Rahmen der vorliegenden Studie können jedoch keine Einblicke in die Begründung der Wahl der einzelnen Begleitforschungsseminare erhalten werden. Eine Erklärung der auftretenden Unterschiede auf Grundlage der jeweiligen Wahl des Begleitforschungsseminars kann damit hier nicht abschließend geklärt werden. In weiteren Untersuchungen könnte die Motivation zur Wahl bestimmter Begleitforschungsseminare weitere Einblicke in die hier beschriebenen Unterschiede geben.

Mögliche Gründe für die fehlende Effektivität des Praxissemesters zur Förderung von EBP-Kompetenzen könnten in einem möglichen Rollenkonflikt der Proband*innen erkannt werden. Im Rahmen des Praxissemesters sollen die Proband*innen einerseits einen Einblick in die Unterrichtspraxis erhalten, andererseits das forschende Lernen in seiner Verbindung aus Theorie und Praxis als Bestandteil der eigenen Profession erkennen (Herzig & Wiethoff, 2019). Im Kontext evidenzbasierter Praxis könnte hier die Unterteilung in *using evidence* und *establishing evidence* erkannt werden (Davies, 1999). Dadurch könnten zwei unterschiedliche Rollenprofile der Proband*innen

angesprochen werden: Einerseits könnte eine Lehrkraft-Perspektive auf das Praxissemester eingenommen werden. Die Proband*innen sind hier in die Unterrichtspraxis eingebunden. Im Sinne evidenzbasierter Praxis müssen dazu Evidenzen recherchiert und mit bestehenden Erfahrungen verknüpft werden. Dazu ist ein grundsätzliches Verständnis von der Relevanz wissenschaftlicher Evidenzen für unterrichtliche Entscheidungen notwendig (Bauer & Kollar, 2023).

Innerhalb dieser Verknüpfung von Theorie und Praxis erkennen Bauer und Kollar (2023) die Gefahr lediglich auf unterstützende Evidenz zurückzugreifen. Im Sinne eines *what-work's* Ansatzes könnten Lehrkräfte lediglich nach unterstützenden Studien für eigene Annahmen suchen. Dabei stellen widerlegende Evidenzen ebenfalls eine wichtige Grundlage für unterrichtliche Entscheidungen dar. Sollten keine unterstützenden Evidenzen gefunden werden, könnte ein Konflikt zu eigenen oder tradierten Erfahrungen entstehen, welcher die generelle Nutzung wissenschaftlicher Evidenz im Sinne eines *using evidence* Ansatzes hemmen könnte.

Diese Perspektive unterstellt die generelle Nutzung von Evidenzen für Unterrichtsplanung im Sinne des *using evidence* Ansatzes. Innerhalb der *EBP implementation scale* gaben die Proband*innen an, häufiger mit Evidenzen aus einer praxisorientierten Perspektive umgegangen zu sein. Dabei bleibt die generelle Anzahl der Ausführung der Tätigkeiten jedoch eher gering. Entsprechend kann den Proband*innen der Umgang mit Evidenzen im Sinne des *using evidence* Ansatzes nur bedingt unterstellt werden.

Andererseits könnte aus einer Forschenden-Perspektive im Sinne des *establishing evidence* Ansatzes die Generierung wissenschaftlicher Evidenz im Rahmen des Praxissemesters als herausfordernd wahrgenommen werden. Im Rahmen von Kap. 2.4 wurden einige Herausforderungen beschrieben, welche für die Generierung von Evidenzen im Bildungssektor bestehen. Diese Vermutung kann durch die seltenere Nutzung von Evidenzen aus forschungsorientierter Perspektive im Rahmen der *EBP implementation scale* unterstützt werden. Aus einer Forschungsperspektive kann die Bedeutung von Evidenzen für unterrichtliche Entscheidungen klar argumentiert werden. So könnte hier ein Rollenkonflikt der Proband*innen entstehen. Einerseits wird die Relevanz von Evidenzen anerkannt (aus Forschenden-Perspektive), andererseits könnte eine fehlende Passung zu eigenen und tradierten Erfahrungen kombiniert mit wenigen unterstützenden Evidenzen (aus Lehrkraft-Perspektive) die Nutzung von Evidenzen eher hemmen. Im Rahmen des Praxissemesters sollen beide Perspektiven von Praxis und Forschung/Theorie zusammengeführt werden, könnten jedoch eher als konkurrierend wahrgenommen werden. Diese wahrgenommene Konkurrenz könnte auch in der institutionalisierten Zuordnung der Perspektivschwerpunkte an einzelne Lernorte (Lehrkraft-Perspektive an Schule und ZfsL, Forschenden-Perspektive an Universität) unterstützt werden. Weitergehende Begründungen sind jedoch auf Grundlage der bestehenden Ergebnisse nicht abzuleiten.

Bauer und Kollar (2023) argumentieren für einen Kulturwechsel, welcher die Bedeutung wissenschaftlicher Evidenz für unterrichtliche Entscheidungen bei Lehrkräften deutlicher in den Fokus setzt. Diese Überzeugungen zum Wert evidenzbasierter Praxis für unterrichtliche Entscheidungen können innerhalb der Sub-Skala Beliefs_EBP_Bildung der *EBP beliefs scale* („Beliefs gegenüber dem generellen Einfluss von EBP auf Bildung“) erkannt werden. Durch das Praxissemester wurden folglich Überzeugungen bzgl. des positiven Einflusses evidenzbasierter Praxis auf Bildungsprozesse gestärkt. Wenngleich dieser Effekt nicht den durch Bauer und Kollar (2023) geforderten Kulturwechsel darstellt, können leichte Veränderungen erkannt werden. Diese könnten dabei auf die unterrichtliche Praxis sowie deren Vernetzung mit wissenschaftlichen Evidenzen im Praxissemester zurückgeführt werden. Dabei schätzen die Proband*innen den Einfluss des Praxissemesters auf deren Kompetenz in evidenzbasierter Praxis positiver ein (vgl. *Beliefs_EBP_Impl_Schule*), als

dies im gekürzten EIDM-ET zu erkennen ist. Auf Grundlage der vorliegenden Ergebnisse dieser Studie könnte der beschriebene Konflikt eine Begründung für die fehlende Kompetenzentwicklung der Proband*innen im Praxissemester darstellen. Dabei können die Ergebnisse, nicht zuletzt aufgrund der verbesserungswürdigen internen Konsistenz des gekürzten EIDM-ET, nur als Hinweise für diese Argumentation betrachtet werden. Vor allem die Betrachtung der Einzelitems zeigt jedoch, dass neben forschungsmethodologischem Wissen auch generelle Prozessschritte (z.B. Bildung einer Forschungsfrage) sowie die Bewertung von Befunden (z.B. bzgl. der Qualität der Quelle oder vorhandenen Validitätsargumenten) für die Proband*innen Herausforderungen darstellen.

In Bezug auf Forschungsfrage I.a wurde in der vorliegenden Studie untersucht, inwiefern Unterschiede in der Kompetenz in evidenzbasierter Praxis für Proband*innen unterschiedlicher Unterrichtsfächer vorliegen. Auf Grundlage der Ergebnisse können keine signifikanten Unterschiede zwischen verschiedenen Fächerbündeln erkannt werden. Weder zum Prä- noch zum Post-Zeitpunkt treten signifikante Unterschiede für den gekürzten EIDM-ET sowie die *EBP beliefs scale* für die Fächerbündel auf. Damit kann ein vermuteter Vorteil von Proband*innen in experimentell geprägten Fächern, z.B. innerhalb des mathematisch-naturwissenschaftlich-technischen Fächerbündels, nicht festgestellt werden. In Bezug auf den Zusammenhang der Kompetenzen in evidenzbasierter Praxis und den zu Grunde liegenden epistemologischen Überzeugungen könnten hier keine fachspezifischen Überzeugungen bei den Proband*innen ausgeprägt sein (Muis et al., 2006). Dabei ergibt sich hier ein Unterschied zu den Ergebnissen von Urhahne und Kremer (2023). Hier muss jedoch berücksichtigt werden, dass die Autor*innen die Proband*innen auf Grundlage des erstgewählten Unterrichtsfachs gebündelt haben. In der vorliegenden Studie wurde auch das zweite und dritte Unterrichtsfach berücksichtigt, um der Interdisziplinarität des Lehramtsstudiums gerecht werden zu können (Rochnia et al., 2020). In Verbindung zur *Theory of integrated Domains in Epistemology* (TIDE) nach Muis et al. (2006) könnten bei den Proband*innen lediglich akademische epistemologische Überzeugungen vorliegen, welche zwar generelle Aspekte der Epistemologie beinhalten, jedoch keine fachspezifische Ausdifferenzierung erfahren haben. Übertragen auf evidenzbasierte Praxis könnten auch hier bisher eher generelle als fachspezifische Aspekte ausgeprägt sein.

Andererseits muss in diesem Zusammenhang der gekürzte EIDM-ET als Testinstrument kritisch diskutiert werden. Vor dem Hintergrund des Modells von Muis et al. (2006) könnten die Items eher auf akademische Überzeugungen, als fachspezifische Überzeugungen abzielen. Als Beispiel kann Item 1 des gekürzten EIDM-ET angeführt werden. Hier sollen die Proband*innen eine Forschungsfrage zu einer unterrichtlichen Situation entwickeln. Dabei kann hier eher ein Bezug zu bildungswissenschaftlichen Evidenzen erzeugt werden, da die Situation nicht an ein einzelnes Unterrichtsfach angegliedert wurde. Möglicherweise stellt diese ausbleibende Angliederung an ein Unterrichtsfach eine Hürde für die Aktivierung fachspezifischer epistemologischer Überzeugungen und daraus resultierenden EBP-Kompetenzen dar. Für die fächerübergreifende Befragung war es andererseits relevant eine Situationsbeschreibung zu wählen, welche für alle Proband*innen nachvollziehbar ist. Die weiteren Items des gekürzten EIDM-ET stehen teils im Zusammenhang mit der Situationsbeschreibung. So könnten hier ähnliche Effekte vermutet werden. Auch die *EBP beliefs scale* fokussiert nicht auf ein spezifisches Unterrichtsfach. Entsprechend könnten auch hier fachspezifische *beliefs* gehemmt worden sein. Für eine erneute Durchführung der vorliegenden Studie sollten so eher fachspezifische Situationsbeschreibungen und Items gewählt werden, welche jedoch über die verschiedenen Fächer hinweg vergleichbar sein müssten. So könnten ggf. eher fachspezifische Überzeugungen und Kompetenzen angesprochen werden. Dabei stellt sich jedoch die Herausforderung der Interdisziplinarität der Studierenden innerhalb des Lehramts (Rochnia et al., 2020). So studieren die Proband*innen mindestens drei Fächer (zwei Unterrichtsfächer +

Bildungswissenschaften). Hieraus ergibt sich auch die Fragestellung, inwiefern fachspezifische Überzeugungen und Kompetenzen bei einzelnen Proband*innen im Kontext der verschiedenen Unterrichtsfächer unterschiedlich ausgeprägt sind. Im Rahmen folgender Untersuchungen könnten so auch Situationsbeschreibungen für die verschiedenen Unterrichtsfächer einzelner Proband*innen eingesetzt werden. So könnten Einblicke gewonnen werden, inwiefern einzelne Proband*innen im Kontext unterschiedlicher Unterrichtsfächer unterschiedlich mit Evidenzen umgehen. Gleichzeitig scheint eine Einteilung der Proband*innen nur auf Grundlage eines ersten gewählten Unterrichtsfachs wie durch Urhahne und Kremer (2023) der diskutierten Komplexität der Überzeugungen sowie der Interdisziplinarität innerhalb des Lehramtstudiums (Rochnia et al., 2020) nicht gerecht werden zu können.

Eine Korrelation der erreichten Bachelornote mit der Punktzahl im EIDM-ET und *Beliefs_EBP_Impl_Schule* weist auf einen Einfluss der erreichten Bachelornote mit EBP-Kompetenzen hin. Dabei stehen höhere Skalenwerte innerhalb der *Beliefs_EBP_Impl_Schule* sowie eine höhere Punktzahl im gekürzten EIDM-ET mit einer schlechteren Bachelornote einher. Auf Grundlage eines zu unterstellendem vertieften Konstruktverständnis von evidenzbasierter Praxis auf Basis einer besseren Bachelornote, scheinen die Ergebnisse von den Erwartungen abzuweichen. Eine mögliche Erklärung könnte dennoch in einem vertieften Konstruktverständnis der Proband*innen mit besseren Bachelornoten erkannt werden. So könnte ein tiefergehendes Konstruktverständnis auch relative Zusammenhänge sowie die Kontextabhängigkeit epistemologischer Überzeugungen sowie folglich evidenzbasierter Praxis enthalten. Solche Überzeugungen könnten dabei die Komplexität des Konstruktes in den Fokus rücken und somit die Beantwortung der Items des gekürzten EIDM-ET erschweren, da hier wenig Kontextinformationen gegeben werden. Tiefergehende Analysen sind jedoch auf Grundlage der bestehenden Datenbasis nicht möglich. Der Zusammenhang eines vertieften Konstruktverständnisses und den EBP-Kompetenzen von Proband*innen könnte in weiteren Studien nachgegangen werden.

Die Untersuchung von Unterschieden auf Basis des Geschlechts als Gruppierungsfaktor zeigt signifikante Unterschiede auf. Dabei können einige der Unterschiede lediglich für die Gruppe von Studierenden beschrieben werden, welche sich als divers identifizieren. Da diese Gruppe aus lediglich 3 Proband*innen besteht, werden diese Zusammenhänge verworfen. Solide statistische Befunden können auf einer solch kleinen Gruppengröße nicht aufgebaut werden. Dennoch könnte der Frage in einer größeren Stichprobe der jeweiligen Gruppierungen nachgegangen werden.

Für die *EBP implementation scale* zeigen sich jedoch Geschlechterunterschiede zwischen männlichen und weiblichen Proband*innen mit einer geringen Effektstärke. Bei diesem Unterschied muss jedoch die Verteilung der Stichprobe auf Geschlechter sowie Studiengänge betrachtet werden. Innerhalb des Studiengangs „Lehramt an Grundschulen“ besteht die Stichprobe zu 89% aus weiblichen Proband*innen. Dieser Überhang weiblicher Studierender im Studiengang „Lehramt für Grundschulen“ kann als typisch beschrieben werden (vgl. Kap.7.2). Gleiches gilt für den Studiengang „Lehramt für sonderpädagogische Förderung“ mit einem Anteil weiblichen Proband*innen von 80%.

Im Rahmen einer Analyse der unterschiedlichen Studiengänge zeigen sich ebenfalls Unterschiede zwischen den beiden beschriebenen Studiengängen mit den anderen untersuchten Studiengängen. Dieser Unterschied lässt sich dabei auf einen Unterschied zwischen der Primarstufe zu den anderen Schulstufen beziehen (vgl. Kap. 6.6.1.3). Es kann nicht ausgeschlossen werden, dass innerhalb des beschriebenen Geschlechtereffekts eigentlich ein Schulstufeneffekt durchschlägt. Auf Grundlage bestehender Untersuchungen zu Geschlechterunterschieden in epistemologischen Überzeugungen könnte diese These untermauert werden. So ergibt sich keine solide

Argumentationsgrundlage, um das Auftreten von Geschlechterunterschieden zu erklären. Für die beschriebenen Unterschiede zwischen den Studiengängen bzw. zwischen den Schulstufen könnten jedoch Erklärungsansätze gefunden werden. Die beschriebenen Unterschiede zeigen sich lediglich für die *EBP implementation scale*. Das bedeutet, dass die Proband*innen der Primarstufe signifikant seltener evidenzbasierte Praktiken einsetzen. Für einen praxisorientierten Umgang mit Evidenzen (*Implement_EBP_Praxis*) besteht dieser Unterschied zur Sekundarstufe 2 sowie der beruflichen Bildung. Für einen forschungsorientierten Umgang mit Evidenzen (*Implement_EBP_Forschung*) kann der Unterschied im Vergleich zu allen anderen Schulstufen formuliert werden. Für *Implement_EBP_Praxis* könnte eine Erklärung im Fokus des Unterrichts erkannt werden. Innerhalb der Primarstufe sollen die Schüler*innen grundlegende Kompetenzen innerhalb der Hauptfächer (Mathematik, Deutsch, Englisch) sowie gebündelter Fächer (Sachunterricht) aufbauen. Dabei steht weniger im Fokus wissenschaftliche Evidenzen zu analysieren sowie zu bewerten (MSB NRW, 2021a). Entsprechend könnte angenommen werden, dass Evidenzen seltener Unterrichtsthema sind. Im Rahmen der Sekundarstufe 2, wie auch der beruflichen Bildung tritt dieser Umgang mit Evidenzen deutlicher in den Vordergrund (MSB NRW, 2007, 2022). Entsprechend könnte der Umgang mit Evidenzen aus einer unterrichtspraktischen Perspektive in der Primarstufe weniger relevant sein. Diese geringere Relevanz könnte dann auch zum selteneren Umgang mit Evidenzen generell führen. So könnten Evidenzen nur seltener mit Kolleg*innen geteilt werden oder ein Einbezug wissenschaftlicher Evidenz in die Unterrichtsplanung erfolgen, wie dies durch *Implement_EBP_Praxis* abgebildet wird.

Gleichzeit könnte die Kontextabhängigkeit als Argument angeführt werden. Zwar finden alle Bildungsprozesse innerhalb des individuellen Kontextes statt (vgl. Kap. 2.4), jedoch kann vor dem Hintergrund des mehrgliedrigen Schulsystems innerhalb der Primarstufe eine größere Vielfalt von Kontexten erwartet werden. So wird die Primarstufe von allen Schüler*innen besucht noch bevor eine Aufteilung auf verschiedene Schulformen zur Sekundarstufe erfolgt. Entsprechend könnten die Klassen durch größere Heterogenität geprägt sein, als dies innerhalb der weiterführenden Schulen der Fall ist. Diese hochgradige Kontextabhängigkeit der Bildungsprozesse innerhalb der Primarstufe könnte den Einbezug von Evidenzen weiter erschweren.

Eine fehlende wahrgenommene Relevanz von wissenschaftlichen Evidenzen für die unterrichtliche Praxis könnte auch einen Einfluss auf den forschungsorientierten Umgang mit Evidenzen (*Implement_EBP_Forschung*) nehmen. Die Diskussion von Evidenzen mit Kolleg*innen oder Eltern könnte so insgesamt ausbleiben. Dabei könnte hier erneut der beschriebene Rollenkonflikt der Proband*innen herangezogen werden. Eine fehlende wahrgenommene Relevanz wissenschaftlicher Erkenntnisse für die Unterrichtsgestaltung könnte hier den Umgang mit Evidenzen von der Tätigkeit einer Lehrkraft trennen (Bauer & Kollar, 2023). Entsprechend könnte auch hier ein Kulturwechsel notwendig werden, um einen deutlicheren Einbezug evidenzbasierter Praxis zu erreichen.

Die vorgestellte Argumentation könnte die eigentlichen Geschlechterunterschiede überzeichnen. Für eine tiefergehende Analyse müsste eine geschlechterkontrollierte Stichprobe für die unterschiedlichen Studiengänge untersucht werden. So könnten Einblicke in die zu Grunde liegende Begründung der beschriebenen Unterschiede erreicht werden. Gleichzeitig bedürfte es der Differenzierung von Unterschieden im Umgang mit Evidenzen für Proband*innen der verschiedenen Studiengänge. Die hier beschriebenen Unterschiede könnten auf Grundlage der bestehenden Rahmenbedingungen angenommen werden, jedoch könnten weitere Aspekte hinzukommen, welche durch die in dieser Studie gewählten Instrumente nicht aufgeklärt werden können.

Auf Grundlage der beschriebenen Diskussionsaspekte sollte die Rolle des Praxissemesters im Sinne der Förderung von Kompetenzen evidenzbasierter Praxis reflektiert werden. Wie sich zeigt, kann das Praxissemester als Lerngelegenheit für den Umgang mit Evidenzen als nicht effektiv beschrieben werden. Die Konzeption des Praxissemesters zeigt hingegen eindeutig die Verknüpfung wissenschaftlicher Evidenz mit unterrichtlicher Praxis als relevantes Lernziel auf (Herzig & Wiethoff, 2019). Scheinbar fehlt es jedoch an effektiven Lernsituationen, um Kompetenzen im Umgang mit Evidenzen aufzubauen. Dabei bietet scheinbar auch das Begleitforschungsseminar nicht ausreichend Gelegenheit einen forschungsorientierten Umgang mit Evidenzen zu trainieren. Andererseits könnte hier auch die fehlende Thematisierung evidenzbasierter Praxis im Rahmen des gesamten Studiums durchschlagen (Stelter & Miethe, 2019). Es ist fraglich, ob notwendige Kompetenzen einzig im Rahmen des Praxissemesters erworben werden können, welches durch eine Vielzahl neuer Eindrücke und Herausforderungen für die Studierenden geprägt ist.

Ein Indiz für die wenig umfangreiche Thematisierung evidenzbasierter Praxis innerhalb des Studiums könnte auch in den nur leicht positiv ausgeprägten *beliefs* gegenüber dem Wert evidenzbasierter Praxis zu Beginn des Praxissemesters erkannt werden. Wie deutlich wird, scheint der von Bauer und Kollar (2023) geforderte Kulturwechsel notwendig zu sein, um Kompetenzen in evidenzbasierter Praxis aufbauen zu können. Dieser bezieht sich dabei nicht nur auf die Proband*innen sondern auch auf die Gestaltung des institutionalisierten Praxissemesters. Es müssen explizite Lernanlässe im Praxissemester erzeugt werden, im Rahmen derer die Studierenden EBP-Kompetenzen erwerben können. Das Begleitforschungsseminar kann dabei aus einer forschungsorientierten Perspektive einen Ansatzpunkt darstellen. Für einen praxisorientierten Umgang mit Evidenzen müssen neue Lernanlässe geschaffen werden. Dabei sollte hier explizit der (fach-)spezifische Kontext einbezogen werden. Der konkrete Anwendungskontext innerhalb der spezifischen Unterrichtsfächer könnte hier gewinnbringend eingesetzt werden. Gleichzeitig könnte die Reflexion spezifischer Situationen vor dem Hintergrund konkreter wissenschaftlicher Evidenzen umgesetzt werden. In diesem Zusammenhang könnten Methoden wie die kollegiale Fallberatung einbezogen werden. Insgesamt gibt es einen Bedarf der Ausschärfung der Rolle des Praxissemesters im Kontext des Kompetenzerwerbs in evidenzbasierter Praxis.

Es wird deutlich, dass neue Förderungsansätze zur Stärkung der EBP-Kompetenzen angehender Lehrkräfte notwendig sind. Dabei scheint eine Förderung von Studierenden aller Fächerbündel notwendig zu sein. Gemeinsame Fördermaßnahmen könnten im Sinne akademischer epistemologischer Überzeugungen (Muis et al., 2006) grundlegende Kompetenzen in evidenzbasierter Praxis fördern. Anschließend könnten fachspezifische Ausdifferenzierungen des Konstruktes in den Blick genommen werden, ohne dabei die Interdisziplinarität der Studierenden zu vernachlässigen. Im Sinne der Kontextabhängigkeit ist ebenfalls fraglich, inwiefern fachspezifische EBP-Kompetenzen auch Einfluss auf akademische EPB-Kompetenzen nehmen könnten. Die Unterrichtsfächer als konkretisierte Anwendungsbereiche evidenzbasierter Praxis könnten einen leichteren Zugang zu dieser ermöglichen. Im Rahmen folgender Untersuchungen sollten daher auch fachspezifische Fördermaßnahmen in den Blick genommen werden.

6.6.3 Ergebnisse der Interventionsstudie zur Förderung angehender Chemielehrkräfte in evidenzbasierter Praxis

6.6.3.1 Ergebnisse des EIDM-ET

Innerhalb des vollständigen EIDM-ET konnten maximal 216 Punkte erreicht werden. Tabelle 25 gibt einen Überblick über die erreichten Punktzahlen zum Prä- und Post-Zeitpunkt.

Tabelle 25 – Deskriptive Statistik des vollständigen EIDM-ET zum Prä- und Post-Zeitpunkt

Testzeitpunkt	N	M	SD
Prä-Test	14	108.71 (50.33%)	30.95
Post-Test	14	121.64 (56.31%)	25.15

Eine Analyse der Einzelitems zeigt Unterschiede zwischen Prä- und Post-Zeitpunkt auf (Tab. 26). Dabei sind für die meisten Items leichte Zunahmen der Mittelwerte zum Post-Zeitpunkt erkennbar. Ein Vergleich der durchschnittlich erreichten Punktzahl zwischen Prä- und Post-Zeitpunkt mittels gepaartem t-Test zeigt einen signifikanten Unterschied mit einer mittelstarken Effektgröße auf ($t(13) = -2.16$, $p = .050$, $d = .578$). Die erreichte durchschnittliche Punktzahl fällt damit am Post-Zeitpunkt signifikant höher aus als zum Prä-Zeitpunkt.

Tabelle 26 – Deskriptive Statistik für die einzelnen Items des vollständigen EIDM-ET zu Prä- und Post-Zeitpunkt

Items des EIDM-ET	M (SD)		Prozentual erreichte Punktzahl	
	Prä	Post	Prä	Post
Item 1_1	9.86 (3.61)	10.07 (4.01)	82.17%	83.92%
Item 1_2	11.36 (2.41)	10.50 (3.06)	94.67%	87.50%
Item 2	12.00 (8.74)	16.57 (7.98)	50.00%	69.04%
Item 3	18.64 (8.25)	21.86 (2.98)	77.67%	91.08%
Item 4	11.14 (9.94)	12.00 (11.53)	46.42%	50.00%
Item 5_1	.57 (2.14)	.57 (2.14)	7.13%	7.13%
Item 5_2	2.29 (3.75)	2.86 (3.98)	28.63%	35.75%
Item 5_3	1.14 (2.91)	1.14 (2.91)	14.25%	14.25%
Item 6_1	6.29 (3.41)	6.29 (3.41)	78.63%	78.63%
Item 6_2	4.57 (4.11)	5.71 (3.75)	57.13%	71.38%
Item 6_3	4.00 (4.15)	6.86 (2.91)	50.00%	85.75%
Item 7	13.36 (6.49)	13.93 (6.22)	55.67%	58.04%
Item 8_1	3.43 (3.88)	3.00 (3.11)	28.58%	25.00%
Item 8_2	3.43 (3.88)	1.71 (2.81)	28.58%	14.25%
Item 9	6.64 (9.72)	8.57 (9.91)	27.67%	35.71%

Anmerkung. N=14

Eine tiefergehende Analyse der Einzelitems zeigt vor allem für Item 5 und 8 einen deutlich unterdurchschnittlichen Mittelwert. Item 5 fokussierte auf forschungsmethodologische Aspekte, indem nach einem geeigneten Studiendesign für verschiedene Themen gefragt wird. Für dieses Item können zum Prä- und Post-Zeitpunkt die geringsten Mittelwerte im Vergleich zur erreichbaren

Gesamtpunktzahl der Items beobachtet werden. Ähnliches gilt für Item 8, welche auf die Signifikanz und Bedeutung von Befunden abstellt. Die Proband*innen sollen hier beschreiben, welche Charakteristika ebendiese aufweisen müssen, damit sie als signifikant und bedeutsam bewertet werden können. Damit fokussieren beide Aspekte auf forschungsmethodologisches Wissen, welches zur Planung und Bewertung von Studien(-ergebnissen) nötig ist. Auf Grundlage der erreichten Punktzahlen für die Items scheint das forschungsmethodologische Wissen bei den Proband*innen eher schwächer ausgeprägt zu sein.

Items 1-4 schneiden im Vergleich deutlich besser ab. Diese Items fokussieren eher auf den Forschungsprozess, in dem Fragestellungen formuliert und nach passenden Evidenzen recherchiert werden soll. Dabei spielt für diese Items das forschungsmethodologische Wissen eine eher untergeordnete Rolle. Die Generierung einer Fragestellung und deren weitere Bearbeitung werden hingegen in einen spezifischen unterrichtlichen Kontext eingebettet.

Für Item 6 können zu Prä- und Post-Zeitpunkt überdurchschnittliche, für Item 7 durchschnittliche Mittelwerte berichtet werden. Beide Items fokussieren ebenfalls auf forschungsmethodologisches Wissen. Während Item 6 unterschiedliche Begriffsdefinition in einem *Multiple-choice* Format abfragt (vgl. Item 4 im gekürzten EIDM-ET, Kap. 6.5.1.1), fokussiert Item 7 auf Validitätsargumente für Studienergebnisse. Im Gegensatz zu Item 5 und 8 werden hier eher oberflächliches Aspekte des forschungsmethodologischen Wissen abgefragt. So wird einerseits Definitionswissen, anderer Seite die bloße Nennung von Validitätsmerkmalen erfragt. Für Item 5 und 8 müssen forschungsmethodologische Aspekte angewendet werden, indem Studiendesigns ausgewählt und tiefergehende Aspekte zur Signifikanz und Bedeutung von Befunden genannt werden sollen.

Auf Grundlage des EIDM-ET lässt sich für den Prä- und Post-Zeitpunkt innerhalb von Phase 1 eine eher mittelmäßige Kompetenzausprägung erkennen. Dabei zeigt sich ein Einfluss auf Items mit einem Schwerpunkt auf forschungsmethodologisches Wissen (Item 6), den kontextualisierten Forschungsprozess (Items 2, 3) sowie die Bewertung von Evidenzen bzgl. deren Implementationsmöglichkeiten innerhalb der Unterrichtspraxis (Item 9).

Für die übrigen Items (4, 5 und 7) sind Mittelwertssteigerungen <1 Punkt zwischen Prä- und Post-Zeitpunkt zu erkennen. Damit scheint die Intervention keinen Einfluss auf das Wissen über Strategien zur Einschränkung/Ausweitung einer Literaturrecherche (Item 4), Studiendesigns (Item 5) sowie die Anwendung von Validitätsargumenten für Studienergebnisse (Item 7) zu haben. Für Items 4 und 5 könnte eine Begründung in der Individualität der durchgeführten Projekte vermutet werden. So könnten Studiendesigns fernab des eigenen Designs ggf. unklar geblieben sein. Innerhalb der Intervention wurden zwar grundlegende Studiendesigns besprochen, jedoch könnte die Anwendung dieser auf den spezifischen Kontext einer Fragestellung herausfordernd gewesen sein, sodass die Proband*innen hier insgesamt eher eine unterdurchschnittliche Punktzahl erreicht haben und diese durch die Intervention nicht zugenommen hat. Die Anwendung von Validitätsargumenten für Studienergebnisse scheint ebenso herausfordernd für die Proband*innen zu sein. Auch die Anwendung von Validitätsargumenten auf Befunde ist im Rahmen der vorliegenden Intervention eher in der Projektdurchführung zu verorten.

Für Items 1 und 8 kann eine Abnahme der erreichten Punktzahl erkannt werden. Dabei zeigt sich die Abnahme für Item 1 nur für die zweite Situationsbeschreibung mit einem Fokus auf den Chemieunterricht. Dabei beträgt die Abnahme der erreichten Punktzahl hier nur unter einem Punkt. Für item 8 hingegen kann eine teils deutliche Abnahme der erreichten Punktzahl erkannt werden. Diese tritt hier besonders deutlich für den zweiten Teil des Items hervor, welcher auf die Bedeutung von Befunden im Kontext der Effektstärke fokussiert. Die Intervention könnte hier eine

Verwirrung des Konzepts der Bedeutsamkeit im empirischen und theoretischen Sinne ausgelöst haben. So ist ein deutlicher Schwerpunkt der Intervention auf der Bedeutung von Evidenzen für Bildungsprozesse zu erkennen. Dabei spielt die empirische Bedeutsamkeit im Sinne der Effektstärke jedoch keine gesonderte Rolle. Folglich könnte es hier zu Verwirrung gekommen sein.

6.6.3.2 Ergebnisse der EBP beliefs scale

Für die *EBP beliefs scale* können zum Post-Zeitpunkt höhere Mittelwerte als für den Prä-Zeitpunkt beschrieben werden (vgl. Tab. 27). Die Mittelwerte liegen (mit Ausnahme von *Beliefs_EBP_Impl_Schule* am Prä-Zeitpunkt) oberhalb des Skalenmittelwerts.

Tabelle 27 – Deskriptive Statistik für die Sub-Skalen der *EBP beliefs scale* zu Prä- und Post-Zeitpunkt

Faktor	<i>M</i>	<i>SD</i>
Beliefs_EBP_Bildung - Beliefs gegenüber dem generellen Einfluss von EBP auf Bildung		
Prä-Zeitpunkt	3.52	.53
Post-Zeitpunkt	3.93	.64
Beliefs_EBP_Impl_Schule - Beliefs gegenüber der Implementation von EBP in der Schule		
Prä-Zeitpunkt	2.45	.79
Post-Zeitpunkt	3.12	.53
Beliefs_EBP_Fähigkeiten - Beliefs gegenüber persönlichen Fähigkeiten in EBP		
Prä-Zeitpunkt	3.20	.56
Post-Zeitpunkt	3.34	.56

Anmerkung. *N*=14

Der Mittelwertsunterschied von *Beliefs_EBP_Bildung* zeigt einen signifikanten Unterschied ($t(13) = -2.80$, $p = .015$) mit einem mittelstarken Effekt ($d = .747$). Der Mittelwert fällt somit zum Post-Zeitpunkt signifikant höher aus als für den Prä-Zeitpunkt. Auch für *Beliefs_EBP_Impl_Schule* kann ein signifikanter Mittelwertsunterschied mittels gepaartem t-Test ermittelt werden ($t(13) = -2.98$, $p = .011$). Die Effektstärke für diese Sub-Skala verpasst mit $d = .796$ nur knapp die Benchmark für einen starken Effekt ($d > .8$). Für *Beliefs_EBP_Fähigkeiten* kann kein signifikanter Mittelwertsunterschied bestimmt werden ($t(13) = -.64$, $p = .531$).

6.6.3.3 Ergebnisse der EBP implementation scale

Im Rahmen der vorliegenden Interventionsstudie werden die Ergebnisse der *EBP implementation scale* für die Gesamtskala berichtet. Eine fehlende Eignung der Daten für eine Faktorenanalyse lässt keine Bildung von Faktoren zu (vgl. Kap. 6.5.2).

Für die *EBP implementation scale* kann zum Prä-Zeitpunkt ein Mittelwert von $M = 1.27$ ($SD = .23$) bestimmt werden. Der Mittelwert zum Post-Zeitpunkt liegt mit $M = 1.56$ ($SD = .32$) leicht höher. Auf Grundlage des fünfstufigen Antwortformats entsprechen die Mittelwerte einer Ausführung evidenzbasierter Tätigkeiten zwischen 0 und 1–3-mal innerhalb der letzten 8 Wochen. Der beschriebene Mittelwertsunterschied fällt signifikant aus ($t(13) = -3.89$, $p = .002$) und zeigt einen starken Effekt auf ($d = 1.04$). Die Proband*innen erreichen damit zum Post-Zeitpunkt einen signifikant höheren Mittelwert, welcher sich jedoch weiterhin im Feld der Ausübung evidenzbasierter Tätigkeiten zwischen 0 bis 1-3-mal innerhalb der letzten 8 Wochen befindet.

6.6.3.4 Ergebnisse der Interviews

Der Ergebnisbericht der Interviews wird entlang der *evidence-use* Mechanismen nach Breckon und Dodson (2016) gegliedert. Wie beschrieben kann dazu sowohl auf ein Interview zum Post-Zeitpunkt zurückgegriffen werden. Die Angabe der Proband*innen vor der Teilnahme an der Interventionsstudie an keinen Workshops o.ä. zum Thema evidenzbasierter Praxis teilgenommen zu haben, kann als Indikator dafür herangezogen werden, dass die beschriebenen Vorstellungen innerhalb der Interviews auf die Intervention zurückgehen könnten.

Mechanismus: Awareness

Im Rahmen des Mechanismus *Awareness* beschreiben die Proband*innen zum Post-Zeitpunkt eine Rolle von Evidenzen für die Schule, welche essentielle Aspekte der Definition evidenzbasierter Praxis aufgreifen. So stellt AXTNMN17 (Pos. 45) die Verknüpfung von wissenschaftlichen Erkenntnissen mit den eigenen Erfahrungen und dem spezifischen Kontext hervor:

Da ist das glaube ich schon wichtig, so was, ja, so schon irgendwie was zu haben, evidenzbasiert, was so eine Aussage macht, ja, ob man irgendwie, weiß ich nicht, ob es da jetzt irgendwas Gutes gibt, was die Schüler motiviert, beispielsweise, oder so was. Genau, (unverständlich.) ich würde glaube ich nicht nur den Fokus darauf legen. Also ich finde es gut, dass man da mal guckt, so, was gibt es da irgendwie für Möglichkeiten, was wurde auch schon gut ausgetestet? Aber trotzdem irgendwie selber für sich auch nochmal guckt, was funktioniert mit MEINEN Schülern gut?

Die Proband*in stellt hier den Einbezug wissenschaftlicher Evidenzen in die Unterrichtsplanung dar. Weitere Proband*innen stellen die Bedeutung von evidenzbasierter Praxis für Schule in den Kontext der Bewertungskompetenz der Schüler*innen (LERFJA31), neue Erkenntnisse in der Fachwissenschaft Chemie (UEASJS22) oder Themen wie den Klimawandel oder die COVID-19 Pandemie (AEUHML12). Insgesamt berichten die Proband*innen von einer großen Bedeutung evidenzbasierter Praxis für Schule. Gleichzeitig werden Erfahrungen aus bereits absolvierten Praktika und dem Praxissemester berichtet, dass die tatsächliche Umsetzung evidenzbasierter Praxis in der Schule seltener beobachtet werden konnte (z.B. BTKELA02, AAAMAA08, HAERCE08). Als ein möglicher Grund wird der Zeitmangel in der Schule beschrieben (DSRFKA28). Dabei wird auch eine abnehmende Relevanz wissenschaftlicher Evidenz mit zunehmenden Dienstjahren berichtet (HAERCE08). In diesem Kontext wird das zunehmende Erfahrungswissen als leitend für Unterrichtsentscheidungen beschrieben. So würden Evidenzen z.B. eher beim Einsatz neuer Methoden benötigt, während bekannte Methoden und Medien aus der Erfahrung heraus eingesetzt werden könnten (AAAMAA08). Hierbei spielen die Proband*innen auch auf ein Verständnis von Evidenzen als etwas „Neues“ an. Evidenzen müssten entsprechend Antworten auf neue Fragen geben. So beschreibt BEHGLA17 (Pos. 101-111) Experimente in einem Lehr-Lern-Kontext nicht als Evidenzen, da sie „schon hundert Mal gemacht [...] [wurden] UND als Student weiß man ja quasi auch, was rauskommen sollte.“ Andere Proband*innen beschreiben, dass Evidenzen so lange Bestand haben, „bis jemand was Besseres beweist“ (UADFJA17, Pos. 51). Hieran schließen die Proband*innen Aussagen zur Vorläufigkeit wissenschaftlicher Erkenntnis an. Es wird beschrieben, dass wissenschaftliche Evidenz niemals endgültig ist: „Man hat ja nie den absoluten Beweis, auch wenn jetzt Leute, die sich damit auskennen sagen, ja, der Meinung bin ich auch.“ (BEHGLA17, Pos. 17). Auf der anderen Seite beschreibt z.B. AEUHML12 Abstufungen von Evidenzen. So werden physikalische Gesetzmäßigkeiten als (nahezu) unumstößlich beschrieben, während Evidenzen in anderen Bereichen eher als vorläufige Gesetzmäßigkeiten beschrieben werden können.

Mechanismus: Agree

Im Unterschied zum Mechanismus *Awareness* fokussiert der Mechanismus *Agree* auf relevante Inhaltsbereiche evidenzbasierter Praxis. Gefragt nach Bereichen im Unterricht, für die Evidenzbasierung wichtig ist, nennen die Proband*innen viele verschiedene Beispiele: Schülervorstellungen (z.B. AAAMAA08, AXTNMN17, BEHGLA17), Unterrichtsaufbau (BEHGLA17), Methoden-/Medienwahl (z.B. AAAMAA08, UADFJA17, UEASJS22, GAJFMS03), Leistungsmessung (z.B. HEUONS29, LAHHRN23), Fachinhalte (z.B. HAERCE08, CAUEEE16). Dabei stellt ein Großteil der Proband*innen die Bedeutung evidenzbasierter Praxis sowohl für fachwissenschaftliche als auch fachdidaktische Bereiche als gleich dar. Im Bereich fachlicher Inhalte beschreibt AXTNMN17 die Bedeutung von Evidenzen hingegen als geringer, da die relevanten Inhalte durch den Kernlehrplan vorgegeben würden. Über den Unterricht hinaus beschreiben die Proband*innen auch die Bedeutung von Evidenzen im alltäglichen Leben. Dabei werden zum Beispiel Gespräche mit Freunden angeführt:

[...] Freunde sagen mir, zum Beispiel jetzt gestern oder vorgestern erst irgendwie, die, die Qualle ist das giftigste Tier der Welt. Und dann schaue ich dann abends schon noch mal, kann man das denn überhaupt so einfach so sagen, ist das nach welchen Kriterien das giftigste Tier? (BEHGLA17, Pos. 35-37)

Andere Proband*innen beschreiben eine fehlende Relevanz von Evidenzen im Alltag außerhalb des Studiums (z.B. UEASJS22). Weiterhin werden unterschiedliche Bereiche des alltäglichen/gesellschaftlichen Lebens genannt: Werbung/Social Media (z.B. BTKELA02, AAAMAA08), Nachrichten/Presse (z.B. CAUEEE16, GAJFMS03, UEASJS22), Produktrezensionen/Stiftung Warentest (HAERCE08), Klimaschutz/-wandel (z.B. BEHGLA17, LERFJA31, CAUEEE16), Politik (z.B. HEUONS29, AAAMAA08), Jura (UADFJA17), Nachhaltigkeit (UADFJA17), Forschung/Wissenschaft (z.B. GAJFMS03, CAUEEE16), Handwerk (BTKELA02) und Medizin/Gesundheit (z.B. HAERCE08, UADFJA17, AXTNMN17). Die Proband*innen zeigen eine große Bandbreite von relevanten Bereichen im Kontext evidenzbasierter Praxis auf.

Im Zusammenhang mit der Rolle von Evidenzen haben die Proband*innen auch Kriterien für Evidenzen formuliert. Diese Bedingungen zielen darauf ab Charakteristika von Evidenzen zu benennen. AXTNMN17 (Pos. 9) formuliert in diesem Zusammenhang: „Ja, also Evidenzen sind für mich, ja, diese empirische Überprüfbarkeit, sage ich jetzt mal.“. Die empirische Überprüfbarkeit, bzw. die Bereitstellung statistischer Kenngrößen zur Beurteilung von Evidenzen wurde vermehrt genannt (z.B. AAAMAA08, AEUHML12, LAHHRN23). Dabei sollten Evidenzen im Idealfall auf mehreren Datensätzen basieren (z.B. AXTNMN17, UADFJA17) und reproduzierbar sein (z.B. HAERCE08, AXTNMN17). Als weiteres Kriterium wird die Veröffentlichung in wissenschaftlichen Journalen beschrieben (z.B. HEUONS29, UADFJA17), im Idealfall mit Peer-review-Verfahren (z.B. BEHGLA17, GAJFMS03). Dabei werden immer wieder auch Meinungen (z.B. von Forschenden, oder Schulpraktiker*innen) erwähnt. Diese werden durch die Proband*innen von Evidenzen grundsätzlich zwar unterschieden (z.B. AXTNMN17), jedoch als ähnliches Konstrukt beschrieben. So würde CAUEEE16 (Pos. 57) Meinungen „[...] als Evidenz bezeichnen, allerdings unter Vorbehalt“. DSRFKA28 (Pos. 56) beschreibt Meinungen als „Hinweis, dass es funktionieren könnte [...]“. Dabei können Unterschiede auf Grundlage der meinungsausßernden Person beobachtet werden. So können mit steigendem akademischen Grad Unterschiede in der Glaubwürdigkeit erkannt werden, wenngleich prinzipiell die gleichen Unsicherheiten durch subjektive Meinungsäußerungen beschrieben werden (z.B. LAHHRN23, GAJFMS03, HAERCE08). Sofern subjektive Meinungen durch eigene Untersuchung oder bestehende Literatur bekräftigt werden können, erkennen die Proband*innen in diesen eher Evidenzen als in unbegründeten Meinungen (z.B. AXTNMN17, LAHHRN23, HEUONS29). Der Evidenz-Grad nimmt so nach den Beschreibungen der Proband*innen mit

bestehenden Begründungen der subjektiven Meinung zu. Dabei spielt ebenso das Zustandekommen der Meinungen eine Rolle. So sollte diesen ein systematisches (Beobachtungs-) Verfahren zugrunde liegen. AEUHML12 hebt die Bedeutung der Konkretheit von Meinungen im Vergleich zu wissenschaftlichen Evidenzen hervor. So könnten in Meinungen spezifische Kontexte beachtet werden, während dies in wissenschaftlichen Evidenzen teils nicht der Fall ist.

Mechanismus: Access and Communication

Innerhalb des Mechanismus *Access and Communication* beschreiben die Proband*innen die Vorbereitung auf den Umgang mit Evidenzen in der Schule durch das universitäre Studium. Die Vorbereitung durch das Studium wurde dabei ambivalent beschrieben. Auf der einen Seite wird der Umgang mit Evidenzen direkt mit der Arbeit im Chemiestudium in Zusammenhang gebracht. Dabei werden vor allem die chemischen Experimente inkl. der daraus abgeleiteten empirischen Belege für Theorien hervorgehoben (z.B. AXTNMN17, AAAMAA08, LERFJA31). Ebenso wird das wissenschaftliche Arbeiten im Sinne von der Recherche und Verweisen auf Literatur innerhalb schriftlicher Arbeiten beschrieben (z.B. AAAMAA08, HAERCE08, LAHHRN23). Auch die Intervention (z.B. AXTNMN17, HEUONS29, HAERCE08) sowie Seminare in anderen Unterrichtsfächern (z.B. UADFJA17, LERFJA31, DSRFKA28) werden durch die Proband*innen genannt. Bei den Seminaren innerhalb der anderen Unterrichtsfächer wurden Methodenseminare z.B. für die Fächer Deutsch, Pädagogik und Ernährungslehre beschrieben. Teilweise beschreiben die Proband*innen, dass die Rolle von Evidenzen in den Fachdidaktiken oder der Bildungswissenschaft deutlicher geworden ist als in den Fachwissenschaften (BTKELA02). Gleichzeitig werden auch Unterschiede für die Unterrichtsfächer beschrieben:

Ja, also es ist jetzt irgendwie sehr schwierig, weil es kommt ja auch irgendwie so total auf das Fach an. (Wie gesagt? unverständlich.) in Ernährungslehre hatten wir ein ganzes Seminar zu Studien, Studienauswertung [...]. Das heißt ich sage mal durch das Fach Ernährungslehre habe ich vielleicht ein Extraschub auch, ja, so eine Extraportion dazu bekommen. (DSRFKA28, Pos 76)

Auf der anderen Seite beschreiben die Proband*innen in einem vergleichbaren Umfang Defizite in der universitären Vorbereitung auf den Umgang mit Evidenzen im Lehrberuf. AAAMAA08 (Pos. 81) hebt dabei hervor, dass zwar der individuelle Umgang mit Evidenzen vorbereitet wurde, jedoch keine Konzepte wie dies an Schüler*innen weitergegeben werden könnte: „Ich glaube was eben schwieriger ist, was die Uni nicht unbedingt beibringt, ist eben wie ich diese Evidenzen im Unterricht sinnvoll einsetze und verwende.“. Ebenso wird bemängelt, dass der Umgang mit Evidenzen nur an ausgewählten Stellen des Studiums aufgegriffen wurde (z.B. AAAMAA08, UEASJS22, UADFJA17). Dabei wurde auf die notwendige Explikation von Evidenzen hingewiesen, welche z.B. im Bachelorstudium (z.B. UEASJS22, CAUEEE16, BTKELA02) oder fachwissenschaftlichen Veranstaltungen (z.B. HEUONS29, BTKELA02, LAHHRN23) ausgeblieben sei. UEASJS22 beschreibt weiter, im Studium selbst keine Evidenzen erzeugt zu haben.

Mechanismus: Interact

Innerhalb des Mechanismus *Interact* sollen die Proband*innen beschreiben, inwiefern auch Schüler*innen mit Evidenzen interagieren könnten. Zum Post-Zeitpunkt beschreiben die Proband*innen neben der Recherche von Evidenzen (z.B. BEHGLA17, UEASJS22, AXTNMN17) auch die Durchführung von Experimenten (z.B. AAAMAA08, LERFJA31, HEUONS29) als wichtige Aspekte des Umgangs mit Evidenzen von Schüler*innen. Dabei hebt UADFJA17 für die Recherche auch die Bewertung von Artikeln/Evidenzen hervor. BEHGLA17 beschreibt in diesem Zusammenhang einen

Unterscheid der generierten Evidenzen. So könnten die Evidenzen, welche von Schüler*innen generiert werden, nur wenig Bedeutung für wissenschaftliche Forschung haben. Für die Schüler*innen hingegen stellen diese wichtigen Ausgangspunkte für Argumentationen dar. AXTNMN17 verortet den Umgang mit Evidenzen von Schüler*innen in Projektarbeiten oder der Facharbeit. GAJFMS03 hingegen beschreibt die Notwendigkeit eines gesamten Unterrichtsvorhabens, um bei den Schüler*innen Kompetenzen im Umgang mit Evidenzen aufzubauen.

Mechanismus: Skills

Innerhalb des Mechanismus *Skills* werden praktische Fähigkeiten evidenzbasierter Praxis mit den Proband*innen diskutiert. So sollen diese bspw. ihr Vorgehen zur Recherche von Evidenzen in fachdidaktisch-pädagogischen sowie fachwissenschaftlichen Zusammenhängen beschreiben.

Grundsätzlich beschreiben die Proband*innen das Vorgehen in beiden Bereichen sehr ähnlich. Ein erster Zugang wird oft im Internet gesucht (z.B. BEHGLA17, HEUONS29). Hier wird durch verschiedene Datenbanken auf Forschungsliteratur fokussiert. Dabei werden für fachdidaktisch-pädagogische Fragestellungen häufig Zeitschriften als Quellentyp genannt (z.B. BTKELA02, HAERCE08, UADFJA17). Für fachwissenschaftliche Fragestellungen werden hingegen seltener spezifische Quellentypen genannt. LERFJA31 (Pos. 71) verweist generell auf einen Unterschied zwischen den beiden Bereichen: „Also ich glaube, man neigt mehr dazu in der FACHWISSENSCHAFT nach etwas Neuem zu suchen, weil bei der Didaktik, das hat man dann vielleicht auch irgendwann gelernt, das funktioniert dann vielleicht irgendwie“.

Gefragt nach den eigenen Fähigkeiten zur Bewertung von Forschungsartikeln beschreiben die Proband*innen meist grundlegende Kompetenzen zu besitzen (z.B. AAAMAA08, UEASJA22, DSRFKA28). Ähnliche Antworten können für die Frage nach den eigenen Fähigkeiten zur Planung einer kleinen Forschungsstudie beobachtet werden (z.B. AXTNMN17, HEUONS29, LAHHRN23). GAJFMS03 bringt in diesem Zusammenhang eine fehlende Motivation zur Durchführung einer solchen Studie ein. Begründet wird dies in der Arbeitsbelastung von Lehrkräften: „Einfach, weil, weil ich an meiner Freundin sehe, wie viel Lehrer einfach zu tun haben und ich glaube nicht, dass man da Zeit hat, seine Freizeit mit sowas zu verbringen.“ (GAJFMS03, Pos. 146-155).

Mechanismus: Structures and Processes

Ein wichtiger Aspekt innerhalb des Mechanismus *Structures and Processes* kann in der Entwicklung evidenzbasierter Praxis mit zunehmender Unterrichtserfahrung von Lehrkräften erkannt werden. Die Proband*innen beschreiben eine abnehmende Bedeutung von Evidenzen mit zunehmender Erfahrung. So könnten bspw. bestimmte Methoden oft durchgeführt worden sein, sodass durch die vorhandene Erfahrung kein Bedarf nach neuen Evidenzen besteht (z.B. LAHHRN23, HAERCE08, AAAMAA08). Andererseits beschreiben einige Proband*innen eine zunehmende Bedeutung für Evidenzen in bestimmten Themengebieten (z.B. AAAMAA08). Dabei unterscheidet sich diese zunehmende Relevanz jedoch von der erwarteten Situation innerhalb der Schulpraxis. So nehmen die Proband*innen insgesamt an, dass z.B. aufgrund von Zeitmangel (LAHHRN23) mit zunehmendem Erfahrungswissen weniger Evidenzen genutzt werden

Bezogen auf vorhandene Schulstrukturen beschreiben die Proband*innen hingegen eine Fülle von Bereichen, für welche eine evidenzbasierte Arbeitsweise besondere Relevanz besitzt: Schulentwicklung (z.B. AXTNMN17), Betreuung/Sozialarbeit (z.B. AXTNMN17, UADFJA17, HAERCE08), Schulleitung (z.B. UEASJS22, GAJFMS03, BTKELA02), individuelle Förderung (UADFJA17), Fachkonferenzen (LERFJA31), Inklusion (HEUONS29). Neben diesen Bereich wird durchgehend der Unterricht durch Lehrkräfte genannt. Hier werden teils Abstufungen für Fächer vorgenommen, welche

jedoch selbst nicht durch die Proband*innen unterrichtet werden. So können für Fächer wie Religion (AAAMAA08) oder Musik (GAJFMS03) keine Aspekte evidenzbasierter Praxis durch die Proband*innen antizipiert werden. AEUHML12 (Pos. 92-94) fasst zusammen, dass evidenzbasierte Praxis „Generell in den Gruppen, wo Entscheidungen getroffen werden für die Schülerinnen und Schüler [...]“ eine wichtige Rolle spielt. Diese Aussage könnte auch die Beschreibungen für Phase 2 der vorliegenden Studie zusammenfassen. Auch hier werden generell Personen oder Gruppen benannt, welche eine besondere Verantwortung und gleichzeitig gewisse Macht im Schulsystem innehaben.

Insgesamt können die Proband*innen am Post-Zeitpunkt einige Merkmale von Evidenzen und dem Umgang mit diesen darstellen. Dabei wird ein gradueller und vorläufiger Charakter von Evidenzen deutlich, welcher jedoch im Rahmen sprachlicher Formulierungen zum Teil weniger klar hervortritt. So beschreiben AXTNMN17 (Pos. 79) z.B.: „[...] in der Chemie haben wir halt diese Theorien sind ne, also nicht unbedingt, wie man in der Mathematik jetzt sagt, das ist BEWIESEN, sondern wir wissen ja, dass es empirisch/haben wir ja festgestellt, dass das so richtig ist“ (Hervorhebung durch den Autor). Ebenso formuliert BEHGLA17 (Pos. 45): „[...] ,wenn ich mir quasi auch schon einen Überblick darüber verschafft habe, ja, was, was da fachwissenschaftlich denn RICHTIG ist, oder so.“. Die Proband*innen beschreiben so Evidenzen, welche richtig sind.

6.6.3.5 Ergebnisse der zweiten Phase der Interventionsstudie

Im Folgenden werden die Ergebnisse der zweiten Phase der Interventionsstudie berichtet. Dabei kann wie berichtet auf 3 Proband*innen zurückgegriffen werden. Entsprechend können lediglich deskriptive Ergebnisse berichtet werden.

EIDM-ET

Die Gesamtpunktzahlen für den EIDM-ET der drei Proband*innen am follow-up-Zeitpunkt liegen zwischen 93-170 Punkten. Mit Ausnahme des Probanden CAUEEE16 mit einer Punktzahl von 170 Punkten liegt der follow-up-Zeitpunkt damit im Bereich einer Standardabweichung um den Mittelwert des Post-Zeitpunktes.

Während für die beiden Probandinnen AAAMAA08 ($M_{\text{post}}=144.00$, $M_{\text{follow-up}}=126.00$) und UADFJA17 ($M_{\text{post}}=148.00$, $M_{\text{follow-up}}=93.00$) eine Abnahme der Punktzahl zwischen Post-Test und follow-up zu erkennen ist, erreicht der Proband CAUEEE16 ($M_{\text{post}}=166.00$, $M_{\text{follow-up}}=170.00$) zum follow-up-Zeitpunkt eine leicht höhere Punktzahl. Aufgrund der sehr kleinen Stichprobengröße am follow-up-Zeitpunkt, können jedoch keine Mittelwertsunterschiede berechnet werden.

EBP beliefs scale

Für die drei Sub-Skalen der *EBP beliefs scale* sind unterschiedliche Entwicklungen zum follow-up-Zeitpunkt zu erkennen. Während für *Beliefs_EBP_Bildung* eine Abnahme zwischen Post- und follow-up-Zeitpunkt für alle drei Proband*innen zu erkennen ist ($M_{\text{AAAMAA08_post}}=2.33$, $M_{\text{AAAMAA08_follow-up}}=2.00$; $M_{\text{UADFJA17_post}}=4.33$, $M_{\text{UADFJA17_follow-up}}=4.00$; $M_{\text{CAUEEE16_post}}=4.33$, $M_{\text{CAUEEE16_follow-up}}=4.00$), zeigt sich für die anderen beiden Subskalen ein anderes Bild. So zeigt sich für die Probandinnen AAAMAA08 und UADFJA17 für die Sub-Skalen *Beliefs_EBP_Impl_Schule* ($M_{\text{AAAMAA08_post}}=2.33$, $M_{\text{AAAMAA08_follow-up}}=2.00$; $M_{\text{UADFJA17_post}}=3.33$, $M_{\text{UADFJA17_follow-up}}=2.67$) und *Beliefs_EBP_Fähigkeiten* ($M_{\text{AAAMAA08_post}}=2.25$, $M_{\text{AAAMAA08_follow-up}}=1.25$; $M_{\text{UADFJA17_post}}=3.25$, $M_{\text{UADFJA17_follow-up}}=3.00$) eine Abnahme zwischen Post- und follow-up-Zeitpunkt, während für CAUEEE16 für beide Sub-Skalen eine Zunahme des Skalenniveaus berichtet werden kann

($M_{\text{CAUEEE16_Beliefs_EBP_Impl_Schule_post}}=2.33$, $M_{\text{CAUEEE16_Beliefs_EBP_Impl_Schule_follow-up}}=4.67$; $M_{\text{CAUEEE16_Beliefs_EBP_Fähigkeiten_post}}=2.00$, $M_{\text{CAUEEE16_Beliefs_EBP_Fähigkeiten_follow-up}}=4.25$).

Für die Probandin AAAMAA08 kann dabei für alle drei Sub-Skalen eine Abnahme des Skalenwerts zwischen Prä-Test und follow-up-Test festgestellt werden. Dabei ist ebenfalls keine Steigerung des Skalenwerts zwischen Prä- und Post-Zeitpunkt zu erkennen. Im Gegenteil dazu kann für Proband CAUEEE16 für *Beliefs_EBP_Impl_Schule* und *Beliefs_EBP_Fähigkeiten* eine durchgängige Steigerung des Skalenwerts beschrieben werden. Ebenso liegen die Skalenwerte am follow-up-Zeitpunkt für diesen Probanden höher als die Mittelwerte der Gesamtstichprobe zum Post-Zeitpunkt. Für Probandin UADFJA17 können jeweils Steigerungen des Skalenwerts zum Post-Test und eine anschließende Abnahme des Skalenwerts zum follow-up-Zeitpunkt erkannt werden.

EBP implementation scale

Auch für die *EBP implementation scale* sind unterschiedliche Entwicklungen für die drei Proband*innen zu beobachten. Für Probandin AAAMAA08 steigt der Mittelwert von Prä- zu follow-up-Zeitpunkt kontinuierlich an ($M_{\text{prä}}=1.00$, $M_{\text{follow-up}}=1.38$). Dabei bleibt der Wert verglichen mit dem Mittelwert der Gesamtstichprobe jedoch unterdurchschnittlich. Probandin UADFJA17 zeigt einen leichten Anstieg zum Post-Zeitpunkt ($M_{\text{prä}}=1.44$, $M_{\text{post}}=1.63$), welcher leicht überdurchschnittlich ausfällt. Zum follow-up-Zeitpunkt sinkt der Skalenwert jedoch wieder ($M_{\text{follow-up}}=1.25$). In diesem Fall sinkt der Skalenwert unter den des Prä-Zeitpunktes. Für den Probanden CAUEEE16 ist zwischen Prä- und follow-up-Zeitpunkt ein deutlicher Anstieg des Skalenwerts erkennbar ($M_{\text{prä}}=1.69$, $M_{\text{follow-up}}=2.31$), wenngleich dieser im Post-Test ($M_{\text{post}}=1.50$) vorerst sinkt. Während der Wert zum Post-Zeitpunkt leicht unterdurchschnittlich ausfällt, zeigt der Proband zum follow-up-Zeitpunkt eine deutliche Steigerung weit über den durchschnittlichen Skalenwert der Gesamtstichprobe am Post-Zeitpunkt.

Interview

Awareness

Innerhalb der zweiten Phase der Interventionsstudie unterscheiden sich die Aussagen der Proband*innen zum Mechanismus *Awareness* nur geringfügig von den Beschreibungen innerhalb des Post-Zeitpunktes. Die Rolle von Evidenzen für die Schule wird nach wie vor als äußerst relevant beschrieben, wenngleich die Bedeutung für das universitäre Studium als noch größer beschrieben wird (UADFJA17_II. Quartal). Im Zusammenhang mit unterrichtlichen Entscheidungen beschreibt CAUEEE16 (II. Quartal, Pos. 47):

Also wenn ich jetzt zum Beispiel hergehe und einfach nur so eine ganz Kleinigkeit habe, dann sagt mir der Kollege ‚Ja, versuch das so und so‘, dann mache ich das einfach mal. Dann gleiche ich das natürlich nicht jedes Mal ab. Das ist klar.

In diesem Zusammenhang beschreibt der Proband die Rolle eigener, oder tradierter Erfahrungen von Kollegen für unterrichtliche Entscheidungen. Gleichzeitig betont CAUEEE16, für einflussreichere Entscheidungen Evidenzen zu recherchieren.

Im Zusammenhang mit der Vorläufigkeit wissenschaftlicher Evidenz sind keine Veränderungen zwischen Post- und follow-up-Zeitpunkt zu erkennen. Die Proband*innen beschreiben jedoch, dass die Abgrenzung gültiger und nicht mehr gültiger Evidenzen im Kontext des Unterrichts relevant sei. So müssten nicht mehr gültige Theorien, also Theorien, welche von dem aktuellen wissenschaftlichen Konsens abweichen, für die Schüler*innen klar benannt werden (UADFJA17_IV. Quartal).

Agree

Für den Mechanismus *Agree* könnten ebenfalls sehr ähnliche Beschreibungen der Proband*innen im Vergleich zum Post-Zeitpunkt erkannt werden. Aus einer unterrichtspraktischen Perspektive bringt CAUEEE16 Evidenzen aus Schulbüchern in die Diskussion ein. Dabei fokussiert der Proband auf fachlich inadäquate Evidenzen, welche z.B. durch Fehldrucke entstanden sein könnten. Hier stellt der Proband die Rolle der Individualität von Evidenzen in den Vordergrund. So sollten die Schüler*innen in diesen Fällen die Evidenzen selbst bewerten. Andererseits beschreibt der Proband die Autorität der Lehrkraft im Unterrichtsraum als relevant. So würden die Schüler*innen eher der Lehrkraft glauben als einem Schulbuch. Dies gelte sogar dann, wenn innerhalb des Schulbuchs eine Aussage durch Personen mit höheren akademischen Abschlüssen getätigt werden. Während der akademische Ausbildungsgrad einen Einfluss auf die Glaubwürdigkeit habe, stelle die physische Anwesenheit und Ansprechbarkeit der Lehrkraft hier einen wichtigen Aspekt dar.

Im Rahmen des follow-up-Interviews (III. Quartal) ergänzt UADFJA17 die Anwendung wissenschaftlicher Gütekriterien auf Evidenzen als weiteres Kriterium für Evidenzen. Halten diese der Prüfung entlang der Kriterien stand, können Sie als Evidenzen bezeichnet werden.

Access and Communication

Bzgl. Des Mechanismus *Access and Communication* unterstützen die Proband*innen zum follow-up-Zeitpunkt eine notwendige frühere und explizitere Thematisierung evidenzbasierter Praxis im Studium (z.B. CAUEEE16_II. Quartal, AAAMAA08_II. Quartal). Gleichzeitig erkennen die Proband*innen rückblickend mehr Situationen innerhalb des Studiums, in denen evidenzbasierte Praxis praktiziert wurden (z.B. CAUEEE16_III. Quartal, UADFJA17_III. Quartal). Dennoch beschreibt AAAMAA08 (II. Quartal, Pos. 59), dass evidenzbasierte Praxis im Studium:

[...] eher zu kurz gekommen beziehungsweise in den Praktika schon fast zurückgesetzt worden [sei], weil gerade im quantitativen Teil, man hat sich durchgemogelt und das war ja einfach nicht mehr, man hat es bewiesen, sondern man hat sich irgendwie ausgerechnet, wie jetzt meine Spanne ist, damit ich noch bestehen muss oder so.“

Die Probandin beschreibt hier ein chemisches Experimentalpraktikum des Fachs Analytische Chemie. Das ‚Durchmogeln‘ habe hier den eigentlichen evidenzbasierten Prozess des Experiments verhindert. CAUEEE19 (IV. Quartal) beschreibt zusätzlich eine wahrgenommene Steigerung der eigenen Kompetenzen auch nach dem Studium bis zum letzten Erhebungszeitpunkt. Auf die Frage, wodurch die Kompetenzen weiter ausgebaut werden konnten, antwortet der Proband:

Das kann ich ganz klar und deutlich beantworten. Das ist einzig und allein durch unsere Interaktion, dadurch, dass wir diese Interviews jetzt mehrfach geführt haben. Wenn man das quasi ins Studium verlagern würde oder zumindest mal als so eine Art ansprechen wurde, dass man sich KONKRET mit diesen Evidenzen, mit dem Sinn von Evidenzen auseinandersetzt, mit Arten von Evidenzen und so weiter und so fort. Und das nicht nur einmal macht, sondern vielleicht über die Zeit immer mal wieder aufgreift, dass das genau das ist, was es für MICH jetzt ausgemacht hat. (CAUEEE16_IV. Quartal, Pos. 144).

Damit hebt CAUEEE16 (IV. Quartal) die Rolle der Reflektion und Explikation evidenzbasierter Praxis für den Unterricht hervor.

Interact

Im Rahmen des follow-up-Interviews werden durch die Proband*innen bzgl. des Mechanismus *Interact* teils sehr konkrete Beispiele für den Umgang mit Evidenzen durch Schüler*innen beschrieben. So stellt UADFJA17 (II. Quartal) ein Experiment zur Konzentrationsbestimmung von Zucker in einem Softgetränk vor. Diese Konzentration sollte dabei vor dem Hintergrund recherchierter Evidenzen bewertet werden:

Und dafür habe ich dann halt auch so ein bisschen im Internet recherchiert und ganz viele Artikel gefunden, wo eben beschrieben wird, warum die Fanta-Kon/ also die Zuckerkonzentration in Fanta unterschiedlich ist in anderen Ländern und wie das auch so mit der Politik zusammenhängt. (UADFJA17_II. Quartal, Pos. 41)

CAUEEE16 (II. Quartal) beschreibt in diesem Zusammenhang die Bedeutung von Evidenzen für Schüler*innen:

Für die ist das komplett neu. Und deswegen muss man hier Evidenzen schaffen oder Evidenzen darlegen, damit die Schülerinnen und Schüler überhaupt erstmal in der Lage sind, das Thema nachzuvollziehen und auch zu akzeptieren. Wenn man einfach nur, ja, irgendetwas in den Raum stellt, ist es für die natürlich erstmal schwer. Klar, wenn ich als Lehrkraft etwas behaupte, dann ist es für die Schülerinnen und Schüler meistens auch schon eine Art Evidenz, denn ich habe für die Schüler sicherlich eine höhere Expertise, oder aus Sicht der Schüler, und dementsprechend werden sie das auch erst einmal so hinnehmen. Ich möchte natürlich auch trotzdem erreichen, dass sie das Ganze hinterfragen und auch selbstständig versuchen nachzuvollziehen. (CAUEEE16_II. Quartal, Pos. 21)

Dabei beschreibt der Proband auch die Bedeutung von Ergebnissen aus Vorstunden für den Erkenntnisprozess der Schüler*innen. Diese sollen in die Argumentation eingebettet werden sodass hier weitere, teils selbst generierte, Evidenzen einbezogen werden. AAAMAA08 (II. Quartal) beschreibt hingegen, dass Evidenzen im Unterricht nicht immer expliziert werden. Dennoch werden Evidenzen z.B. in Form von Arbeitsmaterialien in den Unterricht eingebracht. UADFJA17 (III. Quartal, Pos. 87) stellt jedoch die Tätigkeit der Schüler*innen selbst in den Vordergrund: „Also ich glaube, für die Schüler ist es immer schöner, wenn sie selber etwas tun können oder irgendwie nachvollziehen können und nicht einfach einen Test lesen und das glauben müssen“.

Insgesamt stellen die Proband*innen innerhalb der zweiten Phase der Studie Evidenzen in das Zentrum von Unterricht. Dies wird vor allem mit der Rolle von Evidenzen für das Fach Chemie begründet, welche häufig auf das Experiment abgestellt wird (z.B. UADFJA17_III. Quartal). Dabei beschreibt CAUEEE16 (III. Quartal), dass innerhalb des Unterrichts auch ein Bedarf nach Evidenzen entstehen kann, welcher dann experimentell erfüllt werden kann: „[...] wenn so Sachen im Unterricht aufkommen, wo ich dann selber weiß, die kann man jetzt relativ leicht, schnell und einfach experimentell überprüfen, dass man dann einfach mal spontan noch so ein Experiment mit einbauen kann, funktioniert auch.“ (CAUEEE16_III. Quartal, Pos. 43). Gleichzeitig beschreibt der Proband, dass nicht alle Evidenzen innerhalb des Unterrichts generiert werden könnten. Insbesondere experimentell aufwendige Evidenzen könnten so z.B. nur durch die Lehrkraft in den Unterricht eingebracht werden. So stellt der Proband auch die Rolle der Lehrkraft heraus, um Evidenzen darzulegen. UADFJA17 (IV. Quartal) beschreibt in diesem Zusammenhang einen notwendigen Mix der eingebrachten Evidenzen. So sollten experimentell durch die Schüler*innen generierbare Evidenzen vorgezogen werden. Gleichzeitig beschreibt CAUEEE16 hier jedoch einen Unterschied in der Qualität der Evidenzen.

CAUEEE16 (IV. Quartal) stellt auch die Bedeutung widersprechender Evidenzen für Schüler*innen dar. So könnten innerhalb eines Experiments unterschiedliche Ergebnisse bei unterschiedlichen Schüler*innengruppen auftreten:

Bei einer Gruppe färbt es sich gelb, bei zwei rot oder was auch immer und bei drei grün, aber bei keinem blau und die Lehrkraft sagt am Ende ‚Eigentlich sollte sich das aber orange färben‘. So, dann hat man ganz, ganz viele unterschiedliche Sachen und dann ist es für die Schüler natürlich sehr, sehr schwer, das jetzt nachzuvollziehen. (CAUEEE16_IV. Quartal, Pos. 95)

Zur Auflösung der Situation beschreibt der Proband den Einfluss der Autorität der Lehrkraft in Zusammenhang mit dem Experiment. So sollte das Experiment idealerweise erneut durch die Lehrkraft durchgeführt werden und dabei zu den erwünschten Beobachtungen führen. Zwar werden in diesem Fall gemeinsam von Lehrkraft und Schüler*innen Evidenzen generiert, jedoch wird die Glaubwürdigkeit durch die Durchführung durch die Lehrkraft im Sinne der Autorität erhöht. In diesem Fall könnten belastbare Evidenzen zur Klärung der Fragestellung führen.

Insgesamt können innerhalb von Phase 2 ($M=7.78$ Kodierungen pro Interview) deutlich mehr Kodierungen der Kategorie *Interact* festgestellt werden, als dies innerhalb von Phase 1 ($M=2.77$ Kodierungen pro Interview) der Fall ist.

Skills

Innerhalb von Phase 2 beschreiben die Proband*innen hingegen die Rolle von Evidenzen für die Planung von Unterricht innerhalb des Mechanismus *Skills*. Dabei nennen die Proband*innen vor allem die Unterrichtsvorbereitung als relevante evidenzbasierte Tätigkeit (z.B. AAAMAA08_IV. Quartal, CAUEEE16_IV. Quartal). UADFJA17 (IV. Quartal, Pos. 129) beschreibt weiterhin Schwierigkeiten zur Grenzziehung für relevantes Hintergrundwissen der Lehrkraft, welches durch Evidenzen generiert werden sollte: „Aber dann jetzt erstmal überhaupt nochmal zu entscheiden, was ist denn dieses vielmehr, was die Schüler NICHT wissen, was verwirrt sie? Das finde ich auch schwierig.“.

Structure and Process

In Anlehnung an die Ergebnisse der Post-Interviews zum Mechanismus *Structure and Process* zeigt sich auch im Rahmen der follow-up-Interviews eine erwartete abnehmende Bedeutung von Evidenzen im Vergleich zu eigenen Erfahrungen. Hier beschreibt AAAMAA08 (IV. Quartal, Pos. 105), dass eher fachdidaktische Evidenzen relevant sein könnten: „[...] ich glaube schon, dass ich mich auch noch weiter mit Didaktik auseinandersetzen werde, aber es ist natürlich eine andere Intensität.“.

Die Beschreibung von Evidenzen als ‚richtig‘, welche innerhalb der Post-Interviews beschrieben wurden, kann auch innerhalb der 2. Phase wiedergefunden werden (z.B. AAAMAA08_II. Quartal, UADFJA17_III. Quartal). Auch bzgl. der Beschreibung von Evidenzen und dem Umgang mit diesen sind kaum Unterschiede zwischen Phase 1 und 2 der Studie erkennbar. Jedoch berichten die Proband*innen in der 2. Phase häufiger über unterrichtspraktische Aspekte im Umgang mit Evidenzen. So werden häufig konkrete Beispiele für den Umgang mit Evidenzen von Schüler*innen genannt. Die Bedeutung von evidenzbasierter Praxis für den Unterricht wird durchgehend beschrieben. So spiele evidenzbasierte Praxis für alle an Schule beteiligten Personen eine Rolle, jedoch vor allem für solche, welche einflussreiche Entscheidungen treffen (z.B. Schulleitung). Für den Aufbau eigener Kompetenzen in evidenzbasierter Praxis verweisen die Proband*innen auf Aspekte ihres

Studiums, wie auch die durchgeführte Intervention. Dabei kann in Phase 2 beobachtet werden, dass die Proband*innen ihre Vorbereitung auf die Implementation evidenzbasierter Praxis durch das Studium insgesamt positiver zu bewerten scheinen, als dies innerhalb von Phase 1 der Fall ist.

6.6.4 Zusammenführung der Ergebnisse

Im Rahmen der vorliegenden Interventionsstudie wurden sowohl quantitative als auch qualitative Methoden eingesetzt. Dabei können aus der Zusammenführung der unterschiedlichen Ergebnisse Erkenntnisse abgeleitet werden.

Die im Rahmen des Mechanismus *Awareness* im Interview beschriebene Bedeutung von Evidenzen innerhalb der Schule kann durch das Skalenniveau oberhalb des Skalenmittelwerts der *Beliefs_EBP_Bildung*-Sub-Skala unterstützt werden. Dabei beschreiben die Proband*innen eine generelle Bedeutung evidenzbasierter Praxis für Bildungsprozesse. Dabei können innerhalb des Interviews viele unterschiedliche Themenfelder aufgezählt werden, in welcher evidenzbasierte Praxis eine wichtige Rolle spielt. Gleichzeitig werden jedoch Bedenken formuliert, inwiefern innerhalb der Schulpraxis Zeit bleibt, um evidenzbasierte Tätigkeiten auszuführen. Dabei könnte hier ebenfalls eine Verbindung zu dem geringen Skalenniveau der *EBP implementation scale* erkannt werden. Sowohl am Prä- als auch am Post-Zeitpunkt geben die Proband*innen an wenig mit Evidenzen zu arbeiten. Somit könnte der am Post-Zeitpunkt formulierte Zeitmangel als Validitätsargument für den niedrigen Skalenwert der *EBP implementation scale* angeführt werden.

Innerhalb des Mechanismus *Agree* fokussieren die Proband*innen innerhalb der Interviews auf Kriterien für Evidenzen. Dabei werden im Rahmen der Post-Interviews keine wissenschaftlichen Gütekriterien beschrieben. Dies könnte dabei mit dem durchschnittlichen Abschneiden der Proband*innen in Item 7 in Verbindung gebracht werden. Dieses Item fragt nach Charakteristika von Befunden, damit diese als valide bezeichnet werden können. Das durchschnittliche Abschneiden innerhalb des Items könnte sich dabei auch durch eine fehlende wahrgenommene Bedeutung wissenschaftlicher Gütekriterien innerhalb der Interviews erklären lassen. Dieser Zusammenhang könnte durch das bessere Abschneiden von UADFJA17 im follow-up-Test unterstützt werden, da diese Proband*innen die Bedeutung wissenschaftlicher Gütekriterien hier explizit hervorhebt.

In Bezug auf den Mechanismus *Access and Communication* fokussieren die Interviews vor allem auf die Vorbereitung auf die Implementation evidenzbasierter Tätigkeiten durch das universitäre Studium. Aus den Beschreibungen der Proband*innen ergibt sich ein gemischtes Bild. Während einerseits eine gute Vorbereitung vor allem innerhalb der fachdidaktischen Ausbildung beschrieben wird, bemängeln die Proband*innen vor allem die ausgebliebene Explikation evidenzbasierter Tätigkeiten im Rahmen des Studiums. Diese Ergebnisse können dabei durch die *Beliefs_EBP_Fähigkeiten*-Sub-Skala unterstützt werden. Auch hier erreichen die Proband*innen ein Skalenniveau nur knapp über dem Skalenmittelwert. Dabei können die Ergebnisse der Analyse der Interviews mögliche Begründungen für das Skalenniveau darstellen. Auch die Punktzahl des EIDM-ET zum Prä-Zeitpunkt könnte bzgl. der Vorbereitung auf evidenzbasierte Praxis durch das Studium interpretiert werden.

Der Zusammenhang zwischen der wahrgenommenen Vorbereitung durch das Studium auf die Wahrnehmung der eigenen Fähigkeiten sowie die Punktzahl innerhalb des EIDM-ET könnte für CAUEEE16 im Rahmen der zweiten Phase unterstützt werden. Der Proband beschreibt innerhalb des Interviews eine gute Vorbereitung, welche vor allem auf die Explikation im Rahmen der durchgeführten Interviews zurückgeführt wird. Dabei kann für den Probanden am follow-up-Zeitpunkt auch ein höherer Skalenwert für *Beliefs_EBP_Fähigkeiten* sowie eine höhere Punktzahl innerhalb

des EIDM-ET erkannt werden. Somit könnten hier Validitätsargumente für die Erhebung der Kompetenz durch den EIDM-ET abgeleitet werden.

Der Mechanismus *Interact* stellt den Umgang von Schüler*innen mit Evidenzen in den Mittelpunkt. Dabei werden im Rahmen der zweiten Phase der Interventionsstudie deutlich konkretere Situationen berichtet, als dies in der ersten Phase der Fall ist. Hier könnte ein Zusammenhang zum besseren Abschneiden von AAAMAA08 sowie CAUEEE16 innerhalb der *EBP implementation scale* hergestellt werden. So könnte die häufigere Tätigkeit in evidenzbasierter Praxis mit der Umsetzung im Unterricht und damit konkretisierten Beispielen im Rahmen des Interviews erklärt werden.

Innerhalb des Mechanismus *Skills* werden die Fähigkeiten der Proband*innen in den Fokus gestellt. Dabei könnten Bezüge zwischen den Ergebnissen der Interviews sowie Items 1-3 des EIDM-ET hergestellt werden. Die ersten drei Items des EIDM-ET fokussieren vor allem auf die Durchführung eines Forschungsprozesses auf Grundlage einer spezifischen Situationsbeschreibung. Die Proband*innen schätzen die eigenen Fähigkeiten zur Bewertung der Qualität von Artikeln sowie der Durchführung kleiner Forschungsstudien innerhalb des Post-Interviews als ausreichend ein. Dabei verweisen die Proband*innen jedoch auf fehlende vertiefte Kompetenzen. Verglichen mit der erreichten Punktzahl innerhalb der Items im EIDM-ET könnte diese Einschätzung unterstützt werden. Die Proband*innen erreichen vor allem in Item 1 und 3 überdurchschnittliche Punktzahlen. Dies könnte auf grundlegende Fähigkeiten zur Durchführung eines Forschungsprozesses hinweisen. Für Item 2 kann hingegen nur eine durchschnittliche Punktzahl im Vergleich zu den anderen Items des EIDM-ET erreicht werden. Dabei scheint dies vor dem Hintergrund der häufigeren Nennung von Quellen für fachdidaktische im Vergleich zu fachlichen Fragestellungen zu verwundern.

Andererseits könnte hier ein Einfluss des formulierten graduellen Verständnisses von Evidenzen bzw. der Bedeutung von Evidenzen für unterschiedliche Inhaltsbereiche erkannt werden. So beschreiben die Proband*innen innerhalb der Mechanismen *Skills* sowie *Structures and Processes* von einer abnehmenden Bedeutung von Evidenzen mit zunehmendem Erfahrungswissen. Im EIDM-ET gefragt nach Quelltypen sowie deren Vor- und Nachteilen könnte so eine geringere wahrgenommene Bedeutung von Evidenzen für fachdidaktisch-unterrichtliche Fragestellungen die niedrigere Punktzahl erläutern. Dem gegenüber steht jedoch das berichtete Skalenniveau von *Beliefs_EBP_Bildung* oberhalb des Skalenmittelwerts. Jedoch differenziert diese Sub-Skala der *EBP beliefs scale* nicht zwischen fachdidaktischen und fachwissenschaftlichen Fragestellungen. Entsprechend könnte hier eine generelle Bedeutung evidenzbasierter Praxis für Bildungsprozesse erkannt werden, welche jedoch für den Bereich fachdidaktischer Fragestellungen hinter der Bedeutung von Erfahrungswissen zurücktritt. So könnte die durchschnittliche Punktzahl von Item 2 des EIDM-ET erklärt werden.

Die Beschreibung von lediglich grundlegenden Kompetenzen in der Durchführung einer eigenen Forschungsstudie im Mechanismus *Skills* kann ebenfalls durch das Abschneiden der Proband*innen für Item 5 und 6 des EIDM-ET unterstützt werden. Vor allem für die Frage nach geeigneten Studiendesigns für spezifische Erkenntnisinteressen zeigen die Proband*innen in Item 5 unterdurchschnittliche Punktzahlen. Für die Definition unterschiedlicher forschungsmethodologischer Begriffe in Item 6 sind teils leicht überdurchschnittliche Punktzahlen erreicht worden. Damit könnte ein oberflächliches deklaratives Wissen zu Forschungsmethoden vorhanden sein, während vertiefte Kenntnisse über geeignete Studiendesigns eher unterdurchschnittlich ausgeprägt sind.

6.6.5 Diskussion der Ergebnisse der Interventionsstudie zur Förderung angehender Chemielehrkräfte in evidenzbasierter Praxis

In der vorliegenden Studie wurde der Einfluss einer an *evidence-use* Mechanismen (Breckon & Dodson, 2016) orientierten Intervention auf die Kompetenzen in evidenzbasierter Praxis von angehenden Chemielehrkräften untersucht.

Die Ergebnisse des EIDM-ET zeigen dabei einen Einfluss der Intervention auf das forschungsmethodologische Wissen, Kompetenzen im Forschungsprozess sowie die Bewertung von Evidenzen bzgl. ihrer Umsetzbarkeit in der Unterrichtspraxis auf.

Im forschungsmethodologischen Wissen kann ein relevantes Konzept zur Bewertung von Evidenzen sowie der Planung von Studien erkannt werden. Ohne ein Wissen über grundsätzliche Begriffe und Methoden innerhalb der Forschung ist es nur schwer möglich Studien zu planen. Im Sinne Schommer-Aikins' (2002) Verständnis eines *solid core*, könnte das forschungsmethodologische Wissen die notwendige Ausgangslage zum Denken innerhalb von Forschungszusammenhängen darstellen. Insofern kann ein Einfluss forschungsmethodologischen Wissens auf die Kompetenzen in evidenzbasierter Praxis sowohl auf Grundlage der Ergebnisse der vorliegenden Studie als auch auf theoretischer Ebene begründet werden.

Dabei können auf Grundlage der Ergebnisse dieser Studie Indizien für Unterschiede im forschungsmethodologischen Wissen der Proband*innen erkannt werden. Während eine Förderung des forschungsmethodologischen Wissens auf Ebene von deklarativem Wissen durch die Intervention beobachtet werden konnte, scheint die Anwendung forschungsmethodologischer Konzepte im Sinne der Signifikanz und Bedeutung von Befunden sowie die Entwicklung eines Studiendesigns die Proband*innen auch zum Post-Zeitpunkt vor Herausforderungen zu stellen. Während Stelter und Miethe (2019) auf eine seltenere Thematisierung forschungsmethodologischer Aspekte in der Lehramtsausbildung im Vergleich zu bildungs- und erziehungswissenschaftlichen Studiengängen hinweisen, bedeutet dies nicht zwangsläufig, dass keinerlei forschungsmethodologische Aspekte thematisiert werden. So weisen die Autor*innen auch darauf hin, dass vor allem weniger quantitative und qualitative Methoden innerhalb des Lehramtsstudiums thematisiert werden. Die umfassendere Anwendung forschungsmethodologischer Aspekte könnte dabei vertieftes forschungsmethodologisches Wissen z.B. zu unterschiedlichen Studiendesigns voraussetzen, welches im Rahmen des Studiums und der Intervention nicht erworben werden konnte. Grundsätzliches Wissen zu Gütekriterien wissenschaftlicher Forschung könnten jedoch im Rahmen des Studiums thematisiert worden sein, da diese auch über unterschiedliche Methoden hinweg diskutiert werden können. Somit könnten diese eher Einzug in die Thematisierung forschungsmethodologischer Aspekte innerhalb des Studiums gefunden haben. Um dieser Annahme weitergehend nachgehen zu können, bedarf es der Analyse tatsächlich thematisierter forschungsmethodologischer Inhalte im Studium der Proband*innen. Stelter und Miethe (2019) bemängeln hier, dass auf Grundlage von existierenden Modulhandbüchern keine validen Aussagen über die Inhalte getroffen werden können. Die Handbücher würden zumeist eher abstrakte Konzepte beschreiben, sodass keine Einblicke in die reale Thematisierung der Inhalte möglich sind. Es bedürfte der Erhebung tatsächlicher thematisierter Inhalte, um hier einen Rückbezug zur Beantwortung der unterschiedlichen forschungsmethodologischen Items herzustellen. Festgehalten werden kann jedoch, dass es einen Bedarf an der Förderung forschungsmethodologischem Wissens bei den Proband*innen zu geben scheint, was wiederum Einfluss auf deren Kompetenz in evidenzbasierter Praxis nehmen könnte.

Auf Grundlage der Ergebnisse zur Gestaltung eines Forschungsprozesses aus dem EIDM-ET sowie den Interviews, kann auch die Bedeutung der Kontextualisierung für evidenzbasierte Praxis

diskutiert werden. Vor dem Hintergrund kontextabhängiger epistemologischer Überzeugungen (Schommer-Aikins, 2002), könnte die Situierung als Bedingung für die Implementation evidenzbasierter Praxis erklärt werden. Dabei könnten jedoch auch akademische epistemologische Überzeugungen vorliegen (Muis et al., 2006), welche auch außerhalb spezifischer Kontexte zu einer evidenzbasierten Praxis beitragen könnten. Somit müsste ein grundlegendes Niveau evidenzbasierter Praxis auch aufgrund nicht-kontextualisierter Überzeugungen möglich sein. Indizien für diese Argumentation können in der Bearbeitung der bildungswissenschaftlichen Situationsbeschreibung im EIDM-ET sowie dem generellen Vergleich von Rechercheprozessen in verschiedenen Fachbereichen innerhalb der Interviews erkannt werden. Dabei stellt die beschriebene Situationsbeschreibung auch einen Kontext für evidenzbasierte Praxis dar, welcher jedoch für die Proband*innen nicht im Bereich ihres studierten Unterrichtsfachs liegt. Somit könnten hier allgemeinere – akademische – epistemologische Überzeugungen eine Rolle spielen, als dies für die Situationsbeschreibung aus dem Chemieunterricht erwartet werden könnte.

Die Kontextualisierung könnte weiterhin zur ‚Konkretheit‘ beitragen, welche durch die Proband*innen im Rahmen des Interviews mit Erfahrungswissen in Beziehung gebracht wird. In Bezug auf Entwicklungen im Bereich evidenzbasierter Medizin könnte hier die spezifische Anwendbarkeit von Leitlinien auf individuelle Fälle beschrieben werden. Dabei kann diese Konkretheit zur Umsetzung evidenzbasierter Praxis helfen, jedoch gleichzeitig als entmündigend wahrgenommen werden (Dawes et al., 2005). Vor dem Hintergrund der Ergebnisse aus den Interviews scheint im Bildungsbereich bisher jedoch eher wenig Konkretheit von Evidenzen wahrgenommen zu werden. Somit könnte die Kontextualisierung evidenzbasierter Praxis durch den Einbezug spezifischer Situationen ggf. einen Beitrag zum Aufbau von Kompetenzen in evidenzbasierter Praxis leisten.

Aus Perspektive der Definition evidenzbasierter Praxis könnte ein positiver Einfluss auf die EBP-Kompetenzen von (angehenden) Lehrkräften durch Unterrichtspraxis im Sinne der Kontextualisierung und Konkretheit erklärbar sein. So gilt es im Rahmen evidenzbasierter Praxis wissenschaftliche Evidenzen im Rahmen eines spezifischen Kontextes mit eigenen Erfahrungen zu relationieren. Die tatsächliche Durchführung von Unterricht könnte hier spezifische Fragestellungen aufwerfen, welche im Rahmen evidenzbasierter Prozesse beantwortet werden können. Damit könnte die Kompetenz in evidenzbasierter Praxis durch die tatsächliche Durchführung dieser bzw. die Konfrontation mit relevanten Anwendungssituationen in Phase 2 gefördert werden. Weitere Indizien für diese Argumentation können in der deutlich zunehmenden Konkretheit von beschriebenen Anwendungssituationen evidenzbasierter Praxis innerhalb der Schulpraxis in Phase 2 erkannt werden. Dabei scheint vor allem eine Reflexion der eigenen Praxis einen deutlichen Mehrwert zu leisten. Dadurch könnte eine Explikation entstehen, welche im Post-Interview gefordert und im Rahmen des follow-up-Interviews (im Sinne der Reflexion) als besonders hilfreich beschrieben wird. Inwiefern die Zunahme evidenzbasierter Tätigkeiten innerhalb von Phase 2 (*EBP implementation scale*) mit der Kontextualisierung evidenzbasierter Praxis innerhalb der Schule zusammenhängt kann im Rahmen dieser Studie nicht abschließend geklärt werden. Die vorhandenen Indizien deuten jedoch auf eine Bedeutsamkeit der Schulpraxis für die Ausführung evidenzbasierter Tätigkeiten und damit der EBP-Kompetenz hin. Dabei kann jedoch nicht direkt auf eine Richtung des Zusammenhangs geschlossen werden. So könnte einerseits die Schulpraxis zur Wahrnehmung und Reflexion von evidenzbasierter Tätigkeiten und damit der Kompetenz in ebendieser beitragen, andererseits könnte eine gesteigerte Kompetenz zu häufigeren evidenzbasierten Tätigkeiten innerhalb der Schulpraxis führen.

In Kombination mit den Testergebnissen im EIDM-ET scheint ein fehlender Zuwachs der Selbsteinschätzung der eigenen Fähigkeiten in evidenzbasierter Praxis (*EBP_beliefs_Fähigkeiten*) zu

verwundern. Das bessere Abschneiden im EIDM-ET lässt auf eine gesteigerte Kompetenz zum Post-Zeitpunkt schließen. Diese scheint jedoch nicht mit einer besseren Selbsteinschätzung einherzugehen. Eine mögliche Erklärung könnte in einer fehlenden eigenen Wahrnehmung als kompetent im Umgang mit evidenzbasierten Praktiken liegen. Innerhalb der Intervention wurden an vielen Stellen Evidenzen verarbeitet. Diese Prozesse sind dabei an spezifische Kontexte innerhalb der jeweiligen Projekte gekoppelt. Die *EBP implementation scale* fragt hingegen eher allgemeine Situationen ab. So könnte hier ein Einfluss der Kontextabhängigkeit auf die Selbsteinschätzung entstehen. Die Proband*innen könnten sich in den spezifischen Kontexten ihrer Projekte als kompetent wahrnehmen, dies jedoch nicht auf die Items der Skala übertragen, da diese nicht an den jeweiligen Kontext anschließen.

So bräuchte es ggf. die Abstraktion der Tätigkeiten im eigenen Projekt im Kontext evidenzbasierter Praxis, um diese abstrahierten Konzepte dann auf die Selbsteinschätzung in allgemeinen Situationen übertragen zu können. Im Rahmen der Interviews wird innerhalb des Mechanismus *Skills* ebenfalls eine generelle Fähigkeit zur Bewertung von Artikeln und der Durchführung einer kleinen Studie beschrieben. Gleichzeitig weisen die Proband*innen darauf hin diese Tätigkeiten auf einem basalen Level zu beherrschen. Es werden keine vertieften Kompetenzen berichtet. Dies könnte einerseits als Validitätsargument bzgl. des leicht positiven Skalenwerts von *Beliefs_EBP_Fähigkeiten* der *EBP beliefs scale* gewertet werden. Andererseits kann im Interview ein kontextualisierter Zugang zu evidenzbasierter Praxis erkannt werden, da ein Fokus auf den Chemieunterricht vorherrscht. Ggf. könnten hier fachspezifische epistemologische Überzeugungen aktiviert werden (Muis et al., 2006), welche zu einer konkreteren Beantwortung der Fragen befähigen könnten. Tiefere Einblick in die tatsächliche Selbsteinschätzung der eigenen Fähigkeiten der Proband*innen könnten ggf. durch eine Anpassung der Items der *EBP beliefs scale* erreicht werden.

In Bezug auf die Bedeutung evidenzbasierter Praxis für die Schule (in Anlehnung an *Beliefs_EBP_Impl_Schule*) können innerhalb der Interviews viele verschiedene Beispiele berichtet werden. Dabei wird eine Relevanz des Konstrukts vor allem für Personengruppen beschrieben, welche sich in Machtpositionen befinden (z.B. Schulleitung, etc., vgl. Mechanismus *Structures and Processes*). In Bezug auf zugrundeliegende epistemologische Überzeugungen könnte hier ein Bezug zur Dimension *Source of knowledge* erkannt werden (vgl. Kap. 3.2). So wird durch die Hervorhebung bestimmter Personengruppen ggf. ein Verständnis von Autoritäten innerhalb evidenzbasierter Praxis beschrieben. Dabei kann diese Autorität im institutionalisierten, hierarchischen System Schule wiedererkannt werden. In einem organisatorischen Sinne stellt z.B. die Schulleitung die Funktion von Vorgesetzten dar, welche eine Autorität für Fragen der Schulentwicklung und -organisation darstellt. Diese Autorität scheint hier mit einem erhöhten Bedarf an evidenzbasierter Praxis verknüpft zu werden. Andererseits könnte die Verknüpfung evidenzbasierter Praxis ausschließlich mit schulorganisatorischen Autoritäten zu einem epistemologischen Autoritäten-Verständnis führen. Zwar können Aussagen von Autoritäten, bzw. Expert*innen innerhalb der jeweiligen Bereiche als Evidenzen in einen Bewertungsprozess mit eingehen, jedoch stellen Überzeugungen der weiteren Dimensionen der Epistemologie die Notwendigkeit von weiteren Argumenten klar heraus (vgl. Kap 3.2).

Innerhalb von Phase 2 beschreiben die Proband*innen die Rolle der Lehrkraft als epistemologische Autorität im Unterricht. So könnten widersprüchliche Evidenzen im Klassenraum entstehen, welche durch die Lehrkraft eingeordnet werden müssen. Andererseits beschreiben die Proband*innen, dass diese widersprüchlichen Evidenzen durch die gemeinsame Generierung weiterer Evidenzen näher untersucht werden sollten. Jedoch spielt auch in diesem Zusammenhang die Eingebundenheit der Lehrkraft eine Rolle. So könnten die Evidenzen, welche durch

Zusammenarbeit mit der Lehrkraft erstellt werden, größere Bedeutung haben als selbst erstellte Evidenzen. Die Beschreibung dieses Zusammenhangs durch die Proband*innen lässt Vorstellungen innerhalb der Dimension *Source of knowledge* (Hofer & Pintrich, 1997) erkennen. Demnach sind sich die Proband*innen ihrer Rolle als epistemologische Autorität im Unterricht bewusst, während sie gleichzeitig versuchen die Evidenzen anstelle die Rolle der Lehrkraft im Unterricht in den Fokus zu stellen.

Auch im Sinne des *Model of argument* (Toulmin, 2003) bedarf es weiterer, unterstützender Argumente. So könnte die Empfehlung einer Schulleitung oder die Meinung der Lehrkraft z.B. als *backing* für eine Aussage herangezogen werden. In diesem Sinne würde die Expert*innen-Meinung hier als Unterstützungsaspekt für einen *warrant* genutzt und als solcher einen Einfluss auf den *claim* nehmen. Es bedürfte dennoch *data*, um eine Ausgangslage für die Argumentation zu bilden. Die Meinungen von Expert*innen könnte in verschiedenen Zusammenhängen auch an anderen Stellen des *Model of argument* integriert werden. Das hier beschriebene Autoritäten-Verständnis von Expert*innen könnte dabei auf unterschiedliche Beziehungen ausgeweitet werden. So beschreiben die Proband*innen innerhalb von Phase 2 der Studie Ausbildungslehrkräfte oder Kolleg*innen innerhalb des Referendariats als wichtige Quelle für Hinweise und Tipps. Auch hier könnte ein Autoritäten-Verhältnis zwischen den Personen vorliegen.

Die Rolle von Tipps und Hinweisen von Expert*innen für evidenzbasierte Praxis, welche aus der Erfahrung von Lehrkräften weitergegeben werden, wird auch innerhalb des Mechanismus *Agree* deutlich. Die Proband*innen beschreiben ein graduelles Verständnis von Evidenzen, in welchem Meinungen von Lehrkräften als Hinweise eingehen. Dabei spielt hier erneut die Kontextualisierung und Konkretheit von Erfahrungswissen eine Rolle. Damit bleiben Evidenzen und Meinungen zwar getrennte Entitäten, jedoch können begründete Meinungen graduellen Evidenzcharakter besitzen. In Bezug auf das *Model of argument* nach Toulmin (2003) sind hier Verschiebungen zu erkennen, welche hier in *einem Modified Model of argument* (Abb. 10) zusammengefasst werden können.

An die Stelle der *data* tritt nun die *opinion*. Die Proband*innen anerkennen die Rolle der Meinung von Expert*innen innerhalb der jeweiligen Bereiche. Dabei wird der Expert*innen-Status auch anhand der akademischen Grade sortiert (vgl. Kap. 6.6.3.5). So könnten Professor*innen der jeweiligen Bereiche als glaubwürdiger anerkannt werden als Lehrkräfte ohne zusätzliche wissenschaftliche Qualifikation. Dabei spielt jedoch für alle geäußerten Meinungen die jeweilige Begründung eine Rolle. Diese sollte auf systematischen Verfahren oder der eigenen Rezeption von Literatur beruhen. Entsprechend sollte die *opinion* auf *data* oder *arguments* aufgebaut werden. Entsprechend wird *data* im *Modified Model of argument* (Abb. 10) nicht durch *opinion* ersetzt. *Data* ist weiterhin nötig, um *opinion* zu begründen. Dabei kann sich in der Begründung von *opinion* durch *data* oder *arguments* auch der Glaubwürdigkeitsgrad ausdrücken.

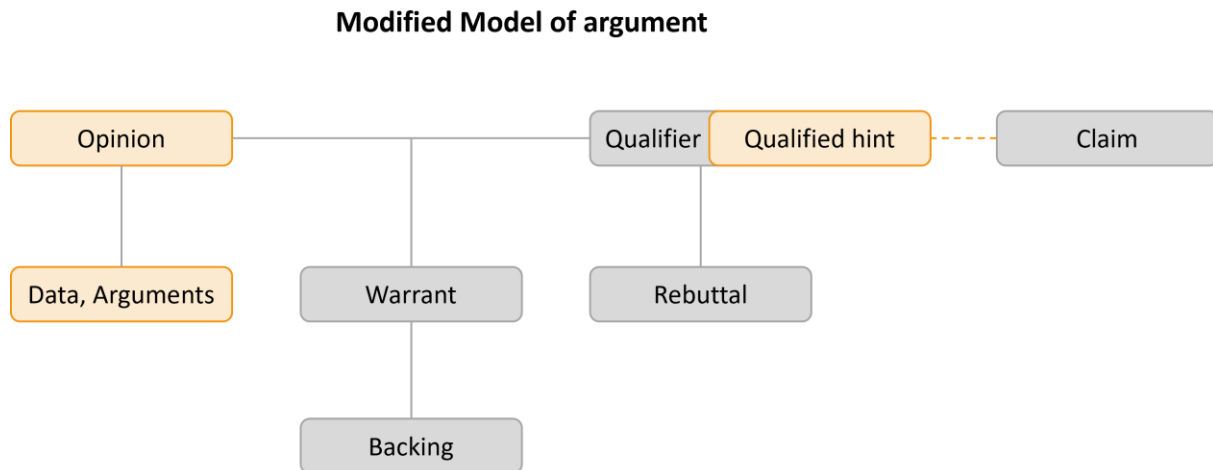


Abbildung 10 - Modifiziertes Model of Argument (Toulmin, 2003)

An die Stelle des *claim* im Modell nach Toulmin (2003) tritt nun ein *qualified hint*. Entsprechend basiert dieser auf *opinion*, aber ebenso auf den Begründungsmechanismen durch *warrant* und *backing* sowie die Ausschlusskriterien des *rebuttal*. Auf Grundlage des *qualified hint* kann dann ein *claim* gebildet werden. Dabei spielt hier insbesondere die Glaubwürdigkeit der Argumentation, *opinion* sowie des Expert*innenstatus hinein. Es stellt sich also die Frage, inwiefern auf die *opinion* vertraut werden kann und diese im Zuge der Argumentation standhält. Wie in Kapitel 2.2 beschrieben, kann die Argumentation dann als Evidenz genutzt werden.

Entsprechend stellen epistemologische Überzeugungen hier eine wichtige Grundlage zur Einschätzung der Glaubwürdigkeit dar. Insbesondere die Dimension *Source of knowledge* wird bereits durch die Proband*innen innerhalb der Interviews angesprochen (akademischer Grad). Jedoch kann auch die Bedeutung der anderen Dimension erkannt werden. So könnten Überzeugungen zur Dimension *Justification of knowledge* grundlegenden Einfluss auf die Verknüpfung der einzelnen Bestandteile des *Modified Model of argument* haben. Vor dem Hintergrund bestehender epistemologischer Überzeugungen formulieren die Proband*innen innerhalb der Interviews einen grundsätzlichen Unterschied zwischen *opinion* und Evidenz (vgl. Mechanismus *Agree*). Dabei spielen die formulierten Kriterien für Evidenzen eine Rolle, welche von Meinungen nicht unbedingt eingehalten werden. Eine besondere Bedeutung scheint die Replizierbarkeit, bzw. die systematische Sammlung von Beobachtungen als Grundlage für die formulierte Meinung zu spielen.

Innerhalb des *Modified Model of argument* werden Meinungen von Expert*innen deutlicher in evidenzbasierte Praxis einbezogen. Hier könnte auch eine Vermittlung der Rolle eigener Erfahrungen und wissenschaftlicher Evidenz erkannt werden. Bereits im Feld evidenzbasierter Medizin wurde die Entmündigung von Praktiker*innen (Dawes et al., 2005) durch den reinen Fokus auf Evidenzen bemängelt (vgl. Kap. 2.3). Ein klarer Einbezug eigener Erfahrungen bzw. der Meinungen von Praktiker*innen im jeweiligen Feld könnte sich hier im *Modified Model of argument* niederschlagen. Dabei kann dennoch eine evidenzbasierte Perspektive erkannt werden, da Evidenzen und Prozeduren zur Erstellung dieser (im Sinne von systematischer Datensammlung, etc.) berücksichtigt werden. Ebenso werden Evidenzen in der Aussagekraft noch über Meinungen und Hinweisen eingeordnet. Dennoch wird die Konkretheit von Hinweisen von Expert*innen im Vergleich zu wissenschaftlichen Evidenzen als Vorteil formuliert (vgl. Kap. 6.6.3.5). Innerhalb des Modells kann der Aushandlungsprozess zwischen Evidenzen und praktischen Erfahrungen erkannt werden.

Zusammenfassend kann in Bezug auf Forschungsfrage II.a festgehalten werden, dass eine Wirksamkeit der Intervention auf die Kompetenzen in evidenzbasierter Praxis (EIDM-ET), die *beliefs* gegenüber evidenzbasierter Praxis allgemein wie auch in der Schule (*Beliefs_EBP_Bildung, Beliefs_EBP_Impl_Schule*) sowie die tatsächliche Implementation evidenzbasierter Tätigkeiten in die Schulpraxis (*EBP implementation scale*) aufgezeigt werden kann. Im Hinblick auf die Stabilität von EBP-Kompetenzen kann nicht abschließend geklärt werden, ob diese durch die Implementation evidenzbasierter Tätigkeiten beeinflusst wird, oder umgekehrt. Insgesamt kann aber ein graduelles Evidenz-Verständnis beschrieben werden, welches Erfahrungen und Meinungen prominenter mit in die Argumentation einbezieht, als dies im *Model of argument* nach Toulmin (2003) der Fall ist.

6.7 Zusammenfassende Diskussion des ersten empirischen Teils

Innerhalb des ersten empirischen Teils wurden zwei Studien berichtet, welche die Kompetenzen in evidenzbasierter Praxis sowie die Einstellung gegenüber dieser bei angehenden Lehrkräften fokussiert haben. Dabei wurde einerseits der Einfluss der studierten Unterrichtsfächer sowie des Praxissemesters als zentrale Lerngelegenheit innerhalb des Studiums fokussiert, andererseits eine Intervention zur fachspezifischen Förderung im Fach Chemie analysiert.

Übergreifend kann festgehalten werden, dass die Proband*innen in beiden Studien Bedarfe für Fördermaßnahmen der EBP-Kompetenz aufzeigen. Die erreichten Punktzahlen innerhalb der Versionen des EIDM-ET zeigen eher mittelmäßige Kompetenzausprägungen auf. Dabei scheint die explizite Förderung im Rahmen fachspezifischer Interventionen vor dem Hintergrund der berichteten Ergebnisse eher dazu geeignet zu sein als das Absolvieren des Praxissemesters. Vor dem Hintergrund der Erkenntnisse aus den Interviews der Interventionsstudie könnte vor allem die Explikation als entscheidendes Merkmal anerkannt werden. Die Proband*innen formulieren einen Bedarf an expliziter Thematisierung evidenzbasierter Praxis bereits für das Studium. Innerhalb der Interventionsstudie kann diese dann erkannt werden. Im Rahmen des Praxissemesters hingegen tritt eine Thematisierung evidenzbasierter Praxis lediglich im Rahmen des Begleitforschungsseminars auf. Dabei muss hier zusätzlich die individuelle Gestaltung dieses innerhalb der unterschiedlichen Fächer berücksichtigt werden. Hier könnte der formulierte Bedarf nach der Schaffung von expliziten Lerngelegenheit innerhalb des Praxissemester erneut aufgegriffen werden (vgl. Kap. 6.6.2). Eine Explikation evidenzbasierter Praxis könnte somit übergreifend als ein Gelingensfaktor für die Förderung von EBP-Kompetenzen beschrieben werden.

Weiterhin kann ein Einfluss der fachspezifischen Perspektiven auf evidenzbasierte Praxis diskutiert werden. Innerhalb des fächerübergreifenden Erhebung konnten keine Unterschiede zwischen unterschiedlichen Fächerbündelkombinationen erkannt werden. Diese Ergebnisse stehen im Kontrast zu der Untersuchung domänenspezifischer epistemologischer Überzeugungen von Urhahne und Kremer (2023). Dabei wurde in Kap. 6.6.2 bereits auf eine mögliche Begründung dieser Unterschiede durch das gewählte Studiendesign eingegangen. Dennoch erscheinen die nicht auftretenden Unterschiede hier eher den Annahmen zu widersprechen. So hätte vermutet werden können, dass evidenzbasierte Praxis innerhalb experimentell geprägter Fächer einen anderen Stellenwert aufweist als in nicht-experimentellen Unterrichtsfächern. Diese Vermutung muss auf Grundlage der Ergebnisse verworfen werden.

Im Rahmen der Interventionsstudie wurde hingegen ein fachspezifischer Zugang zu evidenzbasierter Praxis gewählt, welcher sich in einer Steigerung der EBP-Kompetenzen der Proband*innen deutlich macht. Unterschiede in der Punktzahl für Item 1, könnten hier als weitere Hinweise für

den Vorteil eines fachspezifischen Zugangs gewertet werden. So erreichen die Proband*innen eine höhere Punktzahl für die fachspezifische Situationsbeschreibung als für die eher allgemeine, bildungswissenschaftliche. Diese Vermutung kann durch die beobachteten Unterschiede der EBP-Kompetenz für Proband*innen innerhalb der religiös-philosophischen Begleitforschungsseminare unterstützt werden (vgl. Kap. 6.6.1.1). Somit könnten hier domänenspezifische epistemologische Überzeugungen als leitend für die Wahl eines Begleitforschungsseminars vermutet werden, welche sich dann auch in den EBP-Kompetenzen widerspiegeln könnten.

Innerhalb des Lehramts muss bei dieser Diskussion jedoch die Interdisziplinarität der Studierenden berücksichtigt werden (Rochnia et al., 2020). So können fachspezifische Zugänge immer nur einzelne Unterrichtsfächer in den Blick nehmen, wenngleich die Studierenden durch ihr Studium unterschiedlicher Unterrichtsfächer mehrere verschiedene fachspezifische Perspektiven auf evidenzbasierte Praxis entwickeln könnten. Dabei könnten domänenspezifische epistemologische Überzeugungen für jedes Unterrichtsfach vermutet werden, welche z.B. zur Wahl eines spezifischen Begleitforschungsseminars innerhalb der fächerübergreifenden Befragung führen könnte. Neben den domänenspezifischen epistemologischen Überzeugungen könnten akademische epistemologische Überzeugungen allgemeinen evidenzbasierten Tätigkeiten zur Grunde liegen, welche nicht innerhalb eines spezifischen Unterrichtsfachs kontextualisiert sind (Muis et al., 2006). Dabei könnten sich akademische und domänenspezifische epistemologische Überzeugungen gegenseitig beeinflussen. Entsprechend könnte die Förderung domänenspezifischer epistemologischer Überzeugungen entsprechend dem Zugang der Interventionsstudie sich auch auf akademische epistemologische Überzeugungen auswirken. So könnten fachspezifische Kontexte als Anwendungsbeispiele evidenzbasierter Praxis genutzt werden und auch akademische Überzeugungen beeinflussen. Indizien für diese Argumentation könnten in der gesteigerten Punktzahl des EIDM-ET zum Post-Zeitpunkt der Interventionsstudie erkannt werden. Wie beschrieben, findet nicht für alle Items innerhalb des EIDM-ET eine rein chemiespezifische Kontextualisierung statt. So könnten durch die Intervention auch akademische epistemologische Überzeugungen gefördert worden sein, welche zu einem besseren Abschneiden in (auch) allgemeinen Items des EIDM-ET geführt haben könnten.

Ein Vorteil dieses Zugangs könnte in der Explikation erkannt werden, welche innerhalb spezifischer Kontexte ggf. deutlicher erfolgen könnte. Weitere Indizien für diesen Zusammenhang können in den Reflexionsprozessen der Proband*innen innerhalb der Interviews der Interventionsstudie erkannt werden. So konnte die Reflexion über das Studium des Fachs Chemie weitere Anwendungssituationen evidenzbasierter Praxis aufzeigen. Dabei konnten diese teilweise erst im Rückblick auf das eigene Studium erkannt werden. Die gleichzeitige Nennung von Beispielen aus anderen Unterrichtsfächern (z.B. Deutsch, Pädagogik, Ernährungslehre) deutet auf eine Aktivierung übergreifender Konstrukte mit Anwendung in anderen Unterrichtsfächern in diesem Zusammenhang hin. Damit könnte sich die explizite Thematisierung evidenzbasierter Praxis innerhalb eines Unterrichtsfachs auch auf allgemeinere Kontexte auswirken. In Verbindung mit der Explikation könnte also ein fachspezifischer Zugang zu evidenzbasierter Praxis als weiterer Gelingensfaktor beschrieben werden. In Bezug auf die beschriebene Relativität innerhalb des Konstruktes Epistemologie, könnte ein fachspezifischer Kontext auch hier klare Rahmenbedingungen vorgeben, welche eine Zunahme der Relativität und damit Abstraktheit des Konstruktes verhindern könnte.

Weiterhin tritt vor allem in den Interviews innerhalb der Interventionsstudie klar ein gradueller Charakter von Evidenzen und die damit verbundene große Bedeutung von Meinungen von Praktiker*innen hervor. Entsprechend werden Evidenzen und Meinungen zwar ontologisch von den Proband*innen getrennt, jedoch können Meinungen durch Einhaltung von Kriterien für Evidenzen

einen graduellen Charakter von Evidenzen aufweisen. In Anlehnung an die Entwicklungen im Feld evidenzbasierter Medizin scheint ein klarer Einbezug von Erfahrungen der Proband*innen sowie von Expert*innen bedeutsam zu sein. So könnte einerseits einer Ablehnung evidenzbasierter Praxis als ‚bevormundend‘ begegnet werden. Andererseits könnten so relevante kontextuelle Einflüsse innerhalb der Prozesse berücksichtigt werden. Vor allem für Lehrkräfte am Beginn ihrer Berufslaufbahn könnte die beschriebene Konkretheit von Erfahrungsberichten hilfreich sein. Der Abgleich von Erfahrungen mit wissenschaftlichen Evidenzen könnte so bestehende Subjektivität von Meinungen klar herausstellen. So wird innerhalb des *Modified Model of argument* (an Anlehnung an Toulmin, 2003) die Bedeutung von Meinungen für Begründungsprozesse deutlich, obwohl zeitgleich eine Unterscheidung von Meinungen und Evidenzen dargelegt wird. So könnten über die unterschiedlichen Begründungszusammenhänge innerhalb des Modells ein Einbezug wissenschaftlicher Evidenzen erfolgen. Dabei steht hier vor allem das Zusammenspiel von Evidenzen und Meinungen im Vordergrund. In diesem Zusammenhang könnten epistemologische Überzeugungen zu Autoritäten, etc. aufgegriffen werden. Gleichzeitig könnte der deutliche Einbezug von Erfahrungen die Akzeptanz evidenzbasierter Praxis stärken, wie dies im Feld evidenzbasierter Medizin zu erkennen ist.

Mit Blick auf die Ergebnisse des (gekürzten) EIDM-ET könnte nicht vorhandenes forschungsmethodologisches Wissen den forschungsbezogenen Einsatz von Evidenzen erschweren. So könnte hier die Nutzung von Evidenzen im Sinne des *establishing evidence* (Davies, 1999) erschwert werden. In Anknüpfung an den graduellen Charakter von Evidenzen nutzen die Proband*innen Erfahrungen und Meinungen von Expert*innen zur Planung von Unterricht. Hier ist entsprechend eher der Ansatz des *using evidence* (Davies, 1999) zu erkennen. Wie beschrieben können die beiden Ansätze dabei als eine Lehrkräfte-Perspektive auf Evidenzen zur Unterrichtsplanung und eine Forschenden-Perspektive auf Evidenzen im Sinne der Durchführung von Studien beschrieben werden. Innerhalb der Interviews erfolgt hier eine teilweise Verortung der beiden Perspektiven innerhalb des Studiums. So wird fachchemischen Veranstaltungen eher eine *establishing evidence* Perspektive zugedacht, während im Bereich der Chemiedidaktik eher auf ein *using evidence* zurückgegriffen wird. Während in beiden Bereichen Meinungen von Expert*innen einbezogen werden, formulieren die Proband*innen im Bereich der Unterrichtsdurchführung eine Neigung, Dinge selbst zu erproben. Entsprechend scheinen *qualified hints* hier eine größere Rolle zu spielen als innerhalb fachchemischer Forschung. Dabei könnte hier eine abnehmende Bedeutung von Evidenzen im Sinne der Unterrichtsplanung erkannt werden.

Im Zusammenhang dieser unterschiedlichen Perspektiven auf den Umgang mit Evidenzen scheint der von Bauer und Kollar (2023) beschriebene Kulturwechsel notwendig zu sein. Dabei müssen die beiden Perspektiven innerhalb der Profession der Lehrkräfte verortet werden, sodass beide Zugänge zu Evidenzen in die Tätigkeiten mit aufgenommen werden. Dabei scheint vor allem die Perspektive von *establishing evidence* eher unterrepräsentiert zu sein.

Insgesamt kann der Bedarf einer Förderung angehender Lehrkräfte in deren EBP-Kompetenzen über alle Unterrichtsfächer hinweg beschrieben werden. Dabei könnte ein Fokus auf die Explikation evidenzbasierter Praxis innerhalb eines Unterrichtsfaches einen Einfluss auf generelle, akademische epistemologische sowie weitere fachspezifische epistemologische Überzeugungen nehmen. Die Basierung einer solchen Fördermaßnahme auf einem graduellen Evidenzverständnis im Sinne des Einbezugs von Erfahrungen in den Argumentationsprozess, könnte zu einer höheren Akzeptanz bei den Proband*innen führen. Gleichzeitig solle die Bedeutung der beiden Perspektiven *using evidence* und *establishing evidence* deutlich hervorgehoben werden. Dabei sollten diese Perspektiven nicht einzelnen Bestandteilen des Studiums zugeordnet werden.

7. ZWEITER EMPIRISCHER TEIL - UMGANG MIT ANOMALEN BEOBACHTUNGEN ANGEHENDER LEHRKRÄFTE DES NATURWISSENSCHAFTLICHEN SACHUNTERRICHTS

Diese Teilstudie fokussiert auf die Beobachtung und Deutung von naturwissenschaftlichen Experimenten, die anomale Daten in Form von anomalen Beobachtungen hervorrufen können. Der Umgang mit anomalen Beobachtungen kann hier als Umgang mit widersprechenden Evidenzen betrachtet werden. Dabei stellen anomale Daten widerlegende Evidenzen im Sinne von Bauer und Kollar (2023) dar. Dabei soll einerseits der Umgang mit anomalen Daten an sich, andererseits die Verknüpfung unterschiedlicher Beobachtungs- und Konzeptaspekte betrachtet werden. Durch den Einsatz von sog. *Model-of-data* (MOD) (Chinn & Brewer, 2001) sollen dabei auch die epistemologischen Verknüpfungen zwischen einzelnen Aspekten durch die Proband*innen expliziert werden. So sollen Einblicke in den Umgang mit widerlegenden Evidenzen auf Grundlage derer epistemologischer Verknüpfung mit bestehenden Wissensaspekten erreicht werden. Die übergreifende Forschungsfrage lautet:

III.a Wie reagieren Studierende auf als anomal wahrgenommene Beobachtungen innerhalb eines Experiments?

Die eingesetzten MOD sollen dazu dienen, die mentalen Modelle mit ihren konzeptuellen und epistemologischen Ausprägungen zu externalisieren. Da die Externalisierung von mentalen Modellen generell als eher herausfordernd beschrieben werden kann, stellt sich die weiterführende Frage, inwiefern MOD methodisch dazu geeignet sind mentale Modelle zu externalisieren:

III.b Inwiefern können Studierende ihre im mentalen Modell ablaufende Reaktion auf anomale Beobachtungen in einem Model-of-data darstellen?

Innerhalb dieses zweiten empirischen Teils soll die zur Beantwortung der Forschungsfragen angelegte Studie vorgestellt sowie die Ergebnisse diskutiert werden.

7.1 Studiendesign

Zur Beantwortung der Forschungsfragen wurde eine zweiteilige, digitale Interviewstudie geplant (Abb. 11). Das Interview wurde in eine Trainings- und eine Modellierungsphase eingeteilt. Die Trainingsphase zielte auf das Einüben der Methode der MOD ab, während innerhalb der Modellierungsphase mit Hilfe des lauten Denkens (Sandmann, 2014) das externalisierte mentale Modell validiert werden soll. Innerhalb der Trainingsphase wurden zu Beginn relevante demographische Daten erhoben. Weiterhin wurde ein Präsentationsprogramm, das für die Erstellung der MOD nötig ist, installiert. Anschließend wird die Methode des lauten Denkens eingeführt. Dazu wird im ersten Schritt die Methode beschrieben und auf das Äußern aller Gedanken während der Bearbeitung hingewiesen. Anschließend wurden zwei Übungen durchgeführt. In einem ersten Schritt sollen die Proband*innen beim Multiplizieren von zwei zweistelligen Zahlen laut denken (vgl. Sandmann, 2014). Um eine Trainingssituation zu schaffen, welche verglichen mit der anschließenden Modellierungsphase möglichst authentisch ist, sollen die Proband*innen im zweiten Schritt ihre Gedanken beim Beobachten eines kurzen Videos formulieren. Das Video⁹ zeigt dabei kein chemisches Phänomen, sondern einen Hund, welcher hinter einem Ball herläuft. Zwar stellt

⁹ Video abrufbar unter: <https://www.youtube.com/watch?v=kSEn8ddukAE> © Pascal Pollmeier

dies immer noch einen Unterschied zu chemischen Experimenten aus der Modellierungsphase dar, jedoch können hier bereits Beobachtungen aus einem Video beschrieben werden. Damit sollen Gedanken zu einem beobachteten Prozess geäußert werden, wie es auch in der Modellierungsphase der Fall ist. Die Proband*innen erhalten im Rahmen der Übung die Möglichkeit das Video öfter anzusehen. Das betrachtete Video stellt gleichzeitig den Ausgangspunkt zur Erklärung der Methode MOD (vgl. Kap. 7.1.2) dar.

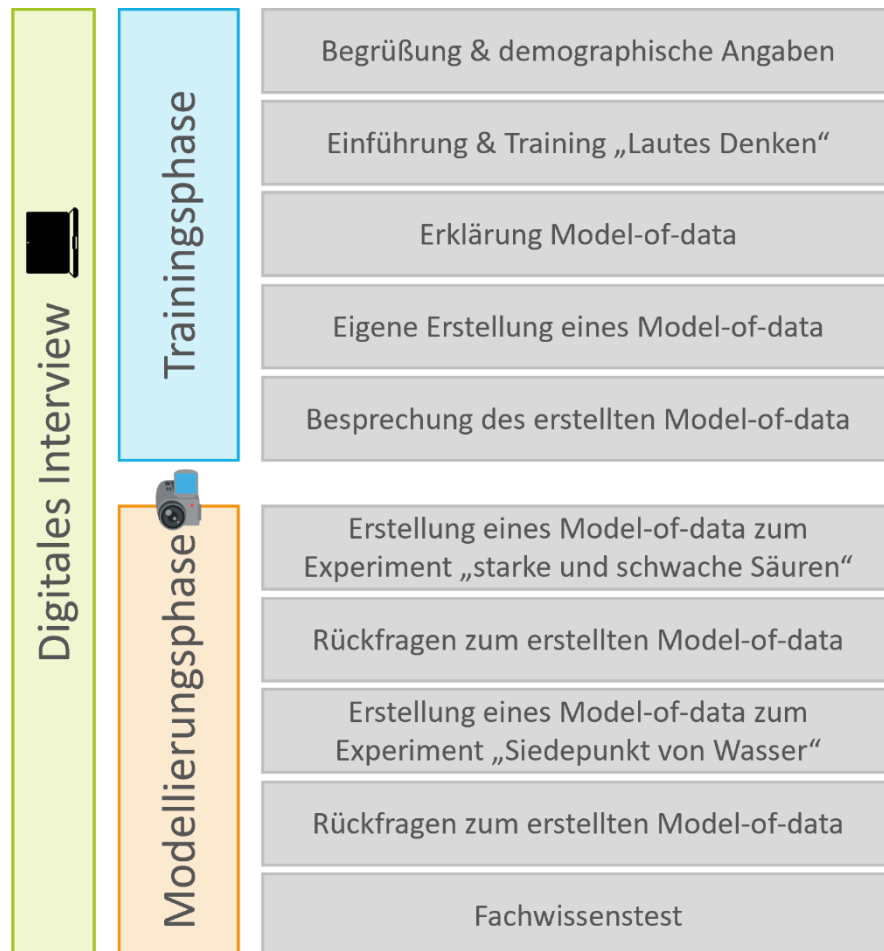


Abbildung 11 - Studiendesign zu Teilstudie 1, genutzte Materialien: Grafik „Laptop clipart“, JoyPixels, creazilla.com (<https://creazilla.com/de/nodes/45348-laptop-clipart>), CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.de>); Grafik „Videokamera clipart“, JoyPixels, creazilla.com (<https://creazilla.com/de/nodes/47947-videokamera-clipart>), CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.de>)

Nachdem Gemeinsamkeiten und Unterschiede zwischen MOD, *Mind-Map* und *Concept Maps* vorgestellt wurden, werden die unterschiedlichen Konstituenten von MOD anhand von konkreten Beispielen in Bezug zum beobachteten Video dargestellt. Ein Vergleich zu *Mind-Maps* und *Concept Maps* wird angestellt, da diese Methoden aus der eigenen Schule erfahren ggf. schon bekannt sind und einige Ähnlichkeiten bestehen. So soll ein leichter Übergang zu den Spezifika von MOD erreicht werden. Zum besseren Verständnis von MOD und deren Erstellung findet anschließend die eigene Konstruktion eines MOD statt. Dabei gehört diese Aufgabe explizit in die Trainingsphase, sodass der Interviewer ggf. eingreifen kann, oder für Fragen zur Verfügung steht.

Als Grundlage des zu bildenden MOD wird ein chemisches Experiment (Löslichkeit von Natriumchlorid in Wasser und Rapsöl¹⁰) durch ein Video beobachtet. Die Proband*innen können das Video

¹⁰ Video abrufbar unter: <https://www.youtube.com/watch?v=NYCWLqvpU9w> © Pascal Pollmeier

dabei jederzeit pausieren, darin vor- oder zurückspulen und das Video beliebig oft ansehen. Mithilfe einer Vorlagendatei im Format eines von drei Präsentationsprogrammen (siehe Anhang) sollen die Proband*innen ein MOD zu dem Experiment erstellen. Dabei stehen Definitionen und Beispiele der unterschiedlichen Konstituenten von MOD jederzeit zur Verfügung. Ebenso können Notizen in einer Gedankensammlung gesammelt werden, welche nicht direkter Bestandteil des MOD ist. Sobald die Proband*innen angeben mit der Erstellung des MOD fertig zu sein, werden offene Fragen besprochen und es findet ein Abgleich mit einer Expert*innen-Lösung (siehe Anhang) statt. In diesem Zusammenhang gibt es erneut die Möglichkeit methodische Fragen bezüglich MOD zu klären. Mit dieser Besprechung kommt die Trainingsphase zum Ende.

Es folgt die Modellierungsphase, welche mithilfe eines Screencasts sowie Audioaufnahmen aufgezeichnet wird. Dabei wird der Bildschirm der Proband*innen gefilmt, um die Konstruktion der MOD nachvollziehen zu können. Innerhalb der Studienphase sollen die Proband*innen je ein MOD zu den Experimenten „Reaktion von Säuren mit Natronlauge“¹¹ und „Siedepunkt von Wasser“¹² erstellen. In dieser Phase können keine Rückfragen an den Interviewer mehr gestellt werden. Der Interviewer greift lediglich bei technischen Problemen oder ausbleibendem lautem Denken ein. Direkt im Anschluss an die Erstellung eines MOD schließt sich ein kurzes Interview an. In diesem Interview wird die Methode des *Member Checkings* eingesetzt, um Unklarheiten in der Darstellung mit den Proband*innen zu klären (Döring & Bortz, 2016). Sind beide Model-of-data erstellt und alle Rückfragen geklärt, schließt sich ein Fachwissenstest (vgl. Kap. 7.4.1) an.

Um im Folgenden einen genaueren Einblick in das Studiendesign sowie die genutzte Lernumgebung zu erhalten, sollen zuerst die genutzten Experimente sowie die Methode der *Model-of-data* vorgestellt werden.

7.1.1 Experimente mit potentiell anomalen Beobachtungen

In der vorliegenden Studie wurden insgesamt drei Experimente genutzt, welche potentiell anomale Beobachtungen hervorrufen können. Dabei kann in diesem Zusammenhang nur von erwarteten anomalen Beobachtungen gesprochen werden, da die individuelle Wahrnehmung von Beobachtungen als anomal sehr unterschiedlich ausfallen kann (Chinn & Brewer, 1993a, 1993b, 1998). Ob eine Beobachtung einen kognitiven Konflikt auslöst oder nicht hängt daher eng mit den bestehenden Lernendenvorstellungen zusammen. In jedem der ausgewählten Experimente können gegensätzliche Beobachtungen getätigt werden. So könnten anomale Beobachtungen aus den einzelnen Experimenten abgeleitet werden.

Dies hängt entscheidend vom konzeptuellen Verständnis der Proband*innen ab. Die erwarteten Beobachtungen können so nur dann als anomal wahrgenommen werden, wenn diese bestehenden Vorstellungen widersprechen bzw. von diesen abweichen. Der Abgleich zwischen den anomalen Beobachtungen und den vorhandenen Vorstellungen kann dann durch die unterschiedlichen Umgangsarten nach Chinn und Brewer (1993b, 1998) beschrieben werden. Dabei können Charakteristika der anomalen Beobachtungen Einfluss auf deren Verarbeitung und den Einfluss auf bestehende Vorstellungen nehmen (Chinn & Brewer, 1993a, 1993b).

Im Folgenden sollen die einzelnen Experimente inkl. der erwarteten Beobachtungen vorgestellt werden. Gleichzeitig sollen bekannte Lernendenvorstellungen erläutert werden, welche in Verbindung mit den anomalen Beobachtungen zu kognitiven Konflikten führen können.

¹¹ Video abrufbar unter: <https://youtu.be/bXYhMJufqp4> © Pascal Pollmeier

¹² Video abrufbar unter: <https://youtu.be/Ig8QwMq9ERw> © Pascal Pollmeier

Löslichkeit von Salz in Wasser und Rapsöl

Dieses Experiment wurde im Rahmen des Modellierungstrainings innerhalb der Trainingssphase genutzt und untersucht die Löslichkeit von Salz in Wasser und Rapsöl. Das Experiment wurde ausgewählt, da eindeutige und gegensätzliche Beobachtungen getätigt werden können. Ebenso stellt die Löslichkeit als Stoffeigenschaft einen Inhalt aus dem Eingangsunterricht im Fach Chemie (MSB NRW, 2019) und damit ein vergleichsweise weniger komplexes Inhaltsfeld dar.

In diesem Experiment wird in zwei Reagenzgläsern je entweder Wasser oder Rapsöl vorgelegt. Anschließend wird ein Spatel Natriumchlorid hinzugegeben und geschüttelt. Die Proband*innen können dabei sowohl die Zugabe des Natriumchlorids als auch die Reagenzgläser nach dem Schütteln in Nahaufnahme beobachten. Anschließend werden beide Reagenzgläser simultan vor schwarzen Karton gehalten und gedreht. Der höhere Kontrast zwischen schwarzem Untergrund und potentiell ungelöstem Salz soll so besser zur Geltung kommen. Eine abschließende Ansicht von unten in beide Reagenzgläser lässt einen klaren Bodensatz im Reagenzglas mit Rapsöl erkennen, während keinerlei Phasenunterschiede im Reagenzglas mit Wasser beobachtet werden können.

Die Löslichkeit eines Stoffes hängt von der Polarität des Stoffes und der des Lösemittels ab. Natriumchlorid stellt aufgrund des Unterschiedes der Elektronegativitäten von Natrium (0.9) und Chlor (3.0) (Riedel & Meyer, 2013) sowie der zugrundeliegenden ionischen Bindung einen polaren Stoff dar. Dieser löst sich im polaren Lösemittel Wasser aufgrund von elektrostatischen Wechselwirkungen gut. Die Anionen und Kationen des NaCl werden durch Wassermoleküle hydratisiert (Riedel & Meyer, 2013). Öl hingegen stellt ein unpolares Lösemittel dar und kann somit keine elektrostatischen Wechselwirkungen gegenüber dem NaCl aufbauen. Somit kann sich Salz nicht in Öl lösen.

Die Beobachtung des unterschiedlichen Lösungsverhaltens von Salz in Wasser und Rapsöl könnte bei den Proband*innen einen kognitiven Konflikt auslösen. Während sich Natriumchlorid in Wasser löst, bleibt im Rapsöl ein Bodensatz zurück. Neben diesem erwarteten kognitiven Konflikt sind weitere Konflikte denkbar. In Anschluss an Schülervorstellungen zum Vernichtungskonzept könnten auch im Lösen des Salzes im Wasser anomale Beobachtungen erkannt werden, welche zu einem kognitiven Konflikt führen könnten (Barke, 2006). Das augenscheinliche Verschwinden des Salzes könnte hier als anomal wahrgenommen werden. Ebenso sind Kombinationen der beschriebenen Konzepte möglich. So könnten die Proband*innen annehmen, dass sich im Rapsöl weniger Salz löst als im Wasser. So könnte der entstehende Bodensatz erklärt werden, ohne dass ein Unterschied in der generellen Löslichkeit von Natriumchlorid zwischen den beiden Lösemitteln diskutiert werden müsste. Je nachdem, welcher kognitive Konflikt ausgelöst wird, könnten unterschiedliche Umgangsarten nach Chinn und Brewer (1993b, 1998) darauffolgen. Dies hängt in erster Linie von den bestehenden Vorstellungen der Proband*innen sowie der Bewertung der wahrgenommenen anomalen Beobachtungen ab.

Reaktion von Säuren mit Natronlauge

In diesem Experiment wird die Reaktion von Natronlauge mit drei unterschiedlichen Säuren untersucht. Die Auswahl dieses Experimentes kann einerseits durch die Relevanz des Themas Säure-Base für die Chemie begründet werden. Andererseits ermöglicht die Färbung des Indikators eine eindeutige Wahrnehmung des pH-Werts, der sich am Ende der drei Telexperimente unterscheidet. Das Experiment kann in drei Telexperimente unterteilt werden, welche analog durchgeführt werden. Zu Beginn werden 10ml einer der drei Säuren (Salzsäure, Salpetersäure & Essigsäure)

mittels Vollpipette in einem Reagenzglas vorgelegt. Die Säuren haben eine Konzentration von $c=1,0 \text{ mol/l}$, was im Video eingeblendet wird. Anschließend werden einige Tropfen Rotkohllindikator hinzugefügt. Das jeweilige Reagenzglas wird aus dem Reagenzglasständer genommen und geschüttelt. Nachdem das Reagenzglas zurück in den Ständer gestellt wird, wird der pH-Wert mithilfe einer eingeblendeten pH-Wert-Skala des Rotkohllindikators abgeschätzt. Die rot-pinke Färbung der Lösung deutet auf einen pH-Wert <5 hin. Im nächsten Schritt werden 10 ml Natronlauge ($c=1,0 \text{ mol/l}$) hinzugefügt. Das Reagenzglas wird dazu aus dem Ständer gehoben und die Natronlauge wird langsam hinzugegeben. Im Reagenzglas sind nun unterschiedliche Farbschichten erkennbar. Während im unteren Bereich eine gelbliche Färbung mit Übergang zu grün vorliegt, ist im oberen Bereich weiterhin eine rot-pinke Färbung zu erkennen. Das Reagenzglas wird erneut geschüttelt und anschließend mit der pH-Wert-Skala des Rotkohllindikators verglichen. Während für Salz- und Salpetersäure eine blaue Färbung vorliegt ($\text{pH}=7$), weist das Telexperiment mit Essigsäure eine grüne Färbung auf ($\text{pH}>7$). Zur besseren Vergleichbarkeit der drei Telexperimente werden am Ende die drei Reagenzgläser nebeneinander in den Reagenzglasständer gestellt und die pH-Wert-Skala erneut eingeblendet.

Die unterschiedliche Färbung der Lösungen am Ende der Telexperiment lässt sich durch die Säurestärke erklären. Während Salzsäure ($\text{pK}_s = -7$) sowie Salpetersäure ($\text{pK}_s = -1.37$) als starke Säuren bezeichnet werden können, handelt es sich bei Essigsäure ($\text{pK}_s = 4.75$) um eine schwache Säure (Riedel & Meyer, 2013). Das chemische Gleichgewicht der Protolysereaktion von Essigsäure liegt deutlich auf der Seite der Edukte. Somit enthält die Essigsäure-Lösung weniger Oxoniumionen als die Salzsäure- oder Salpetersäure-Lösung gleicher Konzentration. Bei der Zugabe der gleichen Menge Natronlauge, welche bei Salz- und Salpetersäure zur Neutralisation mit dem End-pH-Wert=7 führt, schlägt der pH-Wert im Telexperiment mit Essigsäure somit in den basischen Bereich um, da sich ein entsprechendes chemisches Gleichgewicht einstellt. Diese Verschiebung der Gleichgewichtslage kann nur bei schwachen Säuren beobachtet werden.

In diesem Experiment kann die unterschiedliche Färbung von starken und schwachen Säuren am Ende des Experiments beobachtet werden. Dabei könnten Vorstellungen bei den Proband*innen erwartet werden, welche von einer Neutralisation mit dem End-pH-Wert=7 aller Säuren am Ende des Experiments ausgegangen sind. Durch diese Beobachtungen kann ein kognitiver Konflikt vor dem Hintergrund der bestehenden Vorstellungen entstehen. Während bei den beiden erstgenannten Säuren ein neutraler pH-Wert von 7 (blaue Färbung des Rotkohllindikators) angezeigt wird, weist die Grünfärbung des Rotkohllindikators bei der Essigsäure auf einen basischen pH-Wert hin. Diese Beobachtungen lassen sich ohne Wissen über starke und schwache Säuren sowie die Kategorisierung der genutzten Säuren nicht erklären. Entsprechend könnte die grüne Färbung am Ende des Experiments mit Essigsäure als anomal wahrgenommen werden. Dabei kann im Bereich der Säure-Base-Chemie auf eine Reihe von untersuchten Lernendenvorstellungen zurückgegriffen werden (für einen Überblick: Kousathana et al., 2005). In Bezug auf das vorliegende Experiment kann vor allem die Vorstellung aufgegriffen werden, dass eine Neutralisationsreaktion immer zu einer neutralen Lösung führe. Hier könnten die Beobachtungen innerhalb des Experiments als Widersprüche gewertet werden. Ebenso könnte eine Verwechslung der Säurestärke mit der Konzentration der jeweiligen Lösungen entstehen. Durch die Angabe der Konzentrationen der eingesetzten sauren Lösungen könnte hier z.B. eine Verdünnung dieser durch die Zugabe von Natronlauge durch die Proband*innen erkannt werden, welche sich für die unterschiedlichen Lösungen unterschiedlich stark darstellt.

Siedepunkt von Wasser

Das zweite Experiment innerhalb der Modellierungsphase untersucht den Siedepunkt von Wasser. Dieses Experiment wurde aufgrund der verbreiteten Annahme Wasser siede immer bei genau 100°C genutzt. Im vorliegenden Experiment können abweichende Beobachtungen dazu getätigt werden.

Im Experiment wird ein Zweihals-Rundkolben mit Wasser gefüllt. In der zweiten Öffnung des Kolbens ist ein digitales Thermometer platziert, das die Temperatur des Wassers misst. Das Wasser im Kolben wird anschließend mittels Lebensmittelfarbe rot eingefärbt. Dies dient der besseren Sichtbarkeit im Video. Es wird ein Hinweis gegeben, dass die Einfärbung keinen Einfluss auf den weiteren Verlauf des Experiments haben wird. Das Wasser hat eine Ausgangstemperatur von 20,1°C. Im nächsten Schritt wird ein Bunsenbrenner unter dem Kolben entzündet und das Wasser erhitzt. Nun wird ein Becherglas sowie eine große Kristallisierschale mit Kühlwasser bereitgestellt. Nachdem das Wasser einen Moment deutlich gesiedet hat, wird der Gasbrenner ausgeschaltet und entfernt. Sobald kein Blasenwurf mehr zu erkennen ist, wird der Kolben mithilfe eines Stopfens luftdicht verschlossen. Sobald die Wassertemperatur verlässlich unter 100°C abgesunken ist, wird der Kolben in das Kühlwasser in der Kristallisierschale gehangen und mit dem Kühlwasser aus dem Becherglas übergossen.

Die gemessene Temperatur am Siedepunkt während des Erhitzens mit dem Gasbrenner beträgt ca. 101,8°C. Nachdem der Gasbrenner ausgeschaltet wurde, ist im Wasser im Kolben nach und nach weniger Blasenwurf erkennbar, bis dieser ganz zum Erliegen kommt. Die gemessene Temperatur liegt zu diesem Zeitpunkt bei ca. 101,2°C. Nachdem der Kolben in das ca. 20°C kalte Kühlwasser eingesetzt wird, sinkt die Temperatur im Kolben. Zwischen ca. 90-76°C ist nun ein erneutes Sieden des Wassers im Kolben zu beobachten.

Durch dieses Experiment kann die Abhängigkeit des Siedepunkts vom Umgebungsdruck beobachtet werden. Durch das luftdichte Verschießen des Kolbens nach dem ersten Sieden bei ca. 101,8°C wird ein geschlossenes System erzeugt. Die Gasphase des Systems ist hauptsächlich durch Wasserdampf gefüllt. Durch das Abkühlen des Kolbens sinkt der Dampfdruck des Wassers in der Gasphase, wodurch sich die Stoffmenge des Wasserdampfs in der Gasphase reduziert. Als Folge entsteht ein Unterdruck im geschlossenen System. Die aktive Kühlung des Kolbens durch das Kühlwasser beschleunigt diesen Vorgang. Im Zustandsdiagramm (Abb. 12) kann die Dampfdruckkurve erkannt werden, welche den Übergang von flüssigem zu gasförmigem Wasser darstellt. Auf der Dampfdruckkurve kann der Siedepunkt bei 100°C erkannt werden, sofern 1.013bar Umgebungsdruck anliegen. Durch die beschriebene Erzeugung des Unterdrucks im vorliegenden Experiment bei nahezu gleichbleibender Temperatur, wird die Dampfdruckkurve unterschritten und das Wasser beginnt erneut zu Sieden.

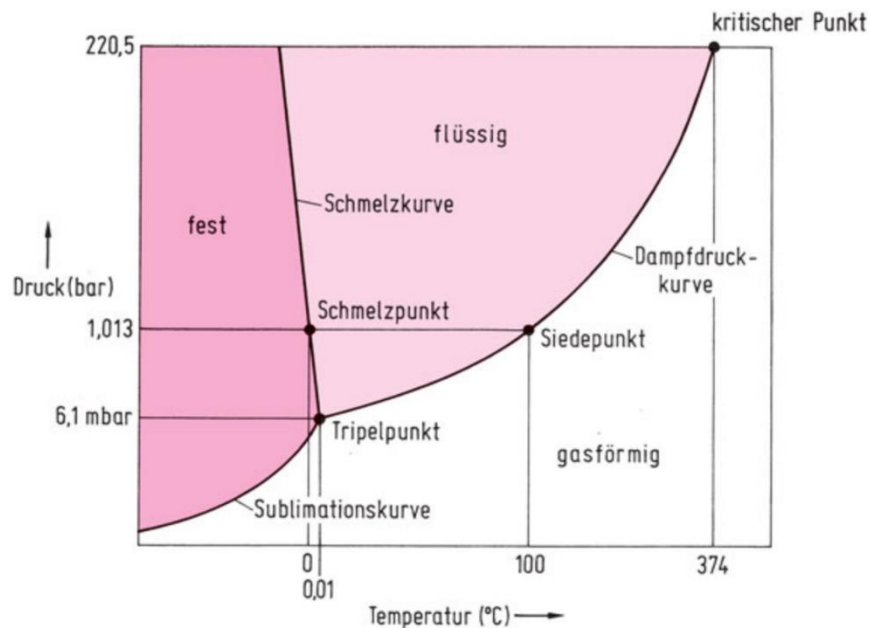


Abbildung 12 - Zustandsdiagramm von Wasser (Riedel & Meyer, 2013, S. 145)

Wie beschrieben kann davon ausgegangen werden, dass die Proband*innen die Siedetemperatur von Wasser bei 100°C erwarten. Vor diesem Hintergrund könnte das Sieden des Wassers für den ersten Siedepunkt bei 101,8°C als anomal wahrgenommen werden. Da der Prozess des Siedens jedoch nach und nach deutlicher ausfällt (Zunahme des Blasenwurfs) und das Wasser nicht erst bei exakt 100°C beginnt zu Sieden, könnte diese Beobachtung als weniger anomal aufgefasst werden. Gleichzeitig könnte dieser höhere Siedepunkt der Vorstellung widersprechen, dass Stoffe keine höheren Temperaturen als an deren Siedepunkt erreichen könnten. Das Sieden bei unter 100°C während der Abkühlung im Experiment kann ebenfalls als anomale Beobachtung beschrieben werden.

Bestehende Untersuchungen zu Schülervorstellungen zum Sieden von Wasser beschäftigen sich zumeist mit der Beschreibung der Gasphase (Barke, 2006; Paschmann et al., 2002). Auf die Frage, woraus die aufsteigenden Gasblasen beim Sieden bestehen, beschreiben die Proband*innen bei Paschmann et al. (2002) unter anderem, dass es sich um Luft oder dessen Bestandteile handle. Andere Proband*innen vermuteten Lufteinschlüsse im Boden des Gefäßes oder eine stattfindende Elektrolyse zu Sauerstoff und Wasserstoff als Ursache der Blasenbildung. Barke (2006) beschreibt ebenfalls die Vorstellung, dass die Teilchenbewegung nicht mit steigender Temperatur zunehme, sondern die Wasserteilchen am Siedepunkt von 100°C spontan auseinanderfliegen. Dennoch wird in den Untersuchungen häufig der Siedepunkt bei 100°C von Lernenden erwartet (Paschmann et al., 2002; Yeo & Zadnik, 2001). Yeo und Zadnik (2001) beschreiben weiterhin Vorstellungen bzgl. der Erhitzung von Stoffen. So könnten Vorstellungen vorliegen, welche die Temperatur am Siedepunkt als die höchste, erreichbare Temperatur eines Stoffes verstehen. Andererseits bestehen Vorstellungen dazu, dass das Erhitzen eines Stoffes zu immer weiterer Temperaturerhöhung führt.

Neben den intendierten kognitiven Konflikten könnten durch die Proband*innen weitere anomale Beobachtungen getätigt werden, welche andere Konflikte auslösen könnten. Wie beschrieben hängt dies von den bestehenden Lernendenvorstellungen im spezifischen Inhaltsbereich ab.

7.1.2 Model-of-data

Die *Model-of-data* (MOD) erfüllen im Rahmen dieser Studie eine doppelte Funktion. Primär dienen diese der Modellierung des mentalen Modells. Mithilfe der MOD sollen die Lernenden ihr mentales Modell externalisieren. Dabei sollen MOD die Externalisierung strukturieren und bei der Verknüpfung und Begründung von Aspekten unterstützen. In der Studie von Chinn und Brewer (2001) wurden MOD nicht in ihrer graphischen Umsetzung zur Externalisierung mentaler Modelle genutzt. Die Proband*innen wurden mit unterschiedlichen Datensätzen, inkl. Erläuterungen konfrontiert und sollten schriftlich argumentieren, warum sie welchen Argumentationen vertrauen. Anschließend wurden die Argumentationslinien aus den schriftlichen Begründungen der Proband*innen vor dem Hintergrund der Rahmentheorie der MOD analysiert. Damit weicht die Umsetzung der Methode in dieser Studie von dem beschriebenen Vorgehen ab. Hier sollen die Proband*innen selbst ein graphisches MOD erstellen und ihr mentales Modell dadurch externalisieren.

Die von Chinn und Brewer (2001) entwickelten MOD zeigen konzeptuelle Aspekte und deren Beziehungen untereinander auf. Insofern gibt es eine Ähnlichkeit zu Concept Maps (Novak & Gowin, 1984), welche z.B. zum formativen Assessment konzeptueller Wissensaspekte genutzt werden können (z.B. Hartmeyer et al., 2018). In Unterscheidung zu Concept Maps können die Verknüpfungen zwischen einzelnen Aspekten innerhalb der MOD durch fünf unterschiedliche Verknüpfungsarten dargestellt werden. Bei den Verknüpfungen innerhalb der MOD handelt es sich um inferentielle Verknüpfungen, welche gleichzeitig eine epistemologische und ontologische Relationierung darstellen (Chinn & Brewer, 2001). Die Verknüpfungen stellen also einen folgernden Zusammenhang zwischen den Aspekten dar. Um die unterschiedlichen Verknüpfungsarten in ihrer unterschiedlichen Natur darzustellen, erfolgt die genauere Erläuterung am Beispiel eines MOD. Abbildung 13 zeigt ein MOD für das Experiment „Löslichkeit von Salz in Wasser und Öl“. Das Experiment wurde im Rahmen der Trainingsphase dieser Studie genutzt und wurde in Kap. 6.5.2 genauer beschrieben. Die orangen Nummerierungen sind nicht Bestandteil des MOD (sie geben insbesondere keine Reihenfolge/Hierarchie an). Diese dienen lediglich zur einfacheren Beschreibung.

Das vorliegende MOD stellt eine Expert*innen-Lösung dar. Dabei wurde ein Ausschnitt aus möglichen Aspekten und Verknüpfungen gewählt. MOD können in ihrem Umfang variieren, je nachdem wo Grenzen bzgl. relevanter Konzepte gezogen werden. Ebenso folgt das dargestellte MOD in Abb. 13 der Durchführung des Experiments. So wurde mit Aspekt 6 eine Verallgemeinerung getätigt, welche in umgekehrter Weise auch für die Aspekte 8, 9 und 11 möglich wäre. Ebenso dient das MOD dazu alle möglichen Verknüpfungsformen deutlich zu machen.

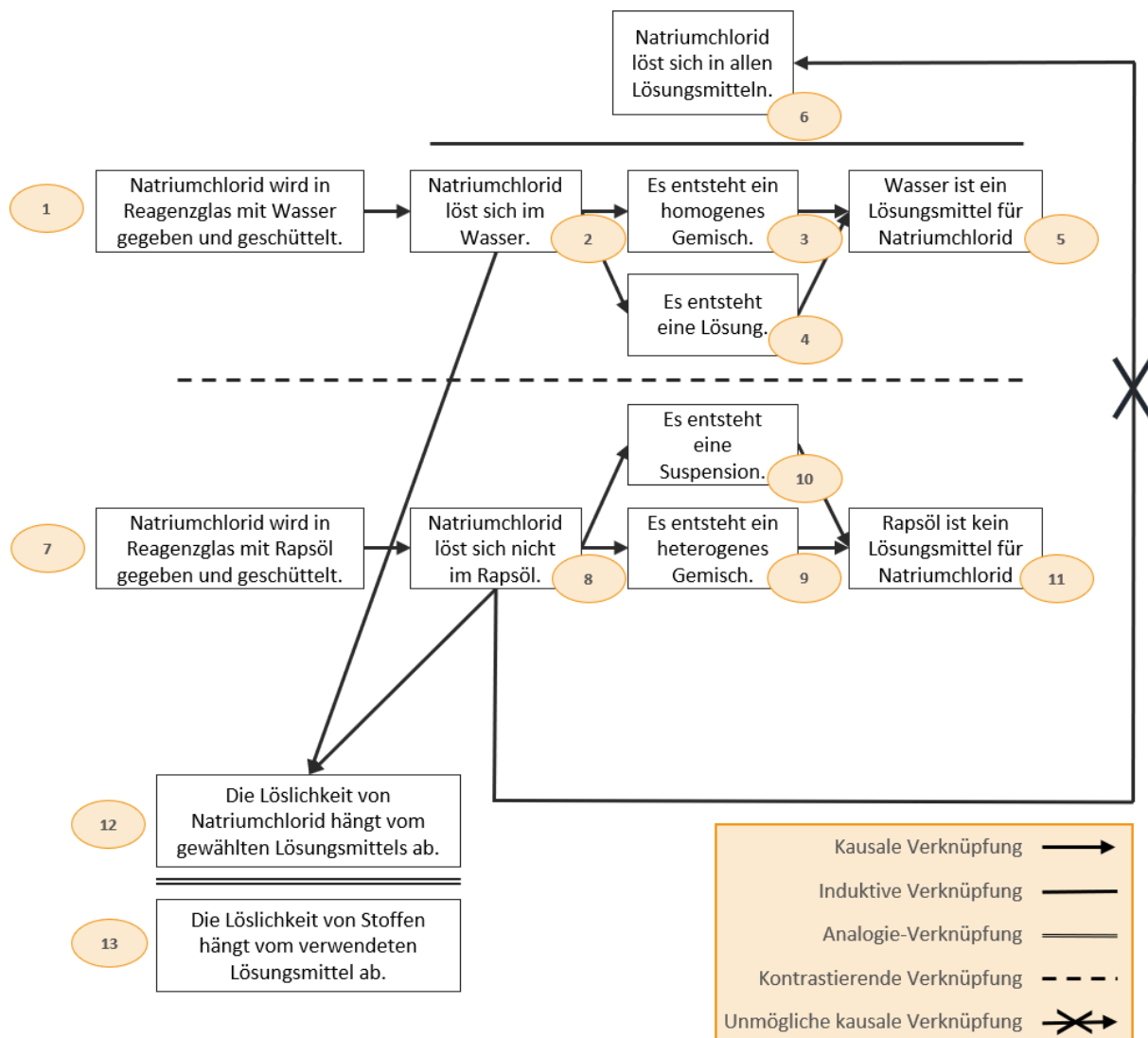


Abbildung 13 - Beispiel eines Model-of-data von Expert*innen zum Experiment "Löslichkeit von Salz in Wasser und Öl". Orange Nummerierung dient nur dem Überblick und der genaueren Beschreibung im Text.

Wie in Abb. 13 zu erkennen ist, nehmen kausale Verknüpfungen den größten Anteil unter der Gesamtzahl der Verknüpfungen ein. Die entspricht dem Postulat von Chinn und Brewer (2001). Kausale Verknüpfungen beziehen einen Effekt auf eine zugrunde liegende Ursache. So kann z.B. zwischen den Aspekten 1 und 2 in Abb. 13 eine kausale Verknüpfung erkannt werden. Dadurch, dass Natriumchlorid in Wasser gegeben und das Reagenzglas geschüttelt wird (Ursache), löst es sich im Wasser (Effekt). Ebenso kann eine kausale Verknüpfung eine Schlussfolgerung darstellen. Hier kann die Schlussfolgerung aus den Aspekten 2 und 8 zu Aspekt 12 als Beispiel angeführt werden. Das Beispiel zeigt ebenfalls die Möglichkeit multipler Verknüpfungen auf. So können zwei Aspekte gemeinsam mit einem weiteren verknüpft werden, oder auch zwei Verknüpfungen aus einem Aspekt folgen (vgl. Aspekte 8). Bei induktiven Verknüpfungen handelt es sich um eine Verallgemeinerung von einem Einzelfall auf eine gesamte Kategorie. Die Verallgemeinerung erstreckt sich dabei auf die nächsthöhere ontologische Kategorie. Als Beispiel kann die Verknüpfung zwischen den Aspekten 2-5 mit dem Aspekt 6 angeführt werden. Die Löslichkeit von Salz in Wasser (Einzelfall) wird hier auf alle Lösungsmittel (Verallgemeinerung) ausgeweitet. Dabei stellt die Kategorie „Lösungsmittel“ die nächsthöhere ontologische Kategorie zum Wasser dar. Die Analogieverknüpfung, welche auch eine Verallgemeinerung darstellt, hingegen bezieht sich auf zwei

unterschiedliche ontologische Kategorien (Chinn & Brewer, 2001). Anhand der Verknüpfung zwischen Aspekten 12 und 13 wird deutlich, dass der Einzelfall (Löslichkeit von Natriumchlorid) auf einen deutlich größeren und damit ontologisch weiter entfernten Zusammenhang (Löslichkeit von allen Stoffen) verallgemeinert wurde. Damit eine solche Verallgemeinerung möglich ist, müssen zwischen den beiden Kategorien einige Ähnlichkeiten vorhanden sein. Kontrastierende Verknüpfungen zeigen Gegensätze auf (Chinn & Brewer, 2001). Es bedarf dazu eines Unterscheidungskriteriums. In der Unterscheidung zwischen den Aspekten 1-6 und 7-11 kann dieses in der Verwendung von Wasser bzw. Rapsöl erkannt werden (Abb. 13). Das Unterscheidungskriterium kann sich entsprechend auf oberflächliche Merkmale beziehen oder auf tiefenstrukturelle Aspekte. So könnte die Unterscheidung auch auf Grundlage der Polarität des jeweiligen Lösungsmittels gebildet werden. In der graphischen Darstellung werden die induktive, Analogie- sowie kontrastierende Verknüpfung immer parallel zwischen den verknüpften Aspekten dargestellt. So wird es möglich, gleichzeitig mehrere Aspekte mit einer Gruppe anderer Aspekte zu verknüpfen. Die letzte verbleibende Verknüpfungsart ist die unmögliche kausale Verknüpfung. Diese Verknüpfung zeigt auf, dass eine Ursache nicht zu einem Effekt führen kann oder dass eine Schlussfolgerung nicht möglich ist. In Abb. 13 kann diese Verknüpfung zwischen den Aspekten 8 und 6 erkannt werden. Wenngleich Aspekt 6 aus einer induktiven Verknüpfung mit den Aspekten 2-5 abgeleitet werden kann, macht Aspekt 8 diesen unmöglich. Dabei ist diese Verknüpfungsart von kontrastierenden Verknüpfungen zu unterscheiden, welche lediglich Gegensätze aufzeigt. Bei unmöglichen kausalen Verknüpfungen werden offene Widersprüche adressiert und ein unmöglicher Zusammenhang dargestellt. Es ist somit möglich Widersprüche in einem MOD darzustellen und ebenso unterschiedliche Erklärungsansätze bzw. Theorien einander gegenüberzustellen (Chinn & Brewer, 2001).

Der Umfang von MOD kann stark variieren. Orientiert an spezifischen Beobachtungen oder Phänomenen können unterschiedliche Aspekte als relevant anerkannt werden. Das oben dargestellte MOD könnte z.B. um die Beschreibung von polaren und unpolaren Lösungsmitteln ergänzt werden. Auch der Löseprozess mit seinen energetischen Abläufen könnte hinzugefügt werden. Die Eingrenzung des spezifischen Umfangs eines MOD obliegt der individuellen Entscheidung und kann durch das zur Verfügung stehende konzeptionelle Wissen bedingt werden. Unterschiede im Umfang von MOD können auch durch die Unterschiede zwischen schulischen und authentisch-wissenschaftlichen Experimenten entstehen (Chinn & Malhotra, 2002b).

Die letztliche Interpretation der Daten bzw. die Akzeptanz dieser (vgl. Chinn & Brewer, 1993b; 1998) kann durch mentale Plausibilitätsprüfungen und Simulationen erreicht werden (Chinn & Brewer, 2001). Dabei beschreiben die Autoren eine besondere Rolle für alternative Verknüpfungen. Sofern es alternative Verknüpfungen und damit Erklärungen für einzelne Aspekte gibt, könnte die Plausibilität in Zweifel gezogen werden. Ebenso könnte in Anlehnung an das *Model of argument* nach Toulmin (2003) eine Bekräftigung durch *warrant* und *backing* stattfinden, welche zur Plausibilitätsprüfung angeführt werden können. Als Beispiel aus dem Expert*innen-MOD könnte die Verknüpfung der Aspekte 2-5 mit Aspekt 6 angeführt werden. Die Verallgemeinerung in Aspekt 6 (*claim*) könnte hier durch *data* (Aspekt 2) erstellt werden. Dabei kann jedoch das Entstehen des homogenen Stoffgemisches als *warrant* agieren, da das Vorliegen eines homogenen Stoffgemisches den Löseprozess begründet. Die Benennung als ‚Lösung‘ könnte hier als *backing* ein Definitionswissen einbringen, welches ebenfalls zur Verallgemeinerung von Aspekt 6 beiträgt.

Die jeweiligen Verknüpfungsarten weisen unterschiedliche Funktionen im Erkenntnisgewinnungsprozess auf und können an epistemologische Überzeugungen angeschlossen werden. Mit Blick auf

die gebildeten Verknüpfungen kann ein unterschiedlicher Anforderungsgrad zur Bildung unterschiedlicher Verknüpfungsarten erkannt werden.

Kausale Verknüpfungen stellen lineare Verknüpfungen von zwei oder mehreren Aspekten mit einem weiteren Aspekt dar. In Bezug auf die Dimension *Simplicity of knowledge* (Hofer & Pintrich, 1997) könnte dieser einfachere Zusammenhang auch mit naiveren epistemologischen Überzeugungen ermöglicht werden. Ebenso bedarf es weniger stark ausgeprägten Überzeugungen bzgl. der Vorläufigkeit von Wissen (*Certainty of knowledge*). Da sich kausale Verknüpfungen in erster Linie auf mind. 2 Aspekte beziehen und diese entweder im Sinne von Ursache-Wirkung-Zusammenhängen oder Schlussfolgerungen aufeinander beziehen, bleiben andere Aspekte/Verknüpfungen dadurch unberührt. Es ist also nicht nötig durch das Postulat einer kausalen Verknüpfung andere Zusammenhänge zu diskreditieren und damit begründetes Wissen aus anderen Bereichen in Zweifel zu ziehen.

Diese Zusammenhänge könnten ähnlich auch auf die kontrastierende Verknüpfung übertragen werden. Die Erstellung kontrastierender Verknüpfungen bedarf ebenso grundlegender Überzeugungen zur Dimension *Justification of knowing* (Hofer & Pintrich, 1997). Es muss begründbar sein, warum ein Kontrast gebildet wird. Dieser kann oberflächlich begründet werden und verlangt nicht zwangsläufig einer tiefenstrukturellen Unterscheidung, wenngleich diese ebenso möglich wäre.

Insbesondere komplexe Verknüpfungen (wie Verallgemeinerungen) zeigen die Bedeutung epistemologischer Überzeugungen auf. Für induktive sowie Analogie-Verknüpfungen könnte ein höherer Grad der Abstraktheit zu höheren Anforderungen bei den Proband*innen führen. Die Nutzung von Verallgemeinerungen bedarf dem Vergleich der beobachteten Entitäten mit solchen aus ontologisch unterschiedlichen Kategorien. So könnten Analogie-Verknüpfungen mit ihren ontologisch deutlich unterschiedlicheren Entitäten schwerer gebildet werden, wenn keine Überzeugungen zur Unsicherheit von Wissen (*Certainty of knowledge*) vorhanden wären. Diese Verallgemeinerungen haben einen heuristischen Anteil. Aufgrund bestimmter Annahmen werden Verallgemeinerungen gebildet, die teils nur auf Grundlage weniger Beobachtungen etc. basieren. Derartige Verallgemeinerungen können teils durch Argumente gestützt, jedoch nicht zu 100% mit Evidenzen argumentiert werden. Dabei stellt die Verknüpfung mit anderen ontologischen Kategorien generell eine Herausforderung dar, welche schon Chi (2008) in ihrem Verständnis von *Conceptual Change* Prozessen beschreibt. Die Autorin beschreibt die nicht angemessene Sortierung von Entitäten in ontologische Kategorien als Ausgangspunkt für einen kognitiven Konflikt und einen notwendigen *Conceptual Change*. Die Proband*innen müssen für das Aufstellen von Verallgemeinerungen zuerst die beobachteten Entitäten in eine ontologische Kategorie zuordnen und anschließend eine davon unterschiedliche Kategorie als Verallgemeinerungsbereich auswählen. Dabei müssen neben den notwendigen ontologischen Unterschieden noch ausreichend Vergleichsmerkmale bestehen, damit die Verallgemeinerung Gültigkeit besitzen kann. Dabei kann hier ebenfalls ein Einfluss der Dimension *Justification of knowing* (Hofer & Pintrich, 1997) erkannt werden. So muss die Vergleichbarkeit und gleichzeitige Unterschiedlichkeit der ontologischen Kategorien begründet werden. Dazu können evidenzbasierte Begründungsprozesse in Anlehnung an das *Model of argument* (Toulmin, 2003) genutzt werden. Es muss dabei gleichzeitig die Komplexität von Wissen innerhalb der Dimension *Simplicity of knowledge* (Hofer & Pintrich, 1997) wahrgenommen werden. Bestehen hingegen Überzeugungen zu einfachem, unvernetzten Wissen, könnte die Konstruktion von Verallgemeinerungen generell gehemmt werden.

Zuletzt kann in unmöglichen kausalen Verknüpfungen ein Bezug zur Dimension *Simplicity of knowledge* erkannt werden. Das Anerkennen von unmöglichen kausalen Verknüpfungen könnte die Komplexität von Wissen aufzeigen. Aspekte, welche mit unmöglichen kausalen Verknüpfungen

verknüpft werden, entstammen häufig anderen Argumentationslinien bzw. Kontexten. Im vorliegenden MOD (Abb. 13) wird die Verallgemeinerung, dass Salz sich in allen Lösungsmitteln löst auf einer Datengrundlage gebildet. Eine Beobachtung innerhalb des zweiten Experimentteils macht diese Aussage nun kausal unmöglich. Damit muss zwischen konkurrierenden Verknüpfungen vermittelt werden, wodurch die Komplexität von Wissen angesprochen wird. Ohne den zweiten Teil des Experiments (Untersuchung der Löslichkeit von Salz in Rapsöl) könnte Aspekt 6 nach wie vor Gültigkeit besitzen. Somit tritt die Kontextabhängigkeit und die Rolle des eigenen Handelns im Zuge der Erkenntnisgewinnung in den Vordergrund. Dabei spielt auch die zeitliche Struktur des Experiments eine Rolle. So könnte wie beschrieben ein gegenteiliger Aspekt zu Aspekt 6 auch aus den Aspekten 8, 9 und 11 gebildet werden. Für unmögliche kausale Verknüpfungen kann die größte Komplexität angenommen werden. Hier könnten auch epistemologische Überzeugungen innerhalb der Dimension *Certainty of knowledge* (Hofer & Pintrich, 1997) deutlich erweiterte Überzeugungen nötig sein. Durch eine unmögliche kausale Verknüpfung werden Aspekte auf Grundlage von anderen Aspekten als unmöglich beschrieben. Dabei können die verknüpften Aspekte jeweils aus längeren Argumentationslinien entspringen und innerhalb dieser als bekräftigte Argumente beschrieben werden. Damit tritt vor allem die Komplexität von Wissen hervor. Es wird also deutlich, dass die unterschiedlichen Verknüpfungsarten eine Interpretation des mentalen Modells auf epistemologischer Grundlage zulassen.

In den Konzeptualisierungen von Chinn und Brewer (2001) wird ein Rückbezug auf die *Conceptual Change* Theorie mit den dahinterliegenden wissenschaftstheoretischen Strömungen deutlich (vgl. Kap. 3.3). Die Entscheidung über die jeweils genutzten Verknüpfungen kann an die Konzeption der *framework theory* angegliedert werden (Vosniadou, 1994, 2014). Die unterschiedlichen epistemologischen und ontologischen Qualitäten der Verknüpfungen hängen eng mit den jeweiligen Überzeugungen zusammen. Diese Überzeugungen können als Bestandteil der *framework theory* angesehen werden und zeigen so einen möglichen Einfluss auf die Erstellung des mentalen Modells auf. Ebenso ist eine Nähe zum *Model of argument* von Toulmin (2003) zu erkennen. Im *Model of argument* wird die Begründung für einen *claim* auf Grundlage von Daten und weiteren Argumenten dargelegt. Diese Begründungsmuster sind für die Erstellung der unterschiedlichen Verknüpfungsarten im Model-of-data notwendig. Die Entscheidung darüber, um welche Verknüpfungsart es sich jeweils handelt, muss vor dem Hintergrund der Daten, wie auch weiterer Argumente (hier: *warrent*, *backing* und *rebuttal*) erfolgen. Die letztliche Begründung könnte vor allem durch die Verknüpfungen unterschiedlicher (konzeptueller) Aspekte erfolgen:

[...] we must bring forward not further data, for about these the same query may immediately be raised again, but propositions of a rather different kind: rules, principles, inference-licences or what you will, instead of additional items of information. Our task is no longer to strengthen the ground on which our argument is constructed, but is rather to show that, taking these data as a starting point, the step to the original claim or conclusion is an appropriate and legitimate one. (Toulmin, 2003, S. 91)

Toulmin (2003) hebt hier ausdrücklich die Rolle der Verknüpfungen und die zugrunde liegenden Regeln als bedeutsam hervor. Die Daten stellen somit den Ausgangspunkt für die Argumentation dar, werden jedoch erst durch die Verknüpfung miteinander und die Begründung dieser Verknüpfungen zum Argument. Hier ist eine starke Ähnlichkeit mit Model-of-data erkennbar, welche ebenfalls die Verbindung von Aspekten mit den jeweiligen Verknüpfungen in das Zentrum des mentalen Modells stellen.

Obgleich die vorgestellte graphische Darstellungsart für diese Studie genutzt wurde, erheben Chinn und Brewer (2001) keinen Anspruch darauf, dass mentale Modelle in ihrer Struktur mit

MOD übereinstimmen: „[MOD] do not make any strong claims about the representational format of the constructed models. The models may be semantic networks with nodes and links, mental models that incorporate mental images, or sets of propositions.” (Chinn & Brewer, 2001, S. 336). Die hier dargestellte graphische Umsetzung stellt eine Möglichkeit der Externalisierung des individuellen mentalen Modells dar und bietet dafür strukturelle Hilfestellungen. Verbindlich sind dabei die unterschiedlichen Verknüpfungsarten. Die spezifische Form, Farbe oder der Umfang des MOD können individuell deutlich abweichen.

7.1.3 Digitale Umsetzung der Modellierungsstudie

Auf Grundlage der Kontaktbeschränkungen im Rahmen der COVID-19-Pandemie, wurde die Studie im Sommersemester 2021 digital umgesetzt. Die Interviews fanden mithilfe eines Videokonferenzsystems (Zoom) virtuell statt. Das Tool wurde zum Zeitpunkt der Studie von der Universität Paderborn zur Nutzung freigegeben. Ebenso wurden die Proband*innen über Datenschutzaspekte rund um das Tool aufgeklärt. Vor dem vereinbarten Interview-Termin erhielten die Proband*innen wichtige Informationen zur technischen Umsetzung via E-Mail. So sollte neben dem Konferenzsystem auch eines der Präsentationsprogramme PowerPoint, Keynote oder Open-Office auf dem PC installiert werden. Ebenso wurden die Proband*innen gebeten die Videokonferenz in einer möglichst ruhigen Umgebung und im Idealfall mit einem Headset abzuhalten.

Innerhalb der Modellierungsphase wurde hauptsächlich die Bildschirmfreigabe vom Interviewer genutzt. Zur eigenen Erstellung eines MOD wurde den Studierenden eine Vorlage im Format ihres genutzten Präsentationsprogramms zur Verfügung gestellt (siehe Anhang).

Die Vorlagendatei besteht aus fünf Bereichen. Im obersten Bereich der Datei wird der persönliche Code eingetragen, um eine spätere Verknüpfung der unterschiedlichen MOD sowie der zugehörigen Videos zu ermöglichen. Auf der linken Seite befindet sich die Gedankenstütze. In dieser werden die Konstituenten eines MOD kurz beschrieben und Beispiele gegeben. Die Beispiele entstammen dabei der vorherigen Erklärungsphase zu MOD. Auf der rechten Seite ist die Gedankensammlung zu finden. Im Rahmen der hier gezeigten Vorlage für das Methodentraining innerhalb der Modellierungsphase sind bereits Beobachtungsaspekte des entsprechenden Experiments eingetragen. Während die restliche Vorlagendatei identisch auch in der Studienphase eingesetzt wird, gibt es im Bereich der Gedankensammlung einen Unterschied. Innerhalb der Studienphase wurden keine Aspekte innerhalb der Gedankensammlung vorgegeben. Die Proband*innen konnten diesen Bereich für eigene Notizen nutzen, z.B. um sich während der Beobachtung im Video relevante Aspekte zu notieren. Im oberen Bereich der Datei befindet sich eine Toolbox. In dieser werden unterschiedliche Formen und Pfeile/Striche zur Verfügung gestellt, welche zur Konstruktion eines MOD benötigt werden. Während rechts neben der Toolbox erneut die unterschiedlichen Pfeile/Striche mit den dazugehörigen Konstituenten eines MOD beschrieben werden, können die Proband*innen sich für die Konstruktion des MOD aus der Toolbox bedienen. Die einzelnen Formen können mit einem Mausklick aus der Toolbox herausgezogen und im MOD positioniert werden. Den Hauptteil der Datei stellt die Arbeitsfläche in der Mitte dar. Hier sollen die Proband*innen ihr MOD erstellen. Die Vorlagendatei wurde den Proband*innen ausführlich vorgestellt. Ebenso wurden die wichtigsten Funktionen des jeweiligen Präsentationsprogramms durch den Interviewer demonstriert. So sollten technische Herausforderungen abgemildert werden.

Die Experimente wurden in Form von Experimentiervideos zur Verfügung gestellt. Dabei wurde darauf geachtet, dass der Experimentierprozess möglichst in Gänze gezeigt wird. So wurde an einzelnen Stellen z.B. auch Zeitrafferaufnahmen zurückgegriffen, um möglichst wenig Bildschnitte zu erzeugen. Die Videos wurden durch ein Videoportal (YouTube) zur Verfügung gestellt. Die

Proband*innen konnten hier im Video spulen, pausieren, oder auch in der Bearbeitung des MOD erneut auf das Video zurückgreifen. Da die Proband*innen während der Erstellung der MOD ihren Bildschirm freigegeben haben, kann die Interaktion mit den Videos nachvollzogen werden.

Innerhalb der Studienphase wurde die Aufzeichnungsfunktion des Videokonferenzsystems genutzt. So wurde der freigegebene Bildschirm der Proband*innen, deren Audiosignal sowie wahlweise auch das Videosignal aufgezeichnet. Auftretende Verbindungsprobleme haben währenddessen teilweise zu Verständnisschwierigkeiten geführt. Diese wurden nach kurzer Pausierung geklärt. Auch bestanden teilweise Probleme in der flüssigen Wiedergabe der Experimentiervideos. Es kam jedoch nicht zu so starken Verbindungsproblemen, dass ein Interview hätte abgebrochen werden müssen.

7.2 Stichprobe der Modellierungsstudie

Die Stichprobe dieser Teilstudie generiert sich aus Studierenden des Lernbereichs Natur- und Gesellschaftswissenschaften (Sachunterricht) im Studiengang für das Lehramt an Grundschulen sowie das Lehramt mit Sonderpädagogischer Förderung (B.Ed.). Die Studierenden belegen im naturwissenschaftlichen Bereich im Anteilsfach Chemie die Veranstaltung „Chemische Grundlagen“. Bei dieser Veranstaltung handelt es sich um eine Großvorlesung, inkl. Übung und Tutorium. Inhaltlich werden vor allem Inhaltsbereiche der Chemie der Sekundarstufe 1 bearbeitet (z.B. Stoffeigenschaften, Redoxreaktionen, Säure-Base-Reaktionen, etc.). Die Teilnahme an der Studie erfolgte freiwillig. Als Anreiz für die Teilnahme, konnte durch diese ein Anteil der qualifizierten Teilnahme-Leistung für die Veranstaltung "Chemische Grundlagen" erworben werden. Für die Erstellung der Datensätze wurden pseudonymisierte Codes für die Proband*innen erstellt.

Die Gesamtstichprobe umfasst dabei $N=59$ Studierende ($M_{\text{Alter}}=21,1$ Jahre, $SD=3,16$ Jahre). Dabei liegen von $N=56$ Proband*innen vollständige Datensätze vor, die als Grundlage für die Analyse genutzt wurden. Der Anteil weiblicher Studierender von ca. 91,5% entspricht einem generellen Trend im Lehramt für die Primarstufe.¹³ Die Studierenden befanden sich zum Erhebungszeitpunkt am Anfang ihres Studiums ($M_{\text{Semester}}=1,6$).

Aufgrund der sehr umfangreichen Datensätze der einzelnen Proband*innen, welche nur im Gesamtzusammenhang interpretiert werden können, erfolgte eine zweistufige Auswertung. Im ersten Schritt wurden eine Grobauswertung der Adäquatheit der beiden konstruierten MOD der Proband*innen vorgenommen. Anschließend wurde für jeden Fall eine durchschnittliche Adäquatheit der beiden konstruierten Model-of-data gebildet. Die Fälle wurden dann wie in Tabelle 6 zu sehen in fünf Adäquatheits-Bereiche eingeteilt. Im nächsten Schritt wurde der prozentuale Anteil an der Gesamtfallzahl für die einzelnen Adäquatheitsbereiche errechnet. Für die vertiefte Analyse sollten nun $N=20$ Fälle ausgewählt werden. Dazu wurde die prozentuale Verteilung der Fälle auf die Adäquatheitsbereiche auf eine Stichprobe von $N=20$ übertragen. So entsteht eine gerundete Stichprobengröße von $N=21$ Fällen für die Analysestichprobe. Die ausgewählten Fälle sollen dabei, in Anlehnung an merkmalspezifisch-repräsentative Stichproben (Döring & Bortz, 2016), die Gesamtstichprobe so gut es geht abbilden. Die Begrenzung auf eine Teilstichprobe von $N=21$ sollte vor dem Hintergrund der Auswahl dieser Fälle betrachtet werden. Baker und Edwards (2012) setzen sich mit der Frage auseinander, wie viele qualitative Interviews ausreichend sind. Neben Daumenregeln wie 12-60 Proband*innen, stellen die Autor*innen vor allem den Kontext der Studie heraus. Auch verweisen die Autor*innen auf die Bedeutung der Auswahl der Stichprobe aufgrund von

¹³ Vergleiche dazu: <https://www.daten.bmfsfj.de/daten/daten/anteil-von-frauen-und-maennern-an-den-erfolgreich-abgelegten-abschlusspruefungen-fuer-das-lehramt-primarbereich-134444>

Vergleichsvariablen. In der vorliegenden Studie wurde dazu die Adäquatheit der gebildeten MOD genutzt. Weiterhin wurden in der vorliegenden Studie neben Interviews auch weitere Daten gesammelt. So kann die Analyse auf eine breite Datenbasis gestellt werden. Guest et al. (2006) beschreiben die Bedeutung von Stichproben-Homogenität für die Auswahl der passenden Stichprobengröße für qualitative Interviews. In der vorliegenden Stichprobe werden ausschließlich Studierende der Veranstaltung „Chemische Grundlagen“ ausgewählt. Diese Studierenden befinden sich in zwei Studiengängen, welche beide auf die Primarstufe fokussieren. Insofern kann von einer relativ homogenen Stichprobe ausgegangen werden, da diese durch den Fokus auf die Primarstufe sowie das Fach Sachunterricht und die Teilnahme an der gleichen Großvorlesung in nah verwandten Studiengängen studieren. Für eine solche Stichprobe und ein spezifisches Erkenntnisinteresse, welches durch ein Kodiermanual beschrieben wird, zeigen die Autoren*innen eine Stichprobe von $N=12$ als ausreichend auf. Dabei ist diese Zahl an die von Guest et al. (2006) durchgeführte Studie gebunden. Dennoch erkennen die Autor*innen Potentiale der Verallgemeinerung, wenngleich der spezifische Kontext berücksichtigt werden müsse. Im Rahmen der Analyse wurde entsprechend auch auf auftretende Sättigung geachtet. Die Begrenzung der Stichprobe auf $N=21$ und damit ca. 1/3 der Gesamtstichprobe kann dadurch begründet werden.

Tabelle 28 – Fallauswahl für die Analyse in Anlehnung an die Verteilung innerhalb der Gesamtstichprobe

Adäquatheits-Bereiche	Fallanzahl	Prozentualer Anteil an Gesamtstichprobe ($N=56$)	Fallanzahl bezogen auf eine geplante Analysenstichprobe ($N=20$)	Gerundete Fallanzahl für Analyse-Stichprobe ($N=21$)
0-20%	1	1,79	0,38	1
20-40%	2	3,57	0,71	1
40-60%	14	25	5	5
60-80%	20	35,71	7,14	7
80-100%	19	33,93	6,79	7

Die Auswahl der einzelnen Fälle aus den jeweiligen Adäquatheitsbereichen wurde auf Grundlage der Verteilung innerhalb der jeweiligen Bereiche umgesetzt (vgl. Tab. 28). Bezogen auf einen Adäquatheits-Bereich wurden die jeweiligen Randextrema sowie eine möglichst gleichmäßige Verteilung über die Spannbreite angestrebt.

7.3 Ablauf der Analyse

Model-of-data und lautes Denken

Zur Analyse der vorliegenden Model-of-data (MOD) sowie der dazugehörigen Videos wurde das Programm MAXQDA genutzt. Zuerst wurde das entwickelte Kategoriensystem im Programm hinterlegt. Anschließend wurden die entwickelten MOD sowie die Videos des lauten Denkens importiert. Während die MOD getrennt für die beiden Experimente der Interviewphase vorlagen, wurden die Videos in einer Datei pro Proband*in importiert. Die Analyse begann mit der Kodierung eines MOD zu einem Experiment. Die importierten Bilddateien wurden in MAXQDA kodiert, indem Umrandungen der zu kodierenden Bereiche erstellt wurden. Diese Umrandungen stellen dann die Kodierungen dar. Anschließend wurde der Videoausschnitt des lauten Denkens der Studienteilnehmenden kodiert. Dieser Zwischschritt wurde aufgrund der Komplexität der erstellten MOD

gewählt. Abweichungen zwischen dem analysierten MOD und dem lauten Denken konnten so im direkten Abgleich zueinander bestimmt werden. So sollte sichergestellt werden, dass alle Abweichungen erkannt werden. Dieser Zweischritt wurde anschließend randomisiert wiederholt, sodass nicht zwei Datensätze der gleichen Studienteilnehmenden nacheinander analysiert wurden. Kodierte Stellen innerhalb der Videos wurden transkribiert, sodass Teiltranskripte entstanden sind. Die Analyse der Datensätze wurde zwischen zwei Kodierenden aufgeteilt, wobei einige Datensätze von beiden Kodierenden analysiert wurden (siehe hierzu Kap. 7.5.1).

Fachwissenstest

Die Analyse des Fachwissenstest sowie der Selbsteinschätzung/Angabe der eigenen Leistungen in verschiedenen Inhaltsfeldern erfolgte im Programm SPSS. Hierzu wurden die digital erhobenen Daten aus dem Umfragetool (Limesurvey) exportiert und in SPSS importiert. Eine anschließende Rekodierung der Variablen (innerhalb von Limesurvey werden die Antworten mit den Optionen A-D ausgegeben) in die entsprechende Punktzahl (falsche Antwort – 0, richtige Antwort – 1) sowie die Beschriftung aller Variablen ermöglichten die weitergehende Analyse.

7.4 Instrumente

Im Folgenden sollen die Instrumente zur Datenerhebung vorgestellt werden. Wie beschrieben, erfüllen die in Kap. 7.1.2 vorgestellten *Model-of-data* (MOD) eine Doppelrolle als Methode zur Externalisierung mentaler Modelle und als Erhebungsinstrument. Die MOD werden innerhalb dieses Kapitels nicht erneut dargestellt. Dafür wird jedoch die Auswertung dieser im Rahmen von Kapitel 7.5.1 vorgestellt.

7.4.1 Fachwissenstest

Der Fachwissenstest kann in zwei Teile gegliedert werden. Einerseits werden Leistungen im Fach Chemie und Einschätzung der eigenen Leistungsfähigkeit der Proband*innen abgefragt. Dazu wurde die jeweils letzte erhaltene Note im Fach Chemie gemeinsam mit dem Zeitpunkt dieser abgefragt. Ebenso sollten die Proband*innen die eigenen Fähigkeiten/Kenntnisse in den Inhaltsbereichen *Säure-Base-Chemie*, *Stoffeigenschaften*, *chemisches Wissen/Kenntnisse allgemein* und *physikalisches Wissen/Kenntnisse allgemein* einschätzen (fünf-stufige Likert-Skala von *sehr schlecht* bis *sehr gut*). Die Abfrage der Abiturnote dient der Erfassung des generellen Leistungsniveaus.

Um andererseits einen Einblick in das chemische Fachwissen der Proband*innen bzgl. der beiden Inhaltsbereiche „Säure-Base“ und „Siedepunkt“ zu erlangen, wird ein Fachwissenstest eingesetzt. Der Fachwissenstest besteht aus zwei Skalen zu Vorstellungen in den Inhaltsbereichen „Säure-Base“ und „Siedepunkt“, die sich inhaltlich jeweils auf die Experimente der Studienphase beziehen. Die Items zum Säure-Base-Experiment wurden einem bestehenden Testinstrument von Kehne (2019) entnommen. Aus den 20 von Kehne (2019) auf Grundlage des Kernlehrplans und eines Expert*innen-Netztes (Martensen & Demuth, 2008) entwickelten Items, wurden 11 zur vorliegenden Studie passende Items ausgewählt. Die Items decken in Form von single-choice Items mit vier Antwortmöglichkeiten zentrale Aspekte zum Säure-Base-Verständnis ab. Tabelle 29 gibt einen Überblick über Beispielitems. Die Testitems zum Siedepunkt-Experiment entstammen zwei unterschiedlichen Testinstrumenten. Der Großteil der Items wurde der *Thermal Concept Evaluation* (Yeo & Zadnik, 2001) entnommen. Aus den 26 Items des Testinstruments von Yeo und Zadnik (2001) wurden sieben relevante Items entnommen. Hierbei handelt es sich ebenfalls um single-choice Items mit 4-5 Antwortmöglichkeiten. Um das Itemformat für das genutzte Testinstrument

gleich zu halten, wurden lediglich Items mit vier Antwortmöglichkeiten ausgewählt. Zusätzlich zu den sieben Items aus der *Thermal Concept Evaluation*, wurde ein Item aus dem *Energy Concept Inventory* (Swackhamer et al., 2005) entnommen. Dieses Testinstrument fokussiert auf Energie in einem breiten Verständnis. Neben Items zu thermischer Energie, werden hier auch Items im Zusammenhang zu Elektrik, Mechanik und Biologie entwickelt. Bei dem ausgewählten Item handelt es sich um ein single-choice Item mit fünf Antwortmöglichkeiten. Zur Vereinheitlichung des Test- und Antwortformats wurde ein Distraktor gestrichen, sodass das Item ebenfalls vier Antwortmöglichkeiten aufweist. Bei dem Item handelt es sich um das zweite Item der Siedepunktskala aus Tab. 32. Der gestrichene Distraktor beschreibt das schnelle Ansteigen der Temperatur durch das vermehrte Kochen und weist damit eine deutliche Ähnlichkeit zu Antwortmöglichkeit 1 auf. Aufgrund der Ähnlichkeit der beiden Distraktoren und einem Überfluss von Antwortmöglichkeiten, welche eine Zunahme der Temperatur beschreiben, wurde dieser Distraktor gestrichen.

Während die Items von Kehne (2019) auf Deutsch vorlagen, mussten die Items der Siedepunktskala aus dem Englischen übersetzt werden. Dazu wurden die englischen Items von einer Anglistin ins Deutsche übersetzt. Um sicher zu stellen, dass die Items in ihrer deutschen Version mit den englischen Items übereinstimmen, wurde die Technik der Rückübersetzung angewendet (Behr et al., 2015). Dazu wurden die Deutschen Items durch eine englische Muttersprachlerin und einen Professor für Fachdidaktik Englisch zurückübersetzt. Im Sinne der Kritik an der Praxis der Rückübersetzung (siehe z.B. Behr et al., 2015) als basierend auf Einzelmeinungen, kann durch die Beteiligung mehrerer Personen an den einzelnen Schritten ein Vorteil erkannt werden. Die Rückübersetzung wurde anschließend mit den Ausgangsitems aus den Ursprungstestinstrumenten verglichen. Es konnten keine relevanten semantischen Abweichungen zwischen den Versionen erkannt werden. So wurden die deutschen Items in das vorliegende Testinstrument übernommen.

Tabelle 29 – Beispielitems des Fachwissenstests

Säure-Base-Skala (Kehne, 2019)	Siedepunkt-Skala (Swackhamer et al., 2005; Yeo & Zadnik, 2001)
Die Reaktion einer Säure und einer Base nennt man...	Auf dem Herd steht ein mit Wasser gefüllter Wasserkessel. Das Wasser hat schnell angefangen zu kochen. Die wahrscheinlichste Temperatur des Wassers ist ungefähr:
<input type="radio"/> ... Neutralisationsreaktion. <input type="radio"/> ... Knallgasreaktion. <input type="radio"/> ... Elektronenreaktion. <input type="radio"/> ... Redoxreaktion.	<input type="radio"/> 88°C <input type="radio"/> 98°C <input type="radio"/> 110°C <input type="radio"/> Keiner der oben genannten Antworten kann richtig sein.
Je mehr Base man zu einer Säure gibt, desto...	Flüssiges Wasser kocht in einem Topf auf dem Gasherd. Der Regler ist auf „NIEDRIG“ gestellt. Während das Wasser kocht, wird der Regler auf „HOCH“ gestellt. Was passiert mit der Temperatur des flüssigen Wassers?
<input type="radio"/> ... konstanter bleibt der pH-Wert. <input type="radio"/> ... höher steigt der pH-Wert. <input type="radio"/> ... eher pendelt sich der pH-Wert bei 0 ein. <input type="radio"/> ... niedriger sinkt der pH-Wert.	<input type="radio"/> Sie nimmt geringfügig ab, weil das Wasser mehr kocht. <input type="radio"/> Sie nimmt zu und tut dies zu einem höheren Tempo als zu dem Zeitpunkt, zu dem der Regler auf „NIEDRIG“ gestellt war. <input type="radio"/> Sie nimmt zu und tut dies im gleichen Tempo als zu dem Zeitpunkt, zu dem der Regler auf „NIEDRIG“ gestellt war. <input type="radio"/> Sie bleibt gleich.

Pro Item kann bei richtiger Beantwortung 1 Punkt erreicht werden. Damit liegt die maximale Gesamtpunktzahl der Säure-Base-Skala bei 11 Punkten, für die Siedepunkt-Skala bei 8 Punkten.

Für das gesamte Testinstrument ergibt sich eine unzufriedenstellende interne Konsistenz in der Stichprobe ($\alpha = .472$, $N=55$). Eine Untersuchung der internen Konsistenz der beiden Subskalen zeigt deutliche Unterschiede zwischen diesen auf (Tab. 30).

Tabelle 30 – Testgüte der Subskalen des eingesetzten Fachwissenstest vor Bearbeitung

Sub-Skala	Itemanzahl	Stichprobengröße N	Interne Konsistenz [Cronbachs α]
Säure-Base-Skala	11	52	.67
Siedepunkt-Skala	8	55	.03

Die interne Konsistenz für die Säure-Base-Skala entspricht in etwa dem vollständigen Testinstrument von Kehne (2019). Eine Betrachtung der einzelnen Items und deren Einfluss auf die interne Konsistenz mittels Item-Skala-Statistiken zeigt keine deutlichen Verbesserungsmöglichkeiten bei der Exkludierung einzelner Items auf. Folglich wird die Subskala in dieser Version zur weiteren Analyse genutzt. Für die Items der Siedepunkt-Skala liegt kein Vergleichswert für die interne Konsistenz vor, da die Items aus zwei unterschiedlichen Testinstrumenten stammen. Um die interne Konsistenz dieser Subskala zu verbessern, wurden Item-Skala-Statistiken gebildet und kritisch reflektiert. Das Ausschließen von Item TP5 zeigt eine deutliche Verbesserung der internen Konsistenz zu $\alpha_{\text{Siedepunkt-Skala}} = .29$ auf (vgl. Tab. 31). Inhaltlich betrachtet kann ein Unterschied dieses Items zu den restlichen Items der Subskala beobachtet werden. Während sich die restlichen Items der Sub-Skala hauptsächlich mit der Temperatur von Wasser und deren Entwicklung in verschiedenen Settings beschäftigen, fokussiert TP5 auf die Zusammensetzung des Wasserdampfs. Das Item fragt ab, woraus die aufsteigenden Blasen in kochendem Wasser bestehen und legt somit einen Fokus auf die stoffliche Zusammensetzung. Damit weicht das Item inhaltlich von der Skala ab. In Bezug auf das beobachtete Experiment besitzt die Zusammensetzung der Gasblasen zwar eine Relevanz, für die Beobachtung eines erneuten Siedens unterhalb 100°C ist die Zusammensetzung jedoch eher nachrangig. Entsprechend kann dieses Item aus der Subskala ausgeschlossen werden. Eine erneute Betrachtung der Item-Skala-Statistiken für die verbleibenden sieben Items zeigt weiteres Verbesserungspotential für die interne Konsistenz auf. So könnte die Exkludierung von Item TP1 zu einer weiteren Steigerung des Alpha-Werts auf $\alpha = .37$ führen. Vor dem Hintergrund des Experiments besitzt dieses Item jedoch eine hohe Relevanz, fragt es doch nach der (wahrscheinlichsten) Temperatur von Wasser am Siedepunkt. Aufgrund der hohen Relevanz des Items wird auf einen Ausschluss verzichtet. Ebenso könnte durch den Ausschluss des Items TP1 keine interne Konsistenz erreicht werden, welche eine belastbarere Analyse der Subskala zulassen würde.

Tabelle 31 – Testgüte der Subskalen des eingesetzten Fachwissenstest nach Bearbeitung

Sub-Skala	Itemanzahl	Stichprobengröße N	Interne Konsistenz [Cronbachs α]
Säure-Base-Skala	11	52	.67
Siedepunkt-Skala	7	55	.29

Bei der Interpretation der Ergebnisse der Siedepunkt-Subskala muss die unzufriedenstellende interne Konsistenz berücksichtigt werden. Mögliche Gründe dafür könnten in der geringen Itemanzahl sowie der relativ kleinen Stichprobe vermutet werden. Vor allem die Itemanzahl kann einen

direkten Einfluss auf Cronbachs Alpha haben (Döring & Bortz, 2016). Schecker (2014) plädiert für einen produktiven Umgang mit niedrigen Alpha-Werten. Dabei stellt er vor allem die Herausforderungen dar, in zeitökonomischen fachdidaktischen Tests hohe Alpha-Werte zu erreichen. Gleichzeitig bestünde die Gefahr, dass die Alpha-Optimierung zu inhaltlichen Einschnitten führen könnte. Schecker (2014) unterscheidet lernzielbezogene und diagnostische Testverfahren. Während diagnostische Verfahren die Kompetenzen von Lernenden in großen Bereichen (oft mit vielen Items) erheben, fokussieren lernzielbezogene Testverfahren auf wesentlich spezifischere Inhaltsbereiche. Wenngleich der hier vorliegende Fachwissenstest das Wissen der Lernenden bzgl. des Siedepunkts von Wasser und dessen Einflussfaktoren erheben soll, stellt dies einen eher kleinen Inhaltsbereich dar. Die Items wurden eng am Experiment orientiert und könnten somit auch als lernzielbezogen verstanden werden. Wenngleich die Unterscheidung nach Schecker (2014) nicht trennscharf zu verstehen ist, zeigen seine Überlegungen die hohe Relevanz inhaltlicher Passung der Items auf. Auf dieser Grundlage könnten die Ergebnisse der Siedepunkt-Subskala vorsichtig interpretiert werden. Ein Bedarf an Überarbeitung der Subskala durch die niedrige interne Konsistenz mit einer größeren Stichprobe und ggf. ergänzenden Items ist dennoch angezeigt. Für die weitere Analyse können daher auch qualitative Zusammenhänge der Beantwortung von Einzelitems und der Gestaltung der MOD betrachtet werden. Eine umfassende, belastbare Interpretation der Subskala im quantitativen Sinne ist auf Grundlage der dargestellten internen Konsistenz nicht möglich.

Der Fachwissenstest soll im Zuge der Analyse als Validitätsargument für die externalisierten MOD genutzt werden. Eine konzeptuelle Aufklärung der Zusammenhänge innerhalb der Experimente setzt ein gewisses Konzeptverständnis voraus (vgl. Kap. 7.1.1). Mithilfe des Fachwissenstest soll dieses konzeptuelle Verständnis innerhalb der Inhaltsbereiche erhoben werden.

7.4.2 Prozessdaten (lautes Denken, Notizen)

Wie durch Chinn und Malhotra (2002a) beschrieben, finden Beobachtungen häufig im Zusammenhang mit eigenen Vorstellungen statt. Diese könnten Einfluss auf die Verarbeitung der Beobachtungen haben. Dieser Einfluss könnte durch die Externalisierung der mentalen Modelle in Form von MOD nur unklar hervortreten. Auch Hinweise auf mögliche Umgangsarten mit anomalen Beobachtungen nach Chinn und Brewer (1993b) könnten in der graphischen Darstellung ggf. schwer erkennbar sein. Um hier weitere Hinweise für die Eignung der Methode der MOD im Sinne von Forschungsfrage III.b (Eignung der MOD um Reaktion auf anomale Beobachtungen darzustellen) zu erhalten, wurde die Methode des lauten Denkens genutzt (für einen Überblick: Sandmann, 2014). Lautes Denken meint die Verbalisierung jeglicher Gedanken, die im Zusammenhang mit einer Frage-/Problemstellung entstehen (Völzke, 2012). Dabei sollen die Proband*innen jegliche Gedanken frei äußern und sich dabei auf die Aufgabe fokussieren. Ericsson und Simon (1998) weisen explizit darauf hin, dass die Proband*innen ihre Gedanken nicht einer anderen Person erklären, sondern lediglich ihren (sequentiellen) Denkprozess verbalisieren sollen. Gleichzeitig bestünde die Gefahr, dass automatisierte Prozesse nicht verbalisiert werden, wodurch sich die Methode des lauten Denkens vor allem bei neuen und komplexen Aufgaben anbieten würde (Völzke, 2012).

Neben dem simultanen lauten Denken, ist auch eine nachträgliche Umsetzung denkbar. Diese wird im Rahmen dieser Studie durch das anschließende Interview eingesetzt. Die Proband*innen sollen hier auf Grundlage ihres MOD Denkprozesse zu bestimmten Aspekten erneut formulieren bzw. erläutern. Das Vorgehen entspricht dabei der Validierungsmethode des *Member Checkings* (Döring & Bortz, 2016). Ziel dieser Phase ist es, Unklarheiten aufzuklären bzw. Denkprozesse zu

verbalisieren, welche simultan nicht umfangreich beschrieben wurden (Völzke, 2012). Es ist dabei nicht Aufgabe der Proband*innen das MOD zu überarbeiten, wenngleich die Formulierung von Abweichungen zur ursprünglichen Darstellung möglich ist. Um im Interview auf offengebliebene Zusammenhänge gezielt eingehen zu können, wurden während der Erstellung des MOD inkl. simultanem lauten Denken Notizen durch den Interviewer erstellt (vgl. Phillippi & Lauderdale, 2018). Die Notizen wurden in einem vorbereiteten Dokument erfasst. Um relevante Aspekte innerhalb der Notizen festzuhalten bzw. entstehende Unklarheiten für das spätere Interview strukturiert zu erfassen, konnte auf eine Fragenstruktur zurückgegriffen werden. Hier wurden die unterschiedlichen Bereiche des MOD wie z.B. die gebildeten Aspekte oder die unterschiedlichen Verknüpfungsarten aufgelistet. Im Rahmen des lauten Denkens sollte diese Strukturierung den Interviewer darin unterstützen relevante Notizen zu sammeln und für das Interview nutzbar zu machen.

7.5 Datenauswertung

7.5.1 Erstellung des Kodiermanuals

Die Entwicklung des Kodiermanuals wurde nah an der Konzeption der *Model-of-data* (MOD) ausgerichtet (Chinn & Brewer, 2001). Entsprechend wurden sowohl die inhaltlichen Aspekte als auch die jeweiligen Verknüpfungen in die Analyse einbezogen. Damit orientiert sich die Kodierung der MOD an Vorschlägen zur Beurteilung von *Concept Maps*. Ruiz-Primo und Shavelson (1996) schlagen zur Beurteilung von Concept Maps drei verschiedene Ansätze vor. Eine erste Möglichkeit wäre die (1) Bewertung der einzelnen Komponenten der gebildeten Map. Dabei beziehen sich die Autor*innen auf ein Scoring System von Novak und Gowin (1984). Hier werden die gebildeten Propositionen (Verbindung von mind. zwei Aspekten), Hierarchielevel sowie Querverweise bepunktet. Dieses Vorgehen könnte zwar generell auch auf die Bewertung von MOD angewendet werden, scheint jedoch z.B. im Punkt der Hierarchie von der Konzeption der MOD abzuweichen. Die zweite von Ruiz-Primo und Shavelson (1996) vorgestellte Bewertungsmöglichkeit besteht im (2) Vergleich mit einer Expert*innen-Map. Unterschiede innerhalb der MOD (z.B. im Umfang) könnten nach Chinn und Malhotra (2002b) auch durch die Authentizität des wissenschaftlichen Vorgehens vermutet werden. Somit könnte eine Expert*innen-Map deutlich von der Lösung der Proband*innen abweichen. Weiterhin würde die Zugrundelegung einer Expert*innen-Map voraussetzen, dass es eine Ideallösung für das MOD gäbe. Der Individualität der mentalen Modelle könnte so ggf. nicht gerecht geworden werden. Zwar könnten essentielle Aspekte und Verknüpfungen fallübergreifend verglichen werden, jedoch könnten individuelle Beobachtungen auftreten (vgl. individuelle Wahrnehmung anomaler Beobachtungen, Kap. 7.1.1), welche nur auf Ebene der individuellen MOD analysiert werden können. Diese Individualität könnte es auch generell erschweren eine Expert*innen-Map zu erstellen. Je nach Perspektive und den epistemologischen Überzeugungen der beteiligten Personen könnten spezifische Zugänge zu bestimmten Mustern führen. Ähnliche Herausforderungen bzgl. dem alleinigen Vergleich mit einer Expert*innen-Map werden auch von Ruiz-Primo und Shavelson (1996) beschrieben. Daher schlagen diese als weiteren Ansatz eine (3) Kombination der beiden vorgestellten Bewertungsmöglichkeiten dar.

Für die Analyse der Model-of-data wird die durch Ruiz-Primo und Shavelson (1996) vorgeschlagene Kombination der individuellen Betrachtung und des Abgleichs mit einem Expert*innen-MOD verwendet. Dazu wurden im ersten Schritt Expert*innen-MOD für die Experimente erstellt. Diese wurden durch Forschende aus dem Bereich Chemiedidaktik in einem ersten Schritt individuell erstellt. Anschließend wurden die einzelnen MOD in ein zusammenfassendes MOD übernommen. Dazu wurden Ähnlichkeiten und Unterschiede der einzelnen MOD untersucht. Das so entstandene

Expert*innen-MOD wurde anschließend erneut mit der Gruppe der Expert*innen diskutiert. Um den erwartbaren Unterschieden zwischen dem Expert*innen-MOD und den MOD der Proband*innen Rechnung zu tragen, wurden die inhaltlichen Aspekte innerhalb der Expert*innen-Lösung differenziert. Abb. 14 zeigt das Expert*innen-MOD zum Versuch „Siedepunkt von Wasser“. Die orange eingefärbten Aspekte zeigen Beobachtungsaspekte aus dem Experimentiervideo auf, grün eingefärbte Aspekte stellen konzeptuelle Aspekte dar. Die konzeptuellen Aspekte können dabei sowohl als Schlussfolgerung aus dem MOD hergeleitet oder als bestehendes Wissen darin integriert werden. Zusätzlich werden Aspekte, die durch eine unmögliche kausale Verknüpfung falsifiziert werden, hervorgehoben. Aspekte mit einem roten Rahmen stellen sog. essentielle Aspekte dar, welche zur Erklärung des Experimentausgangs notwendig sind. Diese wurden festgelegt, um den erwartbaren unterschiedlichen Umfang der Proband*innen-MOD im Vergleich zur Expert*innen-Lösung zu berücksichtigen. Der Vergleich bzgl. inhaltlicher Aspekte fokussiert somit auf die essentiellen Aspekte. Diese wurden zu Argumentationslinien zusammengefasst und unabhängig von den verwendeten Verknüpfungsarten und deren Adäquatheit kodiert. Die jeweiligen Argumentationslinien sind im Kodiermanual im Anhang zu finden. Aspekte, welche über die essentiellen Argumentationslinien hinausgehen, werden als weitere Aspekte unterschieden zwischen in-/adäquaten Beobachtungs-, oder Konzeptaspekten kodiert. Wie in Kapitel 7.1.2 dargelegt, spielen epistemologische Überzeugungen für die Konstruktion der unterschiedlichen Verknüpfungsarten eine wichtige Rolle. Entsprechend werden die epistemologischen Aspekte über die Kodierung der Verknüpfungsarten sowie deren Adäquatheit berücksichtigt.

Im nächsten Schritt wurden die Verknüpfungsarten untersucht. Die individuell von den Proband*innen erstellten Verknüpfungen wurden bzgl. ihrer Adäquatheit bewertet. Dazu wurde der Einsatz der Verknüpfungen mit der theoretischen Beschreibung der Verknüpfungsarten abgeglichen. Sofern die Verknüpfungen adäquat verwendet wurden, wurde diese bepunktet. Dabei unterscheidet sich die vergebene Punktzahl der unterschiedlichen Verknüpfungsarten untereinander. Die unterschiedliche Bepunktung wurde dabei auf Grundlage der Komplexität der Verknüpfungsarten begründet. Entsprechend erhielten die Proband*innen für adäquat gebildete kausale und kontrastierende Verknüpfungen einen Punkt. Diese beiden Verknüpfungsarten können rein auf Ebene der Beobachtungsaspekte gebildet werden und stellen daher die geringsten Anforderungen an die Proband*innen. Für adäquat gebildete induktive- und Analogie-Verknüpfungen erhalten die Proband*innen je zwei Punkte. Für diese Verknüpfungsarten muss mindestens ein Konzeptaspekt verknüpft werden. Somit werden weitere kognitive Prozesse (z.B. Dekontextualisierung/Enkodierung) zur Bildung dieser Verknüpfungen vorausgesetzt. Adäquat verwendete unmögliche kausale Verknüpfungen werden mit drei Punkten bepunktet, da diese die höchste Anforderung an die Proband*innen stellen. Für diese Verknüpfungsart muss ein Widerspruch zwischen unterschiedlichen Argumentationslinien/Aspekten über das gesamte MOD hinweg erkannt werden. Innerhalb der Bepunktung der Adäquatheit der einzelnen Verknüpfungsarten werden keine Graduierungen vorgenommen. Die Begründung einzelner Verknüpfungen kann sehr individuell und vielfältig sein. So fällt es nur aus der Analyse der graphischen MOD schwer Graduierungen in der Bepunktung zu unternehmen. Zusätzlich wurden kausale Verknüpfungen kodiert, welche innerhalb eines Aspekt-Kästchens dargestellt wurden. In diesem Fall werden kausale Verknüpfungen nicht graphisch durch die Nutzung des entsprechenden Pfeils, sondern als Text innerhalb eines Aspekts dargestellt. Durch bestehende Vorarbeiten (Pollmeier, 2019) wurde angenommen, dass derartige kausale Verknüpfungen vorkommen könnten und somit berücksichtigt werden sollten.

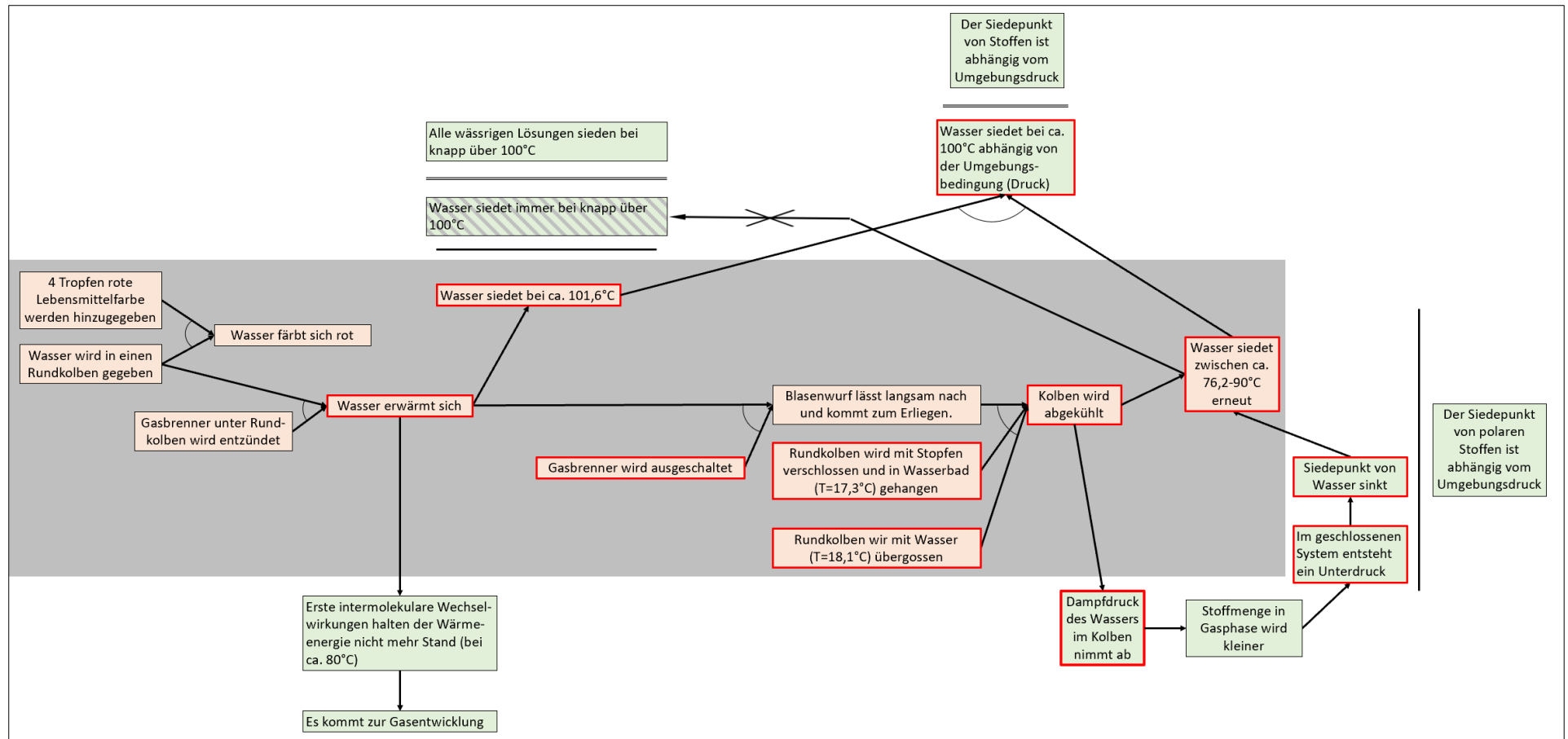


Abbildung 14 - Expert*innen Model-of-data zu Experiment 2, Beobachtungsaspekte (orange Kästen), Konzeptaspekte (grüne Kästen), essentielle Aspekte (rot umrandete Kästen), durch unmögliche kausale Verknüpfungen unmögliche Aspekte (schraffierte Kästen)

Neben dem Vorhandensein essentieller/weiterer Aspekte sowie der Adäquatheit der gebildeten Verknüpfungen wurde das Erkennen anomaler Daten kodiert. Hierzu wurde nach Widersprüchen innerhalb der Model-of-data gesucht. Diese müssen dabei nicht zwangsläufig mit gebildeten unmöglichen kausalen Verknüpfungen übereinstimmen und könnten so auch durch andere Verknüpfungsarten oder rein in der Formulierung von Aspekten erkannt werden. Innerhalb des Kodiermanuals wurden die in Kap. 7.1.1 beschriebenen und auf Grundlage der beschriebenen Vorstellungen erwarteten anomalen Beobachtungen deduktiv eingesetzt. Gleichzeitig wurden induktiv weitere beschriebene anomale Beobachtungen aufgenommen, welche von den Proband*innen als solche wahrgenommen wurden. Durch die Offenheit für weitere anomale Beobachtungen der Proband*innen soll der Individualität möglicherweise entstehender kognitiver Konflikte auf Basis der konzeptuellen Vorstellungen Rechnung getragen werden. Die hier vergebenen Codings wurden nicht bepunktet, sondern als zusätzliche Kategorien notiert.

Aus der Anlage des Kodiermanuals bieten sich unterschiedliche Ansätze zur Quantifizierung der Ergebnisse. Abbildung 15 gibt einen Überblick über die bestimmten Parameter sowie gebildeten quantitativen Maße. In Bezug auf die kodierten essentiellen/weiteren Aspekte kann eine Ratio für das Verhältnis von Beobachtungs- zu Konzeptaspekten gebildet werden. Diese Ratio gibt getrennt für essentielle Argumentationslinien und weitere Aspekte an, ob es einen Schwerpunkt auf einer Art der Aspekte gibt. Für die Untersuchung der genutzten Verknüpfungsarten kann eine prozentuale Adäquatheit angegeben werden, indem die Zahl der adäquaten Verknüpfungen mit der Gesamtzahl der Verknüpfungen im MOD verrechnet wird. Ebenso kann ein Verknüpfungsscore gebildet werden, welcher auf der Summe aller Punkte der Verknüpfungen im Verhältnis zur Gesamtzahl aller Verknüpfungen gebildet wird. Dieser Score kann Aufschluss darüber geben, inwiefern auch komplexere Verknüpfungsarten (s.o.) genutzt wurden.

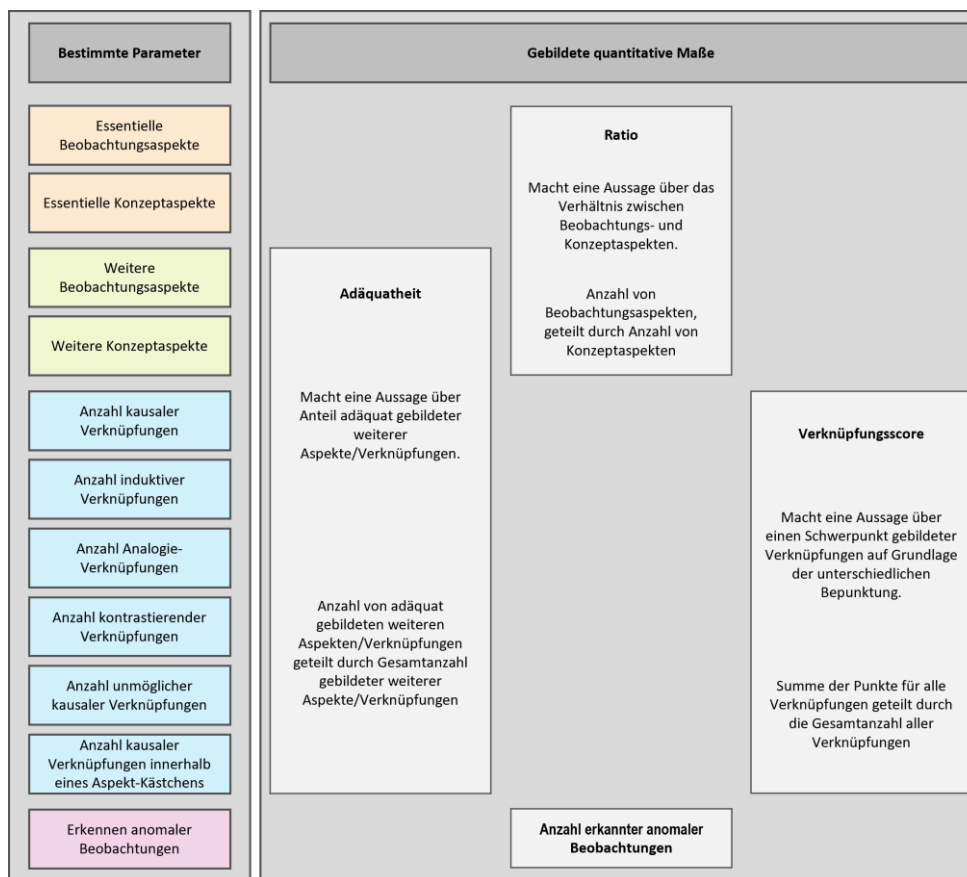


Abbildung 15 - Überblick über bestimmte Parameter sowie gebildete quantitative Maße zur Auswertung der MOD

Wie in Kap. 7.3 beschrieben, wurde die Phase des lauten Denkens (inkl. der Erstellung der MOD) sowie das anschließende Interview als Screencast video- und audiographiert. Um die Abweichungen zwischen dem graphisch dargestellten MOD sowie den Verbalisierungen bzgl. des mentalen Modells genauer untersuchen zu können, wurden diese in das vorhandene Kodiermanual integriert. Dazu wurde den bestehenden Kategorien zur Analyse der graphischen MOD jeweils die Kategorie „Abweichung“ sowie „Formuliert, aber nicht geschrieben“ hinzugefügt. So können sowohl Abweichungen von der graphischen Darstellung als auch verbalisierte Ergänzungen im Material kodiert werden.

7.5.2 Güte der Datenauswertung

Insgesamt sollen Validitätsargumente im Sinne des *Model of argument* (Toulmin, 2003) zur Begründung einer hohen Güte der Auswertung angeführt werden (Kane, 2016). Ein Schwerpunkt auf qualitativen Forschungsmethoden macht die Betrachtung der Güte in einem größeren Zusammenhang notwendig. Dies meint, dass die reine Fixierung auf quantitative Gütekriterien mit Bezug auf die tatsächlich vorliegenden Daten überdacht werden sollte (Frambach et al., 2013). Dabei soll die Triangulation mit Ergebnissen aus den quantitativen Instrumenten zur Steigerung der Güte beitragen. Entsprechend soll im folgenden Kapitel die Güte der Datenauswertung der MOD sowie der Videodaten beschrieben werden.

Um die Güte der Datenauswertung zu steigern, wurden unterschiedliche Qualitätssicherungsschritte implementiert.

Das entwickelte Kodiermanual wurde im ersten Schritt an Daten aus einer Pilotstudie ($N=3$) erprobt. Die Kodierung erfolgte durch zwei Kodierende entlang des in Kap. 7.3 beschriebenen Verfahrens. Es konnten einerseits Unklarheiten innerhalb des Kodiermanuals aufgedeckt werden, andererseits konnten die Kodierenden die Analyse trainieren. Aus der gemeinsamen Diskussion unklarer Kodierungen haben sich sowohl Konkretisierungen des Kodiermanuals (z.B. differenziertere Ankerbeispiele, etc.) als auch die induktive Ergänzung einzelner Kategorien ergeben. Im Sinne semantischer Gültigkeit konnte die Güte des Kodiermanuals so gesteigert werden (Mayring, 2010). Die enge Ausrichtung des Kodiermanuals an dem Rahmenkonzept der *Model-of-data* (MOD) (Chinn & Brewer, 2001) kann als Beitrag zur Konstruktvalidität verstanden werden (Döring & Bortz, 2016). Vor allem die Diskussion über einzelne Kategorien sowie deren Ausschärfung kann zur Steigerung der Konstruktvalidität beitragen.

Eine weitere, verbreitete Möglichkeit, die Güte der Datenauswertung zu überprüfen, stellt die Intercoder-Übereinstimmung dar (Döring & Bortz, 2016; Mayring, 2010). Die Doppelkodierung eines Anteils der Daten (Döring und Bortz (2016) empfehlen 10-20%) kann Aufschluss über den Einfluss der Kodierenden auf die Ergebnisse sowie die Validität des Kodiermanuals geben. Im Sinne des Einflusses der Kodierenden könnten Unterschiede im individuellen Zugang auf die Daten aufgedeckt werden, was zu einer Steigerung der Objektivität führen könnte. Die Validität des Kodiermanuals kann insofern überprüft werden, dass größere Abweichungen zwischen den Kodierenden auch auf unscharfe Kodierregeln zurückführbar sein könnten. So könnten Schwachstellen im Manual entdeckt und überarbeitet werden.

Für die Hauptstudie wurden unterschiedliche Intercoder-Übereinstimmungen untersucht:

- Übereinstimmung der Kodierung der gebildeten Verknüpfungen sowie deren Adäquatheit
- Übereinstimmung gebildeter essentieller/weiterer sowie inadäquater Beobachtungs- und Konzeptaspekte
- Übereinstimmung bei Abweichungen zwischen MOD und lautem Denken

- Übereinstimmung bei im lauten Denken verbalisierten, aber im MOD nicht dargestellten Aspekten/Verknüpfungen

In der vorliegenden Studie wurden insgesamt 6 Datensätze ($\approx 29\%$ der ausgewählten Stichprobe) doppelt kodiert. Die beiden Kodierenden haben dabei das in Kap. 7.3 dargestellte Verfahren unabhängig voneinander durchlaufen. Entsprechend wurde immer ein MOD und anschließend das entsprechende Video des lauten Denkens kodiert. Anschließend wurden diese Kodierungen verglichen. Abweichungen wurden gemeinsam diskutiert und ggf. ein Konsens gefunden.

Innerhalb der graphischen MOD wurde eine Übereinstimmung auf Grundlage der kodierten Bildausschnitte mittels MAXQDA bestimmt. Innerhalb der importierten Bilddateien konnten dazu Ausschnitte des Bildes für einzelne Kodierungen gewählt werden. Das Programm MAXQDA prüft dann für die Übereinstimmung der Kodierungen, ob gleich große Kodierungen an den gleichen Stellen für beide Kodierenden vorhanden sind. Unterschiede in der Größe der Ausschnitte und damit der Kodierung innerhalb der Bilddatei wirken sich so auf die Intercoder-Übereinstimmung im Sinne des Kappa-Koeffizienten aus. So könnte ein um wenige Pixel größer ausgewählter Kodierausschnitt innerhalb der Bilddateien zu durch MAXQDA erkannten Unterschieden führen, welche jedoch keine inhaltliche Abweichung darstellen. Die unterschiedlichen Kodierausschnitte gehen nicht zwingend mit dem Einbezug weiterer Verknüpfungen/Aspekte in die Kodierung einher. Es kann auch „inhaltsleerer“ Raum um die abgezielten Bestandteile des MOD einbezogen werden. Im Rahmen der Übereinstimmungsberechnung können sich diese Unterschiede auswirken. Um die erwarteten Abweichungen in der Größe der kodierten Bildausschnitte in der Analyse zu berücksichtigen, wurde eine notwendige Überlappung von Kodierungen von 50% festgelegt, um als „Übereinstimmung“ gewertet zu werden.

Für die Kodierung der Videos wurden durch die Kodierenden einzelne Szenen der Videos kodiert. Als In- und Out-Points wurde auf Grundlage eines *event-based* Ansatzes die Beschreibung von Abweichungen zum MOD gewählt. Als Übereinstimmungen wurden Kodierungen der beiden Kodierenden gewertet, welche eine 70%-ige Überlappung zeigen. Dadurch kann die unterschiedliche Setzung der In- und Out-Points, welche für die Kodierung der Abweichungen ggf. nicht inhaltstragend sind, ausgeglichen werden. Um sicherzustellen, dass sich keinerlei inhaltstragende Unterschiede durch leicht unterschiedliche In- und Out-Points ergeben, wurden alle kodierten Stellen der Intercoder-Fälle gemeinsam diskutiert. Dazu wurden die einzelnen Events als Grundlage herangezogen. Die unterschiedliche Setzung von In- und Out-Points bezogen auf die Zeitangaben wurden in diesem Zusammenhang nicht diskutiert.

Zur Einschätzung der Intercoder-Übereinstimmung kann neben der prozentualen Übereinstimmung der Kappa-Koeffizient herangezogen werden (Döring & Bortz, 2016). Der Kappa-Koeffizient ist eine um den Zufall bereinigtes Maß zur Schätzung der Intercoder-Reliabilität. Die Zufallsübereinstimmung nimmt dabei mit steigender Kategorienzahl und Komplexität des Manuals ab (Rädiker & Kuckartz, 2019). In der vorliegenden Teilstudie wurde die Intercoder-Übereinstimmung getrennt für die beiden unterschiedlichen Datensätze (MOD und Videos des lauten Denkens) sowie Kategorien bestimmt. Dabei wurde sowohl die prozentuale Übereinstimmung als auch Cohen's Kappa genutzt.

Übereinstimmung für die gebildeten Verknüpfungen und deren Adäquatheit

Innerhalb der MOD wurden die unterschiedlichen Verknüpfungen (auch im Hinblick auf deren Adäquatheit) kodiert. Auf Grundlage einer notwendigen Überlappung von 50% der Ausschnitte

der Kodierenden, kann ein Gesamt-Kappa-Koeffizient von $\kappa = .62$ ($N_{\text{MOD}}=12^{14}$) bestimmt werden. Nach Landis und Koch (1977) kann die Übereinstimmung als substantiell beschrieben werden. Aufgrund der Bestimmung der Übereinstimmung auf Grundlage der gesetzten Bildausschnitte kann angenommen werden, dass der Kappa-Koeffizient die tatsächliche InterCoder-Übereinstimmung in diesem Fall eher unterschätzt.

Tabelle 32 gibt einen Überblick über die Kodierung der Verknüpfungen innerhalb der MOD. Für diese ergibt sich eine hohe (prozentuale) Übereinstimmung von 76.14% auf Grundlage der Kodierung der jeweiligen Verknüpfungsart. Eine Berechnung des Kappa-Koeffizienten limitiert auf die Kodierung der Verknüpfungen erreicht einen Wert von $\kappa = .72$ ($N=12$), welcher nach Landis und Koch (1977) ebenfalls die Benchmark „substantiell“ erreicht. Die Stichprobe besteht dabei aus 6 Proband*innen, von denen jeweils 2 MOD in die Auswertung einbezogen wurden.

Tabelle 32 – Vergabe von Kodierungen sowie InterCoder-Übereinstimmung bzgl. Verknüpfungen innerhalb der Model-of-data

Art der Verknüpfung	Kodierende* r 1 Gesamtanzahl (davon adäquat)	Kodierende* r 2 Gesamtanzahl (davon adäquat)	Prozentuale Übereinstimmung	Cohen's Kappa
Kausale Verknüpfungen (insgesamt)	142 (66)	142 (69)		
Beobachtung – Beobachtung	131 (62)	132 (65)	77.47%	.71
Konzeptwissen – Konzeptwissen	0 (-)	0 (-)	-	-
Beobachtung – Konzeptwissen	11 (4)	10 (4)	63.64%	.53
Induktive Verknüpfungen	9 (3)	9 (2)	77.78%	.71
Analogie-Verknüpfungen	15 (2)	15 (2)	80%	.74
Kontrastierende Verknüpfungen	18 (14)	18 (12)	72.2%	.64
Unmögliche kausale Verknüpfungen	7 (5)	7 (5)	71.43%	.63
Gesamt	191 (90)	191 (90)	76.14%	.74

$N_{\text{MOD}}=12$

Übereinstimmung für die gebildeten essentiellen und weiteren Beobachtungs- und Konzeptaspekte

Nun soll die Übereinstimmung bzgl. der Kodierung essentieller und weiterer Beobachtungs- und Konzeptaspekten sowie inadäquater Aspekte betrachtet werden. Dabei entspringen die essentiellen Aspekte aus den Argumentationslinien des Expert*innen-MOD, während weitere Aspekte individuell durch die Proband*innen gebildet wurden. Bei der Betrachtung der Anzahl der Kodierung der inhaltlichen Aspekte fällt bereits auf, dass hier größere Unterschiede zwischen den

¹⁴ Die Stichprobengröße für den InterCoder-Vergleich umfasst $N=12$. Es handelt sich dabei um je 2 Model-of-data der 6 Proband*innen.

Kodierenden erkannt werden können. Der Kappa-Koeffizient exklusiv für die Kodierungen der Aspekte erreicht einen Wert von $\kappa = .44$ ($N_{\text{MOD}}=12$). Tabelle 36 gibt einen Überblick der Kodierung der einzelnen Aspekt-Kategorien. Quantitative Unterschiede fallen vor allem im Bereich der essentiellen/weiteren Beobachtungsaspekte auf. Eine Begründung dafür kann vor allem in der Definition der Kodiereinheit erkannt werden. Wie in Kapitel 7.5.1 beschrieben, wurden relevante inhaltliche Aspekte in essentielle Argumentationslinien (hier: essentielle Beobachtungs-/Konzeptaspekte) zusammengefasst. Diese essentiellen Aspekte können dabei mehrere Aspekt-Kästchen umfassen. Dies kann in zwei Arten Einfluss auf die Übereinstimmung nehmen: Einerseits könnten durch die individuelle Zusammenfassung von Aspekten zu essentiellen Argumentationslinien Unterschiede in der Kodierung auftreten. So könnten teils Aspekte mit zu einer Argumentationslinie hinzugezählt werden, während der*die jeweils andere Kodierende diese Aspekte vom essentiellen Aspekt ausschließt und somit einzeln als weiteren Aspekt kodiert. Dadurch könnten sich Unterschiede in der Anzahl der jeweils vergebenen Kodierungen erklären lassen. Andererseits könnte die unterschiedliche Zusammenfassung zu essentiellen Aspekten direkten Einfluss auf den bestimmten Kappa-Koeffizienten nehmen. Hier ist erneut ein Bezug zu den ausgewählten Bildausschnitten für die jeweiligen Kodierungen zu erkennen. Abweichungen in der Größe der Bildausschnitte im Programm MAXQDA können sich direkt auf die Übereinstimmung auswirken. Dies wäre ein möglicher Erklärungsansatz für den geringeren Wert des Kappa-Koeffizienten. Weiterhin zeigen sich teils kritische Werte für die Intercoder-Übereinstimmung durch Cohen's Kappa (vgl. Tab. 33) für einzelne Kategorien. Trotz des umfangreichen Trainings der Kodierenden sowie Diskussionen darüber konnte nicht in allen Kategorien eine angemessene Intercoder-Übereinstimmung erreicht werden. Dies könnte durch die Komplexität der MOD sowie des Kodiermanuals erklärt werden. Die Zusammenhänge innerhalb der MOD sind individuell teils sehr unterschiedlich ausgeprägt. Das umfassende Kodiermanual soll hier für alle Zusammenhänge passende Kategorien bereitstellen. Die Komplexität des Manuals schlägt sich dabei auch in den Anforderungen an die Kodierenden nieder. Für die Analyse müssen diese Herausforderungen in der Kodierung berücksichtigt werden.

Tabelle 33 – Vergabe von Kodierungen sowie Intercoder-Übereinstimmung bzgl. inhaltlicher Aspekte innerhalb der Model-of-data

Art der Aspekte	Kodierende*r 1	Kodierende*r 2	Prozentuale Übereinstimmung	Cohen's Kappa
Anzahl essentieller Beobachtungsaspekte	31	41	31.43%	.23
Anzahl essentieller Konzeptaspekte	3	2	80%	.64
Anzahl weiterer Beobachtungsaspekte	86	57	41.06%	.37
Anzahl weiterer Konzeptaspekte	12	6	25%	.17
Anzahl inadäquater Aspekte	11	12	43.48%	.27
Gesamt	143	118	47.93%	.44

Anmerkung. $N_{\text{MOD}}=12$

Zusammenfassend kann für die Übereinstimmung der gebildeten Verknüpfungen sowie der Aspekte ein durchmisches Fazit gezogen werden. Während die Übereinstimmung für die Verknüpfungen nach Landis und Koch (1977) substantiell ausfällt, kann für die gebildeten Aspekte nur eine moderate Übereinstimmung festgestellt werden. Unterschiede können sich zum Teil durch die Konzeption des Kodiermanuals, z.B. bzgl. der Festlegung der Kodiereinheit der essentiellen

Aspekte, erklären lassen. Gleichzeitig muss die hohe Komplexität der gebildeten MOD und die daraus entstehenden Herausforderungen an die Kodierenden berücksichtigt werden.

7.6 Ergebnisse und Diskussion

7.6.1 Analyse der inhaltlichen Aspekte

Die gebildeten inhaltlichen Aspekte können Einblicke in die getätigten Beobachtungen sowie das angewendete Konzeptwissen der Proband*innen geben. Dabei können die gebildeten inhaltlichen Aspekte in unterschiedliche Kategorien unterteilt werden. Hier werden zuerst die essentiellen inhaltlichen Aspekte berichtet, welche zur Erklärung der beobachteten Phänomene erforderlich sind. Anschließend sollen weitere Aspekte betrachtet werden. Diese stellen inhaltliche Aspekte dar, welche nicht zwingend zur Erklärung des Phänomens notwendig sein. Dabei können ebenfalls inadäquate Aspekte gebildet werden, welche gesondert betrachtet werden.

Tabelle 34 gibt einen Überblick über die Anzahl gebildeter essentieller und weiterer Beobachtungs-/Konzeptaspekte je erstelltem MOD. Im Vergleich werden mehr weitere Aspekte als Konzeptaspekte gebildet.

Tabelle 34 – Deskriptive Statistik für essentielle & weitere Beobachtungs-/Konzeptaspekte

Aspektart	N	Anzahl		Ratio	
		Minimum	Maximum	M	SD
Essentielle Beobachtungsaspekte	42	.00	6.00	2.50	1.47
Essentielle Konzeptaspekte	42	.00	3.00	.19	.63
Weitere Beobachtungsaspekte	42	.00	15.0	3.45	3.30
Weitere Konzeptaspekte	42	.00	7.00	.95	1.34

Anmerkung. Die Ratio wurde als Mittelwert der Ratios für die einzelnen MOD bestimmt. Dabei ist die Ratio für zahlreiche einzelne MOD gleich null, da Zähler oder Nenner des Quotienten gleich null sind. So entsteht ein kleiner Mittelwert, welcher von einer errechneten Ratio aus den Mittelwerten der unterschiedlichen Aspekt-Arten abweicht.

Das Verhältnis zwischen der Anzahl innerhalb der MOD gebildeter Beobachtungs- und Konzeptaspekte kann, wie in Kap. 7.5.1 beschrieben auch durch eine Ratio ausgedrückt werden. Dazu wurden die gebildeten Beobachtungsaspekte ausgezählt und durch die Anzahl der gebildeten Konzeptaspekte geteilt. Der Mittelwert für die jeweiligen Ratios aus Beobachtungs- und Konzeptaspekten kann in Tabelle 37 erkannt werden. Ein niedriger Mittelwert der Ratio würde auf eine große Anzahl von Konzept- im Vergleich zu Beobachtungsaspekten hindeuten. Um den niedrigen Mittelwert der Ratio essentieller Aspekte interpretieren zu können, ist jedoch die Betrachtung der absoluten Anzahlen der gebildeten Aspekte relevant (vgl. Tab. 34). Insgesamt wurden deutlich weniger Konzept- als Beobachtungsaspekte gebildet. Auf Ebene der einzelnen MOD ist die Ratio zahlreicher Fälle gleich 0, da keinerlei Konzeptaspekt gebildet wurde (z.B. MASO20CA_Säure-Base, I-ABD29MM_Siedepunkt, etc.). Für die Ratio weiterer Aspekte konnten teilweise keine weiteren Beobachtungsaspekte gebildet werden. Da die Ratio gleich 0 ist, sobald Zähler oder Nenner gleich 0 sind, ergibt sich eine Ratio, welche von einer aus den Mittelwerten (vgl. Tab. 34) der Aspekt-Arten berechneten Ratio abweicht. Folglich lässt sich der geringe Mittelwert der Ratio vor allem

aufgrund wenig gebildeter essentieller Konzeptaspekte sowie vereinzelt ausbleibender weiterer Beobachtungsaspekte begründen.

Die gebildete Ratio auf Grundlage der Gesamtzahl der MOD ($N=42$) muss entsprechend vorsichtig interpretiert werden. Während ein geringer Wert rechnerisch auf eine Überzahl von Konzeptaspekten hinweist, muss dieser in diesem Fall auf fehlende Konzeptaspekte zurückgeführt werden. So wurden nur in 4 von 42 Fällen (z.B. IABD29MM_Säure-Base, SEMS14LA_Säure-Base, etc.) essentielle Konzeptaspekte sowie in 19 Fällen (z.B. AAVR08LA_Siedepunkt, MARF16AN_Säure-Base, etc.) weitere Konzeptaspekte gebildet. Die Berechnung der jeweiligen Ratio auf Grundlage dieser Proband*innen zeigt andere Werte auf ($\text{Ratio}_{\text{essentielle Aspekte}}=1.875$, $N=4$, $SD=.74$; $\text{Ratio}_{\text{weitere Aspekte}}=3.293$, $N=19$, $SD=3.63$). Konzeptaspekte sind damit übergreifend unterrepräsentiert, vor allem wenn essentielle Konzeptaspekte betrachtet werden.

Tabelle 35 gibt einen Überblick über die gebildeten essentiellen Aspekte, getrennt nach Beobachtungs- und Konzeptaspekten. Innerhalb von Experiment 1 kann eine relativ gleichmäßige Bildung der verschiedenen (essentiellen) Beobachtungsaspekte erkannt werden. Alle Proband*innen haben mindestens einen Beobachtungsaspekt gebildet, im Mittel $M=2.52$ ($N=21$, $SD=.66$). Bei Experiment 2 wurden mit im Mittel $M=1.76$ ($N=21$, $SD=1.20$) weniger Beobachtungsaspekte durch die Proband*innen gebildet. Ebenso gibt es einzelne Proband*innen, welche keinerlei essentielle Beobachtungsaspekte gebildet haben. Eine Korrelationsanalyse zwischen der Anzahl aller gebildeter (essentieller) Aspekte in Experiment 1 und 2 der einzelnen Proband*innen zeigt keine Zusammenhänge ($r=-.310$, $p=.172$, $N=21$). Auch sind keine Proband*innen auszumachen, welche sowohl in Experiment 1 also auch Experiment 2 besonders viele Aspekte bilden, oder andersherum. Bei der Betrachtung der Gesamtzahl der gebildeten essentiellen Aspekte über beide Experimente hinweg sind keine großen Unterschiede erkennbar ($M=4.762$, $N=21$, $SD=1.269$).

Auf einer inhaltlichen Ebene kann für Experiment 1 festgehalten werden, dass essentielle Aspekte des Konzeptwissens zur Klärung des Phänomens für zahlreiche Proband*innen fehlen. So erkennen nur zwei Proband*innen das Ausbleiben einer Neutralisationsreaktion mit dem End-pH-Wert von 7 für Essigsäure IABD29MM, IAKS30JA). Lediglich SEMS14LA formuliert das Auftreten einer Neutralisationsreaktion mit dem End-pH-Wert=7 in Abhängigkeit der jeweiligen Säure innerhalb der MOD. SAGT10RL formuliert den Aspekt innerhalb des lauten Denkens. Das Unterscheidungskriterium der Säurestärke der jeweiligen Säuren wird von keiner*m Proband*in berichtet, sodass eine inhaltliche Erklärung des auftretenden Phänomens durch die Proband*innen nicht erfolgt. Auf Ebene der essentiellen Beobachtungsaspekte sind vereinzelt Abweichungen zu erkennen. Diese beruhen zumeist auf Grundlage der Farbwahrnehmung und der daraus resultierenden Einordnung in einen pH-Wert-Bereich.

Für Experiment 2 können bereits auf Ebene der Beobachtungsaspekte deutlichere Unterschiede erkannt werden. Nur rund 50% der Proband*innen beschreiben das Sieden von Wasser bei ca. 100 bzw. 101.6°C. Ähnliches gilt für das luftdichte Verschließen des Kolbens, welches nur von unter 50% der Proband*innen beschrieben wird.

Während das Abkühlen des Rundkolbens von vielen Proband*innen beobachtet wird, beschreibt lediglich ein*e Proband*in das Sieden des Wassers zwischen 76.2-90°C. Im Vergleich zu Experiment 1 sind hier entsprechend bereits auf Ebene der Beobachtungsaspekte deutlichere Unterschiede zu erkennen. Auf Ebene der essentiellen Konzeptaspekte kann lediglich CAJN05DN im Rahmen des lauten Denkens einen Konzeptaspekt formulieren. Dabei wird die Abhängigkeit des Siedepunktes vom Umgebungsdruck formuliert, wenngleich das Sieden bei erniedrigtem Umgebungsdruck auf Ebene der Beobachtungsaspekte nicht beschrieben wird.

Zusammenfassend kann für beide Experimente ein deutlicher Schwerpunkt auf Beobachtungs-Aspekten erkannt werden. Dabei kann innerhalb Experiment 2 nur ein einziger Konzeptaspekt erkannt werden.

Tabelle 35 – Erstellte essentielle Aspekte der Proband*innen

Essentielle Aspekte			Fälle																				
			MASO20CA	RYER27LA	IABD29MM	KNRN23CA	AEDL02JN	INCH29FE	IAKS30JA	AADK15LE	SETS14KA	BAVR30PA	SAGT10RL	MEAM18FA	BTRE19GA	GAWI21JA	MERF13KN	AASA23CA	SEMS14LA	AAVR08LA	MARF16AN	CAJN05DN	MNAR08NE
Experiment 1: Säure-Base	Beobach-	Zugabe & Rotkohllindikator + pH <7	X	X	X	X	X	X		X	X	X	X	X	X		X	X	X	X	X	X	X
		Zugabe NaOH färbt HCl & HNO ₃ bläulich - neutral	X	X	X		X		X	X	X	X	X		X	X	X	X	X		X	X	X
		Zugabe NaOH färbt Essigsäure grün - basisch	X	X	X		X		X	X	X	X	X	X	X	X	X	X		X		X	X
	Konzeptaspekte	Bei HCl & HNO ₃ hat eine Neutralisation mit pH=7 stattgefunden.			X				X										X			X	
		Bei Essigsäure hat keine Neutralisation mit pH=7 stattgefunden.			X				X														
		Protonenübertragungsreaktion																					
		Nicht bei allen Säuren gleiche Protonenkonzentration																					
		Definition starke/schwache Säure (Dissoziationsgrad)																					
		Essigsäure ist schwache Säure, HCl & HNO ₃ sind starke Säuren																					
		NaOH ist eine starke Base.																					
	Ob Neutralisation (pH=7) abläuft, hängt von Säure ab.											X						X					
Experiment 2: Siedepunkt von Wasser	Beobachtungsaspekte	Wassertemperatur steigt, bis es bei ca. 100/101,6°C siedet.			X	X	X	X		X		X		X		X	X	X				X	
		Gasbrenner wird ausgeschaltet, Kolben luftdicht verschlossen.				X	X	X	X	X		X		X		X	X	X					
		Rundkolben wird abgekühlt.	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X		X	X	X
		Wasser siedet zwischen 76.2-90°C.						X															
	Konzept-	Dampfdruck des Wassers sinkt, Unterdruck entsteht.																					
		Siedepunkt von Wasser sinkt.																					
		Der Siedepunkt von Wasser ist abhängig vom Umgebungsdruck.																				X	
Summe:			4	4	7	4	6	5	6	6	4	6	4	5	3	6	5	6	4	2	3	6	4

Anmerkung. Schwarze Kreuze markieren innerhalb der MOD dargestellte Aspekte, rote Kreuze innerhalb des lauten Denkens ergänzte Aspekte

Wie in Tabelle 35 erkannt werden kann, wurden wesentlich mehr weitere Aspekte im Vergleich zu essentiellen Aspekten gebildet. Für tiefergehende Einblicke in die weiteren gebildeten Aspekte (also keine essentiellen Aspekte), wurden die Kategorien der weiteren Aspekte in einem zweiten Durchgang inhaltlich analysiert. Dazu wurden die zu sortierten Kodierungen inhaltlich verglichen und induktiv zu weiteren Unterkategorien zusammengefasst. Tabelle 40 stellt die jeweiligen Aspekte inkl. der Anzahl ihrer Kodierungen dar. Für die weiteren Beobachtungsaspekte der beiden Experimente fallen zwei Ähnlichkeiten auf: Die häufigsten Aspekte spiegeln jeweils eine zentrale Messvariable der Experimente wider. Einmal den pH-Wert über die Indikatorfärbung und einmal die Wassertemperatur. Beide Variablen werden in den Experimentiervideos als zentrale Variablen in den Fokus gestellt. An verschiedenen Stellen (im Falle der Wassertemperatur sogar kontinuierlich) werden Messwerte der jeweiligen Variablen erhoben. Eben diese Variablen stellen nun den größten Anteil weiterer Beobachtungsaspekte dar. Die zweite Ähnlichkeit kann in der Beschreibung der Experimentierschritte erkannt werden. Die Proband*innen erstellen einige Aspekte, die den Experimentierprozess beschreiben. Im Falle des zweiten Experiments werden auch Aspekte des Versuchsaufbaus geschildert. Weitere Beobachtungsaspekte für die beiden Experimente setzen sich mit den zentralen Betrachtungsaspekten auseinander. Im Falle von Experiment 1 werden Aspekte zur Veränderung der Färbung des Indikators wie auch Eigenschaften der Säuren vor Beginn des Experiments beschrieben. Für Experiment 2 wird vor allem das Verdampfen und Kondensieren von Wasser hervorgehoben. Daran anschließend werden vereinzelt Aspekte gebildet, welche sich je nach Proband*in deutlich unterscheiden können.

Ähnliche Tendenzen wie bei der Bildung weiterer Beobachtungsaspekte sind auch bei der Bildung inhaltlich inadäquater Beobachtungsaspekte zu erkennen (Tab. 36). Während in Experiment 1 vor allem inadäquate Zuordnungen eines pH-Werts zur entsprechenden Indikatorfärbung auftreten, wird in Experiment 2 der Siedepunkt bzw. der Abkühlungsprozess thematisiert. Die jeweiligen inadäquaten Aspekte entstammen teilweise dem Hintergrundwissen oder werden als Verallgemeinerungen in die MOD eingebracht.

Im Falle der weiteren Konzeptaspekte sticht bei Experiment 1 erhöhtes Hintergrundwissen zum pH-Wert, wie auch der Färbung des Indikators hervor (Tab. 36). Auch scheinen grundlegende Kenntnisse über den Neutralisationsprozess und die Rolle von Säuren und Basen darin vorhanden zu sein. Für Experiment 2 werden vor allem Verallgemeinerungen auf andere Stoffe oder auch Begründungen für das Abkühlen des Wassers im Rundkolben gebildet.

Tabelle 36 – Unterkategorien für weitere/inadäquate Beobachtungs-/Konzeptaspekte

Weitere (adäquate) Beobachtungsaspekte				Weitere (adäquate) Konzeptaspekte			
Experiment 1: Säure-Base		Experiment 2: Siedepunkt von Wasser		Experiment 1: Säure-Base		Experiment 2: Siedepunkt von Wasser	
Kategorie	N _{Kodier.}	Kategorie	N _{Kodier.}	Kategorie	N _{Kodier.}	Kategorie	N _{Kodier.}
Farbveränderung Indikator/Veränderung pH-Wert	20	Temperaturverlauf	30	Neutralisation/“basisch werden“ von Säuren durch Natronlauge	5	Prozess (des Experiments) läuft so bei allen Flüssigkeiten ab	2
Experimentierschritte	16	Experimentierschritte	29	Benennung des pH-Werts vor Beobachtung der Indikatorfärbung	5	Reduzierung der Umgebungstemperatur führt zu Abkühlung	2
Farbschichtung der Lösung nach Zugabe von Natronlauge	12	Versuchsaufbau	6	Rotfärbung mit Rotkohllindikator inkl. Kategorisierung der Säuren als „sauer“	4	Definition Siedepunkt über Aggregatzustandsveränderung	1
Beschreibung der Eigenschaften der Säuren zu Beginn des Experiments	4	Kondensation von Wasser	4	Abhängigkeit des pH-Werts von der jeweiligen Säure	3	Siedepunkt von Wasser zwischen 79-90°C	1
Unterschied in pH-Wert/Indikatorfärbung	3	Wasserdampf entsteht	4	Trennung des Gemisches nicht möglich	3	Aggregatzustandsveränderungen durch Gasbrenner	1
Nach Zugabe von Natronlauge entsteht keine Säure	3	Sieden von Wasser	3	Homogenes Gemisch entsteht	2	Definition Siedepunkt über konstante Temperatur	1
Beschreibung des Gemisches als homogen	2	Veränderung der Aggregatzustände durch Erhitzen	1	Indikator als Möglichkeit der pH-Wert-Bestimmung	1	Nicht spezifischer Vorwissensaspekt	1
Sortierung pH-Wert zu Indikatorfärbung	2	Weniger Wasserdampf im Kolben	1	pH=7 → neutral	1		
Herstellung Rotkohllindikator	1	Siedepunktabhängigkeit zugegebener Farbstoff	1	Grünfärbung des Indikators mit allen Flüssigkeiten weist auf basischen pH-Wert hin	1		
Probleme Einordnung der Färbung des Rotkohllindikators	1			Höherer pH-Wert von Essigsäure im Vergleich zu Salz- & Salpetersäure	1		
Es liegt keine neutrale Lösung vor	1			Zusammenhang des pH-Wert vor und nach Zugabe von Natronlauge	1		
				pH-Wert von Säuren allgemein <7	1		
				Gemische aus Natronlauge & Säure haben unterschiedliche pH-Werte	1		
				Heterogenes Stoffgemisch	1		

Inhaltlich inadäquate Beobachtungsaspekte				Inhaltlich inadäquate Konzeptaspekte			
Experiment 1: Säure-Base		Experiment 2: Siedepunkt von Wasser		Experiment 1: Säure-Base		Experiment 2: Siedepunkt von Wasser	
Kategorie	N _{Kodier.}	Kategorie	N _{Kodier.}	Kategorie	N _{Kodier.}	Kategorie	N _{Kodier.}
Beschreibung einer inadäquaten Färbung des Indikators/pH-Wert	4	Unterscheidung von Sieden/Kochen	2	Alle Säuren (+ Indikator) werden durch Natronlauge neutralisiert	3	Siedepunkt aller Stoffe liegt bei 100/101°C	6
Inadäquate Einheit	3	Zugabe von Kühlwasser	1	Natronlauge färbt alle Stoffgemische blau	2	Siedepunkt von Wasser liegt bei 100°C	5
Inadäquate Sortierung der Färbung des Indikators zum pH-Wert	2	Weiterer Temperaturanstieg nach Siedepunkt	1	Zugabe von Rotkohllindikator führt zu Neutralisation	2	Wasser liegt bei jeder Temperatur im gleichen Aggregatzustand vor	1
Alle Lösungen mit Natronlauge haben einen pH-Wert=7	1	Abkühlung von noch kochendem Wasser	1	Heterogenes Stoffgemisch	1	Kondensation von Wasserdampf findet immer unter 100°C statt	1
Unverständlicher Aspekt	1	Siedepunkt bei 100°C	1	Base und Säure neutralisieren sich immer einander	1	Der Siedepunkt ist abhängig von der Umgebungstemperatur	1
		Beobachtung von max. 100°C	1	Bei Zugabe von Natronlauge in Säure entsteht immer der gleiche pH-Wert	1	Bei Reduzierung der Umgebungstemperatur dauert es länger, bis Wasser siedet	1
				Säuren mit höheren pH-Wert können durch Natronlauge nicht neutralisiert werden	1		
				Bei unterschiedlichen Säuren laufen mit Natronlauge unterschiedliche (chemische) Reaktionen ab	1		

Neben der Betrachtung essentieller und weiterer Aspekte wurden auch inhaltlich inadäquate Aspekte kodiert. Im Mittel wurden $M=1.10$ ($N=42$, $SD=1.36$) inadäquate Aspekte pro MOD gebildet. Neben einigen (17 von 42) MOD ohne inhaltlich inadäquate Aspekte (z.B. AADK15LA_Säure-Base, BTRE19GA_Siedepunkt, etc.), erreicht MEAN18FA_Säure-Base mit 6 inadäquaten Aspekten das Maximum. Im Bereich der inhaltlich inadäquaten Konzeptaspekte sind vor allem Verallgemeinerungsaspekte erkennbar. Unabhängig von den beiden Experimenten wird das Verhalten einzelner Bestandteile im Experiment auf weitere, durchaus größere Zusammenhänge verallgemeinert.

Zusammenfassend lassen sich innerhalb der gebildeten weiteren/inadäquaten Aspekte gewisse Schwerpunkte ausmachen, welche (meist) auf die relevanten Betrachtungsaspekte der jeweiligen Experimente abheben. Gleichzeitig wird deutlich, dass die Unterscheidung weiterer Aspekte in Beobachtungs- und Konzeptaspekte nicht immer zweifellos möglich ist. So fällt bspw. auf, dass die Beschreibung eines Gemisches als homogen im Bereich der Beobachtungsaspekte sowie im Bereich der Konzeptaspekte verortet ist. Die jeweiligen Einschätzungen durch die beiden Kodierenden basieren nicht allein auf den Aspekten, sondern beziehen teils auch die Struktur des Model-of-data mit ein. So kann es für die Unterscheidung von Beobachtungs- und Konzeptaspekten einen Unterschied machen, in welcher Reihenfolge die Aspekte miteinander verknüpft sind. Zu einem gewissen Anteil obliegt die Entscheidung der Sortierung der jeweiligen Aspekte letztlich jedoch der Interpretation der Kodierenden, ob es sich bei dem jeweiligen Aspekt um Hintergrundwissen oder Beobachtungsschilderungen der Proband*innen handelt.

7.6.2 Beschreibung anomaler Beobachtungen

Einen Sonderfall von dargestellten Beobachtungs- oder Konzeptaspekten stellen Aspekte dar, welche anomale Beobachtungen beschreiben. In Kapitel 7.1.1 wurden die beiden Experimente inkl. der erwarteten Beobachtungen vorgestellt. Tabelle 37 gibt einen Überblick über die gebildeten Aspekte, welche anomale Beobachtungen darstellen. Die blauen Markierungen zeigen dabei die auf Grundlage der Theorie erwarteten anomalen Beobachtungen auf, während die orangenen Markierungen Beobachtungen darstellen, welche lediglich im lauten Denken beschrieben wurden. Dabei können derartige Beobachtungen nur für Experiment 2 beobachtet werden.

Tabelle 37 – Darstellung anomaler Beobachtungen zu Experiment 1 und 2

	Beschriebene anomale Beobachtungen	Anzahl
Experiment 1	Grünfärbung bei Reaktion von Salz- & Salpetersäure mit Natronlauge, andere Färbung bei Reaktion von Essigsäure mit Natronlauge.	15
	Unterschiedliche Indikatorfärbung (und damit pH-Wert) aller drei Säuren nach Zugabe von Natronlauge.	6
	Unterschiedliche pH-Werte der drei Säuren zu Beginn des Experiments (durch Indikatorfärbung).	1
	Zweiter Siedepunkt wird als „Blubbern/Brodeln“ beschrieben.	3
Experiment 2	Zweiter Siedepunkt von Wasser bei unter 100°C wird dargestellt.	1/1
	Wasser blubbert beim Abkühlen bei ca. 80°C.	2
	Wasser blubbert, kocht aber nicht.	1
	Verwirrung, warum Wasserdampf nicht auch rot ist.	1

Anmerkung. Orange hinterlegte Beobachtungen, bzw. orange dargestellte Anzahl von Beobachtungen stammt aus dem lauten Denken, blaue Markierungen stellen theoretisch erwartete anomale Beobachtungen dar.

Innerhalb von Experiment 1 sind in allen MOD Beschreibungen anomaler Beobachtungen erkennbar. Die Mehrheit der Proband*innen beschreibt dabei die beschriebenen erwarteten Beobachtungen (z.B. AADK15LE_Säure-Base, AASA23CA_Säure-Base, IABD29MM_Säure-Base, etc.). Dabei erkennen die Proband*innen einen Unterschied in der Färbung von Salz- und Salpetersäure im Vergleich zu Essigsäure, jeweils nach Zugabe der Natronlauge. Sieben weitere Proband*innen erkennen Unterschiede zwischen allen drei Lösungen nach Zugabe der Natronlauge (z.B. AAVR08LA_Säure-Base, INCH29FE_Säure-Base, MNAR08NE_Säure-Base, etc.). Dabei werden die Lösungen der Salz- und Salpetersäure zumeist relativ nah zueinander bzgl. des pH-Werts kategorisiert (z.B. eine Säure pH=7, die andere pH=7-8), während für die Lösung der Essigsäure zumeist ein pH-Wert von 9-11 angegeben wird. In der Tendenz wird also ein größerer Unterschied zwischen den beiden erstgenannten Säuren und Essigsäure beschrieben. Unterschiede in der Färbung des Indikators erkennt ein*e Proband*in (KNRN23CA_Säure-Base) bereits zu Beginn des Experiments. Hier werden die entstehenden Rot-Töne nach Zugabe von Rotkohllindikator zu den einzelnen Säuren unterschieden. Die beschriebenen Unterschiede sind dabei recht klein und befinden sich im Bereich von pH=3-5.

Bei Experiment 2 wurden insgesamt deutlich weniger anomale Beobachtungen beschrieben. Den Hauptanteil beschriebener Beobachtungen stellt die Beschreibung von „Blubbern/Brodeln“ des Wassers unterhalb 100°C dar (MEAM18FA_Siedepunkt, AASA23CA_Siedepunkt, MNAR08NE_Siedepunkt). Dabei wird die Beobachtung während des Abkühlungsprozesses im Experiment und nicht im Bereich des ‚ersten‘ Siedepunkts während der Erhitzung des Wassers beschrieben. Lediglich ein*e Proband*in konkretisiert die Beobachtung als „Siedepunkt liegt zwischen 75 und 90 Grad“ (INCH29FE_Siedepunkt). Beide Kategorien und damit alle 4 Proband*innen nehmen Bezug auf die theoretisch erwarteten anomalen Beobachtungen, also das Sieden von Wasser unterhalb von 100°C im Abkühlungsprozess. Dabei unterschieden sich die beiden Kategorien jedoch in ihrem Grad der Konkretheit. Während INCH29FE_Siedepunkt das Sieden des Wassers erkennt und folglich den Siedepunkt benennt, beschreiben die anderen 3 Proband*innen innerhalb der MOD lediglich das makroskopische Phänomen. Es kann hier also nicht davon ausgegangen werden, dass das erneute Sieden verstanden wird, jedoch erfolgt zumindest eine makroskopische Beschreibung.

Die fünf Beschreibungen anomaler Daten aus dem lauten Denken gehen auf 3 Proband*innen zurück. Während MEAM18FA sowohl die fehlende Rotfärbung des Wasserdampfes sowie das Blubbern, jedoch ausbleibende Kochen mit einem zweiten Siedepunkt bei unter 100°C in Verbindung bringt, beschreiben SEMS14LA sowie AEDL02JA das Blubbern des Wassers während der Abkühlung bei ca. 80°C. In allen fünf Fällen werden anomale Beobachtungen beschrieben, die zwar teilweise Bezug auf die theoretisch erwarteten anomalen Beobachtungen nehmen, jedoch nicht direkt das Sieden unterhalb 100°C thematisieren.

Insgesamt kann erkannt werden, dass – dem Trend der Anzahl gebildeter Aspekte und Verknüpfungen folgend – für Experiment 2 signifikant weniger anomale Beobachtungen beschrieben werden als für Experiment 1 ($t(20)=2.828$; $p=.010$; $d=.62$). Weiterhin wurden – wie erwartet – auch anomale Beobachtungen beschrieben, welche nicht den erwarteten anomalen Beobachtungen entsprechen. Dies gilt besonders für Experiment 1, in welchem 1/3 der Proband*innen weitere anomale Beobachtungen beschrieben haben.

Basierend auf dem Modell von Chinn und Brewer (1993b, 1998) könnten unterschiedliche Umgangsarten mit anomalen Beobachtungen erwartet werden, welche sich wiederum auf unterschiedliche Art und Weise innerhalb eines MOD äußern könnten (Chinn & Brewer, 2001). Tabelle 38 gibt einen Überblick über die in MOD dargestellten bzw. verbal ergänzten wahrgenommen

anomalen Beobachtungen und den Umgang mit diesen. Dabei kann für 3 Proband*innen keine Einteilung in die postulierten Umgangsarten erfolgen. Bei SAGT10RL kann in Experiment 1 eine unklare Verknüpfung innerhalb des MOD erkannt werden. Auch durch die Erklärungen im Rahmen des lauten Denkens kann die der Umgang mit den anomalen Beobachtungen nicht genauer differenziert werden.

Für die Proband*innen AASA23CA sowie INCH29FE ist innerhalb Experiment 2 unklar, inwiefern die Beobachtungen als anomal wahrgenommen werden. Für die anderen Proband*innen hingegen sind teilweise auch mehrere anomale Beobachtungen in die Analyse eingegangen. So schildert MEAM18FA innerhalb von Experiment 2 z.B. vier unterschiedliche anomale Beobachtungen.

Für Experiment 1 können Umgangsarten mit anomalen Beobachtungen erkannt werden, welche keinerlei Erklärung für diese geben. Hier können die Umgangsarten *Exclusion* und *Abeyance* nach Chinn und Brewer (1993b) beschrieben werden. Im Rahmen der Umgangsart *Exclusion* kann bei den Proband*innen die weitergehende Beschreibung (z.B. KNRN23CA_Säure-Base) bzw. kausale Folgerung der anomalen Beobachtungen (z.B. SETS14KA_Säure-Base) erkannt werden. Die Proband*innen anerkennen und beschreiben die anomalen Beobachtungen im Sinne der unterschiedlichen Färbungen/pH-Werte am Ende des Experiments, geben jedoch keine Erklärung für das Auftreten der anomalen Beobachtungen bzw. wie diese in die bestehenden Vorstellungen integriert werden könnten. Auch bei der Umgangsart *Abeyance* können keinerlei Erklärungsansätze beschrieben werden. Die Proband*innen beschreiben hier einen Konflikt zwischen den eigenen Vorstellungen und den anomalen Daten, können diese jedoch nicht erklären (z.B. BTRE19GA_Säure-Base). Dabei ist die Unterscheidung von der Umgangsart *Exclusion* nur schwer möglich. Eine Unterscheidung könnte die Vernetzung der Aspekte innerhalb des MOD darstellen. Werden anomale Beobachtungen nur beschrieben und isoliert innerhalb einzelner Verknüpfungen dargestellt, ließe sich eher eine *Exclusion* argumentieren. Beiden Umgangsarten ist ähnlich, dass keine umfassende Integration der anomalen Beobachtungen in das MOD erfolgt.

Im Rahmen der *Reinterpretation* werden die Daten wie durch Chinn und Brewer (1993b) beschrieben im Rahmen der bestehenden Vorstellungen erklärt. So kann z.B. bei IAKS30JA_Säure-Base erkannt werden, dass die Probandin an der Neutralisation einer Säure durch eine Natronlauge festhält, jedoch für das Teilexperiment der Essigsäure eine ‚nicht vollständige‘ Neutralisation beschreibt. Damit werden die Beobachtungen reinterpretiert, sodass diese die bestehenden Vorstellungen unterstützt, bzw. diesen nicht widerspricht. Ähnliches ist für Experiment 2 bei MEAM18FA_Siedepunkt zu erkennen. Die Probandin beschreibt hier den zweiten Siedepunkt des Wassers als ‚Blubbern‘, wodurch einerseits die Beobachtungen akzeptiert werden, jedoch andererseits kein Widerspruch zu bestehenden Vorstellungen entsteht. Dabei werden die auftretenden anomalen Beobachtungen hier nicht erklärt.

Für Experiment 1 ist ein Schwerpunkt auf der Theorieveränderung zu erkennen. Dabei können sowohl periphere Veränderungen der Theorie als auch komplette Theoriwechsel beobachtet werden. Eine Unterscheidung der beiden Umgangsarten kann dabei nicht immer eindeutig erfolgen. Ein Beispiel dafür kann bei MERF13KN_Säure-Base erkannt werden. Die Proband*in beschreibt hier eine Abhängigkeit des pH-Werts von den Lösungen am Ende des Experiments von der ausgangs eingesetzten Säure. Damit werden Charakteristika der Säure zur Erklärung des Phänomens eingesetzt. Es erfolgt jedoch keine Erklärung auf theoretischer Ebene. Dadurch ist es unklar, inwiefern eine Theorie völlig verändert wurde, oder nur eine Anpassung vorgenommen wurde. Hinzu kommt eine unklare Verknüpfung der Aspekte innerhalb des graphischen MOD. Es ist jedoch zu erkennen, dass die anomalen Daten in das MOD aufgenommen wurden und mit weiteren Aspekten in Verbindung gebracht werden.

Tabelle 38 – Umgang der Proband*innen mit anomalen Beobachtungen in Experiment 1 und 2

Proband*in	Kodierung anomaler Beobachtungen	Umgangsart mit anomalen Beobachtungen nach Chinn und Brewer (1993b, 1998)
Experiment 1 – Reaktion von Säuren mit Natronlauge		
AAVR08LA	Unterschiedliche/r Färbung/pH-Wert der 3 Säuren nach Exp.	Exclusion
KNRN23CA	Unterschiedlicher pH-Wert zu Beginn	Exclusion
SETS14KA	Unterschiedliche/r Färbung/pH-Wert der 3 Säuren nach Exp.	Exclusion
GAWI21JA	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	Abeyance
CAJN05DN	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	Abeyance
AADK15LE	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	Abeyance
BTRE19GA	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	Abeyance
BAVR30PA	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	Reinterpretation
MARF16AN	Unterschiedliche/r Färbung/pH-Wert der 3 Säuren nach Exp.	Reinterpretation
IAKS30JA	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	Reinterpretation
RYER27LA	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	Reinterpretation
SEMS14LA	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	Peripheral theory change
KNRN23CA	Unterschiedliche/r Färbung/pH-Wert der 3 Säuren nach Exp.	Peripheral theory change
MERF13KN	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	(Peripheral) theory change
MASO20CA	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	Theory change
INCH29FE	Unterschiedliche/r Färbung/pH-Wert der 3 Säuren nach Exp.	Theory change
IABD29MM	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	Theory change
MEAM18FA	Unterschiedliche/r Färbung/pH-Wert der 3 Säuren nach Exp.	Theory change
AASA23CA	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	Theory change
MNAR08NE	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	Theory change
AEDL02IN	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	Theory change
SAGT10RL	Grünfärbung/andere Färbung nach Reaktion mit Essigsäure	unklar
Experiment 2 – Siedepunkt von Wasser		
MEAM18FA	Wieso ist Wasserdampf nicht rot?	Ignoring
MEAM18FA	Zweiter Siedepunkt unter 100°C	Ignoring
SEMS14LA	Wasser blubbert beim Abkühlen bei ca. 80 °C	Ignoring
AEDL02JA	Wasser blubbert beim Abkühlen bei ca. 80 °C	Ignoring
MNAR08NE	Zweiter Siedepunkt wird als "Blubbern/Brodeln" beschrieben	Ignoring
MEAM18FA	Wasser blubbert, kocht aber nicht bei unter 100°C	Teilweises Ignoring
MEAM18FA	Zweiter Siedepunkt wird als "Blubbern/Brodeln" beschrieben	Abeyance
INCH29FE	Zweiter Siedepunkt von Wasser bei unter 100°C	unklar
AASA23CA	Zweiter Siedepunkt wird als "Blubbern/Brodeln" beschrieben	Unklar

Anmerkung. Orange Markierung weisen auf anomale Beobachtungen hin, welche lediglich innerhalb des lauten Denkens formuliert wurden.

Für Experiment 2 kann zusätzlich noch die Umgangsart *Ignoring* beschrieben werden. Die Proband*innen bringen die Beobachtungen hier nicht in die Bearbeitung des Phänomens ein. Dabei kann für einen Großteil der Fälle erkannt werden, dass die anomalen Daten nur innerhalb des lauten Denkens formuliert werden, jedoch keinen Einfluss in das MOD finden (z.B. MEAM18FA_Siedepunkt, SEMS14LA_Siedepunkt). In diesen Fällen werden im Rahmen des lauten Denkens anomale Beobachtungen formuliert, welche nicht weiter diskutiert werden. So beschreibt MEAM18FA_Siedepunkt z.B. eine Verwunderung über die ausbleibende rote Färbung des Wasserdampfes. Da das Wasser rot eingefärbt wird, scheint hier die Beobachtung von farblosen Wasserdampf als anomal erkannt zu werden. Die Probandin beschreibt dabei ihre Verwirrung, geht jedoch weder im lauten Denken noch im MOD weiter auf die Beobachtungen ein. Für die gleiche Probandin kann ebenso ein teilweises Ignorieren anomaler Beobachtungen formuliert werden. So beschreibt die Probandin das ‚Blubbern‘ des Wassers, welches jedoch vom Sieden unterschieden wird im lauten Denken. Innerhalb des MOD wird jedoch nur das ‚Blubbern‘ aufgenommen. Ein Zusammenhang bzw. eine Trennung vom Sieden von Wasser erfolgt nicht mehr.

Übergreifend sind Unterschiede zwischen Experiment 1 und 2 zu erkennen. Während wie beschrieben in Experiment 1 häufiger anomale Beobachtungen getätigt werden, kann hier auch öfter eine Theoriebearbeitung erkannt werden. Innerhalb von Experiment 2 kann die Theoriebearbeitung nicht beobachtet werden. Dafür werden die Beobachtungen in Experiment 2 häufig ignoriert, während diese Umgangsart für Experiment 1 nicht beschrieben werden kann. Auch der Einbezug anomaler Beobachtungen in die Beschreibung des Phänomens bei gleichzeitig ausbleibender Erklärung dieser ist für Experiment 1 deutlich häufiger zu beobachten.

7.6.3 Analyse der gebildeten Verknüpfungen

Im Rahmen der Analyse der MOD müssen die erstellten Verknüpfungen und deren Adäquatheit vor dem Hintergrund epistemologischer Überzeugungen betrachtet werden.

In einem ersten Schritt sollen dazu die quantitativen Maße herangezogen werden, welche in Kapitel 7.5.1 vorgestellt wurden. Tabelle 39 gibt einen ersten Überblick über die Anzahl, den prozentualen Anteil sowie Adäquatheit der gebildeten Verknüpfungen.

Tabelle 39 – Anzahl, prozentualer Anteil sowie Adäquatheit der gebildeten Verknüpfungen

Verknüpfungsart	Gesamtzahl gebildeter Verknüpfungen	Prozentualer Anteil an Gesamtverknüpfungszahl	Gesamtzahl adäquater Verknüpfungen	Prozentualer Anteil adäquater Verknüpfungen
Kausale Verknüpfungen (gesamt)	390	71.298	241	61.795
Beobachtung-Beobachtung	337	61.609	211	62.611
Konzeptwissen-Konzeptwissen	4	0.731	0	0
Beobachtung-Konzeptwissen	49	8.958	30	61.224
Induktive Verknüpfungen	46	8.410	9	19.565
Analogie-Verknüpfung	33	6.033	7	21.212
Kontrastierende Verknüpfung	44	8.044	37	84.091
Unmögliche kausale Verknüpfungen	34	6.216	21	61.765
	547		315	57.587

Anmerkung. Die Stichprobe basiert auf $N_{\text{MOD}}=42$ MOD, welche durch die 21 Proband*innen der Studie erstellt wurden.

Auffällig ist der große Anteil gebildeter kausaler Verknüpfungen an der Gesamtzahl gebildeter Verknüpfungen. Die kausalen Verknüpfungen können dabei zwischen unterschiedlichen Aspekt-Arten gebildet werden. Der Schwerpunkt liegt hier auf kausalen Verknüpfungen zwischen zwei Beobachtungsaspekten. Die Anzahl von kausalen Verknüpfungen zwischen Beobachtungs- und Konzeptaspekten, wie auch zwischen zwei Konzeptaspekten sind deutlich geringer. Hier könnte auch ein Zusammenhang zur Anzahl unterschiedlicher gebildeter Aspekte hergestellt werden. Für die restlichen gebildeten Verknüpfungsarten sind kaum Unterschiede in der Häufigkeit erkennbar. Gemeinsam haben diese jedoch, dass sie deutlich weniger häufig gebildet wurden als kausale Verknüpfungen.

Tabelle 40 – Ergebnisse eines Wilcoxon-Tests für Unterschiede in der Anzahl gebildeter Verknüpfungen zwischen Exp. 1 und 2

Verknüpfungsart		Exp. 1	Exp. 2	z	p	Pearson's r
		M (SD)	M (SD)			
Gesamtzahl der gebildeten Verknüpfungen	Kausale Verknüpfungen					
	Beobachtung-Beobachtung	11.810 (5.938)	4.238 (2.548)	-3.98	< .001	.87
	Konzeptwissen-Konzeptwissen	.095 (.436)	.095 (.301)	.00	1.00	
	Beobachtung-Konzeptwissen	1.857 (2.435)	.476 (.873)	-2.28	.022	.50
	Induktive Verknüpfungen	1.714 (1.978)	.476 (.750)	-2.87	.004	.63
	Analogie-Verknüpfungen	1.238 (1.411)	.333 (.577)	-2.43	.015	.53
	Kontrastierende Verknüpfungen	1.810 (1.327)	.286 (.143)	-3.69	< .001	.81
	Unmögliche kausale Verknüpfungen	1.333 (1.065)	.286 (.561)	-.340	.001	.07
Anzahl adäquat gebildeter Verknüpfungen	Kausale Verknüpfungen					
	Beobachtung-Beobachtung	7.333 (4.553)	2.714 (1.875)	-3.73	< .001	.81
	Konzeptwissen-Konzeptwissen	.0 (.0)	.0 (.0)	.00	1.00	
	Beobachtung-Konzeptwissen	1.143 (1.797)	.286 (.784)	-1.87	.062	
	Induktive Verknüpfungen	.381 (.805)	.048 (.218)	-1.84	.066	
	Analogie-Verknüpfungen	.191 (.402)	.143 (.359)	-.45	.655	
	Kontrastierende Verknüpfungen	1.619 (1.396)	.143 (.359)	-3.57	< .001	.78
	Unmögliche kausale Verknüpfungen	.810 (.928)	.191 (.512)	-2.39	.017	.52

Anmerkung. Die Effektstärken nach Pearson's r wurden nur für Unterschiede berechnet, welche ein Signifikanzniveau von $p=.05$ erreicht haben. $N=21$

In Tabelle 40 können die Mittelwerte der Anzahl gebildeter Verknüpfungen und adäquater Verknüpfungen für die einzelnen Verknüpfungsarten, getrennt nach Experiment, erkannt werden. Die Standardabweichungen zeigen größere Abweichungen im Bereich kausaler Verknüpfungen auf. Im Bereich der restlichen Verknüpfungsarten sind weniger starke Abweichungen erkennbar. Ebenso fällt auf, dass in Exp. 2 deutlich weniger Verknüpfungen gebildet wurden. Für einige Verknüpfungsarten wird dieser Unterschied signifikant (vgl. Tab. 40). Erneut fallen im Bereich kausaler Verknüpfungen große Unterschiede (mit großen Effektstärken) auf. Die Unterschiede der weiteren Verknüpfungsarten zeigen ebenfalls mittelstarke Effektstärken auf ($r>0.5$), sind jedoch

deutlich kleiner als die der kausalen Verknüpfungen. Folglich wurden innerhalb von Experiment 1 signifikant mehr Verknüpfungen zwischen unterschiedlichen Aspekten gebildet (außer kausale Verknüpfungen zwischen zwei Konzeptwissensaspekten). Dabei muss jedoch auf die unterschiedliche Anzahl unterschiedlicher Aspekte innerhalb der beiden Experimente hingewiesen werden. So könnten ggf. weniger Verknüpfungen gebildet werden, da weniger Aspekte gebildet wurden. Entsprechend wurde ebenfalls ein Wilcoxon-Test für die Anzahl gebildeter Verknüpfungen pro gebildeten Aspekten für beide Experimente durchgeführt. Auch hier zeigt sich ein signifikanter Effekt ($z = -3.98$, $p < .001$, $N = 21$) mit einer großen Effektstärke ($r = .87$). Während in Experiment 1 im Mittel $M = 2.51$ ($SD = 1.29$) Verknüpfungen pro Aspekt gebildet wurden, konnten für Experiment 2 im Mittel nur $M = .87$ ($SD = .30$) Verknüpfungen pro gebildetem Aspekt beobachtet werden. Entsprechend wurden innerhalb von Experiment 1 nicht nur mehr Aspekte und Verknüpfungen gebildet, sondern auch mehr Verknüpfungen pro gebildetem Aspekt.

Als weiteres quantitatives Maß zur Analyse der gebildeten Verknüpfungen wurde ein Score für die Nutzung unterschiedlicher Verknüpfungsarten vorgeschlagen (vgl. Kap. 7.5.1). Der Score basiert auf der unterschiedlichen Bepunktung der verschiedenen Verknüpfungsarten durch das Kodiermanual. Ein Wert des Scores nahe 1 weist auf die Nutzung weniger komplexer Verknüpfungsarten hin (kausale und kontrastierende Verknüpfungen). Je höher der Score, desto mehr komplexe Verknüpfungen (induktive, Analogie- und unmögliche kausale Verknüpfungen) wurden gebildet. Da sich der Score auf Grundlage der Anzahl adäquat gebildeter Verknüpfungen errechnet, kann dieser auch gleich Null sein. Andernfalls ist ein Score im Bereich zwischen 1-3 möglich. Ein Vergleich zwischen dem Verknüpfungsscore von Experiment 1 ($M = 1.314$, $SD = .443$, $N = 21$) und Experiment 2 ($M = .941$, $SD = .604$, $N = 21$) zeigt keine signifikanten Unterschiede auf ($z = -1.74$, $p = .081$, $N = 21$). Folglich wurden in keinem der beiden Experimente signifikant mehr komplexe Verknüpfungsarten genutzt als in dem anderen.

Eine genauere Betrachtung der inadäquat gebildeten Verknüpfungen könnte Einblicke in die epistemologischen Überzeugungen der Proband*innen liefern. Entsprechend werden die inadäquat gebildeten Verknüpfungen im Folgenden genauer analysiert.

Bei genauerer Betrachtung nicht adäquat gebildeter Verknüpfungen fallen unterschiedliche Schwerpunkte auf. **Inadäquate kausale Verknüpfungen zwischen zwei Beobachtungsaspekten** stellen zumeist zeitlich nacheinander stattfindende Experimentierschritte oder Beobachtungen dar. Dabei besteht kein kausaler Zusammenhang zwischen den Aspekten. So verknüpft I-AKS30JA_Säure-Base z.B. die Aspekte „Rot gefärbtes Wasser entsteht im Glaskolben“ und „Glaskolben wird mit Hilfe einer Gaszufuhr erhitzt“. Dabei stehen die entstehende Färbung im Kolben und das Erhitzen des Kolbens nicht in einem kausalen Zusammenhang. Die Aspekte können mit den vorgegebenen Aspekten nicht verknüpft werden. Gleichwohl könnten beide gemeinsam z.B. durch kausale Verknüpfungen zu einem weiteren Aspekt verknüpft werden. Ebenso häufig, wie die Beschreibung zeitlicher Abfolgen zwischen Aspekten zu inadäquaten Verknüpfungen führen, werden Verknüpfungen gebildet, welche keinen kausalen Zusammenhang erkennen lassen. Hier handelt es sich um Verknüpfungen, welche weder kausalen Charakter besitzen noch sich in die bestehenden anderen Verknüpfungsarten einordnen lassen könnten. So hat GAWI21JA_Siedepunkt z.B. die Aspekte „Wasser wird in Kolben gefüllt“ und „Temperatur von 20,1 Grad“ kausal miteinander verknüpft. Es bleibt hier unklar, wie sich die Verknüpfung inhaltlich begründet.

Einige weitere inadäquate Kausalverknüpfungen zwischen zwei Beobachtungsaspekten sind auf gekürzte Kausalschlüsse, bzw. das Fehlen von weiteren Bedingungen zurückzuführen. Als Beispiel kann die Verknüpfung der Aspekte „Temperatur konstant bei 101°“ und „Siedepunkt“ von CAJN05DN_Siedepunkt angeführt werden. In diesem Beispiel soll der Siedepunkt des Wassers auf

der Konstanz der Temperatur begründet werden. Unter Annahme des Siedens des Wassers sowie der weiterhin erfolgenden Erhitzung (ohne weitere Temperatursteigerung) kann diese Kausalverknüpfung anerkannt, bzw. nachvollzogen werden. Jedoch fehlen diese Informationen bei der Erstellung der vorliegenden Verknüpfung. Entsprechend muss diese Verknüpfung als inadäquat beschrieben werden, wenngleich nicht unterstellt werden kann, dass das inhaltliche Konzept des Siedepunkts durch CAJN05DN nicht verstanden wurde. Folglich handelt es sich hier um eine gekürzte Kausalverknüpfung, welche vor dem Hintergrund nicht formulierter/verknüpfter Beobachtungen oder Hintergrundwissensaspekte als inadäquat bezeichnet werden muss.

Bei den **inadäquaten kausalen Verknüpfungen zwischen zwei Konzeptaspekten** können ähnliche Gründe der Inadäquatheit erkannt werden, wie bereits beschrieben. So können Verknüpfungen erkannt werden, welche keinen kausalen Zusammenhang darstellen. Als Beispiel kann die gebildete Verknüpfungen zwischen den Aspekten „Siedepunkt von Wasser hängt von der Umgebungstemperatur ab“ und „Durch Reduzierung der Umgebungstemperatur kann das kochende Wasser schneller abkühlen“ von AAVR08LA_Siedepunkt angeführt werden. Die beiden Aspekte beschreiben einmal einen Einfluss auf die Siedetemperatur (inadäquater Konzeptaspekt) sowie auch die Abkühlung des Wassers. Dabei besteht zwischen den Aspekten kein Zusammenhang, welcher durch eine der Verknüpfungsarten beschrieben werden könnte. Auch verkürzte Verknüpfungen können erkannt werden. SEMS14LA_Säure-Base bildet z.B. eine Verknüpfung zwischen den Aspekten „Ph wert (sic!) ist neutralisiert durch Natronlauge“ und „Natronlauge ph wert (sic!) ist basisch (ca. 12)“. Hier wird aus der Neutralisation der Säure durch die Zugabe von Natronlauge auf einen pH-Wert der Lauge von 12 geschlossen. Diese Verknüpfung bedarf jedoch der Kenntnis um welche Säure, in welcher Konzentration es sich handelt. Diese Informationen sind z.T. aus dem Experiment, bzw. dem Hintergrundwissen abrufbar. Aus den beiden angeführten Aspekten kann hingegen nicht direkt eine kausale Verknüpfung erkannt werden. Folglich handelt es sich um eine inadäquate Verknüpfung durch Verkürzung der Argumentation bzw. fehlenden Einbezug weiterer Aspekte.

Neben den beiden Gründen für Inadäquatheit bei kausalen Verknüpfungen zwischen zwei Konzeptaspekten tritt bei den **inadäquaten Verknüpfungen zwischen Beobachtungs- und Konzeptaspekten** die Richtung der gebildeten Verknüpfungen hinzu. So stellt IAKS30JA_Säure-Base eine Verknüpfung zwischen den Aspekten „Die Essigsäure wurde nicht vollständig neutralisiert“ und „Besitzt einen pH-Wert von 8“ auf. Dabei kann das Ausbleiben einer Neutralisation mit dem End-pH-Wert=7 aus dem pH-Wert von 8 gefolgert werden, jedoch nicht andersherum. Hier wurden Ursache und Effekt, bzw. Grundlage und Schlussfolgerung vertauscht. Es handelt sich folglich im Grundsatz um eine kausale Verknüpfung, jedoch in umgekehrter Reihenfolge als dargestellt. Tabelle 41 gibt einen Überblick über die Verteilung auf die jeweiligen Arten inadäquater Verknüpfungen.

Tabelle 41 – Verteilung inadäquater Verknüpfungen auf unterschiedliche Ursachen

	Begründung der Inadäquatheit	Anzahl inadäquater Verknüpfungen für jeweilige Begründungen
Inadäquate kausale Verknüpfungen zwischen zwei Beobachtungsaspekten	Beschreibt keinen kausalen Zusammenhang	55
	Beschreibt zeitliche Abfolge	55
	Verkürzt/Weitere Bedingungen fehlen	14
Inadäquate kausale Verknüpfungen zwischen zwei Konzeptaspekten	Verkürzt/Weitere Bedingungen fehlen	2
	Beschreibt keinen kausalen Zusammenhang	2

	Begründung der Inadäquatheit	Anzahl inadäquater Verknüpfun- gen für jewei- lige Begrün- dungen
Inadäquate kausale Verknüpfungen zwi- schen Beobachtungs- und Konzeptaspek- ten	Beschreibt keinen kausalen Zusammenhang	11
	Verkürzt/Weitere Bedingungen fehlen	4
	Falsche Richtung der Verknüpfung	4
Inadäquate induktive Verknüpfungen	Keine Verallgemeinerung, oder andere Verknüp- fungsart	12
	Kein Einzelfall für Verallgemeinerung vorhanden	9
	Eher kausaler Zusammenhang	7
	Unklar welche Aspekte verknüpft werden sollen	4
	Verallgemeinerung nur aus einem Teil der Aspekte möglich	1
	Eher Analogie-Verknüpfung	1
	Darstellung eines Beispiels	1
	Eher kontrastierende Verknüpfung	1
	Darstellung von Ähnlichkeiten	1
Inadäquate Analogie- Verknüpfungen	Keine Verallgemeinerung, oder andere Verknüp- fungsart	8
	Eher induktive Verknüpfung	4
	Eher kausale Verknüpfung	4
	Kein Einzelfall für Verallgemeinerung vorhanden	3
	Verallgemeinerung nicht aus verknüpftem Aspekt möglich	3
	Unklar welche Aspekte verknüpft werden sollen	2
	Verallgemeinerung nur aus einem Teil der Aspekte möglich	2
Inadäquate kontrastierende Verknüpfungen	Eher Analogie-Verknüpfung	3
	Unklarer Kontrast	2
	Eher induktive Verknüpfung	1
	Unklar welche Aspekte verknüpft werden sollen	1
Inadäquate unmögliche kausale Ver- knüpfungen	Aspekt A macht Aspekt B nicht unmöglich	4
	Unklar, doppelte Verneinung	3
	Unklar welche Aspekte verknüpft werden sollen	2
	Unklarer inhaltlicher Aspekt macht Einordnung nicht möglich	1
	Eher induktive Verknüpfung	1
	Richtung der Verknüpfung unklar	1
	Eher unmögliche Verallgemeinerung	1

Anmerkung. Die hier dargestellten inadäquaten Verknüpfungen wurden aus den je 2 Model-of-data der N=21 Proband*innen gebildet.

Bei den **inadäquaten induktiven Verknüpfungen** wurden am häufigsten Verknüpfungen gebildet, welche nicht als induktive Verallgemeinerungen oder eine andere Verknüpfungsart bezeichnet werden können. So bildet MARF16AN_Siedepunkt eine induktive Verknüpfung zwischen den Aspekten „Offen auf 100°C erhitzt“ und „Durch das ca. 17°C ‚kalte‘ Wasser gekühlt“. Hierbei handelt es sich nicht um eine Verallgemeinerung, jedoch auch nicht um eine der anderen zur Verfügung stehenden Verknüpfungsarten. Vielmehr werden hier zwei Beobachtungsaspekte miteinander verknüpft, welche nicht verknüpft werden können. Ausgehend von diesen Aspekten könnten Verknüpfungen erstellt werden, welche zur Erklärung des beobachteten Phänomens beitragen. Eine

Verknüpfung der Aspekte mit den vorliegenden Verknüpfungsarten ist jedoch nicht möglich. Weiterhin werden Verknüpfungen gebildet, bei denen kein Einzelfall für eine Verallgemeinerung zur Verfügung steht. So verallgemeinert KNRN23CA_Siedepunkt den Aspekt „Wasser wird erhitzt, bis es kocht → Höchsttemperatur ist 101,6 Grad“ sowie weitere Aspekte des Model-of-data zur Aussage „Siedepunkt von Wasser liegt bei circa 101,4 Grad“. Hier werden zwei unterschiedliche Siedetemperaturen genutzt, was eine Verallgemeinerung unmöglich macht. Für die verallgemeinerte Aussage wäre ein Einzelfall (oder Aspekt) nötig, welcher eine Siedetemperatur von 101,4°C aufweist. Dies steht im Kontrast zur formulierten Beobachtung. In anderen Fällen wird der fehlende Einzelfall noch deutlicher. So wird z.B. aus der Hinzufügung von Natronlauge zur Lösung in Exp. 1 verallgemeinert, dass Gemische mit Natronlauge einen pH-Wert von 7 haben (AASA23CA_Säure-Base). Dabei kann aus der bloßen Zugabe von Natronlauge kein pH-Wert etc. verallgemeinert werden. In einigen Fällen würden sich andere Stellen des MOD für die gebildeten Verallgemeinerungen anbieten. AASA23CA_Säure-Base beschreibt in Nähe des verknüpften Aspekts den pH-Wert von 7 für die entstehende Lösung. Eine Verknüpfung dieses Aspekts könnte zur abgezielten Verallgemeinerung führen. Es spielt hier also in einigen Fällen eine Rolle, welche Aspekte für die gebildeten Verallgemeinerungen verknüpft wurden. Weitere inadäquate induktive Verknüpfungen muten eher wie andere Verknüpfungsarten an. Dabei können diese teils nicht eindeutig einer anderen Verknüpfungsart zugeordnet werden, scheinen jedoch von ihrer Natur eher zu diesen Verknüpfungsarten zu passen. Als Beispiel kann die Verknüpfung der Aspekte „P-h-Wert (sic!) 8 (Blau) und „Natronlauge dient der Neutralisation“ von AADK15LE_Säure-Base angeführt werden. Hierbei handelt es sich nicht um eine Verallgemeinerung. Eher wäre eine kausale Verknüpfung der Aspekte (unter Zuhilfenahme weiterer gebildeter Aspekte) denkbar. Die Verknüpfung weist eher den Charakter einer Schlussfolgerung auf. Ähnliche Fälle gibt es auch bzgl. anderer Verknüpfungsarten. In einigen Fällen ist die graphische Zuordnung der verknüpften Aspekte nicht möglich (z.B. GAWI21JA_Säure-Base, MERF13KN_Säure-Base).

Für die gebildeten **inadäquaten Analogie-Verknüpfungen** können ähnliche Begründungen für Inadäquatheit erkannt werden, wie für die inadäquaten induktiven Verknüpfungen. Da es sich bei beiden Verknüpfungsarten um Verallgemeinerungen handelt, sind ähnliche Herausforderungen erkennbar. Eine neue Herausforderung für die Bildung von Analogie-Verknüpfungen stellt die Auswahl der verknüpften Aspekte dar. Zwar konnten ähnliche Herausforderungen schon für inadäquate induktive Verknüpfungen erkannt werden, jedoch sind in diesem Fall keine inadäquaten Aspekte als Ausgangspunkte für die Verallgemeinerungen gemeint. Viel mehr werden einige Aspekte zur Verallgemeinerung herangezogen, aus denen jedoch nur wenige die Verallgemeinerung zulassen. Als Beispiel kann MASA20CA_Säure-Base beschrieben werden. Durch eine Analogie-Verknüpfung wird der Aspekt „Natronlauge färbt alle Stoffgemische Blau (PH (sic!) Neutral)“ erstellt. Die Verallgemeinerung bezieht sich dabei auf die Zugabe von Natronlauge sowie die Verfärbung aller drei Lösungen aus Exp. 1. Damit ist auch der Beobachtungsaspekt „Farbe hat sich von Pink zu einem grün verändert“ des Experimentteils mit Essigsäure eingeschlossen. Dieser kann jedoch zur Verallgemeinerung des beschriebenen Aspekts nicht herangezogen werden. Damit scheidet ein Teil der genutzt Aspekte zur Verallgemeinerung als inadäquat im Sinne der Verallgemeinerung aus.

Neben der Einordnung gebildeter **inadäquater kontrastierender Verknüpfungen** als andere Verknüpfungsarten, stellt die Unklarheit über den zugrundeliegenden Kontrast eine Begründung für Inadäquatheit dar. SETS14KA_Siedepunkt bildet bspw. einen Kontrast zwischen den Aspekten „Kolben mit Wasser wird über einer Flamme erhitzt“ und „Kolben wird mit heißem Wasser verschlossen“. Dabei scheint es keinen oberflächlichen, oder tiefenstrukturellen Unterscheidungspunkt zwischen den Aspekten zu geben. Einzig die unterschiedlichen Beobachtungen

unterscheiden sich voneinander. Diese sind dabei Grundlage zur Bildung zwei getrennter Aspektkästchen, qualifizieren jedoch nicht für eine kontrastierende Verknüpfung.

Zuletzt werden **inadäquate unmögliche kausale Verknüpfungen** häufig auf Grundlage der fehlenden „Unmöglichkeit“ als inadäquat kategorisiert. So macht z.B. der Aspekt „Essigsäure reagiert mit Rotkohllindikator“ die Aussage „Alle Stoffe vermischt mit Rotkohllindikator werden durch Natronlauge neutralisiert (ph (sic!) wert 7)“ nicht unmöglich (SEMS14LA_Säure-Base). In drei Fällen kam es zu doppelter Verneinung und damit verbundener Unklarheit, wie die Verknüpfungen zu verstehen sind. In Rückbezug auf die inadäquaten kausalen Verknüpfungen kann wiederum eine Verknüpfung erkannt werden, deren Richtung inadäquat ist (MASO20CA_Säure-Base).

Neben den beschriebenen Begründungen für inadäquate Verknüpfungen sind weitere Begründungen (in sehr geringen Fallzahlen) vorhanden, welche aus Tab. 44 abgeleitet werden können. Die geringe Zahl dieser Fälle lässt sich meist auf einzelne MOD, bzw. Proband*innen zurückführen. Hier wurden teils bestimmte inadäquate Verknüpfungen gleich mehrfach in einem MOD gebildet.

Für die einzelnen Verknüpfungsarten können unterschiedliche Begründungen für Inadäquatheit berichtet werden. Dabei tritt teils eine Verwechslung mit anderen Verknüpfungsarten auf. In anderen Fällen ist die Richtung oder aber der Einbezug weiterer Aspekte als Grund für Inadäquatheit anzuführen. So wird z.B. bei kausalen und unmöglichen kausalen Verknüpfungen die Richtungen der Verknüpfung vertauscht, während bei verallgemeinernden Verknüpfungsarten vor allem der Einbezug mehrerer Aspekte Herausforderungen darzustellen scheint. Teilweise ist ebenfalls eine Abweichung von der epistemologischen Natur der Verknüpfung zu erkennen. So werden kausale Ursache-Wirkung-Verknüpfungen auch zur Darstellung zeitlicher Abfolgen eingesetzt oder Verallgemeinerungen genutzt, obwohl inhaltlich keine Verallgemeinerung stattfindet.

7.6.4 Ergebnisse der Auswertung des Fachwissenstests

Im Rahmen des Fachwissenstests wurden die beschriebenen Sub-Skalen (vgl. Kap. 7.4.1) sowie Selbsteinschätzungen zur Leistungsfähigkeit abgefragt.

Selbsteinschätzungen

Die Selbsteinschätzung der eigenen Leistungsfähigkeit wurde für unterschiedliche Inhaltsbereiche abgefragt. Tabelle 42 gibt einen Überblick über die deskriptive Statistik für die Selbsteinschätzung in den verschiedenen Inhaltsbereichen. Die Selbsteinschätzung der Proband*innen liegt dabei mit Ausnahme für den Inhaltsbereich „Stoffeigenschaften“ unter dem Skalenmittelwert der 5-stufigen Likert-Skala (1=sehr schlecht, 5=sehr gut).

Tabelle 42 – Deskriptive Statistik für die Selbsteinschätzung der Proband*innen für unterschiedliche Inhaltsbereiche

Inhaltsbereich der Selbsteinschätzung	<i>M</i>	<i>SD</i>
Chemisches Wissen/Kenntnisse	2.71	.78
Physikalisches Wissen/Kenntnisse	2.10	.83
Säure-Base-Chemie	2.19	.87
Stoffeigenschaften	3.14	.73

Anmerkung. N=21

Fachwissen

Für die Skala zum Wissen im Bereich „Säure-Base“ kann ein Mittelwert für die erreichte Punktzahl von $M=7.52$ ($SD=2.36$) ermittelt werden. Damit erreichen die Proband*innen im Mittel 68.36% der maximalen Gesamtpunktzahl. Eine Korrelationsanalyse zeigt keine signifikanten Zusammenhänge mit Merkmalen der MOD auf. Dabei wurden Merkmale wie die Anzahl adäquat gebildeter Verknüpfungen sowie unterschiedlicher Aspekte in die Analyse einbezogen.

Für die Siedepunkt-Skala können auf Grundlage der kritischen internen Konsistenz keine Mittelwerte errechnet werden. Dafür gibt Tabelle 43 eine Übersicht über die Beantwortung der Einzelitems. Besonders für die Items 4, 7 und 8 sind niedrige Mittelwerte zu erkennen. Da die maximal erreichbare Punktzahl der Einzelitems je 1 Punkt ist, entspricht die prozentual erreichte Punktzahl der Itemschwierigkeit, welche zusätzlich in Tabelle 43 dargestellt wird. Die dargestellten Itemschwierigkeiten zeigen vor allem für Items 3-8 psychometrisch schwere Items auf.

Tabelle 43 – Deskriptive Statistik für die Einzelitems der Siedepunkt-Skala

Item	N	M (SD)	Prozentual erreichte Punktzahl	Itemschwierigkeit P_i
1	21	.76 (.44)	76%	76%
2	21	.67 (.48)	67%	67%
3	21	.38 (.50)	38%	38%
4	20	.25 (.44)	25%	25%
6	18	.50 (.51)	50%	50%
7	21	.14 (.36)	14%	14%
8	21	.33 (.48)	33%	33%

Anmerkung. Item 5 wurde auf Grundlage der untersuchten internen Konsistenz aus der Skala ausgeschlossen (vgl. Kap. 7.4.1).

Auf Grundlage der kritischen internen Konsistenz, kann keine Korrelationsanalyse für einen Summenscore erfolgen. Daher sollen im nächsten Schritt die Einzelitems der Siedepunkt-Skala mit Merkmalen der gebildeten MOD korreliert werden. Im Rahmen der Korrelationsanalysen (Rangkorrelation nach Spearman) wurden Zusammenhänge zwischen den erstellten MOD, sowie den Einzelitems der Siedepunkt-Skala gefunden.

Ein negativer Zusammenhang der Anzahl adäquater kausaler Verknüpfungen ist mit Item 8 der Siedepunkt-Skala zu erkennen ($r = -.47$, $p = .032$, $N=21$) Item 8 fragt dabei nach der Veränderung der Temperatur von kochendem Wasser bei weiterer Erhöhung der Temperatur der Heizplatte. Entsprechend zeigt sich ein negativer Zusammenhang zwischen der Beantwortung von Item 8 mit der Anzahl erstellter adäquater kausaler Verknüpfungen.

Ebenso zeigt die Analyse einen Zusammenhang für die Anzahl adäquat formulierter kontrastierender Verknüpfungen mit Item 6 der Siedepunktskala ($r = .61$, $p = .003$, $N=21$). Dieses Item fragt nach den Vorgängen des Abkühlungsprozesses. Dabei erstellen Proband*innen, welche das Item richtig beantworten signifikant mehr adäquate kontrastierende Verknüpfungen und andersherum.

Für Item 1 kann ein korrelativer Zusammenhang mit der Anzahl essentieller Beobachtungsaspekte gezeigt werden ($r = .43$, $p = .050$, $N=21$). Dabei fokussiert das Item auf die wahrscheinlichste Temperatur bei beginnendem Blasenwurf von Wasser.

7.6.5 Ergebnisse zur Validität der erstellen Model-of-data

Nach der Analyse des Umgangs mit anomalen Daten innerhalb von MOD stellt sich die Frage wie valide diese die mentalen Modelle der Proband*innen abbilden können. Dazu sollen im folgenden Ergebnisse zu Forschungsfrage III.b berichtet werden. Ergebnisse der Auswertung der Videos des lauten Denkens

Intercoder-Übereinstimmung für den Vergleich der MOD mit dem lauten Denken

Neben den MOD wurde auch die Intercoder-Übereinstimmung für die Analyse der Videos des lauten Denkens bestimmt. Die Intercoder-Übereinstimmung von $\kappa = .74$ ($N_{\text{Videos}}=6$) kann ebenfalls in den oberen Bereich der Benchmark „substantiell“ eingeordnet werden (Landis & Koch, 1977). Die Stichprobe besteht dabei aus 6 Videos, welche jeweils die beiden MOD der Proband*innen umfassen. Tabelle 44 gibt einen Überblick über kodierte Abweichungen zwischen MOD und lautem Denken sowie Aspekte/Verknüpfungen, welche im lauten Denken verbalisiert, jedoch im MOD nicht dargestellt wurden.

Tabelle 44 – Vergabe von Kodierungen sowie Intercoder-Übereinstimmung bzgl. Abweichungen und verbalisiert, aber nicht notierter Aspekte im lauten Denken

Art der Aspekte	Kodierende*r 1	Kodierende*r 2	Prozentuale Übereinstimmung	Cohen's Kappa
Abweichungen zwischen lautem Denken und MOD	40	41	70.73%	.69
Im lauten Denken verbalisierte, aber MOD nicht dargestellte Aspekte/Verknüpfungen	27	29	82.14%	.81
Gesamt	67	70	75.36%	.74

Anmerkung. $N_{\text{Videos}}=6$

Zusammenfassend kann die Intercoder-Übereinstimmung bei der Analyse der Videos des lauten Denken als zufriedenstellend bezeichnet werden. Abweichungen zwischen den Kodierenden sind teils auf die Verwendung nicht angemessener Kategorien zurückzuführen. Bei erneuter Ansicht und Diskussion der Kodierungen konnte so die Vergabe von Kategorien korrigiert werden. Weiterhin wurden einige Beschreibungen der Proband*innen diskutiert. Hier wurde teils von einer*m Kodierenden eine Abweichung erkannt, von der*dem anderen Kodierenden jedoch nicht. Die gemeinsame Diskussion der Beschreibungen konnte hier teils semantische Klarheit erzeugen und so z.T. zu Anpassungen der Kodierungen führen. So konnte die gemeinsame Diskussion und Konsensfindung der Kodierenden zu einer substantiellen Übereinstimmung beitragen.

Neben den dargestellten Validitätsargumenten und der Untersuchung der Reliabilität der Analyse mithilfe der Intercoder-Reliabilität, kann die pseudonymisierte Auswertung als Beitrag zur Auswertungsobjektivität verstanden werden. Wie in Kap. 7.2 beschrieben, werden für die Proband*innen pseudonymisierte Codes vergeben. Diese ermöglichen die Zuordnung von erstellten MOD und Videos, erlauben jedoch keine Zurückverfolgbarkeit der Personen. So ist den Kodierenden bspw. während der Analyse nicht klar, in welchen Studiengang die Proband*innen studieren o.Ä. So soll die Objektivität der Auswertung gesteigert werden.

Im Rahmen des vorliegenden Designs sind viele Schritte unternommen worden, um die Güte der Daten(-auswertung) zu steigern. Die Validität des Kodiermanuals wurden in engem Abgleich mit

der Literatur zu MOD entwickelt (Chinn & Brewer, 2001) und umfangreich diskutiert. Hieraus können Validitätsargumente im Sinne der Konstruktvalidität für die Analyse abgeleitet werden. Ebenso werden die MOD mithilfe der Videodaten des lauten Denkens abgeglichen. So soll überprüft werden, inwiefern Aspekte des mentalen Modells durch die MOD externalisiert werden. Damit stellt der Abgleich zwischen MOD und den Videodaten ein zentrales Validitätsargument der vorliegenden Studie dar. Dabei muss jedoch berücksichtigt werden, dass auch die Verbalisierung des mentalen Modells durch das laute Denken als Externalisierung verstanden werden muss, welche ggf. nicht alle Aspekte des mentalen Modells verdeutlicht. Durch den Vergleich der MOD sowie des lauten Denkens werden jedoch zwei unterschiedliche Methoden der Externalisierung genutzt, welche in ihrem Zusammenspiel das mentale Modell bestmöglich abbilden sollen. Weiterhin wurde die InterCoder-Übereinstimmung als Reliabilitätsmaß bestimmt. Dazu wurden unterschiedliche Datensätze bzw. Kategorien innerhalb dieser betrachtet. Die durchgeführten Diskussionen unter den Kodierenden konnte dabei zur Steigerung der InterCoder-Reliabilität genutzt werden. In Bezug auf die Übereinstimmung kodierter Aspekte kann ein weiteres Verbesserungspotential erkannt werden. Zuletzt kann die Pseudonymisierung der Proband*innen als Beitrag zur Auswertungsobjektivität beschrieben werden.

Abweichungen und Ergänzungen zwischen lauten Denken und graphischen MOD

Im Rahmen der Analyse der Videos des lauten Denkens sowie der anschließenden Interviewphase wurden Abweichungen zum erstellten MOD kodiert. Ebenso wurden Aspekte, welche im lauten Denken formuliert, jedoch im MOD nicht beschrieben wurden kodiert. Tabelle 45 gibt einen Überblick über die Deskriptive Statistik der Abweichungen sowie der formulierten, jedoch im MOD nicht beschriebenen Aspekte und Verknüpfungen.

Tabelle 45 – Deskriptive Statistik der Abweichungen sowie Ergänzungen von Verknüpfungen und Aspekten für die beiden Experimente

Unterschiede zum MOD im lauten Denken	Exp. 1		Exp. 2	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Abweichungen bzgl. gebildeter Verknüpfungen	3.33	2.61	1.10	1.45
Abweichungen bzgl. gebildeter essentieller Aspekte	0.00	0.00	.10	.30
Abweichungen bzgl. gebildeter weiterer Aspekte	.10	.30	.14	.36
Abweichungen bzgl. gebildeter inadäquater Aspekte	.19	.68	.10	.30
Abweichungen bzgl. beschriebener anomaler Beobachtungen	0.00	0.00	.05	.22
Formulierte, aber nicht gebildete Verknüpfungen	1.05	1.16	.62	1.02
Formulierte, aber nicht gebildete essentielle Aspekte	.10	.30	.10	.30
Formulierte, aber nicht gebildete weitere Aspekte	.43	.81	.57	.60
Formulierte, aber nicht gebildete inadäquate Aspekte	.05	.22	.19	.51
Formulierte, aber nicht gebildete anomale Beobachtungen	0.00	0.00	.24	.70

Größere Unterschiede zwischen MOD und lautem Denken sind nur für die gebildeten Verknüpfungen zu erkennen. Abweichungen zwischen lautem Denken und MOD könnten dabei auch weiteren Aufschluss über die zu Grunde liegenden epistemologischen Überzeugungen geben. Dabei können für Experiment 1 signifikant mehr Abweichungen beschrieben werden als für Experiment 2 ($t(20)=3.548$, $p=.002$, $d=.774$). Auch für die im Rahmen des lauten Denken formulierten, jedoch im MOD nicht beschriebenen Verknüpfungen zeigt sich ein Unterschied zwischen Experiment 1

und 2. Dieser fällt jedoch nicht signifikant aus ($t(20)=1.253$, $p= .225$). Für alle weiteren Kategorien sind kaum Unterschiede zwischen MOD und dem lauten Denken der Proband*innen zu berichten.

Eine vertiefte Analyse der auftretenden Abweichungen sowie formulierten Aspekte kann weitere Einblicke geben. Dabei können für die unterschiedlichen Verknüpfungsarten verschiedene Unterschiede beschrieben werden.

Die Abweichungen ($N_{\text{Kodierungen_Abweichung_kausal}}=34$) für gebildete **kausale Verknüpfungen** basieren auf unterschiedlichen Begründungen. Dabei kann in der Beschreibung von zeitlichen Abläufen ein Schwerpunkt erkannt werden. Die Proband*innen weichen dabei in ihrer Beschreibung der kausalen Verknüpfungen von dem Ursache-Wirkungs-Zusammenhang ab und stellen zeitlich nacheinander folgende Schritte dar (z.B. MEAM18FA, BAVR30PA, RYER27LA). Dabei kann es sich im MOD hier um inadäquate kausale Verknüpfungen handeln, welche durch die hier beschriebenen Abweichungen ebenfalls mit einer anderen Funktion beschrieben werden. In anderen Fällen wurde die Kausalität der Verknüpfungen innerhalb des lauten Denkens reflektiert. Dabei konnte z.B. die Präzisierung von einzelnen Aspekten im lauten Denken dazu beitragen, dass die kausalen Verknüpfungen korrigiert werden konnten. So haben die Proband*innen durch die Konkretisierung im lauten Denken adäquate kausale Verknüpfungen gebildet, während diese auf Ebene des MOD inadäquat erscheinen (AASA23CA, CAJN05DN). MEAM18FA nutzt die kausalen Verknüpfungen auch zum Einbezug von Vorwissen. Dabei stellt die Probandin heraus: „[...] Eigentlich ist das ja Vorwissen. Ich brauche einen Pfeil für Vorwissen. Ich nehme einfach diesen Pfeil. Ehm... Alle Säuren und, das führt dazu, ja eigentlich. Ich mache das einfach so.“ (MEAM18FA, Pos. 12). Auch KNRN23CA nimmt explizit Bezug auf die Darstellung kausaler Zusammenhänge. Die Probandin beschreibt Unsicherheiten, inwiefern mehrere Beobachtungen innerhalb eines Kästchens im MOD dargestellt werden sollen. Dabei nimmt sie auch Bezug auf kausale Verknüpfungen: „[...] weil ich finde, irgendwie einerseits hängt das zusammen die beiden Teile, weil das ja ein Versuch ist. Andererseits kann man ja dann mit der kausalen Verknüpfung, die wär da ja dann Fehl am Platz.“ (KNRN23CA, Pos. 19). Weitere Abweichungen für kausale Verknüpfungen basieren zumeist auf sehr individuellen Herausforderungen, welche nur für einzelne Proband*innen auftreten.

Für im lauten Denken formulierte, jedoch im MOD nicht dargestellte kausale Verknüpfungen ($N_{\text{Kodierungen_Formuliert_kausal}}=10$) können nur wenige Beispiele gefunden werden. Dabei basieren diese zumeist auf dem Einbezug weiterer Aspekte zur multiplen Begründung von Kausalitäten (SAGT10RL), wie auch postulierten Verknüpfungen, welche jedoch keine Kausalität aufweisen (MNAR08NE, AADK15LE).

Für **induktive Verknüpfungen** können ebenfalls einige Unterschiede zwischen lautem Denken und den MOD erkannt werden. Für die auftretenden Abweichungen zwischen lautem Denken und MOD ($N_{\text{Kodierungen_Abweichung_induktiv}}=17$) können keine eindeutigen Schwerpunkte ausgemacht werden. Dabei kann die von der Definition abweichende Nutzung der Verknüpfungsart gebündelt werden. So werden induktive Verknüpfungen genutzt, um Vorkenntnisse einzubringen (MNAE08NE), zeitliche Abfolgen zu beschreiben (MARF16AN) oder kausale Zusammenhänge zu verdeutlichen (MARF16AN). Andere Abweichungen basieren auf Aspekten der Verallgemeinerung. So weichen die Proband*innen teils im lauten Denken vom Umfang der Verallgemeinerung im MOD ab. So werden im lauten Denken weitere inhaltliche Aspekte in die Verallgemeinerung einbezogen (MERF13KN, AASA23CA). GAWI21JA hingegen nutzt die induktive Verknüpfung eher, um Ähnlichkeiten deutlich zu machen: „dass das dasselbe ist“ (Pos. 2). Weiterhin können Abweichungen durch das nachträgliche Verändern unterschiedlicher inhaltlicher Aspekte erkannt werden. Im Zusammenhang mit der jeweiligen Verknüpfungen erfolgen hier teils Veränderungen/Konkretisierungen, die sich auch auf die Verknüpfungen ausweiten können (BAVR30PA, MNAE08NE).

Unterschiede bzgl. im lauten Denken zusätzlich formulierter Verknüpfungen ($N_{\text{Kodierungen_Formuliert_induktiv}}=6$) können teilweise ebenfalls auf die nicht definitionsgemäße Nutzung induktiver Verknüpfungen zurückgeführt werden. So beschreiben die Proband*innen teils induktive Verknüpfungen, welche sie jedoch innerhalb der MOD z.B. als Analogie-Verknüpfung oder kausale Verknüpfung darstellen (MASO20CA). Bei GAWI21JA kann hingegen die Formulierung einer induktiven Verknüpfung erkannt werden, welche später zu einer Analogie-Verknüpfung verändert wird.

Die Unterschiede im Einsatz von **Analogie-Verknüpfungen** in den MOD sowie dem lauten Denken ähneln denen der induktiven Verknüpfungen. So sind Schwerpunkte für die Abweichungen ($N_{\text{Kodierungen_Abweichung_Analogie}}=18$) z.B. für die Verwechslung mit anderen Verknüpfungsarten (AAVR08LA, AADK15LA). Ein Schwerpunkt ist dabei bei der Verwechslung mit der induktiven Verknüpfung zu erkennen (AEDL02JA, MNAE08NE, MASO20CA). Weiterhin kann für zwei Proband*innen die Verwendung der Analogie-Verknüpfung im Sinne von der Darstellung von Ähnlichkeiten erkannt werden. So beschreibt AASA23CA (Pos. 20) im Interview-Teil: „Ehm, ich glaube ich hab das eher dann damit verwendet, dass man, also ich glaube ich habe mich da eher aus dieses Analogie bezogen, dass quasi die Färbung analog für einen pH-Wert steht. Also dass der Begriff gleichbedeutend ist [...]“. Weitere Abweichungen sind. z.B. für die Veränderung von Aspekten (RYER27LA) sowie weitere individuelle Unterschiede zu beobachten.

Für im lauten Denken zusätzlich formulierte Verknüpfungen ($N_{\text{Kodierungen_Formuliert_Analogie}}=12$) ist ebenfalls ein Schwerpunkt auf dem Vertauschen mit anderen Verknüpfungsarten zu erkennen (z.B. KNRN23CA, MARF16AN, IABD29MM). Weiterhin zeigt sich die Nutzung zur Formulierung von Zusammenhängen (AAVR08LA) oder die Formulierung neuer, adäquater Analogie-Verknüpfungen (IABD29MM).

Ein Schwerpunkt der Abweichungen für **kontrastierende Verknüpfungen** ($N_{\text{Kodierungen_Abweichung_kontrastierend}}=10$) kann für die Verwechslung mit Analogie-Verknüpfungen beschrieben werden (AAVR08LA, IABD29MM). Dabei werden hier kontrastierende Verknüpfungen im MOD dargestellt, während eine Analogie-Verknüpfung im lauten Denken formuliert wird. Ebenso können kontrastierende Verknüpfungen beschrieben werden, welche aufgrund der alleinigen Darstellung im MOD nicht nachvollzogen werden können, weil z.B. das Unterscheidungskriterium unklar ist. Im lauten Denken hingegen könnten diese Abweichungen zum Verständnis der Verknüpfungen beitragen (GAWI21JA, SETS14KA).

Für kontrastierende Verknüpfungen können nur wenige formulierte und nicht im MOD dargestellte Aspekte beschrieben werden ($N_{\text{Kodierungen_Formuliert_kontrastierend}}=4$). Dabei kann hier vor allem auf unklare Darstellungen (AASA23CA) sowie ein Vertauschen der Darstellungsarten der Verknüpfungen (AAVR08LA) abgestellt werden. Hier werden teils Verknüpfungen im MOD gebildet, welche jedoch später wieder gelöscht werden und somit im finalen MOD nicht vorzufinden sind.

Zuletzt können auch Unterschiede für **unmögliche kausale Verknüpfungen** formuliert werden. Für die Abweichungen ($N_{\text{Kodierungen_Abweichung_unmöglich_kausal}}=13$) zwischen MOD und lautem Denken könnten dabei unterschiedliche Schwerpunkte beschrieben werden. Dabei kann zuerst ein fehlender Ursache-Wirkungs-Zusammenhang im MOD beschrieben werden, welcher jedoch im Rahmen des lauten Denken erklärt werden kann (AAVR08LA, MARF16AN). Ebenso tritt im Rahmen dieser Verknüpfungsart eine doppelte Verneinung auf. So werden Aspekte verknüpft, welche verneint sind, während die Verknüpfung an sich ebenfalls eine ‚Verneinung‘ im Sinne der Unmöglichkeit der Verknüpfung beinhaltet (MASO20CA, SAGT10RL). Dabei können auch Abweichung bzgl. der Richtung der Verknüpfungen berichtet werden (MASO20CA, IAKS30JA).

Unterschiede basierend auf im lauten Denken formulierten, jedoch im MOD nicht dargestellten Verknüpfungen sind nur sehr vereinzelt zu erkennen ($N_{\text{Kodierungen_Formuliert_unmöglich_kausal}}=3$). Für KNRN23CA kann eine formulierte Verknüpfung erkannt werden, die jedoch auf Grundlage der unkonkreten Formulierung inadäquat ausfällt. AAVR08LA stellt eine unmögliche kausale Verknüpfung stellvertretend für den Zusammenhang mehrerer Teilexperimente dar. Ansätze einer (auch) kontrastierenden Verknüpfung sind in der durch AEDL02JA formulierten unmöglichen kausalen Verknüpfung zu erkennen.

7.7 Diskussion der Ergebnisse der Modellierungsstudie

Die Diskussion der Ergebnisse dieser Studie wird entlang drei Aspekten gegliedert. In einem ersten Abschnitt soll die Rolle von Fachwissen für den Umgang mit anomalen Daten beleuchtet werden. Anschließend erfolgt ein Fokus auf zu Grunde liegende epistemologische Überzeugungen, bevor die Validität der Methode der *Model-of-data* (MOD) diskutiert werden soll. Die Diskussion schließt mit einem Fazit.

Für eine Beleuchtung des Einflusses von Fachwissen auf den Umgang mit anomalen Daten können in einem ersten Schritt die gebildeten inhaltlichen Aspekte untersucht werden. Im Bereich der formulierten inhaltlichen Aspekte ist eine Diskrepanz zwischen der Anzahl gebildeter Beobachtungs- und Konzeptaspekte zu erkennen. Dies gilt sowohl für essentielle, als auch weitere Aspekte. Der Unterschied zwischen den beiden Arten gebildeter Aspekte kann ggf. durch das Studiendesign erklärt werden. Die Beobachtungsaspekte konnten direkt aus den Experimentiervideos abgeleitet werden. Neben einer weitestgehenden Abdeckung der essentiellen Beobachtungsaspekte konnten auch einige weitere Aspekte gebildet werden. Die Konzeptaspekte hingegen mussten teils aus dem Hintergrundwissen, teils aus der Kombination bestimmter Beobachtungsaspekte mit Zusatzinformationen (welche teils im Video angegeben wurden) verknüpft werden. Dabei müssen teils unterschiedliche Verknüpfungsarten eingesetzt werden, um einen Konzeptaspekt begründen zu können. Damit stellt die Konstruktion von Konzeptaspekten höhere Anforderungen an die Proband*innen. Gleichzeitig ist die Bildung von Konzeptaspekten jedoch nötig, um eine Verbindung der Beobachtungen mit den konzeptuellen Vorstellungen zu erreichen. Im Sinne des *Model of argument* nach (Toulmin, 2003) könnten Konzeptaspekt z.B. als *warrant* oder *backing* in die Begründungsprozesse einfließen.

Die Unterschiede in der Anzahl der gebildeten Aspekte könnte sich dabei aus den Experimenten ableiten lassen. So ist der Umfang der beiden Expert*innen-MOD ebenfalls unterschiedlich. Dies könnte in dem unterschiedlichen Umfang der Experimente begründet werden. Während Experiment 1 drei Teilexperimente umfasst, fokussiert Experiment 2 lediglich auf die Prozesse des Erhitzens und des (luftdichten) Abkühlens von Wasser.

Wie beschrieben wurde, wurden durch die Proband*innen innerhalb der MOD deutlich mehr Beobachtungs- als Konzeptaspekte beschrieben. Dies könnte auf ein geringes Konzeptwissen zu den jeweiligen Phänomenen hindeuten. In diesem Zusammenhang sollten auch die Ergebnisse des Fachwissenstests beachtet werden. Im Fachwissen für die Säure-Base-Skala sind Förderbedarfe erkennbar. Dabei beziehen sich diese auf generelle Fachwissensaspekte zum Inhaltsbereich Säure-Base. Es könnte somit auch angenommen werden, dass Fachwissen zur Säurestärke nicht vorhanden ist. Dies könnte durch die ausbleibende Deutung der Beobachtungen innerhalb der MOD unterstützt werden. Für die Siedepunkt-Skala können aufgrund der kritischen internen Konsistenz keine Aussagen zum Fachwissen getroffen werden. Auf Grundlage der MOD sowie der gebildeten Konzeptaspekte lässt sich hier jedoch ebenfalls ein Förderbedarf von Fachwissen zum Thema

Siedepunkt erwarten. Auch für Exp. 2 bleibt entsprechend eine Deutung der Beobachtungen aus. In Anlehnung an diSessa (2002) könnten vorliegende Fachwissensaspekte als *p-prims* getrennt voneinander vorliegen. Eine Erklärung durch die Verknüpfung von *p-prims* in *coordination classes* scheint nicht möglich zu sein. Ebenso könnte das Fehlen von Aspekten des Konzeptwissens die Erklärung erschweren oder ganz verhindern.

Das Fehlen von Fachwissen könnte auch Einfluss auf die Verarbeitung anomaler Beobachtungen nehmen. Für Experiment 2 müssten die Proband*innen Kenntnisse über die Druckabhängigkeit der Siedetemperatur und den Dampfdruck besitzen, um die anomalen Beobachtungen erklären zu können. Für Experiment 1 sind Kenntnisse über die Säure-Stärke nötig. Beide Inhaltsbereiche stellen dabei komplexe Fachwissensbereiche dar, welche nicht ohne Weiteres aus dem Experiment abgeleitet werden können.

In Hinblick auf die Wahrnehmung anomaler Beobachtungen könnten deutliche Unterschiede zwischen Experiment 1 und 2 beobachtet werden. Während die Proband*innen innerhalb von Experiment 1 eine Vielzahl von anomalen Beobachtungen beschreiben, können innerhalb von Experiment 2 nur wenige Beobachtungen beschrieben werden. Hier könnten mehrere Ursachen zu Grunde liegen.

Aus einer fachdidaktischen Perspektive könnten ebenfalls die erwarteten Vorstellungen einen Einfluss auf die Wahrnehmung von anomalen Beobachtungen nehmen. Chinn und Malhotra (2002a) beschreiben die große Bedeutung bestehender Vorstellungen für die Beobachtungen von Phänomenen. Das Säure-Base-Konzept aus Experiment 1 stellt ein fachchemisches Konzept dar, welches den Proband*innen im Rahmen des Schulunterrichts sowie Lehrveranstaltungen an der Universität begegnet sein könnte. Vorstellungen der Proband*innen könnten auch durch den Fachunterricht geprägt werden und nicht der Lebenswelt der Proband*innen entstammen. Für Experiment 2 können jedoch Vorstellungen zum Sieden von Wasser erwartet werden, welche der Lebenswelt entspringen. Dabei findet eine Konfrontation mit dem Konzept des Siedepunktes von Wasser deutlich früher statt und könnte so zu stabileren Vorstellungen führen. In Anlehnung an diSessa (2002) könnten hier bereits stabile Erklärungsnetze bestehen, welche den Siedepunkt von Wasser bei 100°C beinhalten. Auch könnte der Begriff des Siedens oder auch das umgangssprachliche ‚Kochen‘ eine klare ontologische Kategorie darstellen (Chi, 2008). Damit könnten die Vorstellungen zu diesem Inhaltsfeld derart gefestigt sein, dass die Beobachtung aus Perspektive der eigenen Vorstellungen erfolgt. Zwar könnte dies auch für Experiment 1 unterstellt werden, jedoch könnte hier eine andere Stabilität der Vorstellungen vermutet werden. Derart gefestigte Vorstellungen könnten im Sinne der Theoriegeladenheit Einfluss auf die Beobachtung nehmen (Chinn & Malhotra, 2002a).

Ein weiteres Indiz für diese Erklärung könnte in den unterschiedlichen Umgangsarten für Experiment 1 und 2 erkannt werden. Innerhalb von Experiment 1 werden die anomalen Beobachtungen in das Modell aufgenommen und in ca. 50% der Fälle findet eine Theoriebearbeitung statt. Wenn gleich der Umfang dieser nicht immer klar definiert werden kann, findet eine Bearbeitung statt. Dies könnte darauf hindeuten, dass sich das bestehende Konzept durch die Beobachtungen verändern lässt. Für Experiment 2 hingegen erkennen nur wenige Proband*innen überhaupt anomale Beobachtungen. Einige der anomalen Beobachtungen werden dabei ignoriert (vgl. Kap. 7.6.2). Das Ignorieren lässt nach Chinn und Malhotra (2002a, S. 328) auf ein „irrational nonalignment“ schließen, welches auf Grund des Festhaltens an den bestehenden Vorstellungen, dass Wasser immer bei 100°C siedet, valide Beobachtungen außer Acht lässt. Damit könnte das Auftreten des Ignorierens von Beobachtungen im Rahmen von Experiment 2 für einer deutlichere Theoriegeladenheit der Beobachtung schließen lassen. Somit könnten die Vorstellungen zu Experiment 2 stabiler sein,

als die für Experiment 1. Wenngleich – wie beschrieben – die geringe Anzahl von Konzeptaspekten keine vertieften Einblicke in die Vorstellungen der Proband*innen zulässt, könnte hier ein deutlicher Einfluss dieser vermutet werden. So könnten auch lediglich singuläre Vorstellungen zum Siedepunkt von Wasser vorliegen und keine weiteren Vorstellungen, welche in Form von Konzeptaspekten in das MOD integriert werden könnten.

Das Erkennen oder Nicht-Erkennen anomaler Beobachtungen könnte auch durch Aufmerksamkeitslenkung innerhalb der Experimentier-Videos entstanden sein. Hier wurden die Untersuchungsvariablen durch Exponierung präsentiert. Es bedarf – wie beschrieben – der Kombination der Beobachtung der Untersuchungsvariablen sowie der makroskopischen Beobachtungen im Experiment, um die anomalen Beobachtungen wahrzunehmen. Dabei stellt in Experiment 1 vor allem der Experimentierprozess die relevanten Aspekte dar. Die Proband*innen müssen erkennen, dass verschiedene Säuren mit der gleichen Lauge, gleicher Konzentration versetzt wurden. In Kombination mit der Färbung des Indikators (als Untersuchungsvariable) können relevante Beobachtungen getätigt werden, welche die intendierten anomalen Beobachtungen darstellen. Innerhalb von Experiment 2 sind relevante Kontrollvariablen ebenfalls durch den Experimentierprozess zu erkennen. Es handelt sich um den Umgebungsdruck, welcher jedoch nur indirekt durch das Verschließen des Rundkolbens beeinflusst wird. Folglich bedarf es der Übertragung des Experimentierschrittes „Rundkolben wird verschlossen“ in dessen konzeptuellen Einfluss auf den Versuch im Sinne der Veränderung des Drucks. Hier könnte weiteres Konzeptwissen nötig sein, um den Druck als relevante Kontrollvariable zu erkennen. Im Experimentiervideo wird der Druck nicht thematisiert. Hier zeigt sich die notwendige Verknüpfung von der Beobachtung der Untersuchungsvariablen sowie den makroskopischen Beobachtungen (Experiment 1) bzw. impliziten Veränderungen am System (Experiment 2). Die fehlende Verknüpfung zwischen diesen beiden Beobachtungen könnte zum Nicht-Erkennen von anomalen Beobachtungen führen. Entsprechend kann für Experiment 1 ein direkterer Zugang zu den anomalen Beobachtungen im Vergleich zu Experiment 2 beschrieben werden.

In Rückbezug auf Chinn und Brewer (1993b, 1998) können in der vorliegenden Studie nicht alle postulierten Umgangsarten mit anomalen Beobachtungen gefunden werden. So können die Umgangsarten *Rejection* und *Uncertainty* nicht beschrieben werden. Bei der Umgangsart *Rejection* werden Gründe für das Ausschließen der anomalen Beobachtungen aus der Erklärung angegeben. Im Sinne der unterrepräsentierten Konzeptaspekte innerhalb der vorliegenden Studie könnten hier konzeptuelle Gründe für den Ausschluss der Beobachtungen vermutet werden, welche jedoch nicht konkretisiert werden. So wäre es denkbar, dass innerhalb der mentalen Modelle der Proband*innen durchaus Begründungen zur Ablehnung der Beobachtungen stattfinden, welche jedoch nicht externalisiert werden. In diesem Zusammenhang könnten die Proband*innen über mehr Fachwissen verfügen, als dies durch die MOD und den Fachwissenstest erhoben werden kann. Ggf. könnten einige Proband*innen Gründe für das Ignorieren anomaler Beobachtungen haben, welche nicht geäußert werden. Andererseits könnte das Fehlen von Fachwissen auch zu einer fehlenden Wahrnehmung von Beobachtungen als anomal führen. Somit könnte hier eine fehlende Unterscheidbarkeit der Umgangsarten *Ignoring* und *Rejection* vermutet werden. Dabei ergeben sich forschungsmethodologische Herausforderungen dieser Unterscheidung weiter nachzugehen. So müsste z.B. für die Nachfrage nach Gründen des Ausschlusses anomaler Beobachtungen innerhalb eines Interviews eine Beobachtung zuerst als anomal bezeichnet werden. Dabei kann die Bezeichnung von Beobachtungen als anomal nur vor dem Hintergrund der individuellen Vorstellung bedeutsam werden. Durch die Beschreibung der Beobachtungen als anomal könnte dabei wiederum Einfluss auf den Umgang mit diesen genommen werden. So könnten Interviewende als epistemologische Autoritäten wahrgenommen werden (vgl. Dimension *Source of*

knowledge). Somit könnten epistemologische Überzeugungen aktiviert werden, welche die Verarbeitung von anomalen Beobachtungen behindern. Sofern die Proband*innen anomale Beobachtungen als solche benennen, kann die Umgangsart des *Ignoring* ggf. nicht mehr beobachtet werden. Ggf. könnte eine Zusammenfassung der beiden Kategorien eine bessere Unterscheidbarkeit erreichen. Eine Bündelung könnte unter dem Begriff ‚Ausbleibende Beschreibung anomaler Beobachtungen‘ erfolgen. So könnten die beiden Umgangsarten *Ignoring* und *Rejection* zusammengefasst werden.

Auch für andere Umgangsarten ergeben sich teils Herausforderungen bei der Unterscheidung. Somit könnten weitere Bündelungen von Umgangsarten zur besseren Unterscheidung eingesetzt werden. Tabelle 46 gibt einen Überblick über die vorgeschlagene Bündelung unterschiedlicher Umgangsarten in Anlehnung an Chinn und Brewer (1993b, 1998).

Tabelle 46 – Bündelung der Umgangsarten mit anomalen Daten in Anlehnung an Chinn und Brewer (1993b, 1998)

Umgangsarten nach Chinn und Brewer (1993b, 1998)	Vorgeschlagene Bündelung durch den Autor
Ignoring Rejection	Ausbleibende Beschreibung anomaler Beobachtungen
Uncertainty	Uncertainty
Exclusion Abeyance	Ausbleibende Erklärung anomaler Beobachtungen
Reinterpretation	Reinterpretation
Peripheral theory change Theory change	Theoriebearbeitung

Wie beschrieben, ergibt sich für die Unterscheidung zwischen *Exclusion* und *Abeyance* eine Herausforderung die jeweilige Begründung einer ausbleibenden Erklärung zu differenzieren. Während die *Exclusion* auf dem Ausschluss der Beobachtungen aus dem Geltungsbereich der bestehenden Vorstellungen beruht, wird die Erklärbarkeit der Beobachtungen durch die Vorstellungen innerhalb der *Abeyance* für einen späteren Zeitpunkt für möglich erachtet. Ohne eine Begründung der ausbleibenden Erklärung können diese Umgangsarten jedoch nicht unterschieden werden. Im Rahmen von MOD haben die Proband*innen innerhalb der vorliegenden Studie keine entsprechenden Erläuterungen gegeben. Entsprechend könnten diese Umgangsarten zur Kategorie ‚Ausbleibende Erklärung anomaler Daten‘ zusammengefasst werden. Auch für diese Umgangsarten könnte fehlendes Fachwissen zur Argumentation herangezogen werden. Dabei deutet vor allem die Umgangsart *Abeyance* auf fehlendes Fachwissen hin, welches so auch wahrgenommen wird. So wird beschrieben, dass die Beobachtungen erklärt werden können, jedoch zurzeit noch zu wenig Fachwissen vorliegt bzw. die Theorie noch nicht weit genug ‚entwickelt‘ sind.

Auf Ebene epistemologischer Überzeugungen können hier ähnliche Überzeugungen erwartet werden. So scheint eine Relevanz der Beobachtungen wahrgenommen zu werden, damit diese in das

MOD integriert werden. Damit kann ein Bezug zur Dimension *Simplicity of knowledge* erkannt werden. Ein einfaches Ignorieren der Beobachtungen ist nicht möglich. Wenngleich keine Erklärung der Beobachtungen stattfindet, werden diese jedoch als relevant erachtet und in das MOD aufgenommen. Damit könnte die Komplexität von Wissen anerkannt werden, da auch widersprüchliche Beobachtungen eine Berechtigung haben in das MOD aufgenommen zu werden. Dabei können keine Unterschiede zwischen den beiden Umgangsarten festgestellt werden, was mit der Behandlung der anomalen Beobachtungen innerhalb der MOD dieser Studie übereinstimmt.

Eine Unterscheidung der Umgangsarten *Peripheral theory change* und *Theory change* kann auf Grundlage, der nur wenig gebildeten Konzeptaspekte innerhalb der MOD nur schwer erreicht werden. Die bestehenden Vorstellungen der Proband*innen werden aufgrund der nicht gebildeten Konzeptaspekte nicht deutlich. Damit können die Abwandlungen der Theorie durch die Verarbeitung der anomalen Beobachtungen hier nur zusammenfassend als Theoriebearbeitung beschrieben werden. Auch hier könnte entsprechend eine Bündelung zur Kategorie ‚Theoriebearbeitung‘ empfohlen werden. Auch auf Grundlage wissenschaftsphilosophischer Ansätze lässt sich keine einfache Unterscheidung zwischen Veränderungen am *protective belt* oder dem *hard core* der Vorstellungen der Proband*innen formulieren (Lakatos, 1978). So müsste ein umfassender Einblick in die Vorstellungen der Proband*innen erreicht werden, um den Grad der Theoriebearbeitung zu bewerten. Dabei stellt sich ebenfalls die Frage, wie mit fehlendem Fachwissen umgegangen werden sollte. Fachwissen könnte durch die Deutung der Beobachtungen aufgebaut werden. Inwiefern hier jedoch eine periphere oder umfängliche Theoriebearbeitung stattfindet, kann nicht endgültig geklärt werden. Entsprechend könnte auch dieser Fall unter der gebündelten Kategorie ‚Theoriebearbeitung‘ subsumiert werden.

Auch hier können große Gemeinsamkeiten der beiden gebündelten Umgangsarten auf Ebene der epistemologischen Überzeugungen beschrieben werden. Neben Überzeugungen zur Dimension *Simplicity of knowledge* scheint auch die Dimension *Justification of knowing* Relevanz zu besitzen. So werden die anomalen Beobachtungen hier auch in die theoretischen Vorstellungen der Proband*innen integriert. Damit scheint ein Bedarf nach der Begründung theoretischer Annahmen durch empirische Daten unterstrichen zu werden. Ebenso bedarf die Theoriebearbeitung Überzeugungen zur Dimension *Certainty of knowledge*. Die Proband*innen nehmen hier aktiv Veränderungen an bestehenden theoretischen Vorstellungen vor. Dafür sind Überzeugungen zur grundsätzlichen Veränderbarkeit von Wissen nötig.

Neben den beschriebenen Bündelungen könnten die Umgangsarten *Uncertainty* sowie *Reinterpretation* als einzelne Kategorien beibehalten werden. Zwar konnten innerhalb der vorliegenden Studie keine Fälle von *Uncertainty* beobachtet werden, jedoch scheint sich diese Umgangsart von den beschriebenen Bündelungen zu unterscheiden. Auf Ebene epistemologischer Überzeugungen können hier vor allem Aspekte der Dimension *Source of knowledge* erkannt werden. So wird die grundsätzliche Glaubwürdigkeit der Daten angezweifelt. Für die Umgangsart *Reinterpretation* können in der vorliegenden Studie Beispiele erkannt werden. Hier geben die Proband*innen Erklärungen für die anomalen Beobachtungen an, welche jedoch nicht mit einer Theoriebearbeitung einhergehen. Entsprechend könnte auch diese Umgangsart als Kategorie weiterhin Bestand haben. Im Unterschied zur Bündelung der Umgangsarten ‚Theoriebearbeitung‘ könnten hier Überzeugungen zur Dimension *Certainty of knowledge* vorliegen, welche eher von unveränderbarem Wissen ausgehen. Weitere epistemologische Überzeugungen könnten sich denen der ‚Theoriebearbeitung‘ ähneln.

Zwar führt die Bündelung der Umgangsarten zu einer weniger differenzierten Betrachtung, jedoch können auf Grundlage der bestehenden Daten weitere Differenzierungen nicht vorgenommen

werden. Im Rahmen zukünftiger Studien könnte die Bündelung der Umgangsarten forschungsmethodologische Strukturierung ermöglicht werden. Dies meint, dass eine Einordnung in die beschriebenen Bündelungen im Rahmen der Datenerhebung eine tiefergehende Befragung der Proband*innen ermöglichen könnte. So könnte die grundsätzliche Umgangsart in Form der Bündelungen analysiert werden, bevor dann eine weitere Ausdifferenzierung stattfinden kann. So könnten Proband*innen, welche Beobachtungen ignorieren z.B. explizit nach den Beobachtungen gefragt werden, ohne dass diese als anomal beschrieben werden müssten, da die grundsätzliche Umgangsart bereits bekannt ist. Gleiches gilt für die ausbleibende Erklärung, etc. Die Einordnung in Bündelungen könnte unterschiedliche Zugänge und Interviewanlässe für zukünftige Studien strukturieren. Auf Grundlage von den durch die Proband*innen erstellten MOD könnte eine differenziertere Betrachtung der Umgangsarten ggf. nicht möglich sein. Auch scheinen die durch Chinn und Brewer (2001) beschriebenen Zusammenhänge der Umgangsarten mit direkten Verknüpfungen innerhalb des MOD nicht der Komplexität dieser gerecht zu werden. So beschreiben die Autoren z.B., dass die Umgangsart *Ignoring* mit der ausbleibenden Konstruktion eines MOD einhergeht. Jedoch stellen anomale Beobachtungen nur einzelne Aspekte innerhalb der durch die Proband*innen gestalteten MOD dar. Somit stellen die Proband*innen umfassende MOD dar, welche ggf. lediglich keinen Bezug zur den anomalen Beobachtungen aufweisen. Entsprechend wurde für die Auswertung dieser Studie die Rahmentheorie der Autoren (1993b, 1998) zugrunde gelegt und nicht die lockere Anknüpfung an die MOD-Theorie (Chinn & Brewer, 2001).

Zusammenfassend kann die Nützlichkeit der Taxonomie von Chinn und Brewer (1993b, 1998) zur Beschreibung des Umgangs mit anomalen Beobachtungen beschrieben werden. Wenngleich auf der bestehenden Datengrundlage eher die Zuordnung zu verschiedenen Bündelungen der Umgangsarten erfolgen kann, scheint die Taxonomie angemessen zu sein. Die Bündelung kann dabei theoretisch auch durch Unterschiede in den epistemologischen Überzeugungen unterstützt werden. Innerhalb der MOD fehlen jedoch vertiefte Einblicke in das Konzeptwissen der Proband*innen, um eine genauere Differenzierung z.B. für die Theoriebearbeitung vornehmen zu können. Auch der Fachwissenstest kann die individuellen Vorstellungen nicht so umfangreich auflösen, als dass z.B. eine Unterscheidung zwischen einem *Peripheral theory change* sowie einem *Theory change* möglich würde.

Neben dem Fachwissen kann durch die Ergebnisse dieser Studie auch ein Einfluss epistemologischer Überzeugungen auf den Umgang mit anomalen Beobachtungen erkannt werden. Wie gezeigt werden konnte, spielt dabei vor allem die Konstruktion von Verknüpfungen innerhalb der MOD eine entscheidende Rolle. Vertiefte Einblicke in die vorliegenden epistemologischen Überzeugungen der Proband*innen können vor allem auch durch die inadäquat gebildeten Verknüpfungen erreicht werden.

Kausale Verknüpfungen stellen die am häufigsten genutzte Verknüpfungsart in den gebildeten MOD dar. In Rückbezug zu der postulierten Komplexität der unterschiedlichen Verknüpfungsarten (Kap. 7.1.2) könnte hier die geringere Anforderung zur Erstellung kausaler Verknüpfungen unterstützt werden. Der geringere Grad der Abstraktheit könnte hier zu einem größeren Anteil adäquater Verknüpfungen führen. Auch aus Perspektive epistemologischer Überzeugungen könnte sich die häufige Nutzung dieser Verknüpfungsarten erklären lassen. Entsprechend könnten Grundlegende epistemologische Überzeugungen zu den *Dimensionen Simplicity of knowledge* und *Certainty of knowledge* bei den Proband*innen vorliegen. Dabei zeigen sich in der inadäquaten Nutzung kausaler Verknüpfungen auch Abweichungen von der eigentlichen epistemologischen Funktion der Verknüpfungsart. Vor allem die Nutzung kausaler Verknüpfungen zur Darstellung zeitlicher Abfolgen lässt ein fehlendes Verständnis des Ursache-Wirkungs-Zusammenhangs der

Verknüpfungsart erkennen. Dabei müssen hier nicht zwingend naive epistemologische Überzeugungen zu Grunde liegen. Wie teils durch die Proband*innen im lauten Denken formuliert, wurde die kausale Verknüpfung für zeitliche Abfolgen genutzt, obwohl den Proband*innen klar war, dass es sich nicht um eine kausale Verknüpfung handelt. Das Fehlen einer Verknüpfungsart für zeitliche Abfolgen könnte so zu der nicht definitionsgemäßen Nutzung kausaler Verknüpfungen geführt haben.

Innerhalb der MOD können Aspekte auch nebeneinander bzw. untereinander eingesetzt werden. Dabei wird jedoch weniger eine zeitliche Abfolge der Aspekte deutlich als mehr das Vorliegen der Aspekte, welche dann gemeinsam verknüpft werden können. Aufgrund dieses Zusammenhangs, wie auch dem formulierten Bedarf nach einer Verknüpfungsart für zeitliche Abfolgen, sollte eine weitere Verknüpfungsart in die MOD aufgenommen werden. Durch die Aufnahme einer Verknüpfung für zeitliche Abfolgen könnte so ggf. eine Nutzung kausaler Verknüpfungen entgegen ihrer epistemologischen Ausrichtung verhindert werden.

Für kontrastierende Verknüpfungen konnten ähnliche Herausforderungen berichtet werden, wie für die kausalen Verknüpfungen. Wie im Kapitel 7.1.2 beschrieben, sind grundlegende Überzeugungen der Dimension *Justification of knowledge* nötig. Dabei können in den inadäquat gebildeten Verknüpfungen innerhalb dieser Studie Herausforderungen in der Begründung der Kontraste erkannt werden. So fehlen teilweise oberflächliche oder tiefenstrukturelle Begründungen für auftretende Kontraste. Dabei kann nicht abschließend geklärt werden, ob die Inadäquatheit dieser Verknüpfungen durch die epistemologischen Überzeugungen begründet werden kann oder ob z.B. Fachwissen fehlt, um tiefenstrukturelle Kontraste aufzuzeigen. Insgesamt können nur wenige kontrastierende Verknüpfungen als inadäquat bezeichnet werden. Folglich kann der geringere Grad der Anforderung wie in Kap. 7.5.1 postuliert unterstützt werden.

Die seltenere Bildung induktiver sowie von Analogie-Verknüpfungen könnte mit dem höheren Grad der Abstraktheit und daraus resultierenden höheren Anforderungen an die Proband*innen begründet werden. Die Nutzung von Verallgemeinerungen bedarf dem Vergleich der beobachteten Entitäten mit solchen aus ontologisch unterschiedlichen Kategorien. So könnte aus den Reaktionen von Salz- und Salpetersäure mit Natronlauge in Exp. 1 z.B. induktiv verallgemeinert werden, dass die Reaktion von Säuren mit Natronlauge gleicher Konzentration und Volumina immer zu einem pH-Wert von 7 führen (vgl. Expert*innen-MOD Experiment 1 im Anhang). Dabei findet eine Verallgemeinerung der beiden Einzelfälle Salz- und Salpetersäure auf die Kategorie aller Säuren statt, welche jedoch durch das Teilexperiment mit Essigsäure widerlegt wird. Derartige Verallgemeinerungen können in vielen MOD der Proband*innen erkannt werden. Zur Bildung dieser Verallgemeinerung muss eine ontologische Kategorie ausgewählt werden, welche ausreichend Ähnlichkeiten und Unterschiede bietet, damit die Verallgemeinerung Bestand hat. Dabei könnte die Herausforderung zur Bildung von Analogie-Verknüpfungen aufgrund der deutlich unterschiedlichen ontologischen Kategorien für die Verallgemeinerung noch größer sein. Indizien für diese Interpretation können in der selteneren Erstellung von Analogie-Verknüpfungen erkannt werden. Entsprechend könnten hier erweiterte Überzeugungen zur Dimension *Justification of knowledge* nötig werden (vgl. Kap. 7.1.2). Es müssen Argumente für die Unterschiedlichkeit bei gleichzeitiger Ähnlichkeit der Entitäten in den jeweiligen ontologischen Kategorien gebildet werden.

Dabei stellt die Verknüpfung mit anderen ontologischen Kategorien eine Herausforderung dar, welche schon Chi (2008) in ihrem Verständnis von *Conceptual Change* Prozessen beschreibt. Die Autorin beschreibt die nicht angemessene Sortierung von Entitäten in ontologische Kategorien als Ausgangspunkt für einen kognitiven Konflikt und einen notwendigen *Conceptual Change*. Die Proband*innen müssen für das Aufstellen von Verallgemeinerungen zuerst die beobachteten

Entitäten in eine ontologische Kategorie zuordnen und anschließend eine davon unterschiedliche Kategorie als Verallgemeinerungsbereich auswählen. Dabei müssen neben den notwendigen ontologischen Unterschieden noch ausreichend Vergleichsmerkmale bestehen, damit die Verallgemeinerung Gültigkeit besitzen kann.

Im Rahmen der vorliegenden Studie könnte in Experiment 1 eine zu grobe Zuordnung zu ontologischen Kategorien die Verarbeitung der anomalen Beobachtungen erschweren. So ordnen die Proband*innen alle Säuren der Kategorie ‚Säuren‘ zu, ohne zwischen starken und schwachen Säure zu unterteilen. Hier muss jedoch zuerst fehlendes Fachwissen als Herausforderung beschrieben werden. So könnte fehlende Kenntnis über die Säurestärke die ontologische Zuordnung in die jeweiligen Kategorien verhindern. Andererseits kann die nicht angemessene Zuordnung in ontologische Kategorien auch einen Teil der inadäquat gebildeten Verknüpfungen erklären. So kann z.B. eine Verwechslung zwischen induktiven und Analogie-Verknüpfungen erkannt werden, welche hauptsächlich auf der Kategorienzuordnung beruht. Auch der unklare bzw. unpassende Einbezug inhaltliche Aspekte in die Verallgemeinerungen könnte auf Herausforderungen in der ontologischen Zuordnung der Aspekte hinweisen.

Weiterhin konnte in Kap. 7.1.2 die Bedeutung der Dimension *Simplicity of knowledge* dargelegt werden. In Bezug auf die geringe Anzahl gebildeter Verallgemeinerungen innerhalb der MOD könnten hier teils naivere Überzeugungen bei den Proband*innen vorliegen. So könnte ein Verständnis von einfachem, unvernetztem Wissen die Erstellung von Verallgemeinerungen generell gehemmt haben. Dabei könnte ein Indiz für derartige Überzeugungen auch in der hohen Anzahl gebildeter kausaler Verknüpfungen erkannt werden, welche vor allem lineare Verknüpfungen und damit weniger vernetztes Wissen forcieren.

Wie gezeigt werden konnte, basieren die induktive wie auch die Analogie-Verknüpfung auf ähnlichen epistemologischen Überzeugungen. Lediglich die Unterschiedlichkeit der gewählten ontologischen Kategorien stellt einen Entscheidungspunkt dar. Dabei muss jedoch auch die Rolle des Fachwissens beachtet werden. So könnte bei fehlendem Fachwissen die gewählte Kategorisierung vor dem individuellen konzeptuellen Verständnis zutreffend sein, während sie vor dem Hintergrund wissenschaftlich anerkannter Evidenz inadäquat ist. Da es jedoch auf Grundlage der epistemologischen Überzeugungen nur wenig Unterschiede der Verknüpfungsarten gibt, könnte eine Zusammenfassung beider angezeigt sein. So könnte die Nutzung einer zusammenfassenden Verallgemeinerungs-Verknüpfung die notwendige Differenzierung der Unterschiedlichkeit von ontologischen Kategorien zur Unterscheidung zwischen induktiven und Analogie-Verknüpfungen ersetzen. So könnten ggf. auch inadäquate Verknüpfungen durch Verwechslungen der Verallgemeinerungsarten verhindert werden.

Der Gebrauch unmöglicher kausaler Verknüpfungen weicht von der Erwartung und der daraus resultierenden höchsten Bepunktung im Kodiermanual ab. Erwartet wurde, dass vor allem innerhalb der Dimension *Certainty of knowledge* deutlich erweiterte Überzeugungen nötig sind, um unmögliche kausale Verknüpfungen zu erstellen. Durch eine unmögliche kausale Verknüpfung werden Aspekte auf Grundlage von anderen Aspekten als unmöglich beschrieben. Dabei können die verknüpften Aspekte jeweils aus längeren Argumentationslinien entspringen und innerhalb dieser als bekräftigte Argumente beschrieben werden.

Mit Blick auf das Expert*innen-MOD zu Experiment 1 kann z.B. die Verallgemeinerung „Bei der Reaktion von Säuren mit Basen gleicher Volumina und Konzentration entsteht immer eine neutrale Lösung.“ aus den ersten beiden Teilerperimenten begründet werden. Damit könnte dieser Aspekt als ‚gesichertes Wissen‘ angesehen werden, von welchem ausgehend weitere Erkenntnisse

gesammelt werden können. Dies entspricht dem Verständnis des *solid core* nach Schommer-Aikins (2002). Durch das dritte Telexperiment und die daraus entstehende kausale Schlussfolgerung, dass keine Neutralisation mit dem End-pH-Wert=7 stattgefunden hat, kann die vorherige Verallgemeinerung jedoch als kausal unmöglich beschrieben werden. Folglich wird durch die unmögliche kausale Verknüpfung ein Aspekt als unmöglich erklärt, welcher durch zwei Telexperimente und kausalen Schlussfolgerungen daraus unterstützt wird. Ohne ausgeprägte Überzeugungen zur Vorläufigkeit von Wissen könnte eine solche Verknüpfung ggf. nicht gebildet werden. Unmögliche kausale Verknüpfungen dieser Art können vor allem für Experiment 1 häufig innerhalb der Daten berichtet werden (z.B. INCH29FE, MEAM18FA).

In diesem Sinne besitzt diese Verknüpfungsart eine besondere Rolle für den Umgang mit anomalen Beobachtungen. Sofern die anomalen Beobachtungen in das MOD aufgenommen werden, führen diese definitionsgemäß zu Widersprüchen mit bestimmten Aspekten der Theorie oder bestehender Beobachtungen. Im Sinne eines kognitiven Konflikts (vgl. Kap. 3.3.1) könnte dies zur Unzufriedenheit mit der bestehenden Theorie und so zur Theoriebearbeitung führen.

Andererseits weisen unmögliche kausale Verknüpfungen einige Merkmale von kausalen Verknüpfungen auf. So stellen diese lineare Verknüpfungen zwischen mind. 2 Aspekten dar, die nicht zwangsläufig Auswirkungen auf weitere Zusammenhänge haben müssen. Auch hier könnte die Art der verknüpften Aspekte einen Einfluss haben. So können auch Aspekte verknüpft werden, welche nicht innerhalb größerer Argumentationslinien eingebettet sind. Ein Beispiel kann das MOD von SAGT10RL zu Experiment 2 darstellen. Die Probandin erstellt eine unmögliche kausale Verknüpfung zwischen den Aspekten „Kolben mit kochendem Wasser wurde in ein Gefäß mit kaltem Wasser gegeben“ und „Temperatur steigt weiter an“. Dabei ist der erstgenannte Aspekt in Zusammenhang mit weiteren Aspekten gestellt, während der zweite Aspekt isoliert vorgefunden wird. Damit könnte hier die Linearität der unmöglichen kausalen Verknüpfung und damit fehlende Auswirkung auf gesamte Argumentationslinien besonders deutlich werden. Die Proband*innen formulieren so teilweise inhaltliche Aspekte, welche lediglich durch eine unmögliche kausale Verknüpfung verknüpft werden. Dabei sind theoretisch Aspekte möglich, welche einfache Widersprüche zu formulierten Aspekten beinhalten. Diese Aspekte und Verknüpfungen sind damit nicht inhaltstragend im Sinne der Verarbeitung der Beobachtungen. So könnten willkürlich unmögliche kausale Verknüpfungen zu isolierten Aspekten erstellt werden. Dabei könnten hier auch naivere epistemologische Überzeugungen zur Erstellung der Verknüpfungen ausreichen, wie diese für kausale Verknüpfungen beschrieben wurden.

Es zeigt sich folglich, dass unmögliche kausale Verknüpfungen nicht immer am Ende von Argumentationslinien zur Verknüpfung anomaler Beobachtungen genutzt werden. Dabei kann das Fehlen unmöglicher kausaler Verknüpfungen auch mit dem Fehlen widersprechender Aspekte im Sinne anomaler Beobachtungen zusammenhängen. Wie bereits diskutiert, könnten anomale Beobachtungen z.B. aufgrund der Theoriegeladenheit der Beobachtungen nicht in das MOD aufgenommen worden sein. Dadurch könnten dann auch keine unmöglichen kausalen Verknüpfungen erstellt werden. Ein Indiz für den Zusammenhang zwischen dem Beschreiben anomaler Beobachtungen sowie der Bildung unmöglicher kausaler Verknüpfungen kann im Vergleich von Experiment 1 und 2 erkannt werden. So wurden in Experiment 1 signifikant mehr anomale Beobachtungen beschrieben und unmögliche kausale Verknüpfungen erstellt. Folglich kann eine Begründung weniger gebildeter unmöglicher kausaler Verknüpfungen durch die Ergebnisse dieser Studie nicht allein auf die zugrundeliegenden epistemologischen Überzeugungen zurückgeführt werden. Sofern keine anomalen Beobachtungen oder sonstige widersprechenden Aspekte formuliert werden, können auch keine unmöglichen kausalen Verknüpfungen erstellt werden.

In einem Zwischenfazit kann hier die korrespondierende Bedeutung des Fachwissens sowie der epistemologischen Überzeugungen der Proband*innen für den Umgang mit anomalen Beobachtungen festgehalten werden. Dabei scheint das Fachwissen eine wichtige Rolle für die Wahrnehmung anomaler Beobachtungen zu spielen. Im Sinne von theoriegeladenen Beobachtungen könnten anomale Beobachtungen so z.B. ignoriert werden und so keinen Eingang in die epistemologische Vernetzung dieser innerhalb der MOD finden. Andererseits wird deutlich, dass epistemologische Überzeugungen einen wichtigen Einfluss auf die Verknüpfung unterschiedlicher Aspekte und so Einfluss auf die Deutung anomaler Daten haben können. So könnte z.B. die Erstellung von Verallgemeinerungen im Sinne heuristischer Annahmen einen wichtigen Einfluss auf die Deutung anomaler Daten haben. Verhindern epistemologische Überzeugungen jedoch die Erstellung von Verallgemeinerungen, könnte Erkenntnisprozesse hier unterbunden werden.

Die Kombination des Einflusses von Fachwissen und epistemologischen Überzeugungen auf den Umgang mit anomalen Beobachtungen könnte dabei durch das Modell der *framework theories* (Vosniadou, 1994) erklärt werden. Dabei kann das Fachwissen im Rahmen des mentalen Modells Einfluss auf die Wahrnehmung von anomalen Beobachtungen nehmen (vgl. Abb. 5). Dabei können Beobachtungen erst vor dem Hintergrund des Fachwissens als anomal wahrgenommen werden. Der Umgang mit wahrgenommenen anomalen Beobachtungen kann dann jedoch ebenso durch epistemologische Überzeugungen im Rahmen der *framework theories* (vgl. Abb. 5) beeinflusst werden. Dabei konnte im Rahmen dieser Studie die Bedeutung epistemologischer Überzeugungen für die Erstellung von Verknüpfungen zwischen einzelnen inhaltlichen Aspekten diskutiert werden. Die Kombination des Fachwissens und der epistemologischen Überzeugungen könnte dann zu unterschiedlichen Umgangsarten mit anomalen Beobachtungen führen, wie diese in der Bündelung der Taxonomie nach Chinn und Brewer (1993b, 1998) diskutiert wurde.

Mit Blick auf die unterschiedlichen Verknüpfungsarten und deren Nutzung durch die Proband*innen auf Grundlage derer epistemologischer Überzeugungen, wurden bereits Anpassungen der Verknüpfungsarten diskutiert. So könnte z.B. die Einführung einer Verknüpfungsart für zeitliche Abfolgen wie auch die Zusammenfassung verallgemeinernder Verknüpfungsarten zur einfacheren Umsetzung der Methode der MOD beitragen, ohne dabei die differenzierte Bedeutung epistemologischer Überzeugungen für die einzelnen Verknüpfungsarten abzuwerten. Vor dem Hintergrund der Nutzung der MOD stellen sich jedoch auch Fragen zur Validität der erstellten MOD.

Aus der Analyse der Abweichungen und Ergänzungen der MOD durch das laute Denken können Validitätsaspekte abgeleitet werden. Zuerst kann festgehalten werden, dass die Beobachtungen sowie das vorhandene Vorwissen der Proband*innen innerhalb der graphischen MOD abgebildet wurde. Innerhalb des lauten Denkens sind nur weniger Unterschiede zu erkennen. Auch im Rahmen der analysierten anomalen Daten konnten nur wenige Ergänzungen durch das laute Denken beschrieben werden (vgl. Kap. 7.6.2). Folglich könnte das beschriebene fehlende Fachwissen zur Deutung der anomalen Beobachtungen bei den Proband*innen tatsächlich nicht vorhanden und nicht nur durch die Externalisierung des mentalen Modells nicht in den MOD dargestellt worden sein.

Für die gebildeten Verknüpfungen sind deutlichere Abweichungen innerhalb der MOD zu erkennen. So spielen hier an einigen Stellen Verwechslungen zwischen den Verknüpfungsarten eine Rolle. Dabei können teils simple Verwechslungen auf der Darstellungsebene erkannt werden. So konnten einige im MOD inadäquat erstellte Verknüpfungen durch die Erläuterungen im lauten Denken nachvollzogen und korrigiert werden. Hier könnte ein länger angelegtes Training ggf. zur sicheren Nutzung auf der Darstellungsebene beitragen. Dabei könnten auch noch unterschiedlichere Darstellungen für die unterschiedlichen Verknüpfungsarten gewählt werden. Im Rahmen

der vorliegenden Studie wurde auf die von Chinn und Brewer (2001) vorgeschlagenen Darstellungsarten zurückgegriffen. Farbige Differenzierungen könnten zur klareren Unterscheidung auch in kognitiv anspruchsvollen Modellierungsprozessen beitragen.

Ebenso wurden jedoch Verknüpfungsarten auf einer tiefenstrukturellen Ebene vertauscht. Hier ist insbesondere die Unterscheidung zwischen der induktiven und der Analogie-Verknüpfung hervorzuheben. Wie diskutiert, können hier Herausforderungen für die Erstellung von Verallgemeinerungen für die Proband*innen vermutet werden. Andererseits kann eine Unterscheidung der Verknüpfungsarten zur Evaluation der MOD nur mit Kenntnis des Hintergrundwissens der Proband*innen erfolgen. Die verallgemeinerten Entitäten können aufgrund von unterschiedlichen Merkmalen in unterschiedliche ontologische Kategorien sortiert werden. Ohne Kenntnis über dieses Merkmal ist eine Evaluation der Zuordnung nur schwer möglich. Dabei werden die Merkmale jedoch zumeist nicht im Rahmen der graphischen MOD thematisiert bzw. können nur aus dem Kontext des gesamten MOD abgeleitet werden. Hieran schließen sich Validitätsprobleme für die Evaluation der MOD durch die Auswertenden an. Eine oben beschriebene Zusammenfassung der verallgemeinernden Verknüpfungsarten könnte so auch aus Perspektive der internen Validität unterstützt werden.

Im Rahmen des lauten Denken sind einige Abweichungen auch für kausale und kontrastierende Verknüpfungen zu erkennen. Vor allem die Nutzung kausaler Verknüpfungen für die zeitliche Abfolge oder die festgelegten Experimentierschritte spielen dabei – wie oben diskutiert – immer wieder eine Rolle. Auch hier können im lauten Denken Formulierungen erkannt werden, welche eher auf die bewusste Nutzung kausaler Verknüpfungen im Sinne zeitlicher Abfolgen zurückschließen lassen. Auch hier könnte die oben beschriebene Einführung einer neuen Verknüpfungsart zur validen Nutzung der kausalen Verknüpfung im MOD beitragen. Die Proband*innen könnten so die kausale Verknüpfung nur dann einsetzen, wenn es sich um Ursache-Wirkung-Zusammenhänge handelt.

Auf Grundlage des Vergleichs mit dem lauten Denken kann also subsummiert werden, dass die Proband*innen gut in der Lage zu sein scheinen inhaltliche Aspekte sowie anomale Beobachtungen innerhalb eines MOD auszudrücken. Für die Erstellung der Verknüpfungsarten sind Abweichungen im lauten Denken zu erkennen, welche die Hinzufügung einer Verknüpfung für zeitliche Abfolgen wie auch die Zusammenfassung verallgemeinernder Verknüpfungen unterstützen. Die Veränderungen der bereitstehenden Verknüpfungsarten könnten so zu einem valideren Einsatz der Verknüpfungsarten innerhalb der MOD durch die Proband*innen beitragen. Insgesamt könnte so die Eignung der Methode der MOD zur Externalisierung mentaler Modelle weiter gesteigert werden. Neben dem Vergleich mit dem lauten Denken der Proband*innen sollte in der vorliegenden Studie auch der eingesetzte Fachwissenstest als externes Validitätsargument genutzt werden.

Der Fachwissenstest sollte Einblicke in die theoretischen Annahmen der Proband*innen geben. Dabei können jedoch für Experiment 1 keinerlei Korrelationen zwischen den Merkmalen der MOD und dem Abschneiden im Fachwissenstest erkannt werden. Hier hätte z.B. ein Zusammenhang zwischen der Erstellung essentieller Aspekte und dem Fachwissen erwartet werden können. Die fehlende Korrelation zeigt hier jedoch keinen Zusammenhang zum Fachwissen auf. Eine mögliche Erklärung könnte sein, dass die Kompetenz- und Wissensanteile, die zur Konstruktion eines MOD nötig sind, weniger Bestandteil des Fachwissens als eher der *framework-theories* sind. So könnte vor allem der Umgang mit anomalen Beobachtungen durch epistemologische Überzeugungen oder generelle Einstellungen zum Umgang mit Evidenzen generell beeinflusst werden.

Hier muss andererseits auch das Testinstrument kritisch diskutiert werden. So fragt die Skala vor allem grundständige Fachwissensinhalte zum Thema Säure-Base ab. Die Proband*innen schneiden hier mit einer durchschnittlichen Punktzahl von 68% relativ gut ab, wenngleich ebenfalls Förderbedarf erkennbar ist. Für die Erklärung des beobachteten Phänomens müssen hingegen auch Aspekte der Säurestärke sowie des chemischen Gleichgewichts geachtet werden. In diesen Inhaltsbereichen könnte für die genutzte Säure-Base-Skala eine schlechtere Auflösung erwartet werden. Entsprechend könnte ein fehlender Zusammenhang des Fachwissens mit den Merkmalen von MOD auch durch eine unzureichende Abfrage durch das Testinstrument begründet werden. Folglich muss hier die Konstruktvalidität der Säure-Base-Skala kritisch bewertet werden.

Andererseits zeigt sich in der Analyse der essentiellen Konzeptaspekte innerhalb der MOD zu Exp. 1, dass auch grundlegendere Aspekte der Säure-Base-Chemie nicht erstellt werden. So wurden keinerlei Aspekte zum Ablaufen einer Protonenübertragungsreaktion bzw. wenige Aspekte zum Ablaufen einer Neutralisationsreaktion mit dem End-pH-Wert=7 erstellt (vgl. Tab. 35). Auch wurde keinerlei Bezug auf die jeweiligen Ionen für die Ursache einer sauren/basischen Lösung genommen. Diese Fachwissensanteile wurden jedoch im Rahmen des Fachwissenstests abgefragt (siehe Fachwissenstest im Anhang). Somit könnte abgeleitet werden, dass weiteres, grundlegendes Fachwissen zum Bereich Säure-Base bei den Proband*innen vorhanden ist, jedoch nicht im Rahmen der MOD eingesetzt wurde. Hierfür könnten Begründungen in der Relevanz der Fachwissensaspekte für die Beschreibung des Experiments sowie der Deutung der anomalen Beobachtungen erkannt werden. So könnte eine fehlende wahrgenommene Relevanz der konzeptuellen Fachwissensaspekte zum ausbleibenden Einsatz derer innerhalb der MOD geführt haben. Dabei könnte diese Argumentation an die größere Anzahl von Beobachtungs- im Vergleich zu Konzeptaspekten angegliedert werden. Die Proband*innen beschreiben die Phänomene eher auf Ebene der Beobachtungen, als dass konzeptuelle Aspekte zur Erklärung der Beobachtung eingesetzt werden.

Für die Einzelitems der Siedepunkt-Skala lassen sich jedoch Zusammenhänge zum Fachwissen herstellen. Dabei scheinen hier gefestigte Vorstellungen zum Siedepunkt von Wasser (Item 1) im Zusammenhang zur Bildung essentieller Beobachtungsaspekte zu stehen. Entsprechend ergeben sich hier Hinweise, dass Fachwissensaspekte einen Einfluss auf die Merkmale der MOD nehmen könnten. Innerhalb von Item 1 wird nach der wahrscheinlichsten Temperatur des Wassers bei beginnendem Sieden gefragt. Gefestigte Alltagserfahrungen zum Siedepunkt von Wasser bei 100°C wurden hier im Zusammenhang mit dem Erkennen anomaler Beobachtungen bereits diskutiert. Innerhalb von Item 1 steht jedoch eine Temperatur von 100°C dabei nicht zur Auswahl im *Multiple-choice* Format. Entsprechend könnten hier vor allem Proband*innen mit weniger gefestigten Vorstellungen eine richtige Antwort gegeben haben. So ließe sich der Zusammenhang mit dem Erstellen essentieller Beobachtungsaspekte erklären. Im Sinne der Theoriegeladenheit der Beobachtungen könnten hier Proband*innen mit weniger gefestigten Vorstellungen mehr essentielle Beobachtungen tätigen und diese in das MOD überführen. Dabei zeigt sich hier jedoch kein direkter Zusammenhang mit der Erstellung von Aspekten zu anomalen Beobachtungen innerhalb der MOD.

Zusammenhänge zwischen der Bildung adäquater kontrastierender Verknüpfungen zeigen sich für Item 6. So gibt es einen Zusammenhang zwischen erhöhtem Fachwissen zum Abkühlungsprozess und der Erstellung adäquater kontrastierender Verknüpfungen. Proband*innen, welche über vertieftes Fachwissen zum Abkühlungsprozess verfügen, könnten so auch eher tiefenstrukturelle Kontraste zur Erstellung von Verknüpfungen nutzen. Somit könnte sich ggf. die Bedeutung von Fachwissensinhalten für die Erstellung dieser argumentieren lassen. Dabei muss jedoch berücksichtigt werden, dass aus den bestehenden Daten keine Kausalitäten abgeleitet werden können,

wenngleich die hier beschriebene vor dem Hintergrund theoretischer Annahmen argumentierbar wäre.

Ein negativer Zusammenhang zeigt sich in der Korrelationsanalyse zwischen der Anzahl adäquater kausaler Verknüpfungen mit Item 8. Item 8 zielt dabei auf die Temperaturkonstanz am Siedepunkt auch trotz erhöhter Hitzezufuhr ab. Wie beschrieben stellen kausale Verknüpfungen häufig auf lineare Zusammenhänge ab, wie diese auch in item 8 erkannt werden können. Eine Erhöhung der Temperatur der Heizplatte könnte so linear zu einer Erhöhung der Wassertemperatur führen. Die Konstanz der Temperatur am Siedepunkt weicht dabei von einem solchen linearen Zusammenhang ab. Die Unterscheidung dieser beiden linearen Zusammenhänge stellt dabei einmal einen Fokus auf das Fachwissen (Item 8) und einen Fokus auf die epistemologischen Überzeugungen (adäquate kausale Verknüpfungen) dar. Proband*innen welche nicht-lineare Prozesse im Bereich des Fachwissens beschreiben können, könnten dies auf kausale Verknüpfungen vor dem Hintergrund der epistemologischen Überzeugungen übertragen und andersherum. Dabei steht die Linearität der kausalen Verknüpfung der ausbleibenden Linearität der Temperaturerhöhung des Wasser am Siedepunkt gegenüber. Dieser Argumentation folgend könnte eine Übertragung von Prinzipien des Fachwissens auf die epistemologischen Überzeugungen vermutet werden. Im Sinne des Modells der *framework theories* (Vosniadou, 1994) könnten Aspekte des mentalen Modells auf Aspekte des *frameworks* übertragen werden, oder andersherum. Dabei muss die grundsätzlich andere Ausrichtung der beiden Bereiche berücksichtigt werden. Inwiefern eine derartige Übertragung zwischen dem mentalen Modell und dem *framework* vorliegt, kann auf Grundlage der Ergebnisse dieser Studie nicht abschließend geklärt werden.

Insgesamt liegen durch den Abgleich mit dem lauten Denken sowie vereinzelt Zusammenhängen mit dem Fachwissenstest Validitätsargumente für die erstellten MOD vor. Vor allem die erstellten Verknüpfungen weisen auf Validitätsprobleme hin, welche jedoch durch die aus der Analyse des Umgangs mit anomalen Daten sowie dem Abgleich mit dem lauten Denken argumentierten Anpassungen der Verknüpfungsarten für zukünftige Studien überarbeitet werden könnten. In Bezug auf die gebildeten inhaltlichen Aspekte sind nur wenige Abweichungen zum lauten Denken zu erkennen. Unterschiede zwischen dargestelltem Konzeptwissen im MOD und den Ergebnissen des Fachwissenstest für die Säure-Base-Skala könnten auf die wahrgenommene Relevanz der Aspekte zurückgeführt werden. Insofern könnten für die Darstellung inhaltlicher Aspekte starke Validitätsargumente vorliegen.

Zusammenfassend können die Proband*innen die anomal wahrgenommenen Beobachtungen innerhalb der MOD zumeist nicht erklären. Dabei ist insbesondere für Experiment 2 eine ausbleibende Beschreibung der auf Grundlage bestehender Vorstellungsforschung erwarteten anomalen Beobachtungen zu verzeichnen. In diesem Zusammenhang könnte sich ein deutlicher Einfluss der bestehenden Vorstellungen im Sinne der theoriegeladenen Beobachtung zeigen (Chinn & Malhotra, 2002a). Entsprechend könnte ein Einfluss des Fachwissens auf die Beobachtung von Experimenten unterstützt werden. Ebenso könnten epistemologische Überzeugungen Einfluss auf den Umgang mit anomalen Daten genommen haben. So könnten naivere epistemologische Überzeugungen zu Umgangsarten mit anomalen Beobachtungen am oberen Ende der Taxonomie nach Chinn und Brewer (1993b, 1998) geführt haben. Auf Ebene der erstellten Verknüpfungen zeigen sich ebenfalls unterschiedliche epistemologische Überzeugungen, welche zur Inadäquatheit von Verknüpfungen geführt haben könnten.

So muss als Fazit aus der vorliegenden Studie ein Einfluss des Fachwissens sowie der epistemologischen Überzeugungen auf den Umgang mit anomalen Beobachtungen festgehalten werden. Während das Fachwissen schon im Rahmen der Beobachtung einen Einfluss auf den Prozess

nehmen kann, zeigen sich epistemologische Überzeugungen vor allem auch in der Verknüpfung der anomalen Beobachtungen mit weiteren Aspekten bzw. der Begründung der jeweiligen Umgangsart mit anomalen Beobachtungen.

In der vorliegenden Studie erfolgte der Einsatz der MOD zur Externalisierung individueller mentaler Modelle. Die vorgeschlagene Überarbeitung der zur Verfügung stehenden Verknüpfungsarten könnte die Nutzung der Methode der MOD auch für weitere Studien ermöglichen. So könnten Hürden in der Erstellung (valider) MOD abgebaut werden, ohne dabei die epistemologische Begründung unterschiedlicher Verknüpfungsarten weniger differenziert zu ermöglichen. Der Mehrwert der MOD kann in der Darstellung von Fachwissensaspekten und Beobachtungen im Kontext derer epistemologischer Verknüpfung erkannt werden.

8. Zusammenführung und Fazit

Das Ziel dieser Arbeit war es den Umgang mit Evidenzen angehender (Chemie-) Lehrkräfte sowie insbesondere deren Reaktion auf anomale Beobachtungen unter Berücksichtigung epistemologischer Überzeugungen zu untersuchen. Dazu wurden drei Teilstudien konzipiert, welche auf die vorhandenen Kompetenzen in evidenzbasierter Praxis sowie den Einfluss der studierten Unterrichtsfächer in einer querschnittlichen Studie (Teilstudie 1), die Förderung evidenzbasierter Kompetenzen bei angehenden Chemielehrkräften im Rahmen einer längsschnittlichen Studie (Teilstudie 2) und den Umgang mit anomalen Beobachtungen in chemischen Experimenten im Rahmen einer Modellierungsstudie (Teilstudie 3) fokussieren.

Die Bedeutung evidenzbasierter Praxis für den Bildungssektor konnte dabei aus bestehenden Konzeptualisierungen innerhalb der Medizin abgeleitet werden (Cholewa et al., 2015; Dawes et al., 2005; Sackett et al., 1996). Dabei ergeben sich für den Bildungsbereich einerseits ähnliche Herausforderungen für die Implementation evidenzbasierter Praxis. Andererseits könnten Charakteristika des Bildungssektors wie etwa die Interdisziplinarität von (angehenden) Lehrkräften (Rochnia et al., 2020) und Unterschiede in den verfügbaren Evidenzen (Bromme et al., 2014) spezifische Anpassung einer evidenzbasierter Praxis notwendig machen.

Auf Grundlage der theoretisch begründeten Vermutungen zu Unterschieden in der Kompetenz in evidenzbasierter Praxis angehender Lehrkräfte, wurde eine Querschnitt-Studie mit Proband*innen unterschiedlicher Unterrichtsfächer durchgeführt. Dabei sollte ebenso der Einfluss des Praxissemesters als zentrale Lerngelegenheit der Theorie-Praxis-Verknüpfung innerhalb des Lehramtsstudiums Chemie (M.Ed.) untersucht werden. Die Ergebnisse weisen dabei keine Unterschiede für die Kompetenz in evidenzbasierter Praxis für die belegten Unterrichtsfächer aus. In Bezug auf die dargelegten Unterschiede epistemologischer Überzeugungen durch Urhahne und Kremer (2023) kann in der vorliegenden Studie der Einbezug der Interdisziplinarität der Proband*innen durch deren Einteilung in Fächerbündel-Kombinationen einen möglichen Erklärungsansatz darstellen. So belegen die Proband*innen teils Unterrichtsfächer aus unterschiedlichen Bereichen, wodurch ggf. kein rein domänenspezifischer Zugang zu evidenzbasierter Praxis beschrieben werden kann.

Vor dem Hintergrund der Domänenspezifität sowie der genutzten Erhebungsinstrumente (vgl. Kap. 6.5), könnten eher akademische epistemologische Überzeugungen (Muis et al., 2006) bei den Proband*innen aktiviert worden sein, wodurch sich fehlende Unterschiede erklären lassen könnten.

Unterschiede konnten jedoch für die Wahl des Begleitforschungsseminars als besondere Lerngelegenheit evidenzbasierter Praxis innerhalb des Praxissemesters berichtet werden. Die Unterschiede für die Begleitforschungsseminare im religiös-philosophischen Fächerbündel könnten hier ggf. auch domänenspezifische epistemologische Überzeugungen zurückgeführt werden. Dabei könnten vertiefte Einblicke in die Motivation zur Wahl des jeweiligen Begleitforschungsseminars weiter Aufschluss geben, inwiefern auch epistemologische Überzeugungen für die auftretenden Unterschiede eine Rolle spielen.

In Rahmen von Teilstudie 2 wurde eine Interventionsstudie zur Förderung von Kompetenzen in evidenzbasierter Praxis für angehende Chemielehrkräfte durchgeführt. Die Entwicklung einer Intervention spezifisch für Chemielehrkräfte lässt sich dabei durch den aus bestehender Forschung formulierten Bedarf einer fachdidaktischen Perspektive auf EBP begründen (Bauer & Kollar, 2023). Innerhalb dieser Studie konnte eine Förderung der Kompetenzen der Proband*innen innerhalb des fachspezifischen Kontextes aufgezeigt werden. Dabei könnte die Wirksamkeit der chemiespezifischen Intervention ein Indiz für die Bedeutung fachspezifischer Kompetenzen sowie epistemologischer

Überzeugungen darstellen. Es zeigt sich, dass die Intervention übergreifende forschungsmethodologische Aspekte wie die Anwendung von wissenschaftlichen Gütekriterien bei den Proband*innen fördert. Ein Einfluss auf spezifische Aspekte des forschungsmethodologischen Wissens wie z.B. die Auswahl von geeigneten Studiendesigns kann nicht berichtet werden. Ebenso zeigt sich ein Einfluss auf den Forschungsprozess im Sinne der Durchführung eigener Projektarbeiten. Dabei spielt vor allem die Bewertung von Quellen sowie die Bildung relevanter Suchbegriffe eine wichtige Rolle.

Weiterhin kann eine besondere Rolle der Reflexion über Evidenzen und evidenzbasierter Praxis aus den durchgeführten Interviews extrahiert werden. So kann die Reflexion über das eigene Studium Stellen aufzeigen sowie der eigenen Praxis, an welchen mit Evidenzen gearbeitet wurde. Dabei spielt auch der Einbezug von Meinungen von Expert*innen bzw. Kolleg*innen eine Rolle. So ist deutlich eine graduelle Ausprägung von Evidenzen bei den Proband*innen erkennbar. Demnach können Meinungen von Expert*innen durch die Begründung auf eigenen systematischen Erfahrungen Evidenzen ähneln, ohne dabei mit wissenschaftlichen Evidenzen gleichgesetzt zu werden. Die Begründung von Meinungen kann dabei durch das in Anlehnung an Toulmin (2003) entwickelte *Modified Model of argument* (vgl. Kap. 6.6.4) beschrieben werden. Der noch deutlichere Einbezug von Erfahrungen und Meinungen innerhalb dieses Modells im Vergleich zum Feld evidenzbasierter Medizin kann auf Charakteristika des Bildungsbereichs zurückgeführt werden. So können die in Kap. 2.4 beschriebenen Herausforderung einer evidenzbasierten Praxis innerhalb des Bildungsbereichs den Einbezug von Erfahrungswissen nötig machen.

Die Notwendigkeit von Fördermaßnahmen für die Wahrnehmung evidenzbasierter Praxis kann auch durch das geringe Skalenniveau der *EBP implementation scale* innerhalb der fächerübergreifenden Erhebung unterstützt werden. So ist es auch denkbar, dass die Proband*innen deutlich häufiger evidenzbasierte Tätigkeiten ausüben, dies jedoch nicht wahrnehmen. Dabei beschränken sich die evidenzbasierten Tätigkeiten innerhalb der *EBP implementation scale* auf eine Auswahl möglicher Tätigkeiten. Innerhalb des Praxissemesters könnten die Proband*innen der fächerübergreifenden Erhebung weitere evidenzbasierte Tätigkeiten unternommen haben, welche durch das Testinstrument nicht abgefragt wurden. So könnte die Reflexion eigener Tätigkeiten einen Ansatz zur Förderung in evidenzbasierter Praxis durch Explikation darstellen.

Während die ersten beiden Teilstudien auf die Kompetenzen in evidenzbasierter Praxis abzielten, erfolgt in Teilstudie 3 ein deutlicher Fokus auf den Umgang mit anomalen Beobachtungen. Dabei können anomale Beobachtungen im Sinne evidenzbasierter Praxis als widersprechende Evidenzen verstanden werden, welche von Lehrkräften weniger häufig genutzt werden, als Evidenzen die eigene Annahmen unterstützen (Bauer & Kollar, 2023). In dieser Studie wurde der Umgang mit anomalen Beobachtungen von angehenden Lehrkräften des Lernbereichs Sachunterrichts untersucht. Dabei wurden die anomalen Beobachtungen eingesetzt, um einen kognitiven Konflikt auszulösen (Baddock & Bucat, 2008; Bucat, 2015; Posner et al., 1982). Mithilfe der Methode der *Model-of-data* (Chinn & Brewer, 2001) sollten die Proband*innen ihr mentales Modell externalisieren und dabei die Verknüpfungen zwischen unterschiedlichen inhaltlichen Aspekten bzgl. der zugrundeliegenden epistemologischen Zusammenhänge einordnen. Auf Grundlage der Ergebnisse kann erkannt werden, dass ein Teil der Proband*innen ihre theoretischen Vorstellungen durch die Verarbeitung anomaler Beobachtungen verändern konnte. Dabei beschreiben die Proband*innen wie erwartet unterschiedliche anomale Beobachtungen. Diese stimmen dabei nicht immer mit den auf Grundlage der Vorstellungsforschung erwarteten anomalen Beobachtungen überein.

Ergebnisse für die Abhängigkeit des Umgangs mit anomalen Beobachtungen vom Vorwissen zeigen sich für Experiment 2. Hier konnten keine Theoriebearbeitung durch die Proband*innen erkannt werden. Dabei könnten gefestigte Präkonzepte die Wahrnehmung anomaler Beobachtungen im Sinne der

Theoriegeladenheit von Beobachtungen (Chinn & Malhotra, 2002a) verhindert haben. Damit kommt den Beobachtungen von Experimenten eine besondere Bedeutung bei der Wahrnehmung und Bewertung anomaler Beobachtungen als solche zu.

Auf Grundlage des Umgangs mit anomalen Beobachtungen durch die Proband*innen konnte eine Differenzierung folgend der Taxonomie nach (Chinn & Brewer, 1993b, 1998) nicht immer zweifelsfrei vorgenommen werden. Entsprechend wurde eine Zusammenfassung einzelner Kategorien der Taxonomie vorgeschlagen. Dabei können insgesamt Einflüsse des vorhandenen Fachwissens sowie der epistemologischen Überzeugungen berichtet werden. So könnte die reine Beschreibung aber ausbleibende Erklärung anomaler Beobachtungen auf fehlendes Fachwissen der Proband*innen hindeuten, während der Theoriebearbeitung unabhängig des jeweiligen Umfangs erweiterte epistemologische Überzeugungen nötig macht. Die Zusammenfassung unterschiedlicher Umgangsarten basiert somit auf der Rolle von Fachwissen bzw. epistemologischen Überzeugungen, welche für den Umgang mit anomalen Beobachtungen notwendig sind. Zwar könnte durch die Zusammenfassung der Kategorien im ersten Schritt ein Verlust an Differenziertheit erkannt werden, jedoch könnte so eine Einteilung der Umgangsarten vereinfacht werden. So werden nicht immer Begründungen für den spezifischen Umgang mit anomalen Beobachtungen angegeben. Gleichzeitig könnte das Nachfragen nach Begründungen z.B. nach dem impliziten Ignorieren anomaler Beobachtungen zu Reflexionsprozessen führen, welche eine Veränderung des Umgangs erzeugen könnten. Entsprechend könnte die Zuteilung zu den jeweiligen Umgangsarten dadurch erschwert werden.

Ein zweiter Fokus der dritten Teilstudie lag auf der Validität der MOD. Dazu wurde im Abgleich mit Ergebnissen des lauten Denkens dargestellt, dass eine hohe Übereinstimmung für die formulierten Beobachtungs- und Konzeptaspekte sowie die Beschreibung anomaler Beobachtungen vorliegt. Abweichungen hingegen liegen für die formulierten Verknüpfungsarten vor. Dabei sind hier oft Verwechslungen unterschiedlicher Verknüpfungsarten sowohl auf oberflächlicher als auch auf tiefenstruktureller Ebene zu erkennen, welche auf zu Grunde liegende epistemologische Überzeugungen schließen lassen. Um die Verwechslung unterschiedlicher Verknüpfungsarten zu verhindern, wurde die Integration einer neuen Verknüpfung für zeitliche Abfolgen sowie die Zusammenführung der induktiven und Analogie-Verknüpfung im Rahmen der Diskussion vorgeschlagen.

Zusammenfassend kann in allen drei Teilstudien die Bedeutung epistemologischer Überzeugungen sowie deren Einfluss auf evidenzbasierte Praxis festgestellt werden. Angehende (Chemie-) Lehrkräfte verfügen dabei lediglich über basale Kompetenzen im Umgang mit Evidenzen sowie dem Umgang mit anomalen Beobachtungen. Dabei kann eine graduelle Ausprägung des Evidenzverständnisses für die Proband*innen beschrieben werden. Hier wird die Bedeutung von Meinungen und Erfahrung vor dem Hintergrund eines kritischen Umgangs mit diesen für evidenzbasierte Praxis einbezogen. Ebenso scheint das vorhandene Fachwissen eine Rolle für die Wahrnehmung von Beobachtungen zu spielen, während epistemologische Überzeugungen Einfluss auf die Konstruktion von Verknüpfungen zwischen inhaltlichen Aspekten nehmen. Dabei müssen jedoch die diskutierten Limitationen der einzelnen Teilstudien beachtet werden.

Auf Grundlage der Ergebnisse dieser Arbeit könnte eine Betrachtung domänenspezifischer Kompetenzen in evidenzbasierter Praxis angezeigt sein. Indizien dafür leiten sich vor allem aus den beschriebenen unterschiedlichen Kompetenzen in evidenzbasierter Praxis für Proband*innen der Begleitforschungsseminare im religiös-philosophischen Bündel ab. Ebenso könnte die Wirksamkeit der chemiedidaktischen Intervention als Indiz herangezogen werden. Aus der Modellierungsstudie kann der Einfluss des Fachwissens bzw. vorhandener Vorstellungen auf die Aufnahme von Beobachtungen abgeleitet werden. So könnte auch hier domänenspezifisches Fachwissen einen Einfluss auf evidenzbasierte Praxis nehmen. Aus theoretischer Perspektive könnte ebenfalls auf die domänenspezifische

Ausdifferenzierung der nah-verwandten Konstrukte Epistemologie (Muis et al., 2006; Urhahne & Kremer, 2023) sowie *Nature of Science* (N. G. Lederman, 2007; Neumann & Kremer, 2013) verwiesen werden. Vor diesem Hintergrund bedürfte es einer Betrachtung evidenzbasierter Praxis als domänenspezifisch. Dabei ist hier nicht die Unterscheidung zwischen evidenzbasierter Medizin und evidenzbasierter Bildungsprozesse gemeint, sondern eine Ausdifferenzierung im Bildungsbereich selbst. Dabei bieten sich die einzelnen Unterrichtsfächer als Domänen an. Jedoch muss dabei die Interdisziplinarität von Lehramtsstudierenden berücksichtigt werden, welche innerhalb von mind. 2 Unterrichtsfächern sozialisiert werden. Dabei wird hier für einen aktiven Einbezug der Interdisziplinarität plädiert, sodass eine Förderung auch innerhalb der verschiedenen Unterrichtsfächer der Studierenden geschieht und sich nicht auf ein einzelnes Unterrichtsfach beschränkt.

Als Fazit kann die Forderung von Bauer und Kollar (2023) nach einem Kulturwechsel unterstützt werden. Wie sich gezeigt hat, können epistemologische Vorstellungen und ontologische Kategorisierungen neben dem vorhandenem Fachwissen einen entscheidenden Einfluss auf den Umgang mit Evidenzen und anomalen Beobachtungen nehmen. Entsprechend ist es notwendig, dass angehende Lehrkräfte sich ihrer epistemologischen Überzeugungen sowie der Bedeutung dieser für den Umgang mit Evidenzen bewusstwerden.

Gleichzeitig sollte die Rolle bestehenden Fachwissens aus epistemologischer Sicht reflektiert werden, um so den Einfluss des Fachwissens auf Beobachtungen zu diskutieren. In diesem Zusammenhang steht die Rolle gesicherten Wissens als *solid core* nach Schommer-Aikins (2002) im Fokus. So sollte Fachwissen als relevantes Wissen für weiteren Erkenntnisgewinn verstanden werden, jedoch gleichzeitig als veränderbar angenommen werden, damit relevante Beobachtungen getätigt werden können. Entsprechend tritt hier das relativistische Verständnis epistemologischer Überzeugungen hervor, welches im Sinne der *Balance* (Schommer-Aikins, 2002) in die Reflexion mit eingehen sollte.

Für den beschriebenen Kulturwechsel ist es nötig, die Tätigkeiten *using evidence* und *establishing evidence* (Davies, 1999) in der Profession von Lehrkräften zu verorten. So sollte die Nutzung von Evidenzen zur Unterrichtsplanung sowie der Beantwortung von unterrichtlichen Fragestellungen ebenso als Teil des Aufgabengebiets von Lehrkräften verstanden werden, wie die eigene Generierung von Evidenzen. Dabei kann die Generierung von Evidenzen hier auch auf niedrigeren Leveln stattfinden, wie z.B. der Erhebung von Lernendenvorstellungen im eigenen Unterricht, um daraus Implikationen für die Unterrichtsplanung abzuleiten.

In Rückbezug auf die Ergebnisse aus Teilstudie 2 ist es hier relevant, dass die Bedeutung von Evidenzen nicht nur in bestimmten Bereichen der Unterrichtsplanung und -durchführung anerkannt wird. Ebenso sollte keine abnehmende Wahrnehmung der Bedeutung von evidenzbasierter Praxis für Lehrkräfte mit zunehmendem Erfahrungswissen entstehen. In Rückbezug auf das *Modified Model of argument* sollte die Bedeutung von Erfahrungen und wissenschaftlichen Evidenzen sowie die Vermittlung dazwischen als Kernaspekt evidenzbasierter Praxis verstanden werden.

9. Implikationen

9.1 Implikationen für die Forschung

Auf Grundlage der beschriebenen Studien können Implikationen für weiterführende Forschung abgeleitet werden.

Für Teilstudie 1 und 2 ergibt sich zusammenfassend ein Bedarf die Entwicklung von Testinstrumenten für die Kompetenz angehender Lehrkräfte in evidenzbasierter Praxis. Dabei könnte der EIDM-ET als Adaption des Fresno-Tests (Ramos et al., 2003) bzgl. der Testgüte weiterentwickelt werden. Eine Herausforderung kann für den Einbezug der spezifischen Kontexte, bzw. domänenspezifischer Aspekte formuliert werden. So besteht ein Bedarf für ein Testinstrument, welches domänenspezifische Aspekte evidenzbasierter Praxis beinhaltet, jedoch über verschiedene Unterrichtsfächer hinweg vergleichbar ist. So müssten in Anlehnung an den EIDM-ET Situationsbeschreibungen für spezifische Fächer erstellt werden, die vergleichbare Kompetenzen der Proband*innen adressieren. So könnte ein fachspezifischer Kontext für die Proband*innen geschaffen werden, welcher ggf. eine bessere Untersuchung der Kompetenzen in evidenzbasierter Praxis ermöglicht.

Gleichzeitig könnten Einblicke in die tatsächliche Handlung der Lehrkräfte im Unterricht Aufschluss über die Kompetenzen in evidenzbasierter Praxis liefern. So könnten performanzorientierte Testverfahren entwickelt werden, welche die Ausführung evidenzbasierter Praxis in den Blick nimmt.

In Anlehnung an die unterschiedliche Gruppenzuteilung für Teilstudie 1 und die Studie von Urhahne und Kremer (2023) muss dabei jedoch die Interdisziplinarität der Proband*innen (Rochnia et al., 2020) beachtet werden. So könnte die Einführung eines spezifischen Kontextes innerhalb eines Testinstruments für alle Unterrichtsfächer der Proband*innen nötig werden, um deren Kompetenzen ganzheitlich zu untersuchen. Dabei ergeben sich aus den unterschiedlicher Fächerkombinationen zahlreiche Kombinationsmöglichkeiten. Eine Zuordnung zu Fächerbündeln und damit einhergehende Minimierung der Kombinationsmöglichkeiten für das Testinstrument scheint aufgrund der Spezifität jedes Unterrichtsfachs nicht möglich. So vernachlässigt die Einteilung in *soft-sciences* und *hard-sciences* nach Biglan (1973), wie sie von Urhahne und Kremer (2023) genutzt wird, die fachspezifischen Eigenheiten der Unterrichtsfächer innerhalb von Fächerbündeln.

Folglich besteht ein Bedarf nach einem Testinstrument, welches einerseits fachspezifische Besonderheiten evidenzbasierter Praxis einbezieht, andererseits jedoch die Vergleichbarkeit unterschiedlicher Unterrichtsfächer ermöglicht. Die Berücksichtigung der Interdisziplinarität stellt dabei einer große Herausforderung dar.

Ebenso könnte eine Gruppenzuteilung der Proband*innen auf anderen Merkmalen aufgebaut werden. So zeigt sich die Wahl des Begleitforschungsseminars im Rahmen von Teilstudie 1 als Einflussfaktor auf die Punktzahl des gekürzten EIDM-ET auf. Dabei könnten hier zu Grunde liegende Ausprägungen (akademischer) epistemologischer Überzeugungen die Wahl bedingen. Demnach könnte auch eine Gruppeneinteilung aufgrund der Ausprägungen akademischer epistemologischer Überzeugungen möglich sein.

Neben der Erhebung der Kompetenzen in evidenzbasierter Praxis der Proband*innen stellt ebenso die Förderung dieser einen relevanten Forschungsaspekt dar. Auf Grundlage der in Teilstudie 2 präsentierten Interventionsstudie weist die Förderung angehender Lehrkräfte durch eine an *evidence-use* Mechanismen orientierten Intervention im Fach Chemie eine Wirksamkeit auf. Auf Grundlage der nicht nachgewiesenen Unterschiede in der Ausprägung von EBP-Kompetenzen für Proband*innen unterschiedlicher Fächerbündel stellt sich jedoch die Frage, inwiefern auch fächerübergreifende

Interventionsmaßnahmen eine Wirksamkeit zum Aufbau von EBP-Kompetenzen zeigen können. Dabei könnte hier auf akademische epistemologische Überzeugungen (Muis et al., 2006) fokussiert werden, welche keine fächerspezifische Ausprägung erwarten lassen. Ebenso ist offen, inwiefern die dargestellte Intervention auf andere Unterrichtsfächer übertragen werden kann. Dabei könnten übergeordnete Prinzipien der Intervention, wie z.B. die Bearbeitung individueller Projekte sowie der Einbezug forschungsmethodologischer Aspekte auch auf andere Unterrichtsfächer übertragen werden.

In Bezug auf Teilstudie 3 stellt vor allem der Zusammenhang des Fachwissens sowie epistemologischer Überzeugungen der Proband*innen für den Umgang mit anomalen Daten sowie deren Darstellung in MOD ein Desiderat dar. Innerhalb dieser Arbeit konnte ein Zusammenhang zwischen Merkmalen des MOD sowie des Umgangs mit anomalen Beobachtungen mit dem Fachwissen aufgezeigt werden. Dabei könnten in Zusammenhängen zwischen Einzelitems und einzelnen Merkmalen der MOD für Experiment 2 Indizien für einen solchen Zusammenhang formuliert werden. Dabei sollten jedoch auch die epistemologischen Überzeugungen der Proband*innen kontrolliert werden, um auch hier einen möglichen Einfluss genauer untersuchen zu können. Hier stellt sich jedoch die Herausforderung, dass epistemologische Überzeugungen eher im Sinne einer Häufigkeitsverteilung vorliegen (Schommer-Aikins, 2002). So könnten epistemologische Überzeugungen, die durch ein Testinstrument erhoben werden, nicht zwingend mit denen übereinstimmen, die während der Erstellung eines MOD bzw. der Konfrontation mit anomalen Beobachtungen aktiviert werden. Insofern müsste ein Zugang zu epistemologischen Überzeugungen während der Erstellung der MOD gefunden werden, welcher jedoch nicht zu einer zusätzlichen kognitiven Belastung der Proband*innen führt. Entsprechende Maßnahmen, um implizite epistemologische Überzeugungen während der Bearbeitung anomaler Beobachtungen mittels MOD zu extrahieren, sollten somit Gegenstand zukünftiger Forschung sein.

9.2 Implikationen für die Lehrkräftebildung

Neben Implikationen für die Forschung, können aus der vorliegenden Arbeit auch Implikationen für die Lehrkräftebildung abgeleitet werden.

Allem voran kann die Bedeutung evidenzbasierter Praxis sowie zu Grunde liegender epistemologischer Überzeugungen für den Bildungsbereich formuliert werden. Dabei kommt hier der Explikation evidenzbasierter Tätigkeiten eine besondere Bedeutung zu. Entsprechend sollte im Rahmen von Vorlesungen und Seminaren eine Reflexion über genutzte Evidenzen integriert werden, welche gleichzeitig zur bewussten Wahrnehmung evidenzbasierter Tätigkeiten beitragen. Eine besondere Bedeutung könnte hier für die Generierung von Evidenzen durch chemische Experimentalpraktika beschrieben werden. Die Studierenden zeigen hier deutlich evidenzbasierte Tätigkeiten auf, welche jedoch bisher nicht als solche wahrgenommen werden. Die explizite Reflexion über die zu Grunde liegenden Prozesse könnte zu einer Sensibilisierung der Studierenden beitragen.

Weiterhin muss auf Grundlage der Ergebnisse aus Teilstudie 1 die Rolle des Praxissemesters im Lehramtsstudium (M.Ed.) kritisch reflektiert werden. Während das Praxissemester als zentraler Ort für die Theorie-Praxis-Verknüpfung angesehen werden kann, scheinen kaum Lerngelegenheiten für die Implementation evidenzbasierter Praxis zu existieren. Sofern die Theorie-Praxis-Verknüpfung im Sinne evidenzbasierter Praxis einen Fokus des Praxissemesters darstellen soll, müssen hier explizite Lernsituationen geschaffen werden. Dabei sollten diese auch außerhalb des Begleitforschungsseminars liegen, da es ansonsten ggf. zu Rollenkonflikten kommen könnte. So könnte der Ansatz des *establishing evidence* mit dem Begleitforschungsseminar an der Universität verknüpft werden, während innerhalb der Schulpraxis bestenfalls eine Orientierung an *using evidence* (Davies, 1999) stattfindet. Die Verortung beider Ansätze sollte an allen drei Lernorten des Praxissemesters erfolgen, um so die übergreifende Bedeutung evidenzbasierter Praxis für Bildungsprozesse insgesamt zu verdeutlichen. Durch

diese umfassende Thematisierung evidenzbasierter Praxis könnte der von Bauer und Kollar (2023) geforderte Kulturwechsel adressiert werden. Dabei könnte ein Kulturwechsel auf institutionalisierter Ebene des Praxissemesters zu einem Kulturwechsel der angehenden Lehrkräfte beitragen. Hier spielt jedoch ebenso ein Kulturwechsel von Lehrkräften im aktiven Schuldienst eine Rolle, da diese am Lernort Schule einen wichtigen Einfluss auf die Studierenden nehmen könnten. Entsprechend muss auch der Bereich der Lehrkräftefortbildung in diese Prozesse einbezogen werden.

Den hier beschriebenen Aspekten liegen die epistemologischen Überzeugungen von angehenden Lehrkräften zu Grunde. So sollte der Aufbau von adäquaten epistemologischen Überzeugungen ebenfalls im Fokus der Lehrkräfteausbildung liegen, da diese wie beschrieben einen Einfluss auf die Kompetenz in EBP der Lehrkräfte haben.

Zuletzt kann ein Bedarf der Verknüpfung von Erfahrungen und wissenschaftlichen Evidenzen beschrieben werden. Im Sinne des *Modified Model of argument* (vgl. Kap. 6.6.4) bedarf es einer ontologischen Unterscheidung zwischen Evidenzen als funktionale Argumente für einen *claim* und (ggf. unsystematischen) Erfahrungen von Expert*innen. In Anlehnung an die Definition von Evidenzen (vgl. Kap. 2.1) haben Expert*innen-Meinungen durchaus Berechtigung innerhalb von Argumentationsprozessen. Dabei müssen diese jedoch gewisse Qualitätsstandards einhalten und durch hochqualitative Evidenzen ergänzt werden. Dabei sollte kein dichotomes Verständnis der Begriffe im Sinne von wenig qualitativen Meinungen vs. hochqualitativen wissenschaftlichen Evidenzen entstehen. Es braucht die Verknüpfung aus beiden, um im Argumentationsprozess Stand zu halten. Entsprechend sollte eine Verknüpfung beider Konstrukte sowie deren Abgleich miteinander Bestandteil der Lehrkräftebildung sein. Ggf. könnten durch ein solches Verständnis die Ansätze des *using evidence* und *establishing evidence* näher miteinander und an die Profession der Lehrkräfte verknüpft werden. So könnte ein möglicher Kulturwechsel im Rahmen der Lehrkräfteausbildung ggf. zum Erfolg führen.

10 Verzeichnisse

10.1 Literaturverzeichnis

- Adadan, E., Trundle, K. C. & Irving, K. E. (2010). Exploring grade 11 students' conceptual pathways of the particulate nature of matter in the context of multirepresentational instruction. *Journal of Research in Science Teaching*, 47(8), 1004–1035.
- Afadil, A., Rahmawati, S. & Suherman, S. (2022). Conceptual changes of students in thermodynamics in physical chemistry course through the implementation of blended learning. *Jurnal Akademika Kimia*, 10(3), 174–181.
<https://doi.org/10.22487/j24775185.2021.v10.i3.pp174-181>
- Albarqouni, L., Glasziou, P. & Hoffmann, T. (2018). Completeness of the reporting of evidence-based practice educational interventions: a review. *Medical education*, 52(2), 161–170.
<https://doi.org/10.1111/medu.13410>
- Albrecht, C. (2016). *Belastungserleben bei Lehrkräften und Ärzten: Neue Ansätze für berufsspezifische Prävention*. Klinkhardt forschung. Verlag Julius Klinkhardt.
- Allchin, D. (2011). Evaluating knowledge of the nature of (whole) science. *Science Education*, 95(3), 518–542.
- Alpaslan, M. M. (2019). Examining relations between physics-related personal epistemology and motivation in terms of gender. *The Journal of Educational Research*, 112(3), 397–410.
<https://doi.org/10.1080/00220671.2018.1540966>
- Alvermann, D. E. & Qian, G. (2000). Relationship between epistemological beliefs and conceptual change learning. *Reading & Writing Quarterly*, 16(1), 59–74.
<https://doi.org/10.1080/105735600278060>
- Amin, T. G. (2009). Conceptual metaphor meets conceptual change. *Human Development*, 52(3), 165–197. <https://doi.org/10.1159/000213891>
- Amin, T. G., Smith, C. & Wiser, M. (2014). Student conceptions and conceptual change: Three overlapping phases of research. In N. G. Lederman & S. K. Abell (Hrsg.), *Handbook of research on science education* (S. 57–82). Routledge. https://www.researchgate.net/publication/264713095_Student_Conceptions_and_Conceptual_Change_Three_Overlapping_Phases_of_Research
- Appleton, K. (1993). What makes lessons different? A comparison of students' behaviour in two science lessons. *Research in Science Education*, 23(1), 1–9.
- Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. *Offizielle Leitlinien*. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V.
- Argimon-Pallàs, J. M., Flores-Mateo, G., Jiménez-Villa, J. & Pujol-Ribera, E. (2010). Psychometric properties of a test in evidence based practice: the Spanish version of the fresno test. *BMC medical education*, 10, 45. <https://doi.org/10.1186/1472-6920-10-45>
- Ashktorab, T., Pashaeypoor, S., Rassouli, M. & Alavi-Majd, H. (2015). Nursing students' competencies in evidence-based practice and its related factors. *Nursing and midwifery studies*, 4(4), e23047. <https://doi.org/10.17795/nmsjournal23047>
- Baddock, M. & Bucat, R. (2008). Effectiveness of a classroom chemistry demonstration using the cognitive conflict strategy. *International Journal of Science Education*, 30(8), 1115–1128.
<https://doi.org/10.1080/09500690701528824>
- Baker, S. E. & Edwards, R. (2012). *How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research*. National Centre

- for Research Methods (NCRM). <https://research.torrens.edu.au/en/publications/how-many-qualitative-interviews-is-enough-expert-voices-and-early>
- Barke, H.-D. (2006). *Chemiedidaktik: Diagnose und Korrektur von Schülervorstellungen*. Springer-Lehrbuch. Springer-Verlag Berlin Heidelberg. <https://doi.org/10.1007/3-540-29460-0>
- Bauer, J. & Kollar, I. (2023). (Wie) kann die Nutzung bildungswissenschaftlicher Evidenz Lehren und Lernen verbessern? Thesen und Fragen zur Diskussion um evidenzorientiertes Denken und Handeln von Lehrkräften. *Unterrichtswissenschaft*. Vorab-Onlinepublikation. <https://doi.org/10.1007/s42010-023-00166-1>
- Bauer, J. & Prenzel, M. (2012). European teacher training reforms. *Science*, 336(6089), 1642–1643. <https://doi.org/10.1126/science.1218387>
- Bauer, J., Prenzel, M. & Renkl, A. (2015). Evidenzbasierte Praxis - im Lehrerberuf?! Einführung in den Thementeil. *Unterrichtswissenschaft*, 43, 188–192. https://www.researchgate.net/publication/280876279_Evidenzbasierten_Praxis_-_Im_Lehrerberuf_Einfuehrung_in_den_Thementeil
- Baxter Magolda, M. B. (1990). The impact of the freshman year on epistemological development: Gender differences. *The Review of Higher Education*, 13(3), 259–284. <https://doi.org/10.1353/rhe.1990.0018>
- Beelmann, A. (2014). Möglichkeiten und Grenzen systematischer Evidenzkumulation durch Forschungssynthesen in der Bildungsforschung. *Zeitschrift für Erziehungswissenschaft*, 17(S4), 55–78. <https://doi.org/10.1007/s11618-014-0509-2>
- Behr, D., Braun, M. & Dorer, B. (2015). *Messinstrumente in internationalen Studien*. SDM-Survey Guidelines (GESIS Leibniz Institute for the Social Sciences). https://doi.org/10.15465/ge-sis-sg_006
- Belland, B. R., Gu, J., Kim, N. J., Jaden Turner, D. & Mark Weiss, D. (2019). Exploring epistemological approaches and beliefs of middle school students in problem-based learning. *The Journal of Educational Research*, 112(6), 643–655. <https://doi.org/10.1080/00220671.2019.1650701>
- Bellmann, J. & Müller, T. (2011a). Evidenzbasierte Pädagogik - ein Déjà-vu? Einleitende Bemerkungen zur Kritik eines Paradigmas. In J. Bellmann & T. Müller (Hrsg.), *Wissen, was wirkt: Kritik evidenzbasierter Pädagogik* (1. Aufl., S. 9–32). VS Verlag für Sozialwissenschaften / Springer Fachmedien Wiesbaden GmbH Wiesbaden.
- Bellmann, J. & Müller, T. (Hrsg.). (2011b). *Wissen, was wirkt: Kritik evidenzbasierter Pädagogik* (1. Aufl.). VS Verlag für Sozialwissenschaften / Springer Fachmedien Wiesbaden GmbH Wiesbaden. <https://link.springer.com/book/10.1007/978-3-531-93296-5>
- Bendixen, L. D. (2002). A process model of epistemic belief change. In B. K. Hofer & P. R. Pintrich (Hrsg.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (1 Aufl., S. 191–208). Lawrence Erlbaum Associates, Inc.
- Berland, L. K. & Russ, R. S. (2017). Conceptual change through argumentation. In T. G. Amin & O. Levrini (Hrsg.), *Converging perspectives on conceptual change: Mapping an emerging paradigm in the learning sciences* (1 Aufl., S. 180–189). Routledge.
- Berliner, D. C. (2002). Comment: Educational research: The hardest science of all. *Educational Researcher*, 31(8), 18–20. <https://doi.org/10.3102/0013189X031008018>
- Biglan, A. (1973). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology*, 57(3), 195–203. <https://doi.org/10.1037/h0034701>
- Borgetto, B., Tomlin, G. S., Max, S., Brinkmann, M., Spitzer, L. & Pfingsten, A. (2020). Evidenz in der Gesundheitsversorgung: Die Forschungspyramide. In R. Haring (Hrsg.), *Springer Reference Pflege - Therapie - Gesundheit. Gesundheitswissenschaften* (Korrigierte

- Publikation, S. 643–654). Springer. https://link.springer.com/content/pdf/10.1007/978-3-662-54179-1_58-1.pdf
- Breckon, J. & Dodson, J. (2016). *Using evidence - What works? A discussion paper*. Alliance for Useful Evidence. <http://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/Alliance%20Policy%20Using%20evidence%20v4.pdf?ver=2016-04-12-152711-000>
- Broekkamp, H. & van Hout-Wolters, B. (2007). The gap between educational research and practice: A literature review, symposium, and questionnaire. *Educational Research and Evaluation*, 13(3), 203–220. <https://doi.org/10.1080/13803610701626127>
- Bromme, R. & Kienhues, D. (2014). Wissenschaftsverständnis und Wissenschaftskommunikation. In T. Seidel & A. Krapp (Hrsg.), *Pädagogische Psychologie* (6 Aufl., S. 55–81). Beltz Verlag.
- Bromme, R., Prenzel, M. & Jäger, M. (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. *Zeitschrift für Erziehungswissenschaft*, 17(S4), 3–54. <https://doi.org/10.1007/s11618-014-0514-5>
- Brown, S., Montfort, D., Perova-Mello, N., Lutz, B., Berger, A. & Streveler, R. (2018). Framework theory of conceptual change to interpret undergraduate engineering students' explanations about mechanics of materials concepts. *Journal of Engineering Education*, 107(1), 113–139. <https://doi.org/10.1002/jee.20186>
- Bucat, R. (. (2015). Using the cognitive conflict strategy with classroom chemistry demonstrations. In J. García-Martínez & E. Serrano-Torregrosa (Hrsg.), *Chemistry education: Best practices, opportunities and trends* (1 Aufl., S. 447–468). Wiley-VCH Verlag GmbH & Co. KGaA.
- Buehl, M. M. & Alexander, P. A. (2005). Motivation and performance differences in students' domain-specific epistemological belief profiles. *American Educational Research Journal*, 42(4), 697–726. <https://doi.org/10.3102/00028312042004697>
- Bundesministerium für Bildung und Forschung. (2008). *Rahmenprogramm zur Förderung der empirischen Bildungsforschung*. Bundesministerium für Bildung und Forschung (BMBF). Bildungsforschung. https://web.archive.org/web/20130502174428/http://www.bmbf.de/pub/Bildungsforschung_Band_zweiundzwanzig.pdf
- Cakmakaya, O. S., Bati, A. H. & Kolodzie, K. (2021). Cross-cultural adaptation of the fresno test for Turkish language. *PloS one*, 16(1), e0245195. <https://doi.org/10.1371/journal.pone.0245195>
- Canadian Health Services Research Foundation. (2005). *Annual report 2005*. Canadian Health Services Research Foundation.
- Celik, K. N. (2022). *Entwicklung von chemischem Fachwissen in der Sekundarstufe I: Validierung einer Learning Progression für die Basiskonzepte „Struktur der Materie“, „Chemische Reaktion“ und „Energie“ im Kompetenzbereich „Umgang mit Fachwissen“* (1st ed.). Studien Zum Physik- und Chemielernen Ser: v.325. Logos Verlag Berlin. <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=30231069>
- Çetin, P. S., Kaya, E. & Geban, Ö. (2009). Facilitating conceptual change in gases concepts. *Journal of Science Education and Technology*, 18(2), 130–137. <https://doi.org/10.1007/s10956-008-9138-y>
- Cambridge Dictionary (o.D.). Evidence. In *Cambridge Dictionary online*. Abgerufen am 28.09.2022, von <https://dictionary.cambridge.org/de/worterbuch/englisch/evidence>

- Chan, C. K. K. (2001). Peer collaboration and discourse patterns in learning from incompatible information. *Instructional Science*, 29(6), 443–479.
<https://doi.org/10.1023/A:1012099909179>
- Chan, C. K. K., Burtis, J. & Bereiter, C. (1997). Knowledge building as a mediator of conflict in conceptual change. *Cognition and Instruction*, 15(1), 1–40.
https://doi.org/10.1207/s1532690xci1501_1
- Chan, K. & Elliott, R. G. (2004). Epistemological beliefs across cultures: Critique and analysis of beliefs structure studies. *Educational Psychology*, 24(2), 123–142.
<https://doi.org/10.1080/0144341032000160100>
- Charter, R. A. (2003). Study samples are too small to produce sufficiently precise reliability coefficients. *The Journal of general psychology*, 130(2), 117–129.
<https://doi.org/10.1080/00221300309601280>
- Chen, J. A. & Pajares, F. (2010). Implicit theories of ability of Grade 6 science students: Relation to epistemological beliefs and academic motivation and achievement in science. *Contemporary Educational Psychology*, 35(1), 75–87.
<https://doi.org/10.1016/j.cedpsych.2009.10.003>
- Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Hrsg.), *International handbook of research on conceptual change* (S. 61–82). Routledge.
- Chinn, C. A. & Brewer, W. F. (1993a). Factors that influence how people respond to anomalous data. In S. F. Chipman (Hrsg.), *Proceedings of the fifteenth annual conference on the cognitive science society* (S. 318–323). Lawrence Erlbaum Associate Publishers.
- Chinn, C. A. & Brewer, W. F. (1993b). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, 63(1), 1–49. <https://doi.org/10.3102/00346543063001001>
- Chinn, C. A. & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching*, 35(6), 623–654.
[https://doi.org/10.1002/\(SICI\)1098-2736\(199808\)35:6<623::AID-TEA3>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1098-2736(199808)35:6<623::AID-TEA3>3.0.CO;2-O)
- Chinn, C. A. & Brewer, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction*, 19(3), 323–393. https://doi.org/10.1207/S1532690XCI1903_3
- Chinn, C. A. & Malhotra, B. A. (2002a). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology*, 94(2), 327–343.
<https://doi.org/10.1037/0022-0663.94.2.327>
- Chinn, C. A. & Malhotra, B. A. (2002b). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175–218.
- Cholewa, J., Corsten, S., Daniel, S., Grewe, T., Günther, T., Lauer, N., Mantey, S. & Nobis-Bosch, R. (2015). Überprüfung von EBP-Kompetenzen bei Studierenden: Vorstellung der ersten Version des TÜBEP-SE - Test zur Überprüfung der Basiskompetenz in evidenzbasierter Praxis für SprachtherapeutInnen. *Forum Logopädie*, 29(6), 20–27. https://evidenzsst.org/pluginfile.php/121/mod_book/chapter/151/fl_2015_06%20T%C3%9CBEP-ST.pdf
- Commission of the European Communities. (2007, 3. August). *Improving the quality of teacher education*. Commission of the European Communities. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52007DC0392>
- Conley, A. M., Pintrich, P. R., Vekiri, I. & Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. *Epistemological Development and Its Impact on Cognition in Academic Domains*, 29(2), 186–204.

- Coppenrath, V., Filosa, L. A., Akselrod, E. & Carey, K. M. (2017). Adaptation and validation of the fresno test of competence in evidence-based medicine in doctor of pharmacy students. *American journal of pharmaceutical education*, 81(6), 106. <https://doi.org/10.5688/ajpe816106>
- Crawford, T. (2008). Winning the epistemological struggle: Constructing a cultural model of shared authority in an elementary classroom. *Teachers College Record: The Voice of Scholarship in Education*, 110(8), 1706–1736. <https://doi.org/10.1177/016146810811000807>
- Danzer, P. (2022). *Validierung eines Instruments zur Erhebung des Umgangs mit Evidenzen von angehenden Lehrkräften im Chemieunterricht* [unveröffentlichte Masterarbeit]. Universität Paderborn, Paderborn.
- Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, 47(2), 108–121. <https://doi.org/10.1111/1467-8527.00106>
- Dawes, M., Summerskill, W., Glasziou, P., Cartabellotta, A., Martin, J., Hopayian, K., Porzsolt, F., Burls, A. & Osborne, J. (2005). Sicily statement on evidence-based practice. *BMC medical education*, 5(1), 1. <https://doi.org/10.1186/1472-6920-5-1>
- Dega, B. G., Kriek, J. & Mogese, T. F. (2013). Students' conceptual change in electricity and magnetism using simulations: A comparison of cognitive perturbation and cognitive conflict. *Journal of Research in Science Teaching*, 50(6), 677–698. <https://doi.org/10.1002/tea.21096>
- Deutsches Ärzteblatt. (2020, 19. Mai). *US-Präsident Trump nimmt Hydroxychloroquin als Prophylaxe ein* [Pressemitteilung]. <https://www.aerzteblatt.de/nachrichten/113002/US-Praesident-Trump-nimmt-Hydroxychloroquin-als-Prophylaxe-ein>
- diSessa, A. A. (2002). Why “conceptual ecology” is a good idea. In M. Limón & L. Mason (Hrsg.), *Reconsidering conceptual change: Issues in theory and practice* (S. 28–60). Kluwer Academic Publishers.
- Dolfing, R., Bokhove, C., Fechner, S., Rybska, E., Bamber, S., Pollmeier, P. & Dudziak, R. (2022). *Final Report: Research in Teacher Education (RiTE)*.
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-41089-5>
- Dudenredaktion. (o. D.). Evidenz. In *Duden online*. Abgerufen am 28. September 2022, von <https://www.duden.de/rechtschreibung/Evidenz#bedeutungen>
- Duit, R. H. (1999). Conceptual change approaches in science education. In W. Schnotz, S. Vosniadou & M. Carretero (Hrsg.), *New perspectives on conceptual change* (S. 263–282). Pergamon.
- Duit, R. H., Roth, W.-M., Komorek, M. & Wilbert, J. (2001). Fostering conceptual change by analogies - between Scylla and Charybdis. *Learning and Instruction*, 11, 283–303.
- Ebert, W. (2001). *Systemtheorie in der Supervision: Bestandsaufnahme und Perspektiven. Forschung Soziologie: Bd. 109*. Leske + Budrich.
- Elder, A. D. (2002). Characterizing fifth grade students' epistemological beliefs in science. In B. K. Hofer & P. R. Pintrich (Hrsg.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (1 Aufl., S. 347–363). Lawrence Erlbaum Associates, Inc.
- Ericsson, K. A. & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, 5(3), 178–186. https://doi.org/10.1207/s15327884mca0503_3
- Eryilmaz, A. (2002). Effects of conceptual assignments and conceptual change discussions on students' misconceptions and achievement regarding force and motion. *Journal of Research in Science Teaching*, 39(10), 1001–1015. <https://doi.org/10.1002/tea.10054>

- Feldt, t., Karagiannidis, C., Mager, S., Mikolajewska, A., Uhrig, A., Witzke, O. & Wolf, T. (2020). *Erfahrungen im Umgang mit COVID-19-Erkrankten - Hinweise von Klinikern für Kliniker: Gibt es antivirale Substanzen, die zur Behandlung von Civid-19 zur Verfügung stehen?* Robert Koch Institut. https://edoc.rki.de/bitstream/handle/176904/6852.2/COVRIIN_Antivirale%20Therapie_20200708-korr.pdf?sequence=4&isAllowed=y
- Frambach, J., van der Vleuten, C. P. M. & Durning, S. J. (2013). Quality criteria in qualitative and quantitative research. *Academic Medicine*, 88(4), 552. <https://doi.org/10.1097/ACM.0b013e31828abf7f>
- Fritsche, L., Greenhalgh, T [T.], Falck-Ytter, Y [Y.], Neumayer, H.-H. & Kunz, R [R.] (2002). Do short courses in evidence based medicine improve knowledge and skills? Validation of Berlin questionnaire and before and after study of courses in evidence based medicine. *BMJ (Clinical research ed.)*, 325(7376), 1338–1341. <https://doi.org/10.1136/bmj.325.7376.1338>
- Gautret, P., Lagier, J.-C., Parola, P., van Hoang, T., Meddeb, L., Mailhe, M., Doudier, B., Courjon, J., Giordanengo, V., Vieira, V. E., Tissot Dupont, H., Honoré, S., Colson, P., Chabrière, E., La Scola, B., Rolain, J.-M., Brouqui, P. & Raoult, D. (2020). Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *International journal of antimicrobial agents*, 56(1), 105949. <https://doi.org/10.1016/j.ijantimicag.2020.105949>
- Gentner, D. (2001). Mental models, psychology of. In N. J. Smelser & P. B. Baltes (Hrsg.), *International encyclopedia of social & behavioral sciences* (1 Aufl., S. 9683–9687). Elsevier/Per-gamon.
- Gopnik, A. & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, 7(1-2), 145–171. <https://doi.org/10.1111/j.1468-0017.1992.tb00202.x>
- Gorard, S., See, B. H. & Siddiqui, N. (2020). What is the evidence on the best way to get evidence into use in education? *Review of Education*, 8(2), 570–610. <https://doi.org/10.1002/rev3.3200>
- Gossen, M., Fünning, H., Holzbauer, B. & Schipperges, M. (2018). *Zukunft? Jugend fragen! Nachhaltigkeit, Politik, Engagement - eine Studie zu Einstellungen und Alltag junger Erwachsener*. Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit (BMU).
- Gräsel, C., Bormann, I., Schütte, K., Trempler, Kati, Fischbach, Robert & Asseburg, R. (2012). Perspektiven der Forschung im Bereich Bildung für nachhaltige Entwicklung. In M. Adomßent (Hrsg.), *Bildung: Bd. 39. Bildung für nachhaltige Entwicklung: Beiträge der Bildungsforschung* (Bd. 39, S. 7–24). Bundesministerium für Bildung und Forschung (BMBF) Referat Bildungsforschung.
- Greenhalgh, T [Trisha], Howick, J. & Maskrey, N. (2014). Evidence based medicine: a movement in crisis? *BMJ (Clinical research ed.)*, 348, g3725. <https://doi.org/10.1136/bmj.g3725>
- Gropengießer, H. & Marohn, A. (2018). Schülervorstellungen und Conceptual Change. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 49–68). Springer Berlin Heidelberg.
- Guest, G., Bunce, A. & Johnson, L. (2006). How many interviews are enough? *Field Methods*, 18(1), 59–82. <https://doi.org/10.1177/1525822X05279903>
- Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R [Regina], Falck-Ytter, Y [Yngve], Alonso-Coello, P. & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clinical research ed.)*, 336(7650), 924–926. <https://doi.org/10.1136/bmj.39489.470347.AD>
- Haack, S. (2007). *Defending science - within reason: Between scientism and cynicism* (With a new preface by the author). Prometheus Books.

- Halbedl, D. (2011). *Informale Logik und die Untersuchung fehlerhafter Argumente im Rahmen der „Fallacy Theory“* [Diplomarbeit]. Karl-Franzens-Universität, Graz. <https://unipub.uni-graz.at/obvugrhs/content/titleinfo/215582/full.pdf>
- Hammer, D. (2017). The interacting dynamics of epistemology and conceptual understanding. In T. G. Amin & O. Levrini (Hrsg.), *Converging perspectives on conceptual change: Mapping an emerging paradigm in the learning sciences* (1 Aufl., S. 245–252). Routledge.
- Hannover, B., Zander, L. & Wolter, I. (2014). Entwicklung, Sozialisation und Lernen. In T. Seidel & A. Krapp (Hrsg.), *Pädagogische Psychologie* (6 Aufl., S. 139–166). Beltz Verlag.
- Hartmeyer, R., Stevenson, M. P. & Bentsen, P. (2018). A systematic review of concept mapping-based formative assessment processes in primary and secondary science education. *Assessment in Education: Principles, Policy & Practice*, 25(6), 598–619. <https://doi.org/10.1080/0969594X.2017.1377685>
- Heneghan, C. (2015). *Reflections on David Sackett's time at the Centre for Evidence-Based Medicine*. Centre for Evidence-Based Medicine. <https://www.cebm.ox.ac.uk/news/views/reflections-on-david-sacketts-time-at-the-centre-for-evidence-based-medicine>
- Herzig, B. & Wiethoff, C. (2019). Konzeptionelle, strukturelle und inhaltliche Gestaltungsaspekte des Praxissemesters an der Universität Paderborn. In C. Caruso & J. Woppowa (Hrsg.), *Praxissemester (Religion) in NRW: Bilanz und Perspektiven* (S. 6–20). UB-PAD - Paderborn University Library.
- Hetmanek, A., Wecker, C., Kiesewetter, J., Trempler, K., Fischer, M. R., Gräsel, C. & Fischer, F. (2015). Wozu nutzen Lehrkräfte welche Ressourcen? Eine Interviewstudie zur Schnittstelle zwischen bildungswissenschaftlicher Forschung und professionellem Handeln im Bildungsbereich. *Unterrichtswissenschaft*, 43(3), 193–208. https://www.researchgate.net/profile/Andreas-Hetmanek/publication/280932351_Wozu_nutzen_Lehrkraefte_welche_Ressourcen/links/59db841245851508a43a0916/Wozu-nutzen-Lehrkraefte-welche-Ressourcen.pdf
- Hofer, B. K. (2004). Exploring the dimensions of personal epistemology in differing classroom contexts: Student interpretations during the first year of college. *Epistemological Development and Its Impact on Cognition in Academic Domains*, 29(2), 129–163. <https://doi.org/10.1016/j.cedpsych.2004.01.002>
- Hofer, B. K. (2006). Beliefs about knowledge and knowing: Integrating domain specificity and domain generality: A response to Muis, Bendixen, and Haerle (2006). *Educational Psychology Review*, 18(1), 67–76. <https://doi.org/10.1007/s10648-006-9000-9>
- Hofer, B. K. (2017). Identifying the role of epistemic cognition and metacognition in conceptual change. In T. G. Amin & O. Levrini (Hrsg.), *Converging perspectives on conceptual change: Mapping an emerging paradigm in the learning sciences* (1 Aufl., S. 229–236). Routledge.
- Hofer, B. K. & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88–140. <https://doi.org/10.3102/00346543067001088>
- Hofer, B. K. & Pintrich, P. R. (Hrsg.). (2002). *Personal epistemology: The psychology of beliefs about knowledge and knowing* (1. Aufl.). Lawrence Erlbaum Associates, Inc.
- Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen [IQWiG] (o.D.). Aufgaben und Ziele des IQWiG. Abgerufen am 13.12.2022, von: <https://www.iqwig.de/ueber-uns/aufgaben-und-ziele/>
- Jacobson, M. J. & Spiro, R. J. (1995). Hypertext learning environments, cognitive flexibility, and the transfer of complex knowledge: An empirical investigation. *Journal of Educational Computing Research*, 12(4), 301–333. <https://doi.org/10.2190/4T1B-HBP0-3F7E-J4PN>

- Jellinek, G. & Jellinek, W. (1921). *Allgemeine Staatslehre* (3. Auflage unter Verwendung des handschriftlichen Nachlasses durchgesehen und ergänzt). Springer Berlin / Heidelberg. <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=6599026>
- Jiménez-Aleixandre, M. P. & Brocos, P. (2017). Shifts in epistemic status in argumentation and in conceptual change. In T. G. Amin & O. Levrini (Hrsg.), *Converging perspectives on conceptual change: Mapping an emerging paradigm in the learning sciences* (1 Aufl., S. 171–179). Routledge.
- Jin, H., Mikeska, J. N., Hokayem, H. & Mavronikolas, E. (2019). Toward coherence in curriculum, instruction, and assessment: A review of learning progression literature. *Science Education*, 103(5), 1206–1234. <https://doi.org/10.1002/sce.21525>
- Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive science*, 4, 71–115.
- Johnstone, A. H. (1997). Chemistry teaching - science or alchemy? *Journal of Chemical Education*, 74(3), 262. <https://doi.org/10.1021/ed074p262>
- Jones, N., Ross, H., Lynam, T., Perez, P. & Leitch, A. (2011). Mental models: An interdisciplinary synthesis of theory and methods. *Ecology and Society*, 16(1), 46.
- Kane, M. T. (2016). Validation strategies: Delineating and validating proposed interpretations and uses of test scores. In S. Lane, M. R. Raymon & T. M. Haladyna (Hrsg.), *Handbook of test development* (2 Aufl., S. 64–80). Routledge.
- Kang, S., Scharmann, L. C. & Noh, T. (2004). Reexamining the role of cognitive conflict in science concept learning. *Research in Science Education*, 34(1), 71–96.
- Kattmann, U. (2007). Didaktische Rekonstruktion - eine praktische Theorie. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 93–104). Springer-Verlag.
- Kaya, E., Erduran, S., Aksoz, B. & Akgun, S. (2019). Reconceptualised family resemblance approach to nature of science in pre-service science teacher education. *International Journal of Science Education*, 41(1), 21–47. <https://doi.org/10.1080/09500693.2018.1529447>
- Kehne, F. (2019). *Analyse des Transfers von kontextualisiert erworbenem Wissen im Fach Chemie. Studien zum Physik- und Chemielernen*. Logos.
- Kienhues, D., Jucks, R. & Bromme, R. (2020). Sealing the gateways for post-truthism: Reestablishing the epistemic authority of science. *Educational Psychologist*, 55(3), 144–154. <https://doi.org/10.1080/00461520.2020.1784012>
- King, P. M. & Kitchener, K. S. (2002). The reflective judgement model: Twenty years of research on epistemic cognition. In B. K. Hofer & P. R. Pintrich (Hrsg.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (1 Aufl., S. 37–61). Lawrence Erlbaum Associates, Inc.
- Kitchener, R. F. (2002). Folk epistemology: An introduction. *New Ideas in Psychology*, 20(2-3), 89–105. [https://doi.org/10.1016/S0732-118X\(02\)00003-X](https://doi.org/10.1016/S0732-118X(02)00003-X)
- Kizilgunes, B., Tekkaya, C. & Sungur, S. (2009). Modeling the relations among students' epistemological beliefs, motivation, learning approach, and achievement. *The Journal of Educational Research*, 102(4), 243–256. <https://doi.org/10.3200/JOER.102.4.243-256>
- Knefelkamp, L. L. & Slepitz, R. (1976). A cognitive-developmental model of career development - An adaptation of the Perry Scheme. *The Counseling Psychologist*, 6(3), 53–58. <https://doi.org/10.1177/001100007600600310>
- Kousathana, M., Demerouti, M. & Tsaparlis, G. (2005). Instructional misconceptions in acid-base equilibria: An analysis from a history and philosophy of science perspective. *Science & Education*, 14(2), 173–193. <https://doi.org/10.1007/s11191-005-5719-9>
- Kuhl, J. (1983). *Motivation, Konflikt und Handlungskontrolle*. Springer Verlag.

- Kuhn, D., Cheney, R. & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development*, 15(3), 309–328. [https://doi.org/10.1016/S0885-2014\(00\)00030-7](https://doi.org/10.1016/S0885-2014(00)00030-7)
- Kuhn, D. & Weinstock, M. (2002). What is epistemological thinking and why does it matter? In B. K. Hofer & P. R. Pintrich (Hrsg.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (1 Aufl., S. 121–144). Lawrence Erlbaum Associates, Inc.
- Kuhn, T. S. (1996). *The structure of scientific revolutions* (3. ed.). University of Chicago Press. <http://www.loc.gov/catdir/description/uchi051/96013195.html>
- Kultusministerkonferenz. (2019a). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung*. Kultusministerkonferenz (KMK). https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschlusse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf
- Kultusministerkonferenz. (2019b). *Standards für die Lehrerbildung: Bildungswissenschaften*. Kultusministerkonferenz (KMK). https://www.kmk.org/fileadmin/veroeffentlichungen_beschlusse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf
- Kvernbekk, T. (2011). Evidence-based practice and Toulmin. In F. Zenker (Vorsitz), *Argumentation: Cognition & Community: Proceedings of the 9th International Conference of the Ontario Society for the Study of Argumentation (OSSA)*, Windsor, Kanada. https://scholar.u-windsor.ca/ossaarchive/OSSA9/papersandcommentaries/39/?utm_source=scholar.u-windsor.ca%2FOSSAarchive%2FOSSA9%2Fpapersandcommentaries%2F39&utm_medium=PDF&utm_campaign=PDFCoverPages
- Laibhen-Parkes, N., Kimble, L. P., Melnyk, B. M., Sudia, T. & Codone, S. (2018). An adaptation of the original fresno test to measure evidence-based practice competence in paediatric bedside nurses. *Worldviews on Evidence-Based Nursing*, 15(3), 230–240. <https://doi.org/10.1111/wvn.12289>
- Lakatos, I. (1978). The methodology of scientific research programs. In J. Warrall & G. Currie (Hrsg.), *Philosophical papers* (Bd. 1, S. 1–250). Cambridge University Press.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>
- Larsen, C. M., Terkelsen, A. S., Carlsen, A.-M. F. & Kristensen, H. K. (2019). Methods for teaching evidence-based practice: a scoping review. *BMC medical education*, 19(1), 259. <https://doi.org/10.1186/s12909-019-1681-0>
- Lederman, J. S., Lederman, N. G. [N. G.], Bartels, S [S.], Jimenez, J [J.], Acosta, K., Akubo, M [M.], Aly, S [S.], Andrade, M. A. B. S. de, Atanasova, M., Blanquet, E [E.], Blonder, R [R.], Brown, P., Cardoso, R., Castillo-Urueta, P., Chaipidech, P., Concannon, J., Dogan, O. K., El-Deghaidy, H., Elzorkani, A., . . . Wishart, J [J.] (2021). International collaborative follow-up investigation of graduating high school students' understandings of the nature of scientific inquiry: is progress Being made? *International Journal of Science Education*, 43(7), 991–1016. <https://doi.org/10.1080/09500693.2021.1894500>
- Lederman, J., Lederman, N., Bartels, S [Selina], Jimenez, J [Juan], Akubo, M [Mark], Aly, S [Shereen], Bao, C., Blanquet, E [Estelle], Blonder, R [Ron], Bologna Soares de Andrade, M., Buntting, C., Cakir, M., EL-Deghaidy, H., ElZorkani, A., Gaigher, E [Estelle], Guo, S [Shuchen], Hakanen, A., Hamed Al-Lal, S., Han-Tosunoglu, C., . . . Zhou, Q. (2019). An international collaborative investigation of beginning seventh grade students' understandings of scientific inquiry: Establishing a baseline. *Journal of Research in Science Teaching*, 56(4), 486–515. <https://doi.org/10.1002/tea.21512>
- Lederman, N. G. (2006). Syntax of nature of science within inquiry and science instruction. In L. B. Flick & N. G. Lederman (Hrsg.), *Science & Technology Education Library: Bd. 25*.

- Scientific Inquiry and Nature of Science: Implications for Teaching, Learning, and Teacher Education* (Bd. 25, S. 301–317). Springer. https://doi.org/10.1007/978-1-4020-5814-1_14
- Lederman, N. G. (2007). Nature of science: Past, present, and future. In S. K. Abell & N. G. Lederman (Hrsg.), *Handbook of research on science education* (Reprinted., S. 831–879). Routledge.
- Lederman, N. G., Abd-El-Khalick, F., Bell, R. L. & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, 39(6), 479–521.
- Lee, G., Kwon, J., Park, S.-S., Kim, J.-W., Kwon, H.-G. & Park, H.-K. (2003). Development of an instrument for measuring cognitive conflict in secondary-level science classes. *Journal of Research in Science Teaching*, 40(6), 585–603. <https://doi.org/10.1002/tea.10099>
- Lee, S. W.-Y., Luan, H., Lee, M.-H., Chang, H.-Y., Liang, J.-C., Lee, Y.-H., Lin, T.-J., Wu, A.-H., Chiu, Y.-J. & Tsai, C.-C. (2021). Measuring epistemologies in science learning and teaching: A systematic review of the literature. *Science Education*, 105(5), 880–907. <https://doi.org/10.1002/sce.21663>
- Lewis, L. K., Williams, M. T. & Olds, T. S. (2011). Development and psychometric testing of an instrument to evaluate cognitive skills of evidence based practice in student health professionals. *BMC medical education*, 11, 77. <https://doi.org/10.1186/1472-6920-11-77>
- Li, S. C., Law, N. & Lui, K. F. A. (2006). Cognitive perturbation through dynamic modelling: A pedagogical approach to conceptual change in science. *Journal of Computer Assisted Learning*, 22(6), 405–422. <https://doi.org/10.1111/j.1365-2729.2006.00187.x>
- Lichtenegger, K. (2015). Atome, Kerne, Elementarteilchen. In Lichtenegger (Hrsg.), *Schlüsselkonzepte zur Physik* (S. 113–135). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-8274-2385-6_6
- Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, 11(4-5), 357.
- Lin, F. (2018). Characterizing elementary-school students' epistemology of science: Science as collective theory-building process. *The Asia-Pacific Education Researcher*, 27(6), 487–498. <https://doi.org/10.1007/s40299-018-0411-4>
- Lin, J.-W., Yen, M.-H., Liang, J.-C., Chiu, M.-H. & Guo, C.-J. (2016). Examining the factors that influence students' science learning processes and their learning outcomes: 30 years of conceptual change research. *EURASIA Journal of Mathematics, Science & Technology Education*, 12(10), 2617–2646. <https://doi.org/10.12973/eurasia.2016.000600a>
- Liptow, J. (2022). The epistemological relevance of conceptual change. *Grazer Philosophische Studien*, 99(3), 449–473. <https://doi.org/10.1163/18756735-00000175>
- Martensen, M. & Demuth, R. (2008). Wissensdiagnose mit Concept Maps: Entwicklung eines Expertennetzes zum Donator-Akzeptor-Konzept. *Praxis der Naturwissenschaften – Chemie in der Schule*, 57(3), 37–39.
- Mason, L. (2000). Role of anomalous data and epistemological beliefs in middle school students' theory change about two controversial topics. *European Journal of Psychology of Education*, 15(3), 329–346. <https://doi.org/10.1007/BF03173183>
- Mason, L. (2002). Developing epistemological thinking to foster conceptual change in different domains. In M. Limón & L. Mason (Hrsg.), *Reconsidering conceptual change: Issues in theory and practice* (301–335). Kluwer Academic Publishers.
- Mason, L. & Bromme, R. (2010). Situating and relating epistemological beliefs into metacognition: studies on beliefs about knowledge and knowing. *Metacognition and Learning*, 5(1), 1–6. <https://doi.org/10.1007/s11409-009-9050-8>

- Mayring, P. (2010). *Qualitative Inhaltsanalyse Grundlagen und Techniken* (11. Aufl.). Beltz.
- McCluskey, A. & Bishop, B. (2009). The adapted fresno test of competence in evidence-based practice. *The Journal of continuing education in the health professions*, 29(2), 119–126. <https://doi.org/10.1002/chp.20021>
- McComas, W. F. & Olson, J. K. (1998/2002). The nature of science in international science education standards documents. In W. F. McComas (Hrsg.), *Science & Technology Education Library. The nature of science in science education* (Bd. 5, S. 41–52). Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47215-5_2 (Erstveröffentlichung 1998)
- McMaster University (o.D.). About us. Abgerufen am 13.12.2022, von <https://hei.mcmaster.ca/about-us>
- Melnik, B. M., Fineout-Overholt, E. & Mays, M. Z. (2008). The evidence-based practice beliefs and implementation scales: psychometric properties of two new instruments. *Worldviews on Evidence-Based Nursing*, 5(4), 208–216. <https://doi.org/10.1111/j.1741-6787.2008.00126.x>
- Merchant, C. (1989). Entwurf einer ökologischen Ethik (E. Schuhmacher, Übers.). In H.-P. Dürr & W. C. Zimmerli (Hrsg.), *Geist und Natur: Über den Widerspruch zwischen naturwissenschaftlichen Erkenntnissen und philosophischer Welterfahrung* (Zweite Auflage, S. 135–144). Scherz Verlag.
- Verordnung zum Schutz vor Neuinfizierungen mit dem Coronavirus SARS-CoV-2 (Coronaschutzverordnung) (2020 & i.d.F.v. ab dem 18. Dezember 2020 gültige Fassung).
- Verordnung über den Bildungsgang und die Abiturprüfung in der gymnasialen Oberstufe (1998 & i.d.F.v. 23.03.2022). <https://bass.schul-welt.de/9607.htm>
- Gesetz über die Ausbildung für Lehrämter an öffentlichen Schulen (Lehrerausbildungsgesetz) (2009 & i.d.F.v. 01.06.2023). <https://bass.schul-welt.de/9767.htm#1-8p20>
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen. (2019). *Leitlinie Bildung für nachhaltige Entwicklung*. https://www.schulministerium.nrw/sites/default/files/documents/Leitlinie_BNE.pdf
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen. (2021a). *Lehrpläne für die Primarstufe in Nordrhein-Westfalen: Sachunterricht*.
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen. (2021b). *Multiprofessionelle Teams im Gemeinsamen Lernen an Grundschulen und weiterführenden Schulen*. Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (MSB NRW). https://www.schulministerium.nrw/system/files/media/document/file/210505_erlass_multiprofessionelle_teams_gemeinsames_lernen_grundschulen_weiterfuehrende_schulen.pdf
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2007). *Lehrplan für das Berufskolleg in Nordrhein-Westfalen - Chemielaborantin/Chemielaborant*. Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (MSB NRW). https://www.berufsbildung.nrw.de/cms/upload/_lehrplaene/a/chemielaboranten.pdf
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2019). *Kernlehrplan für die Sekundarstufe I Gymnasium/Gesamtschulen für Nordrhein-Westfalen*.
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2022). *Kernlehrplan für die Sekundarstufe II Gymnasium/Gesamtschule in Nordrhein-Westfalen - Chemie*. Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (MSB NRW).
- Mittelsten Scheid, N. & Hößle, C. (2008). Wie Schüler unter Verwendung syllogistischer Elemente argumentieren Eine empirische Studie zu Niveaus von Argumentation im

- naturwissenschaftlichen Unterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 145–165. https://archiv.ipn.uni-kiel.de/zfdn/pdf/14_008_MittelstenScheid.pdf
- Muis, K. R., Bendixen, L. D. & Haerle, F. C. (2006). Domain-generality and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educational Psychology Review*, 18(1), 3–54. <https://doi.org/10.1007/s10648-006-9003-6>
- Murphy, P. K., Alexander, P. A., Greene, J. A. & Hennessey, M. N. (2012). Examining epistemic frames in conceptual change research: Implications for learning and instruction. *Asia Pacific Education Review*, 13(3), 475–486. <https://doi.org/10.1007/s12564-011-9199-0>
- National Research Council. (1996). *National science education standards*. National Academic Press.
- Neumann, I [Irene] & Kremer, K. H. (2013). Nature of science und epistemologische Überzeugungen – Ähnlichkeiten und Unterschiede. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 219–232.
- Niaz, M. (1998). A lakatosian conceptual change teaching strategy based on student ability to build models with varying degrees of conceptual understanding of chemical equilibrium. *Science & Education*, 7(2), 107–127.
- Nitz, S. & Fechner, S. (2018). Mentale Modelle. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (Bd. 10, S. 69–86). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-56320-5_5
- Novak, J. D. & Gowin, D. B. (1984). *Learning how to learn*. Cambridge University Press.
- Nyguen-Kim, Mai Thi. *Corona hat meine Meinung geändert*. YouTube. <https://www.youtube.com/watch?v=Nn2rJrKwENI>
- Odebiyi, O. M. & Choi, Y.-J. (2022). The challenges of measuring epistemic beliefs across cultures: Evidence from Nigerian teacher candidates. *Teaching Education*, 33(2), 214–236. <https://doi.org/10.1080/10476210.2020.1844650>
- Pacaci, C., Ustun, U. & Ozdemir, O. F. (2023). Effectiveness of conceptual change strategies in science education: A meta-analysis. *Journal of Research in Science Teaching*, 1–63. <https://doi.org/10.1002/tea.21887>
- Pant, H. A. (2014). Aufbereitung von Evidenz für bildungspolitische und pädagogische Entscheidungen: Metaanalysen in der Bildungsforschung. *Zeitschrift für Erziehungswissenschaft*, 17(S4), 79–99. <https://doi.org/10.1007/s11618-014-0510-9>
- Parchmann, I., Demuth, R., Ralle, B., Paschmann, A. & Huntemann, H. (2001). Chemie im Kontext: Begründung und Realisierung eines Lernens in sinnstiftenden Kontexten. *Praxis der Naturwissenschaften – Chemie in der Schule*, 50(1), 2–7.
- Parchmann, I. & Kuhn, J. (2018). Lernen im Kontext. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 193–208). Springer Berlin Heidelberg.
- Paschmann, A., Oetken, M. & Vries, T. de (2002). Schülervorstellungen zum Sieden: Der Wechsel des Aggregatzustandes als didaktisches Problem. *Der mathematische und naturwissenschaftliche Unterricht*, 55(7), 413–420.
- Patelarou, A. E., Mechili, E. A., Ruzafa-Martinez, M., Dolezel, J., Gotlib, J., Skela-Savič, B., Ramos-Morcillo, A. J., Finotto, S., Jarosova, D., Smodiš, M., Mecugni, D., Panczyk, M. & Patelarou, E. (2020). Educational interventions for teaching evidence-based practice to undergraduate nursing students: A scoping review. *International Journal of Environmental Research and Public Health*, 17(17). <https://doi.org/10.3390/ijerph17176351>

- Phillippi, J. & Lauderdale, J. (2018). A guide to field notes for qualitative research: Context and conversation. *Qualitative health research*, 28(3), 381–388.
<https://doi.org/10.1177/1049732317697102>
- Pintrich, P. R. (2002). Future challenges and directions for theory and research on personal epistemology. In B. K. Hofer & P. R. Pintrich (Hrsg.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (1 Aufl., S. 389–414). Lawrence Erlbaum Associates, Inc.
- Pintrich, P. R., Marx, R. W. & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, 63(2), 167–199.
<https://doi.org/10.3102/00346543063002167>
- Pollmeier, P. (2019). Kreativität beim Auswerten von Daten – Eine Untersuchung zur Einstellungsänderung von Lernenden aufgrund von anomalen Daten [unveröffentlichte Masterarbeit]. Universität Paderborn, Paderborn.
- Pollmeier, P. & Fechner, S. (2021a). Förderung angehender Lehrkräfte im Umgang mit Evidenzen für den naturwissenschaftlichen Unterricht. In M. Kubsch, S. Sorge, J. Arnold & N. Graulich (Hrsg.), *Lehrkräftebildung neu gedacht. Ein Praxishandbuch für die Lehre in den Naturwissenschaften und deren Didaktiken* (S. 150–154). Waxmann Verlag GmbH.
- Pollmeier, P. & Fechner, S. (2021b). Zwischen den Stühlen? – Verknüpfung von Erfahrungen des Praxissemesters mit Theorien im Lehramtsstudium Chemie. Sukzessiven Kompetenzaufbau nach dem Praxissemester gestalten. In C. Caruso, C. Harteis & A. Gröschner (Hrsg.), *Edition Fachdidaktiken. Theorie und Praxis in der Lehrerbildung* (S. 275–290). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-32568-8_16
- Popper, K. R. (2005). *Logik der Forschung* (11. Aufl., durchges. und erg.). *Gesammelte Werke: in deutscher Sprache*; 3. Mohr Siebeck.
- Posner, G. J., Strike, K. A., Hewson, P. W. & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227.
<https://doi.org/10.1002/sce.3730660207>
- Potvin, P. (2023). Response of science learners to contradicting information: a review of research. *Studies in Science Education*, 59(1), 67–108.
<https://doi.org/10.1080/03057267.2021.2004006>
- Potvin, P., Nenciovici, L., Malenfant-Robichaud, G., Thibault, F., Sy, O., Mahhou, M. A., Bernard, A., Allaire-Duquette, G., Blanchette Sarrasin, J., Brault Foisy, L.-M., Brouillette, N., St-Aubin, A.-A., Charland, P., Masson, S., Riopel, M., Tsai, C.-C., Bélanger, M. & Chasténay, P. (2020). Models of conceptual change in science learning: Establishing an exhaustive inventory based on support given by articles published in major journals. *Studies in Science Education*, 56(2), 157–211. <https://doi.org/10.1080/03057267.2020.1744796>
- Qian, G. & Alvermann, D. (1995). Role of epistemological beliefs and learned helplessness in secondary school students' learning science concepts from text. *Journal of Educational Psychology*, 87(2), 282–292. <https://doi.org/10.1037/0022-0663.87.2.282>
- Rädiker, S. & Kuckartz, U. (2019). *Analyse qualitativer Daten mit MAXQDA*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-22095-2>
- Rahman, T. & Lewis, S. E. (2020). Evaluating the evidence base for evidence-based instructional practices in chemistry through meta-analysis. *Journal of Research in Science Teaching*, 57(5), 765–793. <https://doi.org/10.1002/tea.21610>
- Ramos, K. D., Schafer, S. & Tracz, S. M. (2003). Validation of the fresno test of competence in evidence based medicine. *BMJ (Clinical research ed.)*, 326(7384), 319–321.
<https://doi.org/10.1136/bmj.326.7384.319>

- Richter, M. (2020). *Informationen zum angepassten Schulbetrieb in Corona-Zeiten. Hier: Regelungen für die kommende Woche: Schulmail NRW*. Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (MSB NRW). <https://www.schulministerium.nrw/11122020-informationen-zum-angepassten-schulbetrieb-corona-zeiten-hier-regelungen-fuer-die-kommende>
- Riedel, E. & Meyer, H.-J. (2013). *Allgemeine und Anorganische Chemie* (11. Auflage). Walter de Gruyter.
- Rochnia, M., Trempler, K. & Schellenbach-Zell, J. (2020). Vergleich der Forschungs- sowie Praxisorientierung zwischen Lehramts- und Medizinstudium. *ZeHf – Zeitschrift für empirische Hochschulforschung*, 3(2-2019), 123–138. <https://doi.org/10.3224/zehf.v3i2.03>
- Rock, I. (1985). Perception and knowledge. *Acta Psychologica*, 59(1), 3–22. [https://doi.org/10.1016/0001-6918\(85\)90039-3](https://doi.org/10.1016/0001-6918(85)90039-3)
- Rodrigues Salerno, M., Herrmann, F., Debon, L. M., Soldatelli, M. D., Forte, G. C., Bastos, M. D., Sarria, E. E., Antonello, I. C. F. & Mattiello, R. (2019). Brazilian version of the fresno test of competence in evidence-based medicine: A validation study. *Scientia Medica*, 29(1).
- Rothland, M. (2013). Beruf: Lehrer/Lehrerin – Arbeitsplatz: Schule Charakteristika der Arbeitstätigkeit und Bedingungen der Berufssituation. In M. Rothland (Hrsg.), *Belastung und Beanspruchung im Lehrerberuf* (S. 21–39). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-18990-1_2
- Ruiz-Primo, M. A. & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching*, 33(6), 569–600.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B. & Richardson, W. S. (1996). Evidence bases medicine: What it is and what it isn't: It's about integrating individuals clinical expertise and the best external evidence. *British Medical Journal*, 312(7023), 71–72.
- Samarapungavan, A. (1992). Children's judgments in theory choice tasks: Scientific rationality in childhood. *Cognition*, 45(1), 1–32. [https://doi.org/10.1016/0010-0277\(92\)90021-9](https://doi.org/10.1016/0010-0277(92)90021-9)
- Sandmann, A. (2014). Lautes Denken - die Analyse von Denk-, Lern-, und Problemlöseprozessen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 179–188). Springer.
- Sandoval, W. A. (2017). Situating practices of epistemic cognition. In T. G. Amin & O. Levrini (Hrsg.), *Converging perspectives on conceptual change: Mapping an emerging paradigm in the learning sciences* (1 Aufl., S. 253–260). Routledge.
- Schaeper, H. & Weiß, T. (2016). The conceptualization, development, and validation of an instrument for measuring the formal learning environment in higher education. In H.-P. Blossfeld, J. von Maurice, M. Bayer & J. Skopek (Hrsg.), *Methodological issues of longitudinal surveys: The example of the National Educational Panel Study* (S. 267–290). Springer Science and Business Media. https://doi.org/10.1007/978-3-658-11994-2_16
- Schecker, H. (2014). Überprüfung der Konsistenz von Itemgruppen mit Cronbachs α . In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (online Zusatzmaterial). Springer. https://www.researchgate.net/publication/313220515_Ueberprufung_der_Konsistenz_von_Itemgruppen_mit_Cronbachs_alpha
- Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, 82(3), 498–504. <https://doi.org/10.1037/0022-0663.82.3.498>
- Schommer, M., Crouse, A. & Rhodes, N. (1992). Epistemological beliefs and mathematical text comprehension: Believing it is simple does not make it so. *Journal of Educational Psychology*, 84(4), 435–443. <https://doi.org/10.1037/0022-0663.84.4.435>

- Schommer-Aikins, M. (2002). An evolving theoretical framework for an epistemological belief system. In B. K. Hofer & P. R. Pintrich (Hrsg.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (1 Aufl., S. 103–118). Lawrence Erlbaum Associates, Inc.
- Schommer-Aikins, M., Duell, O. K. & Hutter, R. (2005). Epistemological beliefs, mathematical problem-solving beliefs, and academic performance of middle school students. *The Elementary School Journal*, 105(3), 289–304. <https://doi.org/10.1086/428745>
- Schommer-Aikins, M., Duell, O. K. & Barker, S. (2003). Epistemological beliefs across domains using biglan's classification of academic disciplines. *Research in Higher Education*, 44(3), 347–366. <https://doi.org/10.1023/A:1023081800014>
- Schrader, C. (o.D.). *Kungelei oder klare Beweislage - was ist das eigentlich, „Konsens“?* <https://www.klimafakten.de/meldung/kungelei-oder-klare-beweislage-was-ist-das-eigentlich-konsens>
- Schrader, J. (2014). Analyse und Förderung effektiver Lehr-Lernprozesse unter dem Anspruch evidenzbasierter Bildungsreform. *Zeitschrift für Erziehungswissenschaft*, 17(2), 193–223. <https://doi.org/10.1007/s11618-014-0540-3>
- Schraw, G., Dunkle, M. E. & Bendixen, L. D. (1995). Cognitive processes in well-defined and ill-defined problem solving. *Applied Cognitive Psychology*, 9(6), 523–538. <https://doi.org/10.1002/acp.2350090605>
- Schroeder, M., McKeough, A., Graham, S. A. & Norris, S. P. (2019). Students' views of uncertainty in formal and personal science. *Research in Science and Technological Education*, 37(2), 239–257. <https://doi.org/10.1080/02635143.2018.1541878>
- Schwandt, T. (2009). Toward a practical theory of evidence for evaluation. In S. I. Donaldson, C. A. Christie & M. M. Mark (Hrsg.), *What counts as credible evidence in applied research and evaluation practice?* (S. 197–212). Sage. <https://doi.org/10.4135/9781412995634.d18>
- Schwartz, R. S., Lederman, N. G. & Crawford, B. A. (2004). Developing views of nature of science in an authentic context: An explicit approach to bridging the gap between nature of science and scientific inquiry. *Science Education*, 88(4), 610–645.
- Scurlock-Evans, L., Upton, P. & Upton, D. (2014). Evidence-based practice in physiotherapy: a systematic review of barriers, enablers and interventions. *Physiotherapy*, 100(3), 208–219. <https://doi.org/10.1016/j.physio.2014.03.001>
- Shaneyfelt, T., Baum, K. D., Bell, D., Feldstein, D., Houston, T. K., Kaatz, S., Whelan, C. & Green, M. (2006). Instruments for evaluating education in evidence-based practice: a systematic review. *JAMA*, 296(9), 1116–1127. <https://doi.org/10.1001/jama.296.9.1116>
- Sinatra, G. M. & Chinn, C. A. (2012). Thinking and reasoning in science: Promoting epistemic conceptual change. In A. G. Bus, S. Major & H. L. Swanson (Hrsg.), *Application to learning and teaching* (S. 257–282). American Psychological Association. <https://doi.org/10.1037/13275-011>
- Sinatra, G. M. & Mason, L. (2013). Beyond knowledge: Learner characteristics influencing conceptual change. In S. Vosniadou (Hrsg.), *Educational psychology handbook series. International handbook of research on conceptual change* (Second edition, S. 560–582). Routledge.
- Sinatra, G. M. & Pintrich, P. R. (Hrsg.). (2003). *Intentional conceptual change*. L. Erlbaum.
- Smith, C. L. & Wiser, M. (2015). On the importance of epistemology-disciplinary core concept interactions in LPs. *Science Education*, 99(3), 417–423. <https://doi.org/10.1002/sce.21166>
- Spektrum.de (o.D.) Syllogismus. In *Spektrum Lexikon der Psychologie*. Abgerufen am 05.12.2022, von <https://www.spektrum.de/lexikon/psychologie/syllogismus/15150>

- Spitzer, M. (2010). *Medizin für die Bildung: Ein Weg aus der Krise*. Spektrum-Akademischer-Verlag-Sachbuch. Spektrum Akademischer Verlag.
- Stark, R. (2017). Probleme evidenzbasierter bzw. -orientierter pädagogischer Praxis. *Zeitschrift für Pädagogische Psychologie*, 31(2), 99–110. <https://doi.org/10.1024/1010-0652/a000201>
- Stelter, A. & Miethe, I. (2019). Forschungsmethoden im Lehramtsstudium – aktueller Stand und Konsequenzen. *Erziehungswissenschaft*, 30(58 (1-2019)), 25–33. <https://doi.org/10.3224/ezw.v30i1.03>
- Strike, K. A. & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. A. Duschl & R. J. Hamilton (Hrsg.), *Philosophy of science, cognitive psychology, and educational theory and practice* (147-176). SUNY.
- Swackhamer, G., Dukerich, L. & Hestenes, D. (2005). *Energy Concept Inventory*. <https://energyeducation.eku.edu/sites/energyeducation.eku.edu/files/EnergyConceptInventory.pdf>
- Tack, H. & Vanderlinde, R. (2016). Measuring teacher educators' researcherly disposition: Item development and scale construction. *Vocations and Learning*, 9(1), 43–62. <https://doi.org/10.1007/s12186-016-9148-5>
- Taştan Kırık, Ö. & Boz, Y. (2012). Cooperative learning instruction for conceptual change in the concepts of chemical kinetics. *Chemistry Education Research and Practice*, 13(3), 221–236. <https://doi.org/10.1039/C1RP90072B>
- Tippelt, R. & Reich-Claassen, J. (2010). Stichwort „Evidenzbasierung“. *weiter bilden. DIE Zeitschrift für Erwachsenenbildung*(17), 22–23. <https://doi.org/10.3278/DIE1004W022>
- Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education*, 8(6), 951–970. <https://doi.org/10.1007/s10763-010-9203-x>
- Toulmin, S. E. (2003). *The uses of argument* (Repr). Cambridge University Press. <https://www.cambridge.org/core/books/uses-of-argument/26CF801BC12004587B66778297D5567C> <https://doi.org/10.1017/CBO9780511840005>
- Urhahne, D. (2006). Die Bedeutung domänenspezifischer epistemologischer Überzeugungen für Motivation, Selbstkonzept und Lernstrategien von Studierenden. *Zeitschrift für Pädagogische Psychologie*, 20(3), 189–198. <https://doi.org/10.1024/1010-0652.20.3.189>
- Urhahne, D. & Hopf, M. (2004). Epistemologische Überzeugungen in den Naturwissenschaften und ihre Zusammenhänge mit Motivation, Selbstkonzept und Lernstrategien. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 71–87.
- Urhahne, D. & Kremer, K. (2023). Specificity of epistemic beliefs across school subject domains. *Educational Psychology*, 43(2-3), 99–118. <https://doi.org/10.1080/01443410.2023.2179605>
- Völzke, K. (2012). *Lautes Denken bei kompetenzorientierten Diagnoseaufgaben zur naturwissenschaftlichen Erkenntnisgewinnung: Ausgezeichnet mit dem Martin-Wagenschein-Preis 2012 des ZLB. Reihe Studium und Forschung: Bd. 20*. Kassel Univ. Press. <https://kobra.uni-kassel.de/handle/123456789/2012061541272>
- Vorholzer, A. (2016). *Wie lassen sich Kompetenzen des Experimentellen Denkens und Arbeitens Fördern? Eine Empirische Untersuchung der Wirkung Eines Expliziten und Eines Impliziten Instruktionsansatzes. Studien Zum Physik- und Chemielernen Ser: v.197*. Logos Verlag Berlin. <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=5247125>
- Vorholzer, A. & Aufschnaiter, C. von (2019). Guidance in inquiry-based instruction – an attempt to disentangle a manifold construct. *International Journal of Science Education*, 41(11), 1562–1577. <https://doi.org/10.1080/09500693.2019.1616124>

- Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, 4(1), 45–69.
- Vosniadou, S. (2014). Examining cognitive development from a conceptual change point of view: The framework theory approach. *European Journal of Developmental Psychology*, 11(6), 645–661. <https://doi.org/10.1080/17405629.2014.921153>
- Vosniadou, S. & Brewer, W. F. (1987). Theories of knowledge restructuring in development. *Review of Educational Research*, 57(1), 51–67. <https://doi.org/10.2307/1170356>
- Vosniadou, S. & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24, 535–585.
- Vosniadou, S. & Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive science*, 18(1), 123–183. https://doi.org/10.1207/s15516709cog1801_4
- Vosniadou, S. & Ioannides, C. (1998). From conceptual development to science education: A psychological point of view. *International Journal of Science Education*, 20(10), 1213–1230. <https://doi.org/10.1080/0950069980201004>
- Vosniadou, S., Ioannides, C., Dimitrakopoulou, A. & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction*, 11(4-5), 381–419. [https://doi.org/10.1016/S0959-4752\(00\)00038-4](https://doi.org/10.1016/S0959-4752(00)00038-4)
- Vosniadou, S. & Mason, L. (2012). Conceptual change induced by instruction: A complex interplay of multiple factors. In K. R. Harris (Hrsg.), *APA handbooks in psychology. APA educational psychology handbook: Vol. 2* (1st ed (Online-ausg.), S. 221–246). American Psychological Association. <https://doi.org/10.1037/13274-009>
- Vosniadou, S. & Skopeliti, I. (2014). Conceptual change from the framework theory side of the fence. *Science & Education*, 23(7), 1427–1445. <https://doi.org/10.1007/s11191-013-9640-3>
- Voss, T. (2022). Not useful to inform teaching practice? Student teachers hold skeptical beliefs about evidence from education science. *Frontiers in Education*, 7, Artikel 976791. <https://doi.org/10.3389/feduc.2022.976791>
- Weber, E. U. & Stern, P. C. (2011). Public understanding of climate change in the United States. *The American psychologist*, 66(4), 315–328. <https://doi.org/10.1037/a0023253>
- Weber, K. (2018). *Entwicklung und Validierung einer Learning Progression für das Konzept der chemischen Reaktion in der Sekundarstufe I* [Dissertation, Universität Duisburg-Essen; Logos Verlag Berlin, Berlin]. WorldCat.
- Wenglein, S., Bauer, J., Heininger, S. & Prenzel, M. (2015). Kompetenz angehender Lehrkräfte zum Argumentieren mit Evidenz: Erhöht ein Training von Heuristiken die Argumentationsqualität? *Unterrichtswissenschaft*, 43, 209–224.
- Wilkes, T. & Stark, R. (2022). Probleme evidenzorientierter Unterrichtspraxis. *Unterrichtswissenschaft*. Vorab-Onlinepublikation. <https://doi.org/10.1007/s42010-022-00150-1>
- Yeo, S. & Zadnik, M. (2001). Introductory thermal concept evaluation: assessing students' understanding. *The Physics Teacher*, 39(8), 496–504. <https://doi.org/10.1119/1.1424603>
- Zeng, X., Zhang, Y., Kwong, J. S. W., Zhang, C., Li, S., Sun, F., Niu, Y. & Du, L. (2015). The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. *Journal of evidence-based medicine*, 8(1), 2–10. <https://doi.org/10.1111/jebm.12141>
- Zentrum für Bildungsforschung und Lehrerbildung PLAZ. (2017). *Das Praxissemester in der Ausbildungsregion der Universität Paderborn*. Zentrum für Bildungsforschung und Lehrerbildung PLAZ; Universität Paderborn. https://plaz.uni-paderborn.de/fileadmin/plaz/Praxisphasen/Praxissemester_MA/Infobroschueren/Allgemeiner_Teil.pdf

10.2 Abbildungsverzeichnis

Abbildung 1 - Model of argument nach Toulmin (2003).	7
Abbildung 2 - Übersicht Begriffsklärung Epistemologie	22
Abbildung 3 - Dimensionen epistemologischer Überzeugungen nach Hofer und Pintrich (1997, S. 119)	28
Abbildung 4 - Theory of integrated Domains in Epistemology (TIDE) nach Muis et al. (2006, S. 30)	32
Abbildung 5 - Einbettung mentaler Modelle in framework theories nach Vosniadou (1994)	44
Abbildung 6 - Umgangsarten mit anomalen Daten nach Chinn und Brewer (1993b, 1998)	47
Abbildung 7 - Bestandteile von Nature of science nach McComas und Olson (1998/2002, S. 50). Die Größe der Kreise gibt den relativen Anteil am Konstrukt NOS an.	55
Abbildung 8 - Studiendesign der Interventionsstudie	72
Abbildung 9 - Vergleich der Zeitgestaltung der Intervention für beide Seminare.	74
Abbildung 10 - Modifiziertes Model of Argument (Toulmin, 2003)	130
Abbildung 11 - Studiendesign zu Teilstudie 1	135
Abbildung 12 - Zustandsdiagramm von Wasser (Riedel & Meyer, 2013, S. 145)	140
Abbildung 13 - Beispiel eines Model-of-data von Expert*innen zum Experiment "Löslichkeit von Salz in Wasser und Öl". Orange Nummerierung dient nur dem Überblick und der genaueren Beschreibung im Text.	142
Abbildung 14 - Expert*innen Model-of-data zu Experiment 2, Beobachtungsaspekte (orange Kästen), Konzeptaspekte (grüne Kästen), essentielle Aspekte (rot umrandete Kästen), durch unmögliche kausale Verknüpfungen unmögliche Aspekte (schraffierte Kästen).....	155
Abbildung 15 - Überblick über bestimmte Parameter sowie gebildete quantitative Maße zur Auswertung der MOD	156

10.3 Tabellenverzeichnis

Tabelle 1 – Gemeinsamkeiten und Unterschiede von Gesundheitssektor und Bildungsbereich	14
Tabelle 2 – Evidence-use Mechanismen nach Breckon und Dodson (2016, S. 6)	19
Tabelle 3 – Herausforderungen im Bereich epistemologischer Überzeugungen nach Hofer und Pintrich (1997).....	27
Tabelle 4 – Aspekte von Nature of science nach N. G. Lederman (2006)	56
Tabelle 5 – Aufteilung der Stichprobe auf unterschiedliche Lehramtsstudiengänge inkl. Unterscheidung nach Geschlecht	66
Tabelle 6 – Belegungen der Unterrichtsfächer innerhalb der Unterrichtsfächer-Bündel für die Stichprobe.....	67
Tabelle 7 – Anzahl der Fälle pro Fächerbündelkombination	68
Tabelle 8 – Aufteilung der Stichprobe auf die unterschiedlichen Begleitforschungsseminare.....	69
Tabelle 9 – Verteilung der Gesamtstichprobe des Projekts RiTE auf Fachbereiche und Länder.....	78
Tabelle 10 – Vergleich der Items des Fresno-Tests (Ramos et al., 2003), des TUBE-ST (Cholewa et al., 2015) sowie des EIDM-ET	79
Tabelle 11 – Reliabilität des EIDM-ET an Prä-, Post- und follow-up-Zeitpunkt für die Stichprobe dieser Studie, ergänzt um die zweite Kohorte des Lehramtsstudiums Chemie (B.Ed.) sowie die Gesamtstichprobe im Projekt Research in Teacher Education.	83
Tabelle 12 – Itemschwierigkeiten und -varianz für EIDM-ET sowie Auswahl der Items für das gekürzte Instrument.....	86
Tabelle 13 – Englischsprachige Items der evidence-based practice beliefs/implementation scale.....	91
Tabelle 14 – Faktoren der EBP beliefs scale, zugehörige Items sowie Reliabilität zu verschiedenen Zeitpunkten, bzw. innerhalb verschiedener Substichproben	92
Tabelle 15 – Faktoren der EBP beliefs scale, zugehörige Items sowie Reliabilität der Subskalen.....	93
Tabelle 16 – Faktoren der EBP implementation scale, zugehörige Items sowie Reliabilität der Subskalen	95
Tabelle 17 – Deskriptive Statistik für den gekürzten EIDM-ET an Prä- und Post-Zeitpunkt	98

Tabelle 18 – Deskriptive Statistik für einzelne Items des gekürzten EIDM-ET für Prä- und Post-Zeitpunkt .	98
Tabelle 19 – Mittelwertsunterschiede im gekürzten EIDM-ET und Effektstärken für Fächerbündel	100
Tabelle 20 – Deskriptive Statistik für EBP beliefs scale zu Prä- und Post-Zeitpunkt	100
Tabelle 21 – Ergebnisse einer ANOVA zur Untersuchung von Unterschieden der Sub-Skalen der EBP beliefs scale für verschiedene Studiengänge	101
Tabelle 22 – Ergebnisse einer ANOVA zur Untersuchung von Unterschieden der Faktoren der EBP beliefs scale für verschiedene Fächerbündel-Kombinationen	102
Tabelle 23 – Ergebnisse einer ANOVA zur Untersuchung von Unterschieden der Faktoren der EBP beliefs scale für verschiedene belegte Begleitforschungsseminare	102
Tabelle 24 – Deskriptive Statistik für die Faktoren der EBP implementation scale für den Post-Zeitpunkt	103
Tabelle 25 – Deskriptive Statistik des vollständigen EIDM-ET zum Prä- und Post-Zeitpunkt.....	112
Tabelle 26 – Deskriptive Statistik für die einzelnen Items des vollständigen EIDM-ET zu Prä- und Post-Zeitpunkt	112
Tabelle 27 – Deskriptive Statistik für die Sub-Skalen der EBP beliefs scale zu Prä- und Post-Zeitpunkt.....	114
Tabelle 28 – Fallauswahl für die Analyse in Anlehnung an die Verteilung innerhalb der Gesamtstichprobe	148
Tabelle 29 – Beispielitems des Fachwissenstests.....	150
Tabelle 30 – Testgüte der Subskalen des eingesetzten Fachwissenstest vor Bearbeitung.....	151
Tabelle 31 – Testgüte der Subskalen des eingesetzten Fachwissenstest nach Bearbeitung	151
Tabelle 32 – Vergabe von Kodierungen sowie Intercoder-Übereinstimmung bzgl. Verknüpfungen innerhalb der Model-of-data	159
Tabelle 33 – Vergabe von Kodierungen sowie Intercoder-Übereinstimmung bzgl. inhaltlicher Aspekte innerhalb der Model-of-data	160
Tabelle 34 – Deskriptive Statistik für essentielle & weitere Beobachtungs-/Konzeptaspekte	161
Tabelle 35 – Erstellte essentielle Aspekte der Proband*innen	164
Tabelle 36 – Unterkategorien für weitere/inadäquate Beobachtungs-/Konzeptaspekte	166
Tabelle 37 – Darstellung anomaler Beobachtungen zu Experiment 1 und 2	168
Tabelle 38 – Umgang der Proband*innen mit anomalen Beobachtungen in Experiment 1 und 2.....	171
Tabelle 39 – Anzahl, prozentualer Anteil sowie Adäquatheit der gebildeten Verknüpfungen.....	172
Tabelle 40 – Ergebnisse eines Wilcoxon-Tests für Unterschiede in der Anzahl gebildeter Verknüpfungen zwischen Exp. 1 und 2.....	173
Tabelle 41 – Verteilung inadäquater Verknüpfungen auf unterschiedliche Ursachen	175
Tabelle 42 – Deskriptive Statistik für die Selbsteinschätzung der Proband*innen für unterschiedliche Inhaltsbereiche	178
Tabelle 43 – Deskriptive Statistik für die Einzelitems der Siedepunkt-Skala.....	179
Tabelle 44 – Vergabe von Kodierungen sowie Intercoder-Übereinstimmung bzgl. Abweichungen und verbalisiert, aber nicht notierter Aspekte im lauten Denken	180
Tabelle 45 – Deskriptive Statistik der Abweichungen sowie Ergänzungen von Verknüpfungen und Aspekten für die beiden Experimente	181
Tabelle 46 – Bündelung der Umgangsarten mit anomalen Daten in Anlehnung an Chinn und Brewer (1993b, 1998).....	187

11. Anhang

I. Erster empirischer Teil

i.	Lernmaterialien und Aufgaben der Interventionsphase.....	227
ii.	Evidence-informed decision making in education test (EIDM-ET) – Langfassung	273
iii.	Evidence-informed decision making in education test (EIDM-ET) – Kursfassung.....	280
iv.	Grading rubric des EIDM-ET.....	285
v.	Interviewleitfaden der Interventionsphase.....	291
vi.	Kodiermanual zum leitfadengestützten Interview der Interventionsphase.....	293
vii.	Checkliste zur Analyse von Unterrichtsentwürfen.....	297
viii.	Checkliste zur Analyse von Unterrichtsmitschauen.....	300

II. Zweiter empirischer Teil

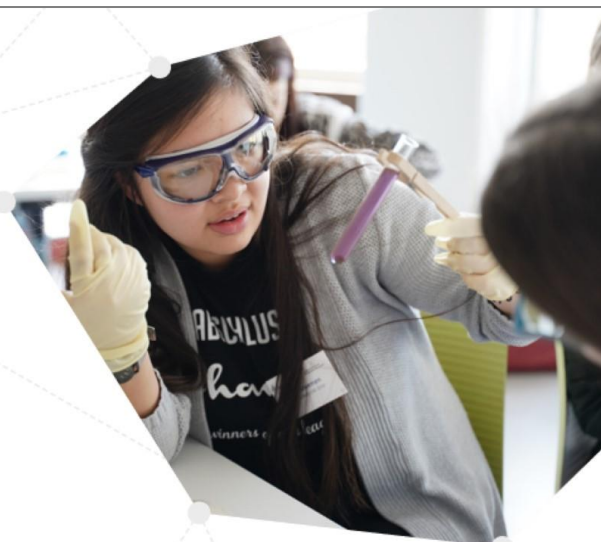
i.	Vorlage zur Erstellung von Model-of-data.....	303
ii.	Materialien der Trainingsphase.....	305
iii.	Videos zu den Experimenten der Trainings- und Modellierungsphase.....	320
iv.	Expert*innen-MOD zu den Experimenten der Modellierungsphase.....	321
v.	Gefährdungsbeurteilungen zu den eingesetzten Experimenten.....	324
vi.	Kodiermanual für MOD und lautes Denken.....	331
vii.	Fachwissenstest.....	336
viii.	Datenblatt für Proband*innen inkl. Vorlage für Notizen während des lauten Denkens.....	342

I. Erster empirischer Teil

i. Lernmaterialien und Aufgaben der Interventionsphase

Im Folgenden werden die Lernmaterialien inkl. der Aufgaben für die Interventionsphase dargestellt. Es handelt sich hier um die Präsentationen im Seminar „Chemische Bildung und chemiedidaktische Konzeptionen“. Die jeweiligen Aufgaben sind mit einer roten Markierung sowie der entsprechenden Nummerierung auf den jeweiligen Folien zu finden.

1. Sitzung



CHEMIEDIDAKTIK

CHEMISCHE BILDUNG UND CHEMIEDIDAKTISCHE KONZEPTIONEN

EINFÜHRUNG

Pascal Pollmeier & Sabine Fechner

Um was geht es?



<https://www.zdf.de/nachrichten/heute/fridays-for-future-wo-steht-die-bewegung-100.html>, <https://truthout.org/articles/as-an-autistic-femme-i-love-greta-thunbergs-resting-autism-face/>, <https://www.jass-potsdam.de/en/blog/2019/09/responding-fridays-future-and-youth-movement-climate-justice>, <https://www.landgard.de/serviceangebote/nachhaltigkeit>, <https://taz.de/Kommentar-Fridays-for-Future/5583849/>

Fünf Kerninfos zum Klimawandel in 20 Worten

ER IST REAL.

WIR SIND DIE URSACHE.

ER IST GEFÄHRLICH.

DIE FACHLEUTE SIND SICH EINIG.

WIR KÖNNEN NOCH ETWAS TUN.

Deutsches Klima-Konsortium; Deutsche Meteorologische Gesellschaft; Deutscher Wetterdienst; Extremwetterkongress Hamburg; Helmholtz-Klima-Initiative; klimafakten.de (Hrsg.). (2020) Was wir heute übers Klima wissen.

3

Ziele der heutigen Sitzung

Organisation

- Prüfungsanforderungen
- Seminarplan
- Online-Abgaben

Forschungsprojekt RiTE

- Vorstellung

Nachhaltigkeit

- Definition
- Gruppenzusammensetzung
- Themenwahl

4

Ziele des Seminars

Entwicklung und Erprobung eines Schülermoduls zum Thema „Nachhaltigkeit“

Die **SuS** können...

...einen chemiebezogenen Aspekt aus einem modernen Thema auf der Grundlage des erworbenen Fachwissens **bewerten** (Argumente formulieren).

Die **StuSt** können...

...in Projektarbeit ein Schülermodul eigenständig entwickeln und erproben, indem sie ihr Thema
1) fachlich vertiefen, 2) mit erhobenen Schülervorstellungen abgleichen und 3) didaktisch strukturieren.

5

Prüfungsleistung

- Vertiefungsmodul Fachdidaktik Chemie Teil 2
 - Chemische Bildung (WS)
 - Spezielle Themen: Digitale Medien (SoSe)
- **Prüfungsleistung → Modulprüfung**
 - Mündliche Prüfung (45 Minuten) zu allen Inhalten der Seminare
- **Qualifizierte Teilnahme**
 - Entwicklung, Erprobung und Dokumentation des Moduls (Online-Abgaben)

6

Seminarinfos

- Projekte finden in 2er/3er-Gruppen statt (→ Eintragung in PANDA)
- Die Gruppen sind für ihre Arbeitsaufteilung und Organisation selbst verantwortlich (siehe Checkliste PANDA)
- Online-Abgaben sind in PANDA zu finden (Deadlines dort vermerkt)
- Synchrone Sitzungstermine sind in Panda angegeben
- Termine zum Experimentieren im Labor werden über PANDA vereinbart

7

Struktur des Projektes

„normales“ Seminar

- Studierendengruppen entwickeln Schüler*innen Modul
- Schülergruppen kommen in die Universität
- Schüler*innen experimentieren in der Uni
- Durchführung wird durch Studierendengruppen reflektiert
- Moduldurchführungen als Prüfungsgrundlage

„Corona“ Seminar

- Studierendengruppen entwickeln Schüler*innen Modul
- Module werden digital durchgeführt
- Durchführung wird durch Studierendengruppen reflektiert
- Moduldurchführungen als Prüfungsgrundlage

Bis auf die Durchführung in der Universität bleibt die Planung bestehen.

8

Struktur des Projektes

Die digitale Durchführung kann unterschiedlich gestaltet werden.

→ **Digitale blended-learning Angebote**

Zum Beispiel:

Eröffnung (synchron)

- Einführung
- Problemaufwurf
- Experiment
- Ausblick
- Techn. Einweisung

WebQuest

- Informationen aufnehmen
- Aufgaben bearbeiten
- ...

Abschluss (synchron)

- Sicherung
- Zusammenfassung
- (Diskussion)
- ...

Rahmenbedingungen:

- Zeitlicher Umfang: Doppelstunde
- Schulklasse muss selbst gefunden werden
- Schüler*innen müssen mit digitalem Endgerät ausgestattet sein (ggf. Computerraum)
- Eröffnung und Abschluss über BigBlueButton

9

Lerngruppe

Für die Erprobung des WebQuests sollen die Projektgruppen sich eigenständig Lerngruppen suchen.

Vor der Erprobung sollen 3 Lernende der Lerngruppe bzgl. Schülervorstellungen interviewt werden. Die so erhobenen Daten fließen wiederum in die Planung des Moduls ein.

Bitte suchen Sie sich zeitnah eine Lernendengruppe und klären Sie ab, dass 3 Lernende für ein kurzes Schülervorstellungs-interview zur Verfügung stehen.

10

WebQuest-Wizard

Was sind WebQuests?

- Übersetzt: „abenteuerliche Spurensuche im Internet“
- Schüler*innen werden als Ausgangspunkte für die Bearbeitung einer Aufgabe, eines Problems usw. Internetquellen zur Verfügung gestellt

Struktur von WebQuests:

1. Einführung
2. Aufgabenstellung
3. Materialien
4. Prozess
5. Evaluation
6. Präsentation

<http://www.webquests.ch/index>

<http://wizard.webquests.ch/docs/Anleitung.pdf>

11

WebQuest-Wizard - Beispiel

Vorstellung des Themas

Arbeitsaufträge, die
bearbeitet werden sollen

Konkrete Hinweise für
die Vorgehensweise

Quellen, die zur Verfügung
stehen

Kritische Reflexion und
Bewertung

Schlusswort



Startseite | Hilfe für Lernende | Hilfe für Lehrkräfte | Impressum |

Die Vielfältigkeit organischer Stoffe

EINFÜHRUNG

Ein Web Quest bietet SchülerInnen die Möglichkeit eigenständig verschiedene Sachverhalte zu erarbeiten.

AUFGABEN

Dafür lässt sich unter den Aspekten der linken Spalte eine Gliederung vorfinden, welche dem Ablauf einer "gewöhnlichen" Unterrichtsstunde ähnelt.

PROZESS

So existiert eine Einführung, in welcher das Thema kurz dargestellt wird, Aufgaben, die von den SchülerInnen bearbeitet werden können, Hinweise zum Arbeitsprozess und natürlich Ressourcen, mit Hilfe dessen die Aufgaben bearbeitet werden können.

RESSOURCEN

Außerdem besteht eine Evaluation, welche die Bewertungskriterien transparent macht und ein Abschluss.

EVALUATION

ABSCHLUSS

http://www.webquests.ch/vielfaeltigkeit_organischer_stoffe.html?page=189654

12

Struktur des Projektes

Umsetzung der Distanzlehre muss individuell abgesprochen werden. Denkbar wäre z.B.:

- Alle Schüler*innen haben ein **eigenes Endgerät** mit Kopfhörern im Raum. Während synchroner Phasen sind alle Mikrofone ausgeschaltet, bis auf Mikrofone der Studierenden. Während asynchroner Phasen können Kameras ausgeschaltet werden, der Chat kann für Fragen genutzt werden.
- Die Schüler*innen nehmen **gemeinsam über ein Endgerät + Beamer** an der Konferenz teil. Der Beamer im Klassenraum zeigt die Studierenden. Während asynchroner Phasen kann ein anderes System genutzt werden.
- ...

13

Struktur des Projektes

Die Durchführung kann auch anders gestaltet werden. Dabei sind synchrone, wie auch asynchrone Modi denkbar.

Verbindliche Vorgaben:

- Das Modul sollte ca. eine Doppelstunde umfassen
- Innerhalb des Moduls sollte ein Experiment eingebettet sein (ob live, Video, etc. ist frei wählbar)
- Es sollten Materialien & Arbeitsaufträge für das Modul erstellt werden.



14

UNIVERSITÄT PADERBORN

CHEMISCHE BILDUNG
CHEMIEDIDAKTISCHE KONZEPTIONEN

Projektarbeit

Was könnte man noch verändern, um es effektiver und motivierender zu machen?

```

graph TD
    PI[Projektinitiative] --> S[Seminar]
    S --> PP[Projektplanung]
    PP --> PD[Projektdurchführung]
    PD --> PA[Präsentation/Aktion]
    PA --> RE[Reflexion & Evaluation]
    RE --> PI
  
```

Erprobungen
Unterrichtsmaterialien

Fachliche Klärung, Schülervorstellungen,
Experimente, Einbettung in den Unterricht

Seminarplan,
Struktur

15

UNIVERSITÄT PADERBORN

RiTE
VORSTELLUNG
RESEARCH IN TEACHER EDUCATION

Forschungsprojekt RiTE

- Forschungsprojekt gefördert durch Erasmus+: **Research in Teacher Education** (<http://ddn-rite.eu/>)
- Ziel:** Lehrer*innen & Lehramtsanwärter*innen fördern und ermutigen evidenzbasierten Unterricht in den MINT-Fächern zu praktizieren.
- Team:** Universitäten Groningen, Southampton, Chester, Poznan & Paderborn
+ Universität Utrecht & DUDOCnetwork NL
- Rolle des Seminars:**
 - Beforschung durch 2 Fragebögen (erste & letzte Sitzung)
 - Kurzinterview
 - Leichte Änderungen an Seminarstruktur
- Unsere Bitte:**
Bitte nehmen Sie an den zwei Fragebögen und dem Kurzinterview teil, um unser Projekt, welches Corona-bedingt mit einigen Einschränkungen zu kämpfen hat, zu unterstützen.

Kofinanziert durch das
Programm Erasmus+
der Europäischen Union ERASMUS+ Programme Agreement number: 2019-1-NL01-KA203-060339

16

Forschungsprojekt RiTE

Unser Zugeständnis:

Wenn Sie an der Studie (2 Fragebögen + Kurzinterview) teilnehmen, entfallen für Sie:

- Abgabe: Dokumentation Interviews
- Abgabe: Checkliste Grafiken
- Abgabe: Literaturliste

Arbeitsschritte ergeben dennoch Sinn, um Ihre Projektarbeit zu strukturieren. An Deadline gebundene Abgaben, sind jedoch in diesem Fall nicht nötig.

Mit den oben stehenden Abgaben würden die meisten Abgaben entfallen. Lediglich Peer-Feedback Prozesse und die Abgabe der Gefährdungsbeurteilungen bleiben obligatorisch.

Forschungsprojekt RiTE

Durchführung des Fragebogens



Gruppenwahl

Bitte bilden Sie 2er – 3er Gruppen. Tragen Sie sich dazu auf PANDA in die jeweiligen Gruppen ein.



19

Nachhaltigkeit

“Humanity has the ability to make human development sustainable – to ensure that it meets the needs of the present without compromising the ability of future generations to meet their own needs.”

UN, 1987

Aufgaben:

1. Definieren Sie schriftlich und in Einzelarbeit, was für Sie Nachhaltigkeit bedeutet.
2. Diskutieren Sie Ihre Definitionen in der Gruppe.
3. Recherchieren Sie das Thema (Text PANDA) und überarbeiten Sie Ihre Definition.

20

Mögliche Themen des Projektes

Wählen Sie in Ihrer Gruppe eins der folgenden Themen aus:

- Abbaubare Kunststoffe
- Kohlenstoffkreislauf und Klima
- „Virtuelles Wasser“ im Wasserkreislauf
- Wasserstofftechnologie und Brennstoffzelle(n)
- Solarzellen
- Lithium-Ionen-Akkus

→ In den jeweiligen Themen können Schwerpunkte selbst gewählt werden.

21

Literaturrecherche

Zeitschriften für den Chemieunterricht:

- Unterricht Chemie (Naturwissenschaften im Unterricht) (Papier in Bibliothek)
- Praxis der Naturwissenschaften – Chemie (Papier in Bibliothek)
- MNU – Der Mathematisch-Naturwissenschaftliche Unterricht (digital im Lernraum)

Suche über Fachportal Pädagogik (FIS Bildung)

<http://www.fachportal-paedagogik.de/start.html>

Fachzeitschriften

Lehrbücher

22

2. Sitzung



UNIVERSITÄT
PADERBORN



CHEMIEDIDAKTIK

CHEMISCHE BILDUNG UND CHEMIEDIDAKTISCHE KONZEPTIONEN

WAS SIND EVIDENZEN?

Pascal Pollmeier & Sabine Fechner



UNIVERSITÄT
PADERBORN

CHEMISCHE BILDUNG
CHEMIEDIDAKTISCHE KONZEPTIONEN

Gruppenphase

Finden Sie sich in Ihren Gruppen zusammen und diskutieren Sie:

- Definition von Nachhaltigkeit (Nutzen Sie auch den Text auf PANDA im Literatur-Ordner unter dem Thema „Einführung“ von Zowada, Niebert & Eilks)
- Eingrenzung des eigenen Themas (z.B. Welche konkreten Teilbereiche des Kohlenstoffkreislaufs sollen bearbeitet werden?)

Zeitraumen: 20 Minuten

UNIVERSITÄT PADERBORN

SP1

CHEMISCHE BILDUNG
CHEMIEDIDAKTISCHE KONZEPTIONEN

Was sind Evidenzen?

1.) Erstellen Sie eine eigene Definition von „Evidenzen“.

2.) Vergleichen Sie Ihre Definition mit der ihrer Kommiliton*innen.

3

UNIVERSITÄT PADERBORN

SP2

CHEMISCHE BILDUNG
CHEMIEDIDAKTISCHE KONZEPTIONEN

Was sind Evidenzen?

Chloroquin – Effektiv zur Behandlung einer SARS-CoV-2 Infektion?

Überlegen Sie:

- Welche Evidenzen müssten vorliegen, um die Wirksamkeit eines Medikaments zur Behandlung einer Infektion bewerten zu können?

Hören Sie sich den Ausschnitt aus dem **NDR-Podcast** mit Prof. Dr. Christian Drosten an.

→ Welche Aspekte des Studiendesigns hält er für relevant?

→ Welche Kriterien muss eine Studie erfüllen, um glaubwürdig zu sein?

NDR Info
CORONA
VIRUS
UPDATE
mit Christian Drosten

Bezugsstudie:
Gautret, P.; Lagier, J.-C.; Parola, P.; Hoang, V. T.; Meddeb, L.; Mailhe, M.; Doudier, B.; Courjon, J.; Giordanengo, V.; Vieira, V. E.; Dupont, H. T.; Honoré, S.; Colson, P.; Chabrière, E.; La Scola, B.; Rolain, J.-M.; Brouqui, P.; Raoult, D. (2020). Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *International Journal of Antimicrobial Agents*, 56(1), S. 1-6.

4

Typen von Evidenzen

Im Falle der Wirksamkeit von Chloroquin gegen das SARS-COV-2 Virus können folgende Evidenz-Typen eine Rolle spielen:

- Anekdotische Evidenz, bzw. Experten Meinungen
- Tier, oder Zellkultur Studien
- Fallberichte, oder Fallserien
- Fall-Kontroll-Studien
- Kohorten-Studien
- Randomisierte kontrollierte Studien (blind, oder doppelt-blind)
- Systematische Reviews

Aufgabe:

- 1) Informieren Sie sich **durch die Infografik** zu den Stärken und Schwächen der jeweiligen Studien. (<https://www.chemie.de/infografiken/pdf/199/199.pdf>)
- 2) Überlegen Sie, welche Evidenz-Typen in der didaktischen Forschung eine Rolle spielen könnten. Welche Stärken und Schwächen haben die jeweiligen Studienformate?

5

Typen von Evidenzen in didaktischer Forschung

Zum Beispiel:

- Fallanalysen
- Vergleichsstudien (Interventionsstudien mit Vergleichs-/Kontrollgruppen)
- Feldstudien in der Schule (Beobachtungsstudien)
- Laborstudien
- Expertenbefragungen
- Videostudien
- Systematische Reviews

Literatur zur Vertiefung:

Krüger, D.; Parchmann, I.; Schecker, H., (Hrsg.) (2014).
„Methoden in der naturwissenschaftsdidaktischen Forschung“.

<https://link.springer.com/content/pdf/10.1007%2F978-3-642-37827-0.pdf>

6

Typen von Evidenzen in didaktischer Forschung

Was bedeutet das für unsere Forschung?

Wie können andere Wissenschaftler überzeugt werden?

7

Gruppenphase

Finden Sie sich in Ihren Gruppen zusammen und planen Sie Ihr weiteres Vorgehen.

Welche Evidenzen sind für die Planung Ihres Projekts nötig?

8

3. Sitzung



CHEMIEDIDAKTIK

CHEMISCHE BILDUNG UND CHEMIEDIDAKTISCHE KONZEPTIONEN

EVIDENZEN FINDEN

Pascal Pollmeier & Sabine Fechner



CHEMISCHE BILDUNG
CHEMIEDIDAKTISCHE KONZEPTIONEN

Datenbanken & Suchstrategien

Das Konzept der Didaktischen Rekonstruktion zeigt auf, dass es in verschiedenen Bereichen wichtig ist, die Unterrichtsplanung auf Evidenzen zu stützen.

Doch...

... wie können passende Evidenzen gefunden werden?

(Die Datenbanken sind ggf. nur im Uni-Netz bzw. über VPN erreichbar.)

APA PsycNET®

FACHPORTAL
PÄDAGOGIK
Erziehungswissenschaft
Bildungsforschung
Fachdidaktik



Clarivate
Web of Science™

Datenbanken & Suchstrategien

Erstellen einer Suchmatrix

Eine Suchmatrix kann zur Strukturierung und systematischen Suche beitragen.

Aspekte	Soziale Netzwerke	Jugendliche	Auswirkung
Synonyme	Soziale Komunität	Heranwachsende	Auswirkung, Einfluss, Bedeutung, Erfolg
Oberbegriffe	Internet, Mobiles Internet	Jugend, Junger Mensch, Junge Menschen	Sozialisation, Soziale Entwicklung, Entwicklung
Unterbegriffe	WhatsApp, Instagram, Facebook, Twitter, Snapchat	Mädchen, Junge, Schüler	Entgleisung, Missbrauch, Medienverhalten
Verwandte Begriffe	Digitale Welt	Pubertierende, Pubertät, Kind, Kinder	Verhalten, Medienverhalten, Vergesellschaftung
Andere Sprachen	Social network, social community	Teenager, Teenie, Teen, Young adult, Adolescent	Effect, Impression, Value, Virulence

Beispiel-Thema: „Soziale Netzwerke und ihre Auswirkungen auf Jugendliche“

Beispiel entnommen: Universität Siegen/Universitätsbibliothek. Suchstrategien – erfolgreich recherchieren. https://www.ub.uni-siegen.de/fileadmin/user_upload/pdf/schulungen/Handout_4.pdf

6

Datenbanken & Suchstrategien

Erstellen einer Suchmatrix

Die Suche nach Synonymen und verwandten Begriffen kann sehr hilfreich sein.

- Suche breit starten (viele verschiedene Begriffe aus der Suchmatrix).
- Nach und nach eingrenzen und Begriffe bewusst nutzen, bzw. auslassen.

Aspekte	Soziale Netzwerke	Jugendliche	Auswirkung
Synonyme	Soziale Komunität	Heranwachsende	Auswirkung, Einfluss, Bedeutung, Erfolg
Oberbegriffe	Internet, Mobiles Internet	Jugend, Junger Mensch, Junge Menschen	Sozialisation, Soziale Entwicklung, Entwicklung
Unterbegriffe	WhatsApp, Instagram, Facebook, Twitter, Snapchat	Mädchen, Junge, Schüler	Entgleisung, Missbrauch, Medienverhalten
Verwandte Begriffe	Digitale Welt	Pubertierende, Pubertät, Kind, Kinder	Verhalten, Medienverhalten, Vergesellschaftung
Andere Sprachen	Social network, social community	Teenager, Teenie, Teen, Young adult, Adolescent	Effect, Impression, Value, Virulence

Beispiel entnommen: Universität Siegen/Universitätsbibliothek. Suchstrategien – erfolgreich recherchieren. https://www.ub.uni-siegen.de/fileadmin/user_upload/pdf/schulungen/Handout_4.pdf

7

Datenbanken & Suchstrategien

Verknüpfung von Suchwörtern

Um mehrere Suchwörter sinnvoll miteinander zu verbinden können **Boolean Operatoren** genutzt werden.

AND – beide Begriffe müssen zusammen vorkommen

OR – entweder der eine, oder der andere Begriff muss vorkommen

NOT – der Begriff darf nicht vorkommen (teilweise auch **AND NOT**)

NEAR/X – der Begriff muss im Abstand von X Wörtern vorkommen
(Kein wirklicher Boolean Operator, jedoch oft möglich)

Beispiel entnommen: Universität Siegen/Universitätsbibliothek. Suchstrategien – erfolgreich recherchieren. https://www.ub.uni-siegen.de/fileadmin/user_upload/pdf/schulungen/Handout_4.pdf

8

Datenbanken & Suchstrategien

Verknüpfung von Suchwörtern

Zur Eingrenzung der Suche können Klammern eingesetzt werden. Die Operatoren beziehen sich dann nur auf die Begriffe in den Klammern.

Möglicher Suchauftrag:

(Soziale Netzwerke **AND** Twitter) **NOT** WhatsApp

→ Je nach Datenbank müssen ggf. weitere Zeichen eingegeben werden, z.B. Web of Science:

TS=(Soziale Netzwerke **AND** Twitter) **NOT** TS=(WhatsApp)

(TS= Topic → es wird im Feld „Topic“ gesucht. Weitere Zeichen für andere Suchfelder)

9

Datenbanken & Suchstrategien

Wortvariationen - Trunkierungen

Teilweise sind nur Wortteile relevant, die Endung, oder der Beginn des Begriffes kann variieren. Dazu kann die Trunkierung eingesetzt werden.

Rechtstrunkierung: z.B. „bürger*“, gefunden wird: bürgerlich, bürgerliche, bürgertum, etc.

Linkstrunkierung: z.B. „*dichtungen“, gefunden wird: Traumdichtungen, Märchendichtungen, etc.

Binnentrunkierung: z.B. „Hi*brand“, gefunden wird: Hildebrand, Hillebrand, Hildbrand, etc.

Teilweise können auch einzelne Buchstaben variiert werden (→ **Maskierung**). Der fragliche Buchstabe wird dann durch ein Symbol ersetzt, z.B. „Poten?ial“. Die Funktion ist jedoch nicht bei allen Datenbanken möglich.

Beispiel entnommen: Universität Siegen/Universitätsbibliothek. Suchstrategien – erfolgreich recherchieren. https://www.ub.uni-siegen.de/fileadmin/user_upload/pdf/schulungen/Handout_4.pdf

10

Datenbanken & Suchstrategien

Wortvariationen - Thesaurus

Ein Thesaurus kann die Suche erleichtern, einige Datenbanken bieten diese Funktion an. Durch ein Thesaurus werden approximative Synonyme zu einem Begriff gesucht.

z.B. Suche nach „education“ in Datenbank ERIC → 264 gefundene Begriffe
→ Wird die Suche um Synonyme und dead terms (veraltete Begriffe) ergänzt → 513 gefundene Begriffe

z.B.: Academic education, Movement education, Educational Opportunities, Bilingual education programmes, etc.

=> Einige der gefundenen Begriffe mögen unpassend sein, andere können zur weiteren Eingrenzung der Suche führen.

Universität Siegen/Universitätsbibliothek. Suchstrategien – erfolgreich recherchieren. https://www.ub.uni-siegen.de/fileadmin/user_upload/pdf/schulungen/Handout_4.pdf

11

Datenbanken & Suchstrategien

Tipps & Tricks

- Singular- & Pluralformen suchen (z.B. durch Trunkierung)
- Unterschiedliche Schreibweisen beachten
- Neue Suchbegriffe aus ersten Treffern generieren
- Suche ggf. auch bestimmte Suchkategorien eingrenzen
- Suchfelder variieren
- Suche ggf. nach Erscheinungsdatum einschränken

Teils entnommen: Universität Siegen/Universitätsbibliothek. Suchstrategien – erfolgreich recherchieren. https://www.ub.uni-siegen.de/fileadmin/user_upload/pdf/schulungen/Handout_4.pdf

12

Wichtige Qualitätskriterien für gefundene Studien

Nicht alle gefundenen Studien weisen zwangsläufig eine hohe Qualität auf.

Was sind Aspekte, auf die bei der Literaturrecherche geachtet werden sollte?

SIGNIFIKANZ

„Ein Ergebnis ist statistisch signifikant (bedeutsam), wenn die Wahrscheinlichkeit, dass es zufällig zustande gekommen ist, klein ist. Das Signifikanzniveau muss definiert werden. Häufig gelten Aussagen, bei denen mit einem Signifikanztest eine Irrtumswahrscheinlichkeit unter 5% gefunden wird, als signifikant.“ (S. 403)

Die Irrtumswahrscheinlichkeit wird meistens mit einem p-Wert (z.B. $p < .05$) angegeben. Ist ein Ergebnis nicht signifikant, ist die Aussagekraft der Studie kritisch zu bewerten.

Krüger, D.; Parchmann, I.; Schecker, H. (Hrsg.). (2014). Methoden in der naturwissenschaftsdidaktischen Forschung. Springer-Verlag, Berlin, Heidelberg.

13

Wichtige Qualitätskriterien für gefundene Studien

Nicht alle gefundenen Studien weisen zwangsläufig eine hohe Qualität auf.

Was sind Aspekte, auf die bei der Literaturrecherche geachtet werden sollte?

BEDEUTUNG FÜR SCHULE/FORSCHUNG

„Effektstärke – Statistisches Maß für die Größe einer Wirkung oder die Relevanz eines Unterschieds in quantitativen Untersuchungen. In der empirischen fachdidaktischen Forschung interessiert nicht nur, *ob* zwei Gruppen sich hinsichtlich eines Merkmals (z.B. Fachleistung) signifikant unterscheiden, oder ob ein Lernzuwachs vorliegt, sondern auch *wie relevant* der Unterschied oder Zuwachs ist.“ (S. 399*)

Die Effektstärke wird oft mit einem d-Wert** angegeben:

d = 0,2 → kleiner Effekt

d = 0,5 → mittlerer Effekt

d = 0,8 → großer Effekt

Neben der Effektstärke muss auch die Bedeutung auch inhaltlich analysiert werden.

*Krüger, D.; Parchmann, I.; Schecker, H. (Hrsg.). (2014). Methoden in der naturwissenschaftsdidaktischen Forschung. Springer-Verlag, Berlin, Heidelberg;
**Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, N.J.: L. Erlbaum Associates

14

Wichtige Qualitätskriterien für gefundene Studien

Nicht alle gefundenen Studien weisen zwangsläufig eine hohe Qualität auf.

Was sind Aspekte, auf die bei der Literaturrecherche geachtet werden sollte?

GÜTEKRITERIEN

Gütekriterien müssen an die jeweilige Studie angepasst sein. Übergeordnet gibt es drei „große“ Kriterien:

Objektivität	Reliabilität	(interne) Validität
The extent to which personal biases are removed and value free information is gathered.	The extent to which the results are consistent if the study would be replicated.	The extent to which observed effects can be attributed to the independent variable.

Frambach, J. M.; Durning, S. J. (2013). AM Last Page: Quality criteria in qualitative and quantitative research. *Academic Medicine*, 88(4), p. 552.

15

Wichtige Qualitätskriterien für gefundene Studien

Es gibt eine Vielzahl weiterer Kriterien die eine Studie erfüllen muss, um als qualitativ hochwertig angesehen zu werden. Diese hängen teilweise aber spezifisch von den genutzten Methoden etc. ab.

Die hier vorgestellten Qualitätskriterien können **für einen ersten Eindruck** während der Recherche hilfreich sein.

16

Aufgabe

Recherchieren Sie Literatur zu dem von Ihnen gewählten Thema.



Beachten Sie dabei die Qualitätskriterien für wissenschaftliche Studien sowie die notwendige Multiperspektivität, welche aus der Didaktischen Rekonstruktion hervor geht.



17

4. Sitzung



CHEMIEDIDAKTIK

CHEMISCHE BILDUNG UND CHEMIEDIDAKTISCHE KONZEPTIONEN

EVIDENZEN ERSTELLEN

Pascal Pollmeier & Sabine Fechner



CHEMISCHE BILDUNG
CHEMIEDIDAKTISCHE KONZEPTIONEN

Inhalt der Sitzung

- Wichtige Aspekte zur Evidenzerstellung
- Erhebungsmethoden
- Erstellung Erhebungsinstrument
- Peer-Feedback

Merkmale wissenschaftlicher Studien

STICHPROBE

- **Stichprobenumfang:** Anzahl aller Objekte oder Personen in der Stichprobe
- **Randomisierung:** zufällige Zuordnung der Teilnehmer zu den Gruppen (Experiment)
- **Verblindung:** Versuchspersonen wissen nicht, ob sie Teil der Experiment- oder Kontrollgruppe sind

Experimentalgruppe:

Gruppe mit neuer Methode/Intervention

Kontrollgruppe:

Gruppe ohne neue Methode/Intervention

3

Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Auflage). Berlin, Heidelberg: Springer Berlin Heidelberg.

Merkmale wissenschaftlicher Studien

UNTERSUCHUNGSDESIGN

- **Quantitative Studie:** Untersuchung theoretisch abgeleiteter Forschungshypothesen an vielen Untersuchungseinheiten mit strukturierten Datenerhebungsmethoden
- **Qualitative Studie:** Sehr detaillierte Untersuchung offener Forschungsfragen an wenigen Untersuchungseinheiten mit unstrukturierten oder teilstrukturierten Datenerhebungsmethoden
- **Mixed-Methods-Studie:** Kombination von qualitativen und quantitativen Vorgehensweisen

4

Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Auflage). Berlin, Heidelberg: Springer Berlin Heidelberg.

Merkmale wissenschaftlicher Studien

UNTERSUCHUNGSDESIGN

- **Explorative Studien:** Genaue Erkundung und Beschreibung eines Sachverhaltes zur Entwicklung wissenschaftlicher Forschungsfragen, Hypothesen und Theorien
- **Deskriptive Studien:** Beschreibung ganzer Populationen im Hinblick auf die Verbreitung ausgewählter Merkmale
- **Explanative Studien:** Überprüfung vorher aufgestellter Hypothesen und Theorien, aus denen die Hypothesen abgeleitet wurden

5

Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Auflage). Berlin, Heidelberg: Springer Berlin Heidelberg.

Merkmale wissenschaftlicher Studien

UNTERSUCHUNGSDESIGN

- **Experimentelle Studie:** Bildung von mindestens zwei Gruppen (**Randomisierung**), experimentelle Manipulation der unabhängigen Variable/n und Messung der in den Experimental- und Kontrollgruppen resultierenden Effekte auf die abhängige/n Variable/n
- **Quasi-experimentelle Studie:** Zurückgreifen auf Gruppen, die einfach vorgefunden oder anderweitig gebildet wurden (**keine Randomisierung**), experimentelle Variation der unabhängigen Variable/n und Messung der in den Experimental- und Kontrollgruppen resultierenden Effekte auf die abhängige/n Variable/n

6

Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Auflage). Berlin, Heidelberg: Springer Berlin Heidelberg.

Merkmale wissenschaftlicher Studien

UNTERSUCHUNGSDESIGN

- **Nicht-experimentelle Studie:** Zurückgreifen auf Gruppen, die einfach vorgefunden oder anderweitig gebildet wurden (**keine Randomisierung**), keine experimentelle Variation der unabhängigen Variable/n, sondern vorgefundene Variation und Betrachtung deren vorgefundenen Unterschiede
- **Querschnittstudie:** Untersuchung einer Stichprobe zu einem Zeitpunkt
- **Trendstudie:** Studie besteht aus mehreren, in zeitlichem Abstand durchgeführten Querschnittstudien, in denen jeweils (zumindest teilweise) dieselben Variablen erhoben werden
- **Längsschnittstudie:** Wiederholte Untersuchung einer Stichprobe über längere Zeit hinweg

7

Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Auflage). Berlin, Heidelberg: Springer Berlin Heidelberg.

Merkmale wissenschaftlicher Studien

DATENERHEBUNGSMETHODEN

- **Beobachtung:** zielgerichtete, systematische und regelgeleitete Erfassung, Dokumentationen und Interpretation von Merkmalen, Ereignissen oder Verhaltensweisen mithilfe menschlicher Sinnesorgane und/ oder technischer Sensoren zum Zeitpunkt ihres Auftretens
- **Interview:** Zielgerichtete, systematische und regelgeleitete Generierung und Erfassung von verbalen Äußerungen einer Befragungsperson/-gruppe zu ausgewählten Aspekten ihres Wissens, Erlebens und Verhaltens in mündlicher Form.
- **Fragebogen:** Zielgerichtete, systematische und regelgeleitete Generierung und Erfassung von Selbstauskünften von Befragungspersonen zu ausgewählten Aspekten ihres Erlebens in schriftlicher Form.

8

Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Auflage). Berlin, Heidelberg: Springer Berlin Heidelberg.

Merkmale wissenschaftlicher Studien

DATENERHEBUNGSMETHODEN

- **Psychologischer Test:** Wissenschaftliches Datenerhebungsverfahren, das aus mehreren Testaufgaben (**Testbogen/ Testmaterial**) sowie festgelegten Regeln zu deren Anwendung und Auswertung (**Testmanual**) besteht zur Erfassung eines latenten psychologischen Merkmals in seiner absoluten oder relativen Ausprägung.
- **Quantitative Dokumentenanalyse bzw. Inhaltsanalyse:**

Zielgerichtete, systematische und regelgeleitete Sammlung und Archivierung von vorhandenen Dokumenten als Manifestationen menschlichen Erlebens und Verhaltens und anschließende Auswertung der Dokumente hinsichtlich ihrer inhaltlichen und formalen Merkmale.

9

Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Auflage). Berlin, Heidelberg: Springer Berlin Heidelberg.

Aufgabe

Erstellen Sie eine erste Version eines Erhebungsinstruments für ihre geplante Studie.

Peer-Feedback

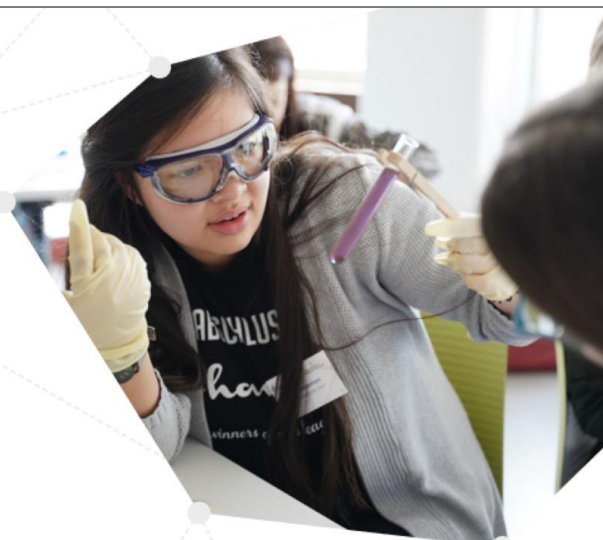
Laden Sie Ihr Instrument auf PANDA hoch. Kommentieren Sie anschließend bis zur angegebenen Deadline das Instrument einer anderen Gruppe.

Hilfen:

- Literatur auf PANDA
- Können bestehende Instrumente genutzt werden?
- Orientierung an bestehenden Instrumenten

10

5. Sitzung



CHEMIEDIDAKTIK

CHEMISCHE BILDUNG UND CHEMIEDIDAKTISCHE KONZEPTIONEN

VERMARKTUNG VON EVIDENZEN

Pascal Pollmeier & Sabine Fechner



CHEMISCHE BILDUNG
CHEMIEDIDAKTISCHE KONZEPTIONEN

Organisation

- Erhebungsinstrumente
- Ablauf des Projekts und mögliche Abgaben
- Dokumentationsbogen
- Ablauf Projektarbeit Dezember/Januar
- Beratungsmöglichkeiten

Vermarktung von Studienergebnissen

Die Art und Weise, wie in Studien über die Erkenntnisse berichtet wird, ist immer auch abhängig von der Zielgruppe. Die zwei vorliegenden Studien haben verschiedene Zielgruppen.

Aufgaben:

- Lesen Sie die gekürzten Studien zum Thema „Nachhaltigkeit“ auf PANDA.
- Analysieren Sie, um welche Zielgruppen es sich jeweils handeln könnte.
- Welche Gestaltungsmittel werden beim Bericht der Erkenntnisse zielgruppenspezifisch eingesetzt?

3

Grafiken

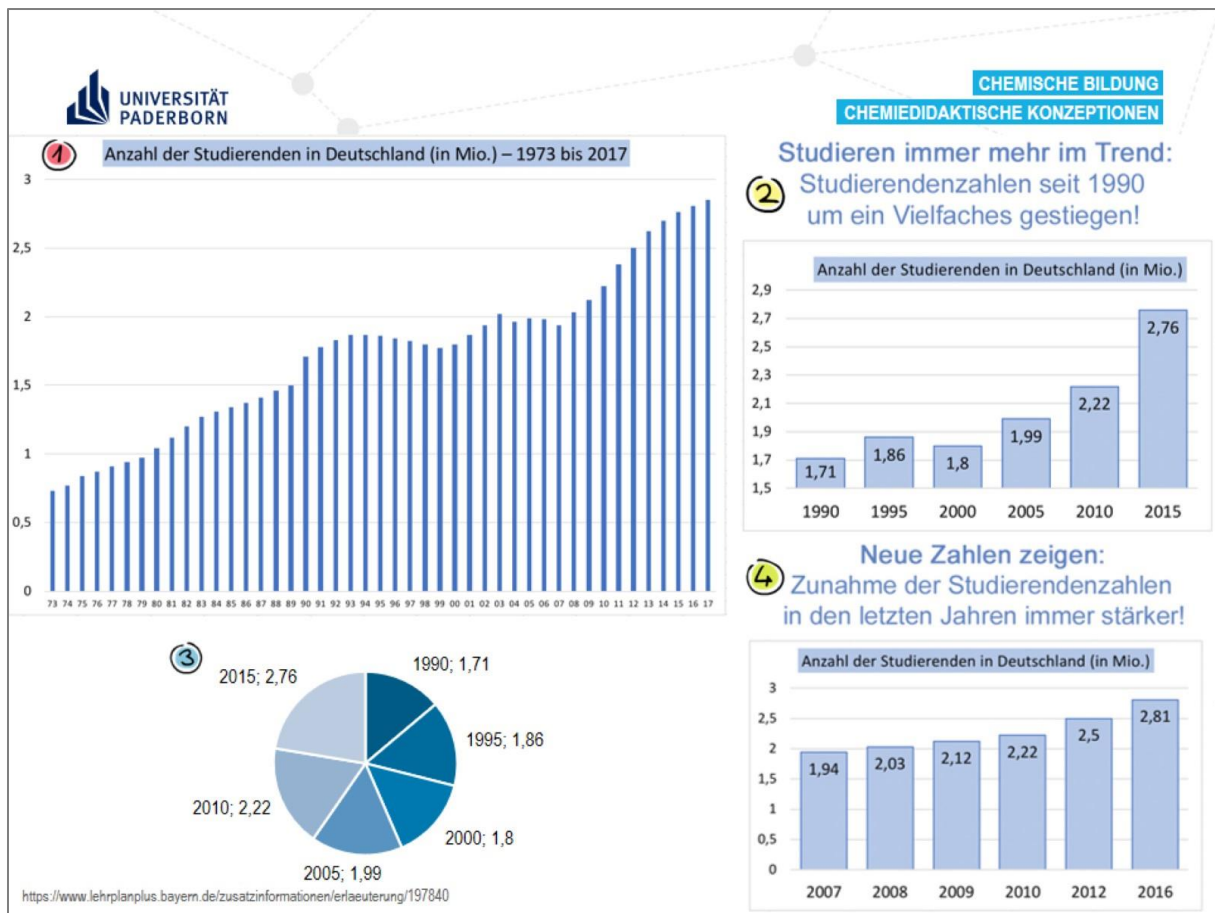
Viele Studien enthalten Grafiken, welche Informationen übersichtlich darstellen sollen. Grafiken beinhalten häufig viele Informationen zeitgleich und können so schnell zu einem Überblick verhelfen.

Ähnlich wie beim Bericht von Erkenntnissen, können Grafiken auch an die Zielgruppe angepasst werden, bzw. von unterschiedlicher Qualität sein.

Aufgaben

- Vergleichen Sie die Grafiken. Welche Grafiken sind qualitativ hochwertiger, welche haben eine geringe Qualität und warum?
- Formulieren Sie Merkmale für qualitativ hochwertige Grafiken.
- Ergänzen Sie die Merkmale mit der Literatur (siehe PANDA).
- Laden Sie Ihre erstellte Checkliste für das Grafikdesign auf PANDA hoch.

4



12. Sitzung



CHEMIEDIDAKTIK

CHEMISCHE BILDUNG UND CHEMIEDIDAKTISCHE KONZEPTIONEN

WIE GUT SIND MEINE MATERIALIEN?

Pascal Pollmeier & Sabine Fechner

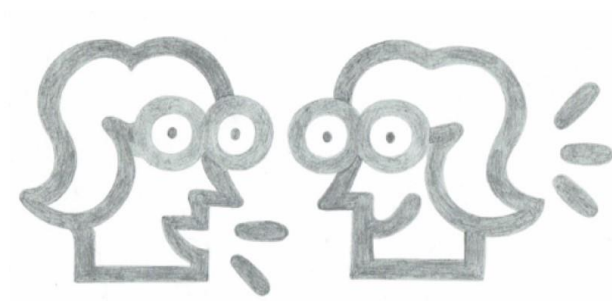


AC3

CHEMISCHE BILDUNG
CHEMIEDIDAKTISCHE KONZEPTIONEN

Feedback in PANDA

Das Feedback zu Ihren Materialien kann jetzt auf PANDA eingesehen werden.



Tauschen Sie sich in Ihrer Projektgruppe über das erhaltene Feedback aus.

- Welche Fragen bleiben offen?
- An welchen Punkten konzentrieren sich die Hinweise?
- Wozu wünschen Sie sich weiteres Feedback?

Bearbeitungszeit:



20min

Speed-Dating

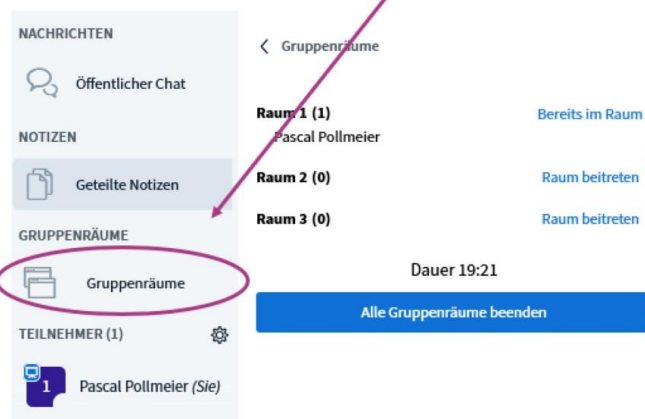
Ablauf

- Sie werden mit Mitgliedern aus anderen Projektgruppen zusammen kommen
- Diskutieren Sie in jeder Runde Ihre Materialien und sammeln sie Verbesserungsvorschläge.
- Nutzen Sie auch die „Bildschirm-Teilen“-Funktion, um Ihre Materialien zu besprechen.
- Als Diskussionsanregung können Ihnen die Leitfragen dienen.
- Pro Runde haben Sie 20min Zeit, davon sind je 10min pro Modul vorgesehen
- Wechseln Sie nach 10 Minuten eigenständig zum jeweils anderen Modul
- Nach 20min wird der Raum gewechselt

3

Speed-Dating

Um den Gruppenraum zu wechseln, verlassen Sie den Gruppenraum und kehren zurück in den Hauptraum. Dort wählen Sie „**Gruppenräume**“ und treten dem entsprechenden Raum bei.



The screenshot shows a mobile application interface with a sidebar menu on the left and a main content area on the right. The sidebar menu includes sections for 'NACHRICHTEN' (with 'Öffentlicher Chat'), 'NOTIZEN' (with 'Geteilte Notizen'), 'GRUPPENRÄUME' (with 'Gruppenräume' circled in red), and 'TEILNEHMER (1)' (with 'Pascal Pollmeier (Sie)'). The main content area shows a list of rooms: 'Raum 1 (1) Pascal Pollmeier', 'Raum 2 (0)', and 'Raum 3 (0)'. To the right of this list are buttons for 'Bereits im Raum', 'Raum beitreten', and 'Raum beitreten'. A blue button at the bottom says 'Alle Gruppenräume beenden'. A status bar at the top indicates 'Dauer 19:21'.

4

UNIVERSITÄT PADERBORN

CHEMISCHE BILDUNG
CHEMIEDIDAKTISCHE KONZEPTIONEN

Speed-Dating

Bearbeitungszeit: 10min Modul A
10min Modul B

Raumaufteilung

	Raum 1 Anna, Eike	Raum 2 Jana	Raum 3 Christiane
Runde #1	Janos	Julia, Lorena	Malin, Matthias
Runde #2	Malin, Matthias	Janos	Julia, Lorena

5

UNIVERSITÄT PADERBORN

CHEMISCHE BILDUNG
CHEMIEDIDAKTISCHE KONZEPTIONEN

Weiterer Verlauf

- Bitte teilen Sie uns Ihren genauen Erprobungszeitpunkt für Ihre Module frühzeitig mit.
- Je nach Format: Geben Sie uns Zugangsmöglichkeiten zum Videokonferenzraum
- Nehmen Sie die Hinweise zur Durchführung von Online-Prüfungen (PANDA) zur Kenntnis.

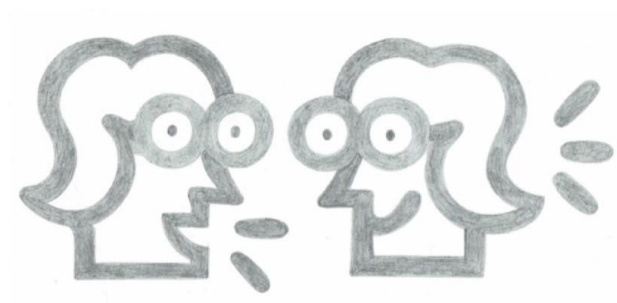
Finale Abgabe aller Materialien: 11.02.2021, 23:59 Uhr

(Letzte synchrone Sitzung: 11.02.2021, 09:15 Uhr)

6

Projektgruppen-Phase

Kehren Sie in Ihre Projektgruppen zurück. Besprechen Sie weiteren Bearbeitungsbedarf und strukturieren Ihr weiteres Vorgehen.



Bei Fragen, oder Diskussionsbedarf, melden Sie sich gerne.

Bearbeitungszeit:



verbleibende Zeit

14. Sitzung



CHEMIEDIDAKTIK

CHEMISCHE BILDUNG UND CHEMIEDIDAKTISCHE KONZEPTIONEN

WARUM SIND EVIDENZEN WICHTIG? WIE KANN EIN KONSENS ERZIELT WERDEN?

Pascal Pollmeier & Sabine Fechner

AG1

CHEMISCHE BILDUNG
CHEMIEDIDAKTISCHE KONZEPTIONEN

Rückblick

An welchen Stellen Ihrer Planung spielen Evidenzen eine Rolle?

Betrachten Sie Ihre finale Planung und analysieren, an welchen Stellen Evidenzen genutzt wurden.

- Wie/Wo/Welche Evidenzen haben Sie gefunden?
- Wie sind sie mit den Evidenzen umgegangen?
- Wie haben Sie zwischen widersprüchlichen Evidenzen vermittelt?
- Wie haben Sie sich auf Quellen und Evidenzen festgelegt?

Rückblick Cholorquin-Studie

Chloroquin – Effektiv zur Behandlung einer SARS-CoV-2 Infektion?

- Testen des Malariamedikaments Chloroquin zur Behandlung von Patienten, die an SARS-CoV-2 erkrankt sind
- Untersuchung von leichten, schwierigen und asymptomatischen Fällen
- Die Experimentalgruppe wurde mit Chloroquin behandelt, die Kontrollgruppe wurde nicht mit Chloroquin behandelt
- Gemessen wurden die Viruskonzentration und die Virusnachweisrate im Hals



Bezugsstudie:

Gautret, P.; Lagier, J.-C.; Parola, P.; Hoang, V. T.; Meddeb, L.; Mailhe, M.; Doudier, B.; Courjon, J.; Giordanengo, V.; Vieira, V. E.; Dupont, H. T.; Honoré, S.; Colson, P.; Chabrière, E.; La Scola, B.; Rolain, J.-M.; Brouqui, P.; Raoult, D. (2020). Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *International Journal of Antimicrobial Agents*, 56(1), S. 1-6.

3

Rückblick Cholorquin-Studie

Chloroquin – Effektiv zur Behandlung einer SARS-CoV-2 Infektion?

Probleme der Studie:

- Keine randomisierte Studie
- Keine Doppelblindstudie
- Unterschiedliche Gruppen
 - Behandelte Patienten im Schnitt 51 Jahre und 2 asymptomatische Patienten
 - Nicht behandelte Patienten im Schnitt 37 Jahre und 4 asymptomatische Patienten
 → Grundgegebenheiten der Patientenrekrutierung komplett unterschiedlich
- Zeitskala: Behandelte Gruppe ist in Wirklichkeit schon weiter fortgeschritten im Verlauf
- Messung der Viruskonzentration und der Virusnachweisrate im Hals und nicht in der Lunge → Krankheit verschwindet schneller



Bezugsstudie:

Gautret, P.; Lagier, J.-C.; Parola, P.; Hoang, V. T.; Meddeb, L.; Mailhe, M.; Doudier, B.; Courjon, J.; Giordanengo, V.; Vieira, V. E.; Dupont, H. T.; Honoré, S.; Colson, P.; Chabrière, E.; La Scola, B.; Rolain, J.-M.; Brouqui, P.; Raoult, D. (2020). Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *International Journal of Antimicrobial Agents*, 56(1), S. 1-6.

4

Konsensbildung in der Wissenschaft

Wie können Schüler*innen durch naturwissenschaftlichen Unterricht darauf vorbereitet werden, gesellschaftliche Diskussionen bzgl. wissenschaftlicher Themen zu führen?

Präzise:

Wie können wir Schüler*innen vorbereiten, um die Aussagekraft der Chloroquin-Studie zu bewerten?

5

Konsensbildung in der Wissenschaft

Wie kann, bzw. sollte überhaupt Konsens in der Wissenschaft gebildet werden?

„[...] funktioniert Wissenschaft natürlich nicht als Demokratie, wo Mehrheiten entscheiden, was richtig ist. Darum hatte Einstein schon recht mit der Relativitätstheorie, bevor er sie überhaupt veröffentlicht hatte. Darum ist es gelegentlich vorgekommen, dass Außenseiter ein etabliertes Lehrgebäude einstürzen lassen, zum Beispiel Alfred Wegener mit seiner These von der Kontinentalverschiebung. Und darum betrachten redliche Wissenschaftler ihre Schlussfolgerungen stets als vorläufig, weil sie in Zukunft durch andere Resultate widerlegt oder zumindest ergänzt und als unvollständig erkannt werden können.“

- Christoph Schrader (<https://www.klimafakten.de/meldung/kungelei-oder-klare-beweislage-was-ist-das-eigentlich-konsens>, Zugriff: 09.10.2020)

6

Konsensbildung in der Wissenschaft

Sehen Sie sich den Ausschnitt:

0:53 min – 6:02 min

des Videos „Corona hat meine Meinung geändert“ von Dr. Mai Thi Nguyen-Kim an.

(<https://www.youtube.com/watch?v=Nn2rJrKwENI>)



- Welche neuen Aspekte zur wissenschaftlichen Konsensbildung werden angesprochen?

ii. Evidence-informed decision making in education test (EIDM-ET) – Langfassung

Der EIDM-ET wurde mittels der Umfragesoftware LimeSurvey digital umgesetzt. Das Seitenlayout des Instruments im Anhang stimmt dementsprechend nicht mit der digitalen Ansicht auf dem jeweiligen Endgerät überein, zeigt jedoch die eingesetzten Items.

Evidence-Informed Decision Making in Education Test (EIDM-ET)

Vielen Dank, dass Sie an dieser Studie teilnehmen. Auf dieser Seite werden allgemeine Angaben zu ihrer Person gesammelt und ein Code generiert. Der Code dient dazu ihre Fragebögen der verschiedenen Befragungszeitpunkte zu verbinden. Alle Informationen werden anonym gesammelt, die Verarbeitung ihrer Daten erfolgt nach den Grundsätzen des DSGVO-Formulars, welches Sie unterschrieben haben.

Code-Generierung:

Bitte erstellen Sie einen Code, welcher uns dabei hilft ihre Daten der verschiedenen Messzeitpunkte zusammen zu führen.

Erster und letzter Buchstabe des Vornamens ihrer Mutter. (e.g. ANNE → AE)		Erster und letzter Buchstabe des Vornamens ihres Vaters. (e.g. PETER → PE)		Erster und letzter Buchstabe ihres eigenen Vornamens. (e.g. JOHN → JN)		Tag des Geburtsdatums ihrer Mutter. (e.g. 02.07.64 → 02)	

Generelle Informationen

Geschlecht: ☐ männlich ☐ weiblich ☐ divers

Alter: _____

Studiengang: _____ (z.B. Master of. Ed. Lehramt Gy/Ge Chemie)

Land: ☐ Niederlande ☐ Vereintes Königreich ☐ Polen ☐ Deutschland

Haben Sie schon einmal einen Workshop zu evidenzbasierter Entscheidungsfindung besucht?

☐ ja ☐ nein

Aufgabe 1

Formulieren Sie eine Forschungsfrage in Bezug auf die folgenden zwei Situationen. Die Frage sollte Ihnen dabei helfen Literatur zur Thematik zu finden und gute Artikel auszusuchen.

Situation 1:

Aufgrund einer globalen Pandemie wurden Sie und Ihre Schüler informiert, alle Unterrichtsaktivitäten auf Online umzustellen. Sie haben Ihren ersten Online-Unterricht begonnen, aber es war sehr schwer herauszufinden, ob die Schüler folgen oder nicht, da Sie sie nicht sehen können. Sie möchten nach Möglichkeiten suchen, Ihre Schüler in den Unterricht einzubeziehen

Forschungsfrage für Situation 1:Situation 2:

Eine Klasse der 7. Jahrgangsstufe zeigt sich in deinem Chemieunterricht wenig motiviert. Ihr bearbeitet gerade Stoffeigenschaften mit einem Schwerpunkt auf theoretische Annahmen. Ein Kollege empfiehlt dir mehr Experimente im Unterricht einzusetzen. Du hast aber auch gerade gelesen, dass die Implementation von Alltagskontexten (wie „Kosmetik“) besonders die Motivation von Mädchen fördern kann.

Forschungsfrage für Situation 2:

Aufgabe 2

Wo können Sie eine mögliche Antwort auf Ihre Frage finden? Nennen Sie 4 Typen von Quellen und geben Sie jeweils einen Vor- und Nachteil an.

Quellentyp	Vorteil	Nachteil
1:		
2:		
3:		
4:		

Aufgabe 3

Stellen Sie sich vor, Sie suchen nach Forschungsergebnissen zu einer Ihrer Fragen (aus Aufgabe 1) in einer internationalen Datenbank. Welche wären ihre besten 3 Suchbegriffe?

Ausgewählte Frage: Situation 1 / Situation 2 / Beispielfrage (Bitte unterstreichen Sie die ausgewählte Frage!)

Beispielfrage (Verwenden Sie diese Frage nur, wenn Sie in Aufgabe 1 keine Frage formuliert haben):

Wie können Nicht-Muttersprachler in einem MINT-Kurs zum Thema Nachhaltigkeit unterstützt werden?

Suchbegriff 1	Suchbegriff 2	Suchbegriff 3

Aufgabe 4

Denken Sie über Wege nach, um Ihre Suche einzugrenzen, bzw. auszuweiten. Geben Sie 2 Beispiele, wie Sie ihre Suche in einer Datenbank eingrenzen/ausweiten können.

Eingrenzen der Suche	Ausweiten der Suche
1:	1:
2:	2:

Die folgenden Aufgaben fokussieren auf Studiendesigns und sind daher unabhängig von Ihren Fragen zu Beginn des Fragebogens.

Aufgabe 5

Studien müssen angemessen gestaltet werden, um passende Daten zu erheben. Welches Studiendesign eignet sich am besten für die folgenden Themen?

Themen	Geeignetes Studiendesign
Effektivität einer didaktischen Methode	
Datensammlung bzgl. Präkonzepten der Lernenden	
Leistung von Lernenden in verschiedenen Ländern	

Aufgabe 6

Was versteht man unter...? (nur eine Antwort möglich)

... Randomisierung?		Ihre Antwort
A	Randomisierung ist ein Verfahren, bei dem die Versuchspersonen unter Verwendung eines Zufallsmechanismus' unterschiedlichen Gruppen zugeordnet werden.	
B	Randomisierung ist ein Verfahren, um kleine Stichproben anhand spezifischer Kriterien zu parallelisieren.	
C	Randomisierung ist ein Verfahren, bei dem einem Untersucher / einer Untersucherin unter Verwendung eines Zufallsmechanismus' die Untersuchungsverfahren zugeteilt werden.	
D	Randomisierung ist ein Verfahren, um auszuschließen, dass die Ergebnisse einer Studie zufällig zustande kommen.	

... Reliabilität?		Ihre Antwort
A	Reliabilität bezeichnet die Zuverlässigkeit der Messergebnisse und die Genauigkeit der Messung.	
B	Reliabilität ist ein Prozess, um Konstrukte messbar zu machen. Den Konstrukten werden beobachtbare Kriterien zugewiesen.	
C	Reliabilität gibt an, inwieweit ein Test- oder Bewertungsverfahren ein interessantes Merkmal so misst, dass es vorhandenen Konstruktdefinitionen und -theorien entspricht.	
D	Reliabilität bezeichnet die intersubjektive Überprüfbarkeit der Messergebnisse. Es macht Messergebnisse durch Standardisierung und Transparenz nachvollziehbar und reproduzierbar.	

... Kontrollgruppe?		Ihre Antwort
A	Eine Kontrollgruppe ist eine Gruppe von Untersucher*innen, die die Validität der Untersuchungsergebnisse überprüft.	
B	Eine Kontrollgruppe ist eine Gruppe von Versuchspersonen, mit der die Ergebnisse einer Studie repliziert werden.	
C	Eine Kontrollgruppe erhält nicht die experimentelle Intervention, ist aber ansonsten mit der Experimentalgruppe vergleichbar.	
D	Eine Kontrollgruppe erhält die experimentelle Intervention und ist auch ansonsten mit der Experimentalgruppe vergleichbar.	

Aufgabe 7

Welche Charakteristika einer Studie sollten in einem Forschungsbericht dargelegt werden, damit Sie die Validität der Befunde akzeptieren? Geben Sie 5 Beispiele.

Beispiel 1:

Beispiel 2:

Beispiel 3:

Beispiel 4:

Beispiel 5:

Aufgabe 8

Nachdem Sie nun über Aspekte nachgedacht haben, welche zur Steigerung der Validität einer Studie beitragen, geht es nun um Signifikanz und Bedeutung von Studien. Geben Sie je zwei Charakteristika von Befunden an, damit sie als signifikant und bedeutungsvoll bezeichnet werden können.

Signifikanz	Bedeutung

Aufgabe 9

Nun haben Sie sich über Validität, Signifikanz und Bedeutung von Studien Gedanken gemacht. In dieser Aufgabe sollen sie 4 Aspekte, inkl. kurzer Begründung angeben, mit denen Sie die Umsetzungsmöglichkeiten der Befunden im schulischen Alltag bewerten würden.

Aspekt	Begründung
1:	
2:	
3:	
4:	

iii. Evidence-informed decision making in education test (EIDM-ET) – Kurzfassung

Die Kurzfassung des EIDM-ET beginnt mit einer Einleitung in die Thematik. Ebenso wird eine Einleitung für die *EBP scales* angeschlossen. Die *EBP scales* befinden sich im Hauptteil dieser Arbeit.

Evidence-Informed Decision Making in Education Test (EIDM-ET)- Kurzfassung

Einleitung:

Momentan spielt der Umgang mit Fakten vs. Meinungen auch im schulischen Umfeld eine große Rolle. Schülerinnen und Schüler müssen auch im Unterricht lernen, welche Informationen verlässlich sind und was im Bereich der Meinungsbildung anzusiedeln ist.

Im folgenden Teil des Fragebogens wollen wir daher Ihre Einstellung gegenüber Evidenzen (die besten verfügbaren, systematisch durch mehrere Studien gesammelten Informationen) erfragen. Auch möchten wir gern wissen, inwiefern Sie mit evidenz-basierter Praxis in Kontakt waren bzw. sie eingesetzt haben.

Aufgabe 1

Formulieren Sie eine Forschungsfrage in Bezug auf die folgende Situation. Die Frage sollte Ihnen dabei helfen, Literatur zur Thematik zu finden und gute Artikel auszusuchen.

Situationsbeschreibung:

Aufgrund einer globalen Pandemie wurden Sie und Ihre Schüler*innen informiert, alle Unterrichtsaktivitäten auf Online umzustellen. Sie haben Ihren ersten Online-Unterricht begonnen, aber es war sehr schwer herauszufinden, ob die Schüler*innen folgen oder nicht, da Sie sie nicht sehen können. Sie möchten nach Möglichkeiten suchen, Ihre Schüler*innen in den Unterricht einzubeziehen.

Forschungsfrage:

Aufgabe 2

Wo könnten Sie eine mögliche Antwort auf Ihre Frage finden? Nennen Sie 4 Typen von Quellen und geben Sie jeweils Vor- und Nachteile an.

Quellentyp	Vorteil	Nachteil
1:		
2:		
3:		
4:		

Aufgabe 3

Stellen Sie sich vor, Sie suchen nach Forschungsergebnissen zu Ihrer Frage (aus Aufgabe 1) in einer internationalen Datenbank. Welche wären Ihre besten 3 Suchbegriffe?

Beispielfrage (Verwenden Sie diese Frage nur, wenn Sie in Aufgabe 1 keine Frage formuliert haben):

Wie können Nicht-Muttersprachler in einem MINT-Kurs zum Thema Nachhaltigkeit unterstützt werden?

Suchbegriff 1	Suchbegriff 2	Suchbegriff 3

Aufgabe 4

Was versteht man unter...? (nur eine Antwort möglich)

➔ Single-Choice-Aufgabe

... Randomisierung?	Ihre Antwort
A Randomisierung ist ein Verfahren, bei dem die Versuchspersonen unter Verwendung eines Zufallsmechanismus unterschiedlichen Gruppen zugeordnet werden.	
B Randomisierung ist ein Verfahren, um kleine Stichproben anhand spezifischer Kriterien zu parallelisieren.	
C Randomisierung ist ein Verfahren, bei dem einer*m Untersucher*in unter Verwendung eines Zufallsmechanismus die Untersuchungsverfahren zugeteilt werden.	
D Randomisierung ist ein Verfahren, um auszuschließen, dass die Ergebnisse einer Studie zufällig zustande kommen.	

... Reliabilität?		Ihre Antwort
A	Reliabilität bezeichnet die Zuverlässigkeit der Messergebnisse und die Genauigkeit der Messung.	
B	Reliabilität ist ein Prozess, um Konstrukte messbar zu machen. Den Konstrukten werden beobachtbare Kriterien zugewiesen.	
C	Reliabilität gibt an, inwieweit ein Test- oder Bewertungsverfahren ein interessantes Merkmal so misst, dass es vorhandenen Konstruktdefinitionen und -theorien entspricht.	
D	Reliabilität bezeichnet die intersubjektive Überprüfbarkeit der Messergebnisse. Es macht Messergebnisse durch Standardisierung und Transparenz nachvollziehbar und reproduzierbar.	

... Kontrollgruppe?		Ihre Antwort
A	Eine Kontrollgruppe ist eine Gruppe von Untersucher*innen, die die Validität der Untersuchungsergebnisse überprüft.	
B	Eine Kontrollgruppe ist eine Gruppe von Versuchspersonen, mit der die Ergebnisse einer Studie repliziert werden.	
C	Eine Kontrollgruppe erhält nicht die experimentelle Intervention, ist aber ansonsten mit der Experimentalgruppe vergleichbar.	
D	Eine Kontrollgruppe erhält die experimentelle Intervention und ist auch ansonsten mit der Experimentalgruppe vergleichbar.	

Aufgabe 5

Welche Charakteristika einer Studie sollten in einem Forschungsbericht dargelegt werden, damit Sie die Validität der Befunde akzeptieren? Geben Sie 5 Beispiele.

Beispiel 1:
Beispiel 2:
Beispiel 3:
Beispiel 4:
Beispiel 5:

Kurzeinleitung Skalen:

Mit den folgenden zwei Aufgaben möchten wir Informationen über Ihren praktischen Umgang sowie Ihre Überzeugungen zu evidenz-basierter Praxis (EBP) erfragen. Evidenz-basierte Praxis meint das Inkludieren eigener Lehr-Lern-Erfahrungen und der besten verfügbaren Informationen aus systematischer Forschung in unterrichtliche (Entscheidungs-) Prozesse. So sollten z.B. nur hochwertige Evidenzen (z.B. hohe Validität, gute Passung zur individuellen Situation, etc.) zur Unterrichtsplanung verwendet, oder Meinungen und Evidenzen auch im Unterricht klar unterschieden werden. Bitte kreuzen Sie die jeweils passende Antwort an.

iv. Grading rubric des Evidence-informed decision making in education test (EIDM-ET)

Der hier dargestellt grading rubric wurde sowohl für die Langfassung, als auch die Kurzfassung des EIDM-ET eingesetzt.

Evidence-Informed Decision Making in Education Test (EIDM-ET)- grading rubric

Adaption nach:

- Fresno Test (Ramos, Kathleen D.; Schafer, Sean; Tracz, Susan M. (2003) Validation of the Fresno test of competence in evidence based medicine. *BMJ* 326:319.)
- TÜBEP-ST (1.0) (Cholewa, Jürgen; Corsten, Sabine; Daniel, Susanne; Grewe, Tanja; Günther, Thomas; Lauer, Norina; Mantey, Stefanie; Nobis-Bosch, Ruth (2015). Überprüfung von EBP-Kompetenzen bei Studierenden. *Forum Logopädie* 6 (29), pp. 20-27.)

Item	Aufgabe	Bewertung
Item 1	<p>Formulieren Sie eine Forschungsfrage in Bezug auf die folgenden zwei Situationen. Die Frage sollte Ihnen dabei helfen Literatur zur Thematik zu finden und gute Artikel auszusuchen.</p> <p><u>Situation 1</u> <i>Aufgrund einer globalen Pandemie wurden Sie und Ihre Schüler informiert, alle Unterrichtsaktivitäten auf Online umzustellen. Sie haben Ihren ersten Online-Unterricht begonnen, aber es war sehr schwer herauszufinden, ob die Schüler folgen oder nicht, da Sie sie nicht sehen können. Sie möchten nach Möglichkeiten suchen, Ihre Schüler in den Unterricht einzubeziehen.</i></p> <p><u>Situation 2</u> <i>Eine Klasse der 7. Jahrgangsstufe zeigt sich in deinem Chemieunterricht wenig motiviert. Ihr bearbeitet gerade Stoffeigenschaften mit einem Schwerpunkt auf theoretische Annahmen. Ein Kollege empfiehlt dir mehr Experimente im Unterricht einzusetzen. Du hast aber auch gerade gelesen, dass die Implementation von Alltagskontexten (wie „Kosmetik“) besonders die Motivation von Mädchen fördern kann.</i></p>	<p><u>Situation 1:</u> <i>Exzellent (12 Punkte):</i> Mehrere relevante Beschreibungen (z.B. „Strategien zum Schülerengagement für den Online-Unterricht“; „Engagement der Schüler in Online-Lernumgebungen“; „Verbesserung des Engagements der Schüler im Online-Unterricht“; ...)</p> <p><i>Stark (8 Punkte):</i> Eine passende Beschreibung (z.B. „Strategien zum Schülerengagement“; „Verbesserung des Online-Unterrichts“; „Schülerengagement online“; ...)</p> <p><i>Eingeschränkt (4 Punkt):</i> eine allgemeine Beschreibung (z.B. „Schülerengagement“; „Online-Unterricht“)</p> <p><i>Nicht vorhanden (0 Punkte):</i> Keins der obigen</p> <p><u>Situation 2:</u> <i>Exzellent (12 Punkte):</i> Mehrere relevante Beschreibungen (e.g. „Motivation Chemie Geschlecht“; „Motivation Alltagskontexte Mädchen“; „Motivation Experimente Chemie“ ...)</p> <p><i>Stark (8 Punkte):</i> Eine passende Beschreibung (z.B. „Motivation Chemie“; „Alltagskontexte“; „chemische Experimente“; ...)</p> <p><i>Eingeschränkt (4 Punkt):</i> eine allgemeine Beschreibung (z.B. „Motivation“; „Chemieunterricht“; „Experimente“; ...)</p>

		<i>Nicht vorhanden (0 Punkte):</i> Keins der obigen
Item 2	Wo können Sie eine mögliche Antwort auf Ihre Frage finden? Nennen Sie 4 Typen von Quellen und geben Sie jeweils einen Vor- und Nachteil an.	<p><i>Exzellent (24 Punkte):</i> 4 Quelltypen + je ein Vor-/Nachteil</p> <p><i>Stark (16 Punkte):</i> 3 Quelltypen + je ein Vor-/Nachteil</p> <p><i>Eingeschränkt (8 Punkte):</i> 2 Quelltypen + je ein Vor-/Nachteil</p> <p><i>Nicht vorhanden (0 Punkte):</i> Keins der obigen</p>
Item 3	Stellen Sie sich vor, Sie suchen nach Forschungsergebnissen <u>zu einer Ihrer Fragen</u> (aus Aufgabe 1) in einer internationalen Datenbank. Welche wären ihre besten 3 Suchbegriffe?	<p><i>Exzellent (24 Punkte):</i> 3 der Beschreibungen aus Item 1/2</p> <p><i>Stark (18 Punkte):</i> 2 der Beschreibungen aus Item 1/2</p> <p><i>Eingeschränkt (9 Punkte):</i> 1 der Beschreibungen aus Item 1/2</p> <p><i>Nicht vorhanden (0 Punkte):</i> Keins der obigen</p>
Item 4	Denken Sie über Wege nach, um Ihre Suche einzugrenzen, bzw. auszuweiten. Geben Sie 2 Beispiele, wie Sie ihre Suche in einer Datenbank eingrenzen/ausweiten können.	<p><i>Exzellent (24 Punkte):</i> 3 (z.B. Boolean Operator; Ein-/Ausschließen von Suchwörtern, allgemeinere/explicitere Suchwörter, Suchfelder, etc.)</p> <p><i>Eingeschränkt (12 Punkte):</i> 1 (z.B. Boolean Operator; Ein-/Ausschließen von Suchwörtern, allgemeinere/explicitere Suchwörter, Suchfelder, etc.)</p> <p><i>Nicht vorhanden (0 Punkte):</i> Keins der obigen</p>
Item 5	<p>Die folgenden Aufgaben fokussieren auf Studiendesigns und sind daher unabhängig von Ihren Fragen zu Beginn des Fragebogens.</p> <p>Studien müssen angemessen gestaltet werden, um passende Daten zu erheben. Welches Studiendesign eignet sich am besten für die folgenden Themen?</p> <ul style="list-style-type: none"> - Effektivität einer didaktischen Methode - Datensammlung bzgl. Präkonzepten der Lernenden - Leistung von Lernenden in verschiedenen Ländern 	<p>Je richtige Antwort: 8 Punkte</p> <ul style="list-style-type: none"> - Prä-post Kontrollgruppen Design - Interview Studie - Querschnittsstudie

Item 6	Was versteht man unter ...	Je richtige Antwort: 8 Punkte
	... Randomisierung	<p>(A) Randomisierung ist ein Verfahren, bei dem die Versuchspersonen unter Verwendung eines Zufallsmechanismus' unterschiedlichen Gruppen zugeordnet werden.</p> <p>(B) Randomisierung ist ein Verfahren, um kleine Stichproben anhand spezifischer Kriterien zu parallelisieren.</p> <p>(C) Randomisierung ist ein Verfahren, bei dem einem Untersucher / einer Untersucherin unter Verwendung eines Zufallsmechanismus' die Untersuchungsverfahren zugeteilt werden.</p> <p>(D) Randomisierung ist ein Verfahren, um auszuschließen, dass die Ergebnisse einer Studie zufällig zustande kommen.</p>
	... Reliabilität	<p>(A) Reliabilität bezeichnet die Zuverlässigkeit der Messergebnisse und die Genauigkeit der Messung.</p> <p>(B) Reliabilität ist ein Prozess, um Konstrukte messbar zu machen. Den Konstrukten werden beobachtbare Kriterien zugewiesen.</p> <p>(C) Reliabilität gibt an, inwieweit ein Test- oder Bewertungsverfahren ein interessantes Merkmal so misst, dass es vorhandenen Konstruktdefinitionen und -theorien entspricht.</p> <p>(D) Reliabilität bezeichnet die intersubjektive Überprüfbarkeit der Messergebnisse. Es macht Messergebnisse durch Standardisierung und Transparenz nachvollziehbar und reproduzierbar.</p>
	... Kontrollgruppe	<p>(A) Eine Kontrollgruppe ist eine Gruppe von Untersucher*innen, die die Validität der Untersuchungsergebnisse überprüft.</p> <p>(B) Eine Kontrollgruppe ist eine Gruppe von Versuchspersonen,</p>

		<p>mit der die Ergebnisse einer Studie repliziert werden.</p> <p>(C) Eine Kontrollgruppe erhält nicht die experimentelle Intervention, ist aber ansonsten mit der Experimentalgruppe vergleichbar.</p> <p>(D) Eine Kontrollgruppe erhält die experimentelle Intervention und ist auch ansonsten mit der Experimentalgruppe vergleichbar.</p>
Item 7	Welche Charakteristika einer Studie sollten in einem Forschungsbericht dargelegt werden, damit Sie die Validität der Befunde akzeptieren? Geben Sie 5 Beispiele.	<p><i>Exzellent (24 Punkte):</i> 5 wichtige Aspekte zur Validität</p> <p><i>Stark (18 Punkte):</i> 3-4 wichtige Aspekte zur Validität</p> <p><i>Eingeschränkt (10 Punkte):</i> 2 wichtige Aspekte zur Validität</p> <p><i>Minimal (5 Punkte):</i> Gib interne Validität oder gibt einen wichtigen Aspekt von den Beispielen</p> <p><i>Nicht vorhanden (0 Punkte):</i> Keins der obigen</p> <p>Beispiele:</p> <ul style="list-style-type: none"> - Angemessenheit des Studiendesigns - Adäquate Verblindung - Anonyme Zuweisungen - Randomisierung der Gruppenzuteilung - Ungültigkeit/Verzerrung der Messung - Bedeutung von Vergleichs-/Kontrollgruppe - Analyseabsicht - Berücksichtigung von Kovariaten (Wurden andere relevante Variablen berücksichtigt?) - Schlüssigkeit des Fazits mit erhobenen Daten - Bedeutung der Nachverfolgung aller Studienteilnehmer*innen - Angemessene statistische Analyse - Stichprobengröße/Power - Sponsoring - Zeitpunkt der Studie - Bestätigung durch andere Studien
Item 8	Nachdem Sie nun über Aspekte nachgedacht haben, welcher zur Steigerung der Validität einer Studie beitragen, geht es nun um Signifikanz und Bedeutung von Studien. Geben Sie je zwei Charakteristika von befunden an, damit sie als signifikant und bedeutungsvoll bezeichnet werden können.	<p><i>Exzellent (je 12 Punkte):</i> 2 Charakteristika</p> <p><i>Stark (je 6 Punkte):</i> 1 Charakteristikum</p> <p><i>Nicht vorhanden (0 Punkte):</i> Keine Charakteristika</p>

		Signifikanz: <ul style="list-style-type: none"> - <i>p-Wert</i> - <i>Vertrauensintervalle</i> - <i>Statistische Power</i> - <i>Genauigkeit der Schätzung</i> - <i>alpha-/beta- Fehler</i> - ... Bedeutung: <ul style="list-style-type: none"> - <i>Effektgröße</i> - <i>Vertrauensintervall</i> - <i>Experimentelle Setting</i> - ...
Item 9	<p>Nun haben Sie sich über Validität, Signifikanz und Bedeutung von Studien Gedanken gemacht. In dieser Aufgabe sollen sie 4 Aspekte, inkl. kurzer Begründung angeben, mit denen Sie die Umsetzungsmöglichkeiten der Befunden im schulischen Alltag bewerten würden.</p>	<p><i>Exzellent (24 Punkte): 4 Aspekte + Begründung</i></p> <p><i>Stark (12 Punkte): 3 Aspekte + Begründung</i></p> <p><i>Eingeschränkt (6 Punkte): 2 Aspekte + Begründung</i></p> <p><i>Minimal (3 Punkte): 1 Aspekt + Begründung</i></p> <p><i>Nicht vorhanden (0 Punkte): Kein Aspekt</i></p> Beispiele: <ul style="list-style-type: none"> - <i>Verfügbarkeit des Verfahrens (z.B. Verfahren ist noch nicht käuflich)</i> - <i>Kosten der Methode/Materialien (z.B. Materialien sind sehr teuer)</i> - <i>Fehlende Qualifikationen</i> - <i>Fachabhängigkeit (z.B. Methode ist für ein anderes Unterrichtsfach gedacht)</i> - <i>erforderliche Infrastruktur ist nicht gegeben nicht vergleichbar (z.B. nur in einem Labor durchführbar)</i> - <i>Lernendenmerkmale (z.B. Methode passt nicht zu den Charakteristika der Lernendengruppe)</i> - <i>Akzeptanz (z.B. Lernende/Peers/Vorgesetzte akzeptieren die Methode nicht)</i> - <i>Zeit (z.B. Methode braucht viel Zeit, diese wird aber zum Erfüllen des Curriculums benötigt)</i> - ...

Max. points: 216

v. Interviewleitfaden der Interventionsstudie

Interview-Guide

Code-Generierung:

Bitte erstellen Sie einen Code, welcher uns dabei hilft ihre Daten der verschiedenen Messzeitpunkte zusammen zu führen.

Erster und letzter Buchstabe des Vornamens ihrer Mutter.		Erster und letzter Buchstabe des Vornamens ihres Vaters.		Erster und letzter Buchstabe ihres eigenen Vornamens.		Tag des Geburtsdatums ihrer Mutter.	
(e.g. ANNE → AE)		(e.g. PETER → PE)		(e.g. JOHN → JN)		(e.g. 02.07.64 → 02)	

1.) Was verstehen Sie unter Evidenzen?

- Haben Sie Beispiele für Evidenzen?
- Was sind die Merkmale von Evidenzen in diesem Beispiel? Worauf bezieht sich dies?
- Kennen Sie andere Merkmale von Evidenzen?

2.) Wo spielen Ihrer Meinung nach Evidenzen eine wichtige Rolle in unserem täglichen Leben?

- In welchen Berufen? (z.B. Arzt, usw.)
- In welchen Bereichen? (z.B. Politik, usw.)
- In welchen Fachgebieten/Bereichen? (z.B. Impfungen, Klimawandel, usw.)

3.) Wann verwenden sie Evidenzen in ihrem täglichen Leben und wie sehen diese Evidenzen aus?

4.) Welche Rolle spielen Evidenzen in Ihrem Lehrerberuf?

- Welche Art von Evidenzen verwenden Sie, um Bewertungsmodi auszuwählen?
- Welche Art von Evidenzen verwenden Sie, um den Inhalt Ihres Unterrichts zu planen?
- Welche Art von Evidenzen verwenden Sie, um die Methoden usw. Ihres Unterrichts zu planen?
- Welche Art von Evidenzen verwenden Sie mit Ihren Schüle*innen in Ihrem Unterricht?
- Beauftragen Sie Ihre Lernenden, Evidenzen zu sammeln?

5.) Wie beurteilen Sie die Rolle von Evidenzen für die Schule?

- In welchen Bereichen der Schule halten Sie Evidenz für wichtig?
- Für wen ist Evidenz in der Schule wichtig?

6.) Wie gut hat Sie Ihre Ausbildung darauf vorbereitet, mit Evidenzen in Ihrem Lehrerberuf umzugehen?

- In welchen Situationen generieren/beobachten/sammeln Sie Evidenzen im Unterricht? Bitte beschreiben/erzählen Sie.
- Glauben Sie, dass sie die Qualität von Forschungsartikeln bewerten können?
- Glauben Sie, dass sie in der Lage sind, Studien im kleinen Maßstab selbst zu planen und durchzuführen?

7.) Gibt es noch etwas, was Sie erwähnen möchten?

vi. Kodiermanual zum leitfadengestützten Interview der Interventionsphase

Das Kodiermanual entstammt dem Erasmus+ Projekt *Research in Teacher Education* und liegt daher in englischer Sprache vor.

Coding manual

Coding unit: Whole sentences (these do not have to be grammatically correct)

For one sentence, multiple codes are possible.

category	definition	example	rule
Awareness	Building awareness and positive attitudes towards evidence use		
Tentativeness of knowledge/evidence	Deals with the possibility of knowledge/evidence to change over time	„You never have an absolute proof, even if people who are knowledgeable say, yes, I agree with them.“ (BEHGLA17, ca. 03:29min)	No reference to the data needed to construct evidence.
Role of evidence for school	Awareness of the role of evidence for school	„In theory, I think it's high, in practice, not so high. So, I think theoretically it is important to refer to empirical evidence, for example, in daily school life and also in all the task areas of schools, i.e. education, training, whatever. But I think that in the practice of schools, this simply comes short.“ (AAAMAA08, ca. 17:49min)	No process steps, understanding or reference to specific themes/methods, etc. More general the awareness of the role of evidence for school
Novelty of evidence	Deals with differences, or opinions on the novelty of evidence.	„ Yes, that what is somehow new is of course more important to me than what has already been done a hundred times [...] So that has no great value if you only repeat an experiment that has already been done.“ (BEHGLA17, ca. 28:12min)	
Agree	Building mutual understanding and agreement on policy-relevant questions and the kind of evidence needed to answer them		
Understanding/perception of evidence	Understanding of evidence, e.g. data needed to build up evidence, study design, format of evidence, examples of evidence, etc.	„[...] if there are numbers behind it and I can understand the numbers, I can understand the methodology behind it, so of course the methodology would always be important for me to know, [...] So is there a study behind it or experiments or something? So, if you can tie it to concrete scientific aspects. Then I would say that they are of good quality.“ (AAAMAA08, ca. 06:58min)	

Evidence in everyday life	Areas/topics of everyday and personal life, where evidence is important.	"So, I think mostly in the media, like in advertising or something." (AAAMAA08, ca. 03:37min)	
Evidence vs. personal views	Relationship between evidence and personal views	"Yes, well, it's more or less a personal opinion, as I actually said at the beginning, so it's not really evidence-based." (BEHGLA17, ca. 14:07min)	No specific reference to the tentativeness of evidence
Evidence in school	Evidence in school in general	"So, for me, science education matter is more accessible than subject matter". (AAAMAA08, ca. 12:39min)	No references to specific areas of lesson planning (cf. Kattmann)
from pure science (CK)	Evidence regarding the scientific area (cf. Kattmann) with specific reference to themes/content knowledge, etc.	"[...] above all, I also convey professional content, yes, and this professional content is also somehow provable, so there must be evidence behind it." (AAAMAA08, ca. 09:18min)	
General education/psychology (PK)	Evidence regarding general education/psychology aspects.	"For example, how many answer options are given in the test, is it a multiple-choice test? That you are aware that the quality of the measurement [...] is kept high." (HEUONS29, ca. 08:50min)	
Science education (PCK)	Evidence regarding the science education/pedagogical area (cf. Kattmann) with specific reference to themes/methods, etc.	"On the other hand, we also had students' conceptions, for example, which can also be taken as evidence somehow, [...]". (BEHGLA17, ca. 09:36min)	
comparison of different quality of evidence	Differences in the quality of evidence	"[...] then I would have said that it was going in that direction. But not really yet, just a tendency towards it." (AXTNMN17, ca. 11:13min)	
Access and communication	Providing communication of, and access to, evidence		
Preparation by the university	Aspects regarding the preparation provided by the university	"I think what is more difficult, what the university doesn't necessarily teach, is how to use this evidence in a meaningful way in the classroom." (AAAMAA08, ca. 21:22min)	
Interact	Facilitating interactions between decision-makers and researchers		
Pupils interaction with evidence	(Process of) Interaction of students with evidence	"[...] Well, I mean, if they do experiments in class, then maybe it's not interesting for science, but for the students it's already/ Well, I mean, maybe they realize, ok, I've seen this in the experiment and it really has to be like that." (BEHGLA17, ca. 19:33min)	
Skills	Supporting decision-makers to develop skills accessing and making sense of evidence		
Searching for evidence	Process of searching for evidence.	"Actually, very similar [science and science education]. [...] Accordingly, I don't really see a big difference". (BEHGLA17, ca. 16:40min)	No reference to specific areas of searching for evidence, or focus on both areas.

scientific	Searching evidence regarding the scientific area (cf. Kattmann)	"Yes, I would also first look for real scientific articles. [...] when I have already gained an overview, so to speak, yes, what is RIGHT there scientifically, or so." BEHGLA17, ca. 10:43min)	
Science education/pedagogical	Searching evidence regarding the science education/pedagogical area (cf. Kattmann)	"I mean, I can look up students' conceptions somehow. For example, in this book by Barke or something. Or simply, I mean, with several years of professional experience, I know, ok, the seventh graders are always like this [...]" BEHGLA17, ca. 13:04min)	
Rating of articles	Personal perception of the ability to rate the quality of research articles.	"[...] So I would say a rough assessment of whether it's good or not. [...]" (AAAMAA08, ca. 25:09min)	
Planning and conducting a small-scale study by themselves	Personal perception of the ability to plan and conduct a small-scale study.	"So, I'll say, we've now got to know different things about how you can do it and so basically I would have said yes, yes." AXTNMN17, ca. 32:27min)	
Structures and processes	Influencing decision-making structures and processes		
Task areas of school with special attention for evidence	Areas, or people in school, who have a special need to work with evidence.	"So maybe it goes in the direction of maybe school management or something like that. But I also don't know/ or maybe something like when you now have liaison teachers or something, who then know, ok, how do I best talk to my pupil now [...]" AXTNMN17, ca. 24:31min)	
Evidence over time as a teacher	Differences/changes over the time as a teacher (even pre-service teacher).	"I don't want to say that it is becoming more important, but I think it is simply changing. But the importance always remains the same, because I think that right now the older teachers have to look for evidence in the area of digital media, for example [...]" (AAAMAA08, ca. 19:06min)	
Policy-driven decision making	Decision based on policy rather than evidence	"[...] I would rather teach electron transfer, because I'm sure that you can actually understand that in the eighth grade. But well, I'm not allowed to do that because it's dictated by politics, so I'd say that's a bit of a disadvantage. (LERFJA31, ca. 28:00min)	
Institutional constraints	Restrictions and structural aspects of the institution, which lead to, or hinder evidence-based practice.	"Simply because I can see from my friend how much teachers have to do and I don't think you have time to spend your free time on something like that." (GAJFMS03, ca. 30:30min)	

vii. Checkliste zur Analyse von Unterrichtsentwürfen

Die Checkliste zur Analyse von Unterrichtsentwürfen entstammt dem Erasmus+ Projekt *Research in Teacher Education* und liegt daher in englischer Sprache vor.

Checklist for lesson planning

Code-Generation:

Please generate your private code which allows us to compare your answers on different time points.

First and last letter of your mother's first name. (e.g. ANNE → AE)		First and last letter of your father's first name. (e.g. PETER → PR)		First and last letter of your own first name. (e.g. JOHN → JN)		Day of your mother's birthday. (e.g. 02.07.64 → 02)	

Was evidence used to plan the lesson? What kind of evidence was used?

Example:

- Evidence is used to plan the lesson (scientific/educational evidence).
- Sources are indicated.

Are students asked to select evidence on a topic themselves?

Examples:

- The topic of the lesson is presented and the students first search for evidence themselves.
- The students do internet searches.

Are Students supported to find evidence?

Examples:

- Students get help where they find evidence.
- Students receive help on how to design their search for evidence (search strategies).
- Students get help on what kind of evidence is useful for the topic.
- Materials on the topic are provided.

Are there periods of reflection on the statements of the evidence used?Examples:

- Students present the main statements of the evidence to each other and discuss them afterwards.
- A panel discussion is held based on the evidence.

--

viii. Checkliste zur Analyse von Unterrichtsmitschauen

Die Checkliste zur Analyse von Unterrichtsmitschauen entstammt dem Erasmus+ Projekt Research in Teacher Education und liegt daher in englischer Sprache vor.

Checklist for lesson observation

Code-Generation:

Please generate your private code which allows us to compare your answers on different time points.

First and last letter of your mother's first name. (e.g. ANNE → AE)		First and last letter of your father's first name. (e.g. PETER → PR)		First and last letter of your own first name. (e.g. JOHN → JN)		Day of your mother's birthday. (e.g. 02.07.64 → 02)	

How is evidence brought into the classroom?

Example:

- In class, evidence is made available as teaching material (see supporting students in using evidence).

--

Are students asked to select evidence on a topic themselves?

Examples:

- The topic of the lesson is presented and the students first look for evidence themselves.
- The students do internet searches.

--

Are students supported in using evidence? If so, how?

Examples:

- Students get help where they find evidence.
- Students receive help on how to design their search for evidence (search strategies).
- Materials on the topic are provided.

--

Is there a discussion with students about what kind of evidence is important for the topic? What aspects are addressed?

Examples:

- Discussions about what type of evidence is useful. (e.g. field studies or laboratory studies)
- Discussions on the advantages and disadvantages of the different types of evidence are held.
- The target group is addressed too.

Is there an open, neutral atmosphere regarding evidence on the topic?

Example:

- The teacher presents evidence by him-/herself.
- The teacher evaluates evidence in conversations to the students.
- The teacher reveals his/her own opinion regarding the topic.
- Teacher behavior

Are the statements of the evidence used discussed with students?

Examples:

- Students present the main statements of the evidence to each other and discuss them afterwards.
- A panel discussion is held based on the evidence.

Is there a discussion about why evidence and the assessment of the quality of evidence are important? Which aspects are addressed?

Examples:

- Making reliable statements.
- Evidence is reflected critically.
- Opinions are not based on the number of people who share this opinion.

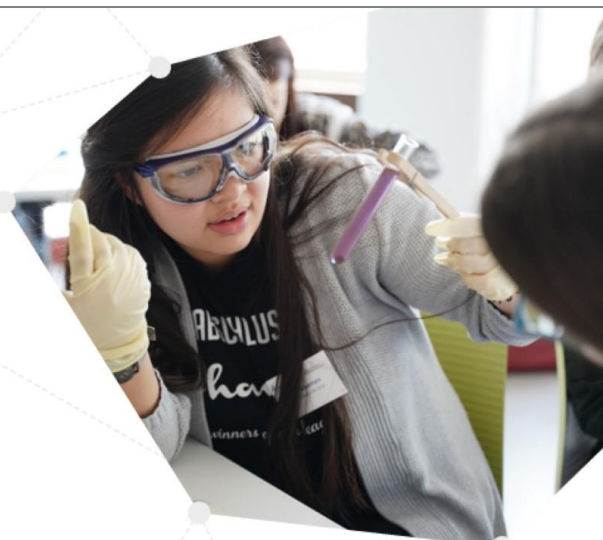
II. Zweiter empirischer Teil

- i. Vorlage zur Erstellung von *Model-of-data*

[illegible]

ii. **Materialien der Trainingsphase**

Die Trainingsphase wurde durch die vorliegende Präsentation strukturiert.



DIDAKTIK DER CHEMIE

HERZLICH WILLKOMMEN

STUDIE ZUM THEMA „KOGNITIVE STRUKTUREN VON EXPERIMENTEN“

Pascal Pollmeier

Ablauf

- Einführung & Organisatorisches
- Methodentraining: Lautes Denken
- Einführung: Model-of-data
- Methodentraining: Model-of-data
- Erstellung Model-of-data
- Rückfragen & Besprechung



Voraussetzungen Software

Haben Sie PowerPoint, Keynote oder OpenOffice auf Ihrem Endgerät installiert?



3

Ablauf

Haben Sie noch Fragen zum Ablauf oder zur Organisation?



4

Vorstellung

Wer bin ich?



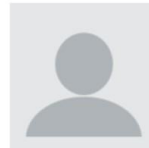
Pascal Pollmeier

26 Jahre

Lehramt für GyGe mit den Fächern
Chemie & Musik (an der HfM Detmold)

Wissenschaftlicher Mitarbeiter/Doktorand
AG Chemiedidaktik Prof. Dr. Sabine Fechner

Wer bist du?



Name

Alter

Studium

Aktuell besuchte Veranstaltung im Bereich Chemie

5

Vorstellung

Um deine Angaben später deinen weiteren Unterlagen zu sortieren zu können, müssten wir deinen persönlichen, anonymen Code generieren.

Erster und letzter Buchstabe des Vornamens der Mutter.	Erster und letzter Buchstabe des Vornamens des Vaters.	Tag des Geburtsdatums der Mutter	Erster und letzter Buchstabe des eigenen Vornamens.
(z.B. Annette → AE)	(z.B. Peter → PR)	(z.B. 04.07.1665 → 04)	(z.B. Björn → BN)



6

Lautes Denken

- Äußere alle Gedanken und Einfälle während der Bearbeitung verbal

Du solltest ...

- ... laut lesen, sofern du etwas liest.
- ... keine Gedanken verschweigen.
- ... deine Gedanken nicht erst ordnen, sondern alles frei mitteilen.
- ... möglichst alle Gedanken äußern.



Es gibt beim lauten Denken kein Richtig oder Falsch!

7

Lautes Denken

Übungsaufgabe

„Multipliziere bitte 12 mal 13 im Kopf und erzähle alles,
was dir dabei durch den Kopf geht.“



8

Lautes Denken

Übungsaufgabe



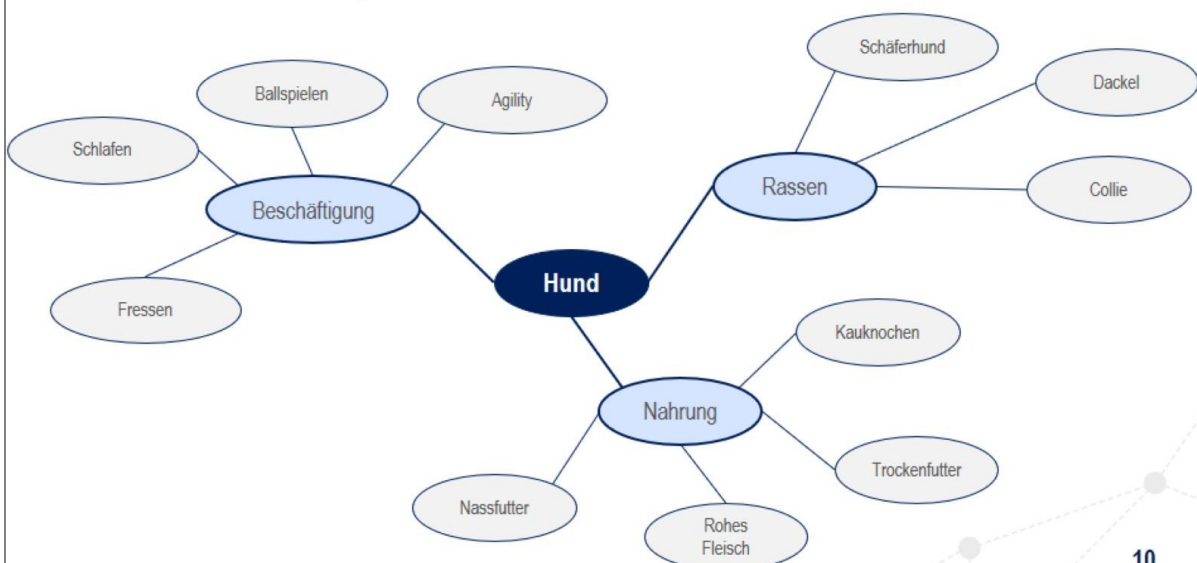
Du siehst gleich ein Video. Sieh dir das Video an und denke dabei laut.



9

Model-of-data

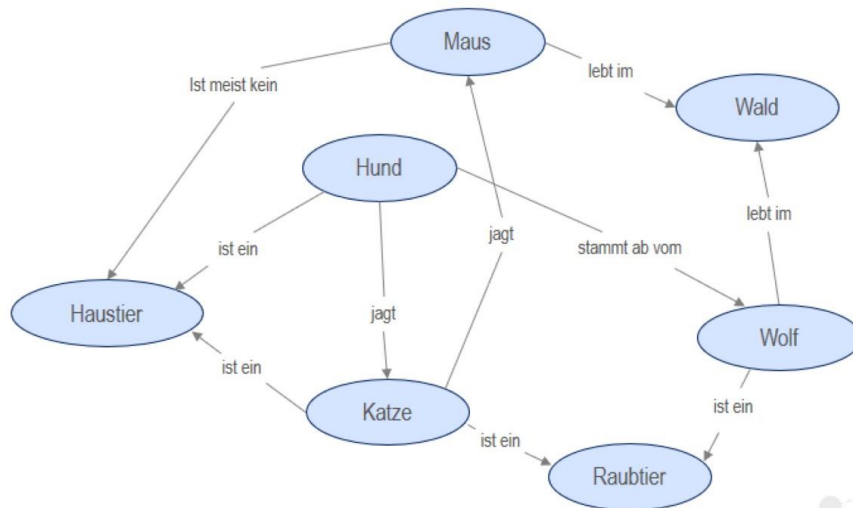
- Ähnlichkeit zu Mind-Maps



10

Model-of-data

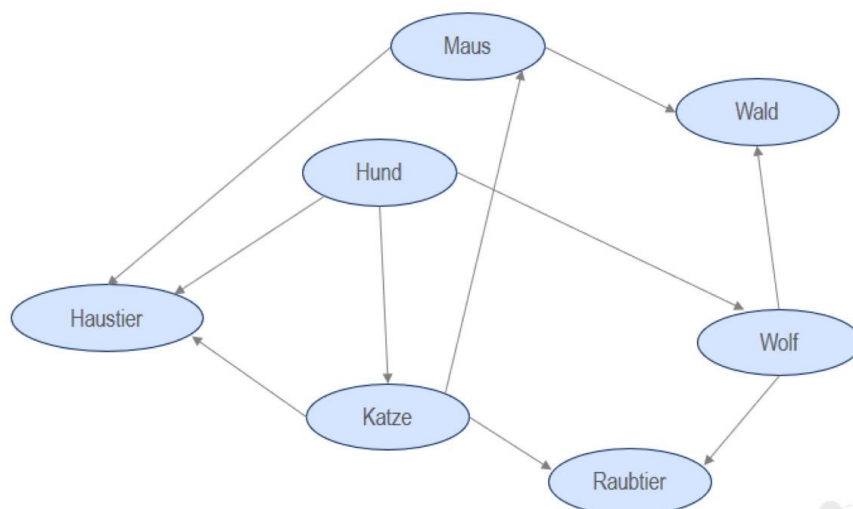
- Ähnlichkeit zu Concept-Maps



11

Model-of-data

Unterschied I: Keine Begriffe auf den Pfeilen zwischen den Kästen.

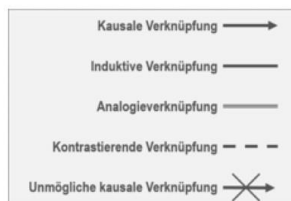


12

Model-of-data

Unterschied II: Unterschiedliche Pfeile & Striche mit unterschiedlichen Aussagen.

Die unterschiedlichen Pfeilarten geben an, **auf welche Art und Weise** zwei Aspekte miteinander verknüpft werden.

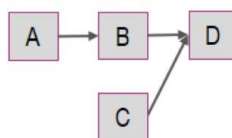


13

Model-of-data

Kästen

Die Kästen beinhalten **Beobachtungen**, oder **Hintergrundwissen** zum Thema.



Die Kästen können innerhalb einer Kette stehen, oder zu einer Kette hinzugefügt werden.

15

Model-of-data

Kausale Verknüpfung

Der Pfeil einer kausalen Verknüpfung zeigt von der Ursache zum Effekt.



Der Ball wird geworfen (Ursache), **dadurch** läuft der kleine Hund los (Effekt).



16

Model-of-data

Induktive Verknüpfung

Induktive Verknüpfungen stellen Verallgemeinerungen dar. Es wird vom Einzelfall auf eine komplette Kategorie geschlossen.



Der kleine Hund kann sehr schnell rennen (Einzelfall). Daher nehme ich an, dass alle kleinen Hunde sehr schnell rennen können (Verallgemeinerung).



17

Model-of-data

Analogie-Verknüpfung

Analogie-Verknüpfungen stellen Verallgemeinerungen dar. Dabei wird nicht innerhalb einer Kategorie verallgemeinert, sondern zwischen 2 Kategorien, welche sich in einigen, wenigen Aspekten ähneln.

Der kleine Hund
kann sehr schnell
rennen.

Alle kleinen Hunde
können sehr
schnell rennen.

Alle kleinen
Vierbeiner können
sehr schnell rennen.

Kleine Hunde können sehr schnell rennen (Einzelfall). Daher nehme ich an, dass alle kleinen Vierbeiner sehr schnell rennen können (Verallgemeinerung).



18

Model-of-data

Kontrastierende Verknüpfung

Kontrastierende Verknüpfungen zeigen die Widersprüche/Gegensätze zwischen verschiedenen Aspekten auf.

Der kleine Hund
kann sehr schnell
rennen.

Alle kleinen Hunde
können sehr
schnell rennen.

Schnecken können sich nur
langsam bewegen.

Alle kleinen Hunde können sehr schnell rennen. Schnecken hingegen können sich nur langsam bewegen.
(Unterscheidungskriterium: Bewegungsgeschwindigkeit)



19

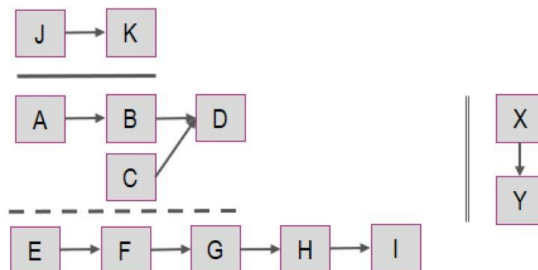
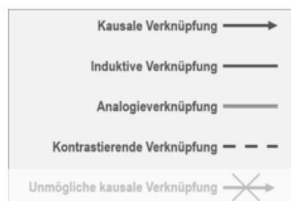
Model-of-data

Verknüpfung längerer Ketten

Induktive, analoge und kontrastive Verknüpfungen können auch zwischen längeren (kausalen) Ketten gebildet werden.

Die Verknüpfungen können in alle Richtungen (horizontal/vertikal) gebildet werden.

Auch die Kästen können variiert werden.



20

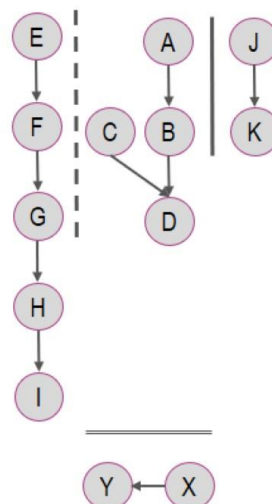
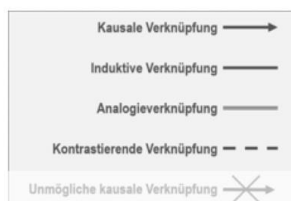
Model-of-data

Verknüpfung längerer Ketten

Induktive, analoge und kontrastive Verknüpfungen können auch zwischen längeren (kausalen) Ketten gebildet werden.

Die Verknüpfungen können in alle Richtungen (horizontal/vertikal) gebildet werden.

Auch die Kästen können variiert werden.



Wichtig ist, dass die Verknüpfungsarten korrekt angewendet werden. !

21

Model-of-data

Unmögliche kausale Verknüpfung

Unmögliche kausale Verknüpfungen zeigen auf, dass ein Effekt nicht aufgrund der genannten Ursache auftreten kann.



Der kleine Hund Felix hat nur drei Beine und kann dadurch nicht schnell rennen (Ursache). Alle kleinen Hunde können schnell rennen (unmöglicher kausaler Effekt).



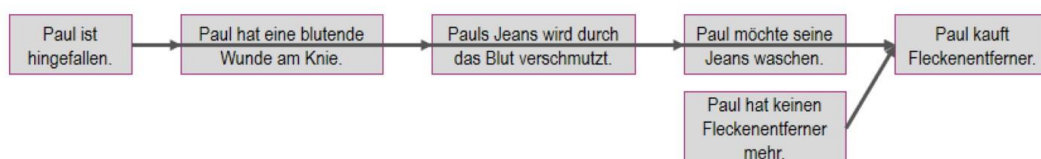
22

Model-of-data

Mögliche Schwierigkeiten

Verkürzte Kausalschlüsse:

Achte bei Kausalschlüssen darauf, dass die Ursache wirklich zum entsprechenden Effekt führt und keine Informationen zwischen zwei Aspekten ausgelassen werden.



23

Model-of-data

Welche Fragen hast du noch zu den unterschiedlichen Verknüpfungsarten, oder zur Erstellung von Model-of-data?



24

Model-of-data

Im nächsten Schritt wirst du selbst ein Model-of-data zu einem Experiment erstellen.

Diese Phase ist als **Training der Methode** gedacht. Du kannst Unklarheiten und Fragen bei der Erstellung jederzeit ansprechen.



25

Model-of-data

Ein Beispiel ...



Denke laut!

Du siehst gleich ein Video zu einem chemischen Experiment. Bitte schaue dir das Video an und äußere währenddessen und im Anschluss alle Gedanken, welche dir durch den Kopf gehen.

Gerne kannst du das Video auch zwischendurch pausieren und erste Gedanken äußern.



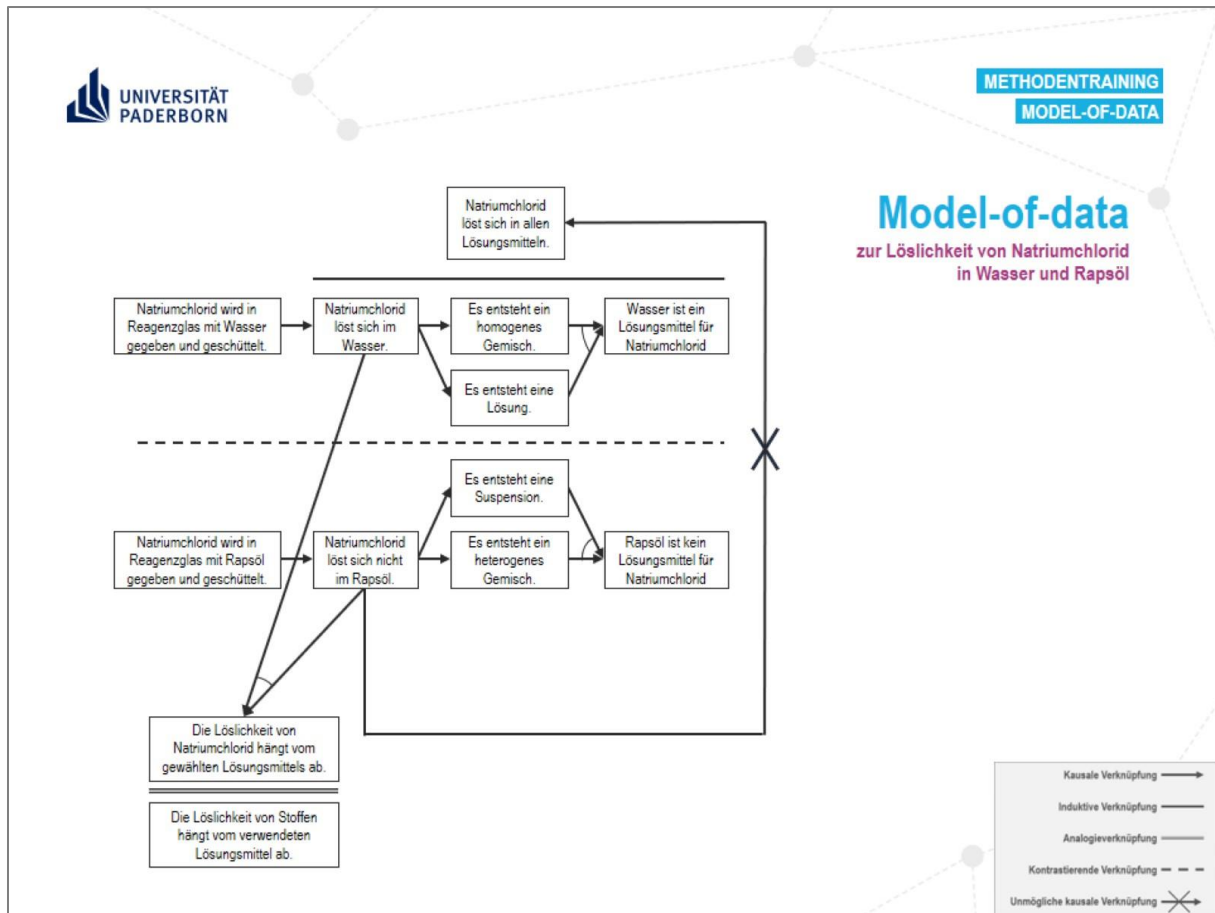
26

Model-of-data

Du erhältst jetzt eine
PowerPoint/OpenOffice/Keynote-Vorlage.



27



iii. Videos zu den Experimenten der Trainings- und Modellierungsphase

Innerhalb von Trainings- und Modellierungsphase wurden insgesamt 4 Videos eingesetzt, welche online unter den unten angegebenen Daten abrufbar sind.

Videos der Trainingsphase

Video zur Einführung in Model-of-data

<https://youtu.be/kSEn8ddukAE>



Video zur Erstellung eines Model-of-data

<https://youtu.be/NYCWLqvpU9w>



Videos der Modellierungsphase

Video zu Experiment 1

<https://youtu.be/bXYhMJufqp4>



Video zu Experiment 2

<https://youtu.be/lg8QwMq9ERw>



iv. Expert*innen-MOD zu den Experimenten der Modellierungsphase

Zur Auswertung der durch die Proband*innen erstellten *Model-of-data* (MOD) wurden Expert*innen-MOD erstellt.

Experten-Model-of-data: Reaktion von Säuren mit Natronlauge

Legende:

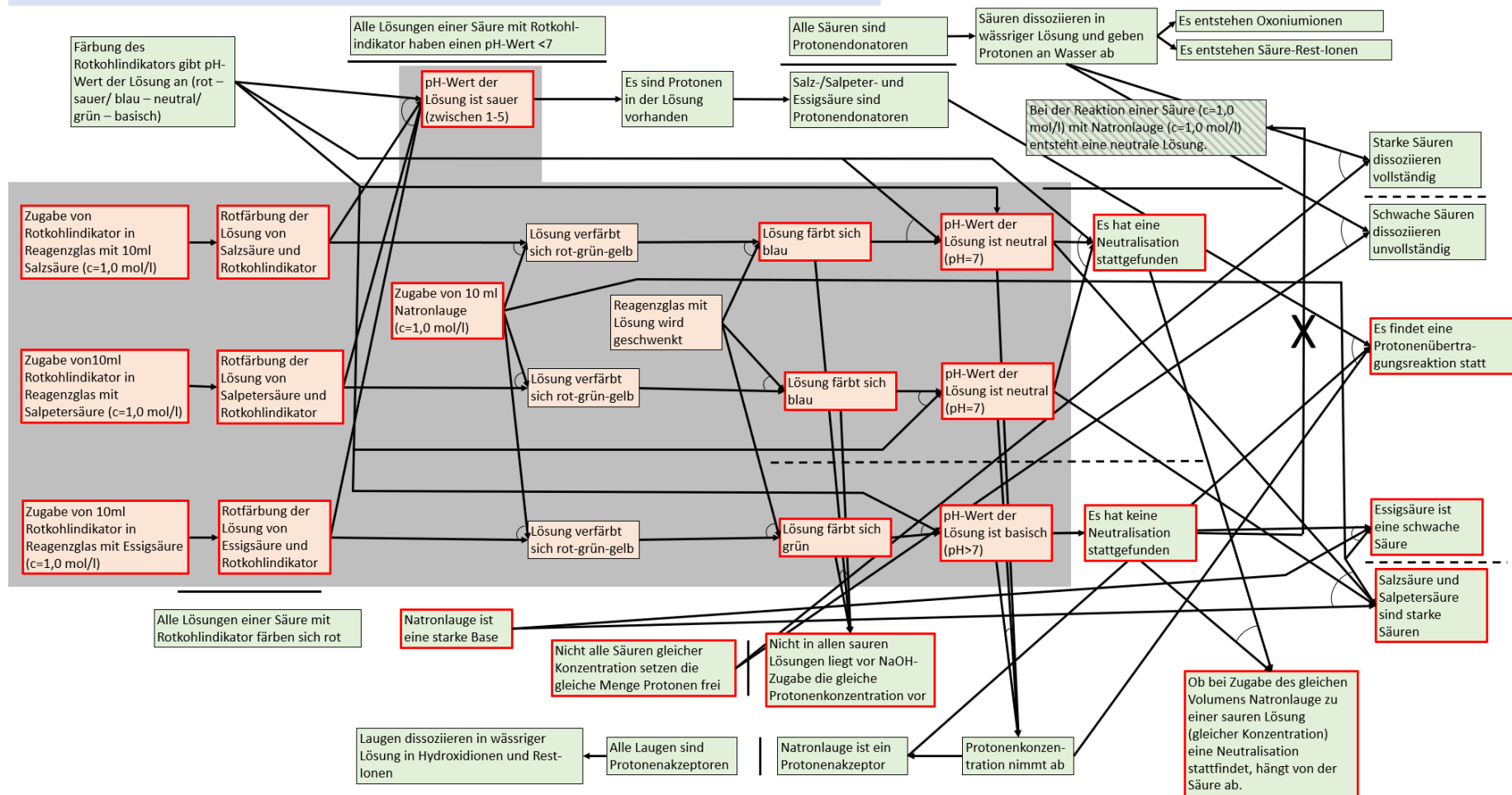
- Kausale Verknüpfung
- * Kausale Verknüpfungen mit mehreren Ursachen werden durch Halbkreise verdeutlicht.
- Induktive Verknüpfung
- Analogieverknüpfung
- - - Kontrastierende Verknüpfung
- ⊗ Unmögliche kausale Verknüpfung

- Beobachtungs-Aspekte
- Bereich der Beobachtungen
- Konzeptwissen-Aspekte
- Konzeptwissen-Aspekt, welcher durch unmöglichen Kausalschluss falsifiziert wird

Rot-umrahmte Kästen markieren essentielle Aspekte, welche zur Erklärung des Versuchs zwingend notwendig sind.



Die kausalen Verknüpfungen zwischen Beobachtungsaspekten sollten immer als verkürzte Verknüpfungen betrachtet werden. Hier könnten Erklärungen auf Ebene des Konzeptwissens integriert werden. Innerhalb des Experten-Model-of-data wurde nur im Bereich der anomalen Daten Konzeptwissen eingebracht.



Experten-Model-of-data: Siedepunkt von Wasser

Legende:

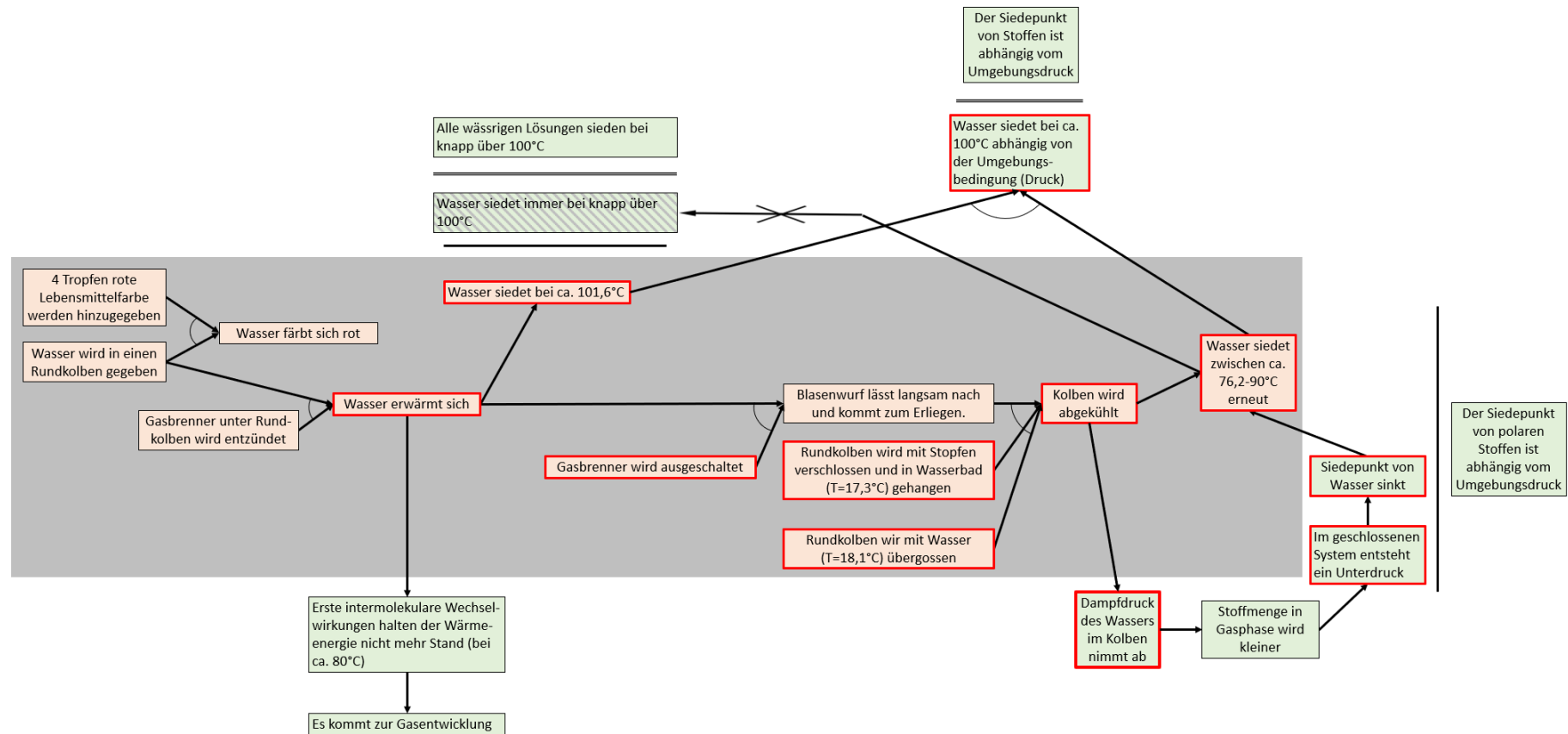
- Kausale Verknüpfung
- *Kausale Verknüpfungen mit mehreren Ursachen werden durch Halbkreise verdeutlicht.
- Induktive Verknüpfung
- Analogieverknüpfung
- - - Kontrastierende Verknüpfung
- ✗ Unmögliche kausale Verknüpfung

- Beobachtungs-Aspekte
- Bereich der Beobachtungen
- Konzeptwissen-Aspekte
- Konzeptwissen-Aspekt, welcher durch unmöglichen Kausalschluss falsifiziert wird

Rot-umrahmte Kästen markieren
essentielle Aspekte, welche zur
Erklärung des Versuchs zwingend
notwendig sind.



Die kausalen Verknüpfungen zwischen
Beobachtungsaspekte sollten immer als verkürzte
Verknüpfungen betrachtet werden. Hier könnten
Erklärungen auf Ebene des Konzeptwissens
integriert werden. Innerhalb des Experten-Model-
of-data wurde nur im Bereich der anomalen Daten
Konzeptwissen eingebracht.



v. Gefährdungsbeurteilungen zu den eingesetzten Experimenten

Für die Experimente der Trainings- sowie Modellierungsphase wurden Gefährdungsbeurteilungen angelegt, welche im Folgenden dargestellt werden.

DOKUMENTATION DER GEFÄHRDUNGSBEURTEILUNG NACH GefStoffV

1. Tätigkeit/Experiment

Löslichkeit von Salz in Rapsöl und Wasser

Jgst:

lfd Nr.:

Tätigkeitsbeschreibung:

Zwei Reagenzgläser werden je ca. zwei fingerbreit mit Wasser bzw. Rapsöl gefüllt. Anschließend wird je eine Spatelspitze Natriumchlorid hinzugegeben und das Lösungsverhalten beobachtet

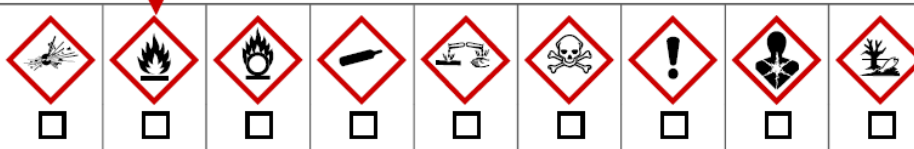
geplanter Einsatz: Lehrerversuch ☒ Schülerversuch ☐ besonders schutzbedürftige Personen ☐

2. Ermittlung der gefährlichen Stoffeigenschaften (Edukte/Produkte/Nebenprodukte)




Stoff/Gemisch	Signalwort	H-Ziffern und H-Sätze
1. Wasser	-	-
2. Rapsöl	-	-
3. Natriumchlorid	-	-

Werden Gasbrenner verwendet? Ja ☐ Nein ☒

Piktogramme der beteiligten Stoffe



3. Beurteilung der Gefahren

Gefahr	zu prüfen mit	vorhanden	nicht vorhanden
durch Haut- und Augenkontakt	Flussdiagramm, S. 3	<input type="checkbox"/>	<input checked="" type="checkbox"/>
durch Einatmen		<input type="checkbox"/>	<input checked="" type="checkbox"/>
durch Brand, Explosion	  	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Sonstige Gefahren	Tabelle „Sonstige Gefahren“, S. 5	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Ggf. Erläuterungen zu weiteren Gefahren:

Ergebnis der verpflichtenden Substitutionsprüfung nach Gefahrstoffverordnung:

Es ist keine weitere Prüfung erforderlich, da keine Gefahrstoffe für das Experiment verwendet werden oder entstehen.







4. Beurteilung des Grads der Gefährdung für die gesamte Tätigkeit

gering ☒ mittel ☐ hoch ☐ sehr hoch ☐

5. Verwendungsverbote und Tätigkeitsbeschränkungen

für die Personen unter 1. werden beachtet (vgl. RISU-NRW). ja ☒

6. Festlegung der Maßnahmen für die geplante Tätigkeit

Mindest- standard vgl. RISU-NRW						
	Schutzbrille	Handschuhe	Abzug	geschl. System	Lüften	Brandschutz
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

weitere Maßnahmen:

☐

Maßnahmen für besonders schutzbedürftige Personen:

☐

7. Entsorgung (optional)

Abfälle können über den Ausguss entsorgt werden.

Anlagen (z. B. Versuchsaufbau):

Datum: 02/2021

Unterschrift:

P. Döllmeier

DOKUMENTATION DER GEFÄHRDUNGSBEURTEILUNG NACH GefStoffV

1. Tätigkeit/Experiment

Reaktion von starken und schwachen Säuren mit Natronlauge

Jgst:

Ild Nr.:

Tätigkeitsbeschreibung:

Je 10ml Salz-, Salpeter- und Essigsäure ($c=0,1\text{ mol/l}$) werden mit je 10ml Natronlauge ($c=0,1\text{ mol/l}$) und Rotkohllindikator versetzt.geplanter Einsatz: Lehrerversuch ☒ Schülerversuch ☐ besonders schutzbedürftige Personen ☐

2. Ermittlung der gefährlichen Stoffeigenschaften (Edukte/Produkte/Nebenprodukte)

Stoff/Gemisch	Signalwort	H-Ziffern und H-Sätze
1. Salzsäure ($c=0,1\text{ mol/l}$)	Achtung	H290
2. Salzpetersäure ($c=0,1\text{ mol/l}$)	-	-
3. Essigsäure ($c=0,1\text{ mol/l}$)	-	-
Natronlauge ($c=0,1\text{ mol/l}$)	Achtung	H290
Rotkohllindikator (Rotkohlsaft)	-	-

Werden Gasbrenner verwendet? Ja ☐ Nein ☒

Piktogramme der beteiligten Stoffe



3. Beurteilung der Gefahren

Gefahr	zu prüfen mit	vorhanden	nicht vorhanden
durch Haut- und Augenkontakt	Flussdiagramm, S. 3	<input checked="" type="checkbox"/>	<input type="checkbox"/>
durch Einatmen		<input type="checkbox"/>	<input checked="" type="checkbox"/>
durch Brand, Explosion		<input type="checkbox"/>	<input checked="" type="checkbox"/>
Sonstige Gefahren	Tabelle „Sonstige Gefahren“, S. 5	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Ggf. Erläuterungen zu weiteren Gefahren:

Ergebnis der verpflichtenden Substitutionsprüfung nach Gefahrstoffverordnung:

Substitution von Gefahrstoffen, Verwendungsformen und -verfahren wurde geprüft. Der Versuch ist zur Vermittlung wesentlicher Lerninhalte nicht verzichtbar und kann unter Einhaltung der in der Versuchsvorschrift genannten Einschränkungen und mit den dort genannten Schutzmaßnahmen durchgeführt werden. Gefährliche Stoffeigenschaften oder andere Gefährdungen, die eine Durchführung durch Schüler/innen oder Lehrkräfte grundsätzlich ausschließen würden, sind nicht bekannt. Die Stoffliste DGV Information 213-098 in degintu.dguv.de wurde berücksichtigt.







4. Beurteilung des Grads der Gefährdung für die gesamte Tätigkeit

gering ☒mittel ☐hoch ☐sehr hoch ☐

5. Verwendungsverbote und Tätigkeitsbeschränkungen

für die Personen unter 1. werden beachtet (vgl. RISU-NRW). ja ☒

6. Festlegung der Maßnahmen für die geplante Tätigkeit

Mindest- standard vgl. RISU-NRW						
	Schutzbrille	Handschuhe	Abzug	geschl. System	Lüften	Brandschutz
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

weitere Maßnahmen:

☐

Maßnahmen für besonders schutzbedürftige Personen:

☐

7. Entsorgung (optional)

Lösungen neutralisieren und über den Abguss entsorgen.

Anlagen (z. B. Versuchsaufbau):

Datum: 02/2021

Unterschrift:

P. Polmi

DOKUMENTATION DER GEFÄHRDUNGSBEURTEILUNG NACH GefStoffV

1. Tätigkeit/Experiment

Druckabhängigkeit der Siedetemperatur von Wasser

Jgst:

Ifd Nr.:

Tätigkeitsbeschreibung:

Wasser wird in einem Glaskolben zu kochen gebracht. Sobald die Temperatur des Wassers nach dem Ausschalten des Brenners unter 100°C fällt, wird der Kolben luftdicht verschlossen. Der Kolben wird in ein Kühlbad gehandelt und mit kaltem Wasser übergossen.

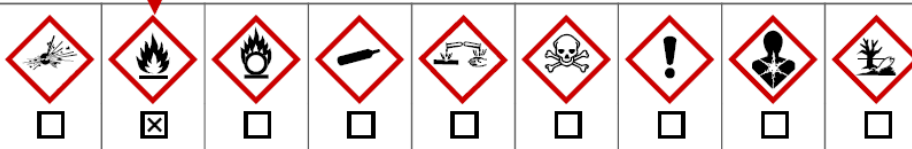
geplanter Einsatz: Lehrerversuch ☒ Schülerversuch ☐ besonders schutzbedürftige Personen ☐

2. Ermittlung der gefährlichen Stoffeigenschaften (Edukte/Produkte/Nebenprodukte)

Stoff/Gemisch	Signalwort	H-Ziffern und H-Sätze
1. Wasser	-	-
2. Propan/Verbrennungsprodukte	Gefahr	H220, H280
3.		

Werden Gasbrenner verwendet? Ja ☒ Nein ☐

Piktogramme der beteiligten Stoffe



3. Beurteilung der Gefahren

Gefahr	zu prüfen mit	vorhanden	nicht vorhanden
durch Haut- und Augenkontakt	Flussdiagramm, S. 3	<input type="checkbox"/>	<input checked="" type="checkbox"/>
durch Einatmen		<input type="checkbox"/>	<input checked="" type="checkbox"/>
durch Brand, Explosion		<input checked="" type="checkbox"/>	<input type="checkbox"/>
Sonstige Gefahren	Tabelle „Sonstige Gefahren“, S. 5	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Ggf. Erläuterungen zu weiteren Gefahren:

Ergebnis der verpflichtenden Substitutionsprüfung nach Gefahrstoffverordnung:

Substitution von Gefahrstoffen, Verwendungsformen und -verfahren wurde geprüft. Der Versuch ist zur Vermittlung wesentlicher Lerninhalte nicht verzichtbar und kann unter Einhaltung der in der Versuchsvorschrift genannten Einschränkungen und mit den dort genannten Schutzmaßnahmen durchgeführt werden. Gefährliche Stoffeigenschaften oder andere Gefährdungen, die eine Durchführung durch Schüler/innen oder Lehrkräfte grundsätzlich ausschließen würden, sind nicht bekannt. Die Stoffliste DGUV Information 213-098 in degutu.dguv.de wurde berücksichtigt.







4. Beurteilung des Grads der Gefährdung für die gesamte Tätigkeit

gering ☐mittel ☒hoch ☐sehr hoch ☐

5. Verwendungsverbote und Tätigkeitsbeschränkungen

für die Personen unter 1. werden beachtet (vgl. RISU-NRW). ja ☒

6. Festlegung der Maßnahmen für die geplante Tätigkeit

Mindest- standard vgl. RISU-NRW						
	Schutzbrille	Handschuhe	Abzug	geschl. System	Lüften	Brandschutz
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

weitere Maßnahmen:

<input type="checkbox"/>	
--------------------------	--

Maßnahmen für besonders schutzbedürftige Personen:

<input type="checkbox"/>	
--------------------------	--

7. Entsorgung (optional)

Das Wasser kann über den Abguss entsorgt werden.

Anlagen (z. B. Versuchsaufbau):

Datum: 02/2021

Unterschrift:

P. Polini

vi. Kodiermanual für MOD und lautes Denken

Das Kodiermanual umfasst sowohl die Kodierung der graphischen *Model-of-data* als auch die Kodierung der Videos.

vii. Kodiermanual für MOD und lautes Denken

	Kategorie	Kriterium	Bepunktung (nur im MOD)	Video	Outcome
Korrektheit der Verknüpfungen	Kausale Verknüpfung				Verknüpfungs-Score Summe der Punkte aller korrekt aufgestellten Verknüpfungen geteilt durch die Gesamtzahl der korrekten Verknüpfungen. Korrektheit Prozentuale Korrektheit einzelner Verknüpfungsraten (auch insgesamt).
	Kausale Verknüpfung in Aspekt-Kästchen				
		Innerhalb eines Kästchens für einen Aspekt wird ein Zusammenhang dargestellt, welcher durch eine kausale Verknüpfung dargestellt werden könnte.	Auszählen	Abweichung Formuliert, aber nicht geschrieben	
	Zwischen zwei Beobachtungs-Aspekten (können verkürzt sein)				
	Adäquat verwendet (1P)	Beziehung von Ursache-Wirkung, Pfeilrichtung korrekt	1 Punkt	Abweichung Formuliert, aber nicht geschrieben	
	Inadäquat verwendet		0 Punkte	Abweichung Formuliert, aber nicht geschrieben	
	Zwischen zwei Konzeptwissen-Aspekten				
	Adäquat verwendet (1P)	Beziehung von Ursache-Wirkung, Pfeilrichtung korrekt	1 Punkt	Abweichung Formuliert, aber nicht geschrieben	
	Inadäquat verwendet		0 Punkte	Abweichung Formuliert, aber nicht geschrieben	
	Zwischen einem Beobachtungs- und einem Konzeptwissen-Aspekt (Richtung egal)				
	Adäquat verwendet (1P)	Beziehung von Ursache-Wirkung, Pfeilrichtung korrekt	1 Punkt	Abweichung Formuliert, aber nicht geschrieben	
	Inadäquat verwendet		0 Punkte	Abweichung Formuliert, aber nicht geschrieben	
	Induktive Verknüpfung				
	Adäquat verwendet (2P)	Verallgemeinerung innerhalb derselben, ontologisch gleichen (nahen) Kategorie aufgrund eines Verallgemeinerungsgegenstandes. (Auch ok, wenn zwei Aspekte verallgemeinert werden, z.B. Eine Säure zu allen Säuren + eine Lauge auf alle Laugen)	2 Punkte	Abweichung Formuliert, aber nicht geschrieben	
	Inadäquat verwendet	z.B. wenn es sich um ein/e Fazit/Zusammenfassung handelt – es muss einen Verallgemeinerungsgegenstand geben.	0 Punkte	Abweichung Formuliert, aber nicht geschrieben	
	Analogie-Verknüpfung				
	Adäquat verwendet (2P)		2 Punkte	Abweichung	

	Kategorie	Kriterium	Bepunktung (nur im MOD)	Video	Outcome
		Verallgemeinerung über eine andere, ontologisch entfernte (ferne) Kategorie aufgrund eines Verallgemeinerungsgegenstandes.		Formuliert, aber nicht geschrieben	
	Inadäquat verwendet	z.B. wenn es sich um ein/e Fazit/Zusammenfassung handelt – es muss einen Verallgemeinerungsgegenstand geben.	0 Punkte	Abweichung	
				Formuliert, aber nicht geschrieben	
	Kontrastierende Verknüpfung				
	Adäquat verwendet (1P)	Unterscheidung auf Grundlage eines Unterscheidungskriteriums.	1 Punkt	Abweichung	
				Formuliert, aber nicht geschrieben	
	Inadäquat verwendet		0 Punkte	Abweichung	
				Formuliert, aber nicht geschrieben	
	Unmögliche kausale Verknüpfung				
	Adäquat verwendet (3P)	Darstellung einer Wirkung/eines Effekts, der nicht aus einer Ursache folgen kann.	3 Punkte	Abweichung	
				Formuliert, aber nicht geschrieben	
	Inadäquat verwendet		0 Punkte	Abweichung	
				Formuliert, aber nicht geschrieben	
Vorhandensein entscheidender Argumentationslinien (Säure-Base)	Zugabe und Rotfärbung von Rotkohllindikator zeigt Anfangs-pH-Wert von <7 (sauer) an (inkl. Einfüllen der Säuren).	Es ist nicht entscheidend in wie viele Kästen der Aspekt aufgeteilt wird. Es ist nicht entscheidend, welche Verknüpfung gewählt wurde, bzw. ob diese korrekt genutzt wurde. Dies wird in der Kategorie „Korrektheit der Verknüpfungen“ bewertet und bepunktet.	Ein Punkt pro vorhandenen Aspekt. Die Summen der Punkte im Bereich Beobachtung und Konzept werden in einer Ratio gegenübergestellt. (Musterlösung: Beobachtung 3 : 6 Konzept)	Abweichung	Ratio essentieller Aspekte Verhältnis von Beobachtungs- und Konzeptaspekten
	Zugabe der gleichen Menge und Konzentration von Natronlauge verfärbt Salzsäure und Salpetersäure bläulich (pH=7, neutral).			Formuliert, aber nicht geschrieben	
	Bei Salpeter- und Salzsäure hat eine Neutralisation mit dem End-pH-Wert=7 stattgefunden.			Abweichung	
	Zugabe der gleichen Menge und Konzentration von Natronlauge verfärbt Essigsäure grün (pH>7, basisch).			Formuliert, aber nicht geschrieben	
	Bei Essigsäure hat keine Neutralisation mit dem End-pH-Wert=7 stattgefunden.			Abweichung	
				Formuliert, aber nicht geschrieben	
				Abweichung	
				Formuliert, aber nicht geschrieben	
				Abweichung	
				Formuliert, aber nicht geschrieben	

	Kategorie	Kriterium	Bepunktung (nur im MOD)	Video	Outcome
	Es findet eine Protonenübertragungsreaktion statt.			Formuliert, aber nicht geschrieben	
	Nicht in allen Säuren liegt vor Zugabe der Natronlauge die gleichen Protonenkonzentration vor.			Abweichung	
	Starke Säuren dissoziieren vollständig, schwache Säure dissoziieren unvollständig.			Formuliert, aber nicht geschrieben	
	Essigsäure ist eine schwache Säure, Salpeter- und Salzsäure sind starke Säuren.			Abweichung	
	Natronlauge ist eine starke Base.			Formuliert, aber nicht geschrieben	
	Ob bei Zugabe des gleichen Volumens Natronlauge (gleicher Konzentration) zu einer Säure eine Neutralisation mit dem End-pH-Wert=7 stattfindet, hängt von der Säure ab.			Abweichung	
				Formuliert, aber nicht geschrieben	
				Abweichung	
Vorhandensein entscheidender Argumentationslinien (Siedepunkt)	Wassertemperatur steigt, bis es bei 101,6°C/ca. 100°C siedet (wird nicht verteilt, wenn "100°C" angegeben werden)	Es ist nicht entscheidend in wie viele Kästen der Aspekt aufgeteilt wird. Es ist nicht entscheidend, welche Verknüpfung gewählt wurde, bzw. ob diese korrekt genutzt wurde. Dies wird in der Kategorie „Korrektheit der Verknüpfungen“ bewertet und bepunktet.	Ein Punkt pro vorhandenen Aspekt. Die Summen der Punkte im Bereich Beobachtung und Konzept werden in einer Ratio gegenübergestellt. (Musterlösung: Beobachtung 4 : 3 Konzept)	Abweichung	Ratio essentieller Aspekte Verhältnis von Beobachtungs- und Konzeptaspekten
	Gasbrenner wird ausgeschaltet, Rundkolben wird luftdicht verschlossen.			Formuliert, aber nicht geschrieben	
	Rundkolben wird abgekühlt.			Abweichung	
	Dampfdruck es Wassers sinkt, Unterdruck entsteht.			Formuliert, aber nicht geschrieben	
	Siedepunkt von Wasser sinkt.			Abweichung	
	Wasser siedet zwischen ca. 76,2-90°C erneut			Formuliert, aber nicht geschrieben	
	Der Siedepunkt von Wasser ist abhängig vom Umgebungsdruck.			Abweichung	
				Formuliert aber nicht geschrieben	

Weitere Aspekte	Beobachtungsaspekte	Jeder Aspekt, der über die entscheidenden Aspekte hinausgeht.	Auszählen	Abweichung	Ratio weiterer Aspekte
	Konzeptaspekte		Auszählen	Formuliert aber nicht geschrieben	
	Inhaltlich inadäquate Beobachtungsaspekte	Beobachtungsaspekte, welche inhaltlich falsch sind.	Auszählen	Abweichung	Verhältnis von Beobachtungs- und Konzeptaspekten
				Formuliert, aber nicht geschrieben	
	Inhaltlich inadäquate Konzeptaspekte	Konzeptaspekte, welche inhaltlich falsch sind.	Auszählen	Abweichung	
				Formuliert, aber nicht geschrieben	
	Für Säure-Base-Experiment				
Grünfärbung/andere Färbung nach Reaktion von Essigsäure und Natronlauge wird dargestellt.	Unterschied der Färbung des Indikators nach Zugabe der Natronlauge bei Essigsäure im Vergleich zu Salz- & Salpetersäure.	Keine Bepunktung	Formuliert, aber nicht geschrieben		
Für Siedepunkt-Experiment:					
Erster Siedepunkt von Wasser bei über 100°C wird dargestellt.	Erster Siedepunkt des Wassers über den antizipierten 100°C (hier 101,6-102°C).	Keine Bepunktung	Formuliert, aber nicht geschrieben		
Zweiter Siedepunkt von Wasser bei unter 100°C wird dargestellt.	Zweiter Siedepunkt wird als solcher benannt, bzw. gleiche Formulierung wie beim ersten Siedepunkt (z.B. „kochen“).		Formuliert, aber nicht geschrieben		

vii. Fachwissenstest

Der Fachwissenstest wurde mittels der Umfragesoftware LimeSurvey digital umgesetzt. Das Seitenlayout des Instruments im Anhang stimmt dementsprechend nicht mit der digitalen Ansicht auf dem jeweiligen Endgerät überein, zeigt jedoch die eingesetzten Items.

1 - Das Wasserstoff-Ion ist in Wasser gelöst ein...

- ☐ ... einfach positiv geladenes Ion.
- ☐ ... zweifach positiv geladenes Ion.
- ☐ ... neutral geladenes Ion.
- ☐ ... einfach negativ geladenes Ion.

2 - Bei der Reaktion einer Base mit ... entstehen Wasser und ein Salz.

- ☐ ... einem Salz ...
- ☐ ... einer Säure ...
- ☐ ... einem Ester ...
- ☐ ... einem Metallhydroxid ...

3 - Auf dem Herd steht ein mit Wasser gefüllter Wasserkessel. Das Wasser hat schnell anfangen zu kochen. Die wahrscheinlichste Temperatur des Wasser ist ungefähr:

- ☐ ... 88°C
- ☐ ... 98°C
- ☐ ... 110°C
- ☐ ... Keiner der oben genannten Antworten kann richtig sein.

4 - Fünf Minuten später kocht das Wasser im Kessel immer noch. Die wahrscheinlichste Temperatur ist:

- ☐ ... 88°C
- ☐ ... 98°C
- ☐ ... 110°C
- ☐ ... 120°C

5 - Die Reaktion einer Säure und einer Base nennt man...

- ☐ ... Neutralisationsreaktion.
- ☐ ... Knallgasreaktion.
- ☐ ... Elektronenreaktion.
- ☐ ... Redoxreaktion.

6 - Wie viel Grad glaubst du hat die Temperatur des Dampfes oberhalb des kochenden Wassers im Kessel?

- ☐ ... 88°C
- ☐ ... 98°C
- ☐ ... 110°C
- ☐ ... 120°C

7 - Was sind typische Basen?

- ☐ Halogene
- ☐ Metallhydroxide
- ☐ Salze
- ☐ Chloride

8 - Man spricht von einer Säure-Base-Reaktion, wenn...

- ☐ ... Elektronen übertragen werden.
- ☐ ... Protonen übertragen werden.
- ☐ ... Ionen übertragen werden.
- ☐ ... Ladungen übertragen werden.

9 - Je mehr Base man zu einer Säure gibt, desto...

- ☐ ... konstanter bleibt der pH-Wert.
- ☐ ... höher steigt der pH-Wert.
- ☐ ... eher pendelt sich der pH-Wert bei 0 ein.
- ☐ ... niedriger sinkt der pH-Wert.

10 - Jim glaube, dass er immer kochendes Wasser nutzen muss, um eine Tasse Tee zu kochen. Er erzählt seinen Freunden: „Ich könnte keinen Tee kochen, wenn ich auf einem hohen Berg zelte, weil Wasser nicht in großen Höhen kocht.“

Wem stimmst du zu?

- ☐ Joy sagt: „Doch, tut es, aber das kochende Wasser wird einfach nicht so heiß wie hier.“
- ☐ Tay sagt: „Das stimmt nicht. Wasser kocht immer bei derselben Temperatur.“
- ☐ Lou sagt: „Der Siedepunkt des Wassers erniedrigt sich, aber das Wasser selbst hat immer noch 100 Grad.“
- ☐ Mai sagt: „Ich stimme Jim zu. Das Wasser kommt nicht bi zu seinem Siedepunkt.“

11 - Welche Ionen sind die Ursache für eine basische Lösung?

- ☐ Chlorid-Ionen
- ☐ Wasserstoff-Ionen
- ☐ Natrium-Ionen
- ☐ Hydroxid-Ionen

12 - Mel kocht Wasser in einem Kochtopf auf dem Herd. Was glaubst du, beinhalten die Blasen, die sich im kochenden Wasser bilden?

- ☐ Luft
- ☐ Sauerstoff- und Wasserstoffgas
- ☐ verdampftes Wasser
- ☐ In den Blasen befindet sich nichts.

13 - Alle Metallhydroxide setzen die gleichen Ionen frei, wenn man sie in Wasser löst. Welche Ionen sind das?

- ☐ H^- - Ionen
- ☐ OH^- - Ionen
- ☐ H^+ - Ionen
- ☐ OH^+ - Ionen

14 - Nachdem er einige Eier im kochenden Wasser gekocht hat, kühlt Mel die Eier ab, indem er sie in eine Schüssel mit kaltem Wasser gibt. Welcher der folgenden Sätze beschreibt den Abkühlungsprozess?

- ☐ O Temperatur wird von den Eiern auf das Wasser transferriert.
- ☐ O Kälte bewegt sich aus dem Wasser in die Eier.
- ☐ O Heiße Objekte kühlen sich natürlich ab.
- ☐ O Energie wird von den Eiern auf das Wasser transferriert.

15 - Ein positiv geladenes Wasserstoff-Ion nennt man auch...

- ☐ O ... Neutron.
- ☐ O ... Proton.
- ☐ O ... Elektron.
- ☐ O ... Nukleon.

16 - Ron stellt fest, dass seine Mutter Suppe in einem Dampfkochtopf kocht, weil sie schneller als in einem normalen Kochtopf kocht. Aber er weiß nicht warum. [Dampfkochtöpfe haben einen verschließbaren Deckel, sodass der Druck innen über dem Atmosphärendruck liegt.]

Welcher Person stimmst du am meisten zu?

- ☐ O Emi sagt: „Das passiert, weil der Druck das Wasser dazu bringt, über 100°C zu kochen.“
- ☐ O Col sagt: „Das passiert, weil der hohe Druck extra Hitze erzeugt.“
- ☐ O Fay sagt: „Das passiert, weil der Dampf eine höhere Temperatur als die kochende Suppe hat.“
- ☐ O Tom sagt: „Das passiert, weil Dampfkochtöpfe die Hitze gleichmäßiger im Essen verteilen.“

17 - Säuren...

- ☐ O ... nehmen im Wasser Hydroxid-Ionen auf.
- ☐ O ... bilden in Wasser Basen.
- ☐ O ... bilden in Wasser H^+ - Ionen.
- ☐ O ... nehmen im Wasser Protonen auf.

18 - Flüssiges Wasser kocht in einem Topf auf dem Gasherd. Der Regler ist auf „NIEDRIG“ gestellt. Während das Wasser kocht, wird der Regler auf „HOCH“ gestellt. Was passiert mit der Temperatur des flüssigen Wassers?

- ☐ Sie nimmt geringfügig ab, weil das Wasser mehr kocht.
- ☐ Sie nimmt zu, und tut dies zu einem höheren Tempo als zu dem Zeitpunkt, zu dem der Regler auf „NIEDRIG“ gestellt war.
- ☐ Sie nimmt zu, und tut dies im gleichen Tempo als zu dem Zeitpunkt, zu dem der Regler auf „NIEDRIG“ gestellt war.
- ☐ Sie bleibt gleich.

19 - Säuren werden auch als ... bezeichnet.

- ☐ ... Protonendonatoren ...
- ☐ ... Protonenakzeptoren ...
- ☐ ... Protonenbinder ...
- ☐ ... Protonenleiter ...

20 - Bitte geben Sie ihre Abiturnote an: _____

21 - Bitte geben Sie an, welche chemische Lehrveranstaltung Sie zuletzt besucht und abgeschlossen haben.

[Sofern Sie an der Universität noch keine chemische Lehrveranstaltung besucht haben, geben Sie bitte an, in welcher Jahrgangsstufe Sie in der Schule zuletzt Chemieunterricht hatten.]

22 - Mit welcher Note haben Sie Ihre letzte chemische Lehrveranstaltung abgeschlossen?

[Sofern Sie an der Universität noch keine chemische Lehrveranstaltung besucht haben, geben Sie bitte die letzte Note an, die Sie in der Schule im Chemieunterricht bekommen haben.]

23 - Bitte stufen Sie Ihre Fähigkeiten/Kenntnisse in den folgenden Gebieten ein:

	Sehr schlecht	schlecht	mittel	gut	Sehr gut
Säure-Base-Chemie	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stoffeigenschaften	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Chemisches Wissen/Kenntnisse allgemein	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Physikalisches Wissen/Kenntnisse allgemein	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

viii. Datenblatt für Proband*innen inkl. Vorlage für Notizen während des lauten Denkens

Das vorliegende Datenblatt wurde zur Erhebung demographischer Daten sowie zur Erstellung von Notizen während der Trainings- und Modellierungsphase eingesetzt.

Datenblatt

Demographische Daten & field-notes während der Durchführung der Studie

Code	
Geschlecht	
Alter	
Studium	
Aktuell besuchte Veranstaltung	
Datum Termin	
Field-notes Teil 1	
Methodische Unklarheiten im Model-of-data	
Field-notes Studienteil I	
Field-notes Studienteil II	

Kurzzusammenfassung

Dissertation Pascal Pollmeier

Abstract – Deutsch

Eine Forderung nach evidenzbasierter Praxis (EBP) im Bildungswesen kann sowohl vor dem Hintergrund der Bedeutung wissenschaftlicher Evidenzen innerhalb gesellschaftlicher Diskurse als auch der Evidenzbasierung in anderen Bereichen formuliert werden. Der Umgang mit und die Bewertung von Evidenzen stellen somit vor dem Hintergrund epistemologischer Überzeugungen relevante Aufgaben für den Chemieunterricht dar. Dabei ist unklar, inwiefern (angehende) Lehrkräfte über Kompetenzen in EBP verfügen. Bestehende Studien weisen auf einen Bedarf der fachdidaktischen Analyse und Förderung von Kompetenzen in EBP bei angehenden Lehrkräften hin.

Im Rahmen dieser Dissertation wurde die Domänenspezifität der Kompetenzen in EBP sowie deren Förderung in einem fachdidaktischen Kontext untersucht. Dabei wurde ein weiterer Schwerpunkt auf den Umgang mit anomalen Beobachtungen im Sinne widersprüchlicher Evidenzen gelegt. Innerhalb von drei Teilstudien wurden unterschiedliche methodische Zugänge zur Kompetenz in EBP auf Grundlage epistemologischer Überzeugungen gewählt.

Die Ergebnisse weisen auf einen Förderbedarf der Kompetenz in EBP für angehende Lehrkräfte aller Unterrichtsfächer hin. Eine entwickelte Intervention für angehende Chemielehrkräfte weist eine Wirksamkeit zur Förderung von Kompetenzen in EBP auf. Im Rahmen einer Modellierungsstudie kann der Einfluss von Fachwissen sowie Indizien für einen Einfluss epistemologischer Überzeugungen auf den Umgang mit anomalen Beobachtungen gezeigt werden.

Abstract – English

The role of evidence-based practice (EBP) in educational contexts can be argued from the necessity of evidence for societal debates, as well as evidence-based practice in other research fields, such as medicine. Therefore, using evidence based on epistemological beliefs is a relevant task for science education. However, it is unclear whether pre-service teachers hold necessary competences in EBP. Existing research indicates a need for the analysis and promotion of competences in EBP for pre-service teachers from a science-education perspective.

This project aims at investigating the domain specificity and promotion of competences in EBP of pre-service teachers. An additional emphasis is put on the reaction to anomalous observations as contradictory evidence. Within three sub-studies, different methodological approaches were used to gather insights into the competence in EBP from the base of epistemological beliefs.

As results show, there is a need for promotion in competences in EBP for pre-service teachers in all subjects. Therefore, the developed intervention for the promotion of pre-service chemistry teachers' competence in EBP shows to be effective. Moreover, the impact of content knowledge and the role of epistemological beliefs on the interpretation of anomalous observations could be demonstrated.