



Technische Universität München

TUM School of Education

Susanne Klatten-Stiftungslehrstuhl für Empirische Bildungsforschung

# **Studien zur Entwicklung und Evaluation eines Trainings für angehende Lehrkräfte zum Nutzen empirischer Studien**

**Sandra Wenglein**

Vollständiger Abdruck der von der Fakultät TUM School of Education der Technischen  
Universität München zur Erlangung des akademischen Grades eines

**Doktors der Philosophie (Dr. phil.)**

genehmigten Dissertation.

**Vorsitzende:** Prof. Dr. Christina Seidel

**Prüfer der Dissertation:**

1. Prof. Dr. Manfred Prenzel
2. Prof. Dr. Johannes Bauer

Diese Dissertation wurde am 05.06.2018 bei der Technischen Universität München  
eingereicht und durch die Fakultät TUM School of Education am 05.07.2018 angenommen.



## Zusammenfassung

Das kompetente Nutzen von Evidenz spielt im Bildungswesen eine immer größere Rolle (Fischer et al. 2014; Wiseman, 2010). Der Praxistransfer hat jedoch bisher in vielen Bereichen nicht, oder nur unzureichend, stattgefunden. Gerade Lehrkräfte begründen ihr Handeln kaum auf Grundlage wissenschaftlicher Evidenz (Hetmanek et al., 2015; Williams & Coles, 2007), obwohl dies in den Standards für Lehrerbildung gefordert wird (KMK, 2004). Ein möglicher Grund ist, dass die Kompetenz zur Bewertung und Nutzung bildungswissenschaftlicher Evidenz im Lehramtsstudium in der Regel kaum systematisch gefördert wird. Im Anschluss an diese Entwicklungen wurde ein Training für Lehramtsstudierende zum kompetenten Umgang mit Evidenz entwickelt. Die Grundidee dieses Trainings besteht darin, innerhalb kurzer Zeit in kooperativen Lernsettings mit problembasierten Fallbeispielen (Gräsel, 2006; Collins et al, 1989) Heuristiken zu zwei relevanten Kompetenzen zu vermitteln: (a) zur Bewertung und (b) zur argumentativen Nutzen von Evidenz.

Die Trainierbarkeit dieser Kompetenzen, sowie die Effektivität des im Rahmen dieser Arbeit entwickelten Trainings wurden mit Hilfe von drei Studien erhoben. Studie 1 diente hierbei als Pilotstudie um erste Ergebnisse bezüglich Effektivität und Akzeptanz zu erzielen. Studie 2 replizierte diese Ergebnisse und konnte auf Grund der höheren Stichprobengröße zudem Informationen hinsichtlich möglicher Einflussfaktoren generieren. In Studie 3 war es Ziel durch eine tiefere Analyse der in Studie 2 durchgeführten Gruppendiskussionen detailliertere Ergebnisse bezüglich ablaufender Prozesse des argumentativen Nutzens von Evidenz zu erhalten.

Als zentrales Ergebnis dieser Arbeit ist zu berichten, dass das Training signifikant große Effekte sowohl auf die Kompetenz des kritischen Hinterfragens von Evidenz ( $d = 1.55$ ,  $p < .001$ ) als auch auf die Kompetenz des argumentativen Nutzens ( $d = 1.36$ ;  $p < .001$ ) erzielen konnte. Auf Basis dieser, sowie weiterer Ergebnisse erweitert die vorliegende Arbeit den Forschungsstand hinsichtlich Informationen zu wirksamen Trainingsmethoden zur Förderung des kritischen Hinterfragens, sowie des argumentativen Nutzens von Evidenz. Zudem tragen die Ergebnisse zu einem tiefergehenden Verständnis der relevanten Prozesse bei, die beim Lernen des Nutzens von Evidenz ablaufen.

## Abstract

The competent use of scientific evidence is becoming increasingly important in educational policy and practice (Fischer et al., 2014; Wiseman, 2010). However, in many fields, the theory-to-practice transfer has either not occurred or is insufficient. Despite public and policy calls to do so (KMK, 2004), teachers rarely use scientific evidence to justify their decisions (Hetmanek et al., 2015; Williams & Coles, 2007). One reason might be that there is currently no systematic way to foster competences in teacher education. Due to these developments, we designed a training program for pre-service teachers. The main idea of the training was to foster two relevant competences within the shortest possible time by using heuristics and collaborative learning with problem-based case vignettes (Gräsel, 2006; Collins et al., 1989): (a) appraising evidence critically and (b) applying it argumentatively.

This thesis presents three studies on the trainability of these competences and the effectiveness of the training that was developed in the scope of this project. Study 1 served as a pilot study to obtain first results on the effectiveness of the training and the participants' motivation and acceptance of it. Study 2 replicated these results with a larger sample. Moreover, we collected data on the potential factors that influence these competences. Study 3 aimed to conduct a deeper analysis of the processes of argumentative use of evidence. Therefore, we analyzed the group discussions that were part of the training within Study 2.

A central finding of this work is that the training had significantly large effects on the participants' competences in appraising ( $d = 1.55$ ,  $p < .001$ ) and applying evidence ( $d = 1.36$ ;  $p < .001$ ). On the basis of these and further results, this work extends the state of research by providing information on effective training methods that foster the ability to critically appraise evidence and apply it argumentatively. Furthermore, these results lead to a deeper understanding of the relevant processes that occur as (pre-service) teachers learn how to use scientific evidence competently.

## Danksagung

An dieser Stelle möchte ich mich bei all den Personen bedanken, die mir durch ihre Unterstützung die Erstellung dieser Dissertation überhaupt erst möglich gemacht haben.

Mein ganz besonderer Dank gilt hierbei meinem Doktorvater Prof. Manfred Prenzel, der mir ein Umfeld mit der perfekten Balance aus konstruktiver Unterstützung und eigenem kreativen Freiraum geschaffen hat. Vielen Dank für die vielen inspirierenden Gespräche und Anregungen, aus denen ich so viel von Ihnen lernen konnte und vielen Dank für das mir entgegengebrachte Vertrauen, dass es mir ermöglicht hat mich stetig weiterzuentwickeln und über mich hinaus zu wachsen.

Zudem gilt mein herzlichster Dank meinem Zweitbetreuer Prof. Johannes Bauer, der mir jederzeit bei all meinen Fragen mit Rat und Tat zur Seite stand. Vielen Dank, dass du immer für mich da warst, mich bei all meinen Plänen unterstützt hast und mir geholfen hast neue Wege zu finden, wenn ich mal nicht weiter wusste.

Des Weiteren danke ich meinem Lehrstuhlteam (dem besten Team, das man sich nur vorstellen kann), sowie all den weiteren Kolleginnen und Kollegen, die mich auf diesem Weg begleitet haben. Besonders danke ich meinen lieben Kolleginnen Susanne, Anne, Franzi und Sara, die mich immer unterstützt und motiviert haben und im Laufe der Jahre richtig gute Freundinnen geworden sind.

Auch meiner Familie und meinen Freunden, insbesondere meinen Eltern, die mich jederzeit bestmöglich unterstützt haben, möchte ich meinen tiefsten Dank aussprechen.

# INHALT

Zusammenfassung .....	1
Danksagung .....	3
1. Einleitung .....	8
1.1. Forderung nach Evidenzbasierung in der Lehrerbildung .....	8
1.2. Forschungsrahmen .....	10
1.2.1. Anknüpfung an aktuelle Forschung .....	10
1.2.2. Einbettung in das Projekt E4teach und Ziele der vorliegenden Arbeit .....	11
1.3. Umfang und Inhalte .....	12
2. Entstehung und Hintergründe evidenzbasierter Praxis .....	17
2.1. Begriffsdefinition Evidenz und Evidenzbasierung .....	17
2.2. Evidenzbasierte Medizin als Vorreiter der evidenzbasierten Praxis .....	18
2.3. Kernkompetenzen für evidenzbasierte Praxis in der Lehrerbildung .....	22
2.3.1. Konzepte zum kompetenten Nutzen von Evidenz .....	22
2.3.2. Einflussfaktoren .....	28
2.4. Evidenzbasierte Praxis in der Lehrerbildung .....	31
2.5. Grenzen und Barrieren von Evidenzbasierung .....	33
2.5.1. Grenzen und Barrieren evidenzbasierter Praxis im Allgemeinen .....	34
2.5.2. Grenzen und Barrieren evidenzbasierter Praxis speziell in der Lehrerbildung und im Lehrertag .....	36
2.6. Evidenzbasierung versus Evidenzorientierung .....	39
2.7. Fazit .....	40
3. Teilkompetenzen zum professionellen Umgang mit Evidenz in Unterricht und Schule ....	42
3.1. Kritisches Hinterfragen von Evidenz .....	42
3.2. Argumentatives Nutzen von Evidenz .....	43
3.2.1. Relevanz evidenzbasierten Argumentierens .....	44
3.2.2. Qualitätsmerkmale evidenzbasierten Argumentierens .....	45
3.2.3. Evidenzbasiertes Argumentieren in Gruppen .....	46
3.2.4. Häufige Schwierigkeiten und Fehler .....	47
3.3. Fazit .....	48
4. Fördern von Kompetenzen zur professionellen Nutzung von Evidenz .....	49
4.1. Ansätze aus anderen Domänen und ihre Übertragbarkeit auf die Lehrerbildung .....	49
4.1.1. Science Education .....	49
4.1.2. Evidenzbasierte Medizin .....	51

4.2. Ansätze aus der internationalen Lehrerbildung zur Förderung der relevanten Kompetenzen zur Nutzung von Evidenz .....	52
4.3. Fokus auf Heuristiken.....	53
4.3.1. Begriffsdefinition Heuristik .....	54
4.3.2. Implementierung von Heuristiken in Kurzzeitinterventionen für Lehramtsstudierende .....	54
4.3.3. Modelle zur kritischen Bewertung von Evidenz.....	54
4.4. Instruktionale Prinzipien .....	59
4.4.1. Problembasiertes und fallbasiertes Lernen .....	59
4.4.2. Kooperatives und dialogisches Lernen.....	59
4.4.3. Kognitive Modellierung durch tutorielle Unterstützung .....	61
4.5. Fazit .....	62
5. Training zum kompetenten Nutzen von Evidenz .....	63
5.1 Aufbau und Inhalt .....	63
5.2. Gruppendiskussionen.....	65
6. Forschungsagenda .....	68
6.1. Forschungsfragen .....	68
6.2. Methodischer Ansatz .....	69
6.2.1. Stichproben und Designs .....	69
6.2.2. Instrumente.....	71
6.3. Relevante Publikationen .....	74
7. Studie I: Pilotstudie zur Überprüfung des entwickelten Trainingsansatzes.....	75
7.1. Forschungsfragen und Hypothesen .....	75
7.1.1. Manifestierung der Heuristiken .....	75
7.1.2. Trainingseffekt .....	75
7.1.3. Konsistenz .....	76
7.1.4. Akzeptanz und Motivation .....	76
7.2. Methoden.....	77
7.2.1. Stichprobe und Design .....	77
7.2.2. Instrumente.....	79
7.2.3. Statistische Analysen.....	81
7.3. Ergebnisse .....	82
7.3.1. Forschungsfragen 1 und 2: Manifestierung der Heuristiken und Effektivität des Trainings .....	82
7.3.2. Forschungsfrage 3: Konsistenz der Ergebnisse .....	86
7.3.3. Forschungsfrage 4: Akzeptanz und Motivation .....	87

7.4. Diskussion.....	87
8. Studie II: Hauptstudie zur Wirksamkeit des Trainings und möglichen Einflussfaktoren ...	89
8.1. Forschungsfragen und Hypothesen .....	89
8.1.1. Trainingseffekte .....	89
8.1.2. Einflussfaktoren .....	89
8.2. Methoden.....	90
8.2.1. Stichprobe und Design .....	90
8.2.2. Instrumente.....	91
8.2.3. Statistische Analysen.....	93
8.3. Ergebnisse .....	94
8.3.1. Forschungsfrage 1: Effektivität des Trainings .....	94
8.3.2. Forschungsfrage 2: Einflussfaktoren .....	98
8.4. Diskussion.....	100
9. Studie III: Tiefergehende Analyse der im Training videografierten Gruppendiskussionen .....	102
9.1. Forschungsfragen und Hypothesen .....	102
9.1.1. Beispiele Erfolgreichen Nutzens.....	102
9.1.2. Häufige Fehler und Probleme .....	102
9.1.3. Gründe und Funktionen .....	103
9.1.4. Transaktivität .....	103
9.2. Methoden.....	103
9.2.1. Stichprobe und Design .....	103
9.2.2. Instrumente.....	104
9.2.3. Statistische Analysen.....	107
9.3. Ergebnisse .....	107
9.3.1. Forschungsfrage 1: Beispiele erfolgreichen Nutzens .....	107
9.3.2. Forschungsfrage 2: Häufige Probleme und Fehler.....	109
9.3.3. Forschungsfrage 3: Gründe und Funktionen .....	111
9.3.4. Forschungsfrage 4: Zusammenhang Transaktivität und Evidenznutzung.....	113
9.4. Diskussion.....	114
10. Gesamtdiskussion .....	116
10.1. Überblick und Diskussion der zentralen Ergebnisse .....	118
10.2. Einschränkungen und Anknüpfungspunkte .....	123
Literaturverzeichnis.....	127
Abbildungsverzeichnis.....	151

Tabellenverzeichnis.....	151
Anhang .....	153
A. Artikel und Manuskripte .....	153
A.1. Veröffentlichtes Manuskript zu Studie 1 .....	153
A.2. Manuskript zu Studie 2 .....	170
B. Testinstrumente.....	192
B.1. Feedbackheft .....	192
B.2. Fragebogen (prä und post) .....	198
B.4. Kodierschema Studie 1 .....	220
B.5. Kooperationskript.....	222
B.6. Kodierschema Studie 2: schriftliche Statements .....	223
B.7. Kodierschema Studie 2: videografierte Gruppendiskussionen.....	226
B.8. Kodierschema Studie 3 .....	230
C. Trainingsmaterialien: .....	232
C.1. Ablauf .....	232
C.2. Checkliste (Heuristiken) .....	235

# 1. Einleitung

## 1.1. Forderung nach Evidenzbasierung in der Lehrerbildung

Der Begriff der Evidenzbasierung gewinnt in vielen professionellen Tätigkeitsfeldern zunehmend an Bedeutung. Nicht nur in der Medizin, sondern auch in Bereichen wie beispielsweise der Politik, der Managementforschung, den Pflegewissenschaften oder in der Physiotherapie wird immer häufiger gefordert, dass zentrale Entscheidungen auf Basis der aktuellen Studienlage getroffen werden sollen (Wiseman, 2010; Sackett, 1996; Dormann 2010; Kuhn, 2010; Fischer et al., 2015). Diese Forderungen erstrecken sich auch auf das Tätigkeitsfeld von Lehrkräften (BMBF, 2008; KMK, 2004; Prenzel, 2010, Spiel, 2009). So sollen sowohl den Unterricht als auch die Schulentwicklung betreffende Entscheidungen durch aktuelle und qualitativ hochwertige Ergebnisse der (Bildungs-)Forschung gestützt werden. Gründe hierfür sind, dass sowohl der Bereich der Schulentwicklung, als auch der Lehrberuf an sich zunehmend komplexer werdende Professionen sind. Zum einen ändern sich die Anforderungen an die Lehrkräfte, sowie an den Unterricht in zunehmendem Maße. Zum anderen wächst auch das Wissen innerhalb der Forschung stetig an und es entwickeln sich immer neue Richtungen und Blickwinkel, welche es zu überblicken und einzuschätzen gilt.

Häufig wird insbesondere den Unterricht betreffend nach dem Prinzip der *common practice* gehandelt (Davies, 1999; Hargreaves, 2000; Weiß et al., 2014). Lehrerinnen und Lehrer nutzen demnach hauptsächlich jene Methoden und Inhalte zur Gestaltung ihrer Unterrichtseinheiten, die sich im Laufe der Zeit bei ihnen selbst, sowie bei erfahrenen Kolleginnen und Kollegen etabliert und bewährt haben. Die Entscheidungen werden auf diese Art zwar durchaus nach bestem Wissen der einzelnen Lehrpersonen getroffen, aktuelle Entwicklungen aus der Forschung werden hierbei jedoch häufig nicht mit berücksichtigt.

Ziel der Forderung nach Evidenzbasierung ist es, dass Lehrkräfte zur Entscheidungsfindung sowohl ihre Erfahrung auf der einen, als auch die beste verfügbare Evidenz auf der anderen Seite nutzen. Sich lediglich auf eine dieser beiden Komponenten zu stützen ist für verantwortungsbewusstes und qualitativ hochwertiges Handeln nicht ausreichend (Sackett, 1996; KMK, 2004). Dabei soll die evidenzbasierte Praxis Lehrerinnen und Lehrer in ihrer Tätigkeit unterstützen und keine bloße zusätzliche Arbeitsbelastung darstellen. Es wird deshalb zunehmend wichtig, möglichst zeiteffizient und routiniert zu erkennen, welche Quellen für die Beantwortung einer speziellen Fragestellung geeignet sind. Hierfür ist es essentiell einerseits einschätzen zu können wie valide und belastbar (und somit praktisch

anwendbar) diese Evidenzquellen sind, sowie andererseits zu wissen, wie man diese Ergebnisse in der Praxis (argumentativ) nutzen kann. Diese Kompetenzen können dabei nicht nur für inhaltliche und fachliche Herausforderungen von Vorteil für Lehrkräfte sein, sondern auch eine Hilfestellung bezüglich persönlicher Fragen, wie etwa die Lehrergesundheit betreffend, sein.

Wenngleich die evidenzbasierte Praxis im Lehrberuf viele gewinnbringende Vorteile zu haben scheint, gibt es jedoch auch Grenzen und Einschränkungen, sowie vorherrschende Vorbehalte und Ressentiments, vor allem auf Seiten der Praktiker, die es zu erschließen gilt. Ein häufig genannter Punkt hierbei ist, dass durch die Fokussierung auf Evidenzbasierung der Unterricht als rein technischer Vorgang angesehen wird und dieser nur noch anhand von *best practice* Methoden wie beispielsweise aus den *What Works in Classroom* Rankings von Hattie (2009) heraus konzipiert wird. Dabei werde jedoch die einzigartige Situation einer jeden Unterrichtsstunde, bestehend aus zahlreichen und häufig zudem stark voneinander abweichenden Bedürfnissen jedes Individuums innerhalb der Klasse, den individuellen Bedürfnissen und Kompetenzen der Lehrkraft, sowie externen Umständen und Gegebenheiten, nicht ausreichend berücksichtigt. Zudem sei es kaum möglich, auf Grund der fragilen und konfligierenden Eigenschaften von Forschung daraus eine hundertprozentig eindeutige Aussage oder Handlungsempfehlung erhalten zu können.

Es ist deshalb wichtig einen Ansatz zu finden, der (zukünftigen) Lehrkräften durch das Vermitteln von Wissen und relevanten Kompetenzen den Zugang zu wissenschaftlicher Forschung erleichtert und sie darüber hinaus motiviert, Evidenz in sinnvollem und angemessenem Maße und demnach im Zusammenspiel mit ihrer individuellen Expertise, sowie der Betrachtung der individuellen Situation nutzen zu können. Deshalb ist es essentiell, besonders die Kompetenzen des kritischen Hinterfragens und argumentativen Nutzens wissenschaftlicher Evidenz bereits im Lehramtsstudium bestmöglich zu fördern. Denn zum einen sind diese Kompetenzen im Lehramtsstudium, wie auch in jedem anderen akademischen Studium hilfreich, um sich mit den relevanten Studieninhalten in angemessener Weise auseinandersetzen zu können. Zum anderen können hier hilfreiche Strategien für die sinngemäße Nutzung von Evidenz in der späteren Berufsausübung erlernt werden. Infolgedessen wird im Rahmen der vorliegenden Arbeit ein erster Schritt in diese Richtung angestrebt, in dem untersucht wird, wie man auf Basis dieser Anforderungen ein Training zum kompetenten Nutzen von Evidenz für Lehramtsstudierende konzipieren kann.

## 1.2. Forschungsrahmen

### 1.2.1. Anknüpfung an aktuelle Forschung

Wie bereits unter 1.1. aufgezeigt, gewinnt kritisches Bewerten und argumentatives Nutzen wissenschaftlicher Evidenz als Teil der Lehrerbildung zunehmend an Bedeutung (Engelmann et al., 2016; Fischer et al., 2014). Obgleich bereits seit Anfang der 2000er Jahre Forderungen nach eben diesen Kompetenzen, sowohl aus diversen bildungspolitischen Einrichtungen (European Commission, 2013; KMK, 2004; Organisation for Economic Cooperation and Development, 2005; U.S. Department of Education, 2000), als auch im Rahmen von Forschung aus den Bildungswissenschaften (Bauer & Prenzel, 2012; Bromme et al., 2014; Hefter et al., 2014) gestellt werden, begründen Lehrerinnen und Lehrer ihre professionellen Entscheidungen nach wie vor hauptsächlich auf Basis des Prinzips der *common practice* (Davies, 1999; Hargreaves, 2000; Weiß et al., 2014). Darüber hinaus konstatieren sowohl internationale Bildungsforscher (Niemi, 2008; Petty, 2009; Parr & Timperley, 2008;) als auch Bildungspolitiker (KMK, 2004; OECD, 2005; U.S. Department of Education, 2007), dass nicht nur die reine Kompetenz, sondern auch eine positive Einstellung gegenüber dem Nutzen von wissenschaftlicher Evidenz eine wichtige Voraussetzung zur Ausübung evidenzbasierter Praxis sind (Hefter et al., 2014). Es ist jedoch zu vermuten, dass bisher unter Lehrkräften mehrheitlich negative Einstellungen gegenüber der Nutzung von Evidenz im professionellen Kontext herrschen (Hargreaves, 2000; Schildkamp & Kuiper, 2010). Sowohl die fehlenden Kompetenzen, als auch die mangelnde Bereitschaft können darin begründet sein, dass darauf in der Lehrerbildung trotz oben genannter Forderungen (Bauer et al., 2012; Weber & Achtenhagen, 2008) nach wie vor zu wenig bis gar kein Fokus gelegt wird (Bauer et al., 2012; Lohmann et al., 2011). Bislang existieren diesbezüglich nur wenige Interventionen, Trainingsprogramme und Curricula für Lehramtsstudierende im internationalen Raum (z.B. Csanadi et al., 2015; Erduran & Garcia-Mila, 2016; Iordanou & Constantinou, 2014). Diese ersten Maßnahmen sind jedoch noch nicht ausreichend, um Lehramtsstudierenden das Prinzip der evidenzbasierten Praxis näher zu bringen, da sie häufig nur sehr kurz (von einer Dauer von ca. 2 Stunden) sind (z.B. Csanadi, et al., 2015), oder sich lediglich auf Themen der *Science Education* und nicht etwa der Bildungsforschung fokussieren (z.B. Iordanou et al., 2014). Bislang gab es kein Training mit ausreichender Intensität, welches auf den Nutzen von Evidenz im professionellen bildungswissenschaftlichen Kontext abzielt.

### **1.2.2. Einbettung in das Projekt E4teach und Ziele der vorliegenden Arbeit**

Vor diesem Hintergrund entstand die vorliegende Arbeit aus dem durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen eines Schwerpunktprogramms (SPP 1409 – Science and the public), geförderten Projektes *E4teach – Evidence for teachers. Modellierung und Training der Kompetenz angehender und aktiver Lehrkräfte zum Umgang mit Evidenz*, welches unter der Leitung von Prof. Dr. phil. Dr. h.c. Manfred Prenzel und Prof. Dr. phil. Johannes Bauer durchgeführt wurde. Ziel dieses Projektes war es, ein Trainingsprogramm sowie ein Messinstrument zur Förderung und Erfassung relevanter Teilkompetenzen für evidenzbasierte Praxis, im Speziellen *Evidenz kritisch bewerten* und *Evidenz (argumentativ) nutzen*, zu erstellen und zu evaluieren.

Ein Instrument zur Messung dieser Kompetenzen wurde im Rahmen einer Dissertation von Susanne Heining er entwickelt (Heining et al., 2016).

Ziel der vorliegenden Arbeit war die theoriebasierte Entwicklung und Evaluierung des Trainingsprogramms innerhalb des Projektkontextes. Dieses Trainingsprogramm ist eine flexibel ins Lehramtsstudium integrierbare Kurzzeitintervention, welches analog zum Messinstrument situationsbasiert und problemorientiert angelegt ist. Der Fokus liegt hierbei auf der Vermittlung von Heuristiken zur kritischen Bewertung und argumentativen Nutzung von Evidenz. Unterstützt wird dies durch Formen kooperativen Lernens sowie schrittweise ausgeblendeter tutorieller Unterstützung. Mit Hilfe der Evaluation des Trainings sollten Informationen dahingehend gewonnen werden, wie man angehende Lehrkräfte bereits im Studium mit dem relevanten „Handwerkszeug“ ausstatten kann um ihnen das Nutzen von Evidenz im späteren beruflichen Alltag zu ermöglichen.

Diese Arbeit beschäftigt sich deshalb schwerpunktmäßig der Frage, wie ein Training für Lehramtsstudierende konzipiert werden kann, das in das bestehende Lehramtscurriculum integrierbar ist und die Kompetenzen des kritischen Hinterfragens von Evidenz, sowie die des argumentativen Nutzens signifikant fördern kann. Darüber hinaus werden Prädiktoren und Einflussfaktoren wie die fachliche Ausrichtung, Vorwissen, oder individuelle Überzeugungen und Einstellungen untersucht. Des Weiteren ist es Ziel dieser Arbeit herauszufinden, welche instruktionalen Prinzipien und Inhalte hilfreich sein können, um eben diese Kompetenzen bestmöglich fördern und bei Lehramtsstudierenden darüber hinaus zu einer möglichst hohen Motivation und Akzeptanz führen. Zudem sollten mit Hilfe der Evaluierung des Trainings weitere Informationen darüber generiert werden, wie man

ein Training dieser Art zukünftig noch verbessern kann um ein in das bestehende Lehramtscurriculum integrierbares Element erschaffen zu können.

### 1.3. Umfang und Inhalte

In der vorliegenden Arbeit soll zunächst ein theoretischer Rahmen hinsichtlich der Förderung relevanter Kompetenzen für das professionelle Nutzen wissenschaftlicher Evidenz bei Lehramtsstudierenden geschaffen werden. Hierfür wird in Kapitel 2 ein einleitender Überblick über die Entstehung und Entwicklung des Evidenzbegriffs und der evidenzbasierten Praxis gegeben. Dabei werden die wichtigsten Konzepte und Hintergründe der evidenzbasierten Praxis sowohl domänenübergreifend, als auch domänenspezifisch für die Lehrerbildung geschaffen. Ziel dieses Kapitels ist die theoretische Herleitung eines Bedingungsmodells, in welchem relevante Kompetenzfacetten, sowie mögliche Einflussfaktoren auf eben diese als Basis für die Entwicklung eines entsprechenden Trainingsprogramms beschrieben sind. Zu diesem Zweck werden zunächst die wesentlichen Begrifflichkeiten Evidenz und Evidenzbasierung sowohl allgemein, als auch im speziellen Kontext der Lehrerbildung definiert. Die evidenzbasierte Praxis hat ihren Ursprung in der Medizin. Deshalb gibt es in dieser Domäne bereits eine große Vielfalt an Forschung und theoretischen Ansätzen, insbesondere zur Anwendung evidenzbasierter Praxis im professionellen Umfeld. Dies wird unter anderem durch das Bereitstellen von Ratingschemata wie etwa der *Cochrane Klassifikation* oder dem *GRADE Rankingschema* realisiert. Diese Schemata können dabei helfen ein erstes Verständnis des Konzeptes der evidenzbasierten Praxis zu generieren und als erste Entwicklungsgrundlage dienen. Aufgrund der Unterschiedlichkeit beider Domänen bedarf es jedoch der Einbindung weiterer theoretischer Ansätze, um ein auf die Lehrerbildung zugeschnittenes Bedingungsmodell erarbeiten zu können. Deshalb wurden weitere Konzepte aus der Medizin, sowie aus anderen Bereichen, in welchen das Rezipieren, beziehungsweise das Nutzen wissenschaftlicher Evidenz bereits zentraler Forschungsgegenstand ist, dargestellt und hinsichtlich ihrer Nutzbarkeit und Übertragbarkeit auf die Lehrerbildung diskutiert. Auf Basis dessen wurden die relevanten Kompetenzfacetten zum erfolgreichen Nutzen von Evidenz für Lehramtsstudierende identifiziert. Um das professionelle Nutzen von Evidenz bei Lehramtsstudierenden effektiv zu fördern, ist jedoch nicht nur die Kenntnis über relevante Kompetenzen von Bedeutung, sondern darüber hinaus auch die Einbindung möglicherweise relevanter Einflussfaktoren. Dabei sind sowohl externe (kontextuelle) Faktoren wie Anwendungsbereiche und Merkmale von Evidenz zu

betrachten, als auch interne (persönliche) Faktoren, wie Vorwissen und Überzeugungen der Lehramtsstudierenden. Diese Einflussfaktoren erweitern das angestrebte Bedingungsmodell, welches als theoretische Grundlage für einen Trainingsansatz zur Förderung des kompetenten Nutzens von Evidenz bei Lehramtsstudierenden dienen kann.

Um die Relevanz und Sinnhaftigkeit einer Förderung der dadurch identifizierten Kompetenzfacetten zu verdeutlichen, wurde im Anschluss daran speziell auf die Bedeutung evidenzbasierter Praxis in der Lehrerbildung, sowie im Lehrerberuf eingegangen. Entscheidend ist hierbei vor allem die Rolle der Universitäten, da diese vornehmlich die entsprechenden Fördermaßnahmen durchführen können. In einem kurzen Exkurs wird anschließend auch ein Ausblick dahingehend gegeben, was Vorteile und Möglichkeiten evidenzbasierte Praxis im Unterricht selbst sein können. Das Nutzen von Evidenz in der Praxis bringt jedoch nicht ausschließlich Vorteile mit sich. Aus diesem Grund werden im Anschluss an die Chancen und Möglichkeiten des Förderns evidenzbasierter Praxis in der Lehrerbildung auch die Grenzen und Barrieren eben dieser näher betrachtet. Dabei wird zunächst auf allgemeine und domänenübergreifende Grenzen eingegangen und anschließend diejenigen erläutert, die spezifisch für die Lehrerbildung und den Lehrerberuf relevant sind. Nach Diskussion der Möglichkeiten und Grenzen evidenzbasierter Praxis wird abschließend der sich etablierende Begriff der evidenzorientierten Praxis diskutiert.

In den dieser Arbeit zugrunde liegenden Studien war es Ziel, vor allem die Kompetenzen des kritischen Hinterfragens, sowie des argumentativen Nutzens von Evidenz zu fördern, da frühere Forschung zeigen konnte, dass speziell diese Aspekte Lehramtsstudierenden die größten Schwierigkeiten zu bereiten scheinen. Aus diesem Grund wird in Kapitel 3 speziell auf diese beiden Kompetenzen eingegangen und beschrieben, wie diese zu definieren sind, was eine hohe Kompetenz in diesem Bereich beschreibt und wo häufige Schwierigkeiten und Fehler liegen. Mit Hilfe dieser Theorien können zwei aneinander angelehnte Kompetenzstufenmodelle sowohl zum kritischen hinterfragen von Evidenz (*appraise*) als auch zum (argumentativen) Nutzen (*apply*) von Evidenz mit jeweils fünf Kompetenzstufen identifiziert werden. Diese können nicht nur als theoretischer Rahmen für den inhaltlichen Aufbau des Trainings dienen, sondern zudem eine Basis für Kodierschemata zur Messung der Kompetenzen darstellen.

Nachdem in den Kapiteln 2 und 3 die zu fördernden Kompetenzen und Hintergründe identifiziert und diskutiert wurden, soll nun in Kapitel 4 der Fokus auf Möglichkeiten

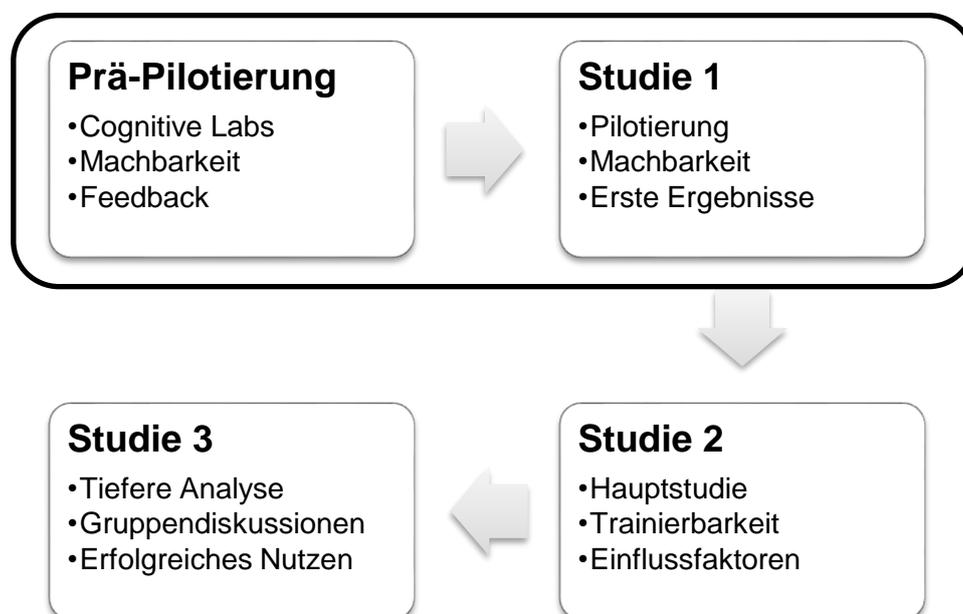
einer effektiven Förderung dieser gelegt werden. Dazu werden zunächst Trainingsansätze aus Domänen, in denen es bereits etablierte Konzepte zur Förderung des kritischen Hinterfragens sowie des (argumentativen) Nutzens von Evidenz gibt erläutert und hinsichtlich ihrer Anwendbarkeit und Übertragbarkeit auf die Lehrerbildung diskutiert werden. So konnten beispielsweise Trainings aus dem Bereich der Science Education zeigen, dass das Nutzen wissenschaftlicher Studien beim Argumentieren durch das Einsetzen von Diskussionen, welche durch Kooperationskripts unterstützt werden, signifikant verbessert werden kann. Trainings aus dem Bereich der evidenzbasierten Medizin konnten zeigen, dass der Einsatz von fallbasiertem Lernen mit abnehmender tutorieller Unterstützung förderlich sein kann beim Erlernen komplexer Strategien wie etwa dem kritischen Hinterfragen von Evidenz. Auch in der internationalen Lehrerbildungsforschung gibt es erste Trainingsansätze, welche ebenso Hinweise dahingehend liefern, dass sich die oben genannten Methodiken positiv auf die Kompetenz bei Lehramtsstudierenden zum Nutzen von Evidenz auswirken können.

Eine weitere bewährte Möglichkeit um komplexe Inhalte möglichst zeiteffizient, effektiv und praktisch nutzbar zu vermitteln ist das Verwenden von Heuristiken. Diese können für Lehramtsstudierende als Daumenregeln fungieren, welche es ihnen ermöglichen können ein wenngleich nicht perfektes, aber dennoch situationsangemessenes Ergebnis zu erzielen. Zugleich können auf diese Weise solch umfangreiche Themen flexibel in das bereits sehr umfassende Lehramtscurriculum integriert werden. Aus diesem Grund wurden bereits existierende Modelle zum kritischen Hinterfragen, sowie zum argumentativen Nutzen von Evidenz dargestellt und hinsichtlich ihrer Nutzbarkeit im Kontext der Lehrerbildung diskutiert.

Im Anschluss daran wurden die als förderlich identifizierten instruktionalen Prinzipien des problem- und fallbasierten Lernens, des kooperativen und dialogischen Lernens, sowie der kognitiven Modellierung durch tutorielle Unterstützung detaillierter beschrieben und ihre möglichen positiven Auswirkungen auf ein Training für Lehramtsstudierende diskutiert. Auf diese Weise konnte eine theoretische Grundlage geschaffen werden, um ein Trainingsprogramm zu konzipieren, welches sich sowohl auf die Motivation als auch auf die Kompetenz Lehramtsstudierender zum Nutzen von Evidenz positiv auswirken kann.

Auf Basis dieser Informationen wurde ein zweitägiges Trainingsprogramm zur Förderung dieser beiden Kompetenzen (*appraise* und *apply*) entwickelt. Dieses wurde zunächst in einer Prä-Pilotierung mit Hilfe von *Cognitive Labs* und lautem

Denken evaluiert um die allgemeine Verständlichkeit zu überprüfen und qualitatives Feedback zur Verbesserung des Trainings zu erhalten (siehe Abbildung 1). Mit Hilfe dessen wurden einzelne Aspekte des Trainings angepasst und eine erste Pilotstudie durchgeführt. Die Prä-Pilotierung, sowie die Pilotierung des Trainings werden zusammengefasst als *Studie 1* dargestellt. Die Ergebnisse der Pilotierung zeigten bereits deutliche Effekte und lieferten somit erste Hinweise bezüglich der Trainierbarkeit des kompetenten Nutzens von Evidenz im Rahmen eines zweitägigen Kurzzeittrainings für Lehramtsstudierende. Zudem konnten hilfreiche Erkenntnisse bezüglich Motivation und Akzeptanz des Trainings generiert werden. Auf Basis dieser Erkenntnisse wurden entsprechende Anpassungen an dem Training, sowie an den Kodierschemata zur Auswertung der Kompetenztests durchgeführt und die Haupterhebung (*Studie 2*) eingeleitet. Diese konnte den positiven Effekt des Trainings auf die Kompetenzen des kritischen Hinterfragens und argumentativen Nutzens von Evidenz bestätigen. Um tiefergehende Einblicke dahingehend zu generieren, wie sich gutes argumentatives Nutzen von Evidenz im Detail äußern kann und wo im Besonderen die Schwierigkeiten und Probleme der Studierenden im Umgang mit Evidenz liegen, wurde im Anschluss daran ein Subsample aus den Teilnehmenden der Hauptstudie gezogen und speziell die videografierten Gruppendiskussionen dieser dahingehend detaillierter analysiert (*Studie 3*).



*Abbildung 1.* Überblick über die Zusammenhänge und Inhalte der drei Einzelstudien im Rahmen dieser Arbeit.

Ziel war es hierbei auch Informationen darüber zu generieren, welche Aspekte innerhalb des Trainings möglicherweise noch unterrepräsentiert sind,

beziehungsweise verbessert oder angepasst werden können, um Lehramtsstudierende mit den bestmöglichen Voraussetzungen auszustatten, um ein kompetentes und situationsangemessenes Nutzen wissenschaftlicher Evidenz im späteren Lehrberuf zu ermöglichen.

## 2. Entstehung und Hintergründe evidenzbasierter Praxis

In diesem Kapitel soll zunächst dargestellt werden, wie die Begriffe *Evidenz* und *Evidenzbasierung* im Allgemeinen, sowie in diesem speziellen Kontext zu definieren sind. Darauf aufbauend wird intensiver auf die Entwicklung und Entstehung der evidenzbasierten Praxis, welche ihren Ursprung im Bereich der Medizin hat, eingegangen. Anschließend werden verschiedene Konzepte aus Bereichen, in denen evidenzbasierte Praxis beziehungsweise das Rezipieren und Nutzen wissenschaftlicher Evidenz stark etabliert und erforscht ist, erläutert und hinsichtlich ihrer Übertragbarkeit auf die Lehrerbildung diskutiert. Ziel ist es hierbei relevante Kompetenzen zu identifizieren, die es im Rahmen des Lehramtsstudiums zu fördern gilt. Darüber hinaus wird durch das Einbeziehen externer sowie interner Einflussfaktoren ein Bedingungsmodell erarbeitet, welches als Grundlage für Förderansätze zum kompetenten Nutzen von Evidenz dienen kann. Im Anschluss daran wird auf Möglichkeiten und Vorteile, sowie auf Einschränkungen und Grenzen des Nutzens wissenschaftlicher Evidenz in der Praxis, sowohl Domänenübergreifend, als auch im Speziellen in der Lehrerbildung eingegangen. Abschließend wird auf Basis dessen die Begrifflichkeit der *evidenzbasierten Praxis* vergleichend zur aktuellen Entwicklung der *evidenzorientierten Praxis* diskutiert.

### 2.1. Begriffsdefinition Evidenz und Evidenzbasierung

Begriffsdefinitionen des Terminus *Evidenz* können oftmals irreführend sein, da es sich hierbei um ein so genanntes *Homonym* (einen Begriff mit mehreren Bedeutungen) handelt. Diese Mehrdeutigkeit resultiert aus der synonymen Verwendung des englischen Wortes *evidence* und dem deutschen Begriff *Evidenz*. Im Deutschen versteht man unter dem Begriff Evidenz die „*vollständige, unmittelbare Einsichtigkeit, Deutlichkeit, [und] Gewissheit*“ (Wermke et al., 2009) und auch im umgangssprachlichen Gebrauch wird der Begriff *evident* häufig verwendet um auszudrücken, dass etwas „*offenkundig und klar ersichtlich ist*“ (ebd.). Im Englischen jedoch meint der Begriff *evidence* „*available facts, circumstances, etc. indicating whether or not a thing is true or valid*“ (Oxford Dictionary of Current English, 1992) und somit keineswegs Eindeutigkeit und Offenkundigkeit. Dies kann zu falschen Erwartungen und Einstellungen bezüglich der Stabilität und Belastbarkeit von Evidenz führen (Raspe, 1999). In der evidenzbasierten Medizin, der Domäne in der die evidenzbasierte Praxis ihren Ursprung hat, wird der Begriff *Evidenz* als Ansammlung aktueller

Forschungsergebnisse aus den besten verfügbaren und zugleich hochgradig validen Studien definiert (DNebM, 2009; Sackett et al., 1996). Der Begriff *Evidenzbasierung* ist als das Stützen auf eben diese beim Treffen wichtiger Entscheidungen zu verstehen (Sackett et al., 1996; Scherfer & Bossmann, 2011). Übertragen auf die Bildungswissenschaften bedeutet *Evidenzbasierung* somit, dass (bildungs-)wissenschaftliche Befunde, welche durch Untersuchungen auf Basis empirischer Forschungsmethoden gewonnen werden, zur Entscheidungsfindung bei komplexen Problemstellungen in der Praxis herangezogen werden (Schneider et al., Shavelson & Towne, 2002; WWC, 2008).

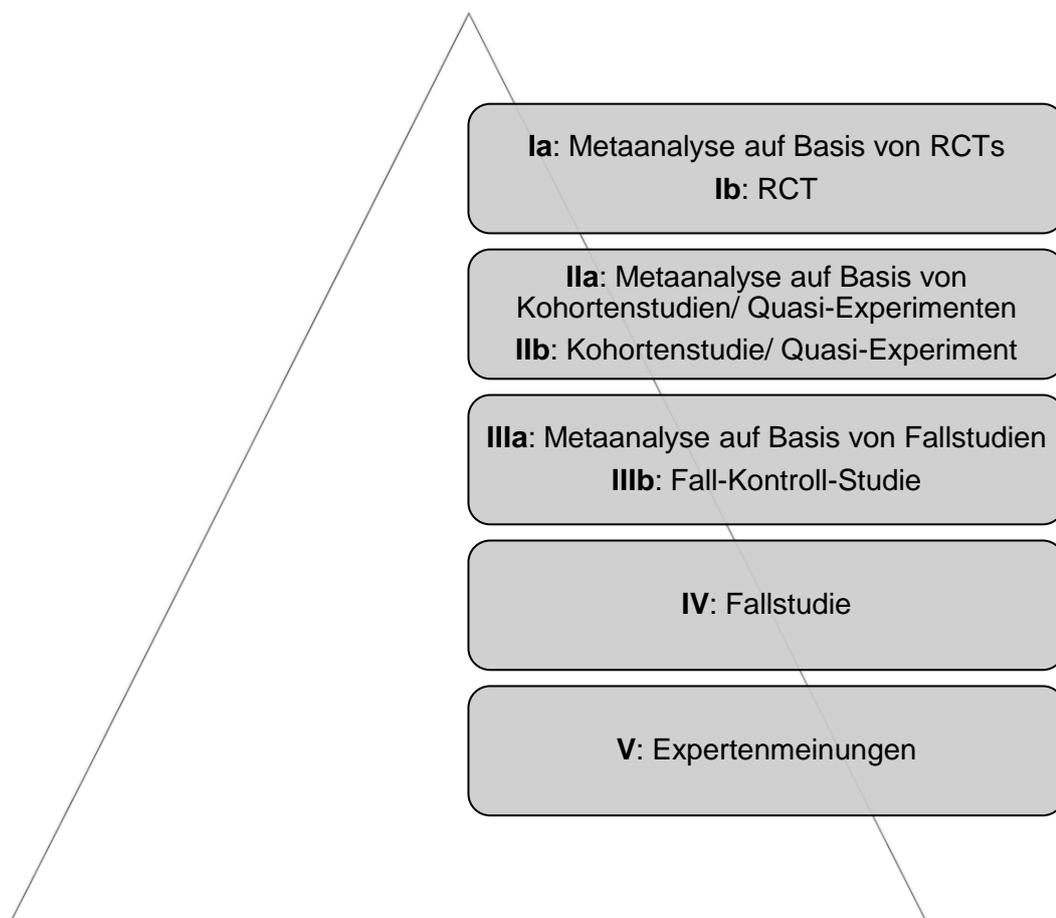
Klar abzugrenzen ist das kompetente Nutzen von Evidenz jedoch von der im Alltag häufig stattfindenden Verwendung *anekdotischer Evidenz* zur Unterstützung einer Behauptung. Diese basiert auf subjektiven Wahrnehmungen und eigenen Erfahrungen und ist deshalb oftmals verzerrt, da sie nicht wissenschaftlich fundiert und überprüft ist und sich lediglich auf das eigene soziale Umfeld und die eigenen Überzeugungen bezieht (Döring & Bortz, 2016). Ein Beispiel für anekdotische und somit nicht-wissenschaftliche Evidenz wäre zu sagen „Das weiß man doch, dass sich die Amerikaner ungesünder ernähren als die Europäer, weil sie viel mehr Fast Food und vor allem Burger essen, das habe ich auch ganz häufig gesehen als ich in den USA im Urlaub war“.

Kompetentes Nutzen von Evidenz im professionellen Kontext bezieht sich in diesem Zusammenhang hauptsächlich auf das kritische Hinterfragen relevanter, wissenschaftlicher Studien und dem situationsangemessenen (argumentativen) Nutzen. Ziel des kompetenten Nutzens von Evidenz ist es demnach nicht, strikte Handlungsvorschriften aus dieser wissenschaftlichen Literatur zu befolgen. Evidenz soll lediglich eine Entscheidungsgrundlage bilden, auf welcher, unter Berücksichtigung der individuellen Situation, sowie der Expertise des Handelnden, eine fundierte Entscheidung getroffen werden kann (siehe auch Kapitel 2.2.) (Hamer & Collinson, 2005; Petty, 2009; Sackett et al., 1996).

## **2.2. Evidenzbasierte Medizin als Vorreiter der evidenzbasierten Praxis**

Nach der Einordnung und Definition des Evidenzbegriffs, soll im Folgenden genauer auf die Entstehung und Entwicklung der evidenzbasierten Praxis eingegangen werden. Die Forderung nach evidenzbasierter Praxis hat, wie bereits in Kapitel 2.1. angedeutet, seinen Ursprung in der Medizin und wurde maßgeblich von David Sackett und Gordon Guyatt entwickelt (Scherfer & Bossmann, 2002; Sackett et al.,

2003). Sie prägten im Rahmen ihrer Forschung seit Anfang der 1990er Jahre den Begriff der evidenzbasierten Medizin (EBM) (Sackett et al., 2003) und definierten ihn als den „*gewissenhafte[n], ausdrückliche[n] und vernünftige[n] Gebrauch der gegenwärtig besten externen, wissenschaftlichen Evidenz für Entscheidungen in der medizinischen Versorgung individueller Patienten*“ (Sackett et al., 1997). Ziel ist es, diejenigen Befunde zur Entscheidungsfindung heranzuziehen, welche forschungsmethodisch die höchstmöglichen Standards erfüllen und somit bestmöglich den Testgütekriterien entsprechen (Atkins et al., 2004; Guyatt et al., 2008; Sackett et al., 1997). In der evidenzbasierten Medizin gibt es diverse Ratingschemata zur kritischen Bewertung der Qualität einer Studie. Die etabliertesten Systeme sind der *GRADE* Ansatz (Grading of Recommendations Assessment, Development and Evaluation), sowie die *Cochrane Klassifikation*. Bei der *Cochrane Klassifikation* werden die Befunde aus Studien in fünf Evidenzklassen, absteigend nach ihrer Qualität, einteilt (siehe Abb. 2) (Hamer & Collison, 2005; Straus et al., 2011; Wirtz et al., 2017).



**Abbildung 2.** Klassen I – V zur kritischen Bewertung von Evidenz aus (klinischen) Studien nach der Cochrane Klassifikation.

Systematische Reviews und Metaanalysen stellen dabei in den Klassen eins bis drei das Ideal dar, jedoch wird auch Einzelstudien eine hohe Bedeutung zugemessen. Am höchsten eingestuft (*Evidenzklasse I*) werden hierbei randomisierte Kontrollstudien (RCTs), welche häufig auch als *Goldstandard* bezeichnet werden (Cordray & Morphy, 2009; Davies, 1999; Straus et al., 2011). Eine Stufe darunter (*Evidenzklasse II*) werden Kohortenstudien und Studien mit quasi-experimentellen Designs eingeordnet. *Evidenzklasse III* beschreibt methodisch hochwertige und valide Fall-Kontroll-Studien und Korrelationsstudien. Darunter (*Evidenzklasse IV*) werden Fall- und Kohortenstudien niedriger Qualität und Belastbarkeit eingestuft. Auf niedrigster Stufe (*Evidenzklasse V*) stehen Einzelfallserien und Expertenmeinungen. Quellen dieser Stufe gilt es jedoch auf Grund der geringen Reliabilität und der damit einhergehenden schwierigen Unterscheidbarkeit zur rein persönlichen Erfahrung zu vermeiden (Straus et al., 2011; Wirtz et al., 2017). Diese Kategorisierung in fünf Evidenzklassen ist zugleich Basis der in der Medizin vielfach gelehrt und weit verbreiteten *Evidenzpyramide* zur kritischen Einschätzung von Studien (Borgetto et al., 2016; Cochrane Collaboration, 2017).

Beim *GRADE* Ansatz wird die Evidenz mit Hilfe diverser Kriterien bewertet (siehe Tabelle 1) und je nach Qualität für jedes einzelne Kriterium auf- beziehungsweise abgewertet. Gründe für eine Abstufung sind Fehleranfälligkeit von Studien (zum Beispiel wenn es sich um eine Fallstudie handelt und nicht um eine RCT), Inkonsistenz innerhalb der Ergebnisse, eine unpräzise Datenlage, sowie eine geringe Generalisierbarkeit und Vergleichbarkeit (Alonso-Coello et al., 2016; Cochrane Collaboration, 2017; WHO, 2010). Ziel dieser Ratingsysteme ist es, die kritische Einschätzung der Evidenz für die praktizierenden Ärztinnen und Ärzte zu erleichtern und somit das evidenzbasierte Entscheiden und Handeln in der Praxis zu ermöglichen. Darüber hinaus werden auf Basis dessen Leitlinien für die Praxis zu entwickelt. Zu eben diesem Zweck haben sich in der Medizin bereits unterstützende Konzepte etabliert. Zum einen gibt es zahlreiche Fortbildungsangebote und Workshops, bei denen die evidenzbasierte Praxis erlernt und trainiert werden kann. Zum anderen werden im Rahmen von *Journal Clubs* nach dem Rotationsprinzip Fachartikel von Spezialisten zusammengefasst, und somit für den entsprechenden Fachbereich aufbereitet und zugänglich gemacht. Ähnlich verfahren wird in den *Medical Grand Rounds* (klinischen Konferenzen, z.B. Cochrane Collaboration, Leitlinien des Deutschen Netzwerks für evidenzbasierte Medizin). Hierbei werden die neuesten Forschungsergebnisse diskutiert und anschließend darauf basierende Leitlinien für die Praktiker veröffentlicht. Inzwischen gibt es hierfür auch feste Einrichtungen (sogenannte *Clearinghouses*), die ausschließlich daran arbeiten die

neuesten wissenschaftlichen Studien einzuschätzen, aufzubereiten und darauf basierende Empfehlungen für die Praxis zu geben (z.B. Cochrane CUE Clearinghouse, What Works Clearinghouse) (Cochrane Collaboration, 2017; DNEbM, 2008; Sackett et al., 1997; WWC, 2008).

Tabelle 1: Ranking für Evidenzquellen nach GRADE.

Qualität der Evidenzquelle	Forschungsdesign	Abwerten:	Aufwerten:
<b>Hoch (4)</b>	RCT oder valide Genauigkeitsstudie für diagnostische Tests	<b>Studienqualität:</b> -1 schwerwiegende Limitationen -2 sehr schwerwiegende Limitationen	<b>Starke Verbindung:</b> +1 Stark, keine plausiblen Störfaktoren, konsistente und direkte Hinweise
<b>Moderat (3)</b>		-1 Inkonsistenz	
<b>Niedrig (2)</b>	Beobachtungsstudie oder indirekte Genauigkeitsstudie für diagnostische Tests	<b>Geradlinigkeit:</b> -1 einige Unsicherheiten -2 große Unsicherheiten	+2 Sehr stark, keine bedeutenden Einschränkungen der Validität und direkte Hinweise
<b>Sehr niedrig (1)</b>		-1 spärliche oder ungenaue Daten  -1 Hohe Wahrscheinlichkeit fehlerhaften Berichts	+1 Hinweise auf einen Dosis-Wirkungs-Gradienten

Ziel all dieser Ansätze ist es jedoch keineswegs Praktiker anzuleiten, sich ausschließlich auf die aktuelle Studienlage zu stützen. Das Konzept der evidenzbasierten Medizin basiert auf drei Säulen: (1) die beste verfügbare externe Evidenz, (2) die interne Evidenz und (3) die Wünsche bzw. Präferenzen des Patienten oder der Patientin (Antes et al., 2003; Greenalgh, 2003; Sackett et al., 1997). *Externe Evidenz* bedeutet in diesem Zusammenhang die aktuellsten und besten verfügbaren Forschungsergebnisse, wohingegen die individuelle Erfahrung der Experten, sowie auch die durch Peers erlernten Strategien, als *interne Evidenz* bezeichnet wird (Sackett et al., 1997). Das Handeln und Argumentieren, welches ausschließlich auf Basis interner Evidenz stattfindet, wird häufig auch als *Eminenzbasierung* bezeichnet (Schmacke, 2004; Wirtz et al., 2017).

Sowohl die Cochrane Evidenzklassen als auch das GRADE Ranking Schema sind auf Heuristiken basierende Konzepte, die von praktizierenden Ärzten (und somit wissenschaftlichen Laien (Schildkamp & Kuiper, 2010)) angewendet werden können um möglichst simpel und zeiteffizient Aussagen hinsichtlich der Belastbarkeit und somit der praktischen Verwendbarkeit einer Studie treffen zu können. Auf Grund der starken Fokussierung auf RCTs können diese Rankings zwar nicht eins zu eins auf die Lehrerbildung übertragen werden. Jedoch können sie als Anregungen dienen, um die relevanten Kompetenzen zum Nutzen von Evidenz bei (zukünftigen) Lehrerinnen und Lehrern zu fördern.

### **2.3. Kernkompetenzen für evidenzbasierte Praxis in der Lehrerbildung**

Nachdem in Kapitel 2.2. erläutert wurde, was hinter dem Begriff der evidenzbasierten Praxis steht und woraus sich dieser entwickelt hat, soll nun genauer darauf eingegangen werden welche Kompetenzen und Teilkompetenzen dafür essentiell sind, sowie welche Ansätze zur Förderung eben dieser herangezogen werden können. Hierfür kann auf etablierte Modelle aus der Medizin und anderen Domänen zurückgegriffen werden, welche auf den Kontext Lehrerbildung angepasst und adaptiert werden können. Darüber hinaus werden diverse Kontextfaktoren betrachtet, welche möglicherweise einen Einfluss auf die Kompetenzentwicklung haben können.

#### **2.3.1. Konzepte zum kompetenten Nutzen von Evidenz**

Im Folgenden werden zunächst drei etablierte Konzepte und Methoden zur Rezeption von Evidenz sowie zum kompetenten Nutzen von Evidenzquellen aus den Bereichen, in denen bereits einschlägige Forschung zum Rezipieren oder Nutzen wissenschaftlicher Evidenz besteht erläutert: der Science Education, den Sozial- und Geisteswissenschaften, sowie der Medizin. Ziel ist es hierbei einen Überblick über bereits erfolgreich umgesetzte Strategien zu generieren und somit diejenige zu identifizieren, welche für ein Kurzzeittraining mit Lehramtsstudierenden am besten geeignet ist. Dadurch soll erschlossen werden, wie möglichst effizient und effektiv Voraussetzungen zur Förderungen der relevanten Kompetenzen in der Lehrerbildung geschaffen werden können.

#### *Kompetenzfacetten aus der evidenzbasierten Medizin*

In der evidenzbasierten Medizin haben sich bereits Handlungsmodelle und Konzepte zum kompetenten Umgang mit Evidenz etabliert (Finkel et al., 2003; Fritsche et al., 2002; Malhotra et al., 2015; Sackett et al., 1996). Diese basieren

vorrangig auf fünf Teilkompetenzen: (1) die Formulierung einer durch Evidenz beantwortbaren Fragestellung (*ask*), (2) die darauf aufbauende Literaturrecherche nach der besten und aktuellsten verfügbaren Evidenz (*acquire*), (3) die kritische Bewertung und Prüfung hinsichtlich Belastbarkeit und Anwendbarkeit der Studienergebnisse (*appraise*), (4) das (argumentative) Nutzen der Evidenz als Grundlage für Entscheidungen (*apply*), sowie (5) die Evaluation der Entscheidungen, die auf Basis der Evidenz getroffen wurden (*evaluate*) (Ramos et al., 2003; Rubin, 2008; Shaneyfelt et al., 2006) (vgl. Abbildung 3, mittlere Spalte).

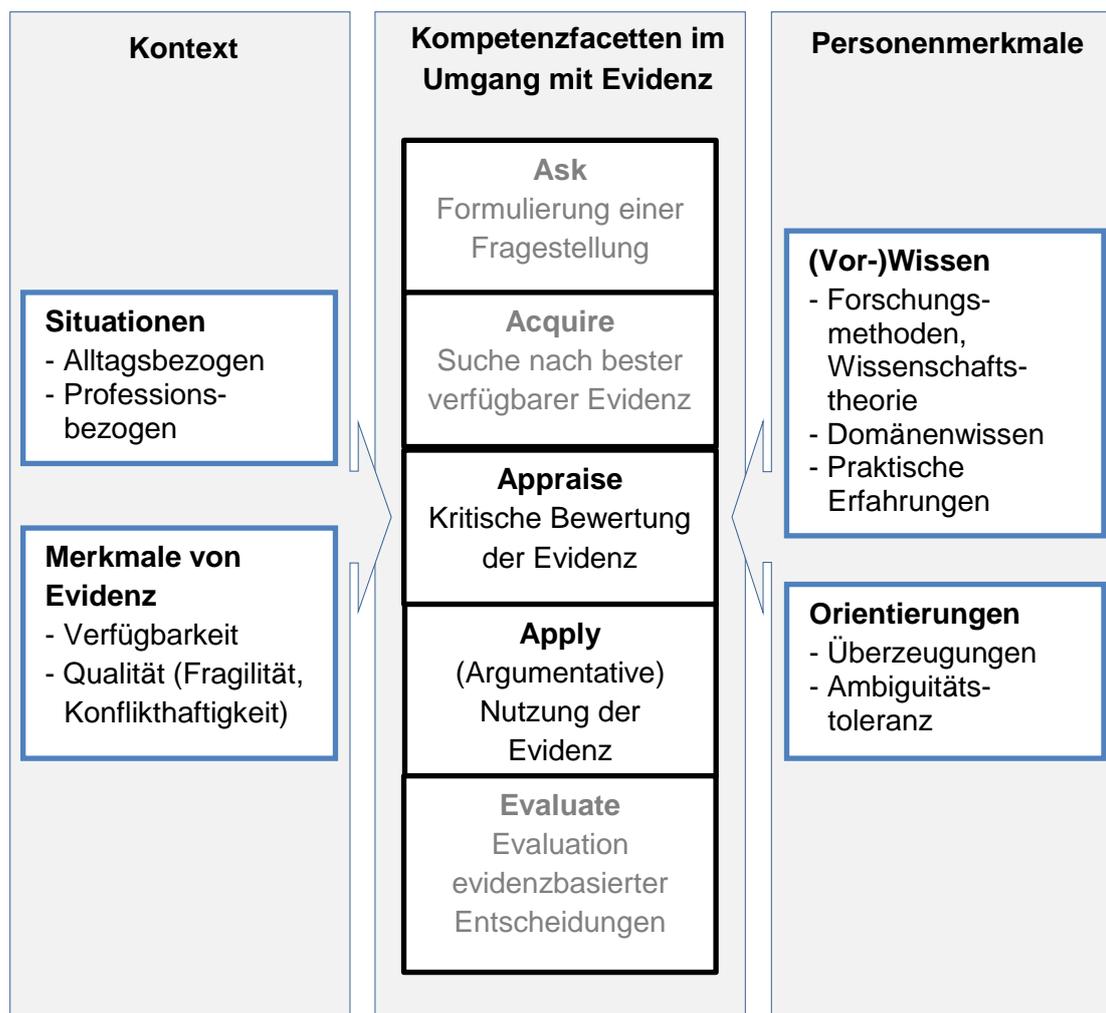


Abbildung 3: Bedingungsmodell zum professionellen Umgang mit Evidenz bei Lehramtsstudierenden.

Um zeiteffizient relevante Inhalte zu eruieren ist es diesem Schema zufolge wichtig, zunächst diejenigen Fragen zu formulieren, die man mit Hilfe der Evidenzquellen beantworten möchte (*ask*). In der Medizin wird hierfür häufig das *PICO* Modell als Hilfestellung zur Generierung recherchierbarer Fragen angewandt. *PICO* steht für *Population, Intervention, Control* und *Outcome*. *Population* bedeutet in diesem Fall

das Formulieren der konkreten Zielgruppe, über die eine Aussage getroffen werden soll. Mit Hilfe dessen kann konkret nach Studien gesucht werden, die eben diese Personengruppe in ihrer Stichprobe erhoben haben. *Intervention* beschreibt den Untersuchungsgegenstand nach dem gesucht werden soll, beispielsweise eine spezielle Technologie, ein bestimmtes Training oder ein spezielles diagnostisches Verfahren. *Control* steht für die Kontrollvariable, beziehungsweise den Vergleichswert an welchem die Intervention gemessen werden soll. *Outcome* beschreibt die Zielgröße und damit das zu messende Ergebnis der Studie. (Hastings & Fisher, 2014; Mamédio da Costa Santos et al., 2007; Miller & Forrest, 2001). Übertragen auf die Bildungswissenschaften könnte eine solche Frage beispielsweise „Wie wirkt sich Teamteaching auf den Unterricht in der Sekundarstufe II aus?“ sein. Die *Population* wären in diesem Fall Schüler der Sekundarstufe II, die *Intervention* wäre der Team teaching Ansatz, die *Control(variable)* der bisherige, traditionelle Unterricht und der *Outcome* könnte beispielsweise die Unterrichtsqualität, das Belastungsempfinden der Lehrkräfte oder die Zufriedenheit und Motivation der Schülerinnen und Schüler sein.

Basierend auf den *PICO* Fragen kann als nächster Schritt die Recherche nach der besten verfügbaren Evidenz zur Beantwortung erfolgen (*acquire*). Mit Hilfe welcher Bibliotheken und elektronischen Datenbanken diese gefunden werden kann unterscheidet sich stark zwischen den Domänen, sowie Standorten (Sackett, 1999; Wirtz et al., 2017). Aus diesem Grund werden oftmals lokale und domänenspezifische Literaturrecherchekurse an den Universitäts- und Fachbibliotheken angeboten, welche die geeignetsten Wege aufzeigen. Da trotz guter Recherche mit zuvor festgelegten Kriterien oftmals eine Vielzahl an Evidenzquellen gefunden wird, müssen diese in einem nächsten Schritt kritisch bewertet werden um diejenigen Quellen herauszufiltern, die die besten, aktuellsten, zur Problemstellung passendsten und zugleich belastbarsten Ergebnisse liefert. Dies kann in der Regel mit Hilfe von zwei Schritten erfolgen. Zunächst werden die Evidenzquellen hinsichtlich ihrer forschungsmethodischen Qualität und Belastbarkeit sortiert (siehe auch Kapitel 2.2.) und anschließend werden diejenigen herausgefiltert, die für das konkrete Problem am geeignetsten anzuwenden, beziehungsweise auf dieses generalisierbar sind (Guyatt et al., 1992; Sackett et al., 1997). Auf Basis dieser Reduzierung und Bewertung kann nun in einem nächsten Schritt entschieden werden, welche Ergebnisse für den vorliegenden individuellen Fall genutzt werden können (*apply*). Dies geschieht im Idealfall durch das Einklang bringen der besten externen Evidenz mit der Expertise des Praktikers und der individuellen Situation (vgl. Kapitel 2.2.) (Guyatt et al., 1992; Sackett et al., 1997;

Witz et al., 2017). Als letzter Schritt werden alle in Schritt eins bis vier getroffenen Entscheidungen nochmals kritisch hinterfragt und bewertet. Diese Selbstbewertung dient dazu, den Prozess des kompetenten Nutzens von Evidenz zu evaluieren und sukzessive zu verbessern und anzupassen (Bellamy et al., 2009; Sackett et al., 1997).

### *Epistemische Aktivitäten*

Um die Prozesse, die für das (argumentative) Nutzen von Evidenz hilfreich sein können besser verstehen zu können ist es hilfreich, ein etabliertes Konzept, welches sich im Bereich der Science Education im Rahmen der Förderung und Evaluation wissenschaftlichen Argumentierens entwickelt hat, zu betrachten. Ausgangspunkt hierfür sind insgesamt acht *epistemic activities*, welche Voraussetzungen für wissenschaftliches Schlussfolgern und Argumentieren (*scientific reasoning and argumentation*, kurz *SRA*) darstellen (Fischer et al., 2014). Die erste epistemische Aktivität und damit ein erster Schritt in Richtung *SRA* ist hierbei die Identifizierung des konkreten Problems (*problem identification*). Darauf folgt, analog zu bestehenden Handlungsmodellen aus der evidenzbasierten Medizin, die Formulierung der Fragestellung (*questioning*). Darauf aufbauend sollen erste Hypothesen bezüglich der Beantwortung der zuvor formulierten Fragestellungen aufgestellt werden (*hypothesis generation*). Mit Hilfe der Hypothesen können nun erste Konstrukte, Prototypen oder Artefakte erstellt werden (*construction and redesign of artefacts*). Demgemäß kann nun Evidenz in Form von beispielsweise experimentellen Studien oder Beobachtungen generiert werden (*evidence generation*). Der nächste Schritt ist darauffolgend diese Evidenz kritisch zu betrachten und hinsichtlich der Tatsache ob und inwieweit sie geeignet ist, um die Fragestellung zu beantworten, beziehungsweise das zu Anfang identifizierte Problem zu lösen, zu bewerten (*evidence evaluation*). Stellt sich die generierte Evidenz als geeignet heraus, können nun auf Basis dessen konkrete Annahmen oder Entscheidungen getroffen werden (*drawing conclusions*). Mit Hilfe von (öffentlichen) Diskussionen mit Experten können diese Annahmen nochmalig überprüft und somit direkt kommuniziert werden (*communicating and scrutinising*) (Fischer et al., 2014; Shavelson & Towne, 2002). Vorzugsweise die Aktivitäten *questioning*, *evidence evaluation* und *drawing conclusions* können nicht nur dabei unterstützen neue Evidenz zu generieren, sondern auch bereits vorhandene Evidenz kompetent zu nutzen und zu hinterfragen.

### *SQ3R Methode*

Ebenso hilfreich für ein theoretisches Verständnis des kompetenten Umgangs mit Evidenz ist es, die bereits im Jahr 1946 erschienene SQ3R Methode von Robinson genauer zu betrachten. SQ3R steht für *Survey* (Überblick), *Question* (Fragestellung), *Read* (Lesen), *Recite* (Wiedergeben) und *Review* (Bewerten) und ist eine Strategie in fünf Schritten, die aus den Sozial- und Geisteswissenschaften stammt und als Unterstützung zum Verständnis komplexer wissenschaftlicher Texte dient. Dieses Modell zur Lesestrategie wissenschaftlicher Texte kann Ansatzpunkte dahingehend liefern, wie eine strukturierte Herangehensweise an wissenschaftliche Texte möglich ist.

Im ersten Schritt (*survey*) geht es zunächst darum einen Überblick über den Text und dessen Inhalt zu bekommen, indem Aspekte wie etwa Autor(en), Titel, Aufbau und Teilüberschriften, Schlagworte, Abstract und Diskussion genauer betrachtet werden. Auf diese Weise entsteht eine erste Idee davon, was die wichtigsten Inhalte und die zentrale(n) Botschaft(en) des Textes sind. Im zweiten Schritt (*question*) soll der Leser selbst Fragen zu den einzelnen Sinnabschnitten des Textes erarbeiten und somit eine Art Gerüst für die wichtigsten Inhalte des Textes erstellen. Mögliche hilfreiche Fragen wären hierbei beispielsweise „Was sind die wichtigsten Fragestellungen und Hypothesen?“, „Was ist die Kernaussage des Textes“, „Welches Forschungsdesign wurde konzipiert um die Forschungsfrage zu beantworten?“. Auf Basis dessen werden im dritten Schritt (*read*) die einzelnen Textabschnitte im Detail gelesen. Wichtige Inhalte innerhalb des Textes können hierfür farbig markiert und den Fragen zugeordnet werden. Nach dem detaillierten Lesen jedes einzelnen Abschnittes soll dieser nochmals rekapituliert werden (*recite*) und so versucht werden, die zuvor aufgestellten Fragen zu beantworten. Dadurch entsteht ein strukturiertes Exzerpt, welches alle wichtigen Inhalte, Antworten und Schlüsselworte beinhaltet. Abschließend wird im letzten Schritt (*review*) der wissenschaftliche Artikel nochmals im Ganzen betrachtet und hinsichtlich Kriterien wie beispielsweise Relevanz, Übertragbarkeit, Nutzbarkeit sowie gelernte Inhalte bewertet. (Döring & Bortz, 2016; Fadness Tadlock, 1978; Mayer, 2005; Robinson, 1961). Auf Basis der SQ3R Methode haben sich im Laufe der Zeit noch weitere Modelle zur Lesestrategie entwickelt, wie zum Beispiel die PQ4R Methode (*Preview, Questions, Read, Reflect, Recite* und *Review*) (Thomas & Robinson, 1972) und das MURDER Schema (*setting the mood, reading for understanding, recalling the material, digest the material, expanding knowledge* und *review*) von Dansereau et al. (1979).

### *Fazit*

Ziel war es, mit Hilfe dieser etablierten Ansätze aus unterschiedlichen Domänen eine Basis für ein Bedingungsmodell zum kompetenten Nutzen von Evidenz bei Lehramtsstudierenden herzuleiten. So kann die SQ3R Methode relevante Einblicke dahingehend liefern, wie komplexe wissenschaftliche Texte rezipiert und nachhaltig verstanden werden können, sowie welche Schritte hierfür nötig sind. Die Verwendung dieser Methodik ist eine hilfreiche Möglichkeit für ein tiefes Elaborieren und Analysieren wissenschaftlicher Texte. Das Fördern des Nutzens von Evidenz bei Lehramtsstudierenden fokussiert sich jedoch auf Strategien, die ein möglichst effektives Rezipieren und Nutzen von Evidenz ermöglichen, um flexibel in den Berufsalltag von Lehrerinnen und Lehrern integrierbar zu sein. Da die SQ3R Methode jedoch sehr zeitintensiv ist, ist diese im Lehrertag nur schwer umsetzbar. Einzelne Teilaspekte, wie beispielsweise das Verschaffen eines Überblicks, das Bewerten des Textes und das (kritische) Hinterfragen des Gelesenen, lassen sich dennoch gut integrieren und können zudem auch für den inhaltlichen Aufbau eines Trainings zur Förderung der relevanten Kompetenzen zum professionellen Nutzen von Evidenz hilfreich sein.

Die epistemischen Aktivitäten beschreiben wichtige Voraussetzungen für das Schlussfolgern und Argumentieren auf Basis von wissenschaftlichen Erkenntnissen. Informationen bezüglich dieser Aktivitäten können wichtige Hinweise dahingehend liefern, welche Prozesse beim erfolgreichen Nutzen von Evidenz relevant sind, sowie welche Kompetenzen hierfür im Speziellen gefördert werden müssen. Besonders die Aktivitäten *questioning* (Formulierung einer Fragestellung), *evidence evaluation* (kritisches Hinterfragen und Bewerten) und *drawing conclusions* (Schlussfolgerungen auf Basis der Evidenz ziehen) können dabei unterstützen ein Konzept zu erstellen um (auch bereits vorhandene) Evidenz kompetent zu hinterfragen und zu nutzen.

Die Kompetenzfacetten zum Nutzen von Evidenz aus dem Bereich der Medizin (*ask, acquire, appraise, apply* und *evaluate*) enthalten viele Aspekte, die auf ähnliche Weise so auch in der SQ3R Methode und den epistemischen Aktivitäten relevant sind. Diese Gemeinsamkeiten finden sich vor allem beim Stellen der richtigen Fragen, dem kritischen Hinterfragen, sowie der (argumentativen) Nutzung der Evidenz. Dies lässt vermuten, dass die im Bereich der Medizin formulierten Kompetenzfacetten auch auf die Bildungswissenschaften generalisierbar sind und demzufolge domänenübergreifend für Lehramtsstudierende einsetzbar sind, auch wenn sich die Textsorten beider Bereiche deutlich unterscheiden (vergleiche Kapitel 2.2.). Demnach können diese Kompetenzfacetten als eine Art „Grundgerüst“ und

somit als zu fördernde (Teil-)Kompetenzen eines Bedingungsmodells für den kompetenten Umgang mit Evidenz bei Lehramtsstudierenden genutzt werden (siehe Abbildung 3, mittlere Spalte).

### 2.3.2. Einflussfaktoren

Neben den relevanten Kompetenzfacetten des erfolgreichen Nutzens wissenschaftlicher Evidenz sind auch mögliche Einflussfaktoren auf eben diese wichtig für die Konzeption eines Bedingungsmodells. Diese Einflussfaktoren können sowohl kontextuellen als auch persönlichen Hintergrunds sein. Im Folgenden werden deshalb beide Bereiche detaillierter ausgeführt.

#### *Kontextuelle Faktoren*

Die kontextuellen, externen Faktoren (siehe Abbildung 3, linke Spalte) beziehen sich auf (a) Anwendungsbereiche (Situationen) und (b) Merkmale von Evidenz. Die Anwendungsbereiche beinhalten zum einen professionelle Situationen und Problemstellungen, die in der Schule auf Lehramtsstudierende zukommen. Zum anderen wird es in Zeiten von „fake news“ und immer provokativer formulierten Zeitungsschlagzeilen auf der einen Seite und des exponentiell wachsenden wissenschaftlichen Wissens auf der anderen Seite, zunehmend wichtiger auch alltagsbezogene Evidenz hinsichtlich ihrer Glaubhaftigkeit und Belastbarkeit kritisch hinterfragen und einschätzen zu können (Bromme, 2008; Kienhues et al., 2011). Dies kann sowohl für die professionelle Vorbildfunktion einer Lehrkraft, als auch für das verantwortungsbewusste Handeln im privaten Alltag relevant sein (Bromme, 2008; Glug, 2009; Sodian & Koerber, 2011; Weinert et al., 2011; Wilson & Sloane, 2000). Ein aktuelles Exempel hierfür wäre beispielsweise die Frage nach den Vorteilen und möglichen Risiken einer Grippeimpfung.

Ob und in wieweit Lehrkräfte in der Lage sind, beziehungsweise dazu motiviert sind Evidenz zu nutzen, hängt zudem maßgeblich von den Merkmalen der jeweiligen Evidenzquelle ab. So verfügen Schulen häufig nicht über ausreichende Zugangsmöglichkeiten zu relevanten Fachbibliotheken und Fachzeitschriften (Reiss et al., 2015), was eine eingeschränkte *Verfügbarkeit* von Evidenz zur Folge hat. Eine weitere mögliche Barriere bezüglich Motivation und Kompetenzentwicklung in der evidenzbasierten Praxis ist die Qualität der Evidenz. Sucht man mit Hilfe von wissenschaftlichen Studien die Antwort auf eine Frage, wird man selten ein hundertprozentig eindeutiges Ergebnis erreichen (Booth et al., 2008). Wissenschaftliche Studien sind oftmals *konfligierend*, also widersprüchlich zu einander. Dies kann unter anderem durch unterschiedliche Studiendesigns und Messmethoden, sowie durch unterschiedliche Zusammensetzungen der

Stichproben begründet sein (Bromme, 2008; Stadler & Bromme, 2007). Evidenz kann zudem in sich *fragil* sein, was bedeutet, dass beispielsweise aufgrund von methodischen Einschränkungen keine ausreichende Generalisierung und somit keine konkrete Aussage bezüglich der Fragestellung getroffen werden kann (Bromme, 2008; Rubin, 2008; Shavelson & Tone, 2008).

### *Persönliche Faktoren*

Einflussfaktoren auf persönlicher Ebene (siehe Abbildung 3, rechte Spalte) sind interne Faktoren, und können durch *Vorwissen* und *Orientierungen* bedingt sein. *Vorwissen* ist ein anerkannter Prädiktor für Weiterentwicklung und Lernerfolg (Gruber & Stamouli, 2009) und bezieht sich hierbei auf Wissen über wissenschaftliche Studien, (empirische) Forschungsmethoden, Statistik und Wissenschaftstheorie, sowie Domänenwissen und praktische Erfahrungen. Vergangene Studien haben gezeigt, dass ein deklaratives Wissen über quantitative Forschungsmethoden und somit ein Grundverständnis von empirischer Forschung eine wichtige Schlüsselfunktion für die evidenzbasierten Praxis darstellt (Davies, 1999; Ramos et al., 2003; Shaneyfelt et al., 2006). Da quantitative Forschungsmethoden hauptsächlich in den Naturwissenschaften verortet sind, ist davon auszugehen, dass ein Grundverständnis von Naturwissenschaften und naturwissenschaftlicher Forschung ebenfalls zu einem besseren Verständnis und einer höheren Akzeptanz von evidenzbasierter Praxis führt (Glug, 2009; Prenzel et al., 2007). Aus diesem Grund kann davon ausgegangen werden, dass *Domänenwissen* ebenfalls ein Einflussfaktor für erfolgreiches Umsetzen und Erlernen evidenzbasierter Praxis sein kann. Domänenwissen steht in diesem Kontext für *prozedurales Wissen* in der Domäne der Naturwissenschaften (Fischer et al., 2014). Besonders in der Lehrerbildung spielt dieser Aspekt aufgrund der Heterogenität in der fachlichen Ausrichtung eine wichtige Rolle. Je nach zu unterrichtender Fächerkombination können (zukünftige) Lehrerinnen und Lehrer einen mathematisch-naturwissenschaftlichen, so genannten *MINT* Hintergrund (Mathematik, Informatik, Naturwissenschaften und Technik) haben, oder einen sprach-, beziehungsweise geisteswissenschaftlichen (*nicht-MINT*) Hintergrund. Demnach haben Lehramtsstudierende möglicherweise ein stark voneinander abweichendes Verständnis und Metawissen von und über Wissenschaft. Folglich ist es anzunehmen, dass Lehramtsstudierende mit mindestens einem *MINT* Fach einen Vorwissensvorteil gegenüber Studierenden haben, die rein geisteswissenschaftlich geprägt sind (Prenzel & Bauer, 2012; Weinert, 2011). Des Weiteren kann auch die bisherige *praktische Erfahrung* ein essentieller

Einflussfaktor auf den kompetenten Umgang mit Evidenz sein. Einerseits kann dies durch die Häufigkeit, mit der ein Studierender bereits mit empirischen Studien aus den Bildungswissenschaften in Kontakt gekommen ist, bedingt sein (Weber & Achtenhagen, 2008). Andererseits durch das im Rahmen des Studiums, beispielsweise in den Fachdidaktiken, erlernte Professionswissen, sowie die aus ersten Unterrichtserfahrungen erlangte Praxiserfahrung (Baumert & Kunter, 2006; Bromme, 1992).

Überdies können sich auch *Orientierungen* auf den Kompetenzzuwachs und die Akzeptanz evidenzbasierter Praxis auswirken. Hierbei sind vor allem Ambiguitätstoleranz, epistemologische Überzeugungen und Überzeugungen zum Nutzen von Evidenz von Relevanz. *Ambiguitätstoleranz* ist per Definition die Fähigkeit Unsicherheit und Mehrdeutigkeit aushalten zu können, beziehungsweise unsichere Situationen auf sich zukommen zu lassen (Dalbert, 1999; Wirtz et al., 2017). Frühere Studien konnten bereits einen Zusammenhang zwischen Ambiguitätstoleranz und der Art und Weise, wie Lehrkräfte mit professionellen Herausforderungen umgehen, aufzeigen (König & Dalbert, 2007). In dem vorliegenden Handlungsmodell ist Ambiguitätstoleranz vor allem im Zusammenhang mit den fragilen und konfligierenden Eigenschaften von Evidenz von Bedeutung. Es kann angenommen werden, dass gerade ambiguitätsintolerante Studierende größere Schwierigkeiten haben sich auf diese unsichere (fragile und konflikthafte) Eigenschaft von Wissenschaft einzulassen (Prenzel & Bauer, 2012).

*Epistemologische Überzeugungen* sind die „*Annahmen einer Person über die Herkunft, Gewissheit, Struktur und Rechtfertigung von Wissen*“ (Wirtz et al., 2017). Es wird davon ausgegangen, dass diese Annahmen beeinflussen, auf welche Art und Weise Menschen neue Informationen aufnehmen und annehmen (Hofer & Pintrich, 2002; Hasanbegovic et al., 2006; Richardson, 1996). Auf Basis dessen ist anzunehmen, dass Studierende, die Wissen in den Bildungswissenschaften als beispielsweise sehr objektiv, eindeutig und stabil einschätzen (Stahl & Bromme, 2007) größere Schwierigkeiten mit Fragilität und Konflikthaftigkeit in bildungswissenschaftlicher Evidenz haben (Prenzel & Bauer, 2012).

Überdies spielen auch die Einstellungen bezüglich evidenzbasierter Praxis eine zentrale Rolle. Zahlreiche Studien konnten zeigen, dass die Annahmen bezüglich der Relevanz und des Nutzens (bildungswissenschaftlicher) Evidenz einen großen Einfluss auf die Akzeptanz und somit auch die Motivation sich näher damit zu beschäftigen hat (Jette et al., 2003; Johnston et al., 2003 Parr & Timperley, 2008). Gerade negative Einstellungen können hier eine bedeutsame Barriere darstellen

(Hetmanek et al., 2015; Hargreaves, 2000; Kühle & Peek, 2007; Parr et al., 2008; Schildkamp et al., 2010; Williams & Coles, 2008; Zembal-Saul, et al., 2003).

### *Fazit*

Um die in Kapitel 2.3.1. identifizierten Kompetenzen effektiv fördern zu können, ist es essentiell zusätzliche Informationen über mögliche Einflussfaktoren auf eben diese mit einzubeziehen. Hierbei können sowohl kontextuelle, als auch persönliche Faktoren von Relevanz sein. Der Grund für das Einbeziehen dieser Einflussfaktoren ist, dass nicht nur das fachliche Wissen, sondern auch die äußeren Umstände, sowie die individuellen Einstellungen und Überzeugungen die Entwicklung der relevanten Kompetenzen und somit das erfolgreiche Nutzen von Evidenz beeinflussen können (siehe auch Kapitel 2.4.). Auf Basis dieser Informationen wurde das Bedingungsmodell (siehe Abbildung 3) zum professionellen und kompetenten Umgang mit Evidenz bei Lehramtsstudierenden entwickelt. Dieses Bedingungsmodell vereint die im Rahmen dieses Kapitels als relevant identifizierten Kompetenzen und Einflussfaktoren. Infolgedessen kann dieses einen theoretischen Rahmen für ein Trainingsprogramm für Lehramtsstudierende zum kompetenten Umgang mit Evidenz bilden, in welchem diese Parameter beachtet und bestmöglich gefördert werden sollen.

## **2.4. Evidenzbasierte Praxis in der Lehrerbildung**

Um die Relevanz, sowie die Gründe für eine notwendige Förderung der in Kapitel 2.3. identifizierten Kompetenzen und Faktoren innerhalb des Bedingungsmodells zum kompetenten Nutzen von Evidenz bei Lehramtsstudierenden weiter zu verdeutlichen, wird im folgenden Kapitel konkret auf deren Bedeutung in der Lehrerverberuf eingegangen. Dabei gilt es zwischen der Verwendung von Evidenz in der Lehrerbildung und Evidenz im Unterricht zu unterscheiden.

### *Lehrerbildung*

Zunächst wird hierfür die Rolle der Universitäten bei der Integration evidenzbasierter Praxis in den Lehrberuf näher betrachtet. Deren Aufgabe ist es zum einen die Lehramtsstudierenden dahingehend zu motivieren, einen Bezug zu wissenschaftlichen Befunden zu bekommen und die Relevanz dessen für den Berufsalltag zu erkennen (Schrader, 2014). Darüber hinaus ist es essentiell, die relevanten Kernkompetenzen zum kompetenten Umgang mit wissenschaftlichen Studien zu vermitteln. Die Wertschätzung wissenschaftlichen Wissens ist auch dabei eine zentrale Voraussetzung für den *theory-to-practice* Transfer, denn ohne

eine positive Einstellung (*will*) wird trotz erlernten Inhalten zum kompetenten Umgang mit Evidenz (*skill*) keine Anwendung in der Praxis stattfinden (Hefter et al., 2014; Schrader, 2014). Demzufolge kann die „*Implementierung von wissenschaftlichen Erkenntnissen in der Praxis [...] nur auf Seiten der Praktiker geleistet werden*“ (Hetmanek et al., 2015).

Zum anderen ist es notwendig, dass die Forschung im Bereich der Bildungswissenschaften sowohl methodisch (Bromme et al., 2016; Prenzel, 2012; Schrader, 2014; Slavin, 2008) als auch inhaltlich (Hargreaves, 2007a, 2007b) erweitert wird. Darüber hinaus gilt es zu eruieren, welche Eigenschaften Evidenzquellen haben müssen, um von Praktikern als relevant und nützlich angesehen zu werden (Fischer & Wecker, 2006). Hierfür ist es erforderlich mehr darüber herauszufinden, wie genau Evidenz in der Praxis genutzt wird. Zwar gaben Lehrkräfte im Rahmen einer Studie von Hetmanek et al (2015) an, Ergebnisse der Bildungswissenschaften häufig als Ressource, vor allem zur Vorbereitung auf den Umgang mit problematischen Lernvoraussetzungen, sowie den Unterrichtsinhalt, zu verwenden, jedoch beinhaltet dies nicht notwendiger Weise auch zugleich ein kritisches Hinterfragen und somit kompetentes Nutzen von Evidenz. Die Förderung dieser relevanten Kompetenzen zum Umgang mit (bildungs-)wissenschaftlichen Studien im Rahmen des Lehramtsstudiums wird sowohl national als auch international von Forschern, wie auch von politischen Einrichtungen gefordert (Bauer & Prenzel, 2012; Dederich, 2009; European Commission, 2013; KMK, 2004; Prenzel, 2013; Schrader, 2014; U.S. Department of Education, 2007). Jedoch gibt es bislang keine systematische Förderung und keine strukturierte Umsetzung im deutschen Lehramtsstudium (Bauer et al., 2012; Bauer, 2010). Auch im internationalen Raum existieren bislang nur wenige Trainings- und Förderungsansätze (siehe Kapitel 4.2.).

Betrachtet man diese Forderungen, lässt sich zusammenfassend daraus schließen, dass neben der Einführung von Trainings zur Förderung und Motivation zur kompetenten Anwendung evidenzbasierter Praxis bei Studierenden sowohl eine stärkere Zusammenarbeit zwischen Forschung und Schule (*boundary crossing*) angestrebt werden muss, als auch die bereits verfügbaren Forschungsergebnisse für die praktizierenden Lehrkräfte besser aufbereitet, angepasst und verfügbar gemacht werden müssen (Hartmann et al., 2016). Ziel in der Lehrerbildung ist es demnach zukünftige Lehrerinnen und Lehrer dahingehend zu unterstützen, die relevanten Voraussetzungen für das kompetente Nutzen wissenschaftlicher Evidenz zu erlernen.

### *Unterricht*

Wie vorausgehend beschrieben, wird sowohl vom deutschen Kultusministerium (KMK, 2004), als auch von der European Commission (2013), sowie vom U.S. Department of Education (2007) im Rahmen des *No Child Left Behind Acts* (2001) gefordert, den Unterricht durch wissenschaftlich gestützte Handlungspraktiken anzureichern, in dem Lehrerinnen und Lehrer aktuelle Befunde der Bildungsforschung kritisch hinterfragen und bei Bedarf in der Praxis nutzen. Es wäre dabei jedoch nicht realistisch zu erwarten, dass Lehrkräfte in ihrem Berufsalltag ausschließlich oder überwiegend evidenzbasiert arbeiten (Haas, 2005; Seel, 1997). Ziel soll es demnach sein, dass Lehrerinnen und Lehrer vor allem bei weitreichenden und wichtigen Entscheidungen (wie zum Beispiel bei der Frage nach der Sinnhaftigkeit von Hausaufgaben, der Einführung von Ganztageschulen oder eines Teamteaching Ansatzes), oder bei speziellen Einzelfallentscheidungen (wie zum Beispiel individuelle Lern- und Leistungsstörungen oder Mobbingfälle), in der Lage sind die beste verfügbare Evidenz zu Rate zu ziehen. Auf diese Weise kann die Nutzung von Evidenz eventuelle Fehlentscheidungen vermeiden (Whitehurst, 2002) und zur Findung einer optimalen Lösung beitragen.

Zudem sollten Lehrerinnen und Lehrer auch in darüber hinaus gehenden Bereichen in der Lage sein, Evidenzquellen kritisch einschätzen zu können (Moore, 2007). Denn innerhalb unserer postmodernen Wissensgesellschaft, sowie eines Zeitalters in dem Begriffe wie *Fake News* und *Alternative Facts* immer größere Bedeutung bekommen, ist es unabdingbar, dass die Lehrperson eine Vorbildfunktion zum kritischen Hinterfragen und bei Bedarf zum begründeten Ablehnen eben dieser, einnehmen kann. Die Kompetenz des kritischen Hinterfragens ist folglich nicht ausschließlich auf wissenschaftliche Studien begrenzt, sondern kann ebenso auf Zeitungsartikel, TV Berichte und andere Medien übertragen werden.

Ziel der evidenzbasierten Praxis in der Bildungsforschung ist es demnach Lehrerinnen und Lehrer zu befähigen, wichtige Entscheidungen auf Basis der besten verfügbaren Evidenz treffen zu können, Evidenz inhaltlich im Rahmen ihres Unterrichts zu nutzen, sowie als Vorbild innerhalb einer Wissensgesellschaft zu fungieren und darüber hinaus ihre Schülerinnen und Schüler innerhalb des Unterrichts mit wissenschaftlichem Denken und Handeln vertraut zu machen (Klein et al., 2012).

## **2.5. Grenzen und Barrieren von Evidenzbasierung**

Nachdem in den vorherigen Kapiteln ausführlich über die Entstehung, sowie die Vorteile des Nutzens wissenschaftlicher Evidenz berichtet wurde, soll nun im

Folgenden darauf eingegangen werden, welche möglichen Grenzen und Schwierigkeiten in der praktische Anwendung auftreten können. Dabei wird zunächst auf die Evidenzbasierung im Allgemeinen eingegangen. Im Anschluss daran werden Grenzen und Barrieren, die für die Lehrerbildung spezifisch sind, aufgezeigt.

### **2.5.1. Grenzen und Barrieren evidenzbasierter Praxis im Allgemeinen**

Zunächst sollen jene Grenzen und Kritikpunkte näher betrachtet werden, welche die evidenzbasierte Praxis im Allgemeinen, also sowohl die Medizin als auch die Bildungswissenschaften betreffen.

#### *Mangelnde Akzeptanz der Praktiker*

Wie in Kapitel 2.3. bereits erwähnt, existieren oftmals negative Einstellungen gegenüber dem Nutzen wissenschaftlicher Studien, sowohl in der evidenzbasierten Medizin, als auch in der Lehrerbildung (e.g. Hetmanek et al., 2015; Slavin, 2002). Dabei gilt es zu unterscheiden zwischen negativen Überzeugungen bezüglich der Glaubhaftigkeit und Belastbarkeit (bildungswissenschaftlicher) Evidenz (Stahl & Bromme, 2007), negativen Einstellungen hinsichtlich des praktischen Nutzwertes von Evidenz (Jette et al., 2003; Johnston et al., 2003) und wahrgenommenen Barrieren die Ausübung evidenzbasierter Praxis betreffend (Jette et al., 2003; Johnston et al., 2003; Young & Ward, 2001). Zudem kann der Forschungsstand, das heißt sowohl die Tatsache ob ausreichend Forschung zu der jeweiligen Problemstellung vorhanden ist, als auch die praktische Nutzbarkeit der Ergebnisse, eine solche Barriere darstellen (Bromme et al., 2016; Hargreaves, 2007a, 2007b; Prenzel, 2012; Schrader, 2014; Slavin, 2008). Die Akzeptanz stellt jedoch eine grundlegende Voraussetzung zur erfolgreichen Umsetzung evidenzbasierter Praxis dar (Slavin, 2002). Es ist deshalb unumgänglich diese Aspekte in zukünftige Interventionen und Trainingsprogramme zu integrieren und die Bedenken und Barrieren der Praktiker inhaltlich aufzugreifen um Lösungsansätze hierfür finden zu können. Nur wenn Praktiker von der Sinnhaftigkeit und des Nutzens von Evidenz tatsächlich überzeugt sind und darüber hinaus der Zugang zu wissenschaftlichen Ergebnissen für die Praxis weiter erleichtert und gefördert wird, werden sie diese auch im Berufsalltag anwenden (Cook et al., 2008).

#### *Mangelnde Kompetenzen zum professionellen Umgang mit Evidenz*

Neben der Motivation und Akzeptanz, sind die Kompetenzen zum kritischen Bewerten und Nutzen von Evidenz ebenso essentiell für ein erfolgreiches Ausführen evidenzbasierter Praxis (siehe auch Kapitel 4). Sowohl in der Medizin, als auch im

Lehramt sind Statistik und empirische Forschungsmethoden bislang kein zentraler Bestandteil, sondern spielen oftmals wenn überhaupt eine stark untergeordnete Rolle. Folglich haben praktizierende Ärztinnen und Ärzte, sowie auch Lehrerinnen und Lehrer häufig Schwierigkeiten, die Evidenzquellen kritisch einzuschätzen und somit die beste verfügbare Evidenz zu identifizieren (Biesta, 2010; Bromme, 2008; Prenzel & Bauer, 2011; Sackett et al., 1996). Dies kann sich wiederum negativ auf Motivation und Akzeptanz auswirken.

#### *Befürchtete Abwertung individueller Expertise*

Viele Praktiker ziehen nach wie vor überwiegend individuelle Expertise und Erfahrungen, sowie innerhalb eines Instituts tradiertes Wissen (Eminenzbasierung, siehe Kapitel 2.2.) nach dem Prinzip der *common practice* für das Treffen von Entscheidungen heran (Davies, 1999; Hargreaves, 2000; Petty, 2009). Evidenzbasierte Entscheidungen werden hingegen oftmals unter der Begründung abgelehnt, dass dies den individuellen Entscheidungsspielraum im Einzelfall einschränken würde und somit eine auf die Situation speziell ausgerichtete Lösung (beispielsweise in der Medizin eine auf individuell auf die Bedürfnisse des Patienten zugeschnittene Behandlung) kaum möglich sei (Cook, 2006). Wichtig ist deshalb den Studierenden zu vermitteln, dass Evidenzbasierung das in Einklang bringen von individueller Expertise und Evidenz zum Ziel hat (vergleiche Kapitel 2.6.).

#### *Hoher Kostenfaktor*

Je valider und belastbarer eine Studie ist, desto höher ist vor allem in der Medizin durch die Starke Aussortierung mit Hilfe der in Kapitel 2.2. beschriebenen Ratingsysteme, die Wahrscheinlichkeit der Empfehlung zur Anwendung der Ergebnisse in der Praxis (Sackett et al., 1996). Vor allem in der Medizin werden inzwischen fast ausschließlich Ergebnisse als Handlungsgrundlage genutzt, die aus groß angelegten, randomisierten Kontrollstudien stammen. Diese Studien sind mit deutlich höherem Aufwand und höheren Kosten verbunden (Bassant et al., 2002). Aus diesem Grund sehen sowohl Forscher als auch Ärzte die Gefahr, dass diese kostenintensiven Studien weniger von Universitäten und immer mehr von den Pharmazieunternehmen durchgeführt, beziehungsweise in Auftrag gegeben werden. Dies wiederum könnte dazu führen, dass bevorzugt monetär gewinnbringende Methoden beforscht und publiziert werden und die Neutralität der Forschung unter diesem Umstand leidet (Bassant et al., 2002; Horrobin, 2000; Kienle et al., 2003). Dies ist jedoch ein sehr für die Medizin spezifischer Aspekt, der sich nicht direkt auf andere Domänen wie beispielsweise die Bildungswissenschaften übertragen lässt.

Dennoch ist es wichtig dies im Blick zu behalten und bei Evidenzquellen auch die Interessen dahinter sowie das durchführende Institut mit in die jeweiligen Entscheidungen einzubeziehen.

#### *Qualitative Studien und Einzelfallstudien*

Die oben genannte Favorisierung quantitativer randomisierter Kontrollstudien hat neben dem Kostenfaktor noch eine weitere Auswirkung. So werden dadurch automatisch Studien, die niedriger eingestuft werden, wie beispielsweise qualitative Studien oder Einzelfallstudien, weniger Beachtung geschenkt. Eben diese Studien können allerdings sehr wertvoll für spezielle Einzelfälle sein, sowohl in der Medizin als auch in den Bildungswissenschaften. Darüber hinaus wird oftmals das Fehlen einer randomisierten Kontrollstudie fälschlicherweise mit dem Fehlen von Wirksamkeit gleichgesetzt (Altmann & Bland, 1995).

### **2.5.2. Grenzen und Barrieren evidenzbasierter Praxis speziell in der Lehrerbildung und im Lehreralltag**

Im Folgenden werden nun weitere Grenzen und Schwierigkeiten evidenzbasierter Praxis diskutiert, die sich speziell auf den Lehrberuf beziehen.

#### *Übertragbarkeit evidenzbasierter Medizin auf die Lehrprofession*

Ein zentraler Kritikpunkt an der Forderung nach evidenzbasierter Praxis in der Lehrprofession ist die Frage nach der Übertragbarkeit eines aus der Medizin stammenden Konzeptes auf die Bildungswissenschaften (Davies, 1999). Hierzu ist anzumerken, dass zwar beide Domänen deutliche Unterschiede, vor allem inhaltlicher und forschungsmethodischer Art, aufweisen, jedoch ist die Art der zu treffenden Entscheidungen, sowie die Situation des Konfliktes zwischen Evidenz- und Eminenzbasierung durchaus vergleichbar. In der evidenzbasierten Medizin ist es Ziel, die beste verfügbare externe Evidenz mit der individuellen Expertise und dem Wunsch des Patienten in Einklang zu bringen (siehe Kapitel 2.2.). Übertragen auf den Lehrberuf, wäre dies das Handeln auf Basis der besten verfügbaren bildungswissenschaftlichen Forschungsergebnisse, der individuellen Erfahrung und Expertise von Lehrkräften und der individuellen Situation innerhalb der Klasse. Darüber hinaus handelt es sich sowohl beim Medizin- als auch beim Lehramtsstudium um ein wissenschaftliches Studium, in welchem sowohl das Rezipieren als auch das Nutzen wissenschaftlicher Forschung eine entscheidende Rolle spielt. Grundsätzlich lässt sich sagen, dass das Konzept zur evidenzbasierten Praxis aus der Medizin nicht eins zu eins auf die Bildungswissenschaften

übertragbar ist, aber unter Einbindung relevanter Anpassungen durchaus vergleichbar sein kann (Finkel, 2003) (vergleiche Kapitel 2.3.).

### *Überzeugungen und Einstellungen bezüglich der Belastbarkeit und Generalisierbarkeit bildungswissenschaftlicher Evidenz*

Wie bereits beschrieben, haben Lehrkräfte oft Vorbehalte gegenüber bildungswissenschaftlicher Evidenz. Häufig wird einerseits Forschung im Allgemeinen, sowie Forschung in der Medizin und den Naturwissenschaften als belastbar und objektiv eingeschätzt, gleichzeitig aber die Forschung in den Bildungswissenschaften als subjektiv und instabil wahrgenommen (Bromme, 2008; Gallagher, 2002). Im Vergleich zur Medizin gibt es in den Bildungswissenschaften kaum randomisierte Kontrollstudien und systematische Reviews. Bestrebungen gehen jedoch vermehrt in die Richtung, die wissenschaftlichen Methoden in der Bildungsforschung dementsprechend weiterzuentwickeln. Dies ermöglicht eine langsam wachsende Anzahl an entscheidungsrelevanter Evidenz, welche für die Praxis geeignet ist (Bromme et al., 2014; Hartmann et al., 2016; Schrader, 2014). Dennoch ist der in der Medizin angestrebte *Goldstandard* nicht eins zu eins auf die Bildungswissenschaften übertragbar, da ein Großteil der Studien im Feld (wie beispielsweise in der Schule) und nicht im Labor stattfinden. Vielmehr wird deshalb eine methodische Vielfalt aus (*quasi-*) *experimentellen* Studien und *mixed-methods* Designs angestrebt (vgl. Kapitel 2.2.) (Bromme et al., 2016; Prenzel, 2012; Schrader, 2014). Abgesehen von der methodischen Qualität wird aber darüber hinaus auch die Relevanz für die Praxis von Lehrkräften als eher gering eingeschätzt (Gough, 2004). So werden bildungswissenschaftliche Studien von Lehrkräften oft als zu theoretisch und wenig hilfreich für die Praxis angesehen. Aus diesem Grund wird gegenwärtig, nach dem Vorbild der *What Works Clearinghouses* aus der Medizin, im Rahmen eines Clearing Houses für Unterricht angestrebt, aktuelle Forschungsergebnisse dergestalt aufzubereiten, dass es Lehrerinnen und Lehrern möglich ist, sich trotz eingeschränkter Zeit und forschungsmethodischem Hintergrundwissen, über die neueste Studienlage zu informieren (Clearing House Unterricht, 2015; Seidel et al., 2016). Voraussetzung hierfür ist jedoch eine ausreichend erforschte Studienlage hinsichtlich der jeweiligen Problemstellungen. Ein weiterer Lösungsansatz im internationalen Raum sind die Aufbereitung von Studienergebnissen in Metaanalysen und beispielsweise die daraus entstandenen *What Works in Classroom* Rankings von John Hattie (2008). Mit Hilfe dieser praxisorientierten Aufbereitungen können Lehrerinnen und Lehrer effektiv und zeitsparend einen Eindruck über den aktuellen Forschungsstand

gewinnen, was sich wiederum positiv auf ihre Einstellungen (vor allem bezüglich der Generalisierbarkeit und dem praktischen Nutzen wissenschaftlicher Forschung) auswirken kann.

#### *Mangelnde Zeit zum Nutzen von Evidenz in der Praxis*

Eine häufig genannte wahrgenommene Barriere zur Ausübung evidenzbasierter Praxis im Lehreralltag ist der Zeitfaktor. Studien zeigen, dass Lehrerinnen und Lehrer oftmals der Überzeugung sind, evidenzbasierte Praxis sei extrem zeitintensiv und deshalb wäre es nahezu unmöglich diese in den Lehreralltag zu integrieren (Porsch & Bromme, 2011). Daraus resultieren starke Ressentiments, da auf diese Weise evidenzbasierte Praxis mit einem erheblichen Mehraufwand in Verbindung gebracht wird. Auch hierfür können die Synthesen des *Clearing House Unterrichts* (CHU), sowie weitere systematischen Reviews und Metaanalysen nutzbringend sein.

#### *Mangelnde Verfügbarkeit*

Darüber hinaus haben Lehrkräfte häufig Probleme überhaupt an aktuelle Forschungsergebnisse zu gelangen, da viele Schulen die benötigten Zugänge und Abonnements für relevante Fachzeitschriften nicht bereitstellen (können) (Reiss et al., 2015). Überdies haben die Forschenden an den Universitäten oftmals wenig Kontakt mit Lehrkräften und sind somit unzureichend mit den realen Problemstellungen der Praxis vertraut. So gibt es für zahlreiche Fragestellungen nach wie vor wenig beziehungsweise unzureichende Evidenz (Hargreaves, 2007a, 2007b).

#### *Gefahr der überkritischen Bewertung*

Ein weiterer Aspekt der bedacht werden sollte und der vor allem bei der Konzeption von Trainingsprogrammen zur evidenzbasierten Praxis beachtenswert ist, ist die Entwicklung der Teilnehmenden dahingehend, bei der kritischen Bewertung von Evidenz zu kritisch zu werden. Wissenschaftliche Forschung hat, wie bereits erwähnt, die Eigenschaft konfligierend und fragil zu sein (Bromme, 2008). Es gibt demnach nicht die „perfekte“ Studie mit perfekten Rahmenbedingungen, einer perfekten Stichprobenszusammensetzung und perfekten Daten und Ergebnissen. Es ist deshalb wichtig darauf zu achten, die Praktiker dahingehend zu sensibilisieren Forschung zwar kritisch zu hinterfragen aber dennoch keine Perfektion als Maßstab anzusetzen.

### *Fazit*

Zusammenfassend lässt sich sagen, dass sowohl die empirische Erforschung von für den Lehrberuf relevanten Problemstellungen, als auch die Förderung der relevanten Kompetenzen zum kritischen bewerten und nutzen eben dieser Evidenz, bereits deutliche Fortschritte aufzuweisen hat. Diese zeigen sich vor allem durch den Ausbau empirischer Forschung in den Bildungswissenschaften, sowie durch den Aufbau von Institutionen wie dem Clearing House Unterricht, welche den Zugang zu Forschungsergebnisse für Praktiker weiter erleichtern. Jedoch sind zugleich noch zahlreiche Herausforderungen zu bestreiten. Diese bestehen einerseits in der Verbesserung und der Ausweitung empirischer Forschungsmethoden in den Bildungswissenschaften. Andererseits werden zudem weitere Institute benötigt, die es Praktikern ermöglichen mit strukturierten Kurzzusammenfassungen zeiteffizient praxisrelevante Inhalte aus der aktuellen Forschung in ihr professionelles Handeln einzubeziehen. Darüber hinaus werden weitere Trainings und Workshops zur Vermittlung der Kernkompetenzen der evidenzbasierten Praxis aber auch zur Steigerung der Akzeptanz und Motivation bei (zukünftigen) Lehrerinnen und Lehrern benötigt.

## **2.6. Evidenzbasierung versus Evidenzorientierung**

Wie bereits in Kapitel 2.1. angedeutet, führt der Terminus evidenzbasierte Praxis häufig zu Missverständnissen dahingehend, dass Entscheidungen einzig und allein auf Basis von externer, wissenschaftlicher Evidenz getroffen werden sollen. Dieses Missverständnis begründet unter anderem, die sowohl bei praktizierenden Lehrerinnen und Lehrern, als auch bei Lehramtsstudierenden vorherrschenden Ressentiments in Bezug auf evidenzbasierte Praxis (Hetmanek et al., 2015; Hargreaves, 2000; Parr et al., 2008; Williams & Coles, 2008; Zembal-Saul et al., 2003). Aus diesem Grund kann es förderlich sein, an Stelle dessen von *evidenzinformierter* (Biesta, 2010) oder *evidenzorientierter Praxis* (Stark, 2017) zu sprechen. Ziel ist es dabei zu vermitteln, dass unter der Begrifflichkeit das situationsangemessene Berücksichtigen wissenschaftlicher Forschungsergebnisse im Rahmen des professionellen Handelns (vgl. Bauer et al., 2015) zu verstehen ist. Demnach soll nur dann wissenschaftliche Evidenz als Unterstützung hinzugezogen werden, wenn es tatsächlich sinnvoll ist. Wichtig ist darüber hinaus, dass auch im Falle des Hinzuziehens von Evidenz nicht ausschließlich die darin enthaltenen Forschungsergebnisse zu berücksichtigen sind. Die in Kapitel 2.2. beschriebenen drei Säulen evidenzbasierter Medizin, bestehend aus der besten verfügbaren externen Evidenz, der internen Evidenz, sowie den Wünschen bzw. Präferenzen

des Patienten oder der Patientin (Antes et al., 2003; Greenalgh, 2003; Sackett et al., 1997) können sinngemäß auch auf die Bildungsforschung übertragen werden (siehe Abbildung 4).

Um wichtige und weitreichende Entscheidungen im Lehrberuf sinnvoll und situationsangemessen auf Basis von Evidenz treffen zu können ist es folglich unumgänglich, die Ergebnisse aus den besten und validesten verfügbaren Evidenzquellen mit der eigenen Expertise und Erfahrung, die sich Lehrkräfte im Laufe ihres Berufslebens aneignen, in Einklang zu bringen und diesbezüglich Abwägungen zu treffen. Darüber hinaus ist immer die individuell vorliegende Situation, wie beispielsweise spezielle Hintergründe oder Bedürfnisse der Schülerinnen und Schüler, zu beachten.

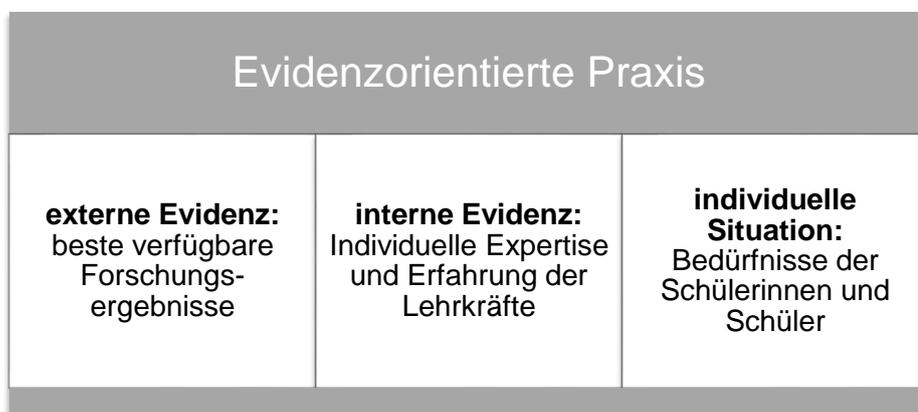


Abbildung 4: Darstellung der drei Säulen für evidenzorientierte Praxis im Lehrberuf

Folglich kann es sinnvoll sein, im Kontext der Lehrerbildung an Stelle von evidenzbasierter Praxis von evidenzorientierter Praxis zu sprechen, da dieser Begriff einerseits weniger vorurteilsbelastet ist und andererseits die Intention des Konzeptes (nämlich einer situationsangemessenen Verwendung von Evidenz) besser beschreiben kann.

## 2.7. Fazit

In diesem Kapitel wurde zunächst die Entwicklung des Ansatzes des Nutzens wissenschaftlicher Evidenz für Problemstellungen in der Praxis, sowie die dafür relevanten Kompetenzen mit ihren möglichen Einflussfaktoren basierend auf verschiedenen Modellen unterschiedlicher Domänen erläutert. Auf Basis dessen wurde ein Bedingungsmodell erarbeitet, welches als Grundlage für das Fördern des kompetenten Nutzens von Evidenz in der Lehrerbildung genutzt werden kann (siehe Abbildung 3). Zur Identifizierung der relevanten Kompetenzen hat scheint ein

Kompetenzmodell aus der Medizin (Ramos et al., 2003; Rubin, 2008; Shaneyfelt et al., 2006) am praktikabelsten. Hinzu kommen sowohl interne Einflussfaktoren, wie (Vor-)Wissen und Orientierungen, als auch externe Faktoren, wie Anwendungsbereiche und Merkmale von Evidenz.

Ziel dieser Arbeit ist es Informationen darüber zu generieren, wie vor allem das kritische Hinterfragen und das argumentative Nutzen bei Studierenden im Rahmen des Lehramtsstudiums effektiv gefördert werden kann. Weitere Schritte, wie etwa Aus der Expertiseforschung, Forschung zu Problemlösestrategien oder Entscheidungskompetenz wurden in diesem Zusammenhang nicht miteinbezogen. Dies wäre jedoch ein denkbarer nächster Schritt.

Im Anschluss daran wurde auf die Möglichkeiten und Grenzen des Nutzens von Evidenz in der Praxis sowohl im Allgemeinen, als auch im Speziellen in der Lehrerbildung eingegangen. Zusammenfassend lässt sich sagen, dass die empirische Bildungsforschung ein expandierendes Forschungsfeld ist, welches unter anderem zum Ziel hat relevante Forschung zu Problemstellungen der Praxis zu generieren. Dies ist von großer Wichtigkeit, denn sowohl Metaanalysen wie beispielsweise die von Hattie, als auch für die Praxis aufbereitete Forschungssynthesen von Institutionen wie dem Clearing House Unterricht können nur auf tatsächlich vorhandene Forschungsergebnisse zurückgreifen.

Abschließend wurde auf Basis der zuvor erarbeiteten Abwägung von Chancen und Einschränkungen des Nutzens wissenschaftlicher Evidenz, der Ansatz der situationsangemessenen und dadurch *evidenzorientierten* Praxis diskutiert, welcher ein geeignetes und motivierendes Ziel bezüglich der Förderung relevanter Kompetenzen im Lehramtsstudium sein kann.

## 3. Teilkompetenzen zum professionellen Umgang mit Evidenz in Unterricht und Schule

In Kapitel 2 wurde bereits beschrieben wo das Konzept der evidenzbasierten Praxis seinen Ursprung hat und welche Kompetenzen hierfür nötig sind. Im Rahmen der dieser Arbeit zugrunde liegenden Studien wurde im Besonderen auf zwei dieser Kompetenzen eingegangen, (1) das kritische Hinterfragen von Evidenz (*appraise*), sowie (2) das argumentative Nutzen (*apply*). Der Grund hierfür liegt darin, dass diese beiden Teilkompetenzen diejenigen sind, mit denen die Studierenden die größten Schwierigkeiten zu scheinen haben (Fischer et al., 2014). Im folgenden Kapitel wird deshalb auf nochmals detaillierter erläutert, wie diese zu definieren sind und wie der aktuelle Forschungsstand diesbezüglich ist.

### 3.1. Kritisches Hinterfragen von Evidenz

Kritisches Hinterfragen von Evidenz beschreibt den Prozess des Beurteilens und Interpretierens wissenschaftlicher Evidenz bei systematischer Betrachtung der Ergebnisse, sowie der Belastbarkeit und der Relevanz der vorliegenden Studien (Burls, 2009). Dies ist von zentraler Bedeutung, da oftmals nicht auf den ersten Blick ersichtlich ist, welche Studien sich zur Beantwortung einer Fragestellung eignen und welche Quellen verlässlich und belastbar sind. Im Bereich der Medizin gibt es, wie bereits in Kapitel 2 beschrieben, zahlreiche Ratingsysteme zur kritischen Einschätzung (z.B. Chochrane Collaboration, 2017; DNEbM, 2017). Diese können jedoch nicht eins zu eins auf die Bildungswissenschaften übertragen werden, da randomisierte Kontrollstudien, die den Goldstandard in der evidenzbasierten Medizin bilden, in der empirischen Bildungsforschung kaum stattfinden.

Um das Kompetenzlevel der Studierenden zu bestimmen wurde deshalb auf Basis eines Schemas zur Bewertungskompetenz, welches aus dem Bereich Science Education stammt (Wilson, 2005), unter Integration der für die Bildungswissenschaften relevanten Parameter (Bromme, 2014; Kuhn & Crowell, 2011) ein Modell mit fünf Kompetenzstufen entwickelt (siehe Tabelle 2). Die höchste zu erreichende Kompetenzstufe ist demnach, wenn Studierende die für die Problemstellung relevanten Inhalte aus einer Evidenzquelle erfassen und diese auf Basis der gelehrten Heuristiken (siehe Kapitel 4.3.) kritisch hinterfragen und bewerten können. Liegt mehr als eine Evidenzquelle vor, ist zudem das Erkennen der Fragilität und Konflikthaftigkeit des Forschungsfeldes erforderlich, um auf Basis

einer darauf beruhenden Abwägung der Vor- und Nachteile der verschiedenen Studien eine wissenschaftlich fundierte Entscheidung treffen zu können. Wird dies nicht erkannt und berücksichtigt, bewegt sich der Teilnehmende auf Kompetenzstufe 4. Wird die Evidenz lediglich stellenweise oder teilweise fehlerhaft kritisch bewertet, lässt sich dies in Kompetenzstufe 3 einordnen. Bei durchgängig falscher oder gänzlich fehlender kritischer Bewertung wäre ein Studierender auf Stufe 2 einzuordnen. Werden zudem auch noch die Inhalte der Studien missverstanden oder falsch zitiert, wäre dies die niedrigste Kompetenzstufe 1.

*Tabelle 2.* Kompetenzstufen zum kritischen Bewerten von Evidenz

<i>Level 5</i>	Der Studierende nennt die relevanten und wichtigsten Inhalte und bewertet diese kritisch. Darüber hinaus erkennt er die Fragilität und Konflikthaftigkeit des Forschungsfeldes.
<i>Level 4</i>	Der Studierende nennt die relevanten und wichtigsten Inhalte und bewertet diese durchgängig kritisch.
<i>Level 3</i>	Der Studierende nennt die relevanten Inhalte und bewertet diese teilweise kritisch.
<i>Level 2</i>	Der Studierende nennt Inhalte der Studie ohne diese kritisch zu bewerten.
<i>Level 1</i>	Der Studierende nennt irrelevante oder falsche Inhalte der Studie. Zudem bewertet er diese nicht kritisch oder sogar falsch.

Frühere Studien haben gezeigt, dass gerade die konfligierenden und fragilen Aspekte von Evidenz oftmals die größte Herausforderungen sowohl für Studierende als auch für praktizierende Lehrkräfte sein können (Bromme, 2008; Fischer et al., 2014; Hetmanek et al., 2015; Kuhn & Crowell, 2011).

### **3.2. Argumentatives Nutzen von Evidenz**

Nachdem in Kapitel 3.1. beschrieben wurde, wie bildungswissenschaftliche Evidenz kritisch hinterfragt und bewertet werden kann, soll nun genauer darauf eingegangen werden, wie man die Evidenz auf Basis dieser kritischen Einschätzung argumentativ nutzen kann. Dies ist wiederum ein essentieller Vorgang beim Treffen wissenschaftlich fundierter Entscheidungen. Dabei wird zunächst genauer auf die Relevanz des evidenzbasierten Argumentierens für Lehrkräfte eingegangen. Anschließend wird erläutert, welche Parameter gutes evidenzbasiertes Argumentieren bedingen und welche darüber hinaus für gutes Argumentieren

innerhalb von Dialogen oder Gruppendiskussionen von Bedeutung sind. Überdies werden häufige Probleme und Fehler, die beim evidenzbasierten Argumentieren auftreten können und somit von Relevanz bei der Erstellung von Trainingsprogrammen sein können, beschrieben.

### **3.2.1. Relevanz evidenzbasierten Argumentierens**

Evidenzbasiertes Argumentieren ist in diversen Kontexten von zentraler Bedeutung. Zum einen ist es essentiell für die professionelle Arbeit von Lehrkräften. Die Kultusministerkonferenz hat bereits 2004 postuliert, dass Lehrerinnen und Lehrer ihren Beruf „als ständige Lernaufgabe“ verstehen und in diesem Zusammenhang „*Ergebnisse der Bildungsforschung für die eigene Tätigkeit*“ nutzen, rezipieren und bewerten sollen (KMK, 2004, Kompetenz 10). Zum anderen sollen Lehrkräfte in der Lage sein „*Ergebnisse der Unterrichts- und Bildungsforschung auf die Schulentwicklung*“ anzuwenden (KMK, 2004, Kompetenz 11). Aber auch im internationalen Forschungsraum besteht die Forderung nach evidenzbasierter Praxis im Lehrerberuf (z.B. Cetin et al., 2010; Cochran-Smith & Lytle, 1999; Erduran et al., 2015; Fischer et al., 2014; Hennessey et al., 2013; Kuhn, 2008; Ozdem et al., 2011; Zembal-Saul et al., 2003;). Dabei ist evidenzbasiertes Argumentieren nicht nur eine Kernkompetenz um den Lehrerberuf an sich auszuüben. Es wird zudem gefordert dies darüber hinaus bereits frühzeitig als Element innerhalb des Unterrichts zu nutzen, um Schülerinnen und Schüler frühzeitig mit wissenschaftlichem Argumentieren vertraut zu machen (Driver et al., 2000; Hemberger et al. 2017; Kuhn, 2015; Kuhn & Crowell, 2011; Kuhn et al., 2008). Der Grund hierfür ist, dass evidenzbasiertes Argumentieren eine zentrale Voraussetzung ist um wissenschaftliches Denken und somit auch Wissenschaft selbst verstehen zu können (Belland et al., Bricker & Bell, 2008; Kuhn et al., 2015). Überdies ist wissenschaftliches Argumentieren, ebenso wie das kritische Bewerten von Evidenz, eine zentrale Kompetenz um auch im privaten Alltag fundierte und informierte Entscheidungen treffen zu können (Felton, 2004; Kelly et al., 1998; Kulatunga et al., 2014; Mercier et al., 2011; Rapanta & Macagno, 2016; Zohar et al., 2002). Durch evidenzbasiertes Argumentieren und Begründen von Handlungen kann zudem eine nachträgliche Rechtfertigung getroffener Entscheidungen, auch im juristischen Sinne, erleichtert werden. Dies ist besonders im Bereich der evidenzbasierten Medizin von essentieller Bedeutung (Chisholm, 1980; Curd & Cover, 1998; Sacket et al., 1996), jedoch auch im Lehrerberuf denkbar.

### 3.2.2. Qualitätsmerkmale evidenzbasierten Argumentierens

Gutes evidenzbasiertes Argumentieren bedeutet eine Behauptung (*claim*) aufzustellen und diese mit Hilfe von wissenschaftlicher Evidenz (als *backing* oder *warrant*) zu begründen und zu untermauern (Booth, Colomb & Williams, 2008; Osbourne, 2010; Sandoval & Reiser, 2004; Toulmin, 1958). Analog zu den in Kapitel 3.1. dargestellten Kompetenzstufen für kritisches Hinterfragen von Evidenz kann auch die Kompetenz des argumentativen Nutzen, durch fünf Kompetenzlevel dargestellt werden (Kuhn & Crowell, 2011; Wilson, 2005) (siehe Tabelle 3).

Die höchste Stufe (Kompetenzstufe 5) ist erreicht, wenn der Studierende durchgängig seine Behauptungen mit Evidenz begründet und dabei die Evidenzquelle(n) zugleich kritisch hinterfragt und bewertet. Das erfolgreiche kritische bewerten zeigt sich in diesem Fall durch eine innerhalb des Arguments enthaltene wissenschaftliche Aussage über die Evidenzquelle (*single-warrant*) oder gar mehrere Aussagen dieser Art (*multiple warrant*). Darüber hinaus ist das Argument klar strukturiert und nachvollziehbar (Bell et al., 2010; Toulmin, 1978). Die Evidenz wird in reflektierter Weise in die Behauptung integriert und dadurch ein sinnvoller Argumentationszusammenhang geschaffen (Kuhn & Crowell, 2011; Wilson & Sloane, 2000). Darüber hinaus werden innerhalb der Argumentation alle Aspekte, die sowohl für, als auch gegen die eigene Behauptung sprechen betrachtet (*two-sided argumentation*) (Kuhn & Crowell, 2011). Für ein erfolgreiches Nutzen von Evidenz muss die Meinung der Quelle jedoch nicht zwangsläufig geteilt werden. Den Ergebnissen der Evidenzquelle kann ebenso durch begründete Kritik (beispielsweise aus forschungsmethodischen Gründen oder auf Grund fehlender Generalisierbarkeit) widersprochen werden und so durch Ablehnen der Ergebnisse als *backing* für die eigene Behauptung verwendet werden (Hemberger et al. 2017; Kuhn, 2015). Eine Stufe darunter (Kompetenzstufe 4) wird erreicht, wenn der Studierende zwar durchgängig evidenzbasiert argumentiert und dafür auch relevante und angemessene Evidenz verwendet, diese aber nicht kritisch hinterfragt. Argumentiert der Studierende nur stellenweise und nicht in Bezug auf alle Elemente evidenzbasiert kann er auf Kompetenzstufe 3 eingeordnet werden. Lediglich Kompetenzstufe 2 erreichen Studierende, die kaum Evidenz verwenden um ihre Begründung zu stützen oder dabei irrelevante Evidenz heranziehen, sowie inhaltliche Fehler begehen. Ebenso wird diese Stufe erreicht, wenn der Studierende die Evidenzquelle lediglich zitiert ohne sie in einen sachgemäßen und logischen Zusammenhang mit einer Behauptung zu bringen. Ein weiteres Kriterium zur Einordnung auf Kompetenzlevel 2 ist es, wenn Studierende lediglich diejenigen Argumente aus der Evidenzquelle herausziehen, welcher sich mit ihrer eigenen

Meinung decken (*one-sided argumentation*) und diejenigen kommentarlos auslassen, welche der eigenen Meinung widersprechen (Kuhn & Crowell, 2011). Die unterste Stufe (Kompetenzstufe 1) bezeichnet den Fall, wenn durchgängig keinerlei Evidenz verwendet werden und ausschließlich unbegründete Behauptungen aufgestellt werden (Wilson, 2005).

*Tabelle 3.* Kompetenzstufen zum argumentativen Nutzen von Evidenz

<i>Level 5</i>	Der Studierende begründet durchgängig evidenzbasiert und hinterfragt die Evidenz kritisch.
<i>Level 4</i>	Der Studierende begründet durchgängig evidenzbasiert und verwendet dabei relevante und angemessene Evidenz.
<i>Level 3</i>	Der Studierende begründet teilweise evidenzbasiert, aber nicht in Bezug auf alle Elemente.
<i>Level 2</i>	Der Studierende zitiert irrelevante Evidenz oder macht inhaltliche Fehler.
<i>Level 1</i>	Der Studierende nennt durchgängig keine Begründungen und führt keine Evidenz auf.

### **3.2.3. Evidenzbasiertes Argumentieren in Gruppen**

Wird nicht (schriftlich) auf Individualebene argumentiert, sondern innerhalb von Dialogen oder Gruppendiskussionen, gibt es weitere Kriterien und Parameter, die für eine hohe Qualität evidenzbasierten Argumentierens relevant sind. Bislang gibt es international zahlreiche Studien, welche evidenzbasiertes Argumentieren unterschiedlichster Zielgruppen (z.B. Schülern, wissenschaftlichen Laien und Lehrkräften) in dialogischen Diskussionen (Zwiegesprächen) untersuchen (z.B. Asterhan & Schwarz, 2009; Billig, 1987; Felton & Kuhn, 2001; Graff, 2003; Kuhn, 2005; Kulatunga et al., 2014; Rips et al., 1999; Teasley 1997). Es gibt jedoch bislang kaum Untersuchungen zur Argumentationsqualität in größeren Gruppen ab 3 Teilnehmenden.

In Dialogen und Gruppendiskussionen kann Evidenz aus unterschiedlichen Beweggründen genutzt werden: Sie kann den Zweck haben die eigene Position zu bestärken (*M+*), oder die Position anderer zu schwächen (*O-*), sowie die Position anderer zu unterstützen (*O+*), als auch um die eigene Position zu schwächen (*M-*) (Kuhn, 2015; Kuhn & Crowell, 2011). Diese sind qualitativ zwar als gleichwertig anzusehen, letzteres (*M-*) setzt jedoch das höchste Maß an Bereitschaft zum Rezipieren und Nutzen von Evidenz voraus, da hierbei die eigene Meinung durch

die Überzeugungskraft der Evidenz verändert wird. (Hemberger et al., 2017; Kuhn & Crowell, 2011).

Einen zentralen Einfluss auf die Argumentationsqualität hat zudem die Art und Weise, in welcher ein Studierender darauf reagiert, wenn er die Meinung des Mitdiskutanten nicht teilt. Dies kann idealerweise sowohl durch eine begründete Kritik am Argument des anderen zum Ausdruck gebracht werden (*Counter-C*), als auch durch ein stichhaltiges, evidenzbasiertes Gegenargument (*Counter-A*). Eine niedrige Qualität beim Ausdruck von Uneinigkeit wäre hingegen das bloße Ansprechen der Nichtübereinstimmung (*disagreement*), wie beispielsweise indem lediglich gesagt wird „das stimmt doch nicht“, oder „nein das finde ich nicht“ ohne dass hierfür eine weitere Begründung geliefert wird. (Kuhn & Crowell, 2011).

Ein weiteres Kriterium für erfolgreiches evidenzbasiertes Argumentieren in Gruppen ist die Transaktivität. Diese beschreibt, ob sich das Argument eines Teilnehmenden innerhalb einer Diskussion auf das bezieht, was der vorherige Sprecher gesagt hat, oder ob ein völlig neuer Aspekt ohne jeglichen Bezug zu dem vorherigen Argument vorgebracht wird. Transaktive Aussagen können beispielsweise Gegenargumente (*Counter-A* oder *Counter-C*) sein, sowie Ergänzungen zu der zuvor getroffenen Aussage oder Argumente die an das Gesagte des Vorredners anknüpfen (Felton & Kuhn, 2001; Kuhn & Udell, 2003). Transaktivität ist in Gruppendiskussionen und Dialogen eine wichtige Voraussetzung zur gemeinsamen Entscheidungsfindung (Felton & Kuhn, 2001).

#### **3.2.4. Häufige Schwierigkeiten und Fehler**

Nachdem die relevanten Hintergründe zum evidenzbasierten Argumentieren auf Individual- sowie auf Gruppenebene beschrieben wurden, sollen im Folgenden häufig auftretende Probleme und Schwierigkeiten im argumentativen Umgang mit Evidenz dargelegt werden. Beim evidenzbasierten Argumentieren auf der Individualebene zeigen sich häufig Schwierigkeiten bei der Verbindung einer Behauptung (*claim*) mit der Evidenz (*evidence*). Oftmals werden hierbei Behauptungen mit irrelevanter, unpassender oder falsch zitierter Evidenz untermauert. Zudem zeigte sich häufig, dass Behauptungen alleine und ohne jegliche Evidenz aufgestellt wurden. Aber auch ein bloßes Zitieren einer Quelle ohne das Aufstellen einer Behauptung und somit ohne jeglichen Argumentationszusammenhang, ist eine in der Literatur bekannte Fehlerquelle (Sandoval & Millwood, 2005).

Des Weiteren ist ein Missinterpretieren der Evidenz, so dass sie zum eigenen Claim passt und die damit einhergehende fehlende Neutralität, eines der zentralsten

Probleme und Schwierigkeiten in der evidenzbasierten Praxis. Hierbei wird die Evidenzquelle nicht kritisch bewertet und objektiv genutzt, sondern es werden lediglich Ergebnisse, welche die eigene Meinung teilen, herangezogen oder sogar der eigenen Meinung widersprechende Ergebnisse falsch zitiert um die eigene, persönliche Meinung zu unterstützen (Kuhn & Moore, 2015). Zudem stützen Lehramtsstudierende ihre Argumente und Entscheidungen oftmals lediglich auf Alltagswissen („das weiß man doch, dass...“) (siehe auch „anekdotische Evidenz“, Kapitel 2.1.) oder fehlerhafte Theorien (Stark, 2005)

Auf Gruppenebene in Dialogen oder Diskussionen ist einer der häufigsten Fehler und Schwierigkeiten beim evidenzbasierten Argumentieren das Fehlen von Transaktivität. Häufig konzentrieren sich Teilnehmende nahezu ausschließlich auf den eigenen Argumentationsstrang und lassen dadurch die Aussagen der anderen Diskussionsteilnehmer außer Acht. Dadurch kann keine qualitativ hochwertige und fließende Diskussion entstehen sondern im Extremfall lediglich sich abwechselnde unzusammenhängende Einzelargumente (Felton & Kuhn, 2001; Kuhn & Udell, 2003; Kuhn et al., 2008).

Bei der Konzeption neuer Trainingsprogramme zum argumentativen Nutzen von Evidenz kann es folglich sehr förderlich sein diese bereits bekannten Schwierigkeiten und Probleme aufzugreifen und Lösungsansätze hierfür zu implementieren.

### **3.3. Fazit**

Ziel dieses Kapitels war es, diejenigen Kompetenzfacetten detaillierter darzulegen, die im Rahmen der dieser Arbeit zu Grunde liegenden Studien speziell gefördert und gemessen wurden. Die in diesem Kapitel beschriebenen Kompetenzmodelle und Ansätze können sowohl für die inhaltliche Konzeption eines Trainingsprogramms und die Erstellung von unterstützenden Heuristiken hilfreich sein, als auch als Grundlage für anschließende Kodierschemata zur Erfassung der Kompetenz zum kritischen Hinterfragen und argumentativen Nutzen von Evidenz dienen.

## 4. Fördern von Kompetenzen zur professionellen Nutzung von Evidenz

In Kapitel 2 wurde die relevante Hintergrundforschung zum professionellen und kompetenten Nutzen von Evidenz im Allgemeinen und im Speziellen in der Lehrerbildung dargestellt. Auf Grundlage dessen wurde ein Bedingungsmodell mit den relevanten Kompetenzen und Einflussfaktoren erstellt. Im Folgenden Kapitel soll anknüpfend daran näher darauf eingegangen werden, wie diese Kompetenzen effektiv gefördert und trainiert werden können, um gezielte Veranstaltungen mit Trainingscharakter konzipieren zu können, die flexibel in das bestehende Lehramtscurriculum integrierbar sind.

Dabei werden zunächst etablierte Trainingsprogramme aus den Bereichen der Science Education, sowie der Medizin betrachtet. Zudem gibt es auch erste Förderungsansätze aus der Lehrerbildung im internationalen Raum. Anschließend werden hilfreiche Verfahren, wie das Verwenden von Heuristiken, sowie weitere förderliche instruktionale Prinzipien beschrieben und hinsichtlich ihrer Nutzbarkeit und Übertragbarkeit diskutiert. Der Fokus wird hierbei, wie in Kapitel 3 beschrieben, auf die beiden Kompetenzen Evidenz kritisch bewerten (*appraise*) und Evidenz (argumentativ) nutzen (*apply*) gelegt, da diese beiden im Rahmen der vorliegenden Studien gefördert und gemessen wurden.

### 4.1. Ansätze aus anderen Domänen und ihre Übertragbarkeit auf die Lehrerbildung

Im Folgenden Kapitel werden zunächst bereits erprobte Trainingskonzepte zu Teilbereichen der evidenzbasierten Praxis aus Domänen mit ähnlichen Zielsetzungen, in diesem der Falle Science Education und Medizin, dargestellt. Ziel ist es dabei, erfolgreiche Ansätze und Methoden zu eruieren, die auf das Training zukünftiger Lehrerinnen und Lehrer übertragbar sind.

#### 4.1.1. Science Education

Zunächst wird hierfür näher auf relevante Forschung aus dem Bereich der Science Education eingegangen. Science Education umfasst das Lernen und Lehren von und über Wissenschaft und richtet sich an alle „Nicht-Wissenschaftler“, beziehungsweise wissenschaftliche Laien. Dabei kann es sich um Kinder, Schülerinnen und Schüler, Studierende oder auch Erwachsene handeln. Enthalten sind Elemente zu wissenschaftlichen Methoden und Designs, sowie wissenschaftlichen Inhalten und Prozessen, die sich überwiegend auf die Sozial-

und Naturwissenschaften beziehen. (Hurd, 1991; Jenkins, 1985; Taber, 2009). Für diesen Bereich ist es essentiell, dass Lernende eine Idee vom Konzept empirischer Forschung generieren und dadurch ein Grundverständnis von (natur-)wissenschaftlicher Forschung erlangen. Aus diesem Grund gibt es innerhalb dieser Domäne bereits einige Trainingsansätze, die sich mit dem Verständnis von Wissenschaft, dem kritischen Hinterfragen, sowie dem argumentativen Nutzen von Evidenz befassen. Im Folgenden werden beispielhaft einige dieser Trainingskonzepte genannt und daraus resultierend Transfermöglichkeiten auf die Lehrerbildung diskutiert.

Im Bereich der Science Education existieren unter anderem hilfreiche Trainings zur Förderung des (argumentativen) Nutzens von (natur-)wissenschaftlicher Evidenz. Darunter befinden sich Trainings und Interventionen mit computerbasierten Lernumgebungen, in welchen Lernende mit Hilfe von Kooperationskripts (CSCL) wissenschaftliches argumentieren erlernen und weiterentwickeln. Diese konnten im Bereich der Science Education bereits signifikante Effekte im Rahmen einer Interventionsstudie mit Psychologiestudentinnen und -studenten hervorbringen (z. B. Kopp & Mandl, 2011; Fischer et al., 2007, Weinberger et al., 2003). Darüber hinaus konnten zahlreiche weitere Trainings und Interventionsstudien zeigen, dass kooperatives Lernen in problemorientierten Umgebungen in Kombination mit dialogischen Diskursen (sowohl virtuell über Chat-Funktionen am Computer als auch persönlich), in welchen wissenschaftliches Argumentieren trainiert werden kann, zu bedeutsamen Fortschritten bei Studierenden, sowie auch bei Schülerinnen und Schülern unterschiedlichen Alters hinsichtlich ihrer (wissenschaftlichen) Argumentationskompetenz führen kann (Kuhn, 2010; Mayweg-Paus et al., 2016; Osbourne, 2010; Smyrniou et al., 2015).

Überdies konnten beispielsweise Hutschinson & Gigerenzer (2005; Gigerenzer, 2008) im Rahmen ihren Studien die Wirksamkeit des Einsetzens von Heuristiken (in diesem Beispiel im Bereich der Biologie) aufzeigen, wenn es um das Treffen komplexer Entscheidungen geht (siehe auch Kapitel 3.3.).

Dies zeigt, dass sowohl das Nutzen von Kooperationskripts und problemorientierten Fallbeispielen, als auch das Trainieren wissenschaftlichen Argumentierens innerhalb von Diskussionen signifikante Lernfortschritte bringen kann. Darüber hinaus können Heuristiken beim Treffen fundierter Entscheidungen unterstützen. Beide Fähigkeiten sind Voraussetzungen zum kompetenten Nutzen von Evidenz (*apply*). Sie sind thematisch individuell anpassbar, für wissenschaftliche Laien jeglichen Vorwissensstands geeignet und zeitlich in ein

Kurzzeittraining integrierbar. Aus diesem Grund scheint ein Adaptieren dieser Methoden in ein Training für zukünftige Lehrkräfte sinnvoll.

#### **4.1.2. Evidenzbasierte Medizin**

Auch in der Medizin gibt es, neben zahlreichen Trainings und Workshops zur Forschungskompetenz, diverse Trainingsansätze zur Förderung der relevanten Kompetenzen für die Ausübung evidenzbasierter Praxis. Überwiegend basieren diese auf den in Kapitel 2.2. beschriebenen Ratingsystemen (z.B. Higgins et al., 2011; Howick et al., 2011) und sind deshalb schwer auf andere Domänen übertragbar. Jedoch gibt es dennoch einige Ansätze, die weniger domänenspezifisch sind. Im Folgenden soll deshalb ein Einblick dahingehend gegeben werden, welche hilfreichen Zugänge zum Erlernen der notwendigen Theorien und Methoden in dieser Domäne vorherrschen. Einige dieser Trainings werden im Folgenden kurz skizziert. Ziel dessen ist es Elemente zu eruieren, die in ein Training für Lehramtsstudierende integrierbar sind.

So implementierten Finkel et al. (2003) einen viermal je 2-stündigen Kurs in das Medizinstudium. Im Rahmen dessen bekamen die Studierenden nach einer kurzen Inputphase die Aufgabe in Gruppen für 25 Minuten einen aktuellen Fall zu diskutieren, anschließend für 25 Minuten in den relevanten Datenbanken nach Evidenz zu suchen und danach weitere 25 Minuten eine Synthese für einen klinischen Plan zu erstellen. Dies ist ähnlich zu den Diskussionsphasen aus der Science Education, angereichert mit einer eigenen Recherchephase. Sowohl die Recherche- als auch die Diskussionsphasen können auf Interventionen für Lehramtsstudierende übertragbar sein, da auch Lehrkräfte in ihrem Berufsalltag in die Situation gelangen können selbst relevante und aktuelle Forschungsergebnisse recherchieren zu müssen, sowie die Ergebnisse dieser in Diskussionen zu nutzen beziehungsweise zu rezipieren (wie beispielsweise in Lehrerkonferenzen oder Lehrer-Eltern-Gesprächen).

Weitere Trainings und Interventionsstudien zum Lehren der Kernkompetenzen evidenzbasierter Medizin wurden, dem eigenen Goldstandard entsprechend, im Rahmen randomisierter Kontrollstudien getestet und stützten sich hauptsächlich auf *blended learning* (BL) Ansätze (z. B. Gerhardt-Szep et al., 2017; Ilic et al., 2015). Die Kontrollgruppe erhielt einen konservativeren Unterricht in Form von *didaktischem Lernen* (DL). Didaktisches Lernen beschreibt in diesem Fall das Verwenden herkömmlicher Unterrichtsmethoden, die nicht auf online-basierten Ansätzen beruhen. Hierbei konnten jedoch bei den EBM relevanten Kompetenzen keine signifikanten Unterschiede festgestellt werden. Folglich gilt es diesen Ansatz zunächst weiterhin dahingehend zu überprüfen, ob der durch das Erstellen einer

*blended learning* Umgebung zu leistende Mehraufwand tatsächlich lohnenswert ist, bevor über ein Übertragen in andere Domänen sinnvoll erscheint.

Zudem zeigten Trainings zum Fördern der Argumentationskompetenz beim Stellen von Diagnosen (*medical reasoning*), dass Konzepte fallbasierten Lernens mit instruktionalen Ansätzen wie *Cognitive Apprenticeship* förderlich beim Erlernen der relevanten Strategien sein können (Eva, 2004; Gräsel & Mandl, 1993)

## **4.2. Ansätze aus der internationalen Lehrerbildung zur Förderung der relevanten Kompetenzen zur Nutzung von Evidenz**

Nachdem in Kapitel 4.1. Ansätze zum Fördern der Kompetenzen zum kritischen Hinterfragen und (argumentativen) Nutzen wissenschaftlicher Evidenz aus anderen Domänen beschrieben wurden, wird nun im Folgenden näher auf erste Förderansätze innerhalb der Lehrerbildung eingegangen. Dabei werden zunächst aktuelle Interventionen im internationalen Raum dargelegt und im Anschluss daran die Nutzbarkeit und Übertragbarkeit dieser diskutiert.

Wie bereits angedeutet, wurden in jüngster Zeit auch im Bereich der Lehrerbildung erste Trainings- und Interventionsansätze entwickelt und getestet. Einige davon werden im Folgenden exemplarisch dargestellt. So untersuchten Csanadi et al., (2016; 2017) in ihrer Interventionsstudie unter anderem, ob kooperatives Lernen das wissenschaftliche Argumentieren zukünftiger Lehrerinnen und Lehrer fördern kann. Der Fokus wurde dabei auf das Messen der epistemischen Aktivitäten (siehe Kapitel 2.3.) der Teilnehmenden gelegt. Ihre Studien konnten zeigen, dass dyadisches Argumentieren komplexere Vorgänge wissenschaftlichen Argumentierens hervorbringen kann als individuelles. Dies stützt somit die Ergebnisse der Ansätze aus der Science Education.

Ein weiterer Trainingsansatz wurde von Iordanou & Constantinou (2014) entwickelt und hatte zum Ziel, evidenzbasiertes Argumentieren bei zukünftigen Lehrkräften bezüglich sozialwissenschaftlicher Fragestellungen zu fördern. Dieser Ansatz basiert auf dem Modell zur *meta-level awareness* (Kuhn, 2001), welches besagt, dass Metawissen bezüglich des Nutzens von Evidenz besonders in Diskussionen zu einer Erhöhung der Anzahl evidenzbasierter Argumente führen kann. Aus diesem Grund wurden die Studierenden mit Hilfe von *prompts* dazu aufgefordert die geführten Diskussionen im Anschluss dahingehend zu reflektieren, ob sie bei ihren Argumenten, beziehungsweise Gegenargumenten Evidenz verwendet haben und welche Qualität diese hatte. Darüber hinaus sollte daran gearbeitet werden, wie

man diese Argumente und Gegenargumente noch weiter hinsichtlich eines qualitativ hochwertigen Nutzens von Evidenz, verbessern könnte. Als Lernumgebung wurde dafür eine webbasierte Plattform für „*reflective, inquiry-based teaching and learning*“ entwickelt und bereitgestellt. Mit Hilfe dessen absolvierten die Studierenden ein einwöchiges Programm mit thematischen Einführungen und praktischen Übungen in Form von Diskussionen (Iordanou & Constantinou, 2014).

Des Weiteren entwickelten Wagner et al. einen Trainingsansatz zur Theorienutzungs-kompetenz bei Lehramtsstudierenden. Ziel war es hierbei im Rahmen eines erweiterten Seminars das wissenschaftliche reflektierte Handeln Lehramtsstudierender zu fördern. Hierfür wurde eine auf *advokatorischen Fehlern* aufbauende computerbasierte Lernumgebung entwickelt. Der *blended learning* Ansatz wurde zusätzlich durch nachlassende tutorielle Unterstützung (*fading*), sowie durch fehlerbasiertes kooperatives Lernen unterstützt. Mit Hilfe dieses Trainings konnte die praktische Anwendung bildungswissenschaftlichen Wissens gefördert werden. (Klein et al., 2015; Wagner et al., 2014).

Folglich gibt es im internationalen Raum bereits erste erfolgreiche Ansätze hinsichtlich der Entwicklung von Trainings zur Förderung des kompetenten (argumentativen) Nutzens von Evidenz bei Lehramtsstudierenden. Als besonders förderlich hat sich hierbei sowohl der Einsatz von (dyadischen) Diskussionen, als auch das Nutzen von nachlassender tutorieller Unterstützung erwiesen. Diese instruktionalen Prinzipien sind auch auf die Förderung der in dieser Arbeit fokussierten Kompetenzen der kritischen Hinterfragens und im Besonderen des argumentativen Nutzens von Evidenz übertragbar und nutzbar. Darüber hinaus gibt es aus den Bereichen der Science Education und der Evidenzbasierten Medizin (siehe Kapitel 4.1.) etablierte Trainingsprogramme zur Förderung des Verständnisses von Wissenschaft und (empirischer) Forschung und somit der Fähigkeit des kritischen Bewertens und Hinterfragens von Evidenzquellen.

### **4.3. Fokus auf Heuristiken**

Folglich gibt es in anderen Domänen bereits wirksame Ansätze, um Kompetenzen, die für den professionellen Umgang mit Evidenz essenziell sind, zu trainieren und zu fördern. Häufig werden dabei zusätzlich so genannte *Heuristiken* eingesetzt. Im folgenden Kapitel soll nun erörtert werden wie der Begriff *Heuristik* in diesem Zusammenhang zu definieren ist und inwieweit man diese gezielt innerhalb eines Trainings für Lehramtsstudierende zum professionellen Nutzen von Evidenz einsetzen kann. Darüber hinaus soll ein Überblick darüber gegeben werden, welche

Heuristiken im speziellen für ein Training zum kompetenten Nutzen von Evidenz herangezogen werden können.

#### **4.3.1. Begriffsdefinition Heuristik**

Im weiteren Sinne bezeichnet der von dem griechischen Wort *εὐρίσκειν* (*heuriskein*, zu Deutsch: *entdecken*) abgeleitete Terminus *Heuristik* „...*Begriffe und Methoden, die unsere Erkenntnis erweitern, aber selbst noch keine Beweise oder Lösungen geben [...] Es sind Annahmen, Arbeitshypothesen, vermutete Zusammenhänge oder Modelle, die einen heuristischen Wert als Entwürfe oder Finderegeln haben.*“ (Wirtz, 2017). In diesem spezifischen Kontext bezeichnen Heuristiken hilfreiche Daumenregeln und andere Strategien, welche es Laien ermöglichen sollen, trotz eingeschränkten Hintergrundwissens, schnelle und wissenschaftlich fundierte Entscheidungen treffen zu können (Bromme, 2014).

#### **4.3.2. Implementierung von Heuristiken in Kurzzeitinterventionen für Lehramtsstudierende**

Das Lehramtsstudium in Deutschland besteht aus einer großen Bandbreite komplexer Inhalte, die sowohl fachlicher (die Inhalte der angestrebten Unterrichtsfächer) als auch erziehungswissenschaftlicher Art sind. Dies führt dazu, dass für den Erwerb neuer Kompetenzen das bereits bestehende, sehr umfangreiche und enge Lehramtscurriculum beachtet werden muss (Bauer et al., 2012; Lohmann et al., 2011). Demzufolge ist es kaum möglich einen umfassenden Kompetenzerwerb im Sinne von Handlungsmodellen zur evidenzbasierten Praxis zu implementieren (Rubin, 2008). Folglich können Heuristiken eine zielführende Alternative sein, denn dadurch können Lehramtsstudierende dabei unterstützt werden, zu einer Lösung zu gelangen, die zwar nicht perfekt, aber dennoch gut genug für die jeweilige individuelle Situation ist. Dass diese Methode hilfreich sein kann, zeigen ähnliche Trainings aus den Bereichen der Science Education (Fischer et al., 2014; Gigerenzer, 2008; Hutchinson & Gigerenzer, 2005; Kopp & Mandl, 2011; Kuhn, 2010; Osbourne, 2010) sowie der Evidenzbasierten Medizin (Finkel et al., 2003; Higgins & Green, 2011; Howick et al., 2011; Ilic et al., 2015; Malhotra et al., 2015) (vgl. Kapitel 4.1.).

#### **4.3.3. Modelle zur kritischen Bewertung von Evidenz**

Da es auf Grund der Komplexität und Diversität des Forschungsfeldes in der Lehrerbildung kaum möglich erscheint, ein enges Ratingsystem, wie es beispielsweise in der Medizin verwendet wird, zur kritischen Einschätzung von Evidenz heranzuziehen, müssen diese für eine Anwendung im Lehreralltag generalisiert und angepasst werden. Im Folgenden werden deshalb zunächst

weitere Möglichkeiten für Heuristiken zur kritischen Bewertung beschrieben, welche sich in unterschiedlichen Domänen, neben den Cochrane Evidenzklassen und dem GRADE Schema (siehe Kapitel 2.2.) entwickelt haben.

### *Jadad-Skala*

Die Jadad-Skala (auch Jadad Score oder Oxford-Skala genannt) ist ein in der Medizin entwickeltes Ratingschema zur kritischen Bewertung klinischer Interventionsstudien. Sie beinhaltet fünf dichotome Fragen (Ja/Nein) deren Beantwortung mit Hilfe eines Punktesystems ausgewertet werden kann (Jadad et al., 2000). Zu stellen sind hierbei die folgenden Fragen:

„Wurde die Studie als randomisiert beschrieben?“ (Ja: +1 Punkt; Nein: +0 Punkte)

„War die Randomisierung sachgerecht?“ (Ja: +1 Punkt; Nein: -1 Punkt)

„Wurde die Studie als doppelblind beschrieben?“ (Ja: +1 Punkt; Nein: +0 Punkte)

„War die Verblindung sachgerecht?“ (Ja: +1 Punkt; Nein: -1 Punkt)

„Wurden Ausfälle begründet?“ (Ja: +1 Punkt; Nein: +0 Punkte)

(EBN Netzwerk Österreich, 2007; Jadad et al., 2000) Erreicht eine Studie auf dieser Skala einen Wert von unter 3 Punkten, wird sie als qualitativ schlecht eingestuft und somit von einer Verwendung der Studie als Evidenzquelle abgeraten. Um systematische Fehler bei der Einschätzung zu vermeiden wird vorgegeben, dass jede Studie von mindestens zwei Ratern bewertet werden sollte. (Jadad et al., 2000; Oremus et al., 2001). Auf Grund der Beschränkung auf Randomisierung, Verblindung und Drop-outs ist diese Skala jedoch nur sehr eingeschränkt innerhalb der Bildungsforschung anwendbar, da sich die Forschungslage diesbezüglich deutlich von der Medizin unterscheidet.

### *MAGIC*

Die vom Psychologen Robert Abelson entwickelten *MAGIC* Kriterien (1995) beziehen sich auf die Einschätzung von statistischen Analysen und die daraus folgende Aussagekraft hinsichtlich der Relevanz einer Studie, sowie der Möglichkeit zur Generalisierung auf die jeweilige Population. *MAGIC* ist hierbei ein Akronym aus den Begriffen *Magnitude*, *Articulation*, *Generality*, *Interestingness* und *Credibility*. Zu jedem Begriff gibt es je eine spezifische Fragestellung um die Qualität wissenschaftlicher Ergebnisse einschätzen zu können:

*Magnitude*: Wie hoch sind die beobachteten Effekte? (Große Effektstärken sind kleineren vorzuziehen)

*Articulation:* Wie detailliert sind die Befunde berichtet? (Präzise Informationen sind hierbei als ideal anzusehen)

*Generality:* In welchem Maße sind die Ergebnisse generalisierbar? (Effekte, die zu einem höheren Maß generalisierbar sind und somit eine breitere Zielgruppe betreffen sind hierbei vorzuziehen)

*Interestingness:* Wie überraschend sind die Befunde und hinterfragen sie geltende Überzeugungen? (Je höher die Interessantheit und Neuartigkeit der Ergebnisse desto höher die qualitative Einschätzung)

*Credibility:* Wie glaubwürdig und belastbar sind die auf Basis der Ergebnisse getroffenen Aussagen? (Sind die Aussagen methodisch solide und inhaltlich kohärent?)

(Abelson, 1995). Diese Kriterien sind deutlich offener gehalten und somit weniger domänenspezifisch. Dadurch sind sie leichter auf andere Bereiche übertragbar und wurden beispielsweise in den Wirtschaftswissenschaften (Qian, 2014), sowie der Statistik (Boomsma, 2000) bereits erfolgreich eingesetzt. Eine Übertragbarkeit auf die Bildungswissenschaften scheint deshalb praktikabel.

#### *Elsevier Reviewers Guide*

Des Weiteren enthalten die Richtlinien für Reviewer von Fachzeitschriften zur Bewertung einer Studie hilfreiche Anhaltspunkte, welche auch in der Praxis genutzt werden können. Im Folgenden werden beispielhaft die Kriterien des Elsevier Verlags dargestellt. Diese unterteilen sich in die Bereiche Originalität, Organisation und Klarheit, sowie ethische Grundsätze:

*Originalität:* Vergleichbar zum Aspekt *Interestingness* von Abelson soll die Neuartigkeit der Ergebnisse bewertet werden und somit der *added value*, welcher den Wert und die Bedeutung der Ergebnisse für die jeweilige wissenschaftliche Community bezeichnet.

*Organisation und Klarheit:* Im Rahmen dieses Kriteriums soll die strukturelle Aufarbeitung eines Artikels bewertet werden. Positiv zu bewerten ist es beispielsweise, wenn der Titel passend und ansprechend gewählt ist, der Abstract alle relevanten Informationen enthält, die Einleitung auf das zu untersuchende Problem hinführt und einen Überblick über den aktuellen Forschungsstand gibt. Zudem sollen die Hypothesen und Forschungsfragen klar und verständlich dargelegt sein, sowie eine detaillierte Beschreibung derer Methoden, die zur Beantwortung der Forschungsfragen verwendet wurden. Überdies soll im Rahmen der Möglichkeiten geprüft werden, ob die statistischen Methoden geeignet sind, um die Ergebnisse in der vorliegenden

Art darzustellen und ob alle relevanten Ergebnisse objektiv und neutral berichtet wurden. Darüber hinaus sollen innerhalb der Diskussion alle Behauptungen durch die Ergebnisse belegbar sein und Anknüpfungspunkte für zukünftige Forschung daraus erkennbar werden.

*Ethische Grundsätze:* Final ist zu bewerten ob alle ethischen Grundsätze im Rahmen der Datenerhebung eingehalten wurden. Häufig wird sich dabei an der in der Medizin entwickelten *Declaration of Helsinki* (WMA, 2000) orientiert.

(Elsevier, 2018). Da auch diese Kriterien nicht domänenspezifisch sind, scheint eine Übertragbarkeit auf bildungswissenschaftliche Studien möglich.

### *Glaubwürdigkeitsbeurteilung*

Im Rahmen einer Studie zur „Glaubwürdigkeitsbeurteilung medizinischer Fachinformationen im Internet durch Laien“ wurden von Wittwer et al. (2004) sieben zentrale Dimensionen der Glaubwürdigkeit identifiziert. Diese zeigen an, auf welcher Basis fachliche Laien die Glaubwürdigkeit einer Informationsquelle einschätzen. Diese Dimensionen sind:

1. *Aussageinhalt* (die Art und Weise, wie der Inhalt und die Aussagen präsentiert sind)
2. *Vermittlungskompetenz* („der journalistisch sorgfältige Umgang mit Mitteilungen“)
3. *Objektivität der Berichterstattung*
4. *Unparteilichkeit*
5. *Vollständigkeit*
6. *Verständlichkeit der Informationen*
7. *Das Kommunikationsmedium selbst*

(Wittwer et al., 2004). Diese Dimensionen können hilfreich sein um zu verstehen, wie Laien (wissenschaftliche) Texte tatsächlich bewerten. In Folge dessen kann dies als Anknüpfungspunkt für weitere Kriterien, welche innerhalb eines Trainings erlernt werden, dienen.

### *Fazit*

Ähnlich wie die *Cochrane Evidenzklassen* und das *GRADE* Rating Schema (Kapitel 2.2.) ist die *Jadad-Skala* auf Grund der Fokussierung auf die in der Medizin als Goldstandard angesehenen randomisierten Kontrollstudien nur begrenzt auf die Bildungswissenschaften oder ähnliche Domänen generalisierbar. Jedoch sind einzelne Aspekte wie das Bewerten einer forschungsmethodisch angemessenen

Durchführung, sowie die Wichtigkeit einer Begründung von Abweichungen wie etwa Drop-outs durchaus übertragbar. Daran anknüpfend liefern auch die *MAGIC* Kriterien hilfreiche Fragen zur kritischen Bewertung der forschungsmethodischen Qualität einer Studie. Dadurch, dass diese Kriterien im Gegensatz zur *Jadad Skala* nicht domänenspezifisch sind, können diese domänenübergreifend für empirische Studien unterschiedlicher Bereiche und folglich auch auf die Bildungswissenschaften angewendet werden. Um darüber hinaus auch Aussagen hinsichtlich der Qualität des Berichtens von Studienergebnissen treffen zu können, können Richtlinien etablierter Fachzeitschriften für Reviewer, wie etwa dem *Elsevier Reviewers Guide*, hilfreiche Strategien bereitstellen. Darüber hinaus können Informationen zur Glaubwürdigkeitsbeurteilung wissenschaftlicher Studien durch Laien und somit Informationen dahingehend, wann Laien wissenschaftliche Texte als besonders glaubwürdig erachten, ein wichtiger Anknüpfungspunkt zur Vermittlung von Heuristiken zum kritischen Hinterfragen wissenschaftlicher Evidenz sein. Folglich können die in diesem Kapitel beschriebenen Konzepte als Hintergrund dienen um Heuristiken zur kritischen Bewertung (bildungswissenschaftlicher) Studien für Lehramtsstudierende sowohl inhaltlich, als auch strukturell zu entwickeln.

Diese Heuristiken können für Lehramtsstudierende einerseits als Strategien zum effizienten bewerten wissenschaftlicher Studien in Form von Daumenregeln genutzt werden und andererseits auch gleichzeitig als eine Art „Eisbrecher“ dienen, um Studierende mit der Nutzung von Evidenz vertraut zu machen und Vorurteile hinsichtlich der zeitlichen Realisierbarkeit des Nutzens von Evidenz in der Praxis abzubauen. Ziel ist es, die Studierenden durch das Abbauen dieser Ressentiments und anderer wahrgenommener Barrieren der Studierenden zu ersten Schritten in Richtung einer evidenzbasierten Praxis zu motivieren. Dieses Ziel basiert auf zahlreichen Studien, welche gezeigt haben, dass Lehramtsstudierende oft Probleme damit haben empirische Studien richtig zu durchdringen und darüber hinaus häufig negative Einstellungen bezüglich Relevanz und praktischem Nutzen bildungswissenschaftlicher Evidenz haben (Hetmanek et al., 2015; Hargreaves, 2000; Kühle & Peek, 2007; Parr et al., 2008; Schildkamp et al., 2010; Williams & Coles, 2008; Zembal-Saul et al., 2003). Ähnliche Ergebnisse gibt es diesbezüglich auch aus der Medizin – auch hier haben viele Studierende Verständnisprobleme, sowie negative Einstellungen bezüglich des Nutzens von Evidenz (Donner-Banzhoff, 2007; Johnston et al., 2003; Young & Ward, 2001).

## 4.4. Instruktionale Prinzipien

Im folgenden Kapitel soll näher auf die Hintergründe der in Kapitel 4.2. als wirksam identifizierten instruktionalen Prinzipien und ihre Anwendbarkeit in einem Training zur Förderung der Kompetenzen zum kritischen hinterfragen und argumentativen Nutzen von Evidenz für Lehramtsstudierende eingegangen werden. Um sowohl die Akzeptanz, als auch die Kompetenz zum professionellen Umgang mit Evidenz zu fördern, werden drei instruktionalen Schwerpunkte gesetzt: (1) Fall- und problembasiertes Lernen, (2) kooperatives und dialogisches Lernen, sowie (3) kognitive Modellierung durch *cognitive apprenticeship*.

### 4.4.1. Problembasiertes und fallbasiertes Lernen

*Problem-basiertes Lernen (PBL)* ist eine Lernform, die darauf abzielt die Lernenden auf Problemkontexte der realen Welt vorzubereiten, indem sie dazu angeleitet werden weitestgehend selbständig Lösungsansätze für ein instruiertes Problem zu erarbeiten (Barrows, 1996; Hung et al., 2008; Jonassen, 2000). Die Darbietung authentischer Fallvignetten kann hierfür zusätzlich unterstützend sein (Zumbach et al., 2008). Ein problembasierter Einstieg in eine neue Thematik, mit Hilfe eines konkreten, alltagsrelevanten Fallbeispiels, kann sich zudem positiv auf Motivation und Akzeptanz der Teilnehmenden auswirken (Kahn & Coomarasamy, 2006; Fischer et al., 2014). Darüber hinaus kann durch die damit einhergehende Förderung der essentiellen kognitiven Voraussetzungen der Transfer zu professionsrelevanten Problemstellungen erleichtert werden (Coomarasamy & Kahn, 2004; Gräsel, 2006; Hmelo-Silver & Barrows, 2006; Mähler & Stern, 2010). Dies kann durch den Einsatz von problembasierten Fallvignetten zusätzlich unterstützt werden (Gräsel, 1997). Hierbei ist zu beachten, dass die Fallvignetten so konstruiert sind, dass sie für die Lernenden auf Grund ihrer Interessantheit intrinsisch motivierend sind, um eine intensive Auseinandersetzung mit den Lerninhalten zu erzielen (Deci & Ryan, 1993; Stark & Mandl, 2000). Überdies zeigten frühere Studien, dass problembasierte Fallvignetten die Lernenden dabei unterstützen, ihre Fähigkeiten zum kritischen Hinterfragen von Informationen weiterzuentwickeln (Savery, 2006).

### 4.4.2. Kooperatives und dialogisches Lernen

Besonders in Trainings, welche die Entwicklung und Förderung von (evidenzbasierten) Argumentationskompetenzen beinhalten, hat sich die Implementierung kooperativer Methoden als besonders hilfreich erwiesen (Fischer et al., 2014; Iordanou et al., 2014; Kopp et al., 2011; Kuhn, 2010; Osbourne, 2010) (siehe Kapitel 3.1. und 3.2.). *Kooperatives Lernen* hat zum Ziel, dass die Lernenden

in Form von Gruppenarbeiten gemeinsam neue Inhalte erarbeiten und sich dabei gegenseitig unterstützen. Dadurch können sowohl fachliche als auch soziale Kompetenzen maßgeblich gefördert werden (Green & Green, 2005). Mit Hilfe von kooperativem Lernen ist es zudem möglich, auch bei sehr heterogenen Gruppenzusammensetzungen (was bei Lehramtsstudierenden aufgrund der unterschiedlichen fachlichen Hintergründe meist der Fall ist) individuelle Lernfortschritte beziehungsweise einen individuellen Kompetenzzuwachs zu erreichen (Slavin, 1995). Speziell im Kontext evidenzbasierter Praxis können kooperative Lerneinheiten dabei unterstützen, verschiedene Perspektiven einzunehmen, sowie mit kognitiven Konflikten umzugehen und somit einen Perspektivenwechsel erleichtern (Vygotsky, 1978). Dies wiederum ist eine grundlegende Voraussetzung für den kompetenten Umgang mit den fragilen und konfligierenden Elementen von Evidenz (Bromme, 2008; Desimone, 2009; Gräsel et al., 2006; Meier et al., 2007).

Eine weitere etablierte Methode um eben diese Kompetenzen zu fördern ist *dialogisches Lernen*. Dialoge ermöglichen es den Lernenden, die Perspektiven anderer in Betracht zu ziehen und dadurch die eigene kritisch zu hinterfragen, was essentiell ist für die Bewertung von Wahrheitsansprüchen (Paul, 1992) und somit im weiteren Sinne auch für ein erfolgreiches Nutzen von Evidenz. Darüber hinaus zeigen zahlreiche Interventionsstudien, dass ein dialogischer Ansatz zu einer gesteigerten Motivation, sowie einer erhöhten aktiven Teilnahme der Lernenden führen kann (z.B. Chinn et al., 2001; Osborne et al., 2013; Pehmer et al., 2015). Dies kann beispielsweise durch angeleitete Gruppendiskussionen realisiert werden. Die Anleitung dieser kann dabei idealerweise sowohl durch tutorielle Unterstützung, als auch durch den Einsatz von *Kooperationsskripts* erfolgen. *Kooperationsskripts* fungieren als eine Art Gerüst (*scaffolding*) zur Strukturierung kooperativen Arbeitens. Durch eine klare Definition der einzelnen Aufgabenschritte, sowie der jeweiligen Rollen innerhalb der Gruppe können typische Probleme, wie beispielsweise eine ungleiche Verteilung der Arbeitslast oder das Auftreten von „Trittbrettfahrern“ vermieden werden. Auch auf unterschiedliche Voraussetzungen, beispielsweise aufgrund unterschiedlichen Vorwissens, kann dadurch eingegangen werden und somit der Lerneffekt für alle Gruppenmitglieder erhöht werden (Ertl & Mandl, 2004). Darüber hinaus kann auf diese Weise eine hohe Qualität der Ergebnisse sichergestellt werden (Fischer et al., 2014; Kollar et al., 2006).

#### 4.4.3. Kognitive Modellierung durch tutorielle Unterstützung

Kognitive Modellierung durch abnehmende tutorielle Unterstützung ist eine wirkungsvolle Methode um Studierende beim Erlernen und Erproben der komplexen epistemischen Prozesse wissenschaftlichen Denkens zu unterstützen (Collins et al., 1989; Iordanou et al., 2014; Kopp et al., 2011; Schworm & Renkl, 2007; Zembal-Saul et al., 2003). Bei der kognitiven Modellierung werden komplexe Arbeitsschritte von einem Experten vorgeführt, welcher dabei *laut denkt* und somit die Lernenden an allen damit einhergehenden Prozessen seiner kognitiven Verarbeitung teilhaben lässt (Collins et al., 1989; Seidel & Krapp, 2014). Durch das Teilen eben dieser beim Lösen komplexer Problemfälle kann sowohl das allgemeine Verständnis, als auch der Kompetenzerwerb gefördert werden (Seidel & Krapp, 2014). Darüber hinaus kann hierdurch eine Basis für darauf aufbauendes, zunehmend selbstgesteuertes und kooperatives Lernen geschaffen werden (Collins et al., 1989; Johnston et al., 2003; Meats et al., 2009). Frühere Studien konnten zudem zeigen, dass Trainingsprogramme, die kognitive Modellierung enthalten, zu einem höheren Kompetenzzuwachs führen können als Trainings die sich ausschließlich auf selbstgesteuertes Lernen stützen (Johnston et al., 2009).

Eine Möglichkeit zur tutoriellen Unterstützung ist der *Cognitive Apprenticeship* Ansatz. Diese Methode ist angelehnt an das praktische Lernen innerhalb einer Lehrling-Meister-Beziehung, in welcher anhand konkreter Fallbeispiele Problemstellungen gelöst werden. Dabei übernimmt der „Meister“ zu Anfang den führenden Anteil der Arbeit und nimmt sich im Laufe der Zeit mehr und mehr zurück. Um dies zu erreichen basiert *Cognitive Apprenticeship* auf insgesamt vier unterschiedlichen Phasen der Unterstützung. Am Anfang steht die Modellierungsphase (*modeling*), in welcher, wie bereits oben beschrieben, durch einen hohen Steuerungsanteil des Lehrenden, sowie durch Verbalisierung seiner Denkprozesse, der Lernende dabei unterstützt wird sich erstmals mit den Problemstellungen auseinanderzusetzen. Darauf folgt die Betreuungsphase (*coaching*), in welcher der Lernende nun mit eigenen Problemstellungen konfrontiert wird. Bei der Lösung dieser wird er von dem Lehrenden weitgehend unterstützt und beraten. Auf Basis dessen kann in die Begleitphase (*scaffolding*) übergeleitet werden. Hierbei arbeitet der Lernende zunehmend selbstgesteuert und bearbeitet nun auch komplexer werdende Fälle. Der Lehrende steht ihm dabei beratend zur Seite und unterstützt somit den Lernenden darin, ein eigenständiges Grundgerüst für zukünftige Aufgaben zu entwickeln. Die letzte Phase leitet das vollständige Zurücktreten des Lehrenden ein (*fading*). Der Lernende arbeitet nun selbstgesteuert darauf hin ohne Unterstützung des Lehrenden agieren zu können, was sowohl zu

einem weiteren Kompetenzzuwachs, als auch zu einem Anstieg des Selbstvertrauens und der Lernmotivation führen kann. Während den gesamten vier Phasen soll der Lernende zudem dabei unterstützt werden seine eigenen Denkprozesse zu verbalisieren (*articulation*), über neu Erlerntes zu reflektieren (*reflection*), sowie die neu erlernten Fähigkeiten eigenständig und praktisch anzuwenden (*exploration*) (Collins et al., 1989; Gräsel, 1997; Iordanou et al., 2014; Kopp et al., 2011; Schworm & Renkl, 2007; Seidel & Krapp, 2014; Wirtz et al., 2017; Zembal-Saul et al., 2003). Vor allem in der Medizin hat sich der Einsatz dieser Methode bereits vielfach bewährt (Gräsel et al., 1997).

#### **4.5. Fazit**

Ziel dieses Kapitels war es, einen Überblick über etablierte Trainingsprogramme und Interventionen zur Förderung der relevanten Kompetenzen zum professionellen Nutzen wissenschaftlicher Evidenz aus unterschiedlichen Domänen zu generieren. Auf Basis dessen wurde diskutiert, welche Inhalte dieser Ansätze für ein Training für Lehramtsstudierende zum kritischen Hinterfragen und argumentativen Nutzen von Evidenz adaptiert und genutzt werden können. Anschließend wurden instruktionale Prinzipien und Strategien, wie beispielsweise das Einbinden von Heuristiken oder das Einsetzen problemorientierten Lernens, welche sich in ähnlichen Bereichen bereits als förderlich und motivierend erwiesen haben, erläutert. Ziel war es dabei, Informationen dahingehend zu generieren, wie ein Training für Lehramtsstudierende zum kompetenten Nutzen von Evidenz, welches flexibel in bestehende Lehramtscurricula integrierbar ist, möglichst effektiv und motivierend gestaltet werden kann. Zusammenfassend lässt sich sagen, dass etablierte Trainingsansätze anderer Domänen zeigen, dass die Einbindung von Heuristiken, fallbasiertem und problemorientierten Lernens, kooperativer Lernformen, sowie kognitiver Modellierung mit abnehmendem tutoriellen Support einen positiven Einfluss auf den Kompetenzerwerb im Rahmen evidenzbasierter Praxis und wissenschaftlichen Denkens haben kann.

## 5. Training zum kompetenten Nutzen von Evidenz

Die in Kapitel zwei bis vier beschriebenen Inhalte bilden die theoretische Grundlage für die Entwicklung eines flexibel in bestehende Lehramtscurricula integrierbaren Trainings zur Förderung der Kompetenzen zum kritischen Hinterfragen und argumentativen Nutzen von Evidenz. Im folgenden Kapitel werden der Aufbau sowie der inhaltlichen Schwerpunkte und die hierfür verwendeten instruktionalen Prinzipien erläutert.

### 5.1 Aufbau und Inhalt

Die Dauer des Trainings beträgt 2 Tage zu je circa sieben Stunden. Inhaltlich kann das Training dabei in drei Hauptkomponenten aufgeteilt werden: (1) das Fördern der Motivation zum Nutzen von Evidenz durch Erklärungen und Demonstrationen hinsichtlich der Relevanz dessen, (2) Heuristiken zum kritischen Bewerten von Evidenz und (3) Heuristiken zum argumentativen Nutzen von Evidenz (siehe Tabelle 4).

Tabelle 4. Inhaltlicher Aufbau des Trainingsprogramms.

Motivation und Verständnis der Relevanz evidenzbasierter Praxis		Heuristiken zum kritischen Bewerten von Evidenz		Heuristiken zum argumentativen Nutzen von Evidenz	
Problemorientierter Einstieg mit alltagsrelevanten Fallbeispielen	Transfer zu professionsrelevanten Themen aus der Bildungsforschung	Allgemeine Bewertungskriterien und forschungsmethodische Grundlagen	Tutoriell angeleitete Fallbearbeitung (beide Domänen)	Grundlagen evidenzbasierter Argumentierens, Umgang mit Fragilität und Konflikthaftigkeit	Gruppendiskussionen mit Fallbeispielen beider Domänen

Der erste Teil (*Motivation und Verständnis*) begann mit einem problembasierten Einstieg zu in den Medien bekannten Inhalten, die durch falsche Durchführung von Studien oder Missinterpretationen von Daten falsche Informationen liefern. Verwendete Themen waren beispielsweise der „Supersize me“ Film, typische

fehlinterpretierte Unfall-Statistiken aus Automobilzeitschriften, Zeitungsschlagzeilen zu „Anstieg der Kriminalität durch Zuwanderung“, oder der „Gefährlichkeit des Nutzens von Computern und anderen Medien für Kinder und Jugendliche“. Ziel war es von Beginn an eine motivierende Lernatmosphäre zu schaffen und mit Hilfe von Themenbereichen, welche persönliche und alltägliche Problemkontexte der Teilnehmenden ansprechen, eine positive Einstellung in Bezug auf evidenzbasierte Praxis zu fördern. Über das gesamte Training hinweg wurden deshalb alternierend alltagsrelevante (gesundheitswissenschaftliche) und professionsrelevante (bildungswissenschaftliche) Fallbeispiele genutzt. Durch die Abwechslung alltags- und professionsrelevanter Fallbeispiele können Motivation und Akzeptanz gesteigert werden, da anzunehmen ist, dass das Interesse sich mit Inhalten zu befassen, welche die persönliche, private Lebenswelt betreffen, höher ist als bei professionsrelevanten Themen (Prenzel & Bauer, 2012). Ziel war es zudem, mit Hilfe dieser alltagsrelevanten Themen den Transfer zu berufsrelevanten Themen zu erleichtern.

Auch im zweiten Teil (*kritisches Bewerten*) wurde deshalb abwechselnd mit gesundheitswissenschaftlichen und bildungswissenschaftlichen Studien gearbeitet. Dieser Teil des Trainings diente als eine Art forschungsmethodischer Crashkurs, da Lehrkräfte die evidenzbasierte Praxis betreffend überwiegend als Laien eingestuft werden können (Schildkamp & Kuiper, 2010). In diesem wurde den Studierenden auf Basis von Heuristiken das Wissen vermittelt, dass sie mindestens benötigen um empirische Studien verstehen und bis zu einem gewissen Maß auch kritisch bewerten zu können. Grundlage für die Heuristiken zur Teilkompetenz *Evidenz kritisch bewerten (appraise)* basieren auf Literatur zur Einführung in Statistik und empirischen Forschungsmethoden (Bühner & Ziegler, 2009; Greenhalgh, 2010; Oestreich & Romberg, 2012; Scherfer & Bossmann, 2011). Darüber hinaus wurden bereits etablierte Heuristiken (Abelson, 1995) und Konzepte (Wittwer et al., 2004), sowie auf Empfehlungen für Gutachter internationaler Fachzeitschriften (Elsevier, 2018; Sternberg, 2005) implementiert (siehe Kapitel 4.3.). Sie helfen dabei Fragen zu beantworten wie beispielsweise:

*Ist die Fragestellung klar und präzise formuliert?*

*Welche Daten werden erhoben und passen diese zur Fragestellung?*

*Ist das Design geeignet um die Fragestellung zu beantworten?*

*Sind die Zusammensetzung, sowie die Größe der Stichprobe angemessen und dadurch repräsentativ?*

*Gibt es Aussagen bezüglich der Testgütekriterien?*

*Sind die Ergebnisse signifikant und wenn ja, sind sie dies in erwarteter Richtung?*

*Gibt es mindestens ein berichtetes Ergebnis pro Fragestellung?*

*Wie groß sind eventuelle Effektstärken?*

*Werden in der Diskussion Limitationen adäquat genannt?*

*Werden alle Fragestellungen beantwortet und sind diese auf Basis der Ergebnisse angemessen?*

*Gibt es unzulässige Generalisierungen bezüglich Kausalität, Population oder Situation?*

*Ist die Relevanz erkennbar?*

Mit Hilfe der Beantwortungen dieser Fragen lässt sich eine erste Einschätzung bezüglich Glaubwürdigkeit und Belastbarkeit der vorliegenden Studie treffen, ohne dabei über tieferegehende statistische und forschungsmethodische Kenntnisse zu verfügen. Instruktoral wurde dies durch *Cognitive Apprenticeship* und einen abnehmenden tutoriellen Support unterstützt.

## **5.2. Gruppendiskussionen**

Im dritten Teil des Trainings (*argumentatives Nutzen*) lag der Fokus auf dem argumentativen Nutzen wissenschaftlicher Evidenz. Hierfür wurden Heuristiken, welche auf Forschung zu wissenschaftlichem Argumentieren basieren (Booth et al., 2008; Khine, 2012; Kuhn et al., 1997; Wilson, 2005) eingesetzt. Ziel dieser Heuristiken ist es, die Lehramtsstudierenden dabei zu unterstützen eine Behauptung (*claim*) zu verfassen und diese in eine sinnvolle Verbindung zu einer wissenschaftlichen Begründung (*backing*), sowie einem wissenschaftlichen Beleg (*evidence*) zu bringen (Booth et al., 2008; Osbourne, 2010; Toulmin, 1978). Des Weiteren sollten die Studierenden dabei unterstützt werden, nicht nur auf Basis von Evidenz zu argumentieren, sondern auch mit begründeter Kritik entgegen der Aussage der Evidenzquelle entscheiden zu können (Bell, 2010; Kuhn et al., 2011; Sandoval & Reiser, 2004) (vgl. Kapitel 4.2). Auch diese Heuristiken wurden analog zu denen des kritischen Hinterfragens, durch problemorientierte Fallbeispiele und *Cognitive Apprenticeship* instruiert.

Um die Anwendung dieser Heuristiken zu üben, wurden im Anschluss daran zwei Gruppendiskussionen (siehe Tabelle 5, linke Spalte) eingesetzt. Analog zum Aufbau des Trainings wurde zuerst ein gesundheitswissenschaftliches Thema diskutiert und im Anschluss daran ein bildungswissenschaftliches. Hierfür wurden die Studierenden zunächst in Gruppen von 3-5 Teilnehmenden aufgeteilt. Als Einstieg wurden erneut problembasierte Fallvignetten präsentiert. Zur Lösung der dadurch

aufgezeigten Problemstellungen erhielt jeder Teilnehmende innerhalb einer Gruppe je eine unterschiedliche wissenschaftliche Evidenzquelle in Form einer (experimentellen) Studie, einer Metaanalyse, oder eines systematischen Reviews mit konfligierenden und fragilen Elementen, sowie variierender Qualität und Belastbarkeit.

*Tabelle 5.* Überblick über die Domänen und Fallbeispiele in den Gruppendiskussionen sowie in den schriftlichen Statements

	<b>Gruppendiskussionen<sup>1</sup></b>		<b>schriftliche Statements<sup>2</sup></b>		
<i>Domäne</i>	Gesundheits- wissen- schaften	Bildungs- wissen- schaften	Bildungs- wissen- schaften	Bildungs- wissen- schaften	Bildungs- wissen- schaften
	—————→		—————→		
	<i>Transfer<sup>3</sup></i>		<i>Transfer<sup>4</sup></i>		
<i>Fall- beispiel</i>	Hypertonie und erhöhte Salzzufuhr	Geschlechter- unterschiede im Umgang mit Computern	Hausaufgaben (pro und contra)	ideale Klassengröße	Team- teaching (pro und contra)

Die erste, gesundheitswissenschaftliche Gruppendiskussion thematisierte eventuelle gesundheitliche Risiken durch eine zu hohe Salzzufuhr. Als Einstieg in die Diskussion wurde den Teilnehmenden eine Zeitungsschlagzeile der Münchner Tageszeitung (tz) mit dem Titel „Salz – tödliche Gefahr fürs Herz“ gezeigt. Um herauszufinden, ob tatsächlich Maßnahmen ergriffen werden müssen, um den Salzkonsum in der Bevölkerung zu reduzieren, erhielten die Studierenden insgesamt fünf Studien unterschiedlicher Art, die sich mit dem Zusammenhang von erhöhter Salzzufuhr und Herz-Kreislauf-Erkrankungen, wie beispielsweise Hypertonie befassen. Die Studien waren zudem von unterschiedlicher forschungsmethodischer Qualität und widersprachen sich teilweise in ihren Ergebnissen. Die zweite (bildungswissenschaftliche) Diskussion hatte zum Ziel, den Transfer zum professionellen Kontext zu schaffen. Dafür wurde den Studierenden

<sup>1</sup> Beide Fallbeispiele in eben dieser Reihenfolge.

<sup>2</sup> Je zwei von drei in randomisierter Reihenfolge.

<sup>3</sup> Transfer von den Gesundheitswissenschaften zu dem Bildungswissenschaften.

<sup>4</sup> Transfer von mündlichem Diskutieren in der Gruppe hin zu individuellen schriftlichen Statements. Nur bei der Experimentalgruppe.

ein bewusst provokatives Beispiel für eine Diskussion innerhalb einer Lehrerkonferenz gezeigt, in welcher ein Informatiklehrer anmerkte, dass Mädchen scheinbar schlechtere Leistungen im Umgang mit Computern aufweisen. Aus diesem Grund empfehle er die Einführung eines zusätzlichen Förderkurses um die vorhandenen Defizite aufholen zu können. Um einen Lösungsansatz hierfür diskutieren zu können erhielten die Teilnehmenden analog zur ersten Diskussion fünf Studien, die sich sowohl in ihrer Qualität, als auch in ihren Schlussfolgerungen unterschieden. Thematisch behandelten diese Studien Genderunterschiede in den Naturwissenschaften und speziell im Informatikunterricht. Der Grund für die Wahl dieses möglicherweise emotionsbehafteten Beispiels war einen Anreiz zu schaffen, dennoch evidenzbasiert zu argumentieren und sich nicht ausschließlich auf persönliche und anekdotische Evidenz zu stützen.

Mit Hilfe von Kooperationskripts (Kollar et al., 2006) wurden die Studierenden in beiden Diskussionen angeleitet, ihr Exemplar zunächst individuell zu bearbeiten indem die Evidenzquelle gelesen, kritisch bewertet und für die Diskussion aufbereitet wird. Anschließend wurde in der Gruppe von jedem Teilnehmenden die jeweilige Evidenzquelle den anderen hinsichtlich der wichtigsten Inhalte, sowie einer kritischen Einschätzung berichtet. Anschließend wurde auf Basis aller Berichte in einer Diskussion versucht, einen gemeinsamen Lösungsansatz mit Hilfe der vorliegenden Evidenz zu entwickeln. Die Lösungsansätze wurden anschließend im Plenum diskutiert und eventuelle Probleme und Schwierigkeiten angesprochen.

## 6. Forschungsagenda

### 6.1. Forschungsfragen

Ziel der im Rahmen dieser Arbeit durchgeführten Studien war die Entwicklung und Evaluation eines Trainings zur Förderung des kompetenten Nutzens von Evidenz bei Lehramtsstudierenden. In Studie 1 wurden deshalb im Rahmen einer Pilotierung erste Ergebnisse hinsichtlich der Effektivität des entwickelten Trainings, der allgemeinen Machbarkeit, sowie der Akzeptanz und Motivation der Teilnehmenden angestrebt. Zudem sollten erste Informationen auf mögliche Einflussfaktoren auf die relevanten Kompetenzen generiert werden. Konkret wurden in Studie 1 die folgenden Forschungsfragen gestellt:

*Forschungsfrage 1:* „Wie manifestieren sich die erlernten Heuristiken [zum kritischen Hinterfragen, sowie] zur argumentativen Nutzung von Evidenz (a) in den problembezogenen Gruppendiskussionen während des Trainings und (b) in der nachfolgenden Testsituation, in der Evidenz individuell bewertet und problembezogen argumentativ genutzt werden muss?“

*Forschungsfrage 2:* „Trägt das Training bei der Trainingsgruppe (TG) gegenüber der Kontrollgruppe (KG) zu einer gesteigerten Argumentationsqualität bei?“

*Forschungsfrage 3:* „Wie konsistent ist die [Bewertungs- und] Argumentationsqualität (a) zwischen dem kooperativen Setting (Gruppendiskussion) und der individuellen Situation im schriftlichen Test sowie (b) über die alltags- und professionsrelevanten Problemkontexte in den Gruppendiskussionen hinweg? Gelingt ein Transfer der Heuristiken?“<sup>5</sup>

*Forschungsfrage 4:* Wie ist die die Akzeptanz und Motivation der Teilnehmenden, sowie deren subjektiv wahrgenommener Lernzuwachs innerhalb des Trainings? Sind die einzelnen Inhalte allgemein verständlich?

Mit Hilfe der Ergebnisse aus Studie 1 konnten sowohl das Training, als auch die Testinstrumente weiterentwickelt werden. Ziel von Studie 2 war es, die Ergebnisse der Pilotstudie zu replizieren und durch die Durchführung der Studie in einer größeren Stichprobe detailliertere Ergebnisse hinsichtlich der Trainierbarkeit und möglicher Einflussfaktoren zu erhalten. Aus diesem Grund wurden in Studie 2 die folgenden Forschungsfragen gestellt:

*Forschungsfrage 1:* Trägt ein auf Heuristiken basierendes Kurzzeittraining bei der Trainingsgruppe gegenüber der Kontrollgruppe zu einer

---

<sup>5</sup> Forschungsfragen 1-3: Veröffentlicht in Wenglein et al., 2015

gesteigerten Bewertungs- und Argumentationskompetenz (*appraise* und *apply*) bei? Wie konsistent ist die Bewertungs- und Argumentationsqualität über verschiedene Settings und Problemkontexte hinweg?

*Forschungsfrage 2:* Welche Rolle spielen hierfür das wissenschaftstheoretische Vorwissen, die Fachausrichtung, die Überzeugungen, die Lesegewohnheiten und die Ambiguitätstoleranz der Studierenden?

Ziel von Studie 3 war es, auf Basis der Daten eines Subsamples von Studie 2 tiefergehend zu analysieren, wie genau sich erfolgreicher Nutzen von Evidenz bei Lehramtsstudierenden äußert, wo häufige Probleme und Schwierigkeiten liegen und wie auf Basis dessen ein solches Training in Zukunft noch weiter verbessert und auf die Bedürfnisse der Studierenden abgestimmt werden kann. Demgemäß wurden in Studie 3 die folgenden Forschungsfragen gestellt:

*Forschungsfrage 1:* Wie zeigt sich erfolgreicher Nutzen von Evidenz in Gruppendiskussionen?

*Forschungsfrage 2:* Was sind häufige Fehler und Schwierigkeiten beim Nutzen wissenschaftlicher Evidenz in Gruppendiskussionen?

*Forschungsfrage 3:* Was sind die Gründe, aus denen Evidenz in Gruppendiskussionen genutzt wird und welche Funktion hat die Evidenz bei erfolgreichem Nutzen?

*Forschungsfrage 4:* Welchen Einfluss hat das erfolgreiche Nutzen von Evidenz auf die Transaktivität innerhalb von Gruppendiskussionen?

## 6.2. Methodischer Ansatz

### 6.2.1. Stichproben und Designs

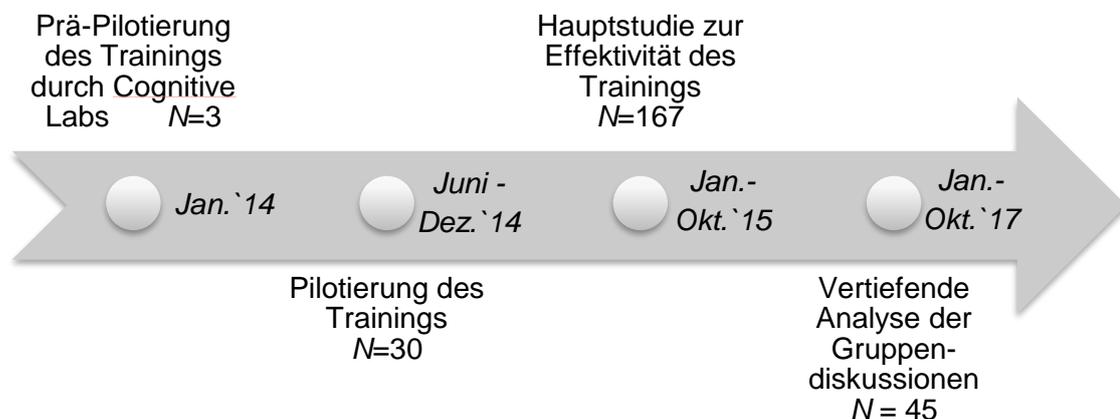
#### *Stichproben*

Die Daten wurden zu drei Studien mit insgesamt  $N = 200$  Studierenden erhoben. Dies setzt sich zusammen aus einer Prä-Pilotierung mit  $N = 3$  Teilnehmenden, der Pilotierung des Trainings mit  $N = 30$  Teilnehmenden und der Hauptstudie zur Effektivität des Trainings mit  $N = 167$  Studierenden. Die Teilnehmenden studierten alle gymnasiales Lehramt an einer der folgenden 14 Universitäten: Technische Universität München, Otto-Friedrich-Universität Bamberg, Universität Regensburg, Universität Bayreuth, Universität Würzburg, Universität Augsburg, Ludwig-Maximilians Universität München, Universität Erlangen, Universität Bonn, Universität Duisburg-Essen, Christian-Albrechts-Universität zu Kiel, Goethe-Universität Frankfurt, Universität Münster und Humboldt-Universität zu Berlin. Den Teilnehmenden wurde zu Beginn jeder Erhebung ein Code zugeteilt, der nicht mit

ihren persönlichen Daten in Verbindung zu bringen ist und sowohl bei den Fragebögen, als auch bei den Testinstrumenten inklusive des Kompetenztests und den videografierten Gruppendiskussion verwendet wurde. Zudem wurden die Teilnehmenden über die Datenschutzrichtlinien aufgeklärt und unterschrieben eine Einverständniserklärung zur Aufzeichnung der Videodaten. Sämtliche Bestimmungen und ethische Grundsätze der *Declaration of Helsinki* (MWA, 2000) wurden dadurch eingehalten.

### *Forschungsdesigns*

Das Training wurde zunächst in einer Prä-Pilotierung und einer darauf folgenden Pilotierung hinsichtlich der Machbarkeit getestet (siehe Abbildung 5). Diese Studien werden in Kapitel 7 detailliert beschrieben. Auf Basis dieser Ergebnisse wurden einzelne Details des Trainings, sowie der Auswertungsinstrumente nachjustiert und in der Hauptstudie (siehe Kapitel 8) die Effektivität der Trainings erhoben. Abschließend wurden mit Hilfe eines Subsample aus der Hauptstudie nochmals detailliertere Aspekte hinsichtlich des evidenzbasierten Argumentierens in den videografierten Gruppendiskussionen analysiert. Diese tiefere Analyse wird in Kapitel 9 dargestellt.



*Abbildung 5.* Überblick über die Zeiträume der Datenerhebung.

Sowohl die Pilotierung, als auch die Hauptstudie waren als 2 x 2 (quasi-) experimentelles Design angelegt. In der Experimentalgruppe erhielten die Teilnehmenden beider Studien zunächst einen Fragebogen zu relevanten Hintergrundvariablen, Einstellungen und Überzeugungen (siehe Tabelle 4). Darauf folgte ein kurzer Test zum wissenschaftstheoretischen Vorwissen. Anschließend

begann das zweitägige Training mit zwei daran anschließenden, videografierten Gruppendiskussionen. Im Anschluss daran wurde ein circa 90minütiger Kompetenztest durchgeführt. Abschließend wurden relevante Einstellungen und Überzeugungen nochmals in einem Fragebogen erhoben. Darüber hinaus wurden motivationale Aspekte hinsichtlich der Trainingsinhalte und –durchführung erfragt. Die Kontrollgruppe in der Pilotstudie erhielt kein Training, in der Hauptstudie erhielt die Kontrollgruppe das Training nach Durchführung des Kompetenztests.

*Tabelle 6.* Ablaufbeschreibung des Experiments in der Experimental- und Kontrollgruppe.

<b>Experimentalgruppe</b>	<b>Kontrollgruppe</b>
<i>Fragebogen (prä)</i>	<i>Fragebogen (prä)</i>
<i>Vorwissenstest</i>	<i>Vorwissenstest</i>
Training (2 Tage)	Kompetenztest
Gruppendiskussionen	Training (2 Tage)
Kompetenztest	Gruppendiskussionen
<i>Fragebogen (post)</i>	<i>Fragebogen (post)</i>

### 6.2.2. Instrumente

#### *Fragebogen*

Mit Hilfe eines Fragebogens wurden in beiden Erhebungen sowohl vor, als auch nach dem Training (siehe Tabelle 6, Fragebogen prä und Fragebogen post) relevante Informationen zu Einstellungen und Überzeugungen der Teilnehmenden erhoben. Im Folgenden werden die für die Auswertung relevantesten Skalen beschrieben.

Darunter befand sich ein Testinstrument für *Connotative Aspects of Epistemic Beliefs* (CAEB, Stahl & Bromme, 2007), mit welchem die epistemischen Überzeugungen sowohl zu Bildungs- als auch zu Gesundheitswissenschaften erhoben wurden. Die Items innerhalb dieses Tests lassen sich auf zwei Hauptfaktoren aufteilen: *texture* (Überzeugungen über die Struktur von Wissenschaft, 10 items) und *variability* (Einstellungen über die Stabilität und Falsifizierbarkeit von Wissenschaft, 7 items). Beantwortet wurden die Items mit Hilfe eines siebenstufigen semantischen Differentials. Die Reliabilität lag aggregiert über alle Bedingungen (Gesundheits- und Bildungswissenschaften; vor und nach dem Training) bei  $\alpha = .78$  (Vor dem Training:  $\alpha = .81$ .; nach dem Training:  $\alpha = .77$ ).

Zudem wurden *Überzeugungen zum Nutzen von Evidenz* erhoben. Diese wurden von bereits getesteten Skalen adaptiert (Jette et al., 2003; Johnston et al., 2003 & Young et al., 2001). Mit Hilfe dessen wurden Überzeugungen und Einstellungen hinsichtlich des Nutzens von bildungswissenschaftlicher empirischer Evidenz erhoben. Die Skala enthielt insgesamt 4 items, welche auf einer vierstufigen Likert Skala beantwortet werden konnten. Die mittlere Reliabilität lag hierbei bei  $\alpha = .57$ . Ergänzend dazu wurden ein einer Skala mit 9 items (4-stufige Likert Skala) die *wahrgenommenen Barrieren* zur Nutzung von Evidenz erhoben. Diese wurde ebenfalls in Anlehnung an Jette et al., 2003; Johnston et al., 2003 & Young et al., 2001 entwickelt. Die Reliabilität dieser Skala lag bei  $\alpha = .65$ .

Darüber hinaus wurden weitere möglicherweise relevante Hintergrundvariablen einmalig in dem Fragebogen vor dem Training (*Fragebogen prä*) erhoben. Darunter befanden sich diverse Variablen zum *sozioökonomischen Hintergrund* wie Geschlecht, Alter, Muttersprache, Domäne (studiert der oder die Teilnehmende mindestens ein MINT Fach oder nicht), Semesteranzahl, sowie Vorerfahrung aus anderen Studiengängen, Praktika oder Ausbildungen (Bauer et al., 2010).

Des Weiteren wurden die *Lesegewohnheiten* erhoben und dabei auch im Speziellen die *Lesemotivation* und die *Lesehäufigkeit* wissenschaftlicher Texte erhoben. Erhoben wurden diese mit 6 items durch eine vierstufige Likert Skala (Prenzel et al., 2007). Die Reliabilität dieser Skala lag bei  $\alpha = .69$ . Darüber hinaus war auch der *Nutzen* sowie der *Zugang zu Fachinformationsquellen* Inhalt des Fragebogens (Eigenentwicklung in Anlehnung an Jette et al., 2003; Johnston et al., 2003 & Young & Ward, 2001).

Im Anschluss daran folgte eine Skala zur Selbsteinschätzung des *Verständnisses fachbezogener Forschungsbegriffe* wie etwa *Evidenz*, *Signifikanz* oder *Validität* (10 items, 4-stufige Likert Skala). (Eigenentwicklung in Anlehnung an Jette et al., 2003; Johnston et al., 2003 & Young et al., 2001). Ziel war es hierbei, Informationen über bereits vorhandenes Domänenwissen zu generieren. Die Reliabilität dieser Skala betrug  $\alpha = .81$ .

Zudem wurde die *Ambiguitätstoleranz* (Dalbert et al., 2010) erhoben. Diese Skala misst, inwieweit Unvorhergesehenes Uneindeutiges toleriert wird (6 items, 4-stufige Likert Skala). Die Reliabilität dieser Skala betrug  $\alpha = .71$ .

#### *Test zum wissenschaftstheoretischen Vorwissen*

Vor dem Training erhielten alle Teilnehmenden beider Bedingungen und beider Erhebungen einen kurzen Test zum wissenschaftstheoretischen Vorwissen. Dieser

enthielt vier Aufgaben zum Verständnis von (natur)wissenschaftlichen, experimentellen Studien und basierte auf Aufgaben, die im Rahmen des NEPS Panels entwickelt wurden (Weinert, 2011). Die Bearbeitungszeit hierfür betrug 15 Minuten. Maximal waren 24 Punkte zu erreichen. Die Studierenden erreichten dabei einen durchschnittlichen Wert von  $M = 13,1$ ;  $SD = 5.03$  (range  $0 < 24$ ). Dabei erreichten 19,9% der Teilnehmenden weniger als 10 Punkte, 70,2% erzielten 10-20 Punkte und 9% erreichten über 20 Punkte.

#### *Videografierte Gruppendiskussionen*

Die Gruppendiskussionen wurden videografiert und eine Verbindung zu den Fragebögen, Vorwissenstests und Kompetenztests durch einen von den Teilnehmenden in die Kamera gezeigten Code ermöglicht. Eine Diskussion dauerte im Durchschnitt 30 Minuten. Die Vorbereitungszeit zum individuellen Lesen und kritischen Bewerten der Studie als Vorbereitung auf die Diskussion betrug 45 Minuten.

#### *Kompetenztest*

Der Kompetenztest wurde in beiden Erhebungen (bei der Experimental nach und bei der Kontrollgruppe vor dem Training) durchgeführt (siehe Tabelle 5, rechte Spalte). Dabei wurden den Teilnehmenden, analog zu den Gruppendiskussionen, mit Hilfe von Fallvignetten, Problemstellungen dargeboten. Zu diesen erhielten sie strukturierte Kurzfassungen drei verschiedener Evidenzquellen mit fragilen und konfligierenden Elementen, sowie variierender Qualität. Diese sollten zunächst einzeln kritisch bewertet werden. Im Anschluss daran wurden die Teilnehmenden angeleitet ein individuelles schriftliches Statement von einem Umfang von circa 500 Wörtern zu verfassen und darin die aktuelle Evidenzlage zu diskutieren und auf Basis dessen Empfehlung oder einen Lösungsansatz zu verfassen (Heininger et al., 2016).

### 6.3. Relevante Publikationen

Teile der Befunde aus Studie 1 wurden bereits in einem Artikel in der Zeitschrift für Unterrichtswissenschaft veröffentlicht:

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.

Ziel ist es zudem, die zentralen Ergebnisse der Studie 2 in einer internationalen Fachzeitschrift zu veröffentlichen:

Wenglein, S., Bauer, J., & Prenzel, M. (in prep.) How to foster pre-service teachers' argumentation quality in evidence-based contexts.

## 7. Studie I: Pilotstudie zur Überprüfung des entwickelten Trainingsansatzes

### 7.1. Forschungsfragen und Hypothesen

Ziel der ersten Studie war es, das entwickelte Trainingsprogramm zu pilotieren und dadurch erste Ergebnisse bezüglich der Effektivität und Wirksamkeit zu generieren. Darüber hinaus wurde aktiv und intensiv das Feedback der Teilnehmenden erfragt, um Informationen über eventuelle Schwierigkeiten, Missverständnisse, oder demotivierende Aspekte zu erhalten. Diese Informationen sollten als Hilfestellung dienen, um das Training weiterzuentwickeln und es im Zuge dessen noch besser an die Bedürfnisse der Lehramtsstudierenden anzupassen.

Um die Effektivität und Wirksamkeit zu testen, wurden die folgenden, zum Teil bereits so im Rahmen eines Artikels in der Zeitschrift *Unterrichtswissenschaft* veröffentlichten (Wenglein et al., 2015) Forschungsfragen, erarbeitet.

#### 7.1.1. Manifestierung der Heuristiken

*Forschungsfrage 1:* „Wie manifestieren sich die erlernten Heuristiken [zum kritischen Hinterfragen, sowie] zur argumentativen Nutzung von Evidenz (a) in den problembezogenen Gruppendiskussionen während des Trainings und (b) in der nachfolgenden Testsituation, in der Evidenz individuell bewertet und problembezogen argumentativ genutzt werden muss?“

*Hypothese 1.1.:* Die Heuristiken unterstützen die Lernenden dabei, die vorliegenden Evidenzquellen kritisch zu hinterfragen und diese hinsichtlich ihrer Qualität und Nutzbarkeit einzuschätzen.

*Hypothese 1.2.:* Die Heuristiken unterstützen die Lernenden dabei, die vorliegenden Evidenzquellen als Argumentationsgrundlage zu nutzen und auf Basis dessen fundierte Entscheidungen treffen zu können.

#### 7.1.2. Trainingseffekt

*Forschungsfrage 2:* „Trägt das Training bei der Trainingsgruppe (TG) gegenüber der Kontrollgruppe (KG) zu einer gesteigerten Argumentationsqualität bei?“

*Hypothese 2.1.:* Teilnehmende der Trainingsgruppe weisen eine durchschnittlich höhere Kompetenzstufe im kritischen Bewerten von Evidenz auf als Teilnehmende der Kontrollgruppe.

*Hypothese 2.2.:* Teilnehmende der Trainingsgruppe weisen eine durchschnittlich höhere Kompetenzstufe im argumentativen Nutzen von Evidenz auf als Teilnehmende der Kontrollgruppe.

### **7.1.3. Konsistenz**

*Forschungsfrage 3:* „Wie konsistent ist die [Bewertungs- und] Argumentationsqualität (a) zwischen dem kooperativen Setting (Gruppendiskussion) und der individuellen Situation im schriftlichen Test, sowie (b) über die alltags- und professionsrelevanten Problemkontexte in den Gruppendiskussionen hinweg? Gelingt ein Transfer der Heuristiken?“

*Hypothese 3.1.:* Die Kompetenz zum kritischen Bewerten von Evidenz ist über beide Domänen (Gesundheitswissenschaften und Bildungswissenschaften) hinweg konsistent.

*Hypothese 3.2.:* Die Kompetenz zum argumentativen Nutzen von Evidenz ist über beide Domänen (Gesundheitswissenschaften und Bildungswissenschaften) hinweg konsistent.

*Hypothese 3.3.:* Die Kompetenz zum kritischen Bewerten von Evidenz ist über beide Settings (mündliche Gruppendiskussionen und individuelle, schriftliche Statements) hinweg konsistent.

*Hypothese 3.4.:* Die Kompetenz zum argumentativen Nutzen von Evidenz ist über beide Settings (mündliche Gruppendiskussionen und individuelle, schriftliche Statements) hinweg konsistent.

### **7.1.4. Akzeptanz und Motivation**

*Forschungsfrage 4:* Wie ist die Akzeptanz und Motivation der Teilnehmenden, sowie deren subjektiv wahrgenommener Lernzuwachs innerhalb des Trainings? Sind die einzelnen Inhalte allgemein verständlich?

*Hypothese 4.1.:* Die Trainingsinhalte sind klar, allgemeinverständlich und eindeutig aufbereitet.

*Hypothese 4.2.:* Die Teilnehmenden sind motiviert sich im Rahmen des Trainings mit den relevanten Inhalten zum kompetenten Nutzen von Evidenz zu beschäftigen.

*Hypothese 4.3.:* Die Teilnehmenden haben nach Absolvieren des Trainings einen subjektiven Lernzuwachs erreicht.

## 7.2. Methoden

### 7.2.1. Stichprobe und Design

Um eine optimale Nutzbarkeit der Ergebnisse zu gewährleisten, wurde das Training zunächst im Rahmen einer einmaligen Präpilotierung getestet, bevor die Pilotierung gestartet wurde.

#### *Präpilotierung*

In der Präpilotierung wurde das Training mit einer Stichprobe von  $N = 3$  Lehramtsstudierenden der Technischen Universität München, sowie der Ludwig-Maximilians Universität München ( $N = 1$  männlich,  $N = 2$  weiblich) getestet. Design und Aufbau entsprachen dabei dem der Pilotierung, jedoch ohne Kontrollgruppe. Die Teilnehmenden bearbeiteten das Training mit zusätzlichen Elementen lauten Denkens (*cognitive labs*), sowie dem Ausfüllen eines ausführlichen Feedbackheftes nach jedem Sinnabschnitt (siehe Tabelle 7, rechte Spalte).

#### *Pilotierung*

Die Pilotierung des Trainings wurde mit Studierenden des Lehramtes an Gymnasien ( $N = 30$ ) der Technischen Universität München, der Otto-Friedrich-Universität Bamberg, der Universität Regensburg, der Universität Bayreuth, der Universität Würzburg, der Universität Augsburg, sowie der Ludwig-Maximilians Universität München durchgeführt. Davon waren  $N = 14$  Studierende in der Trainings- und  $N = 16$  in der Kontrollgruppe. Die Teilnehmenden befanden sich durchschnittlich im 6. Studiensemester ( $M = 6.27$  Semester,  $SD = 2.61$ ) und waren  $M = 23.36$  Jahre alt ( $SD = 2.98$ ). Bezüglich des Fächerhintergrundes studierten 80% der Probanden mindestens ein MINT-Fach, folglich hatten lediglich 20% einen rein sprach- oder geisteswissenschaftlichen Hintergrund. 75% der Teilnehmenden waren weiblich. Die Zuweisung bezüglich Experimental- und Kontrollgruppe erfolgte nach Anmeldung und somit nicht randomisiert (*haphazard assignment*; Shadish et al., 2002). Dennoch waren Hintergrundmerkmale wie Alter, Studienfortschritt und Geschlecht gleichmäßig über Kontroll- und Experimentalgruppe verteilt (alle  $p > .20$ ). Bei dem Fächerhintergrund zeigten sich jedoch signifikante Unterschiede: In der Kontrollgruppe studierten 100% der Teilnehmenden mindestens ein MINT-Fach, wohingegen in der Trainingsgruppe ein relativ ausgeglichenes Verhältnis vorherrschte und nur 57% der Teilnehmenden einen MINT Hintergrund aufwiesen. Die Daten wurden im Rahmen eines quasi-experimentellen Designs erhoben, die Kontrollgruppe erhielt dabei kein Treatment, sondern nur die Test- und Messinstrumente (siehe Tabelle 7). Beide Gruppen erhielten zu Beginn einen

Fragebogen zur Erhebung der relevanten Hintergrundvariablen. Im Anschluss daran erfolgte ein kurzer Test zum wissenschaftstheoretischen Verständnis und Vorwissen. Im Anschluss daran absolvierte die Experimentalgruppe das Trainingsprogramm. Sowohl nach Abschluss des ersten Tages, als auch am Ende des Trainings an Tag zwei, wurde erneut ein Fragebogen mit relevanten Inhalten wie Einstellungen, Motivation und Akzeptanz bearbeitet (siehe Instrumente, Kapitel 6.2.2.). Im Anschluss an das Training führten die Teilnehmenden je eine Gruppendiskussion im Bereich der Gesundheitswissenschaften, sowie der Bildungswissenschaften (vergleiche Tabelle 5 in Kapitel 5). Diese wurden videografiert und ausgewertet. Abschließend führten alle Teilnehmenden beider Bedingungen den schriftlichen Kompetenztest durch. Darüber hinaus beantworteten die Studierenden in der Trainingsgruppe über das gesamte Training hinweg mit Hilfe eines zu Anfang ausgehändigten Feedbackheftes Fragen bezüglich Verständnis, Motivation und Akzeptanz der Inhalte. Darüber hinaus wurde das gesamte Training videografiert um eventuell relevante Prozessdaten bezüglich einer Weiterentwicklung des Trainings zu erheben.

*Tabelle 7: Ablaufbeschreibung des quasi-experimentellen Designs in der Pilotierung, sowie der Präpilotierung.*

<b>Experimentalgruppe</b> <b>N = 14</b>	<b>Kontrollgruppe</b> <b>N = 16</b>	<b>Präpilotierung</b> <b>N = 3</b>
<i>Feedbackheft<sup>6</sup> (gekürzt)</i>	---	<i>Feedbackheft</i>
<i>Fragebogen (prä)</i>	<i>Fragebogen (prä)</i>	<i>Fragebogen (prä)</i>
<i>Vorwissenstest</i>	<i>Vorwissenstest</i>	<i>Vorwissenstest</i>
Training (Tag 1)	---	Training (Tag 1)
<i>Fragebogen (post 1)</i>	---	<i>Fragebogen (post 1)</i>
Training (Tag 2)	---	Training (Tag 2)
Gruppendiskussionen	---	Gruppendiskussionen
Kompetenztest	Kompetenztest	Kompetenztest
<i>Fragebogen (post2)</i>	---	<i>Fragebogen (post 2)</i>

<sup>6</sup> Das Feedbackheft wurde sowohl in der Pilotierung als auch in der Prä-Pilotierung kontinuierlich während des gesamten Trainings bearbeitet.

### 7.2.2. Instrumente

Zusätzlich zu den in Kapitel 6.2.2. beschriebenen Instrumenten, wurden im Rahmen dieser Studie die folgenden Forschungsinstrumente zur Erhebung der für die Beantwortung der Forschungsfragen relevanten Daten verwendet.

#### *Präpilotierung*

In der Präpilotierung wurde mit Hilfe von Cognitive Labs und Feedbackheften detaillierte Informationen bezüglich der Klarheit und Verständlichkeit der Inhalte, sowie der Motivation und Akzeptanz der Teilnehmenden generiert.

*Cognitive Labs:* Im Rahmen von *Cognitive Labs* werden mit Hilfe von lautem Denken der Testpersonen die kognitiven Prozesse beim Bearbeiten der Trainingsinhalte erfasst. Dabei werden die Teilnehmenden dazu angewiesen, sämtliche Gedanken und Denkprozesse laut auszusprechen. Diese wiederum werden videografiert und anschließend inhaltlich qualitativ ausgewertet (National Center for Education Statistics, 2002; Wilson, 2005).

*Feedbackheft:* Mit Hilfe eines ausführlichen, eigenentwickelten Feedbackheftes wurden die Teilnehmenden dazu angeleitet, alle drei Sinnabschnitte des Trainings (Motivation und Verständnis der Relevanz evidenzbasierter Praxis, kritisches Bewerten von Evidenz und argumentatives Nutzen von Evidenz; siehe Tabelle 4 in Kapitel 5) zu reflektieren und Fragen hinsichtlich Verständlichkeit, Klarheit, Lernzuwachs und Motivation zu beantworten.

#### *Pilotierung*

In der Pilotierung wurde analog zur Präpilotierung ein gekürztes Feedbackheft mit dem Ziel der Generierung qualitativen Feedbacks eingesetzt. Zudem wurde das in Kapitel 3.2.2. (Tabelle 3) beschriebene Kompetenzmodell zum argumentativen Nutzen von Evidenz nach Wilson adaptiert und durch Hinzufügen niedriginferenter Items ein Kategorienschema, welches die erfolgreiche Verwendung der im Training gelehrt Heuristiken abbildet, entwickelt.

*Feedbackheft (gekürzt):* In dem gekürzten Feedbackheft wurden die Studierenden dazu angeleitet zu reflektieren, ähnlich wie in Lernportfolios, was in den einzelnen Lernabschnitten des Trainings gelernt wurde und wie diese Inhalte auf die Praxis übertragbar sind.

*Kategorienschema:* Zur optimalen Abbildung der Kompetenzen, sowie der Anwendung der im Training vermittelten Heuristiken, wurden zwei aneinander angelehnte Kategorienschemata entwickelt. Diese dienten zur Analyse der

schriftlichen Statements innerhalb des Kompetenztests, sowie der videografierten Gruppendiskussionen (siehe Tabelle 8).

*Tabelle 8: Kategorienschemata zur Auswertung der schriftlichen Statements, sowie der Gruppendiskussionen in der Pilotierungsstudie.*

<b>Indikator</b>	<b>Stufen/Kategorien</b>	<b>Quellen</b>
<i>Schriftliche Statements und Gruppendiskussionen</i>		
Globalrating	5 = Der Studierende begründet durchgängig evidenzbasiert und hinterfragt die Evidenz kritisch 4 = Der Studierende begründet durchgängig evidenzbasiert und verwendet dabei relevante und angemessene Evidenz 3 = Der Studierende begründet teilweise evidenzbasiert, aber nicht in Bezug auf alle Elemente 2 = Der Studierende zitiert irrelevante Evidenz oder macht inhaltliche Fehler 1 = Der Studierende gibt durchgängig keine Begründungen und führt keine Evidenz auf	Wilson (2005), Kuhn & Crowell (2011)
Beleg von Argumenten mit Evidenz	<ul style="list-style-type: none"> <li>• Ohne Beleg</li> <li>• Mit Beleg</li> <li>• Mit Beleg und kritischer Bewertung</li> </ul>	Toulmin (1958), Kuhn & Crowell (2011)
Typ der Evidenznutzung	<ul style="list-style-type: none"> <li>• Keine Evidenz</li> <li>• Irrelevant (benutzt irrelevante Evidenz)</li> <li>• Beschreibend (Nacherzählen des Inhalts)</li> <li>• Single-Warrant (enthält eine wissenschaftliche Annahme über die Evidenzquelle)</li> <li>• Multiple-Warrant (enthält mehr als eine wissenschaftliche Annahme über die Evidenzquelle)</li> </ul>	Bell (2010), Sandoval & Reiser (2004)
Evidenzquelle	<ul style="list-style-type: none"> <li>• Keine</li> <li>• Persönliche Erfahrung (zitiert Situationen aus dem Alltag)</li> <li>• Individuell (benutzt eigene Evidenzquelle)</li> <li>• Instruiert (benutzt die ihm vorgelegte Evidenzquelle)</li> <li>• Beides (benutzt vorgegebene und eigene Evidenzquellen)</li> </ul>	Bell (2010)
Klarheit der Argumente	1 = sehr unklar – 4 = sehr klar	Bell (2010)
<i>Zusätzlich erhoben bei den Gruppendiskussionen</i>		
Vorhandensein der Bewertungsheuristiken	<ul style="list-style-type: none"> <li>• Fragestellung ...genannt (ja/nein) ...richtig erkannt (ja/nein) ...kritisch bewertet (ja/nein)</li> <li>• Methoden und Design ...genannt (ja/nein) ...richtig erkannt (ja/nein) ...kritisch bewertet (ja/nein)</li> </ul>	Abelson (1995), Elsevier (2018), Sternberg (2005)

---

	<ul style="list-style-type: none"> <li>• Ergebnisse <ul style="list-style-type: none"> <li>...genannt (ja/nein)</li> <li>...richtig erkannt (ja/nein)</li> <li>...kritisch bewertet (ja/nein)</li> </ul> </li> <li>• Schlussfolgerungen <ul style="list-style-type: none"> <li>...genannt (ja/nein)</li> <li>...richtig erkannt (ja/nein)</li> <li>...kritisch bewertet (ja/nein)</li> </ul> </li> </ul>	
Argumentations- qualität	<p>1 = stimmt überhaupt nicht; 4 = stimmt voll und ganz:</p> <ul style="list-style-type: none"> <li>• Der Studierende verwendet schlüssige Argumente.</li> <li>• Der Studierende begründet und belegt seine Argumente durchgängig anhand der Studie(n).</li> <li>• Der Studierende schätzt die Studie(n) kritisch hinsichtlich ihrer Qualität ein.</li> <li>• Die Einschätzung der Qualität der Studie fließt in die Argumentation mit ein.</li> <li>• Der Studierende berücksichtigt den praktischen Nutzen der Studie.</li> <li>• Der Studierende berücksichtigt die von den anderen Lernenden genannten Befunde.</li> <li>• Der Studierende bewertet die von den anderen Lernenden genannten Befunde kritisch.</li> <li>• Der Studierende zieht am Ende weiterführende Schlussfolgerungen für die Ausgangsfrage der Gruppendiskussion.</li> </ul>	Toulmin (1958), Kuhn & Crowell (2011)

---

In Anlehnung an das Kompetenzmodell von Wilson (2005) (siehe Kapitel 4), wurden hochinferente Globalratings zur Einschätzung der Kompetenz entwickelt. Darüber hinaus wurden zusätzliche niedriginferente Items hinzugefügt, um die Anwendung der Heuristiken besser abbilden und die Kompetenzen detaillierter messen zu können. Zudem diente dies gleichzeitig als Möglichkeit zur Validierung des Globalratings. Als Basis dafür dienten bereits existierende Ratingschemata (Bell, 2010; Sandoval & Reiser, 2004), sowie theoriebasierte Konzepte (Abelson, 1995; Elsevier, 2018; Kuhn & Crowell, 2011; Sternberg, 2005; Toulmin, 1958).

### 7.2.3. Statistische Analysen

Die Kodierungen wurden von zwei im Vorfeld trainierten, unabhängigen Ratern durchgeführt. Dabei wurde über alle Items eine hohe Interrater-Reliabilität erreicht, welche durch Intraklassen-korrelationen (McGraw & Wong, 1996) berechnet wurde. Bei den Auswertungen der Videos wurde eine Übereinstimmung von  $ICC (C, 2) = .93$  (range  $.75 < .95$ ), bei den schriftlichen Statements eine Übereinstimmung von  $ICC (C, 2) = .95$  (range  $.90 < .95$ ) erreicht. Um die Vergleichbarkeit beider

Gruppendiskussionen, sowie beider Statements, zu berechnen wurde die Reliabilität über die jeweiligen Globalratings berechnet. Diese lag bei den videografierten Diskussionen bei Cronbachs  $\alpha = .70$  und bei den schriftlichen Statements bei  $\alpha = .92$ . Aufgrund der geringen Größe, sowie der Verteilung innerhalb der Stichprobe, wurden nonparametrische Tests (Mann-Whitney  $U$ , einseitig: TG > KG) zur Analyse der Gruppenunterschiede verwendet. Um die Konsistenz der Ergebnisse der Globalratings zwischen den Settings zu analysieren, wurden Konsistenz-Intraklassenkorrelationen berechnet.

## 7.3. Ergebnisse

### 7.3.1. Forschungsfragen 1 und 2: Manifestierung der Heuristiken und Effektivität des Trainings

#### *Globalrating (hochinferent)*

Abbildung 6 zeigt die Gruppenunterschiede in den Globalratings zwischen Kontroll- und Experimentalgruppe. Die Probanden der Experimentalgruppe zeigen alle (bis auf einen Ausreißer) eine deutlich höhere Argumentationskompetenz als die Teilnehmenden der Kontrollgruppe. Das durchschnittliche Kompetenzlevel in der Experimentalgruppe lag zwischen 3 („Der Studierende begründet teilweise evidenzbasiert, aber nicht in Bezug auf alle Elemente“) und 4 („Der Studierende begründet durchgängig evidenzbasiert und verwendet dabei relevante und angemessene Evidenz“) ( $M = 3.43$ ,  $SD = .75$ ). Das durchschnittliche Kompetenzlevel der Kontrollgruppe lag hingegen bei Kompetenzstufe 2 („Der Studierende zitiert irrelevante Evidenz oder macht inhaltliche Fehler“) ( $M = 2.03$ ,  $SD = .89$ ). Aus diesem signifikanten Unterschied zwischen den beiden Gruppen ergibt sich eine große Effektstärke des Trainings ( $d = 1.67$ ,  $p < .001$ ).

Aufgrund der ungleichen Verteilung des Fächerhintergrundes unter den Teilnehmenden beider Bedingungen, wurde zudem ein weiterer Vergleich berechnet, in welchen ausschließlich Studierende mit mindestens einem MINT Fach einbezogen wurden. Auch hier zeigte sich ein signifikanter Effekt (Experimentalgruppe:  $M = 2.95$ ,  $SD = .91$ ; Kontrollgruppe:  $M = 2.03$ ,  $SD = 0.89$ ;  $U = 64$ ,  $p = .025$ ,  $d = 1.03$ ). Das höchste Kompetenzlevel („Der Studierende begründet durchgängig evidenzbasiert und hinterfragt die Evidenz kritisch“) wurde von keinem der Teilnehmenden erreicht.

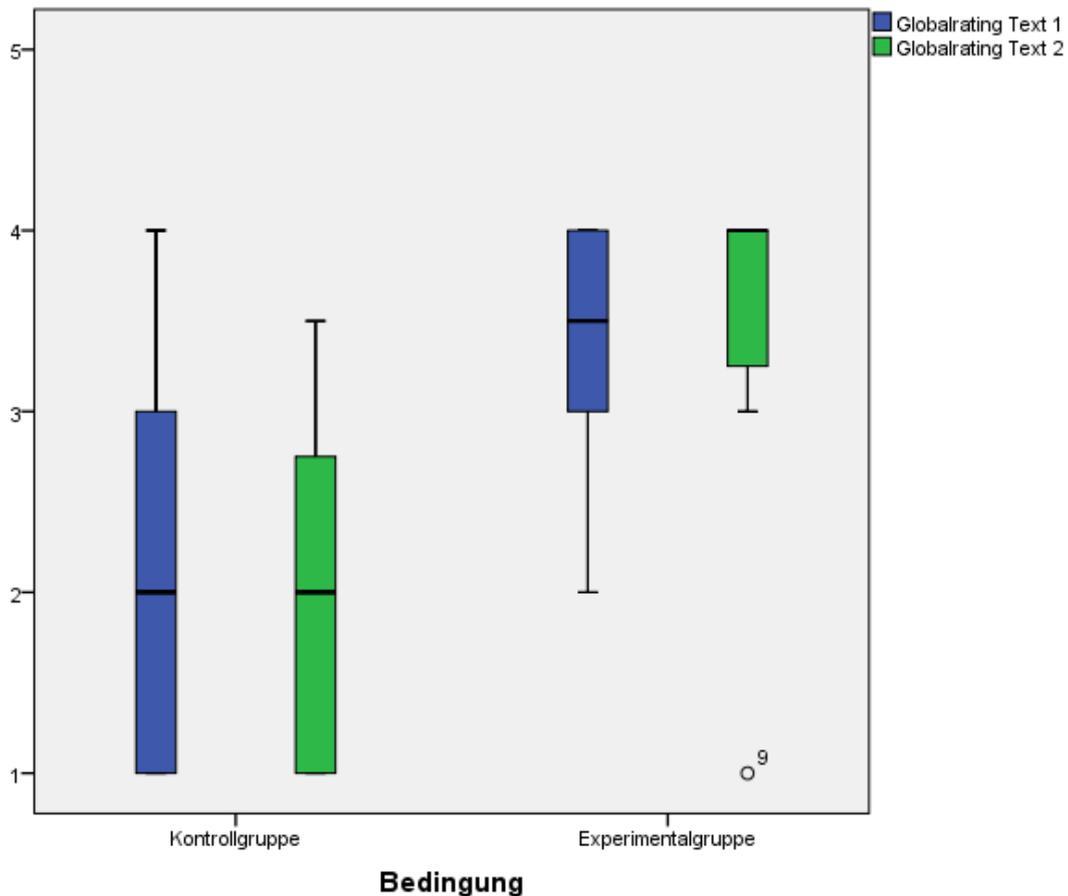


Abbildung 6: Boxplot Diagramm: Darstellung der Globalratings der Kontrollgruppe im Vergleich zur Experimentalgruppe

#### *Detaillierte Ergebnisse (niedriginferent) und Manifestierung der Argumentationsheuristiken*

In Tabelle 9 werden die Ergebnisse der einzelnen, niedriginferenten Items, jeweils aggregiert über beide Messungen (Durchschnittswerte beider Gruppendiskussionen, sowie beider schriftlicher Statements), dargestellt. Auch hier zeigt sich, dass die Teilnehmenden der Experimentalgruppe (EG) deutlich häufiger evidenzbasiert argumentieren als die Teilnehmenden der Kontrollgruppe (KG) (EG:  $M = 61.5\%$ ,  $SD = 20.4$ ; KG:  $M = 20.4\%$ ,  $SD = 24.6$ ;  $U = 158.5$ ,  $p < .001$ ). Ein zusätzlich kritisches Hinterfragen der Evidenz konnte am häufigsten in den Gruppendiskussionen (GD) gemessen werden (GD:  $M = 23.0\%$ ,  $SD = 18.1$ ; EG:  $M = 16.3\%$ ,  $SD = 17.8$ ; KG:  $M = 10.0\%$ ;  $SD = 15.9$ ;  $U = 104.5$ ;  $p = .21$ ).

Tabelle 9: Detaillierte Ergebnisse über beide Settings hinweg (so veröffentlicht in Wenglein et al., 2015)

Variable	Experimentalgruppe				Kontroll- gruppe		Texte: TG vs. KG <sup>2</sup>
	Gruppen- diskussionen		Texte		Texte		
	<i>M</i> <sup>1</sup>	<i>SD</i>	<i>M</i> <sup>1</sup>	<i>SD</i>	<i>M</i> <sup>1</sup>	<i>SD</i>	
<i>Beleg von Argumenten mit Evidenz</i>							
Ohne Evidenz	15.5%	16.5	22.3 %	26.1	69.6 %	26.5	<i>U</i> = 18 <i>p</i> < .001
Mit Evidenz	61.4%	25.3	61.5 %	20.4	20.4 %	24.6	<i>U</i> = 158.5 <i>p</i> < .001
Mit Evidenz und kritischer Bewertung	23.0%	18.1	16.3 %	17.8	10.0 %	15.9	<i>U</i> = 104.5 <i>p</i> = .21
<i>Art der Evidenznutzung</i>							
Keine	3.3%	8.6	22.7 %	36.2	67.3 %	33.2	<i>U</i> = 32.5 <i>p</i> = .003
Irrelevant	15.6%	24.9	0.0%	0.0	2.2%	5.2	<i>U</i> = 71.5 <i>p</i> = .21
Beschreibend	73.4%	29.4	1.0%	3.4	0.0%	0.0	<i>U</i> = 96 <i>p</i> = .36
Single Warrant	7.8%	8.3	63.8 %	38.4	27.7 %	32.5	<i>U</i> = 132 <i>p</i> = .01
Multiple Warrant	0.0%	0.0	12.6 %	22.4	2.8%	6.1	<i>U</i> = 107.5 <i>p</i> = .17
<i>Evidenzquelle</i>							
Keine	22.2%	26.0	22.6 %	39.0	67.3 %	33.2	<i>U</i> = 36 <i>p</i> = .005
Persönlich	1.4%	4.5	0.0%	0.0	2.3%	7.0	<i>U</i> = 77 <i>p</i> = .31
Individuell	69.9%	26.0	0.0%	0.0	1.0%	2.89	<i>U</i> = 77 <i>p</i> = .305
Instruiert	0.9%	2.6	77.4 %	39.0	29.4 %	33.3	<i>U</i> = 141 <i>p</i> = .004
Instruiert und individuell Klarheit <sup>3</sup>	5.4%	12.1	0.0%	0.0	0.0%	0.0	---
	3.38	0.47	3.16	0.94	1.73	0.79	<i>U</i> = 153 <i>p</i> < .001

Anmerkungen. <sup>1</sup>Über zwei Rater und je zwei Gruppendiskussionen bzw. Textaufgaben gemittelt; <sup>2</sup>Einseitiger Mann-Whitney-Test; <sup>3</sup>1 = sehr unklar, 4 = sehr klar.

Die häufigste Art Evidenz zu nutzen war in den Gruppendiskussionen diese *Beschreibend* einzusetzen ( $M = 73.4\%$ ;  $SD = 29.4$ ). In den schriftlichen Statements der Experimentalgruppe wurde dem überwiegend eine (*Single-Warrant*) oder mehrere (*Multiple Warrant*) wissenschaftliche Behauptung(en) über die Evidenz hinzugefügt (*Single Warrant*:  $M = 63.8\%$ ,  $SD = 38.4$ ). In der Kontrollgruppe wurde überwiegend gar keine Evidenz verwendet ( $M = 67.3\%$ ,  $SD = 26.4$ ).

Im Falle der Evidenznutzung im Rahmen des Argumentierens, nutzten sowohl die Teilnehmenden der Experimental-, als auch die der Kontrollgruppe am häufigsten die ihnen dargebotenen Evidenzquellen, letztere jedoch deutlich seltener (EG:  $M = 77.4\%$ ,  $SD = 39.0$ ; KG:  $M = 29.4\%$ ;  $SD = 33.3$ ). In den Gruppendiskussionen hingegen wurde überwiegend individuelle Evidenz genutzt ( $M = 69.9\%$ ;  $SD = 26.0$ ). Die Argumente der Studierenden in der Experimentalgruppe waren überwiegend klar und verständlich aufgebaut und strukturiert, wohingegen die der Studierenden in der Kontrollgruppe als eher unklar eingeschätzt wurden (GD:  $M = 3.38$ ,  $SD = 0.47$ ; EG:  $M = 3.16$ ,  $SD = 0.94$ ; KG:  $M = 1.73$ ,  $SD = 0.79$ ;  $U = 153$ ,  $p < .001$ ).

#### *Manifestierung der Bewertungsheuristiken*

Mit Hilfe der Videoaufnahmen beider Gruppendiskussionen wurde zusätzlich die Manifestierung der Bewertungsheuristiken erfasst. Tabelle 10 zeigt, dass sowohl in der gesundheitswissenschaftlichen (GeWi), ersten Gruppendiskussion, als auch in der zweiten, bildungswissenschaftlichen (BiWi) Diskussion am häufigsten Informationen aus Ergebnis- und Diskussionsteilen der vorgegebenen Evidenzquellen zur Stützung der eigenen Behauptung herangezogen wurden (GeWi:  $M = 92.2\%$ ,  $SD = .32$ ; BiWi:  $M = 85.7\%$ ,  $SD = .37$ ).

Zudem ist erkennbar, dass alle vier Bereiche (Fragestellung, Methoden und Design, Ergebnisse und Schlussfolgerungen) zwar mit großer Häufigkeit genannt und richtig erkannt wurden, jedoch deutlich weniger oft auch zudem kritisch bewertet. (Am Beispiel Schlussfolgerungen Diskussion 1 (GeWi): genannt und richtig erkannt:  $M = 92.9\%$ ,  $SD = .32$ ; kritisch bewertet:  $M = 28.6\%$ ,  $SD = .23$ ).

Tabelle 10: Anwendung der Heuristiken in den Gruppendiskussionen

	Video 1 (GeWi)			Video 2 (BiWi)		
	Ja	Nein	SD	Ja	Nein	SD
Fragestellung						
...genannt	78,6%	21,4%	.43	57,1%	42,9%	.51
...richtig erkannt	78,6%	21,4%	.43	57,1%	42,9%	.51
...kritisch bewertet	28,6%	71,4%	.32	14,3%	85,7%	.18
Methoden und Design						
...genannt	71,4%	28,6%	.47	57,1%	42,9%	.51
...richtig erkannt	71,4%	28,6%	.47	57,1%	42,9%	.48
...kritisch bewertet	50,0%	50,0%	.41	21,4%	78,6%	.37
Ergebnisse						
...genannt	92,9%	7,1%	.32	85,7%	14,3%	.38
...richtig erkannt	92,9%	7,1%	.32	85,7%	14,3%	.38
...kritisch bewertet	57,1%	42,9%	.36	42,9%	57,1%	.37
Schlussfolgerungen						
...genannt	92,9%	7,1%	.32	85,7%	14,3%	.37
...richtig erkannt	92,9%	7,1%	.32	85,7%	14,3%	.37
...kritisch bewertet	28,6%	71,4%	.23	42,9%	57,1%	.42

### 7.3.2. Forschungsfrage 3: Konsistenz der Ergebnisse

Tabelle 11 zeigt die Übereinstimmung der Ergebnisse der Globalratings über alle Settings hinweg. So war die Konsistenz der Ergebnisse zwischen beiden Gruppendiskussionen im akzeptablen Bereich [ $ICC (C, 2) = .69, p = .02$ ]. Die Konsistenz zwischen den beiden schriftlichen Statements war dagegen sehr hoch [ $ICC (C, 2) = .92, p = .00$ ]. Die Konsistenz über beide Gruppendiskussionen und über beide Statements hinweg lag ebenfalls im akzeptablen Bereich [ $ICC (C, 4) = .71, p = .004$ ].

Tabelle 11: Konsistenz der Ergebnisse der Globalratings über die Settings hinweg

	ICC	p
Video1 (GeWi) und Video 2 (BiWi)	.696	.020
Statement 1 und Statement 2	.924	.000
Videos und Statements	.711	.004

### 7.3.3. Forschungsfrage 4: Akzeptanz und Motivation

Tabelle 12 zeigt, dass an beiden Trainingstagen sowohl der subjektiv wahrgenommene Lernzuwachs, als auch die selbst eingeschätzte Motivation, emotionale Befindlichkeit und allgemeine Zufriedenheit, hoch waren und sich alle Indikatoren geringfügig von Tag eins zu Tag zwei hin steigerten. Die einzelnen Items wurden mit Hilfe einer vierstufigen Likert Skala erhoben. Dabei lagen alle Werte zwischen 3 („stimmt eher“) und 4 („stimmt ganz genau“). Zudem waren die Reliabilitätskoeffizienten aller Skalen im verwertbaren Bereich ( $\alpha$  range .70 < .93). Auch die Aussagen aus der qualitativen Auswertung der Feedbackhefte, sowie der Aussagen, die mit Hilfe der Cognitive Labs in der Präpilotierung erhoben wurden, stützen diese Ergebnisse.

*Tabelle 12:* Überblick über Motivation und Akzeptanz an beiden Trainingstagen

	Items <sup>7</sup>	Tag 1			Tag 2			gesamt		
		M	SD	$\alpha$	M	SD	$\alpha$	M	SD	$\alpha$
Lernzuwachs	3	3.33	0.14	0.70	3.48	0.11	0.78	3.40	0.14	0.84
Motivation	6	3.48	0.11	0.72	3.70	0.08	0.83	3.59	0.15	0.87
emot. Befindlichk.	10	3.32	0.26	0.79	3.47	0.26	0.83	3.39	0.27	0.89
Zufriedenheit	16	3.52	0.26	0.84	3.58	0.18	0.93	3.55	0.22	0.92

## 7.4. Diskussion

Die Förderung der relevanten Kompetenzen zum professionellen Umgang mit Evidenz für (zukünftige) Lehrkräfte gewinnt sowohl national als auch international zunehmend an Bedeutung. (EC, 2013; KMK, 2004/2014; Parr & Timperley, 2008; U.S. Department of Education, 2007). Da dies bislang kein fester Bestandteil des Lehramtsstudium ist, werden effektive Trainings, die flexibel in bestehende Lehramtscurricula integrierbar sind benötigt (Bauer et al., 2012; Bauer, 2010). Aus diesem Grund wurde das, im Rahmen des E4teach Projekts entwickelte und auf Heuristiken basierendes Training, im Rahmen dieser Pilotstudie erstmals getestet und überprüft.

<sup>7</sup> Beantwortung auf einer 4-stufigen Likert-Skala (4= stimmt ganz genau, 3 = stimmt eher, 2 = stimmt eher nicht, 1 = stimmt überhaupt nicht)

Erste Ergebnisse zur Nutzungs- und Argumentationskompetenz, sowie zur Manifestierung der Heuristiken zeigten bereits deutliche Effekte. So wiesen die Teilnehmenden der Experimentalgruppe eine deutlich höhere Kompetenz auf, als die Teilnehmenden in der Kontrollgruppe. Dies zeigte sich sowohl in den hochinferenten Globalratings, als auch in den niedriginferenten Items. Während die Studierenden in der Kontrollgruppe im Durchschnitt lediglich Kompetenzstufe 2 erreichten, bewegten sich die Studierenden in der Experimentalgruppe zwischen Stufe 3 und 4. Die höchste Kompetenzstufe 5 (Kuhn & Crowell, 2011; Wilson, 2005) konnte jedoch von keinem der Teilnehmenden erreicht werden. Zudem wurde die Evidenz zwar häufig genutzt, jedoch vergleichsweise selten auch kritisch hinterfragt und bewertet (Bell, 2010; Sandoval & Reiser, 2004). Dies sind vermutlich die Grenzen einer solchen Kurzzeitintervention. Um ein tieferes Verständnis, sowie eine nachhaltige Anwendung zu fördern, sind weitere Lerngelegenheiten notwendig (Desimone, 2009). Die konsistenten Ergebnisse, sowohl über die schriftlichen Statements hinweg, als auch über die beiden Gruppendiskussionen, sowie die Gesamtheit dieser zeigt, dass den Teilnehmenden der Transfer zunächst von den Gesundheitswissenschaften in den Gruppendiskussionen und anschließend der Transfer von mündlichen Diskussionen in der Gruppe, hin zu individuell verfassten, schriftlichen Statements, größtenteils gelungen ist.

Auch die Ergebnisse zur Motivation und Akzeptanz der Teilnehmenden sind durchweg positiv. Diese ersten Ergebnisse zeigen, dass der Trainingsansatz sowohl die Kompetenz, als auch die Motivation hinsichtlich eines ersten Schrittes zu einer evidenzbasierten Praxis fördern kann.

Jedoch gibt es Einschränkungen bezüglich der Interpretation dieser Ergebnisse. Diese betreffen dabei vor allem die Größe, sowie die Zusammensetzung der Stichprobe. Um Fächereffekte zu vermeiden und ein repräsentativeres Ergebnis über die tatsächliche Effektivität der Trainings zu generieren ist es deshalb notwendig, diese nochmals mit einer deutlich größeren Stichprobe zu replizieren. Zudem kann eine Überarbeitung der Kodierschemata dahingehend hilfreich sein, die Heuristiken in beiden Settings noch systematischer und vergleichbarer abzudecken, sowie deren Komplexität etwas zu reduzieren. Die Ergebnisse dieser Pilotstudie dienten als Ansatzpunkt für die Konzeption der Hauptstudie (siehe Kapitel 8), in welcher sowohl das Trainingsprogramm als auch die Testinstrumente auf Basis der Ergebnisse dieser Studie weiterentwickelt wurden.

## 8. Studie II: Hauptstudie zur Wirksamkeit des Trainings und möglichen Einflussfaktoren

### 8.1. Forschungsfragen und Hypothesen

Ziel von Studie 2 war es, die Ergebnisse aus Studie 1 mit Hilfe einer größeren Stichprobe zu replizieren, sowie darüber hinaus Informationen über mögliche Einflussfaktoren hinsichtlich der Kompetenzentwicklung zum kritischen Hinterfragen und argumentativen Nutzen von Evidenz zu generieren. Daraus ergaben sich die folgenden beiden übergeordneten Forschungsfragen:

#### 8.1.1. Trainingseffekte

*Forschungsfrage 1:* Trägt ein auf Heuristiken basierendes Kurzzeittraining bei der Trainingsgruppe gegenüber der Kontrollgruppe zu einer gesteigerten Bewertungs- und Argumentationskompetenz (*appraise* und *apply*) bei? Wie konsistent ist die Bewertungs- und Argumentationsqualität über verschiedene Settings und Problemkontexte hinweg?

*Hypothese 1.1.:* Die Trainingsgruppe zeigt ein höheres Kompetenzniveau beim kritischen Hinterfragen von Evidenz (*appraise*) als die Kontrollgruppe.

*Hypothese 1.2.:* Die Trainingsgruppe zeigt ein höheres Kompetenzniveau beim argumentativen Nutzen von Evidenz (*apply*) als die Kontrollgruppe.

*Hypothese 1.3.:* Die Bewertungskompetenz und die Argumentationskompetenz korrelieren untereinander. Demzufolge haben Studierende, die ein hohes Kompetenzlevel beim kritischen Hinterfragen von Evidenz erreichen auch ein hohes Niveau beim argumentativen Nutzen der Evidenzquellen.

*Hypothese 1.4.:* Die Kompetenzen sind über die verschiedenen Settings hinweg konsistent. Somit zeigen Studierende sowohl beim schriftlichen und individuellen Bearbeiten des Testinstruments, als auch bei den mündlichen Diskussionen in der Gruppe ein ähnliches Niveau beider Kompetenzen.

*Hypothese 1.5.:* Die Kompetenzen sind über verschiedene Problemkontexte hinweg konsistent. Demnach zeigen die Teilnehmenden ein ähnliches Kompetenzniveau sowohl bei gesundheitswissenschaftlichen Problemstellungen, als auch bei professionsrelevanten, bildungswissenschaftlichen Kontexten.

#### 8.1.2. Einflussfaktoren

*Forschungsfrage 2:* Welche Rolle spielen hierfür das wissenschaftstheoretische Vorwissen, die Fachausrichtung, die Überzeugungen, die Lesegewohnheiten und die Ambiguitätstoleranz der Studierenden?

Hypothese 2.1.: Studierende, die mindestens ein MINT-Fach studieren und somit bereits vertraut mit den Methoden empirischer, beziehungsweise naturwissenschaftlicher Forschung, sind als Studierende mit rein sprach- oder geisteswissenschaftlichem Hintergrund, zeigen durch diesen Vorteil sowohl in der Experimental- als auch in der Trainingsgruppe ein höheres Kompetenzlevel.

Hypothese 2.2.: Wissenschaftstheoretisches Vorwissen beeinflusst den Lernzuwachs innerhalb des Trainings.

Hypothese 2.3.: Epistemologische Einstellungen über Wissenschaft in den Gesundheitswissenschaften, sowie den Bildungswissenschaften beeinflussen den Lernzuwachs innerhalb des Trainings.

*Hypothese 2.4.:* Überzeugungen und (wahrgenommene) Barrieren zum Nutzen von Evidenz beeinflussen den Lernzuwachs innerhalb des Trainings.

*Hypothese 2.5.:* Lesegewohnheiten in Bezug auf Forschung und Fachartikel beeinflussen den Lernzuwachs innerhalb des Trainings.

*Hypothese 2.6.:* Die individuelle Ambiguitätstoleranz beeinflusst den Lernzuwachs innerhalb des Trainings.

## **8.2. Methoden**

### **8.2.1. Stichprobe und Design**

Die Studie wurde als 2x2 experimentelles Design konzipiert. Der randomisierende Faktor war hierbei die Bedingung (Experimental- versus Kontrollgruppe) und der beobachtete Faktor war das durch den Fächerhintergrund erlangte Domänenwissen (MINT versus nicht-MINT).

Es nahmen insgesamt  $N = 167$  Studierende des Lehramts an Gymnasien von insgesamt 10 verschiedenen Universitäten innerhalb Deutschlands an der experimentellen Studie teil. Dafür wurde das Training 21-mal durchgeführt, davon 11-mal in der Experimental- und 10-mal in der Kontrollbedingung. Welche Bedingung zu welchem Termin stattfand wurde unter Berücksichtigung einer ausgeglichenen Verteilung der Standorte zufällig ausgewählt (*batch randomization*). Demnach befanden sich  $N = 84$  Teilnehmende in der Experimental- und  $N = 83$  in der Kontrollgruppe. Die Rekrutierung der Studierenden erfolgte über Aushänge und Werbung in Seminaren für Lehramtsstudierende mit psychologischem oder pädagogischem Inhalt. Für diejenigen Teilnehmenden, die von ihrem Institut weder Pflicht- noch Kür- ECTS Credits erhalten konnten, wurde ein Amazon Gutschein im Wert von je 40 Euro bereitgestellt, um einem möglichen motivational bedingten Bias entgegenzuwirken. Insgesamt erhielten  $N = 13$  Teilnehmende (7,8%) im Rahmen

ihres Instituts bereitgestellte Kür-ECTS Credits,  $N = 46$  Teilnehmende (27,5%) erhielten Pflicht-ECTS Credits durch die Integration des Trainings in das Seminar „Forschendes Lernen“ an der Technischen Universität München und  $N = 108$  (64%) bekamen als Ausgleich den Gutschein ausgehändigt. Die Probanden waren zu 57,1% weiblich ( $N = 96$ ) und das durchschnittliche Alter betrug  $M = 24,8$  Jahre ( $SD = 5.90$ , *range*  $18 < 51$ ). Im Schnitt befanden sie sich, wie angestrebt, im 5. Fachsemester ihres Lehramtsstudiums ( $M = 4,68$ ,  $SD = 2.67$ , *range*  $1 < 14$ ).  $N = 94$  (56%) studierten mindestens ein MINT Fach,  $N = 74$  (44%) hatten einen rein geistes- und/oder sprachwissenschaftlichen Hintergrund. 54,2% der Teilnehmenden gaben an keine bis nahezu keine Erfahrung mit Texten der empirischen Forschung zu haben. Darüber hinaus gaben 56,6% an mindestens ein- bis zweimal die Woche oder öfter wissenschaftliche Artikel zu lesen. Bildungswissenschaftliche Artikel wurden jedoch nur von 7,1% mit der gleichen Häufigkeit gelesen. Trotz der Randomisierung der Termine hinsichtlich Kontroll- und Experimentalgruppe wurden die Teilnehmenden beider Gruppen hinsichtlich ihrer Hintergrundvariablen verglichen, um potenzielle zufällige Unterschiede kontrollieren zu können. Über alle Variablen hinweg zeigten sich keine statistisch signifikanten Unterschieden zwischen beiden Gruppen (alle Werte  $p > .14$ ).

Die Studierenden beider Bedingungen erhielten zu Beginn einen individuellen, papierbasierten Fragebogen zu relevanten Einstellungen, Überzeugungen und (sozioökonomischen) Hintergrundvariablen, sowie einen kurzen Test zum wissenschaftstheoretischen Vorwissen. Die Experimentalgruppe begann anschließend mit dem Trainingsprogramm und den darauffolgenden Gruppendiskussionen, während die Kontrollgruppe zuerst das Testinstrument bearbeitete (siehe Tabelle 6, Kapitel 5). Abschließend beantworteten die Teilnehmenden beider Bedingungen einen abschließenden Fragebogen.

### **8.2.2. Instrumente**

Zusätzlich zu den in Kapitel 6.2.2. beschriebenen Instrumenten wurde zur Auswertung der Ergebnisse, ein auf Basis der Erkenntnisse in Studie 1 (siehe Kapitel 7.2.2.) weiterentwickeltes Kodierschema verwendet. Ziel war es ein einheitliches Schema zu entwickeln, welches sowohl zur Auswertung der schriftlichen Statements aus den Kompetenztests, als auch der videografierten Gruppendiskussionen verwendet werden kann. Dies ermöglicht eine bessere Vergleichbarkeit über die verschiedenen Settings hinweg (siehe Tabelle 13).

*Tabelle 13:* Kategorienschemata zur Auswertung der schriftlichen Statements, sowie der Gruppendiskussionen in der Hauptstudie

Indikator	Stufen/Kategorien	Quellen
Hochinferente Globalratings		
Bewertungskompetenz ( <i>appraise</i> )	<p>Level 5: Der Studierende nennt die relevanten und wichtigsten Inhalte und bewertet diese kritisch. Darüber hinaus erkennt er die Fragilität und Konflikthaftigkeit des Forschungsfeldes.</p> <p>Level 4: Der Studierende nennt die relevanten und wichtigsten Inhalte und bewertet diese durchgängig kritisch.</p> <p>Level 3: Der Studierende nennt die relevanten Inhalte und bewertet diese teilweise kritisch.</p> <p>Level 2: Der Studierende nennt Inhalte der Studie ohne diese kritisch zu bewerten</p> <p>Level 1: Der Studierende nennt irrelevante oder falsche Inhalte der Studie. Zudem bewertet er diese nicht kritisch oder sogar falsch.</p>	Eigenentwicklung in Anlehnung an Wilson (2005), Bromme (2010)
Argumentationskompetenz ( <i>apply</i> )	<p>Level 5: Der Studierende begründet durchgängig evidenzbasiert und hinterfragt die Evidenz kritisch.</p> <p>Level 4: Der Studierende begründet durchgängig evidenzbasiert und verwendet dabei relevante und angemessene Evidenz.</p> <p>Level 3: Der Studierende begründet teilweise evidenzbasiert, aber nicht in Bezug auf alle Elemente.</p> <p>Level 2: Der Studierende zitiert irrelevante Evidenz oder macht inhaltliche Fehler.</p> <p>Level 1: Der Studierende nennt durchgängig keine Begründungen und führt keine Evidenz auf.</p>	Eigenentwicklung in Anlehnung an Wilson (2005), Kuhn & Crowell (2011)
Niedriginferente Ratings:		
Beleg mit Evidenz	<p>1 = Behauptung</p> <p>2 = Behauptung mit Beleg</p> <p>3 = Behauptung mit Beleg und kritischer Bewertung der Evidenzquelle</p>	Toulmin (1958), Kuhn & C. (2011)
Länge der Arg.	1 = kurz (max. 1 Satz); 2 = mittel (1-2 Sätze); 3 = lang (3 oder mehr Sätze)	Bell (2010)
Warrant	1 = keiner; 2 = irrelevant (benutzt irrelevante Evidenz); 3 = beschreibend (bloßes Nacherzählen des Inhalts); 4 = single-warrant (enthält eine wissenschaftliche Annahme über die Evidenzquelle); 5 = multiple-warrant (enthält mehr als eine wissenschaftliche Annahme über die Evidenzquelle)	Bell (2010), Sandoval & Reiser (2004)
Backing	1 = keine; 2 = persönliche Erfahrung (zitiert Situationen aus dem Alltag; 3 = individuell (benutzt eigene Evidenzquelle); 4 = instruiert (benutzt die ihm vorgelegte Evidenzquelle);	Bell (2010)
Klarheit	1 = sehr unklar; 2 = eher unklar; 3 = eher klar; 4 = sehr klar	Bell (2010)
Position	1 = einseitig; 2 = zweiseitig (beachtet beide Seiten/ Pro und Contra)	Kuhn & C. (2011)

Verhalten bei Uneinigkeit 1 = bloße Ablehnung; 2 = Counter-A (Gegenargument); 3 = Counter-C (Kritik am Argument des anderen); 4 = Counter-A und Counter-C Kuhn & Crowell (2011)

Im Unterschied zu den in Studie 1 angewendeten Schemas wurde hier ein zweites Globalrating in Anlehnung an das Kompetenzmodell von Wilson (2005) entwickelt, um die Argumentationskompetenz (*apply*) vergleichend zur Bewertungskompetenz (*appraise*) erheben zu können. Zudem wurden weitere niedriginferente Ratings hinzugefügt, um das Nutzen der im Training vermittelten Heuristiken noch besser abbilden zu können.

Dieses Kodierschema wurde von zwei unabhängigen und geschulten Ratern angewendet. Um die Übereinstimmung der beiden Rater zu überprüfen, wurden Intraklassen-Korrelationen für jedes einzelne Item des Schemas berechnet (mittlerer *ICC* = .87, range .70 < .97). Bei denjenigen Items, welche eine klare Entscheidung voraussetzen, wie beispielsweise „Position“ und somit kein Mittelwert zweier Rater sinnvoll ist, wurde mit Hilfe eines dritten Raters konsensvalidiert und somit eine gemeinsame Entscheidung getroffen.

### 8.2.3. Statistische Analysen

Um die Effektivität des Trainings und dessen Einflussfaktoren auf die beiden Kompetenzen (*appraise* und *apply*) bewerten zu können, wurden drei hierarchische Regressionsmodelle für die hochinferenten Globalratings berechnet. Modell 1 analysiert dabei die Trainingseffekte ohne Kontrollvariablen (Forschungsfrage 1). Modell 2 bezieht die Domäneneffekte (MINT versus nicht MINT) mit ein (Forschungsfrage 2). Als Kontrollvariablen wurden hierfür das (wissenschaftstheoretische) Vorwissen, sowie die Lesegewohnheiten bezüglich wissenschaftlicher Texte herangezogen. Um ein detaillierteres Ergebnis hinsichtlich der Unterschiede zwischen Trainings- und Kontrollgruppe zu generieren, wurden zudem weitere Gruppenvergleiche (einseitige ANOVA) für die niedriginferenten Ratings berechnet. In Modell 3 wurde zusätzlich ein möglicher Einfluss der epistemologischen Überzeugungen (*CAEB texture* und *CAEB variability*), den Überzeugungen zum Nutzen von Evidenz, sowie der Ambiguitätstoleranz und (wahrgenommenen) Barrieren zum Nutzen von Evidenz kontrolliert. Die statistischen Analysen wurden mit Hilfe des Statistikprogramms MPlus 7 (Muthén & Muthén, 1998-2015) unter Korrektur der Standardfehler für die Mehrebenenstruktur des Experiments (unterschiedliche Trainingszeitpunkte, verschiedene Universitäten und Standorte und unterschiedliche Gruppen innerhalb des Trainings) berechnet. Hierfür wurde ein alpha level von 0.05 für alle statistischen Tests verwendet.

## 8.3. Ergebnisse

### 8.3.1. Forschungsfrage 1: Effektivität des Trainings

#### *Ergebnisse der hochinferenten Globalratings*

Ziel war es zunächst, Informationen bezüglich der Wirksamkeit und Effektivität des Trainingsprogramms zu generieren. Betrachtet man die hochinferenten Globalratings, zeigen sich hierbei signifikante große Effekte (siehe Tabelle 14, Modell 1).

*Tabelle 14:* Hierarchische Regressionsmodelle zur Berechnung von trainings- und Domäneneffekten. Sowie weiteren Einflussfaktoren (signifikante Effekte sind fett gedruckt)

	Bewertungskompetenz ( <i>appraise</i> )			Argumentationskompetenz ( <i>apply</i> )		
	Modell 1	Modell 2	Modell 3	Modell 1	Modell 2	Modell 3
	$\beta$	(SE)	$\beta$	(SE)	$\beta$	(SE)
<b>Bedingung</b>	<b>.59</b>	(.15)	<b>.56</b>	(.23)	<b>.58</b>	(.16)
<b>MINT</b>			.04	(.22)	.06	(.23)
<b>Vorwissen</b>			.01	(.02)	-.03	(.02)
<b>Lesegew.</b>			-.02	(.15)	-.05	(.16)
<b>Bed. x MINT</b>			.08	(.31)	.08	(.31)
<b>CAEB text.</b>					-.01	(.14)
<b>CAEB var.</b>					.01	(.12)
<b>Ambiguit.</b>					-.03	(.13)
<b>Nutzen</b>					.11	(.18)
<b>Barrieren</b>					-.06	(.13)
<b>R<sup>2</sup></b>	.40		.38		.40	
					.38	
					.34	
					.36	

Um die Unterschiede zwischen Experimental- und Kontrollgruppe noch deutlicher herauszuarbeiten, wurde die Effektstärke durch Cohen's  $d$  berechnet. Auch hier zeigen sich signifikante große Effekte für beide Kompetenzen: Die Effektstärke des Trainings hinsichtlich der Bewertungskompetenz (*appraise*) beträgt  $d = 1.55$ , die der Argumentationskompetenz (*apply*) liegt bei  $d = 1.36$ . Folglich unterscheiden sich die Teilnehmenden der Kontrollgruppe deutlich hinsichtlich ihres Kompetenzlevels von den Teilnehmenden der Experimentalgruppe. Das mittlere Kompetenzlevel Studierenden in der Kontrollgruppe lag bei  $M = 2.63$  ( $SD = 1.10$ ), das der Studierenden in der Experimentalgruppe bei  $M = 4.05$  ( $SD = .88$ ). Die Ergebnisse wurden hierfür auf Individualebene über beide schriftlichen Statements aggregiert (siehe Abbildung 7). Betrachtet man die Ergebnisse beider Statements einzeln für beide Bedingungen, waren beide Kompetenzen bei der Kontrollgruppe im zweiten Statement etwa eine halbe Kompetenzstufe höher als beim erstem Statement (Bewertungskompetenz Text 1:  $M = 2.01$ ,  $SD = .86$  versus Text 2:  $M = 2.62$ ,  $SD = 1.14$ ; Argumentationskompetenz Text 1:  $M = 2.37$ ,  $SD = 1.22$  versus Text 2:  $M = 2.77$ ,  $SD = 1.24$ ). Zudem zeigen diese Ergebnisse, dass die Argumentationskompetenz leicht höher war als die Bewertungskompetenz. Bei der Experimentalgruppe waren diese Unterschiede ebenso sichtbar, jedoch deutlich geringer ausgeprägt. (Bewertungskompetenz Text 1:  $M = 3.81$ ,  $SD = 1.01$  versus Text 2:  $M = 3.98$ ,  $SD = 1.01$ ; Argumentationskompetenz Text 1:  $M = 4.02$ ,  $SD = .91$  versus Text 2:  $M = 4.23$ ,  $SD = .75$ ).

Zur Messung der Effektivität des Trainings konnten ausschließlich die Ergebnisse aus den schriftlichen Statements herangezogen werden. Die videografierten Gruppendiskussionen waren Bestandteil des Trainings und wurden somit bei beiden Bedingungen erst abschließend durchgeführt. Aus diesem Grund ist ein entsprechender Vergleich hierzu nicht möglich. Die Teilnehmenden der Kontrollgruppe bearbeiteten den Kompetenztest vor dem Training und somit auch vor den Gruppendiskussionen, während dieser in der Experimentalbedingung erst nach den Diskussionen durchgeführt wurde. Somit könnte in diesem Fall durch einen möglichen Lerneffekt des Tests ein Vorteil bei der Kontrollgruppe bestehen, jedoch zeigten sich keine signifikanten Unterschiede in den Kompetenzen. In der Experimentalgruppe lag die Bewertungskompetenz in den Gruppendiskussionen bei  $M = 3.88$  ( $SD = .77$ ) in Diskussion 1 (über das gesundheitswissenschaftliche Thema) und bei  $M = 4.06$  ( $SD = .79$ ) in Diskussion 2 (über das bildungswissenschaftliche Thema). Die Argumentationskompetenz war durchschnittlich bei  $M = 4.09$  ( $SD = .84$ ) (Diskussion 1) beziehungsweise  $M = 4.25$  ( $SD = .86$ ). In der Kontrollgruppe lag die Bewertungskompetenz im Durchschnitt bei

$M = 3.77$  ( $SD = .82$ ) (Diskussion1) und  $M = 4.04$  ( $SD = .78$ ) (Diskussion 2) und die Argumentationskompetenz bei  $M = 3.99$  ( $SD = .84$ ) (Diskussion 1) und  $M = 4.58$  ( $SD = .57$ ) (Diskussion 2). Folglich zeigen auch diese Ergebnisse ähnlich zu denen der schriftlichen Statements einen geringfügigen Anstieg der Kompetenz von Diskussion 1 zu Diskussion 2, sowie einen leicht höheren Wert in der Argumentationskompetenz im Vergleich zur Bewertungskompetenz.

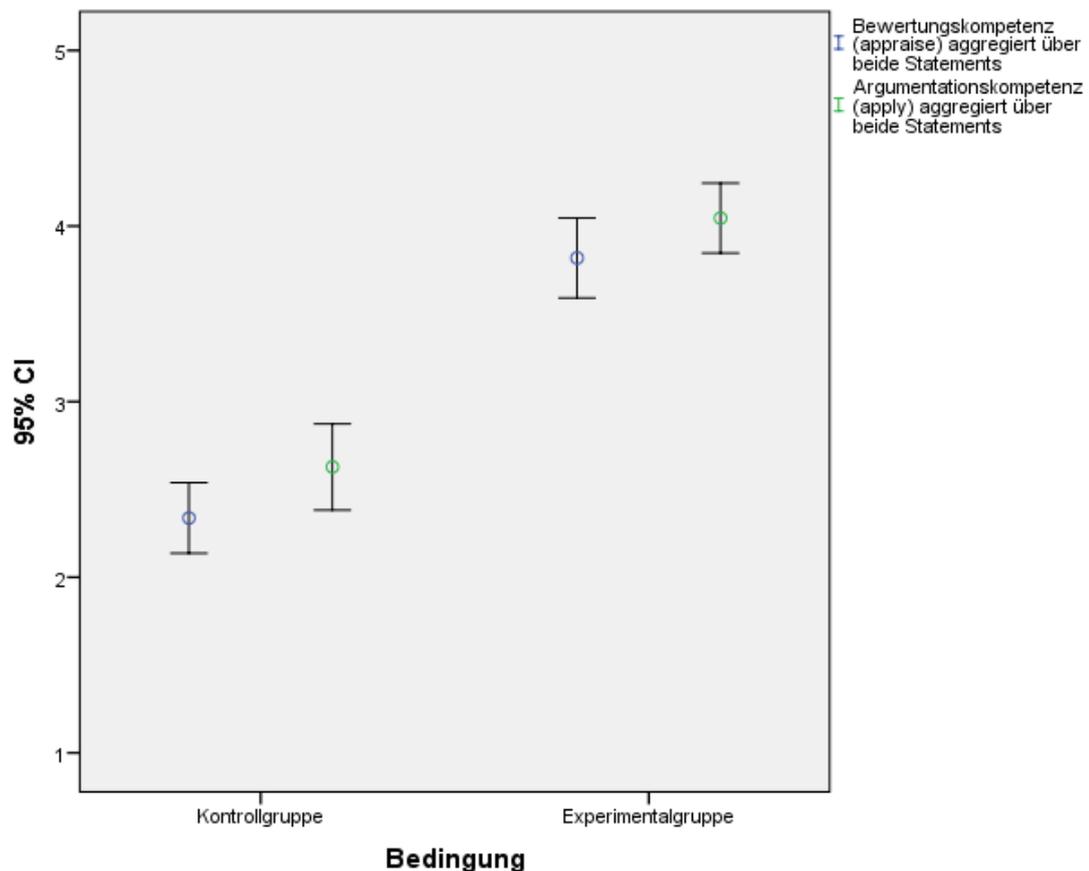


Abbildung 7: Fehlerbalkendiagramm: Vergleich der Ergebnisse der hochinferenten Globalratings beider Bedingungen, aggregiert über beide Statements

#### Ergebnisse der niedriginferenten Ratings

Auch in den niedriginferenten Ratings sind die Unterschiede zwischen Experimental (Ex)- und Kontrollgruppe (K) sichtbar (siehe Tabelle 15). Die Studierenden in der Experimentalgruppe belegten ihre Behauptungen häufiger mit Evidenz (Ex.:  $M = 49,24\%$  versus K.:  $M = 34,63\%$ ) und hinterfragten diese häufiger kritisch (Ex.:  $M = 22,36\%$  versus K.:  $M = 14,54\%$ ). Darüber hinaus enthielten ihre Argumente häufiger wissenschaftliche Annahmen über die Evidenzquelle (single, beziehungsweise multiple warrant) (Ex.:  $M = 71,95\%$  versus K.:  $M = 45,40\%$ ). Zudem verwendeten die Teilnehmenden häufiger die ihnen bereitgestellten wissenschaftlichen

Evidenzquellen (Ex.:  $M=71,95\%$  versus K.:  $M=46,10\%$ ) und verwendeten seltener anekdotische (persönliche) Evidenz (Ex.:  $M=4,45\%$  versus K.:  $M=12,75\%$ ). Außerdem waren die Argumente insgesamt klarer strukturiert (Ex.:  $M=3,32$  versus K.:  $M=2,51$ ) und es wurde häufiger beidseitig argumentiert (Ex.:  $M=84,85\%$  versus K.:  $M=71,35\%$ ). Im Falle von Uneinigkeit verwendeten die Studierenden in der Experimentalgruppe geringfügig seltener Behauptungen mit bloßer Ablehnung anstelle von Gegenargumenten oder begründeten Kritiken an Argumenten anderer (Ex.:  $M=4,15\%$  versus K.:  $M=8,90\%$ ).

*Tabelle 15:* Ergebnisse der niedriginferenten Ergebnisse aggregiert über beide Statements

<b>Variable</b>	<b>Experimentalgruppe</b> % bzw. <i>M</i>	<b>Kontrollgruppe</b> % bzw. <i>M</i>	<b>Ex vs. K</b> $p^8$
<b>Argumente</b>			
Behauptung	28.41%	50.83%	<.001
... mit Beleg	49.24%	34.63%	.001
...und kritischem Hinterfragen	22.36%	14.54%	=.342
<b>Warrant</b>			
keine	16.45%	35.75%	=.001
irrelevant	4.45%	15.05%	
beschreibend	7.15%	3.80%	
single warrant	71.95%	42.40%	
multiple warrant	0.00%	3.00%	
<b>Backing</b>			
keine	16.45%	36.60%	=.001
persönlich	4.80%	12.75%	
individuell	6.80%	4.60%	
instruiert	71.95%	46.10%	
<b>Länge</b>	2.01	2.14	=.329
<b>Klarheit</b>	3.32	2.51	<.001
<b>Position</b>			
one-sided	15.15%	28.70%	<.001
two-sided	84.85%	71.35%	
<b>Disagreement</b>			
Ablehnung	4.15%	8.90%	<.001
Counter-A	21.65%	35.40%	
Counter-C	68.45%	53.45%	
Counter-A und Counter-C	5.75%	2.25%	<.001
<i>Konsistenz der Ergebnisse</i>			

<sup>8</sup> Signifikanz der Skalen

Neben der Effektivität des Trainings war die Konsistenz der Kompetenzen über die verschiedenen Settings und Problemkontexte hinweg ein weiteres Ziel dieser Datenerhebung. Mit Hilfe des in Kapitel 8.2.2. beschriebenen Kodierschemas wurden die Kompetenzen in beiden schriftlichen Statements (je eine bildungswissenschaftliche Problemstellung), sowie in beiden Gruppendiskussionen (eine gesundheitswissenschaftliche und eine bildungswissenschaftliche Problemstellung) erhoben und konnten somit durch Intraklassen-Korrelationen (*ICC*) verglichen werden (siehe Tabelle 16).

*Tabelle 16:* Konsistenz der Ergebnisse auf Individualebene über die verschiedenen Settings und Problemkontexte hinweg

	<b>Statements<sup>9</sup></b>	<b>Gruppen- diskussionen<sup>10</sup></b>	<b>Statements und Diskussionen</b>
<b>Bewertungs- kompetenz</b>	<i>ICC</i> = .77	<i>ICC</i> = .49	<i>ICC</i> = .52
<b>Argumentations- kompetenz</b>	<i>ICC</i> = .78	<i>ICC</i> = .31	<i>ICC</i> = .25
<b>Über beide Kompetenzen</b>	<i>ICC</i> = .90	<i>ICC</i> = .41	<i>ICC</i> = .54

Die Ergebnisse innerhalb der schriftlichen Statements des Testinstruments sind auf Individualebene sowohl einzeln, als auch über beide Kompetenzen hinweg konsistent. Ein Studierender mit hoher Bewertungskompetenz in Statement eins zeigte demnach auch häufig eine hohe Bewertungskompetenz in Statement zwei (*ICC* = .77), analog dazu sind auch die Ergebnisse der Argumentationskompetenz (*ICC* = .78). Darüber zeigte sich eine Konsistenz beider Kompetenzen in den Ergebnissen des Testinstruments (*ICC* = .90). Ein Studierender mit hoher Bewertungskompetenz zeigte demzufolge auch zugleich eine hohe Argumentationskompetenz. In den Ergebnissen der Gruppendiskussionen konnte diese Konsistenz nicht festgestellt werden. Auch über beide Settings hinweg zeigte sich keine Konsistenz auf Individualebene.

### **8.3.2. Forschungsfrage 2: Einflussfaktoren**

Des Weiteren war es Ziel dieser Studie Informationen darüber zu generieren, ob Faktoren wie der Fachhintergrund (MINT versus nicht-MINT), wissenschaftstheoretisches Vorwissen, Lesegewohnheiten in Bezug auf

<sup>9</sup> Ergebnisse der Globalratings aus Statement 1 zu Statement 2

<sup>10</sup> Ergebnisse der Globalratings aus Gruppendiskussion 1 (GeWi) zu Gruppendiskussion 2 (BiWi)

wissenschaftliche Artikel, epistemologische Überzeugungen, Ambiguitätstoleranz, Überzeugungen zum Nutzen von Evidenz und (wahrgenommene) Barrieren zum Nutzen von Evidenz einen Einfluss auf den Trainingseffekt haben können. Die Regressionsmodelle 2 und 3 (siehe Tabelle 14) zeigen, dass entgegen vorheriger Annahmen keiner der Faktoren einen Einfluss auf die Kompetenzentwicklung hatte. Darüber hinaus zeigen sich auch im deskriptiven Vergleich keine signifikanten Unterschiede. So erreichten Studierende mit mindestens einem MINT Fach in den videografierten Gruppendiskussionen annähernd die gleiche Kompetenzstufe wie Studierende mit rein sprach- und geisteswissenschaftlichem Hintergrund (siehe Tabelle 17).

*Tabelle 17: Vergleich der Kompetenzen mit Domänenwissen (durch Fachhintergrund und Vorwissen) auf Basis der Ergebnisse aus den Gruppendiskussionen (aggregiert über beide Diskussionen)*

		<b>Bewertungskompetenz</b>	<b>Argumentationskompetenz</b>
<b>MINT</b>	ja	M = 3.97 (SD = .60)	M = 4.20 (SD = .74)
	nein	M = 3.87 (SD = .72)	M = 4.26 (SD = .52)
<b>Vorwissen</b>	gering <sup>11</sup>	M = 3.82 (SD = .92)	M = 4.06 (SD = .42)
	mittel <sup>12</sup>	M = 3.98 (SD = .66)	M = 4.27 (SD = .64)
	hoch <sup>13</sup>	M = 3.86 (SD = .66)	M = 4.21 (SD = .79)

Auch das wissenschaftstheoretische Vorwissen zeigte keine Auswirkungen auf die Kompetenz nach dem Training. Zudem wurden alle weiteren möglichen Einflussfaktoren einzeln kontrolliert, jedoch zeigte keiner dieser Faktoren einen Einfluss auf die Kompetenzstufen nach dem Training. Vergleicht man die Ergebnisse aus den schriftlichen Statements der Kontrollgruppe ( $N = 80$ ) miteinander, unterscheiden sich auch hier die Studierenden mit mindestens einem MINT Fach kaum von den sprach- und geisteswissenschaftlich geprägten Studierenden. Das Vorwissen jedoch zeigt einen messbaren Unterschied von bis zu einer ganzen Kompetenzstufe (am Beispiel Argumentationskompetenz: Vorwissen gering:  $M = 2.27$ ,  $SD = 1.03$  versus Vorwissen hoch:  $M = 3.20$ ,  $SD = .85$ ; siehe Tabelle 18).

<sup>11</sup> 14 Punkte oder weniger im Test zum wissenschaftstheoretischen Vorwissen erzielt

<sup>12</sup> 15 bis 19 Punkte im Test zum wissenschaftstheoretischen Vorwissen erzielt

<sup>13</sup> 20 Punkte oder mehr im Test zum wissenschaftstheoretischen Vorwissen erzielt (max. 24)

*Tabelle 18:* Vergleich der Kompetenzen vor dem Training mit Domänenwissen auf Basis der Ergebnisse aus den schriftlichen Statements (aggregiert über beide Statements)

		<b>Bewertungskompetenz</b>	<b>Argumentationskompetenz</b>
<b>MINT</b>	ja	M = 2.38 (SD = .76)	M = 2.62 (SD = 1.10)
	nein	M = 2.28 (SD = 1.07)	M = 2.64 (SD = 1.11)
<b>Vorwissen</b>	gering	M = 2.11 (SD = .74)	M = 2.27 (SD = 1.03)
	mittel	M = 2.30 (SD = .94)	M = 2.58 (SD = 1.14)
	hoch	M = 2.75 (SD = .87)	M = 3.20 (SD = .85)

## 8.4. Diskussion

Die Ergebnisse dieser Studie zeigen, dass sowohl die Kompetenz des kritischen Hinterfragens und Bewertens, als auch des argumentativen Nutzens wissenschaftlicher Evidenzquellen mit Hilfe des auf Heuristiken basierenden Trainings effektiv gefördert werden können. Dies zeigt sich vor allem durch die große Effektstärke des Training (Forschungsfrage 1). Obwohl die Teilnehmenden der Trainingsgruppe deutlich höhere Kompetenzstufen erreichten als die Teilnehmenden der Kontrollgruppe, erreichte dennoch, analog zu den Ergebnissen aus Studie 1, kaum einer die höchste Kompetenzstufe. Diese Stufe setzt nicht nur ein reines Anwenden der gelernten Heuristiken, sondern darüber hinaus ein elaboriertes Verständnis, sowie ein selbst-reflektiertes Hinterfragen voraus. Dies zu erreichen ist vermutlich nur durch mehrfaches Üben und Anwenden möglich und somit innerhalb eines zweitägigen Trainings kaum erreichbar. Darüber hinaus zeigten die Ergebnisse dieser Studie, dass weder Domänenwissen noch andere Faktoren wie (epistemologische) Überzeugungen, Einstellungen und Ambiguitätstoleranz einen Einfluss auf die nach dem Training erzielte Kompetenzstufe haben (Forschungsfrage 2). Diese Ergebnisse widersprachen zwar den zuvor aufgestellten Erwartungen, jedoch zeigen sie, dass diese basalen Kompetenzen zum Nutzen von Evidenz universell für Studierende jeglichen Hintergrunds trainierbar sein können.

Die Tatsache, dass Studierende nach einem zweitägigen Training in daran anschließenden Testsituationen in der Lage sind die soeben erlernten Heuristiken anzuwenden, bedeutet jedoch nicht, dass dies auch so in der zukünftigen Berufspraxis zur Anwendung kommt. Aus diesem Grund wäre eine Replikationsstudie mit entsprechenden Posttests notwendig um diesbezüglich Informationen erhalten zu können. Des Weiteren muss bei der Interpretation der

hohen Effektstärken des Trainings die Tatsache mit einbezogen werden, dass die Kontrollgruppe keinerlei Intervention vor dem Kompetenztest erhalten hat. Aus diesem Grund war ein Effekt, wenn auch nicht in dieser Größe, zu erwarten. Darüber hinaus lag der Fokus in der vorliegenden Arbeit auf der Förderung von zwei der für evidenzbasierte Praxis notwendigen fünf Kompetenzen (siehe Abbildung 3, Kapitel 2.3.1.). In einem nächsten Schritt kann es deshalb zudem sinnvoll sein, auch die drei weiteren Kompetenzen (dem Formulieren der Fragestellung, dem suchen nach der besten verfügbaren Evidenz und der Evaluation evidenzbasierter Entscheidungen) in ein Training mit einzubeziehen.

Trotz dieser Limitationen trägt diese Studie dazu bei, den aktuellen Forschungsstand durch Informationen dahingehend, wie ein möglichst effektives und (zeit-)effizientes Kurzzeittraining gestaltet werden kann um die notwendigen Kompetenzen zum Nutzen von Evidenz zu fördern, zu erweitern. Das erfolgreiche Einsetzen der Heuristiken zeigt, dass Lehramtsstudierende kein (forschungsmethodisches) Expertenwissen benötigen um Evidenz – zumindest auf einer basalen Ebene – situationsangemessen kompetent kritisch hinterfragen und argumentativ nutzen zu können. Folglich können Trainings dieser Art dazu beitragen dem Nutzen wissenschaftlicher Evidenz in der Praxis einen Schritt näher zu kommen.

## 9. Studie III: Tiefergehende Analyse der im Training videografierten Gruppendiskussionen

### 9.1. Forschungsfragen und Hypothesen

Ziel dieser Studie war es, durch eine tiefergehende Analyse der videografierten Gruppendiskussionen detailliertere Informationen dahingehend zu generieren, wie genau sich erfolgreiches Nutzen von Evidenz bei Lehramtsstudierenden innerhalb von Gruppendiskussionen äußern kann und was andererseits besonders häufig auftretende Probleme und Fehler sind. Frühere Studien mit Elementen dialogischer Diskussionen liefern erste Hinweise dahingehend, worin diese bestehen können (vergleiche Kapitel 3.2.4.). Ziel ist es zu eruieren, ob sich diese auch in Gruppendiskussionen bei Lehramtsstudierenden zeigen, sowie ob darüber hinaus noch weitere Schwierigkeiten und Probleme auftreten. Darüber hinaus wurde untersucht, aus welchen Gründen die Studierenden Evidenz als Beleg ihrer Argumente hinzugezogen haben und wie sich dies auf den allgemeinen Verlauf der Diskussion auswirkt (vor allem die Transaktivität betreffend). Aus diesem Grund wurden die folgenden Forschungsfragen untersucht:

#### 9.1.1. Beispiele Erfolgreichen Nutzens

*Forschungsfrage 1:* Wie zeigt sich erfolgreiches Nutzen von Evidenz in Gruppendiskussionen?

Zur Beantwortung dieser Forschungsfrage sollen typische qualitative Beispiele aus den transkribierten Diskussionen dargestellt werden. Ausgewählt wurde diese mit Hilfe der durch das Kodierschema erzielten Werte wie *Kompetenzstufe*, *Transaktivität* und *Funktion*.

#### 9.1.2. Häufige Fehler und Probleme

*Forschungsfrage 2:* Was sind häufige Fehler und Schwierigkeiten beim Nutzen wissenschaftlicher Evidenz in Gruppendiskussionen?

Hypothese 2.1.: Studierende haben häufig Schwierigkeiten damit, ihre Behauptung mit angemessener Evidenz sinnvoll zu verknüpfen.

Hypothese 2.2.: Evidenz wird hauptsächlich dazu benutzt, um die eigene Meinung zu stützen und dadurch bestimmte Ergebnisse gezielt bevorzugt.

Hypothese 2.3.: Ergebnisse der Evidenzquelle werden missinterpretiert, so dass sie zur eigenen Behauptung passen.

Hypothese 2.4.: Die (evidenzbasierten) Argumente der Mitdiskutanten werden oftmals übergangen oder nicht beachtet.

Hypothese 2.5.: Die Studierenden verwenden häufig ausschließlich anekdotische Evidenz um ihre Behauptung zu belegen.

### **9.1.3. Gründe und Funktionen**

*Forschungsfrage 3:* Aus welchen Gründen nutzen Lehramtsstudierende Evidenz in Gruppendiskussionen?

*Hypothese 3.1.:* Evidenz wird hauptsächlich dazu verwendet, um die eigenen Behauptungen zu stützen.

*Hypothese 3.2.:* Evidenz wird zudem häufig dazu genutzt, um Argumente von Mitdiskutanten zu stützen oder zu schwächen.

*Hypothese 3.3:* Evidenz wird kaum dazu verwendet, die eigene Meinung kritisch zu hinterfragen oder diese bei der eigenen Meinung widersprechender Studienlage sogar zu ändern.

### **9.1.4. Transaktivität**

*Forschungsfrage 4:* Gibt es einen Zusammenhang zwischen erfolgreichem Nutzen von Evidenz und Transaktivität innerhalb der Gruppendiskussionen?

*Hypothese 4.1.:* Es besteht ein Zusammenhang zwischen Transaktivität und dem erfolgreichen Nutzen von Evidenz.

*Hypothese 4.1.:* Studierende haben Schwierigkeiten dabei, sich gleichzeitig einerseits auf das Formulieren eigener evidenzbasierter Argumente zu konzentrieren und andererseits transaktiv und kritisch auf die Argumente ihrer Mitdiskutanten einzugehen. Deshalb besteht die Annahme, dass die Transaktivität geringer ausfällt, je höher die Qualität der (evidenzbasierten) Argumente ist.

## **9.2. Methoden**

### **9.2.1. Stichprobe und Design**

Diese Studie wurde anhand eines Subsamples der Haupterhebung (Studie 2) durchgeführt. Im Rahmen der Haupterhebung wurden insgesamt 33 Gruppen á 3-5 Teilnehmenden videografiert und mit Hilfe des in Studie 2 beschriebenen Kodierschemas ausgewertet. Auf Basis dessen wurden die Teilnehmenden hinsichtlich ihrer Bewertungs- und Argumentationskompetenz auf Gruppenebene in drei Kategorien eingeteilt: 11 Gruppen mit den insgesamt niedrigsten Kompetenzniveaus, 11 Gruppen mit mittleren Werten und 11 Gruppen mit den insgesamt höchsten Kompetenzniveaus. Aus diesen 3 Leistungsgruppen wurden je 4 Gruppen zufällig gezogen. Aufgrund des großen Trainingseffektes waren die Unterschiede zwischen den Leistungsgruppen jedoch eher gering ausgeprägt (LG1

(hoch):  $M^{14} = 4,64$ ,  $SD = .41$ ; LG2 (mittel):  $M = 4,02$ ,  $SD = .38$ ; LG3 (niedrig):  $M = 3,34$ ,  $SD = .27$ ). Folglich erzielten die Teilnehmenden aus Leistungsgruppe 3 zwar leicht niedrigere Ergebnisse als die übrigen Teilnehmenden, befanden sich aber dennoch durchschnittlich zwischen Kompetenzstufe 3 und 4 und somit auf einem mittleren bis hohen Kompetenzniveau. Auf diese Weise entsandt ein Subsample aus  $N = 48$  Teilnehmenden und  $N = 12$  Gruppen mit jeweils zwei Diskussionen (vergleiche Tabelle 5 in Kapitel 5.2.). Diese insgesamt 24 ca. dreißigminütigen videografierten Gruppendiskussionen wurden vollständig transkribiert.

### 9.2.2. Instrumente

Um die transkribierten Gruppendiskussionen auf Individualebene hinsichtlich der oben beschriebenen Forschungsfragen auszuwerten, wurden zusätzlich zu der Auswertung in Studie 2 drei zusätzliche Aspekte mit Hilfe von weiteren Kodierschemata auf Statement-Ebene erhoben. So wurde bei jedem neuen Statement festgehalten, ob dieses evidenzbasiert ist und sich inhaltlich auf das vorherige bezieht und somit transaktiv ist oder nicht (*Transaktivität: ja/nein* in Kombination mit *Evidenz: ja/nein*). Darüber hinaus wurde ein Kodierschema nach Hemberger et al. (2015) adaptiert, um die Gründe der Evidenznutzung in den einzelnen Argumenten zu erheben (siehe Tabelle 19). Dadurch konnte für jedes einzelne Statement bestimmt werden, ob es sich um ein Argument mit erfolgreichem Nutzen von Evidenz handelt und wenn dies zutrifft, aus welchem Grund die Evidenz verwendet wurde. Hierbei wurde zwischen vier Hauptgründen unterschieden: Das Nutzen von Evidenz um die eigene Position zu stärken (M+), um die Position anderer zu schwächen (O-), um die Position anderer zu stärken (O+) und um die eigene Position zu schwächen (M-) (vergleiche Kapitel 3.2.3.).

Zudem wurde mit Hilfe eines weiteren etablierten Auswertungsschemas nach Kuhn & Crowell, 2011 die exakte Funktion jedes einzelnen Statements bestimmt (siehe Tabelle 20). Diese insgesamt 26 Funktionen lassen sich in Fragen, transaktive Statements und nicht-transaktive Statements untergliedern. Mit Hilfe der Transkription und der dadurch ermöglichten tiefergehenden Analyse der Gruppendiskussionen können reale Beispiele erfolgreichen evidenzbasierten Argumentierens identifiziert werden. Ziel ist es dabei nicht lehrbuchartige und perfekte Argumente zu verfassen, sondern eine professionelle Diskussion über fachliche Problemstellungen mit dem situationsangemessenen Verwenden relevanter Evidenz zu führen. Da die Möglichkeiten hierfür vielfältig sind und es somit keine allgemeine Musterlösung geben kann, sollen die ausgewählten Argumentationen als Ankerbeispiele dafür dienen, wie sich erfolgreiches Nutzen von

---

<sup>14</sup> Ergebnisse aus den Globalratings, aggregiert über beide Diskussionen und Kompetenzen

Evidenz nach einem zweitägigen Kurzzeittraining bei Lehramtsstudierenden konkret zeigen kann.

*Tabelle 19: Kodierschema zur Bestimmung der Qualität, sowie der Gründe der Evidenznutzung*

<b>Level</b>	<b>Beschreibung</b>
1	<i>Erfolgreiches Nutzen von Evidenz um eine Behauptung zu untermauern</i>
1.1a	...um die eigene Position zu bestärken (M+)
1.1b	...um die Position anderer zu schwächen (O-)
1.1c	...um die Position anderer zu unterstützen (O+)
1.1d	...um die eigene Position zu schwächen (M-)
1.2.	<i>Erfolgreiches Nutzen von Evidenz um eine Behauptung zu untermauern, teilt aber die Meinung der Evidenzquelle nicht</i>
1.2a	...um die eigene Position zu bestärken (M+)
1.2b	...um die Position anderer zu schwächen (O-)
1.2c	...um die Position anderer zu unterstützen (O+)
1.2d	...um die eigene Position zu schwächen (M-)
2.	<i>Versuch aber dennoch kein erfolgreiches Nutzen von Evidenz um eine Behauptung zu untermauern</i>
2.1.	Es ist unklar warum die Evidenzquelle zitiert wird (Passung zur Behauptung)
2.2.	Die Evidenzquelle wird fehlinterpretiert bzw. eine falsche Schlussfolgerung gezogen
3	Bloßes zitieren einer Evidenzquelle (ohne Behauptung)
4	Falsches zitieren oder fehlinterpretieren der Evidenzquelle (ohne eine Behauptung)

*Tabelle 20:* Kodierschema zur Bestimmung der Funktion der Evidenznutzung (Kuhn & Crowell, 2011)

<b>Transaktive Fragen</b>		
1.1	Agree-?	Nachfrage, ob eigenen Behauptung zugestimmt wird
1.2	Case-?	Bitte eine bestimmte Position in einem bestimmten Szenario oder Fallbeispiel einzunehmen
1.3	Clarify-?	Bitte um Klärung, nachfrage
1.4	Justify-?	Frage nach Evidenz/Beleg für die getätigte Aussage
1.5	Meta-?	Eine Frage die die Diskussion selbst betrifft (nicht deren Inhalt)
1.6	Position-?	Bitte eine Position zu beziehen, Frage nach Position/ Meinung
1.7	Question-?	Informelle Frage die sich nicht auf die vorherige Äußerung der anderen bezieht
1.8	Respond-?	Bitte um Antwort
<b>Transaktive Statements</b>		
2.1	Counter-A	Zeigt ‚nicht-zustimmen‘ durch Gegenargument
2.2	Counter-C	Zeigt ‚nicht-zustimmen‘ durch Kritik an dem Argument des anderen
2.3	Disagree	Zeigt ‚nicht-zustimmen‘ ohne Gegenargument oder Kritik
2.4	Dismiss	Gibt an, dass man die zuvor getätigte Aussage des anderen für irrelevant hält
2.5	Clarify	Klärung der eigenen Position, Klärung von Fragen, klarstellen
2.6	Agree	Zustimmung
2.7	Advance	Ergänzung zu einer zuvor von anderen angeführten Arguments
2.8	Add	Ergänzung zu einer zuvor von anderen getroffenen Aussage
2.9	Aside	Kommentar, der nicht die Aussage anderer ergänzt
2.10	Coopt	Behauptung, dass die zuvor getätigte Aussage der anderen entgegen der eigentlichen Position ist
2.11	Interpret	Paraphrase
2.12	Meta	Äußerung über die Diskussion selbst (nicht deren Inhalt)
2.13	Null	Äußerung, die nicht zum Thema gehört
2.14	Refuse	Explizites Ablehnen auf eine Frage zu antworten
2.15	Substantiate	Eine Äußerung die anbietet die zuvor getätigte Aussage der anderen zu unterstützen
<b>Nicht-transaktive Statements</b>		
3.1	Counter-A	Gegenargument
3.2	Continue	Weiterführen der zuvor getätigten eigenen Aussage ohne darauf zu achten was andere in der Zwischenzeit gesagt haben
3.3	Unconnected	Eine Aussage die sich nicht auf die zuvor getätigten Äußerungen der anderen bezieht

### 9.2.3. Statistische Analysen

Die Kodierungen auf Statement-Ebene wurden anhand der Transkripte von zwei geschulten und unabhängigen Ratern, die zudem keine Informationen dahingehend hatten welches Kompetenzniveau die Gruppe in dem Globalrating in Studie 2 erreichte, mit Hilfe der in Kapitel 8.2.2. beschriebenen Kodierschemata ausgewertet. Dabei wurde eine hohe Interrater-Reliabilität über alle Items hinweg erreicht ( $ICC = .88$ ; range  $.76 < .98$ ). Im Falle von Uneinigkeit der Rater wurde mit Hilfe eines dritten Raters eine Konsensvalidierung durchgeführt. Aufgrund der kleineren Stichprobe wurden zur tiefergehenden Analyse der Gruppendiskussionen sowohl qualitative als auch quantitative Ansätze und Methoden gewählt. So wurden sowohl quantitative Analysen auf Basis der Kodierschemata durchgeführt (Forschungsfragen 2, 3 und 4), als auch qualitative Untersuchungen bei detaillierterer Analyse der Transkripte und einzelner Statements (Forschungsfragen 1 und 2).

## 9.3. Ergebnisse

### 9.3.1. Forschungsfrage 1: Beispiele erfolgreichen Nutzens

Die Tabellen 21 und 22 zeigen zwei Beispiele aus der Gruppendiskussion über die gesundheitswissenschaftliche Problemstellung, in welcher ein möglicher Zusammenhang zwischen Hypertonie und erhöhter Kochsalzzufuhr anhand aktueller Forschung diskutiert wurde. Beide Beispiele zeigen Ausschnitte, in welchen die Teilnehmenden Aspekte der Evidenzquellen, wie etwa die Wahl der Instrumente (Tabelle 21) oder das Studiendesign (Tabelle 22) kritisch hinterfragen und die Studienergebnisse nutzen, um auf Basis dessen gemeinsam mit ihren Mitdiskutanten zu einer fundierten Entscheidung zu gelangen.

Tabelle 23 zeigt einen Ausschnitt aus einer Gruppendiskussion mit bildungswissenschaftlicher Problemstellung. Hier wurde darüber diskutiert, ob ein, wie von einem fiktiven Kollegen vorgeschlagener extra Kurs in Informatik speziell für Mädchen angeboten werden soll (siehe auch Kapitel 5.2.). Auch in diesem Beispiel wird deutlich wie es gelingen kann, konstruktiv mit den konfligierenden und fragilen Elementen der Evidenzquellen umzugehen und die Forschungsergebnisse in Einklang mit eigenen Erfahrungen und Überzeugungen zu bringen.

*Tabelle 21:* Beispiel für erfolgreiches evidenzbasiertes Argumentieren (Domäne: Gesundheitswissenschaften)<sup>15</sup>

<b>Gesundheitswissenschaften: Hypertonie und überhöhte Salzzufuhr</b>		
<b>TN Code</b>	<b>Statement</b>	<b>Rating</b>
164	Ich denke dass eine zu hohe Salzzufuhr schädlich ist für den Körper, das belegt auch die Studie XXX, die haben das mit einer Langzeitstudie über insgesamt 20 Jahre belegen können.	M+, Ev.
172	Aber diese Langzeitstudie war doch in Taiwan oder? Das muss doch dann nicht zwangsweise für uns hier auch gelten, oder?	Counter C, Tr.
159	Nee, des ist global. Das war in Japan, Taiwan und Deutschland	Clarify, Tr.
165	Das stimmt, aber in der Studie XXX ist herausgefunden worden, dass das Vorgehen dort, also die Salzzufuhr über Urinausscheidung zu messen, nicht richtig ist, da das nicht konstant ist. In deiner Studie ist das aber damit gemessen worden. Deshalb kann man nicht sagen dass es zwangsweise am Salz liegt.	Counter C, O-, Ev., Tr.

*Tabelle 22:* Beispiel für kritisches Hinterfragen von Evidenz (Domäne: Gesundheitswissenschaften)

<b>Gesundheitswissenschaften: Hypertonie und überhöhte Salzzufuhr</b>		
<b>TN Code</b>	<b>Statement</b>	<b>Rating</b>
165	Ähm aber ich hätte eine Frage zu der Studie auf die du dich beziehst. Also wie und was haben die gemacht? Also was war das Design?	Clarify-?, Tr.
159	Das war eine experimentelle Studie...	M+, Clarify, Tr.
172	Aber es gab doch keine Kontrollgruppe...	O-, Counter-C, Tr..
164	Und die Stichprobe war eigentlich viel zu klein und es sind nur Männer...	O+, Add, Tr.
165	Also ist das nicht repräsentativ für die Gesellschaft	O-, Add, Tr.

<sup>15</sup> Teilnehmende aus Leistungsgruppe 1 (höchstes Kompetenzniveau)

Tabelle 23: Beispiel für erfolgreiches evidenzbasiertes Argumentieren (Domäne: Bildungswissenschaften)<sup>16</sup>

<b>Bildungswissenschaften: Mädchen und Computer – erhöhter Förderbedarf?</b>		
<b>TN Code</b>	<b>Statement</b>	<b>Rating</b>
70	Der Kollege Müller der hat ja gesagt, die Mädchen, die, machen bei uns nicht so mit in Informatik und ich finde auch, dass wir dagegen was tun können. Aufgrund der mir vorliegenden Metastudie, ähm, wäre es jetzt plausibel, ein Extrakurs für Mädchen anzubieten so wie der Markus das vorgeschlagen hat.	O+, M+, Ev.
67	Hier in meinem Text wird Mathematik und Physik, hat sich`s aber herausgestellt, dass Mädchen und Jungs eigentlich gleichwertig agieren und, und auch gleichwertig gut sind.	Counter-A, Ev., Tr.
69	Ähm, ich frage mich halt, wenn wir sagen würden, Mädchen fördern wir speziell, ob wir damit dann nicht gerade diese Geschlechterrollen, diese Stereotypen, die ja irgendwo noch in den Köpfen verhaftet sind, weiter fördern, weil wir damit irgendwo noch demonstrieren, ok, Mädchen können das prinzipiell nicht so gut also fördern wir sie. Wenn ich das jetzt mal aus der Perspektive mit den Fächern Mathe und Chemie sehen kann auch, auch im Hintergrund vielleicht die Studie, wobei ich die ehrlich gesagt nicht so sonderlich aussagekräftig finde	M+, Counter-C, Advance, Ev., Tr.
71	Also ich glaube eben, dass diese, diese Stereotypenbilder ganz stark noch verhaftet sind. Und das hat auch so ein bisschen im Ansatz die Studie angeschnitten, dass das so ist, dass sich die Mädchen eben da wesentlich schlechter einschneiden, äh, einschätzen. Was auch deine Studie zeigt. Und ich frag mich halt, wenn wir sagen, wir trennen das, ob wir dann nicht irgendwo das weiter fördern.	M+, Add, Ev., Tr.

### 9.3.2. Forschungsfrage 2: Häufige Probleme und Fehler

Am häufigsten hatten die Studierenden dieser Studie Schwierigkeiten damit, ihre Behauptung sinnvoll mit der Evidenz zu verbinden ( $M = 77\%$ ). Dabei wurde entweder die Evidenzquelle lediglich zitiert ohne diese mit einer Behauptung zu verknüpfen ( $M = 28\%$ ) (siehe Tabelle 19, Stufe 3) oder eine Behauptung ohne jeglichen Beleg aufgestellt ( $M = 49\%$ ). Des Weiteren fehlte häufig der Bezug zu den vorherigen Argumenten der Mitdiskutierenden (*Transaktivität*) ( $M = 76\%$ ). Darüber hinaus wurde häufig ausschließlich anekdotische Evidenz verwendet um eine

<sup>16</sup> Teilnehmende aus Leistungsgruppe 1 (höchstes Kompetenzniveau)

Behauptung zu belegen ( $M = 73\%$ ). Zudem nutzten die Studierenden in  $M = 41\%$  der Fälle ausschließlich Evidenz, die ihre persönliche Meinung stützt, was eine fehlende Neutralität bedingt. Auch der Aufbau eines klar strukturierten und nachvollziehbaren Arguments bereitete den Teilnehmenden zum Teil Schwierigkeiten ( $M = 29\%$ ). In  $M = 5\%$  der Fälle wurde die Evidenz dahingehend fehlinterpretiert, so dass sie passend zur eigenen Behauptung ist (siehe Tabelle 19, Stufe 4). In den Tabellen 24 und 25 werden exemplarisch einige der häufigsten Probleme und Schwierigkeiten beim argumentativen Nutzen von Evidenz dargestellt.

*Tabelle 24: Beispiele aus den Diskussionen für häufige Fehler und Schwierigkeiten<sup>17</sup>*

### **Schwierigkeiten bei der Verknüpfung von Behauptung und Evidenz**

*„Also in der Studie steht dass, ...“ (Reines Zitat, ohne Anschluss oder Zusammenhang)*

---

#### **Evidenz fehlinterpretieren, so dass sie zur eigenen Behauptung passt**

*„Die Amerikaner ernähren sich ja prinzipiell alle ungesund mit zuviel Salz weil die alle viel zu viele Burger essen. Das steht ja auch hier in der Studie.“ (Anm: Diese Aussage befand sich in keiner der Studien)*

*„Genau. Und es ist ja auch erwiesen – also kognitive Studien gibt’s ja da genug – dass Männer oder Jungs, das männliche Geschlecht eher visuell denkt. Und dass es dann möglich wär, dass eben dadurch, dass wir so viel mehr Grundschullehrerinnen als Grundschullehrer haben der Leseunterricht schon vorgeprägt ist“*

---

#### **Ausschließlich anekdotische Evidenz**

*„Jungs haben keine speziellen Bedürfnisse im Informatikunterricht, bei mir hat sich das auch ganz natürlich entwickelt“*

---

#### **Unklare Argumentationsstruktur**

*„Natürlich, klar, wenn’s so um sozioökonomische Faktoren geht, dann kann man das ja unmöglich pauschalisieren. Das ist aber glaub ich ein generelles Problem, also wenn wir das jetzt mal auf Deutschland beziehen“*

---

<sup>17</sup> Diese Aussagen stammen von Teilnehmenden aller drei Leistungsgruppen

Tabelle 25: Beispiel für fehlende Transaktivität<sup>18</sup>

<b>Bildungswissenschaften: Mädchen und Computer – erhöhter Förderbedarf?</b>		
<b>TN Code</b>	<b>Statement</b>	<b>Rating</b>
049	Artikel XY sagt dass es viel zu wenige weibliche Rollenvorbilder im Informatikunterricht gibt.	Ev., bloßes Zitieren
054	Wir haben damals im Informatikunterricht Mandalas gemalt	Unconnect., Anekdot. Ev.
049	Deshalb müsste man vielleicht mehr Frauen anstellen	Continue
084	Haha cool das durften wir nicht	Continue2

### 9.3.3. Forschungsfrage 3: Gründe und Funktionen

Die Studierenden verwendeten Evidenz überwiegend um ihre eigenen Meinungen und Behauptungen zu stützen (Diskussion 1:  $M = 71,40\%$ ; Diskussion 2:  $M = 76,40\%$ ). Der zweithäufigste Grund war, die Argumentation eines Mitdiskutanten zu unterstützen (Diskussion 1:  $M = 19,80\%$ ; Diskussion 2:  $M = 18,10\%$ ). Weniger häufig nutzten die Teilnehmenden Evidenz um die Position der Mitdiskutanten zu widerlegen oder zu schwächen (Diskussion 1:  $M = 9,50\%$ ; Diskussion 2:  $M = 5,50\%$ ). Das Nutzen von Evidenz zum Hinterfragen der eigenen Position war dagegen selten bis gar nicht vorhanden (Diskussion 1:  $M = 0,30\%$ ; Diskussion 2:  $M = 0,00\%$ ) (siehe Tabelle 26).

Tabelle 26: Gründe für das argumentative Nutzen von Evidenz

	<b>Diskussion 1 (GeWi<sup>19</sup>)</b>	<b>Diskussion 2 (BiWi<sup>20</sup>)</b>
	<b><i>M</i></b>	<b><i>M</i></b>
<b>M+</b> <sup>21</sup>	71,40%	76,40%
<b>M-</b> <sup>22</sup>	0,30%	0,00%
<b>O+</b> <sup>23</sup>	19,80%	18,10%
<b>O-</b> <sup>24</sup>	9,50%	5,50%

Vergleichend über die drei Leistungsgruppen hinweg zeigt sich, dass die Teilnehmenden aus Leistungsgruppe 1 (höchstes Kompetenzniveau) häufiger

<sup>18</sup> Teilnehmende aus Leistungsgruppe 3 (niedrigstes Kompetenzniveau)

<sup>19</sup> Gesundheitswissenschaftliche Problemstellung (Zusammenhang zwischen Hypertonie und erhöhter Salzzufuhr)

<sup>20</sup> Bildungswissenschaftliche Problemstellung (Förderbedarf von Mädchen im Umgang mit Computern)

<sup>21</sup> Unterstützen der eigenen Behauptung

<sup>22</sup> Schwächen oder Hinterfragen der eigenen Behauptung

<sup>23</sup> Stärken der Behauptung anderer

<sup>24</sup> Schwächen der Behauptung anderer

Evidenz verwendet haben um ihre eigene Meinung zu bestärken (M+) und weniger häufig um die Aussagen anderer zu stärken (O+) oder zu schwächen (O-).

Tabelle 27: Gründe für das Nutzen von Evidenz aufgeteilt nach Leistungsgruppen

	Leistungsgruppe 1 <sup>25</sup>	Leistungsgruppe 2	Leistungsgruppe 3
	<i>M</i>	<i>M</i>	<i>M</i>
<b>M+</b>	81,40%	69,30%	71,0%
<b>M-</b>	0,30%	0,00%	0,00%
<b>O+</b>	14,80%	18,20%	22,5%
<b>O-</b>	3,50%	12,50%	6,5%

Neben den Gründen für das argumentative Nutzen von Evidenz wurde zusätzlich auch die Bedeutung der (evidenzbasierten) Aussagen kategorisiert und ausgewertet (vergleiche Tabelle 20). Wenn Fragen in Bezug auf die Evidenzquellen gestellt wurden, waren dies überwiegend Verständnisfragen (*Clarify, Question & Respond*), sowie die konkrete Nachfrage nach Belegen für die soeben getätigte Aussage (*Justify*) (siehe Abbildung 8).

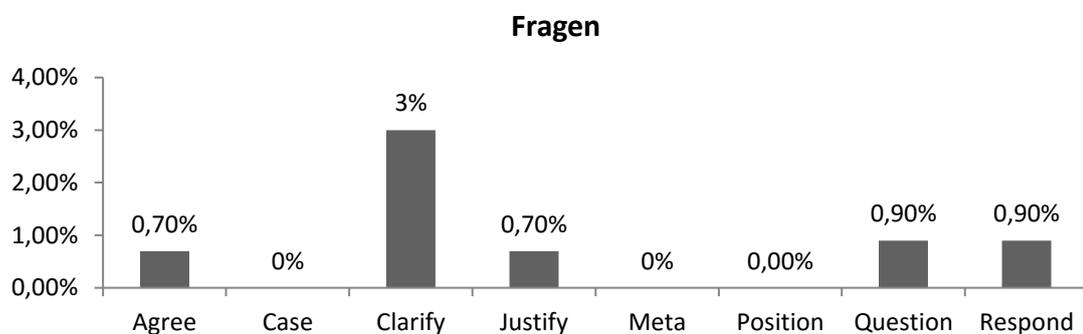
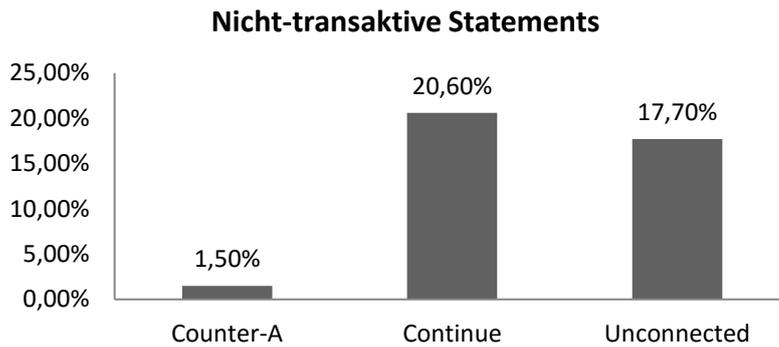


Abbildung 8: Funktionen des argumentativen Nutzens von Evidenz: Fragen

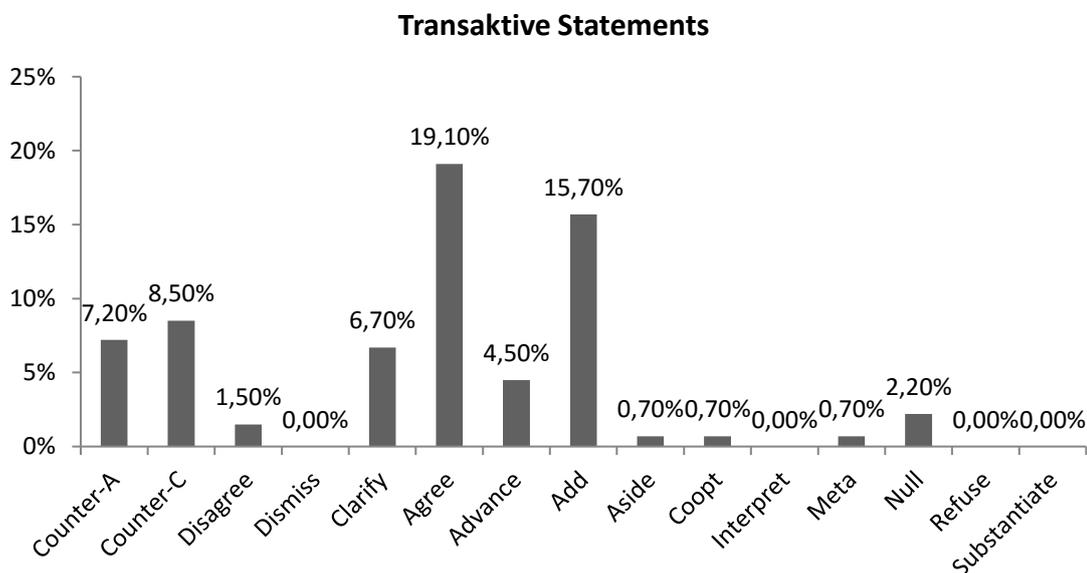
Bei nicht-transaktiven Statements wurde häufig der eigene Argumentationsstrang weitergeführt, ohne auf zwischenzeitlich getätigte Argumente und Aussagen der Mitdiskutierenden einzugehen (*Continue*). Oftmals wurde auch ein neues Argument ohne Zusammenhang zum vorher Diskutierten angeführt (siehe Abbildung 9).

<sup>25</sup> Ergebnisse aggregiert über beide Diskussionen



**Abbildung 9:** Funktionen des argumentativen Nutzens von Evidenz: nicht-transaktive Statements

Bei transaktiven Statements waren die häufigsten Funktionen das Zustimmung für (*Agree*) und Ergänzen (*Add*) von Aussagen und Argumenten anderer. Des Weiteren wurde häufig Evidenz genutzt um ein Gegenargument (*Counter-A*), oder eine begründete Kritik an einem Argument der Mitdiskutierenden (*Counter-C*) zu äußern (siehe Abbildung 10).



**Abbildung 10:** Funktionen des argumentativen Nutzens von Evidenz: transaktive Statements

#### 9.3.4. Forschungsfrage 4: Zusammenhang Transaktivität und Evidenznutzung

Das erfolgreiche Nutzen von Evidenz innerhalb eines Arguments und die Tatsache ob ein Argument sich auf das vorher gesagt der Mitdiskutierende bezieht korrelieren signifikant negativ miteinander (Gesamt:  $r = -.36$ ,  $p < .001$ ; LG1:  $r = -.11$ ,  $p = .69$ ;

LG2:  $r = -.56$ ;  $p = .03$ ; LG3:  $r = -.18$ ,  $p = .49$ ). Das bedeutet, dass im Falle von erfolgreichem Nutzen von Evidenz häufig nicht darauf geachtet wurde, was die Diskussionspartner vorher argumentiert haben. In den Transkripten zeigen sich hier teilweise deutlich erkennbare „parallele Monologe“ (vergleiche auch Tabelle 25).

## 9.4. Diskussion

Kompetentes Nutzen von Evidenz in (Gruppen-)Diskussionen ist eine wichtige Voraussetzung für evidenzbasierte Praxis in der Lehrerverberufung. Im Berufsalltag kann dies für Lehrkräfte beispielsweise bei Elterngesprächen, bei Lehrerkonferenzen, oder im Austausch mit Kollegen von Relevanz sein. Inwieweit man Lehramtsstudierende im Studium auf diese Situation vorbereiten kann ist noch weitgehend unerforscht. Zwar gibt es einige Studien zum evidenzbasierten oder wissenschaftlichen Argumentieren in dialogischen Diskussionen (z.B. Billig, 1987; Kuhn, 2005; Graff, 2003; Teasley 1997; Asterhan & Schwarz, 2009; Felton & Kuhn, 2001; Rips, Brem, & Bailenson, 1999), jedoch gibt es noch keine gezielten Interventionen zur Förderung des evidenzbasierten Argumentierens in Gruppen (mit mehr als zwei Diskutierenden) bei Lehramtsstudierenden. Ziel dieser Studie war es deshalb, erste Informationen darüber zu generieren, wie und aus welchen Gründen Lehramtsstudierende Evidenz argumentativ in Gruppendiskussionen nutzen, sowie welche Schwierigkeiten und Probleme dabei häufig auftreten. Auf Basis dieser Informationen können weitere Ansätze zur Verbesserung und Erweiterung eines Trainings zum kompetenten Nutzen von Evidenz für Lehramtsstudierende entwickelt werden.

Erfolgreiches Nutzen von Evidenz in Diskussionen bedeutet jedoch nicht, dass diese (rein mechanisch) für jedes einzelne Argument genutzt, sondern situationsangemessen und sinnvoll eingesetzt wird. Um ein besseres Verständnis dafür zu generieren, wie sich ein solches sinnvolles und situationsangemessenes Nutzen von Evidenz äußern kann, wurden zunächst einige positive qualitative Beispiele aus den Transkriptionen der videografierten Diskussionen dargestellt (Forschungsfrage 1). Diese Beispiele können sowohl zu einem besseren Verständnis dahingehend beitragen, wie sich kompetentes argumentatives Nutzen von Evidenz in der Praxis äußern kann, als auch als Ankerbeispiele für zukünftige Trainings sowie Kodierschemata dienen. Ergänzend dazu können Informationen und konkrete Beispiele zu den häufigsten Fehlern und Schwierigkeiten, die bei Lehramtsstudierenden auftreten, genutzt werden (Forschungsfrage 2). Auch die detaillierte Analyse der Funktionen und Gründe (Forschungsfrage 3), aus welchen Studierende Evidenz in Gruppendiskussionen nutzen kann relevante Informationen

zur Weiterentwicklung entsprechender Trainingsprogramme bereitstellen. Wie erwartet haben die Studierenden in dieser Studie die ihnen vorliegende Evidenz hauptsächlich dazu verwendet, um ihre eigene Position zu stärken. Selten bis gar nicht jedoch wurde Evidenz verwendet um die eigene Meinung zu hinterfragen und möglicherweise zu ändern. Daraus lässt sich schließen, dass in zukünftigen Trainings noch weitere Lerngelegenheiten für einen reflektierten und elaborierten Umgang mit Evidenz notwendig sind. Das Üben eines selbstreflektierten Umgangs mit Evidenz kann darüber hinaus auch dem entgegenwirken, dass Evidenz ausschließlich gezielt hinsichtlich der Resultate ausgewählt wird um die eigene Meinung zu stärken. Ist dies der Fall, findet keine objektive Betrachtung statt, sondern eine missbräuchliche Anwendung von Evidenz. Dass Studierende weitere Lerngelegenheiten benötigen, die über ein zweitägiges Training hinausgehen bestätigt auch das Ergebnis, dass Studierende deutlich weniger transaktiv argumentieren (sich also deutlich weniger auf die Inhalte der Diskussion und die Argumente ihrer Mitdiskutierenden konzentrieren), je qualitativ hochwertiger ihre eigenen Argumente sind (Forschungsfrage 4).

Bei der Interpretation der Ergebnisse muss jedoch beachtet werden, dass sich diese auf ein instruiertes Setting im Rahmen eines Trainings, bei welchem die Lehramtsstudierenden konkret angeleitet wurden Evidenz, welche ihnen vorgelegt wurde, zu nutzen. Aus diesem Grund können daraus noch keine Schlussfolgerungen hinsichtlich des realen Nutzens von Evidenz in der Praxis gezogen werden. Zudem waren auf Grund der Stichprobengröße und des hohen Trainingseffektes auf alle Studierenden keine ausreichenden Möglichkeiten gegeben, um bedeutsame Vergleiche zwischen Teilnehmenden verschiedener Leistungsstufen aufzeigen zu können. Hierfür könnte es möglicherweise hilfreich sein, zusätzliche Daten einer untrainierten Vergleichsgruppe zu erheben.

Trotz dieser Limitationen können mit Hilfe dieser Ergebnisse konkrete Hinweise zur Verbesserung und Weiterentwicklung zukünftiger Trainings zur Förderung des argumentativen Nutzens von Evidenz gewonnen werden. So sollte verstärkt darauf geachtet werden, die Studierenden dergestalt anzuleiten, dass Evidenz in Diskussionen nicht rein mechanisch in Bezug auf die eigenen Aussagen genutzt werden sollte, sondern dabei auch die Argumente der Mitdiskutanten mit einzubeziehen sind. Des Weiteren sollten Studierende bereits von Anfang an darauf sensibilisiert werden, die beste verfügbare Evidenz hinsichtlich ihrer Belastbarkeit auszuwählen und keinesfalls ausschließlich hinsichtlich ihrer Passung zur eigenen Meinung. Zudem bedarf es weiterer Lerngelegenheiten zum Üben des argumentativen Nutzens von Evidenz.

## 10. Gesamtdiskussion

Das kritische Hinterfragen und Bewerten wissenschaftlicher Evidenz, sowie dessen argumentatives Nutzen sind wichtige Kernkompetenzen für (zukünftige) Lehrerinnen und Lehrer. Dies wird sowohl von internationalen Bildungsforschern, als auch aus der Politik heraus gefordert (Bauer et al., 2012; EC, 2013; KMK, 2004/2014; Parr et al., 2008). Besonders hinsichtlich einer durch das Nutzen von Evidenz gegebenen Legitimationsfunktion bei weitreichenden Entscheidungen, sowohl auf Schüler- als auch auf Schulebene, sind diese Kompetenzen von hoher Relevanz (Fischer et al., 2014; Higgins & Green, 2011; Rubin, 2008). Aktuelle Studien zeigen dennoch, dass aktive Lehrkräfte kaum Evidenz in ihrem professionellen Umfeld nutzen (Williams & Coles, 2007). Ein möglicher Grund hierfür ist, dass trotz internationaler wie nationaler Forderungen aus Politik und Wissenschaft, die notwendigen Kompetenzen in der Lehrerausbildung noch nicht ausreichend gefördert werden. Im Allgemeinen zeigt sich zudem, dass die bildungswissenschaftlichen Anteile in der Lehrerbildung in Deutschland bislang sehr gering sind und deshalb kaum Möglichkeiten geboten werden, den kompetenten Umgang mit wissenschaftlicher Evidenz konkret zu fördern. (Bauer et al., 2012; Lohmann et al., 2011). Da die Curricula des Lehramtsstudiums jedoch bereits sehr umfangreich sind, vor allem auf Grund der vielfältigen fachlichen Inhalte in den jeweiligen Unterrichtsfächern, wäre es hilfreich, diese Kompetenzen im Rahmen eines flexibel in bestehende Lehramtscurricula integrierbaren Kurzzeittrainings fördern zu können. Neben der fehlenden Entwicklung der relevanten Kompetenzen, herrschen zudem sowohl unter aktiven Lehrkräfte als auch unter Lehramtsstudierenden oftmals negative Einstellungen und Vorurteile bezüglich des Nutzens von Evidenz in der beruflichen Praxis vor (Hetmanek et al., 2015; Hargreaves, 2000; Kühle et al., 2007; Parr et al., 2008; Schildkamp et al., 2010; Williams et al., 2008; Zembal-Saul et al., 2003). Wichtig für ein solches Kurzzeittraining ist deshalb nicht nur die Vermittlung der fachlichen Inhalte, sondern auch das Aufzeigen der Relevanz, sowie der praktischen Nutzbarkeit. Bislang existieren im internationalen Raum jedoch nur wenige Interventionsstudien bezüglich der Trainierbarkeit und Förderung dieser Kompetenzbereiche. Im Rahmen dieser Arbeit wurde deshalb ein zweitägiges Trainingsprogramm entwickelt, welches sowohl die Relevanz des Nutzens von Evidenz in der Praxis aufzeigt, als auch diejenigen Inhalte vermittelt, die mindestens notwendig sind um Evidenz kritisch zu hinterfragen und argumentativ nutzen zu können. Ziel war es dabei, zunächst einen theoretischen Rahmen zur Identifizierung der relevanten Kompetenzen, sowie zur bestmöglichen Förderung eben dieser zu schaffen.

Da das professionelle Nutzen von Evidenz in anderen Domänen, wie etwa der Medizin (z.B. Atkins et al., 2004; Guyatt et al., 2008; Sackett et al., 1997), oder der Science Education (z.B. Fischer et al., 2014; Shavelson & Tone, 2002; Weinberger et al., 2003), bereits etablierter ist als in der Lehrerbildung, wurden zunächst erfolgreiche Ansätze und Konzepte aus diesen Bereichen dargestellt und hinsichtlich ihrer Übertragbarkeit auf die Lehrerbildung diskutiert. Mit Hilfe dieser Ansätze, konnte ein Bedingungsmodell auf Basis von fünf Kompetenzfacetten zum professionellen Nutzen von Evidenz identifiziert werden. Diese umfassen die Formulierung einer Fragestellung, die Suche nach der besten verfügbaren Evidenz, die kritische Bewertung der Evidenz, das argumentative Nutzen der Evidenz und die Evaluation evidenzbasierter Entscheidungen (Ramos et al., 2003; Rubin, 2008; Shaneyfeldt et al., 2006). Auf Basis früherer Studien anderer Domänen ist anzunehmen, dass diese Kompetenzfacetten durch interne Faktoren wie (Vor-) Wissen und Orientierungen (z.B. Davies, 1999; Prenzel & Bauer, 2012; Ramos et al., 2003; Weber & Achtenhagen, 2008), sowie durch externe Faktoren wie Anwendungsbereiche und Merkmale von Evidenz (z.B. Bromme, 2008; Kienhues et al., 2001; Wilson & Sloane, 2000) beeinflusst werden können. Das hieraus entstandene Bedingungsmodell diente als theoretische Basis zur Entwicklung des Trainingsprogramms. Im Rahmen dessen wurden im Speziellen zwei dieser Kompetenzfacetten fokussiert: das kritische Bewerten von Evidenz und das argumentative Nutzen dessen. Gründe für diese Fokussierung sind sowohl der Aspekt der zeitlichen Begrenzung in welcher das Training stattfinden sollte, als auch Ergebnisse früherer Studien, welche besagen, dass Studierende vorrangig mit eben diesen Kompetenzfacetten die größten Schwierigkeiten zu haben scheinen (Fischer et al., 2014).

Um diese beiden Kompetenzen sowohl effektiv fördern, als auch vergleichbar messen zu können, wurden auf Basis von etablierten Kompetenzmodellen aus der Science Education zwei Stufenmodelle mit je fünf Kompetenzstufen zum kritischen Hinterfragen, sowie zum argumentativen Nutzen von Evidenz entwickelt (Bromme, 2014; Kuhn & Crowell, 2011; Wilson, 2005). Demzufolge zeichnet sich eine hohe Kompetenz im kritischen Bewerten von Evidenz dadurch aus, dass sowohl die relevanten Inhalte genannt und kritisch bewertet werden können, sowie darüber hinaus die Fragilität und Konflikthaftigkeit des Forschungsfeldes erkannt wird. Eine hohe Kompetenz bezüglich des argumentativen Nutzens von Evidenz besteht, wenn Argumente durchgängig mit Evidenz belegt sind und diese dabei zugleich kritisch hinterfragt und bewertet wird. Um diese Kompetenzstufen erreichen zu können, bedarf es eines entsprechenden Hintergrundwissens. Da es jedoch kaum

umsetzbar wäre, Lehramtsstudierende mit einem tiefgehenden (vorrangig statistischen und forschungsmethodischen) Fachwissen diesbezüglich auszustatten, wurden gezielte Heuristiken eingesetzt. Ziel dieser Heuristiken war es, den Studierenden das Wissen zu vermitteln, das mindestens notwendig ist um Evidenz kompetent und situationsangemessen im beruflichen Alltag nutzen zu können (Bromme, 2014). Der Einsatz solcher Heuristiken hat sich in anderen Domänen wie beispielsweise der Medizin (Finkel et al., 2003; Higgins & Green, 2011; Howick et al., 2011; Ilic et al., Malhotra et al., 2015) oder der Science Education (Fischer et al., 2014; Gigerenzer, 2008; Hutchinson & Gigerenzer, 2005; Kopp & Mandl, 2011; Kuhn, 2010; Osbourne, 2010) bereits bewährt. Auf Grund der Unterschiedlichkeit der Forschung innerhalb der Domänen, können diese jedoch nicht eins zu eins auf die Bildungswissenschaften übertragen werden. Deshalb wurden verschiedene Ansätze für Heuristiken aus anderen Domänen zunächst beschrieben und daraus diejenigen ausgewählt und adaptiert, die für den speziellen Kontext der Förderung des kritischen Hinterfragens und des argumentativen Nutzens von Evidenz bei Lehramtsstudierenden situationsangemessen und anwendbar scheinen (z.B. Abelson, 1995; Elsevier, 2018; Kuhn & Crowell, 2011; Toulmin, 1978).

Für eine bestmögliche Vermittlung dieser Heuristiken wurden darauf folgend erfolgreiche Methoden etablierter Trainings dargestellt und analog zu den Heuristiken hinsichtlich ihrer Übertragbarkeit auf die Lehrerbildung diskutiert. Hierbei deutete sich an, dass vor allem problem- und fallbasiertes Lernen (z.B. Coomarasamy & Kahn, 2004; Gräsel, 2006; Hmelo-Silver & Barrows, 2006), Formen kooperativen Lernens (z.B. Fischer et al., Iordanou et al., 2014; Kopp et al., 2011; Osbourne, 2010) und kognitive Modellierung durch tutorielle Unterstützung (z.B. Collins et al., 1989; Iordanou et al., 2014; Kopp et al., 2011; Zembal-Saul et al., 2003) am besten geeignet scheinen, um die Kompetenzen des kritischen Hinterfragens, sowie des argumentativen Nutzens im Rahmen eines Kurzzeitrainings effektiv zu fördern. Die Ergebnisse der Evaluation dieses Trainings zeigen, dass auf diese Weise beide Kompetenzen effektiv trainiert werden konnten. Dies deutet darauf hin, dass die Kombination aus den oben genannten instruktionalen Prinzipien mit der Verwendung von Heuristiken wirksame Lernvoraussetzungen zur Weiterentwicklung der relevanten Kompetenzen schaffen können.

## **10.1. Überblick und Diskussion der zentralen Ergebnisse**

Zur Evaluation und Entwicklung des Trainings wurden im Rahmen dieser Arbeit drei Studien durchgeführt. Ziel der Studie 1 war es zunächst, erste Ergebnisse

hinsichtlich der Trainierbarkeit der Kompetenzen, sowie der Machbarkeit des Trainings und der Akzeptanz durch die Teilnehmenden zu erlangen. Hierbei zeigten sich bezüglich der Bewertungs- wie auch der Argumentationskompetenz bereits deutliche Effekte. Die Teilnehmenden der Trainingsgruppen wiesen sowohl in den hochinferenten Globalratings, als auch in den niedriginferenten Items eine deutlich höhere Kompetenz hinsichtlich des kritischen Hinterfragens von Evidenzquellen, sowie deren argumentativer Nutzung auf, als die Teilnehmenden der Kontrollgruppe. Jedoch erreichte kein Teilnehmender die höchste Kompetenzstufe 5, in welcher zusätzlich zum kritischen bewerten und Nutzen der Evidenzquelle auch die Fragilität und Konflikthaftigkeit des Forschungsfeldes erkannt und benannt werden soll (Kuhn & Crowell, 2011; Wilson, 2005). Diese Ergebnisse zeigen, dass innerhalb einer zweitägigen Kurzzeitintervention zwar basale Kenntnisse und Techniken zum kompetenten Nutzen von Evidenz vermittelt werden konnten, ein reflektiertes und elaboriertes Verständnis der Natur wissenschaftlicher Studien jedoch erwartungsgemäß nicht erreicht werden konnte. Eine häufige Schwierigkeit ist hierbei die Überwindung der durch die konfligierenden und fragilen Eigenschaften von Wissenschaft entstehenden kognitiven Konflikte bei den Teilnehmenden (Bromme, 2008; Kuhn & Crowell, 2011). Es ist anzunehmen, dass hierfür weitere Lerngelegenheiten, die über ein zweitägiges Training hinausgehen, notwendig sind (Desimone, 2009).

Neben dem Erreichen der Kompetenzstufe auf Individualebene innerhalb jedes Settings wurde auch die Konsistenz dieser über die vier verschiedenen Testsituationen (je zwei schriftliche Statements und je zwei Gruppendiskussionen) gemessen. Die Ergebnisse zeigten, dass die Ergebnisse über alle Settings hinweg annähernd konsistent waren. Dies zeigt, dass die Teilnehmenden ihr Wissen sowohl in einer individuellen Situation schriftlich anwenden konnten, als auch den Transfer zu mündlichen Diskussionen unterschiedlicher Domänen (zuerst gesundheitswissenschaftlichen und anschließend bildungswissenschaftlichen Inhalts) erzielen konnten.

Die mit Hilfe der Fragebögen, sowie der Feedbackhefte erfassten Ergebnisse hinsichtlich der Motivation, der Akzeptanz und dem subjektiv wahrgenommenen Lernzuwachs waren bei den Teilnehmenden des Trainings durchgängig positiv. Dies weist darauf hin, dass unter anderem durch die Verwendung problemorientierter Fallbeispiele, sowie durch die Verwendung von Fallbeispielen, die Themen behandeln, welche für die Teilnehmenden nicht nur professions- sondern auch alltagsrelevant sind, die Wichtigkeit und Relevanz der Entwicklung der notwendigen Kompetenzen zum kritischen Hinterfragen und argumentativen Nutzen von Evidenz

vermittelt werden konnte. Jedoch wurden durch das Training keine epistemologischen Überzeugungen und Einstellungen verändert, diese blieben erwartungsgemäß über beide Messzeitpunkte hinweg (vor und nach dem Training) stabil (Stahl & Bromme, 2007).

Die Ergebnisse sind jedoch, vor allem auf Grund der geringen Größe der Stichprobe, sowie der hohen Anzahl an Teilnehmenden mit naturwissenschaftlichem Hintergrund nur als erste Hinweise die es zu replizieren gilt zu interpretieren. Dies war Ziel der Studie 2. Vor Beginn der Studie 2 wurden die Testinstrumente, sowie die einzelne Trainingsmaterialien mit Hilfe der Ergebnisse und den Erfahrungen die aus Studie 1 hervorgingen angepasst und weiterentwickelt. So wurde ein Fallbeispiel zum Thema Lehrergesundheit, welches von den Studierenden als weniger interessant (weil bereits in anderen Seminaren sehr oft behandelt) beurteilt wurde ausgetauscht. Auch das Kodierschema wurde angepasst und ein zweites Globalrating zur Bewertung der Argumentationskompetenz (analog zur Bewertungskompetenz) entwickelt, um eine bessere Vergleichbarkeit erreichen zu können. Zudem wurden einige Formulierungen der niedriginferenten Items angepasst, die bei den Ratern zu Missverständnissen geführt hatten, sowie weitere Items eingefügt um die Verwendung der innerhalb des Trainings vermittelten Heuristiken besser abbilden zu können.

Ziel der Studie 2 war die Replikation der Ergebnisse von Studie 1, sowie das Erfassen möglicher Einflussfaktoren auf die relevanten Kompetenzen zum Nutzen von Evidenz. Ziel war es zudem, den aktuellen Forschungsstand durch das generieren von Informationen bezüglich der Voraussetzungen, sowie der Möglichkeiten des effektiven Förderns von relevanten Kompetenzen zum (argumentativen) Nutzen von Evidenz in der Lehrerbildung zu erweitern (Iordanou et al., 2014; Parr et al., 2008; Zohar, 2008). Dementsprechend wurde ein Trainingsansatz, welcher diverse förderliche Methoden zum kompetenten Nutzen von Evidenz unterschiedlicher Domänen wie der Medizin, der Science Education und der Lehrerbildung enthält, experimentell getestet. Darüber hinaus wurden mögliche relevante Einflussfaktoren auf die relevanten Kompetenzen erhoben.

Die Ergebnisse zeigten analog zu Studie 1 eine deutlich höhere Kompetenz bei den Teilnehmenden der Trainingsgruppe als bei den Teilnehmenden der Kontrollgruppe. Das Training zeigte einen nachweislich signifikant hohen Effekt auf die Kompetenzen des kritischen Hinterfragens, sowie des argumentativen Nutzens. Ebenso konnte bei den trainierten Teilnehmenden zwar eine hohe Kompetenzstufe erzielt werden, die höchste Kompetenzstufe wurde jedoch nur selten erreicht. Dies

bestärkt die Vermutung, dass zur Erlangung dieser weitere Lerngelegenheiten nötig sind (Glug, 2009; Prenzel et al., 2007; Prenzel & Bauer, 2012; Weinert, 2011).

Die getesteten Einflussfaktoren Domänenwissen (durch Fächerhintergrund), wissenschaftstheoretisches Vorwissen, Lesehäufigkeit wissenschaftlicher Texte, epistemologische Überzeugungen, Überzeugungen zum Nutzen von Evidenz, Ambiguitätstoleranz und wahrgenommene Barrieren hinsichtlich der Verwendung von Evidenz in der Praxis zeigten, entgegen vorheriger Vermutungen, keinerlei Einfluss auf die erzielte Kompetenzstufe nach dem Training. Die Hypothese war, dass Studierende mit naturwissenschaftlichem Hintergrund (mindestens einem MINT Fach) auf Grund der größeren Vertrautheit mit empirischen, experimentellen Studien einen Vorteil gegenüber Studierenden mit rein sprach- oder geisteswissenschaftlichem Hintergrund haben. Diese Annahme resultiert aus Forschungsansätzen zum Zusammenhang von Domänenwissen und wissenschaftlichem Argumentieren (Chinnappan et al., 2011; Fischer et al., 2014; Schunn & Anderson, 1999). Dies konnte im Rahmen dieser Studie jedoch nicht bestätigt werden. Ein Grund hierfür könnte sein, dass Studierende mit geisteswissenschaftlichem Hintergrund eine höhere Kompetenz im Umgang mit komplexen Texten aufweisen, dies gilt es jedoch gesondert zu untersuchen.

Diese Ergebnisse lassen jedoch vermuten, dass die Kompetenzen zum professionellen Nutzen von Evidenz mit Hilfe dieses Trainings bei Lehramtsstudierenden jeglichen Fachhintergrunds effektiv gefördert werden können. Im Gegensatz zu Studie 1 waren die Ergebnisse der Globalratings in Studie 2 über die verschiedenen Settings hinweg nicht konsistent. Zwar waren die Ergebnisse auf Individualebene über beide Kompetenzen (*appraise* und *apply*), sowie über beide Statements hinweg konsistent, jedoch nicht über die Ergebnisse der Statements und der Gruppendiskussionen. Der Transfer von der individuellen, schriftlichen Situation hin zu einer mündlichen Diskussion auf Gruppenebene war demnach nicht notwendigerweise gegeben. Ebenso zeigte sich keine Konsistenz in den Ergebnissen der Globalratings zwischen dem ersten (gesundheitswissenschaftlichen) und dem zweiten (bildungswissenschaftlichen) Fallbeispiel. Ein Grund hierfür könnte sein, dass die Studierenden bei der gesundheitswissenschaftlichen Diskussion über den Zusammenhang zwischen erhöhter Salzzufuhr und Hypertonie weniger Erfahrung und Hintergrundwissen aufwiesen, als bei dem zweiten bildungswissenschaftlichen Diskussion und sich deshalb mehr auf die ihnen vorliegenden Evidenzquellen bezogen.

Grundsätzlich ist vor allem die hohe Effektstärke des Trainings sehr aussichtsreich, jedoch muss bei der Interpretation dieser Ergebnisse beachtet werden, dass die

Testinstrumente direkt im Anschluss an ein zweitägiges Training durchgeführt wurden. Folglich ist es nicht möglich, dies hinsichtlich des tatsächlichen Nutzens von Evidenz in der späteren Berufspraxis zu generalisieren. Um hierüber Aussagen treffen zu können werden entsprechende Langzeitstudien mit mehreren Messzeitpunkten benötigt. Des Weiteren kann ein „mechanisches“ Nutzen von Evidenz sowohl schriftlich als auch in Diskussionen kein Ziel für zukünftige Lehrerinnen und Lehrer sein. Vielmehr soll den Studierenden die Fähigkeit zum kritischen Hinterfragen sowie zum situationsangemessenen Nutzen von Evidenz vermittelt werden. Hierfür gilt jedoch zu klären, wann das Nutzen von Evidenz situationsangemessen ist, sowie wie sich erfolgreiches Nutzen von Evidenz in der Praxis tatsächlich zeigen kann. Um erste Hinweise dahingehend zu generieren wie Lehramtsstudierende Evidenz tatsächlich nutzen, wurden in Studie 3 die videografierten Gruppendiskussionen eines Subsamples aus Studie 2 nochmals detaillierter analysiert.

In Studie 3 wurden zunächst plakative Beispiele aus den Transkripten der Gruppendiskussionen dargestellt, um zu illustrieren, wie sich erfolgreiches Nutzen von Evidenz im Rahmen einer Gruppendiskussion zeigen kann. Diese Beispiele können sowohl zu Demonstrationszwecken in zukünftigen Trainings, als auch als Ankerbeispiele für zukünftige Auswertungen dienen. Um zukünftige Trainings weiterzuentwickeln und noch besser an die Bedürfnisse der Lehramtsstudierenden anzupassen, wurden zudem häufige Probleme und Fehler der Teilnehmenden analysiert und zusätzlich durch Beispiele dargestellt. Die Ergebnisse liefern zudem Hinweise hinsichtlich eines tiefergehenden Verständnisses der relevanten Prozesse, die beim Lernen des Nutzens von Evidenz ablaufen. Die häufigsten Probleme lagen darin, eine Behauptung auf angemessene Art und Weise mit einer Evidenzquelle zu verknüpfen. Des Weiteren wurde Evidenz oftmals ausschließlich dazu verwendet um die eigene Meinung zu untermauern und im Zuge dessen mögliche Gegenpositionen außer Acht gelassen oder gar missinterpretiert, so dass es keine Notwendigkeit gibt die eigene Behauptung zu hinterfragen, beziehungsweise zu ändern. Speziell dies müsste in zukünftigen Trainings verstärkt aufgegriffen werden, um die Studierenden für einen solchen Missbrauch von Evidenz zu sensibilisieren.

Auch eine Analyse dessen, aus welchen Gründen Studierende Evidenz nutzen und welche Funktionen die Argumente im Falle eines erfolgreichen Nutzens von Evidenz einnehmen, liefert zusätzliche relevante Informationen. Diese können dazu dienen um zu verstehen, warum und wie genau Lehramtsstudierende Evidenz argumentativ

nutzen. Dies wiederum sind hilfreiche Informationen für die Entwicklung zukünftiger Trainings und Testinstrumente.

Darüber hinaus wurde im Rahmen dieser Studie untersucht, ob ein Zusammenhang zwischen erfolgreichem Nutzen von Evidenz und Transaktivität besteht. Die Ergebnisse zeigten erwartungsgemäß eine signifikante negative Korrelation was bedeutet, dass Studierende beim erfolgreichen Nutzen von Evidenz oftmals nicht auf die Argumente ihrer Mitdiskutanten eingegangen sind. Ein Grund hierfür ist möglicherweise, dass das Verfassen eines evidenzbasierten Arguments auf Basis der neu erlernten Heuristiken eine hohe Konzentration der Studierenden voraussetzt, so dass ein Eingehen auf die Mitdiskutierenden schwer möglich wird. Auch dies könnte möglicherweise durch weitere Lerngelegenheiten außerhalb des Trainings verbessert werden.

Zu beachten ist, dass die Ergebnisse als Hinweise und Beispiele dahingehend zu interpretieren sind, wie und warum Lehramtsstudierende Evidenz argumentativ in Diskussionssituationen nutzen. Im Rahmen dieser Studie wurden die Teilnehmenden detailliert mit Hilfe von Kooperationskripts dazu angeleitet die instruierten Evidenzquellen kritisch zu bewerten und zu nutzen. Diese angeleitete Testsituation kann folglich nicht eins zu eins auf eine reale Situation übertragen werden.

## **10.2. Einschränkungen und Anknüpfungspunkte**

Wie bereits in der Diskussion der zentralen Ergebnisse der einzelnen Studien beschrieben, gilt es die Limitationen der vorliegenden Studien bei der Frage nach der Generalisierbarkeit ihrer Ergebnisse zu beachten. Generell ist zu beachten, dass das im Rahmen dieser Arbeit entwickelte Training nicht konkret zu evidenzbasiertem Handeln im Lehrerberuf führen kann. Es kann ein erster Schritt in diese Richtung sein, in dem Studierende das nötige Basiswissen, welches mindestens notwendig ist um Evidenz kompetent nutzen zu können, kennen lernen. Um eine evidenzbasierte, beziehungsweise eine evidenzorientierte Praxis zu fördern, bedarf es jedoch weiterer Lerngelegenheiten, sowie positiver Einstellungen und Überzeugungen hinsichtlich des Nutzens von Evidenz. Die Einstellungen und Überzeugungen zur Belastbarkeit, sowie zum Nutzen von Evidenz in der Praxis zeigten sich vor allem die Bildungswissenschaften betreffend überwiegend negativ. Diese Einstellungen und Überzeugungen blieben auch durch das zweitägige Training erwartungsgemäß unverändert. Dennoch hatten sie keinerlei Einfluss auf den Lernzuwachs innerhalb des Trainings. Dies lässt vermuten, dass die Überzeugungen nicht notwendigerweise relevant für den Lernerfolg sind, jedoch

sind sie für das tatsächliche Nutzen von Evidenz in der späteren Berufspraxis essentiell.

Des Weiteren gilt zu beachten, dass in dem der vorliegenden Arbeit zu Grunde liegenden Training der Fokus auf das Fördern derer Kompetenzen lag, die Studien zufolge Praktikern die größten Schwierigkeiten bereiten. Somit wurden nicht alle Kompetenzfacetten, die für ein professionelles Nutzen von Evidenz voraussetzend sind, trainiert. Es sind folglich zusätzliche Erhebungen und Trainings notwendig um zu prüfen, inwieweit die Kompetenzfacetten des Formulierens der richtigen Fragestellung (*ask*), der Suche nach der besten verfügbaren Evidenz (*acquire*), sowie der Evaluation evidenzbasierter Entscheidungen (*evaluate*) (siehe Abbildung 3) bereits erfolgreich umgesetzt werden können. Zudem sind weitere Posttests zur Überprüfung der Nachhaltigkeit der innerhalb des Trainings erlernten Inhalte notwendig, um weitere Aussagen hinsichtlich des tatsächlichen Nutzens von Evidenz treffen zu können.

Überdies ist zu beachten, dass die Experimentalgruppe ausschließlich mit einer zu diesem Zeitpunkt noch untrainierten Kontrollgruppe verglichen wurde. Um weitere Aussagen hinsichtlich der Wirksamkeit der Heuristiken, sowie der instruktionalen Prinzipien im Speziellen treffen zu können, müssten deshalb weitere Vergleiche mit alternativen Methoden und Inhalten erhoben werden. Darüber hinaus könnte auch eine Replikationsstudie mit einer deutlich größeren Stichprobengröße und einer ausgeglichenen Verteilung über verschiedene Universitäten hinweg, unter Einbeziehung der dort vorherrschenden Anteile bildungswissenschaftlicher Inhalte, weitere Erkenntnisse zur Trainierbarkeit und Nachhaltigkeit des Trainings liefern. Besonders für Informationen hinsichtlich der Nachhaltigkeit des Trainings, wären Langzeitstudien mit entsprechenden Posttests notwendig. Darüber hinaus würde es zur Verbesserung der Ergebnisse beitragen, wenn alle Teilnehmenden die gleichen extrinsisch-motivationalen Voraussetzungen einer Teilnahme, wie etwa in Form von ECTS Credits erhalten würden. Im Rahmen dieser Studie nahm ein Großteil der Studierenden freiwillig an dem zweitägigen Training teil. Zwar erhielten diese als Aufwandsentschädigung einen 40€ Gutschein eines Online-Versandhandels, jedoch ist dennoch anzunehmen, dass diese Teilnehmenden überdurchschnittlich motiviert und interessiert an der trainierten Thematik waren.

Zudem könnten zukünftige Trainings davon profitieren, einen noch stärkeren Fokus auf den speziellen Umgang mit den fragilen und konfligierenden Eigenschaften von Evidenz zu setzen. Frühere Studien zeigten, dass dies oftmals eine der größten Herausforderungen für Studierende darstellt (Bromme, 2008; Fischer et al., 2014; Hetmanek et al., 2015; Kuhn & Crowell, 2011). Darauf weisen auch die Ergebnisse

der Evaluation des dieser Arbeit zu Grunde liegenden Trainings hin, da die höchste Kompetenzstufe zum kritischen Hinterfragen von Evidenz, welche ein kompetentes Auseinandersetzen mit eben diesen Eigenschaften voraussetzt, nur extrem selten erreicht wurde.

Trotz dieser Einschränkungen zeigen die großen Effekte des Trainings, dass es möglich ist, innerhalb eines flexibel in bestehende Lehramtscurricula integrierbaren Kurzzeittrainings, eine solide Basis für einen kompetenten Umgang mit Evidenz bei Lehramtsstudierenden zu schaffen und die relevanten Inhalte mit Hilfe von Heuristiken zu vermitteln. Dies wiederum kann ein erster Schritt in Richtung eines professionellen Nutzens von Evidenz im späteren Berufsalltag der Lehramtsstudierenden sein.

Für die Lehrerbildung bedeuten diese Ergebnisse, dass (zukünftige) Lehrerinnen und Lehrer nicht notwendigerweise ein fundiertes statistisches und forschungsmethodisches Expertenwissen benötigen, um Evidenz erfolgreich und kompetent nutzen zu können. Ein derartiges Kurzzeittraining kann Lehramtsstudierende dazu befähigen Lösungen erarbeiten, die zwar möglicherweise nicht perfekt, jedoch der individuellen Situation angemessen sind. Es bedarf jedoch weiterer Lerngelegenheiten, um ein routiniertes Nutzen von Evidenz in der Praxis anzuregen. Hierfür könnten beispielsweise spezielle Diskussionsrunden, ähnlich wie die Grand Rounds oder Journal Clubs in der Medizin, bei dem sich Fachpersonal trifft um gemeinsam auf Basis der aktuellen Forschung Lösungsansätze für aktuelle Problemstellungen zu finden und die relevante Literatur für weitere Kolleginnen und Kollegen entsprechend praxisnah aufbereiten, eine hilfreiche und praktikable Möglichkeit sein. Auch für Lehrerinnen und Lehrer wäre es denkbar, spezielle Fachgruppen einzuführen, in welchen Ergebnisse aktueller Forschung sowohl im Allgemeinen, als auch hinsichtlich spezifischer Problemstellungen, diskutiert werden.

Ziel zukünftiger Forschung sollte deshalb auch nicht nur die zuvor diskutierte Weiterentwicklung entsprechender Trainings zur Förderung der Kompetenzen sein. Es sollte zudem angestrebt werden (zukünftigen) Lehrkräften ausreichend empirische Untersuchungen zu relevanten Problemstellungen der Praxis zur Verfügung zu stellen. Zu vielen Problemstellungen aus dem Berufsalltag aktiver Lehrerinnen und Lehrer ist der aktuelle Forschungsstand nach wie vor unzureichend (Hargreaves, 2007a, 2007b). Ein ausreichender Forschungsstand ist zudem eine essentielle Grundlage für Institutionen wie das Clearing House Unterricht oder Metaanalysen wie diejenigen von John Hattie (2008), welche zum Ziel haben

Forschungsergebnisse praxisrelevant aufzubereiten. Des Weiteren sollte ein Fokus darauf gelegt werden, Lehrkräften und Schulen im Allgemeinen einen besseren Zugang zu aktueller Forschung zu gewähren, da eine häufig genannte Barriere zum professionellen Nutzen von Evidenz der fehlende oder sehr beschwerliche Zugang zu entsprechenden Fachinformationsquellen ist (Reiss et al., 2015).

## Literaturverzeichnis

- Abelson, R. P. (1995). *Statistics as principled argument*. Mahwah, NJ: LEA.
- Alonso-Coello, P., Schünemann, H. J., Moberg, J., Brignardello-Petersen, R., Alk, E. A. et al. (2016). GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ*, 353.
- Altman, D. G. & Bland, J. M. (1995). Absence of evidence is not evidence of absence. *BMJ*, 311, 485.
- Antes G, Bassler D, Forster J. (Hrsg.) (2003). *Evidenz-basierte Medizin (EbM). Praxis-Handbuch für Verständnis und Anwendung der EbM*. Stuttgart: Thieme.
- Asterhan, C. & Schwarz, B. (2009). Argumentation and Explanation in Conceptual Change: Indications From Protocol Analyses of Peer-to-Peer Dialog. *Cognitive Science*. 33. 374-400.
- Atkins D., Best D., Briss P. A., Eccles M., Falck-Ytter Y., Flottorp S., et al. GRADE Working Group (2004). Grading quality of evidence and strength of recommendations. *BMJ*, 328 (7454), 1490.
- Barrows, H. S. (1996). Problem-based Learning in medicine and beyond: A brief overview. In L. Wilkerson & W. H. Gijsselaers (Hrsg), *Bring problem-based learning to higher education: Theory and practice*, 68, 3-12. San Francisco: Jossey-Bass.
- Bassand, J. P., Rydén, L., Simoons, M. & European Society of Cardiology (2002). The need for resources for clinical research: the European Society of Cardiology calls for European, international collaboration. *Clinical Research in Europe*, 360 (9348).
- Bauer, J., Möller, J. & Prenzel, M. (2010). PaLea: Panel zum Lehramtsstudium. Professionsbezogene Entwicklung von Lehramtsstudierenden in den neuen Studienstrukturen. In B. Schwarz, P. Nenniger & R. S. Jäger (Hrsg.), *Erziehungswissenschaftliche Forschung – nachhaltige Bildung. Beiträge zur 5.*

- DGfE-Sektionstagung "Empirische Bildungsforschung"/AEPF-KBBB im Frühjahr 2009 (S. 178-183). Landau: VEP.
- Bauer, J., Diercks, U., Rösler, L., Möller, J. & Prenzel, M. (2012). Lehramtsausbildung in Deutschland. *Unterrichtswissenschaft*, 40, 101-120.
- Bauer, J. & Prenzel, M. (2012). European teacher training reforms. *Science*, 336, 1642-1643.
- Baumert, J. & Kunter, M. (2006). Stichwort Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9, 469-520.
- Bell, P. (2010). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education*, 22, 797-817.
- Bellamy, J. L., Bledsoe, S. E. & Mullen, E. J. (2009). The Cycle of Evidence-Based Practice. In H. U. Otto, Polutta, A. & H. Ziegler (Hrsg.) (2009). *Evidence-Based Practice – Modernising the Knowledge Base of Social Work?* Leverkusen: Verlag Barbara Budrich.
- Belland, B. R., Glazewski, K. D., Richardson, J. C. (2011). Problem-based learning and argumentation: testing a scaffolding framework to support middle school students' creation of evidence-based arguments. *Instructional Science*, 39, 667-694.
- Biesta, G. J. J. (2010). Why 'what works' still won't work: From evidenced-based education to value-based education. *Studies in Philosophy and Education*, 29, 491-503.
- Billig, M. (1987). *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge, UK: Cambridge University Press.
- BMBF (Hrsg.). (2008). *Wissen für Handeln. Forschungsstrategien für eine evidenzbasierte Bildungspolitik*. Bonn: BMBF.
- Boomsma, A. (2000). Reporting Analysis of Covariance Studies. *Structural Equation Modeling*, 7, 461-483.

- Booth, W. C., Colomb, G. G. & Williams, J. M. (2008). *The craft of research*. Chicago: University of Chicago Press.
- Borgetto, B., Spitzer, L., Pfingsten, A. (2016). Die Forschungspyramide. Evidenz für die logopädische Praxis brauchbar machen. *Forum Logopädie*, 1, 24-28.
- Bricker, L. A., & Bell, P. (2008). Conceptualizations of argumentation from science studies and the Learning Sciences and their implications for the practices of science education. *Science Education*, 92 (3), 473–498
- Bromme, R. (2008). *Wissenschaft und Öffentlichkeit: Das Verständnis fragiler und konfligierender wissenschaftlicher Evidenz* [Antrag an die DFG]. Retrieved from: <http://wissenschaftundoeffentlichkeit.de/DFG-SPP1409>
- Bromme, R., Prenzel, M. & Jäger, M. (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. *Zeitschrift für Erziehungswissenschaft (Sonderheft 27)*, 3-54.
- Bromme, R., Prenzel, M. & Jäger, M. (2016). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. *Zeitschrift für Erziehungswissenschaft*, 19 (1), 129-146.
- Bühner, M., Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson.
- Burls, A. (2009). What is critical appraisal? *Evidence-based medicine: What is ...? series*. NPR09/1113, 1-8.
- Cetin, P., Erduran, S. & Kaya, E. (2010). Understanding the Nature of Chemistry and Argumentation: the Case of Pre-service Chemistry Teachers. *Journal of Kirsehir Education Faculty*. 11, 41-59.
- Chinn, C. A., Anderson, R. C., Waggoner M. A. (2001). Patterns of discourse in two kinds of literature discussion. *Reading Research Quarterly*, 36 (4), 378-441.
- Chinnapan, M., Ekanayake, M. B. & Brown, C. (2011). Specific and general knowledge in geometric proof development. *SAARC Journal of Educational Research*, 8, 1-28.

- Chisholm, R. M. (1980). Evidenz. In J. Speck (Hrsg.), *Handbuch wissenschaftstheoretischer Begriffe* (Bd. 1, S. 196-197). Göttingen: Vandenhoeck.
- Clearing House Unterricht (2015). *Qualitätsoffensive – Claringhouse Unterricht*. Entnommen am 24.03.2018 von <https://www.clearinghouse.edu.tum.de/>
- Cochrane Collaboration (2017). GRADE approach to evaluating the quality of evidence: a pathway. Retrieved from: <http://training.cochrane.org/path/grade-approach-evaluating-quality-evidence-pathway>
- Cochran-Smith, M. & Lytle, S.. (1999). The Teacher Research Movement: A Decade Later. *Educational Researcher*. 28. 15-25.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Hrsg.), *Knowing, learning and instruction* (S. 453–494). Hillsdale, NJ: LEA.
- Cook, B. G., Tankersley, M. & Harjusola-Webb, S. (2008). Evidence-based special education and professional wisdom: Putting it all together. *Intervention in School and Clinic*, 44 (2), 105-111.
- Coomarasamy, A. & Kahn (2004). What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. *BMJ*, 329(7473), 1-5.
- Cordray, D. S., Morphy, P. (2009). Research synthesis and public policy. In: H. M. Cooper, L. V. Hedges & J. C. Valentine (Hrsg.), *The handbook of research synthesis and meta-analysis*. (S. 473-493). New York: Sage.
- Csanadi, A., Kollar, I., & Fischer, F. (2015). *Teacher students' scientific reasoning: how can social context and internal scripts as predictors inform instructional design?*. Paper presented at the 19th Annual Conference for Junior Researchers of the European Association for Research on Learning and Instruction (JURE), Limassol, Cyprus.

- Csanadi, A., Kollar, I., Fischer, F. (2016). Scientific reasoning and problem solving in a practical domain: Are two heads better than one? In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: Presentation at the International Conference of the Learning Sciences (ICLS)*, 1, 50-57. Singapore: International Society of the Learning Sciences.
- (14) Collaborative and Individual Scientific Reasoning of Pre-Service Teachers: New Insights through Epistemic Network Analysis (ENA). Available from: [https://www.researchgate.net/publication/316090489\\_Collaborative\\_and\\_Individual\\_Scientific\\_Reasoning\\_of\\_Pre-Service\\_Teachers\\_New\\_Insights\\_through\\_Epistemic\\_Network\\_Analysis\\_ENA](https://www.researchgate.net/publication/316090489_Collaborative_and_Individual_Scientific_Reasoning_of_Pre-Service_Teachers_New_Insights_through_Epistemic_Network_Analysis_ENA) [accessed Mar 31 2018].
- Csanadi, A., Eagan, B., Williamson Shaffer, D., Kollar, I. & Fischer, F. (2017). Collaborative and Individual Scientific Reasoning of Pre-Service Teachers. Conference Paper at *New Insights through Epistemic Network Analysis (ENA)*.
- Curd, M. & Cover, J. A. (Hrsg.). (1998). *Philosophy of science*. New York, NY: Norton.
- Dalbert, C. & Radant, M. (2010). Ungewissheitstoleranz bei Lehrkräften. *Journal für LehrerInnenbildung*, 10, 53-57.
- Dansereau, D. F., Collins, K. W., McDonald, B. A., Holly, C. D., Garland, J., Diekhoff, G. & Evans, H. (1979). Development and evaluation of a learning strategy training program. *Journal of Educational Psychology*, 71, 1, 64-73.
- Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, 47, 108-121.
- Dedering, K. (2009). Evidence-based education policy: Lip service or common practice? *European Educational Research Journal*, 8, 484-496.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development. *Educational Researcher*, 38(3), 181–199.

- Deutsches Netzwerk für evidenzbasierte Medizin DNEbM (2009). Leitlinie evidenzbasierte Gesundheitsinformation. Evidenzbasierte Leitlinie. Version 1.0. Entnommen am 09.03.2018 von: [http://www.ebm-netzwerk.de/was-wir-tun/publikationen/Leitlinie Evidenzbasierte Gesundheitsinformation.pdf](http://www.ebm-netzwerk.de/was-wir-tun/publikationen/Leitlinie-Evidenzbasierte-Gesundheitsinformation.pdf)
- Donner-Banzhoff, N. (2007). Die Evidenzbasierte Medizin: Gescheitert oder einfach nur anders? *Zeitschrift für ärztliche Fortbildung und Qualität im Gesundheitswesen*, 101, 441-444.
- Döring, N., Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. 5. Auflage. Berlin: Springer.
- Driver, R., Newton, P., Osbourne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84 (3).
- Eva, K. W. (2004). What every teacher needs to know about clinical reasoning. *Medical Education*, 39 (1), 98-106.
- Evidence-based Nursing – Netzwerk Österreich (EBN) (2007). *Jadad-Score*. Entnommen am 31.03.2018 von [https://web.archive.org/web/20111120215443/http://www.ebn.at/cms/dokumente/10089199\\_1486478/68ad7abd/Jadad-Score\\_\\_.pdf](https://web.archive.org/web/20111120215443/http://www.ebn.at/cms/dokumente/10089199_1486478/68ad7abd/Jadad-Score__.pdf)
- Elsevier (2018). Guidelines for Reviewers. Entnommen am 31.03.2018 von <https://www.journals.elsevier.com/healthcare-the-journal-of-delivery-science-and-innovation/policies/guidelines-for-reviewers>.
- Engelmann, K., Neuhaus, B. & Fischer, F. (2016). Fostering scientific reasoning in education – meta-analytic evidence from intervention studies. *Educational research and evaluation*. 22(5-6), 333-349.
- Erduran, S. & Garcia-Mila, M. (2015). Epistemic practices and thinking in science: Fostering teachers' development in scientific argumentation.

- Ertl, B., & Mandl, H. (2004). *Kooperationsskripts als Lernstrategie*. (Forschungsbericht Nr. 172). München: Ludwig-Maximilians-Universität, Lehrstuhl für empirische Pädagogik und Pädagogische Psychologie.
- European Commission [EC] (2007). *Towards more knowledge-based policy and practice in education and training*. Retrieved from <http://www.eipee.eu/Default.aspx?tabid=2533&language=en-US>
- European Commission [EC] (2013). *Supporting teacher competence development for better learning outcomes*. Retrieved from [ec.europa.eu/education/policy/school/doc/teachercomp\\_en.pdf](http://ec.europa.eu/education/policy/school/doc/teachercomp_en.pdf)
- Fadness Tadlock, D. (1978). SQ3R: Why it Works, Based on an Information Processing Theory of Learning. *Journal of Reading*, 22 (29), 110-112.
- Felton, M , & Kuhn, D. (2001). The development of argumentive discourse skills. *Discourse Processes*, 32, 135-153.
- Felton, M. K. (2004). The development of discourse strategies in adolescent argumentation. *Cognitive Development*, 19, 35-52.
- Finkel, M.; Brown, H.; Gerber, L.; Supino, P. (2003): Teaching evidence-based medicine to medical students. In: *Medical teacher* 25 (2), S. 202–204. DOI: 10.1080/0142159031000092634.
- Fischer, F. & Wecker, C. (2006). Pasteurs Quadrant und die Diskussion in den USA um die Verbesserung des praktischen Nutzens der Bildungsforschung. In A. Brüggemann & R. Bromme (Hrsg.), *Entwicklung und Bewertung von anwendungsorientierter Grundlagenforschung in der Psychologie (Reihe Rundgespräche und Kolloquien der DFG)* (S. 27-37). Berlin: Akademie Verlag.
- Fischer, F., Kollar, I., Mandl, H. & Haake, J. M. (2007). *Scripting computer-supported collaborative learning – Cognitive, computational and educational perspectives*. Berlin: Springer.
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., ... & Eberle, J. (2014). *Scientific Reasoning and Argumentation: Advancing an*

- Interdisciplinary Research Agenda in Education. *Frontline Learning Research*, 3, 1-18.
- Fritsche, L.; Greenhalgh, T.; Falck-Ytter, Y.; Neumayer, H.-H.; Kunz, 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. In: *British Medical Journal* 325, S. 1338–1341.
- Gerhardt-Szep, S., Dreher, S., Rüttermann, S. & Weberschock, T. (2017). Konzeption und Implementierung eines neuartigen E-Learning-Moduls mit EbM-Lerninhalten im Fach Zahnerhaltungskunde. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*. 127-128.
- Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, 3, 20-29.
- Glug, I. (2009). *Entwicklung und Validierung eines Multiple-Choice-Tests zur Erfassung prozessbezogener naturwissenschaftlicher Grundbildung*. Dissertation: Universität Kiel.
- Gough, D. (2004). Systematic research synthesis. In G. Thomas & R. Pring (Hrsg.), *Conducting educational research. Evidence-based practice in education*, 44-62. Maidenhead: Open University Press.
- Graff, G. (2003). *Clueless in academe: How schooling obscures the life of the mind*. New Haven: Yale University Press.
- Gräsel, C. & Mandl, H. (1993). Förderung des Erwerbs diagnostischer Strategien in fallbasierten Lernumgebungen. *Unterrichtswissenschaft* 21, 4, 335-369.
- Gräsel, C. (1997). Problemorientiertes Lernen. Strategieranwendung und Gestaltungsmöglichkeiten. In N. Havers, H. Tschamler & W. Tunner (Hrsg). *Münchener Universitätsschriften Psychologie und Pädagogik*. Göttingen: Hogrefe.

- Gräsel, C. (2006). Gestaltung problemorientierter Lernumgebungen. In K. H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (S. 335-339). Bad Heilbrunn: Klinkhardt.
- Green, N., & Green, K. (2005). *Kooperatives Lernen im Klassenraum und im Kollegium: Das Trainingsbuch*. Seelze-Verlber: Kallmeyer.
- Greenhalgh, T. (2003). *Einführung in die Evidence-based Medicine: kritische Beurteilung klinischer Studien als Basis einer rationalen Medizin*. Göttingen: Huber.
- Greenhalgh, T. (2010). *How to read a paper: the basics of evidence based medicine*. London: Blackwell BMJ Books.
- Gruber, H. & Stamouli, E. (2009). Intelligenz und Vorwissen. In J. Möller & E. Wild (Hrsg.), *Pädagogische Psychologie* (S.27-48). Berlin: Springer.
- Guyatt, G., Cairns, J. Churchill, D., Cook, D., Haynes, B. et al. (1992). Evidence-Based Medicine. A New Approach to Teaching the Practice of Medicine. *JAMA*, 268 (17), 2420-2425.
- Guyatt, G.H., Oxman A.D., Vist G.E., Kunz R., Falck-Ytter Y., Alonso-Coello P., Schünemann H.J. et al. GRADE Working Group (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, 336, (7650), 924-6.
- Haas, A. (2005). Unterrichtsplanung im Alltag von Lehrerinnen und Lehrern. In: A. A. Huber (Hrsg.), *Vom Wissen zum Handeln. Ansätze zur Überwindung der Theorie-Praxis-Kluft in der Schule und Erwachsenenbildung* (S. 5-19). Tübingen:\_ Ingeborg Huber Verlag.
- Hamer, S., Collison, G. (2005). *Achieving Evidence-based Practice. A Handbook for Practitioners*. London: Elsevier
- Hargreaves, D. (2000). Teaching as a research-based profession. In B. Moon (Hrsg.), *Leading professional development in education* (S. 200-210). London: Routledge.

- Hargreaves, D. (2007a). Teaching as a research-based profession: possibilities and prospects. In: M. Hammersley (Hrsg.), *Educational Research and Evidence-Based Practice* (3-17). London: Sage.
- Hargreaves, D. (2007b). In defence of research for evidence-based teaching: a rejoinder to Martyn Hammersley. In: M. Hammersley (Hrsg.), *Educational Research and Evidence-Based Practice* (43-60). London: Sage.
- Hartmann, U., Decristan, J. & Klieme, E. (2016). Unterricht als Feld evidenzbasierter Bildungspraxis? Herausforderungen und Potentiale für einen wechselseitigen Austausch von Wissenschaft und Schulpraxis. *Zeitschrift für Erziehungswissenschaften*, Sonderheft 31, 179-199.
- Hasanbegovic, J., Bauer, J., Gruber, H. & Rehl, M. (2006). The two-fold role of epistemological beliefs in higher education. In P. Tynjälä, J. Välimaa & G. Boulton-Lewis (Hrsg.), *Higher education and working life* (S. 163-176). Oxford: Pergamon.
- Hastings, C. & Fisher, C. A. (2014). Searching for proof: Creating and using an actionable PICO question. *Nursing Management*, 45 (8), 9-12.
- Hattie, J. (2008). *Visible learning. A synthesis of over 800 Meta-analyse relating to achievement*. Abingdon: Routledge.
- Hefter, M. H., Berthold, K., Renkl, A., Riess, W., Schmid, S., & Fries, S. (2014). Effects of a training intervention to foster argumentation skills while processing conflicting scientific positions. *Instructional Science*, 42(6), 929-947.
- Heininger, S., Bauer, J., Wenglein, S. & Prenzel, M. (2016, März). *Lehrkräfte bewerten Evidenz – Entwicklung einer heuristisch orientierten Skala zur Bewertung bildungswissenschaftlicher Evidenz*. Vortrag auf der 4. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF), Berlin.
- Hemberger, L., Kuhn, D., Matos, F. & Shi, Y. (2017). A Dialogic Path to Evidence-Based Argumentive Writing. *Journal of the Learning Sciences*, 26 (4), 575-607.

- Hennessey, M., Murphy, P. K. & Kulikowich, J. M. (2013). Investigating teachers' beliefs about the utility of epistemic practices: A pilot study of a new assessment. *Instructional Science*. 41 (10).
- Hetmanek, A., Wecker, C., Gräsel, C., Kieseewetter, J., Trempler, K. Fischer, M.R., & Fischer, F. (2015). Ressourcen nutzen Lehrkräfte wann? Eine Interviewstudie zur Schnittstelle zwischen Unterrichtsalltag und Wissenschaft, *Unterrichtswissenschaft*, 43 (3).
- Higgins, J. P. T. & Green, S. (Hrsg.). (2011). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 5.1.0). Retrieved from: [www.cochrane-handbook.org](http://www.cochrane-handbook.org)
- Hmelo-Silver, C. E., Barrows, H. S. (2006). Goals and Strategies of e Problem-based Learning Facilitator. *Interdisciplinary Journal of Problem-based Learning*, 1 (1), 5-22.
- Hofer, B.K. & Pintrich, P.R. (2002). *Personal epistemology. The psychology of beliefs about knowledge and knowing*. Mahwah, NJ: Erlbaum.
- Horrobin, D. F. (2000). Innovation in the pharmaceutical industry. *J R Soc Med*, 93, 341–5.
- Howick, J., Chalmers, I., Glasziou, P., Greenhalgh, T, Heneghan, C., Liberati, A., ... & Thornton, H. (2011). *The 2011 Oxford CEBM Levels of Evidence: Introductory Document*. Retrieved from: <http://www.cebm.net/index.aspx?o=5653>
- Hung, W., Jonassen, D. H. & Liu, R. (2008). Problem-based Learning. In M. Spector, D. Merrill, J. van Merriënboer & M. Driscoll (Hrsg), *Handbook of research on educational communications and technology*, 485-506. New York: Erlbaum.
- Hurd, P.D. (1991). Closing the educational gaps between science, technology, and society. *Theory into Practice*. 30 (4), 251–9.

- Hutchinson, J. C.; Gigerenzer, G. (2005): Simple heuristics and rules of thumb: where psychologists and behavioural biologists might meet. In: *Behavioural processes* 69 (2), S. 97–124. DOI: 10.1016/j.beproc.2005.02.019.
- Ilic, D., Nordin, R., Glasziou, P., Tilson, J., Villanueva, E. (2015): A randomised controlled trial of a blended learning education intervention for teaching evidence-based medicine. In: *BMC medical education* 15, S. 39.
- Iordanou, K. & Constantinou, C. P. (2014). Developing pre-service teachers' evidence-based argumentation skills on socio-scientific issues. *Learning and Instruction, 34*, 42-57.
- Jadad, A.R., Haynes, R.B., Hunt, D., & Browman, G.P.(2000). The internet and evidence-based decision-making: a needed synergy for efficient knowledge management in health care. *Canadian Medical Association Journal*, 162(3), 362-365. In J. Behrens & G. Langer (2004). *Evidence-based Nursing*, S. 197-198.
- Jenkins, E. (1985). History of science education. In T. Husén, T. N. Postlethwaite. *International encyclopedia of education*. Oxford: Pergamon Press, 4453–4456.
- Jette D, Bacon K, Batty C, Carlson M, Ferland A, Hemingway RD, et al. Evidence-based practice: beliefs, attitudes, knowledge, and behaviors of physical therapists'. *Physical Therapy* 2003; 83(9):786-805.
- Johnston, J. M., Leung, G. M., Fielding, R., Tin, Y. K. T. & Ho, L.-M. (2003). The development and validation of a knowledge, attitude and behaviour questionnaire to assess undergraduate evidencebased practice teaching and learning. *Medical Education, 37*, 992-1000.
- Johnston, J. M., Schooling, C. M. & Leung, G. M. (2009). A randomised-controlled trial of two educational modes for undergraduate evidence-based medicine learning in Asia. *BMC Medical Education, 9*(1), 63.
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology and Development, 48*(4), 63-85.

- Kelly, G., Druker, S. & Chen, C. (1998). Students' reasoning about electricity: Combining performance assessments with argumentation analysis. *International Journal of Science Education*, 20, 849-871.
- Khan, K. S. & Coomarasamy, A. (2006). A hierarchy of effective teaching and learning to acquire competence in evidenced-based medicine. *BMC Medical Education*, 6, 59.
- Khine, M. (2012). *Perspectives on scientific argumentation*. Dordrecht: Springer.
- Kienhues, D., Stadtler, M. & Bromme, R. (2011). Dealing with conflicting or consistent medical information on the Web: When expert information breeds laypersons' doubts about experts. *Learning and Instruction*, 21, 193-204.
- Kienle, G., Karutz, M., Matthes, H., Matthiessen, H., Petersen, P. & Kiene, H. (2003). Evidenzbasierte Medizin: Konkurs der ärztlichen Urteilskraft? *Deutsches Ärzteblatt*, 100 (33).
- Kim, M., Anthony, R., Blades, D. (2014). Decision making through dialogur: a case study of analyzing preservice teachers' argumentation on socioscientific issues. *Research in Science Education*. Online publication.
- Klein, M., Wagner, K., Klopp, E. & Stark, R. (2015). Förderung anwendbaren bildungswissenschaftlichen Wissens bei Lehramtsstudierenden anhand fehlerbasierten kollaborativen Lernens: Eine Studie zur Replikation bisheriger Befunde sowie zur Nachhaltigkeit und Erweiterung der Trainingsmaßnahmen. *Unterrichtswissenschaft*, 43 (3), 225-244.
- KMK (2004/2014). *Gemeinsamer Rahmen der Länder für die frühe Bildung in Kindertageseinrichtungen*. Beschluss der Kultusministerkonferenz. Berlin: KMK.
- Kollar, I., Fischer, F. & Hesse, F. W. (2006). Collaboration scripts - a conceptual analysis. *Educational Psychology Review*, 18, 159-185.
- König, S. & Dalbert, C. (2007). Ungewissheitstoleranz und der Umgang mit beruflichen Anforderungen im Lehramt. *Empirische Pädagogik*, 21, 306-321.

- Kopp, B. & Mandl, H. (2011). Fostering argument justification using collaboration scripts and content schemes. *Learning and Instruction, 21*, 636-649.
- Kühle, B. & Peek, R. (2007). Lernstandserhebungen in Nordrhein-Westfalen. Evaluationsbefunde zur Rezeption und zum Umgang mit Ergebnissrückmeldungen in Schulen. *Empirische Pädagogik, 21*, 428-447.
- Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction, 15*, 287–315.
- Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74*, 1245-1260.
- Kuhn, D., Gog, W., Iordanou, K., Shaenfield, D. (2008). Arguing on the Computer: A Microgenetic Study of Developing Argument Skills in a Computer-Supported Environment. *Child Development, 79* (5), 1310-1328.
- Kuhn, D. (2010). Teaching and learning science as argument. *Science Education, 94*, 810-824.
- Kuhn, D. & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science, 22*, 545–552.
- Kuhn, D., Hemberger, L., Khait, V. (2015). Dialogic argumentation as a bridge to argumentative thinking and writing/ La argumentacion dialectica como Puente para el pensamiento y la escritura argumentative. *Infancia y aprendizaje, 39* (1).
- Kulatunga, U., Moog, R. & Lewis, J. (2014). Use of Toulmin's argumentation scheme for student discourse to gain insight about guided inquiry activities in college chemistry. *Journal of college science teaching, 43*, 78-86.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern:*

- Ausbildung und Beruf. *Zeitschrift für Pädagogik*. (51. Beiheft), 47–70.  
Weinheim: Beltz
- Lohmann, V., Seidel, V. & Terhart, E. (2011). Bildungswissenschaften in der universitären Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, 4, 271–302.
- Mähler, C. & Stern, E. (2010). Transfer. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 859-869). Weinheim: Beltz.
- Malhotra, A.; Maughan, D.; Ansell, J.; Lehman, R.; Henderson, A.; Gray, M. et al. (2015): Choosing Wisely in the UK: the Academy of Medical Royal Colleges' initiative to reduce the harms of too much medicine. In: *BMJ (Clinical research ed.)* 350, S. h2308.
- Mamédio da Costa Santos, C., Andrucio de Mattos Pimenta, C., Cuce Nobre, M. R. (2007). The PICO strategy for the research question construction and evidence search. *Revista Latino-Americana de Enfermagem*, 15 (3).
- Mayer, H. O. (2005). *Einführung in die Wahrnehmungs-, Lern- und Werbe-Psychologie*. München: Oldenburg Verlag.
- Mayweg-Paus, E., Thiebach, M. & Jucks, R. (2016). Let me critically question this! – Insights from a training study on the role of questioning on argumentative discourse. *International Journal of Educational Research*. 79.
- McGraw, K. O. & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30–46.
- Meats, E., Heneghan, C., Crilly, M. & Glasziou, P. (2009). Evidence-based medicine teaching in UK medical schools. *Medical Teacher*, 31, 332-337.
- Meier, A., Spada, H. & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, 2, 63-86.
- Mercier, H. (2011). Reasoning serves argumentation in children. *Cognitive Development*, 26, 177-191.

- Miller, S. A. & Forrest, J. L. (2001). Enhancing your practice through evidence-based decision making: PICO, learning how to ask good questions. *Journal of Evidence Based Dental Practice*, 1 (2), 136-141.
- Moore, A. (2007). Beyond reflection: contingency, idiosyncrasy and reflexivity in initial teacher education. In M. Hammersley (Hrsg.), *Educational Research and Evidence-Based Practice* (S. 121-138). London: Sage.
- Muthén, L., & Muthén, B. (1998-2015). Mplus User's Guide (8<sup>th</sup> ed.) Los Angeles, CA: Muthén & Muthén.
- National Center for Education Statistics (2002). *The measurement of instructional background indicators: Cognitive laboratory investigations of the responses of fourth and eighth grade students and teachers to questionnaire items*. Washington, DC: U.S. Department of Education.
- Niemi, H. (2008). Research-based teacher education for teachers' lifelong learning. *Lifelong Learning in Europe*, 13, 61–69.
- OECD (2005). *Teachers Matter. Attracting, developing and retaining effective teachers*. Retrieved from OECD Publishing website: <http://www.oecd.org/edu/school/34990905.pdf>
- Oestreich, M., Romberg, O. (2012). *Keine Panik vor Statistik! Erfolg und Spaß im Horrorfach nichttechnischer Studiengänge*. 4. Auflage. Wiesbaden: Springer Spektrum.
- Oremus, M., Wolfson, C., Perrault, A., Semers, L., Momoli, F. & Moride, Y. (2001). Interrater Reliability of the Modified Jadad Quality Scale for Systematic Reviews of Alzheimer's Disease Drug Trials. *Dementia and Geriatric Cognitive Disorders*, 12, 232-236.
- Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, 328, 463-466.
- Osborne, J., Simon, S., Christodoulou, A., Howell-Richardson, C., Richardson, K. (2013). Learning to argue: A study of four school and their attempt to develop

- the use of argumentation as a common instructional practice and its impact on students. *Journal of Research in Science Teaching*, 50 (3), 315-347.
- Ozdem Yilmaz, Y., Ertepinar, H., Cakiroglu, J. & Erduran, S. (2011). The Nature of Pre-service Science Teachers' Argumentation in Inquiry-oriented Laboratory Context. *International Journal of Science Education*, 10.
- Parr, J.-M. & Timperley, H. S. (2008). Teachers, schools and using evidence: Considerations of preparedness. *Assessment in Education: Principles, Policy & Practice*, 15, 57-71.
- Paul, R. C. (1992). *Critical thinking: What every person needs to survive in a rapidly changing world*. Santa Rosa: Foundation for Critical Thinking.
- Pehmer, A. K., Gröschner, A., Seidel, T. (2015). Fostering and scaffolding student engagement in productive classroom discourse: Teachers' practice changes and reflections in light of teachers' professional development. *Learning Culture and Social Interaction*, 7 (2015), 12-27
- Petty, G. (2009). *Evidence-based teaching*. Cheltenham: Nelson Thornes.
- Porsch, T. & Bromme, R. (2011). Effects of epistemological sensitization on source choices. *Instructional Science*, 39(6), 805-819.
- Prenzel, M., Schöps, K., Rönnebeck, S., Senkbeil, M., Walter, O. Carstensen, C. H. et al. (2007). Naturwissenschaftliche Kompetenz im internationalen Vergleich. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hamman, E. Klieme et al. (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 63-205). Münster: Waxmann.
- Prenzel, M. (2010). Challenges facing the educational system. *Zeitschrift für Pädagogische Historiographie*, 16, 96-99.
- Prenzel, M. & Bauer, J. (2011). Modellierung und Training der Kompetenz angehender und aktiver Lehrkräfte zum Umgang mit Evidenz: Antrag auf Gewährung einer Sachbeihilfe. Unveröffentlichtes Dokument. Technische Universität München.

- Prenzel, M. (2012). Empirische Bildungsforschung morgen: Reichen unsere bisherigen Forschungsansätze aus? In M. Gläser-Zikuda, T. Seidel, C. Rohlf, A. Gröschner & S. Ziegelbauer (Hrsg.), *Mixed Methods in der empirischen Bildungsforschung*, 273-285. Münster: Waxmann.
- Prenzel, M. (2013). Initiativen und Perspektiven zur Weiterentwicklung der Lehrerbildung. In W. Benz, J. Kohler & K. Landfried (Hrsg.) *Handbuch Qualität in Studium und Lehre. Beitrag E. 8.9 (1-22)*. Berlin: Raabe Verlag.
- Prenzel, M. & Bauer, J. (2016). *Abschlussbericht zum DFG-Projekt E4teach – Evidence for teachers. Modellierung und Training der Kompetenz angehender und aktiver Lehrkräfte zum Umgang mit Evidenz*. Unveröffentlichter Bericht. Technische Universität München.
- Qian, S. (2014). Statistics in ecology is for making a "principled argument". *Landscape Ecology*. 29, 42.
- Raspe, H. (1999). Geleitwort. In R. Kunf & L. Fritsche (Hrsg.) D. L. Sackett, S. W. Richardson, W. Rosenberg, R. B. Hanyes (1999). *Evidenzbasierte Medizin. EBM-Umsetzung und –Vermittlung*. München: Zuckschwerdt
- Ramos, K. D., Schafer, S. & Tracz, S. M. (2003). Validation of the Fresno Test of competence in evidence based medicine. *BMJ*, 326, 319-321.
- Rapanta, C. & Macagno, F. (2016). Argumentation methods in educational contexts: Introduction to the special issue. *International Journal of Educational Research*. 79.
- Reiss, K., Sälzer, C., Schiepe-Tiska, A., Klieme, E., Köller, O. (Hrsg.) (2016). *PISA 2015. Eine Studie zwischen Kontinuität und Innovation*. München: Waxmann.
- Rips, L. J., Brem, S. K., Bailenson, J. N. (1999). Reasoning Dialogues. *Current Directions in Psychological Science*, 8 (6), 172-177.
- Robinson, F. P. (1961). *Effective study. Revised Edition*. New York: Harper & Row.
- Rubin, A. (2008). *Practitioner's guide to using research for evidence-based practice*. Hoboken, N.J.: Wiley.

- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B. & Richardson, W. S. (1996). Evidence-based medicine: What it is and what it is not. *BMJ*, 312, 71-72.
- Sackett, D. L. (1997). Evidence-based medicine. *Seminars in Perinatology*, 21, 3-5.
- Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg W. M. C. & Haynes R. B. (2003). *Evidence-Based Medicine: How to Practice and Teach EBM*. London: BMJ Books.
- Sandoval, W. A. & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88, 345-372.
- Savery, J., R. (2006). Overview of Problem-based Learning: Definitions and Distinctions. *Interdisciplinary Journal of Problem-based Learning*, 1(1).
- Schildkamp, K. & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, 26, 482-496.
- Schneider, B., Cornoy, M., Kilpatrick, J., Schmidt, W.H. & Shavelson, R.J. (2007). *Estimating causal effects using experimental and observational designs*. Washington, DC: AERA.
- Schmacke, N. (2004). Innovationen in der Medizin : das magische Dreieck von Evidenz, Ressourcen und Werten. *Zeitschrift für qualitative Bildungs-, Beratungs- und Sozialforschung* 5, 2, 337-360.
- Schrader, J. (2014). Analyse und Förderung effektiver Lehr-Lernprozesse unter dem Anspruch evidenzbasierter Bildungsreform. *Zeitschrift für Erziehungswissenschaft*, 17, 193-223.
- Schunn, C. D., Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science*, 23 (3), 337-370.

- Schworm, S. & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, 99, 285-296.
- Seel, A. (1997). Von der Unterrichtsplanung zum konkreten Lehrerhandeln – Eine Untersuchung zum Zusammenhang von Planung und Durchführung von Unterricht bei Hauptschullehrerstudentinnen. *Unterrichtswissenschaft*, 25, 257-273.
- Seidel, T., Reiss, K., Bauer, J., Bannert, M., Blasini, B., Hubwieser, P., Jurik, V., Knogler, M. et al. (2016). Kompetenzorientierte und evidenzbasierte Lehrerinnen- und Lehrerbildung: Didaktische Weiterentwicklung im Projekt Teach@TUM. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 34 (2).
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shaneyfelt, T., Baum, K. D., Bell, D., Feldstein, D., Houston, T. K., Kaatz, S. ... & Green, M. (2006). Instruments for evaluating education in evidence-based practice: a systematic review. *JAMA*, 296, 1116-1127.
- Shavelson, R., J., & Towne, E. (Hrsg.). (2002). *Scientific Research in Education*. Washington, DC: National Academic Press.
- Slavin, R. E. (1995): *Cooperative Learning: Theory, Research and Practice*. Boston: Allyn and Bacon.
- Slavin, R. E. (2002). Evidence-based education policies: transforming educational practice and research. *Educational Researcher*, 31, 15-21.
- Slavin, R. R. (2008). Perspectives on Evidence-Based Research in Education. What Works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37 (1), 5-14.

- Smyrniou, Z., Petropoulou, E. & Sotiriou, M. (2015). Applying Argumentation Approach in STEM Education: A Case Study of the European Student Parliaments Project in Greece. *American Journal of Educational Research*, 3 (12), 1618-1628.
- Sodian, B. & Koerber, S. (2011). Hypothesenprüfung und Evidenzevaluation im Grundschulalter. *Unterrichtswissenschaft*, 39, 21-34.
- Spiel, C. (2009). Evidenzbasierte Bildungspolitik und Bildungspraxis – eine Fiktion? Problemaufriss – Thesen – Anregungen. *Psychologische Rundschau*, 60, 255-256.
- Stadtler, M. & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning*, 2, 191-210.
- Stahl, E. & Bromme, R. (2007). The CAEB: An instrument for measuring connotative aspects of epistemological beliefs. *Learning and Instruction*, 17(6), 773-785.
- Stark, R. & Mandl, H. (2000). Konzeptualisierung von Motivation und Motivierung im Kontext situierten Lernens. In U. Schiefele, K. P. Wild (Hrsg.). *Interesse und Lernmotivation*. Münster: Waxmann, 95-115.
- Stark, R. (2005). Constructing arguments in educational discourses. In H. Gruber, C. Harteis, R. Mulder & M. Rehl (Hrsg.), *Bridging individual, organisational and cultural aspects of professional learning* (64-71). Regensburg: S. Roderer.
- Stark, R. (2017). Probleme evidenzbasierter bzw. –orientierter pädagogischer Praxis. *Zeitschrift für Pädagogische Psychologie*, 31, 99-110.
- Sternberg, R. J. (Hrsg.). (2005). *Reviewing scientific works in psychology*. Washington, DC: APA.
- Straus, S. E., Glasziou, P., Richardson, W. S., Haynes, B. S. (2011). Evidence-based medicine. How to practice and teach it. Churchill Livingstone: Elsevier.

- Taber, K.S. (2011). Constructivism as educational theory: Contingency in learning, and optimally guided instruction. In J. Hassaskhah. *Educational Theory. Education in a Competitive and Globalizing World*. Nova, 39-62.
- Teasley, S. D. (1997). Talking about reasoning: How important is the peer in peer collaboration? In L. B. Resnick, R. Säljö, C. Pontecorvo, B. Burge (Hrsg). *Discourse, tools and reasoning: Essays on situated cognition*.
- Thomas, E. L. & Robinson, H. A. (1972). *Improving reading in every class: A sourcebook for teachers*. Boston: Houghton Mifflin.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- U.S. Department of Education. (2007). *Strategic Plan for Fiscal Years 2007–12*. Washington, D.C.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Walma van der Molen, J., van Aalderen-Smeets, S. (2013): Investigating and stimulating primary teachers' attitudes towards science: Summary of a large-scale research project. In: *Frontline Learning Research* 1, S. 1–9.
- Weber, S. & Achtenhagen, F. (2008). Forschungs- und evidenzbasierte Lehrerbildung. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität* (S. 477-487). Weinheim: Beltz.
- Weinberger, A., Stegmann, C. & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior*, 26, 506-515.
- Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T. & Carstensen, C.H. (2011). Development of competencies across the life span. *Zeitschrift für Erziehungswissenschaft*, 14, Supplement 2, 67-86.

- 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.
- Weiß, S., Schramm, S. & Kiel, E. (2014). Was sollen Lehrerinnen und Lehrer können? Anforderungen an den Lehrer/innenberuf aus Sicht von Lehrkräften und Ausbildungspersonen. *Forum: Qualitative Sozialforschung*, 15 (3).
- What Works Clearinghouse (2014). Procedures and Standards Handbook Version 3.0. Entnommen am 29.03.2018 von [https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc\\_procedures\\_v3\\_0\\_standards\\_handbook.pdf](https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf)
- Whitehurst, G. J. (2002). *Evidence-based education (EBE)*. Presentation at the conference on student achievement and school accountability, Orlando, USA. Entnommen am 26.03.2018 von <https://www2.ed.gov/nclb/methods/whatworks/eb/edlite-slide001.html>
- Williams, D. & Coles, L. (2007). Evidence-based practice in teaching: an information perspective. *Journal of Documentation*, 63, 812-835.
- Wilson, M. (2005). *Constructing measures*. New York: Taylor & Francis.
- Wilson, M. & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education*, 13, 181-208.
- Wirtz, M. A. (Ed.) (2017). *Dorsch – Lexikon der Psychologie*. Bern: hogrefe Verlag.
- Wiseman, A. W. (2010). The uses of evidence for educational policymaking: Global contexts and international trends. *Review of Research in Education*, 34, 1-24.
- Wittwer, J., Bromme, R. & Jucks, R. (2004). Kann man dem Internet trauen, wenn es um die Gesundheit geht? Die Glaubwürdigkeitsbeurteilung medizinischer Fachinformationen im Internet durch Laien. *Zeitschrift für Medienpsychologie*, 16 (2), 48-56.
- World Health Organization (2010). *Antiretroviral Therapy for HIV Infection in Infants and Children: Towards Universal Access. Recommendations for a Public Health Approach. 2010 Revision*. Genf: WHO Press.

- World Medical Association (WMA) (2000). *Declaration of Helsinki. Ethical Principles for Medical Research involving human subjects*. Adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964. Entnommen am 31.03.2018 von <http://www.who.int/bulletin/archives/79%284%29373.pdf>
- WWC What Works Clearinghouse (2008). *WWC Procedures and Standards Handbook. Version 3.0*. Entnommen am 09.03.2018 von: [https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc\\_procedures\\_v3\\_0\\_standards\\_handbook.pdf](https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf)
- Young, J. M. & Ward, J. E. (2001). Evidence-based medicine in general practice: beliefs and barriers among Australian GPs. *Journal of Evaluation in Clinical Practice*, 7, 201-210.
- Zemal-Saul, C., Munford, D., Crawford, B., Friedrichsen, P., & Land, S. (2003). Scaffolding preservice science teachers' evidence-based arguments during an investigation of natural selection. *Research in Science Education*, 32, 437-463.
- Zohar, A. (2008). Science teacher education and professional development in argumentation. In S. Erduran & M. P. Jimenez-Aleixandre (Hrsg.), *Argumentation in science education* (S. 245-268). Dordrecht: Springer.
- Zumbach, J. , Haider, K. , Mandl, H. (2008). Fallbasiertes Lernen: Theoretischer Hintergrund und praktische Anwendung. In J. Zumbach & H. Mandl (Hrsg.), *Pädagogische Psychologie in Theorie und Praxis. Ein fallbasiertes Lehrbuch* (S. 1–11). Göttingen: Hogrefe.

## Abbildungsverzeichnis

<i>Abbildung 1.</i> Überblick über die Zusammenhänge und Inhalte der drei Einzelstudien im Rahmen dieser Arbeit.....	15
<i>Abbildung 2.</i> Klassen I – V zur kritischen Bewertung von Evidenz aus (klinischen) Studien nach der Cochrane Klassifikation.....	19
<i>Abbildung 3:</i> Bedingungsmodell zum professionellen Umgang mit Evidenz bei Lehramtsstudierenden.....	23
<i>Abbildung 4:</i> Darstellung der drei Säulen für evidenzorientierte Praxis im Lehrberuf .....	40
<i>Abbildung 5.</i> Überblick über die Zeiträume der Datenerhebung.....	70
<i>Abbildung 6:</i> Boxplot Diagramm: Darstellung der Globalratings der Kontrollgruppe im Vergleich zur Experimentalgruppe .....	83
<i>Abbildung 7:</i> Fehlerbalkendiagramm: Vergleich der Ergebnisse der hochinferenten Globalratings beider Bedingungen, aggregiert über beide Statements .....	96
<i>Abbildung 8:</i> Funktionen des argumentativen Nutzens von Evidenz: Fragen .....	112
<i>Abbildung 9:</i> Funktionen des argumentativen Nutzens von Evidenz: nicht-transaktive Statements .....	113
<i>Abbildung 10:</i> Funktionen des argumentativen Nutzens von Evidenz: transaktive Statements .....	113

## Tabellenverzeichnis

<i>Tabelle 1:</i> Ranking für Evidenzquellen nach GRADE. ....	21
<i>Tabelle 2.</i> Kompetenzstufen zum kritischen Bewerten von Evidenz.....	43
<i>Tabelle 3.</i> Kompetenzstufen zum argumentativen Nutzen von Evidenz .....	46
<i>Tabelle 4.</i> Inhaltlicher Aufbau des Trainingsprogramms.....	63
<i>Tabelle 5.</i> Überblick über die Domänen und Fallbeispiele in den Gruppendiskussionen sowie in den schriftlichen Statements .....	66
<i>Tabelle 6.</i> Ablaufbeschreibung des Experiments in der Experimental- und Kontrollgruppe.....	71
<i>Tabelle 7:</i> Ablaufbeschreibung des quasi-experimentellen Designs in der Pilotierung, sowie der Präpilotierung.....	78
<i>Tabelle 8:</i> Kategorienschemata zur Auswertung der schriftlichen Statements, sowie der Gruppendiskussionen in der Pilotierungsstudie. ....	80
<i>Tabelle 9:</i> Detaillierte Ergebnisse über beide Settings hinweg (so veröffentlicht in Wenglein et al., 2015).....	84
<i>Tabelle 10:</i> Anwendung der Heuristiken in den Gruppendiskussionen .....	86
<i>Tabelle 11:</i> Konsistenz der Ergebnisse der Globalratings über die Settings hinweg	86
<i>Tabelle 12:</i> Überblick über Motivation und Akzeptanz an beiden Trainingstagen ....	87
<i>Tabelle 13:</i> Kategorienschemata zur Auswertung der schriftlichen Statements, sowie der Gruppendiskussionen in der Hauptstudie.....	92

<i>Tabelle 14:</i> Hierarchische Regressionsmodelle zur Berechnung von trainings- und Domäneneffekten. Sowie weiteren Einflussfaktoren (signifikante Effekte sind fett gedruckt) .....	94
<i>Tabelle 15:</i> Ergebnisse der niedrinferenten Ergebnisse aggregiert über beide Statements .....	97
<i>Tabelle 16:</i> Konsistenz der Ergebnisse auf Individualebene über die verschiedenen Settings und Problemkontexte hinweg.....	98
<i>Tabelle 17:</i> Vergleich der Kompetenzen mit Domänenwissen (durch Fachhintergrund und Vorwissen) auf Basis der Ergebnisse aus den Gruppendiskussionen (aggregiert über beide Diskussionen).....	99
<i>Tabelle 18:</i> Vergleich der Kompetenzen vor dem Training mit Domänenwissen auf Basis der Ergebnisse aus den schriftlichen Statements (aggregiert über beide Statements) .....	100
<i>Tabelle 19:</i> Kodierschema zur Bestimmung der Qualität, sowie der Gründe der Evidenznutzung .....	105
<i>Tabelle 20:</i> Kodierschema zur Bestimmung der Funktion der Evidenznutzung (Kuhn & Crowell, 2011) .....	106
<i>Tabelle 21:</i> Beispiel für erfolgreiches evidenzbasiertes Argumentieren (Domäne: Gesundheitswissenschaften).....	108
<i>Tabelle 22:</i> Beispiel für kritisches Hinterfragen von Evidenz (Domäne: Gesundheitswissenschaften).....	108
<i>Tabelle 23:</i> Beispiel für erfolgreiches evidenzbasiertes Argumentieren (Domäne: Bildungswissenschaften).....	109
<i>Tabelle 24:</i> Beispiele aus den Diskussionen für häufige Fehler und Schwierigkeiten .....	110
<i>Tabelle 25:</i> Beispiel für fehlende Transaktivität.....	111
<i>Tabelle 26:</i> Gründe für das argumentative Nutzen von Evidenz .....	111
<i>Tabelle 27:</i> Gründe für das Nutzen von Evidenz aufgeteilt nach Leistungsgruppen .....	112

# Anhang

## A. Artikel und Manuskripte

### A.1. Veröffentlichtes Manuskript zu Studie 1

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.

Sandra Wenglein, Johannes Bauer, Susanne Heiningler,  
Manfred Prenzel

## **Kompetenz angehender Lehrkräfte zum Argumentieren mit Evidenz: Erhöht ein Training von Heuristiken die Argumentationsqualität?**

Pre-Service Teachers' Evidence-Based Argumentation Competence:  
Can a Training of Heuristics Improve Argumentative Quality?

In Diskussionen um evidenzbasierte Praxis im Bildungswesen werden zunehmend Forderungen artikuliert, Lehrkräfte sollten bildungswissenschaftliche Forschungsbefunde bewerten und für schulrelevante Fragestellungen nutzen können. In der vorliegenden quasi-experimentellen Studie erprobten wir eine Intervention, die Lehramtsstudierenden grundlegende Heuristiken zur Bewertung und argumentativen Nutzung von Evidenz vermittelt. Fragestellungen waren, (a) wie sich die Intervention in der Argumentationsqualität in kooperativen Gruppendiskussionen während des Trainings und in einer nachfolgenden individuellen Testsituation niederschlägt, (b) ob die Teilnehmenden der Trainingsgruppe gegenüber einer Kontrollgruppe eine gesteigerte Argumentationsqualität aufwiesen und (c) wie konsistent die Argumentationsqualität der Trainingsgruppe über die verschiedenen Situationen (Gruppendiskussion vs. Test) und Problemkontexte (alltags- vs. professionsbezogen) ausfällt. Hierzu wurden Daten von  $N = 30$  Lehramtsstudierenden aus videografierten Gruppendiskussionen und der Testbearbeitung analysiert. Die Ergebnisse zeigten eine insgesamt hohe Argumentationsqualität in der Trainingsgruppe, die ihre Argumente systematischer und besser mit Evidenz begründete als die Kontrollgruppe. Innerhalb der Trainingsgruppe fiel die Argumentationsqualität über die vier Messgelegenheiten hinweg konsistent aus.

**Schlüsselwörter:** evidenzbasierte Praxis, Argumentation, Lehramtsstudierende, Intervention

Current debates on evidence-based practice in education increasingly demand that teachers should be able to evaluate and use scientific evidence to inform their practice and decision-making. The present quasi-experimental study investigated a training program on pre-service teachers' competence in evidence-based argumentation. We analyzed (a) the argumentative quality that the training instigated in group discussions during the training and in a subsequent individual written test, (b) whether the trained students' argumentation differed from that of the control group, and (c) the consistency of the quality of the trained students' argumentation in the various settings (group discussion vs. test) and problem contexts (health- vs. school-related problems). For this purpose, we analyzed the group discussions and test answers of  $N = 30$  participants. The results showed an overall high argumentative quality in the training group. Trained students justified their arguments better and more systematically with evidence than students from the control group. Moreover, in the training group, argumentative quality was consistent across the four measurement occasions.

**Keywords:** evidence-based practice, argumentation, pre-service teachers, intervention

## 1 Einleitung

Die internationale Diskussion um Standards von Lehrerprofessionalität betrachtet Lehrkräfte zunehmend – ähnlich wie z. B. Medizinerinnen bzw. Mediziner oder Ingenieurinnen bzw. Ingenieure – als Angehörige einer hochausgebildeten und forschungsbasierten Profession (z. B. Bauer & Prenzel, 2012; Darling-Hammond & Bransford, 2005; European Commission [EC], 2007, 2013). Dabei gewinnt die Forderung nach evidenzbasierter Praxis (z. B. Rubin, 2008; Sackett, Rosenberg, Gray, Haynes & Richardson, 1996), wie in vielen anderen professionellen Tätigkeitsfeldern, auch im Bildungsbereich an Bedeutung (zusammenfassend: Bromme, Prenzel & Jäger, 2014). Evidenzbasierte Praxis bezieht sich dabei auf die Integration individueller Expertise mit dem gegenwärtig besten verfügbaren wissenschaftlichen Wissen (Davies, 1999; Petty, 2009; Sackett et al., 1996; Slavin, 2002). Professionelles Handeln soll also nicht lediglich auf Basis individuell-subjektiver Überzeugungen und Erfahrung erfolgen, sondern am jeweils besten verfügbaren wissenschaftlichen Wissen orientiert bzw. kritisch daran reflektiert sein (Bromme et al., 2014; Sackett et al., 1996). Evidenz ersetzt dabei nicht individuelle Erfahrung und Expertise, sondern dient als Ergänzung oder Korrektiv und muss situativ interpretiert werden.

Von Seiten der Bildungspolitik (EC, 2013; Kultusministerkonferenz [KMK], 2004) und -forschung (z. B. Niemi, 2008; Parr & Timperley, 2008; Petty, 2009) wurden im Kontext dieser Diskussionen Forderungen laut, Lehrkräfte müssten über förderliche Einstellungen gegenüber evidenzbasierter Praxis sowie über grundlegende Kompetenzen zur kritischen Bewertung und argumentativen Nutzung wissenschaftlicher Evidenz verfügen. Wie solche Kompetenzen bei (angehenden) Lehrkräften tatsächlich ausgeprägt sind, ist weitgehend ungeklärt. Allerdings gibt es wenig Grund, hier optimistisch zu sein. Erstens ist in der bisherigen Schulpraxis eine konsequente Ausrichtung an wissenschaftlichen Erkenntnissen keineswegs selbstverständlich (Hargreaves, 2000; Schildkamp & Kuiper, 2010). Lehrkräfte orientieren sich offenbar stark an vorgefundenen Praxisformen, Unterrichtsskripts und eigener Erfahrung (Hargreaves, 2000). Zweitens ist aufgrund der geringen bildungswissenschaftlichen Studienanteile (Bauer, Diercks, Rösler, Möller & Prenzel, 2012; Lohmann, Seidel & Terhart, 2011) nicht anzunehmen, dass Kompetenzen zur Bewertung und Nutzung bildungswissenschaftlicher Evidenz in der Lehramtsausbildung systematisch gefördert werden, auch wenn sich inzwischen zunehmend eine empirische Forschungsorientierung feststellen lässt (Bauer & Prenzel, 2012; Weber & Achtenhagen, 2008).

Zudem existieren bislang nur wenige Interventionsstudien, die zeigen wie relevante Kompetenzbereiche gefördert werden könnten. Beispielsweise legten Iordanou und Constantinou (2014) eine Trainingsstudie zum evidenzbasierten Argumentieren bei Lehramtsstudierenden vor. Diese Arbeit fokussierte auf das Argumentieren zu fachlichen Themen (*Science Education: Klimawandel und Energiegewinnung*) in computerbasierten Diskussionen. Inwiefern die Ergebnisse solcher Studien auf die Integration und argumentative Nutzung verschiedener Evidenzquellen zu schulischen Problemstellungen im Kontext evidenzbasierten Handelns von Lehrkräften übertragbar sind, ist eine offene Frage.

Das Projekt *E4teach – Evidence for Teachers* greift diesen Forschungsbedarf auf und untersucht, wie die Kompetenz zur Bewertung und die argumentative Nutzung (bildungs-)wissenschaftlicher Evidenz bei angehenden Lehrkräften gemessen und im Rahmen eines Trainings gefördert werden kann. Hierzu entwickelten wir eine Kurzzeitintervention, die flexi-

bel in bestehende Lehramtscurricula integrierbar ist und auf der Vermittlung von Heuristiken basiert. In diesem Beitrag präsentieren wir Ergebnisse aus einer ersten quasi-experimentellen Überprüfung dieses Trainings mit Fokus auf der argumentativen Nutzung von Evidenz. Dabei sollte geklärt werden, (a) wie sich die erlernten Heuristiken in der Argumentationsqualität in Gruppendiskussionen während des Trainings und in einer nachfolgenden individuellen Testsituation niederschlagen, (b) ob die Teilnehmenden der Trainingsgruppe gegenüber einer Kontrollgruppe eine gesteigerte Argumentationsqualität aufweisen und (c) wie konsistent die Argumentationsqualität der Trainingsgruppe über die verschiedenen Situationen und Problemkontexte hinweg ausfällt.

### 1.1 Argumentative Nutzung von Evidenz

Die Kompetenz, Evidenz für einen vorliegenden Problemkontext in ihrer Relevanz und Aussagekraft bewerten und argumentativ nutzen zu können, ist ein zentraler Bestandteil typischer Handlungsmodelle evidenzbasierter Praxis (Ramos, Schafer & Tracz, 2003; Rubin, 2008; Shaneyfelt et al., 2006). Vorgelagerte Facetten sind das Formulieren einer Fragestellung und eine systematische Recherche relevanter Evidenz; nachgelagert ist die Evaluation evidenzbasierter Entscheidungen. Im finnischen Modell einer forschungsbasierten Lehrerbildung wird zudem der Fähigkeit von Lehrkräften, selbst handlungsbezogene Forschung durchführen zu können (z. B. zur Evaluation eigener Interventionen), eine hohe Bedeutung beigemessen (Niemi, 2008). Evidenz dient dabei erstens dazu, Entscheidungen oder Handlungen direkt zu informieren. Da im Handlungsmodell evidenzbasierter Praxis die rationale Begründung eine große Rolle spielt (Sackett et al., 1996), kommt der Evidenz zweitens eine argumentative Legitimationsfunktion für eine Entscheidung oder Handlung zu (Chisholm, 1980; Curd & Cover, 1998; Sackett et al., 1996). Dies wird auch im Argumentationsmodell von Toulmin (1958) ersichtlich. Evidenz dient darin primär zur faktischen Untermauerung der Begründung für eine Behauptung (*Ground*). Darüber hinaus kann sie als *Backing* die Geltung einer Schlussregel (*Warrant*) unterstützen (vgl. Booth, Colomb & Williams, 2008; Osbourne, 2010). Einschränkungen der Aussagekraft von Evidenz oder Widersprüchlichkeiten zwischen Evidenzquellen können dabei explizit anerkannt (*Qualifier*) oder in einer Widerlegung (*Rebuttal*) entkräftet werden.

Dabei ist anzunehmen, dass eine höhere Argumentationsqualität mit einer systematischeren und reflektierteren Integration von Evidenz in Argumentationszusammenhängen einhergeht (Kuhn & Crowell, 2011; Wilson & Sloane, 2000). Im Bereich der *Science Education* legte Wilson (2005; Wilson & Sloane, 2000) einen entsprechenden Ansatz vor, der spezifische Qualitäten von Argumentationsmustern Kompetenzstufen der Nutzung von Evidenz zuordnet. Antworten auf einem niedrigen Niveau basieren auf subjektiven Meinungen, irrelevanter Evidenz oder liefern überhaupt keine Begründungen. Ein hohes Kompetenzniveau ist dagegen durch die Untermauerung von Aussagen mit relevanter und akkurater Evidenz gekennzeichnet, wobei die verwendete Evidenz zusätzlich kritisch hinterfragt wird (vgl. Bell, 2010; Sandoval & Reiser, 2004). Für eine hohe Argumentationsqualität müssen also auch vorliegende Einschränkungen in der Aussagekraft von Evidenz (Fragilität) oder Konflikte in den Aussagen verschiedener Evidenzquellen berücksichtigt werden (Bromme, 2008; Kuhn & Crowell, 2011).

Für die vorliegende Studie ermöglicht das beschriebene Modell von Wilson (2005) eine globale Einschätzung der Kompetenz zum evidenzbasierten Argumentieren. Um Argumenta-

tionsqualität tiefergehend zu analysieren, können zusätzlich einzelne Argumente detaillierter betrachtet werden. Relevante Merkmale betreffen erstens die Klarheit des angeführten Arguments, also wie explizit logische Verbindungen zwischen Aussagen gemacht werden (Bell, 2010; Sandoval & Reiser, 2004). Zweitens können die Qualität der verwendeten Evidenz und die Art der Nutzung weiter differenziert werden (Bell, 2010; Brown, Nagashima, Fu, Timms & Wilson, 2010; Felton & Kuhn, 2001; Kuhn & Crowell, 2011; Sandoval & Reiser, 2004). Bell (2000) grenzt diesbezüglich fünf Qualitäten ab: (a) Argumente ohne Untermauerung durch Evidenz, (b) mit Nutzung irrelevanter Evidenz, (c) eine rein beschreibende Nutzung, in der eine Evidenzquelle lediglich nacherzählt wird, sowie Argumente, in denen Evidenz mit (d) einer oder (e) mehreren wissenschaftlichen Annahmen stringent verbunden wird (*Single/Multiple Warrant*). Darüber hinaus kann analysiert werden, welche Evidenzquellen herangezogen werden, z. B. ob es sich nur um persönliche Alltagserfahrungen (*Pseudoevidence*; Kuhn, 1991) oder um wissenschaftliche Evidenz handelt (Bell, 2010).

## 1.2 Training von Kompetenzen zum evidenzbasierten Argumentieren

Trotz der regen Debatten um evidenzbasierte Praxis existieren bislang nur wenige Arbeiten zur Förderung relevanter Kompetenzen bei Lehramtsstudierenden (Iordanou & Constantinou, 2014). Für die Gestaltung entsprechender Trainingsmaßnahmen ist zu erwägen, welche Inhalte und lehr-lerntheoretischen Gestaltungsprinzipien vielversprechend erscheinen.

*Fokus auf Heuristiken.* Bezüglich der Trainingsziele und -inhalte muss der relativ enge Rahmen berücksichtigt werden, den Lehramtscurricula bieten (Bauer et al., 2012; Lohmann et al., 2011). Der Erwerb umfassender Kompetenzen im Sinn von Handlungsmodellen evidenzbasierter Praxis (Rubin, 2008) ist aufgrund der Vielzahl an Studieninhalten kaum realistisch. Ziel unseres Projekts war deshalb die Entwicklung einer effektiven Kurzzeitintervention (2 Tage), die flexibel in Lehramtscurricula integrierbar ist oder extracurricular angeboten werden kann. Ein solches Training kann die Funktion eines *Eisbrechers* haben, indem es grundlegende Heuristiken vermittelt, die Studierenden die Rezeption von Evidenz erleichtern. Dies ist auch vor dem Hintergrund zu sehen, dass (angehende) Lehrkräfte offenbar häufig Verständnisprobleme auch mit einfachen Formen empirischer Ergebnisdarstellung, geringe Kompetenzen im evidenzbasierten Argumentieren sowie dysfunktionale Einstellungen gegenüber Ergebnissen der Bildungsforschung aufweisen (Hargreaves, 2000; Kühle & Peek, 2007; Parr & Timperley, 2008; Schildkamp & Kuiper, 2010; Williams & Coles, 2007; Zembal-Saul, Munford, Crawford, Friedrichsen & Land, 2003). Diese Probleme treffen jedoch nicht nur auf Lehrkräfte, sondern z. B. auch auf Medizinerinnen und Mediziner zu (Donner-Banzhoff, 2007; Johnston, Leung, Fielding, Tin & Ho, 2003; Young & Ward, 2001). Evidenzbasierte Praxis basiert deshalb in der Medizin stark auf der Verwendung von Heuristiken. Ein Beispiel hierfür sind hierarchische Modelle, die mit Evidenzstufen eine Rangordnung der Aussagekraft von Studientypen vorschlagen und somit Praktikerinnen und Praktikern eine Einschätzung von Studien erlauben (z. B. Higgins & Green, 2011; Howick et al., 2011). Solche Werkzeuge haben den Charakter einfacher Heuristiken (Gigerenzer, 2008), denn sie erlauben auf Basis begrenzten Wissens zu einer Lösung zu kommen, die möglicherweise nicht optimal, in vielen Fällen aber akzeptabel bezüglich des Ausgangsproblems ist.

Entsprechend zielt das entwickelte Training auf den Erwerb von Heuristiken ab, die helfen, die Aussagekraft empirischer Studien einzuschätzen und ihre Ergebnisse argumentativ für

einen gegebenen Problemkontext zu verwenden. Diese Heuristiken basieren erstens auf gängigen Begutachungskriterien zur Einschätzung empirischer Studien (z. B. Abelson, 1995; Sternberg, 2005). Zweitens wird bezüglich der argumentativen Nutzung von Evidenz ein Argumentationsschema vermittelt, das auf dem oben dargestellten Modell Toulmins (1958) basiert (vgl. Booth et al., 2008). Auch dieses Argumentationsschema hat im vorliegenden Kontext die Funktion einer Heuristik, indem es die Struktur eines evidenzbasierten Arguments und damit generelle Möglichkeiten zur argumentativen Nutzung von Evidenz aufzeigt.

*Instruktionale Prinzipien.* Das Trainingsdesign basiert wesentlich auf Ansätzen zum Training evidenzbasierter Praxis in der Medizinerbildung (z. B. Khan & Coomarasamy, 2006) sowie auf Fördermaßnahmen zum evidenzbasierten Argumentieren aus der *Science Education* (Fischer et al., 2014; Kopp & Mandl, 2011; Kuhn, 2010; Osbourne, 2010) und deren Anwendung auf (angehende) Lehrkräfte (Iordanou & Constantinou, 2014; Zembal-Saul et al., 2003; Zohar, 2008). Diese Auswahl erfolgte unter der Annahme, dass die entsprechenden Lehr-Lernprinzipien auch für den Erwerb von Fähigkeiten zur Bewertung und argumentativen Nutzung von Evidenz für professionsbezogene Problemstellungen des Lehrberufs anwendbar sind.

Auf dieser Grundlage folgte das Trainingsdesign primär drei instruktionalen Prinzipien: (a) Durch die *Situierung in authentische, professions- und alltagsrelevante Problemfälle* kann ein Anwendungsbezug geschaffen werden, der sich günstig auf motivationale Voraussetzungen und die kognitive Verarbeitung auswirkt (Fischer et al., 2014; Kahn & Coomarasamy, 2006). Entsprechend wurde zur Motivierung der Teilnehmenden ein problemorientierter Einstieg umgesetzt (zunächst zu einer alltagsrelevanten gesundheitsbezogenen Frage, dann weiterführend zu bildungswissenschaftlichen Problemstellungen). (b) *Kollaboratives und dialogisches Lernen* unterstützt die Entwicklung argumentativer Kompetenzen (Fischer et al., 2014; Iordanou & Constantinou, 2014; Kopp & Mandl, 2011; Kuhn, 2010; Osbourne, 2010). Solche Lernanteile fördern eine produktive Auseinandersetzung mit kognitiven Konflikten im Umgang mit Evidenz sowie das systematische Einnehmen multipler Perspektiven. Zudem können hierdurch Entwicklungen auf der sozialen Ebene angeregt werden (Vygotsky, 1978). (c) *Kognitive Modellierung und Unterstützung (Scaffolding)* haben sich für das Erlernen der komplexen epistemischen Prozesse wissenschaftlichen Denkens als hilfreich erwiesen (Collins, Brown & Newman, 1989; Iordanou & Constantinou, 2014; Kopp & Mandl, 2011; Schworm & Renkl, 2007; Zembal-Saul et al., 2003). Zudem kann eine qualitativ hochwertige Kooperation durch die Verwendung von Kooperationsskripts unterstützt werden (Fischer et al., 2014; Kollar, Fischer & Hesse, 2006). Den genannten Prinzipien folgend wurden im Training die Heuristiken zur Bewertung und argumentativen Nutzung von Evidenz also problemorientiert erarbeitet, tutoriell modelliert und mit ausblendender Unterstützung in kooperativen Übungen vertieft (s. Trainingsablauf).

## 2 Fragestellungen und Hypothesen

Ziel der vorliegenden Studie war die erste Überprüfung des entwickelten Trainingsansatzes auf Basis der diskutierten Gestaltungsprinzipien in einem quasi-experimentellen Design. Dabei sollten folgende Fragestellungen adressiert werden:

1. Wie manifestieren sich die erlernten Heuristiken zur argumentativen Nutzung von Evidenz (a) in den problembezogenen Gruppendiskussionen während des Trainings und (b) in der nachfolgenden Testsituation, in der Evidenz individuell bewertet und problembezogen argumentativ genutzt werden muss? Wie ist die Argumentationsqualität kriterial einzuschätzen?
2. Trägt das Training bei der Trainingsgruppe (TG) gegenüber der Kontrollgruppe (KG) zu einer gesteigerten Argumentationsqualität bei?
3. Wie konsistent ist die Argumentationsqualität (a) zwischen dem kooperativen Setting (Gruppendiskussion) und der individuellen Situation im schriftlichen Test sowie (b) über die alltags- und professionsrelevanten Problemkontexte in den Gruppendiskussionen hinweg? Gelingt ein Transfer der Heuristiken?

Fragestellung 1 bezieht sich auf die Beschreibung und kriteriale Einschätzung der Argumentationsqualität bei den untersuchten Lehramtsstudierenden. Diese Einschätzung erfolgte einerseits anhand des Stufenmodells nach Wilson (2005), andererseits anhand von relativen Häufigkeiten einer qualitativ hochwertigen Nutzung von Evidenz (Bell, 2010; Sandoval & Reiser, 2004). Fragestellung 2 zielt auf eine Einschätzung der Trainingswirksamkeit. Hier vermuteten wir eine insgesamt höhere Qualität in der argumentativen Nutzung von Evidenz in der TG. Fragestellung 3 betrifft eine detailliertere Analyse der Argumentationsqualität in der TG über die verschiedenen Messgelegenheiten. Sollte sich, wie von uns vermutet, eine relativ hohe Konsistenz über die verschiedenen Settings bzw. Problemkontexte hinweg zeigen, wäre dies ein Indiz für einen gelingenden Transfer der Heuristiken (vgl. Mähler & Stern, 2010).

### 3 Methoden

#### 3.1 Stichprobe

An der Studie nahmen  $N = 30$  Studierende des Lehramtes an Gymnasien teil (TG:  $n = 14$ , KG:  $n = 16$ ). Die Rekrutierung erfolgte über einschlägige Lehrveranstaltungen. Die Teilnehmenden waren zu 75 % weiblich, durchschnittlich  $M = 23.36$  Jahre alt ( $SD = 2.98$ ), im sechsten Studiensemester ( $M = 6.27$  Semester,  $SD = 2.61$ ), studierten zu 80 % mindestens ein MINT-Fach und gaben zu 73 % an, 1-5 wissenschaftliche Artikel pro Woche zu lesen. Da die Zuweisung zu den experimentellen Bedingungen nicht zufällig erfolgte, wurden die Gruppen anhand der genannten Hintergrundmerkmale verglichen. Dabei zeigten sich keine bedeutsamen Unterschiede (alle  $p > .20$ ), mit Ausnahme der Fächerverteilung. Diesbezüglich studierten alle Teilnehmenden der KG mindestens ein MINT-Fach, während dies in der TG nur zu 57 % der Fall war. Aufgrund dieser Verteilung können potenzielle Fächereffekte in der vorliegenden Studie nur annäherungsweise durch einen Vergleich der Trainingseffekte in der Gesamtgruppe und in der Teilgruppe von Studierenden mit mindestens einem MINT-Fach analysiert werden.

#### 3.2 Design

Die Studie war als Quasi-Experiment angelegt. Die Zuweisung erfolgte über die Anmeldung der Teilnehmenden zu verschiedenen angebotenen Trainingsterminen (*haphazard assignment*; Shadish, Cook & Campbell, 2002). Vor dem Experiment bearbeiteten alle Teilneh-

menden einen Fragebogen zu relevanten Hintergrundvariablen. Die KG bearbeitete nach dem Eingangsfragebogen den schriftlichen Test zum Umgang mit Evidenz. Die TG nahm im Anschluss an dem zweitägigen Training zum kompetenten Umgang mit Evidenz teil (ca. 6 Stunden aktive Lernzeit am ersten und 4 Stunden am zweiten Tag; Ablauf s. Abbildung 1). Um relevante Prozessdaten zu erheben, wurden die Trainings komplett videografiert. Nach Beendigung des Trainings bearbeiteten die Teilnehmenden den im Projekt entwickelten schriftlichen Test zur Nutzung von Evidenz sowie einen Abschlussfragebogen.

Training: „Kompetenter Umgang mit Evidenz“						
Eingangsfragebogen	Verständnis für Nutzen von Evidenz	Heuristiken zur kritischen Bewertung von Evidenz		Heuristiken zum argumentativen Nutzen von Evidenz		Test
	Problemorientierter Einstieg: Alltag (Gesundheit) und Lehrerberuf	Allgemeine Bewertungskriterien, methodische Grundlagen	Tutoriel angeleitete Fallbearbeitung (beide Domänen)	Grundlagen evidenzbasierter Argumentierens, Umgang mit Fragilität und Konflikthaftigkeit	Gruppendiskussionen mit Fällen aus beiden Domänen	

Abbildung 1: Überblick zu den Trainingsbausteinen.

### 3.3 Trainingsablauf

Das Training beginnt am ersten Tag mit einem problemorientierten Einstieg anhand einer Situationsvignette mit gesundheitswissenschaftlicher Thematik (Gesundheitsrisiken durch Fastfood). Dies dient erstens der Schaffung einer günstigen motivationalen Ausgangsbasis, da wir davon ausgehen, dass die Motivation, sich mit einem alltagsrelevanten Thema zu beschäftigen, zunächst höher ist als bei einem bildungswissenschaftlichen Thema. Zweitens soll dadurch Vorwissen über Evidenznutzung aktiviert werden. Im Anschluss folgt eine Einführung in Kernprinzipien evidenzbasierter Praxis, wobei immer wieder Analogien zum medizinischen Kontext gezogen werden. In dieser Phase werden die Heuristiken zur Bewertung und argumentativen Nutzung von Evidenz problemorientiert erarbeitet und in kooperativen Übungen vertieft. Das erworbene Wissen wenden die Teilnehmenden mit abnehmender tutorieller Unterstützung auf weitere Situationsvignetten an (abwechselnd gesundheits- und bildungswissenschaftlich). Am zweiten Trainingstag finden zwei größer angelegte Übungen mit Gruppendiskussionen zu einer gesundheits- und einer schulbezogenen Problemstellung statt (Zusammenhang von Hypertonie und Salzzufuhr; Förderbedarf von Mädchen im Umgang mit Computern). Die Lernenden erhalten jeweils eine Fallvignette und zugeordnete Texte, die unterschiedlich aussagekräftige Studien mit teilweise konfligierenden Informationen beinhalten. Diese Evidenzquellen werden von den Lernenden zunächst gelesen und hinsichtlich ihrer Belastbarkeit kritisch bewertet. Im Anschluss daran sollen sie in einer Gruppendiskussion (3-4 Personen pro Gruppe) eine gemeinsame Lösung finden. Die Gruppendiskussionen werden mit Hilfe eines Kooperationskripts unterstützt, das die Vorbereitung und den Ablauf strukturiert, ohne allerdings bestimmte Positionen für die Argumentation vorzugeben.

### 3.4 Instrumente und Auswertung

*Qualität der argumentativen Nutzung von Evidenz.* Die Argumentationsqualität wurde über die videografierten Gruppendiskussionen sowie den im Projekt entwickelten Test zum Umgang mit Evidenz erfasst, sodass insgesamt vier Argumentationssituationen vorliegen: zwei Gruppendiskussionen (gesundheits- und schulbezogen) und zwei textbasierte Aufgaben (beide schulbezogen). Die Aufgaben im Test sind strukturell vergleichbar zu den Aufgaben im Training. Den Probandinnen und Probanden werden zwei Fallvignetten zu schulbezogenen Problemen und jeweils zugeordnete Evidenzquellen dargeboten. Letztere bestehen pro Vignette aus drei Texten im Umfang von 2-3 Seiten, die insgesamt vier empirische Studien beschreiben. Sie folgen der Struktur empirischer Originalarbeiten, sind jedoch in ihrer Komplexität reduziert. Die Teilnehmenden haben die Aufgabe, die Evidenzquellen zu bewerten und anschließend auf dieser Basis ein schriftliches Statement zum Ausgangsproblem abzugeben (ca. ½-1 Seite). Die Bearbeitungsdauer des Tests beträgt etwa zwei Stunden.

Zur Auswertung der Videos und der Testantworten wurde ein Kategorienschema mit Indikatoren entwickelt, das die Argumentationsqualität mit einer Kombination aus hoch- und niedriginferenten Ratings abbildet (Tabelle 1). Dabei wurde die Argumentationsqualität im ersten Kodierdurchlauf pro Messgelegenheit mit einem hochinferenten Globalrating einge-

Tabelle 1: Übersicht zu den Qualitätsindikatoren der argumentativen Nutzung von Evidenz

Indikator	Stufen/Kategorien	Quellen
Globalrating	1 = Der Studierende gibt durchgängig keine Begründungen und führt keine Evidenz auf 2 = Der Studierende zitiert irrelevante Evidenz oder macht inhaltliche Fehler 3 = Der Studierende begründet teilweise evidenzbasiert, aber nicht in Bezug auf alle Elemente 4 = Der Studierende begründet durchgängig evidenzbasiert und verwendet dabei relevante und angemessene Evidenz 5 = Der Studierende begründet durchgängig evidenzbasiert und hinterfragt die Evidenz kritisch	Kuhn & Crowell (2011), Wilson (2005)
Beleg von Argumenten mit Evidenz	<ul style="list-style-type: none"> <li>• Ohne Beleg</li> <li>• Mit Beleg</li> <li>• Mit Beleg und kritischer Bewertung</li> </ul>	Kuhn & Crowell (2011), Toulmin (1958)
Typ der Evidenznutzung	<ul style="list-style-type: none"> <li>• Keine Evidenz</li> <li>• Irrelevant (benutzt irrelevante Evidenz)</li> <li>• Beschreibend (Nacherzählen des Inhalts)</li> <li>• Single-Warrant (enthält eine wissenschaftliche Annahme über die Evidenzquelle)</li> <li>• Multiple-Warrant (enthält mehr als eine wissenschaftliche Annahme über die Evidenzquelle)</li> </ul>	Bell (2010), Sandoval & Reiser (2004)
Evidenzquelle	<ul style="list-style-type: none"> <li>• Keine</li> <li>• Persönliche Erfahrung (zitiert Situationen aus dem Alltag)</li> <li>• Individuell (benutzt eigene Evidenzquelle)</li> <li>• Instruiert (benutzt die ihm vorgelegte Evidenzquelle)</li> <li>• Beides (benutzt vorgegebene und eigene Evidenzquellen)</li> </ul>	Bell (2010)
Klarheit der Argumente	1 = sehr unklar – 4 = sehr klar	Bell (2010)

schätzt, das die oben beschriebenen Stufen abbildet (Wilson, 2005). Anschließend wurden weitere Aspekte der Argumentation niedriginferent kodiert (Beleg von Argumenten mit Evidenz, Typ der Evidenznutzung, Evidenzquelle) und die Klarheit der Argumente bewertet. Die Daten der Gruppendiskussionen wurden auf der Individualebene in mehreren Durchläufen direkt anhand der Videos kodiert. Den Ratern lagen dabei neben den Videos das Kategoriensystem und detaillierte Instruktionen vor.

Alle Kodierungen wurden von zwei unabhängigen, trainierten Ratern vorgenommen, die eine hohe Interrater-Reliabilität erzielten. Über alle Items hinweg zeigten sich hohe Intraklassenkorrelationen (McGraw & Wong, 1996) für die Auswertung der Videos, ICC (C, 2) = .93, und der Texte, ICC (C, 2) = .95. Die ICCs für die einzelnen Indikatoren erreichten  $.75 < .95$  bei den Videos sowie  $.90 < .97$  bei den Texten. Auf dieser Basis wurden die Werte über beide Rater aggregiert und für die weiteren Analysen verwendet. Die Reliabilität des Globalratings lag für die über beide Videos gemittelten Werte bei Cronbachs  $\alpha = .70$  und für die über beide Texte gemittelten Werte bei  $\alpha = .92$ .

*Hintergrundvariablen.* Im Eingangsfragebogen wurden relevante Hintergrundvariablen zur Soziodemographie (Alter, Geschlecht), zum Studium (Studiensemester, Fächer) und zur Häufigkeit des Lesens wissenschaftlicher Literatur erfasst.

### 3.5 Analysen

Zur Analyse von Gruppenunterschieden wurden aufgrund der kleinen Stichprobe Mann-Whitney U-Tests verwendet (einseitig: TG > KG). Für die Bestimmung der Konsistenz der Argumentationsqualität wurden Konsistenz-Intraklassenkorrelationen berechnet.

## 4 Ergebnisse

### 4.1 Fragestellungen 1 und 2: Argumentationsqualität und Unterschiede zwischen Trainings- und Kontrollgruppe

*Globale Argumentationsqualität.* In Abbildung 2 sind getrennt nach TG und KG Boxplots zum Globalrating der Argumentationsqualität für die einzelnen Messgelegenheiten (2 Gruppendiskussionen, 2 textbasierte Aufgaben) sowie aggregiert nach Gruppendiskussionen und Textaufgaben abgebildet. Die TG liegt durchschnittlich in allen Fällen zwischen Stufe 3 (*Der Studierende begründet teilweise evidenzbasiert, aber nicht in Bezug auf alle Elemente*) und Stufe 4 (*Der Studierende begründet durchgängig evidenzbasiert und verwendet dabei relevante und angemessene Evidenz*).

Die Studierenden der KG zeigten in beiden Textaufgaben eine niedrigere Argumentationsqualität als die TG (Stufe 2: *Der Student zitiert irrelevante Evidenz oder macht inhaltliche Fehler*). Vergleicht man die über beide Textaufgaben gemittelte Argumentationsqualität zwischen den Gruppen, so zeigt sich ein statistisch signifikanter Unterschied zugunsten der TG mit einer großen Effektstärke (TG:  $M = 3.43$ ,  $SD = .75$ ; KG:  $M = 2.03$ ,  $SD = 0.89$ ;  $U = 158$ ,  $p < .001$ ,  $d = 1.67$ ). Zur vorläufigen Einschätzung des Facheffektes wurde zusätzlich ein Teilgruppenvergleich nur für die Studierenden mit mindestens einem MINT-Fach durchgeführt. Auch in dieser Teilgruppe zeigt sich ein statistisch signifikanter Unterschied zugun-

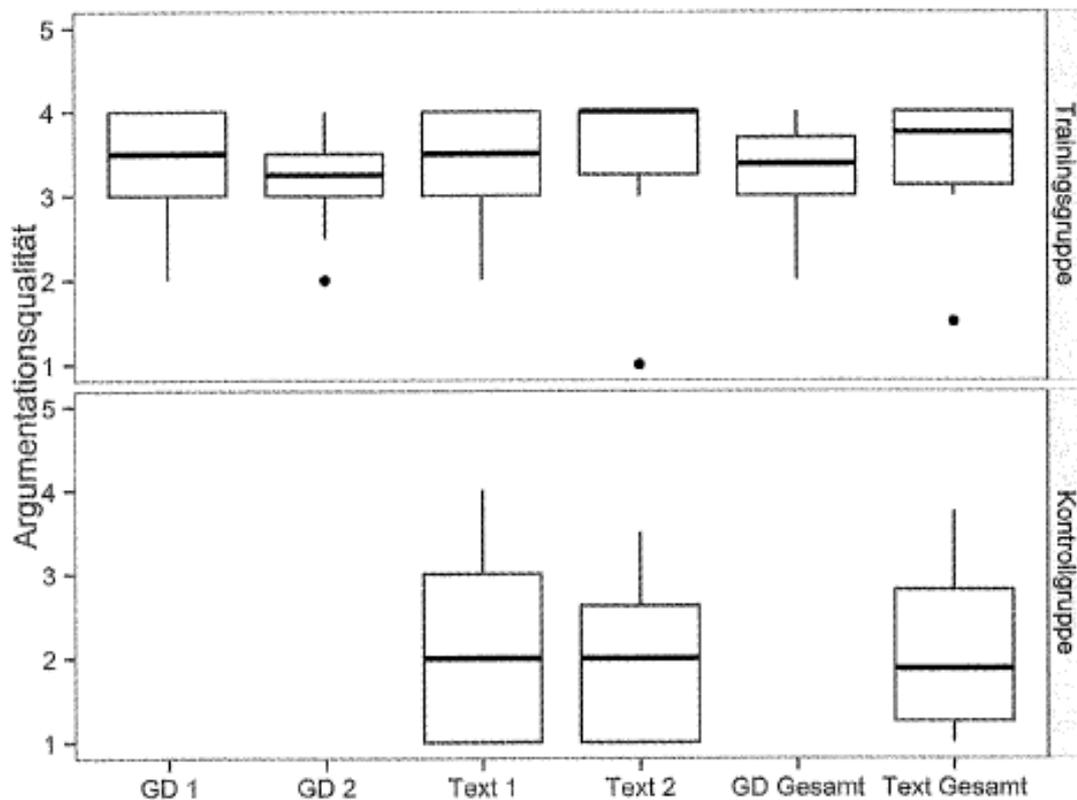


Abbildung 2: Boxplots zum Globalrating der Argumentationsqualität in den einzelnen Gruppendiskussionen (= GD; nur in der Trainingsgruppe) und textbasierte Aufgaben sowie aggregiert für Gruppendiskussionen und Textaufgaben (5 = höchste Stufe) getrennt für Trainings- ( $n = 14$ ) und Kontrollgruppe ( $n = 16$ ).

ten der TG, der kleiner ausfällt als in der Gesamtgruppe, aber dennoch eine große Effektstärke aufweist (TG:  $M = 2.95$ ,  $SD = .91$ ; KG:  $M = 2.03$ ,  $SD = 0.89$ ;  $U = 64$ ,  $p = .025$ ,  $d = 1.03$ ). (Die Teilgruppenanalyse wurde auch für die weiteren Indikatoren durchgeführt. Dabei zeigte sich deskriptiv dasselbe Bild eines im Vergleich zur Gesamtgruppe reduzierten, aber dennoch großen Effekts.)

*Detaillierte Aspekte der Argumentationsqualität.* Tabelle 2 enthält die Ergebnisse zu den weiteren Aspekten der Argumentationsqualität, jeweils aggregiert für die Gruppendiskussionen und Textaufgaben. Die niedriginferenten Kodierungen zum Beleg von Argumenten mit Evidenz dienen zunächst der Validierung des hochinferenten Globalratings. Sie zeigen, dass die Teilnehmenden der TG überwiegend evidenzbasiert argumentieren, während dies auf mehr als zwei Drittel der KG nicht zutrifft. Argumente, in denen zusätzlich zur Evidenznutzung auch eine kritische Bewertung zu finden ist, sind in beiden Gruppen vergleichsweise selten.

Bezüglich der Art der Evidenznutzung zeigte die TG während der Gruppendiskussionen noch bei der überwiegenden Zahl der Argumente eine rein nacherzählende Nutzung von Evidenz (*beschreibend*). Im Test wurden dann überwiegend eine oder mehrere Evidenzquellen im Kontext der Begründung für eine Behauptung verwendet (*Single-* bzw. *Multiple-Warrant*).

Die KG rekurrierte in mehr als zwei Drittel der Argumente nicht auf Evidenz, während dies in der TG nur bei knapp einem Viertel der Fall war.

Das Bild zur Art der Evidenzquelle gestaltet sich ähnlich. Während die TG in den Gruppendiskussionen noch überwiegend auf individuelle Evidenzquellen zurückgriff, wurde im Test mehrheitlich auf die instruierten Quellen verwiesen. Knapp ein Viertel der Argumente wurde allerdings nicht mit Evidenz untermauert. Letzteres traf in der KG für etwa zwei Drit-

Tabelle 2: Aspekte der Argumentationsqualität

Variable	Trainingsgruppe				Kontrollgruppe		Texte: TG vs. KG <sup>2</sup>
	Gruppen- diskussionen		Texte		Texte		
	M <sup>a</sup>	SD	M <sup>a</sup>	SD	M <sup>a</sup>	SD	
<b>Beleg von Argumenten mit Evidenz</b>							
Ohne Evidenz	15.5 %	16.5	22.3 %	26.1	69.6 %	26.5	$U = 18$ $p < .001$
Mit Evidenz	61.4 %	25.3	61.5 %	20.4	20.4 %	24.6	$U = 158.5$ $p < .001$
Mit Evidenz und kritischer Bewertung	23.0 %	18.1	16.3 %	17.8	10.0 %	15.9	$U = 104.5$ $p = .21$
<b>Typ der Evidenznutzung</b>							
Keine	3.3 %	8.6	22.7 %	36.2	67.3 %	33.2	$U = 32.5$ $p = .003$
Irrelevant	15.6 %	24.9	0.0 %	0.0	2.2 %	5.2	$U = 71.5$ $p = .21$
Beschreibend	73.4 %	29.4	1.0 %	3.4	0.0 %	0.0	$U = 96$ $p = .36$
Single-Warrant	7.8 %	8.3	63.8 %	38.4	27.7 %	32.5	$U = 132$ $p = .01$
Multiple-Warrant	0.0 %	0.0	12.6 %	22.4	2.8 %	6.1	$U = 107.5$ $p = .17$
<b>Evidenzquelle</b>							
Keine	22.2 %	26.0	22.6 %	39.0	67.3 %	33.2	$U = 36$ $p = .005$
Persönlich	1.4 %	4.5	0.0 %	0.0	2.3 %	7.0	$U = 77$ $p = .31$
Individuell	69.9 %	26.0	0.0 %	0.0	1.0 %	2.89	$U = 77$ $p = .305$
Instruiert	0.9 %	2.6	77.4 %	39.0	29.4 %	33.3	$U = 141$ $p = .004$
Instruiert und individuell	5.4 %	12.1	0.0 %	0.0	0.0 %	0.0	—
Klarheit <sup>3</sup>	3.38	0.47	3.16	0.94	1.73	0.79	$U = 153$ $p < .001$

Anmerkungen: TG = Trainingsgruppe, KG = Kontrollgruppe.

<sup>1</sup> Über zwei Rater und je zwei Gruppendiskussionen bzw. Textaufgaben gemittelt.

<sup>2</sup> Einseitiger Mann-Whitney-Test.

<sup>3</sup> 1 = sehr unklar, 4 = sehr klar.

tel der Argumente zu, während in knapp 30 % der Argumente die vorgegebenen Evidenzquellen benutzt wurden.

Die Klarheit der Argumente wurde in der TG sowohl bei den Gruppendiskussionen als auch bei den Texten als überwiegend hoch bis sehr hoch eingeschätzt. Bei den Texten zeigte sich ein statistisch signifikanter Unterschied zur KG mit einer großen Effektstärke ( $d = 1.68$ ).

*Qualitative Charakterisierung der Argumente.* Um die Unterschiede in der Argumentation zwischen TG und KG nochmals anhand der Originaldaten zu illustrieren, sind in Tabelle 3 zwei typische Argumente aus den schriftlichen Statements sowie ihre Bewertung aufgeführt. Die Person aus der TG nimmt darin Bezug auf eine instruierte Evidenzquelle (*Single Warrant*) und bewertet diese kritisch. Dagegen argumentiert die Person aus der Kontrollgruppe auf Basis persönlicher Erfahrung und bezieht sich nicht auf die vorgegebenen Evidenzquellen.

#### 4.2 Fragestellung 3: Konsistenz der Argumentationsqualität über die Settings und Problemkontexte

Die Auswertungen zur Konsistenz der globalen Argumentationsqualität zeigen eine relativ hohe Übereinstimmung über die vier Messgelegenheiten hinweg, ICC (C, 4) = .71. Studierende, deren Argumentation in den beiden Gruppendiskussionen eine hohe Argumentationsqualität aufweist, argumentieren demzufolge tendenziell auch in den schriftlichen Statements besser. Dasselbe gilt entsprechend auch zwischen den Gruppendiskussionen mit gesundheits- vs. schulbezogenem Problemkontext, ICC (C, 2) = .70.

Tabelle 3: Beispielhafte Argumentationen aus dem Test und ihre Bewertung

Zitat	Bewertung
<i>Trainingsgruppe</i>	
„Roth (2012) zeigte, dass eine zusätzliche Unterrichtsstunde pro Woche denselben positiven Effekt auf die Schulleistung hat wie Hausaufgaben. Ersteres hat aber den Vorteil, dass das Belastungserleben der Schüler wesentlich geringer ist. Einschränkend ist jedoch zu bemerken, dass hier nur Hausaufgaben im Fach Mathematik untersucht wurden. Es bleibt fraglich, inwieweit die Ergebnisse auf andere Fächer übertragbar sind.“	Globalrating Stufe 4; nutzt die vorgegebene Evidenzquelle
<i>Kontrollgruppe</i>	
„In meinen Augen wird oft vergessen, dass Hausaufgaben ein Hilfsmittel sind, dass ohne eigenem Interesse des Schülers nur zu einem Leistungsdruck führen. Natürlich kann diese Aussage nicht so stark verallgemeinert werden, da es immer Lernende gibt die in Herausforderungen Probleme sehen, wohingegen bei anderen Schülern sich diese Perspektive genau umdreht. Aber wenn ich hier an meine eigene Schulzeit zurückdenke, dann ist Hausaufgabe für mich nur immer dann förderlich gewesen, wenn ich die behandelte Materie schon verstanden habe, oder aber mein eigenes Interesse am Thema so groß war, dass ich mich selbst nochmal damit vertieft auseinandergesetzt habe.“	Globalrating Stufe 1; belegt das Argument ausschließlich durch persönliche Erfahrung

Anmerkung: Rechtschreib- und Grammatikfehler im Original.

## 5 Diskussion

Die Kompetenz, wissenschaftliche Evidenz interpretieren und argumentativ als Legitimation für Entscheidungen oder Handlungen nutzen zu können, ist im Kontext evidenzbasierter Praxis hoch relevant (Fischer et al., 2014; Higgins & Green, 2011; Rubin, 2008). Ihr Erwerb wird deshalb in Diskussionen um Standards von Lehrerprofessionalität zunehmend gefordert (EC, 2013; KMK, 2004; Parr & Timperley, 2008). Allerdings ist davon auszugehen, dass sie in der Lehramtsausbildung nicht systematisch aufgebaut wird. Um den anfänglichen Kompetenzerwerb zu unterstützen, entwickelten wir ein flexibel einsetzbares Kurzzeittraining, das auf der Vermittlung von Heuristiken fokussiert. Die vorliegende Studie zielte auf eine erste Überprüfung, wie sich dieses Training auf die Qualität der argumentativen Nutzung von Evidenz auswirkt. Sie erweitert damit den Stand der Forschung zu Voraussetzungen evidenzbasierter Praxis und zu effektiven Designprinzipien von Fördermaßnahmen zum evidenzbasierten Argumentieren bei Lehramtsstudierenden (Jordanou & Constantinou, 2014; Parr & Timperley, 2008; Zohar, 2008).

Die Ergebnisse zur kriterialen Einschätzung der Argumentationsqualität (Fragestellung 1) zeigten bei der TG eine hohe Argumentationsqualität in allen untersuchten Bereichen. Dabei bestätigten die niedriginferenten Kodierungen zum Beleg von Argumenten mit Evidenz das Bild des hochinferenten Globalratings. Die TG begründete sowohl in den Gruppendiskussionen als auch im schriftlichen Test ihre Argumente mehrheitlich mit relevanter und akkurater Evidenz. Allerdings erreichte keiner der Studierenden die höchste Stufe des verwendeten Globalratings, auf der *durchgängig* Evidenz herangezogen und zusätzlich kritisch hinterfragt wird (Bell, 2010; Kuhn & Crowell, 2011; Sandoval & Reiser, 2004; Wilson, 2005). Die niedriginferenten Kodierungen ergänzten hierzu, dass kritische Bewertungen durchaus zu finden waren, allerdings vergleichsweise selten. Hier zeigen sich offenbar Grenzen der Kurzzeitintervention, die sicherlich nur eine Starthilfe bei der Kompetenzentwicklung leisten konnte. Um nachhaltigere und weitreichendere Lernprozesse in Gang zu setzen, sind vermutlich wiederholte Lerngelegenheiten erforderlich (Desimone, 2009).

Im Vergleich der schriftlichen Testergebnisse zwischen den experimentellen Gruppen (Fragestellung 2) konnten Unterschiede in der Evidenznutzung mit substantiellen Effektstärken nachgewiesen werden. Die Studierenden der KG lagen im Globalrating etwa zwei Rating-Stufen unter der TG und belegten ihre Argumente (bei gleicher Aufforderung im Test) seltener mit wissenschaftlicher Evidenz. Stattdessen griffen sie häufiger als die TG auf persönliche Erfahrung als Legitimationsgrundlage zurück. Dies belegt die Einschätzung, dass normale Lehrpläne die avisierten Kompetenzen eher nicht fördern und bestärkt frühere Befunde, dass (angehende) Lehrkräfte Schwierigkeiten mit evidenzbasierter Argumentation haben (Zemal-Saul et al., 2003; Zohar, 2008). Bei der Interpretation dieser Ergebnisse ist allerdings die erwähnte unglückliche Fächerkombination der experimentellen Gruppen zu berücksichtigen, aufgrund derer Facheffekte nur eingeschränkt analysiert werden konnten. Diesbezügliche Subgruppenanalysen für Studierende mit mindestens einem MINT-Fach zeigten ebenfalls einen großen, im Vergleich zur Gesamtgruppe aber reduzierten Trainingseffekt. Um dies genauer zu untersuchen, sollte die Fächerkombination in zukünftigen Studien systematisch variiert werden.

Bezüglich Fragestellung 3 weisen die Befunde auf eine konsistente Argumentationsqualität in den Gruppendiskussionen und den schriftlichen Statements ebenso wie im Übergang von alltags- zu gesundheitsbezogenen Problemstellungen hin. Dies zeigt, dass die Studierenden die trainierten Heuristiken zwischen den verschiedenen Situationen transferieren konnten.

Diese ersten Erkenntnisse zur Trainingswirksamkeit sind insgesamt erfreulich und konsistent mit früheren Forschungsarbeiten zur Wirksamkeit der diskutierten Designprinzipien der Situierung in Problemkontexte, kooperativ-dialogischen Lernens und der kognitiven Modellierung für die Förderung argumentativer Kompetenzen (Fischer et al., 2014; Iordanou & Constantinou, 2014; Kopp & Mandl, 2011; Kuhn, 2010; Osbourne, 2010; Schworm & Renkl, 2007). Sie weisen auch auf die Wirksamkeit der vermittelten Heuristiken hin.

Bei der Interpretation der Ergebnisse sollten aber neben den bereits diskutierten Einschränkungen weitere Limitationen unserer Studie berücksichtigt werden. Dies betrifft zunächst die kleine Stichprobe. Hier stellte sich zwar aufgrund der großen beobachteten Effekte kein Problem mit der statistischen Power. Allerdings konnten mit dieser Stichprobengröße weder Mehrebenenanalysen für die Daten aus den Gruppendiskussionen durchgeführt werden, noch war es sinnvoll das Alpha-Niveau für die Signifikanztests zu adjustieren, weil dies eine unverhältnismäßige Reduzierung der Power bedeutet hätte. Deshalb wären Replikationsstudien mit größeren Stichproben wünschenswert. Dabei sollten relevante Kovariaten auf Individual- und Kontextebene systematischer einbezogen werden. Zudem wäre in nachfolgenden Studien eine randomisierte Zuweisung wichtig, um die interne Validität zu erhöhen. Nach unserer Erfahrung gestaltet sich dies in der praktischen Umsetzung aber aufgrund der komplexen Organisation des Lehramtsstudiums extrem schwierig. Durch die Vielzahl an Veranstaltungen in verschiedenen Fächern (ggf. an verschiedenen Standorten) sind gemeinsame freie Zeitslots nur sehr mühsam zu finden. Schließlich ist auf inhaltlicher Ebene nochmals zu betonen, dass wir lediglich Teilkompetenzen des kompetenten Umgangs mit Evidenz untersuchten. So fokussierten wir auf die *Rezeption* von Evidenz und klammerten im Vergleich zu Handlungsmodellen evidenzbasierter Praxis (Rubin, 2008) Schritte der Definition einer Frage und der Informationssuche, aber auch des Transfers in die Handlungsumsetzung aus. Ebenso konnten auch die komplexen kognitiven Prozesse der Integration multipler Evidenzquellen (Stadtler, Bromme & Rouet, 2014) nicht berücksichtigt werden.

Aus praktischer Sicht stellt sich abschließend die Frage, wie basale Kompetenzen zur Nutzung von Evidenz und positive Einstellung gegenüber evidenzbasierter Praxis (EC, 2013; KMK, 2004; Parr & Timperley, 2008) weiter gefördert werden können. Wichtige Schritte dazu sind aus unserer Sicht, erstens den *Zugriff* auf wissenschaftliche Befunde zu erleichtern. Eine leichtere Erreichbarkeit scheint mit einer erhöhten Akzeptanz von Forschung und evidenzbasierter Praxis einherzugehen (Bauer, Sälzer & Prenzel, 2015). Zweitens könnte die praxisbezogene *Aufbereitung* und Synthese von Forschungsergebnissen verständnisbezogene Barrieren reduzieren. Zu beiden Zielen kann die Etablierung von *Clearing Houses*, wie sie in der Medizin existieren, einen Beitrag leisten (im Bildungsbereich z. B. *IES What Works Clearinghouse*). Schließlich wäre weiter zu untersuchen, welchen Beitrag eine forschungsbasierte Lehrerbildung (Niemi, 2008) leisten kann, die sowohl im Studium als auch im Referendariat Prinzipien evidenzbasierten Handelns aufgreift.

## Literatur

- Abelson, R. P. (1995). *Statistics as principled argument*. Mahwah, NJ: LEA.
- Bauer, J., Diercks, U., Rösler, L., Möller, J. & Prenzel, M. (2012). Lehramtsausbildung in Deutschland. *Unterrichtswissenschaft*, 40, 101-120.
- Bauer, J. & Prenzel, M. (2012). European teacher training reforms. *Science*, 336, 1642-1643.
- Bauer, J., Sälzer, C. & Prenzel, M. (2015, April). *Teachers' attitudes towards evidence-based practice*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, USA.

- Bell, P. (2010). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education*, 22, 797-817.
- Booth, W. C., Colomb, G. G. & Williams, J. M. (2008). *The craft of research*. Chicago, IL: University of Chicago Press.
- Bromme, R. (2008). *Wissenschaft und Öffentlichkeit: Das Verständnis fragiler und konfliktgeleiteter wissenschaftlicher Evidenz* [Antrag an die DFG]. Zugriff am 2.2.2011 unter <http://wissenschaftundoeffentlichkeit.de/DFG-SPP1409>
- Bromme, R., Prenzel, M. & Jäger, M. (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. *Zeitschrift für Erziehungswissenschaft, Sonderheft 27*, 3-54.
- Brown, N. J. S., Nagashima, S. O., Fu, A., Timms, M. & Wilson, M. (2010). A framework for analyzing scientific reasoning in assessments. *Educational Assessment*, 15, 142-174.
- Chisholm, R. M. (1980). Evidenz. In J. Speck (Hrsg.), *Handbuch wissenschaftstheoretischer Begriffe* (Bd. 1, S. 196-197). Göttingen: Vandenhoeck.
- Collins, A., Brown, J. S. & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning and instruction* (pp. 453-494). Hillsdale, NJ: LEA.
- Curd, M. & Cover, J. A. (Eds.). (1998). *Philosophy of science*. New York, NY: Norton.
- Darling-Hammond, L. & Bransford, J. (Eds.). (2005). *Preparing teachers for a changing world*. San Francisco, CA: Jossey-Bass.
- Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, 47, 108-121.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development. *Educational Researcher*, 38, 181-199.
- Donner-Banzhoff, N. (2007). Die Evidenzbasierte Medizin: Gescheitert oder einfach nur anders? *Zeitschrift für ärztliche Fortbildung und Qualität im Gesundheitswesen*, 101, 441-444.
- European Commission. (2007). *Towards more knowledge-based policy and practice in education and training*. Zugriff am 16.7.2014 unter <http://www.eipcc.eu/Default.aspx?tabid=2533&language=en-US>
- European Commission. (2013). *Supporting teacher competence development for better learning outcomes*. Zugriff am 9.10.2014 unter [ec.europa.eu/education/policy/school/doc/teachercomp\\_en.pdf](http://ec.europa.eu/education/policy/school/doc/teachercomp_en.pdf)
- Felton, M. & Kuhn, D. (2001). The development of argumentative discourse skill. *Discourse Processes*, 32, 135-153.
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., ... Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, 3, 1-18.
- Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, 3, 20-29.
- Hargreaves, D. (2000). Teaching as a research-based profession. In B. Moon (Ed.), *Leading professional development in education* (pp. 200-210). London: Routledge.
- Higgins, J. P. T. & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions (version 5.1.0)*. Zugriff am 25.11.2014 unter [www.cochrane-handbook.org](http://www.cochrane-handbook.org)
- Howick, J., Chalmers, I., Glasziou, P., Greenhalgh, T., Heneghan, C., Liberati, A., ... Thornton, H. (2011). *The 2011 Oxford CEBM levels of evidence: Introductory document*. Zugriff am 25.11.2014 unter <http://www.cebm.net/index.aspx?o=5653>
- Iordanou, K. & Constantinou, C. P. (2014). Developing pre-service teachers' evidence-based argumentation skills on socio-scientific issues. *Learning and Instruction*, 34, 42-57.
- Johnston, J. M., Leung, G. M., Fielding, R., Tin, Y. K. T. & Ho, L.-M. (2003). The development and validation of a knowledge, attitude and behaviour questionnaire to assess undergraduate evidencebased practice teaching and learning. *Medical Education*, 37, 992-1000.
- Khan, K. S. & Coomarasamy, A. (2006). A hierarchy of effective teaching and learning to acquire competence in evidenced-based medicine. *BMC Medical Education*, 6, 59.
- Kollar, I., Fischer, F. & Hesse, F. W. (2006). Collaboration scripts – A conceptual analysis. *Educational Psychology Review*, 18, 159-185.
- Kopp, B. & Mandl, H. (2011). Fostering argument justification using collaboration scripts and content schemes. *Learning and Instruction*, 21, 636-649.
- Kühle, B. & Peek, R. (2007). Lernstandserhebungen in Nordrhein-Westfalen. Evaluationsbefunde zur Rezeption und zum Umgang mit Ergebnisrückmeldungen in Schulen. *Empirische Pädagogik*, 21, 428-447.
- Kuhn, D. (1991). *The skills of argument*. New York, NY: Cambridge University Press.
- Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, 94, 810-824.
- Kuhn, D. & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science*, 22, 545-552.
- Kultusministerkonferenz. (2004). *Gemeinsamer Rahmen der Länder für die frühe Bildung in Kindertageseinrichtungen (Beschluss der Jugendministerkonferenz vom 13./14.05.2004/Beschluss der Kultusministerkonferenz vom 03./04.06.2004)*. Berlin: KMK.
- Lohmann, V., Seidel, V. & Terhart, E. (2011). Bildungswissenschaften in der universitären Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, 4, 271-302.
- Mühler, C. & Stern, E. (2010). Transfer. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 859-869). Weinheim: Beltz.
- McGraw, K. O. & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30-46.

- Niemi, H. (2008). Research-based teacher education for teachers' lifelong learning. *Lifelong Learning in Europe*, 13, 61-69.
- Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, 328, 463-466.
- Patr, J.-M. & Timperley, H. S. (2008). Teachers, schools and using evidence: Considerations of preparedness. *Assessment in Education: Principles, Policy & Practice*, 15, 57-71.
- Petty, G. (2009). *Evidence-based teaching*. Cheltenham: Nelson Thornes.
- Ramos, K. D., Schafer, S. & Tracz, S. M. (2003). Validation of the Fresno test of competence in evidence based medicine. *BMJ*, 326, 319-321.
- Rubin, A. (2008). *Practitioner's guide to using research for evidence-based practice*. Hoboken, NJ: Wiley.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B. & Richardson, W. S. (1996). Evidence-based medicine: What it is and what it is not. *BMJ*, 312, 71-72.
- Sandoval, W. A. & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88, 345-372.
- Schildkamp, K. & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, 26, 482-496.
- Schworm, S. & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, 99, 285-296.
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Shaneyfelt, T., Baum, K. D., Bell, D., Feldstein, D., Houston, T. K., Kaatz, S., ... Green, M. (2006). Instruments for evaluating education in evidence-based practice: A systematic review. *JAMA*, 296, 1116-1127.
- Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31, 15-21.
- Stadtler, M., Bromme, R. & Rouet, J.-F. (2014). „Science meets Reading“: Worin bestehen die Kompetenzen zum Lesen multipler Dokumente zu Wissenschaftsthemen und wie fördert man sie? *Unterrichtswissenschaft*, 42, 55-68.
- Sternberg, R. J. (Hrsg.). (2005). *Reviewing scientific works in psychology*. Washington, DC: APA.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Weber, S. & Achtenhagen, F. (2008). Forschungs- und evidenzbasierte Lehrerbildung. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität* (S. 477-487). Weinheim: Beltz.
- Williams, D. & Coles, L. (2007). Evidence-based practice in teaching: An information perspective. *Journal of Documentation*, 63, 812-835.
- Wilson, M. (2005). *Constructing measures*. New York, NY: Taylor & Francis.
- Wilson, M. & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education*, 13, 181-208.
- Young, J. M. & Ward, J. E. (2001). Evidence-based medicine in general practice: Beliefs and barriers among Australian GPs. *Journal of Evaluation in Clinical Practice*, 7, 201-210.
- Zemba-Saul, C., Munford, D., Crawford, B., Friedrichsen, P. & Land, S. (2003). Scaffolding preservice science teachers' evidence-based arguments during an investigation of natural selection. *Research in Science Education*, 32, 437-463.
- Zohar, A. (2008). Science teacher education and professional development in argumentation. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education* (pp. 245-268). Dordrecht: Springer.

#### Autorenhinweise

Das Projekt *E4teach – Evidence for teachers: Modellierung und Training der Kompetenz angehender und aktiver Lehrkräfte zum Umgang mit Evidenz* wird von der Deutschen Forschungsgemeinschaft (DFG) im Schwerpunktprogramm 1409 *Wissenschaft und Öffentlichkeit* gefördert (Projektleitung: Prof. Dr. Manfred Prenzel und PD Dr. Johannes Bauer; Fördernummer: PR 473/8-1).

#### Anschriften der Autorinnen und Autoren

Sandra Wenglein  
 PD Dr. Johannes Bauer  
 Susanne Heining  
 Prof. Dr. Manfred Prenzel  
 Technische Universität München, TUM School of Education  
 Susanne Klatten-Stiftungslehrstuhl für Empirische Bildungsforschung  
 Arcisstr. 21, D-80333 München, E-Mail: jbauer@tum.de

*Beitrag eingegangen:* 28.11.2014, *revidiert:* 2.3.2015, *angenommen:* 26.3.2015.

## A.2. Manuskript zu Studie 2

Wenglein, S., Bauer, J., & Prenzel, M. (in prep.) How to foster pre-service teachers' argumentation quality in evidence-based contexts.

### How to foster pre-service teachers' argumentation quality in evidence-based contexts

#### Abstract

*Despite policy calls, teachers often don't utilize evidence for their professional reasoning. Therefore, an increased interest in research on training related competences in teacher education has emerged. Thus, our study addressed the following questions: 1) To what degree can a two-day training based on heuristics be effective at helping pre-service teachers critically appraise evidence and apply it argumentatively? Further, is there a difference between pre-service teachers with background in at least one STEM subject versus students without a STEM background? 2) Do additional factors like personal background, epistemic beliefs, beliefs in using evidence, prior knowledge, domain knowledge and tolerance of ambiguity influence training success? To analyze the use of heuristics and argumentation quality, we developed a scoring rubric based on competence models of scientific argumentation as well as review criteria. The results showed that the training had significant large effects on the participants' competences in appraising ( $d = 1.55$ ;  $p < .001$ ) and applying ( $d = 1.36$ ;  $p < .001$ ) evidence argumentatively. Furthermore, we found that background variables did not relate to training success. Thus, this study showed that basic competences in appraising and using evidence can be trained effectively with reasonably short trainings.*

Keywords: evidence-based practice; argumentation; intervention; pre-service teachers

## 1. Introduction

Fostering teachers' competences in critically appraising scientific evidence and in applying it argumentatively is becoming increasingly important (Engelmann, Neuhaus & Fischer, 2016; Fischer et al., 2014;). Even though evidence-based practice for teachers is demanded from several different sources in educational policy (European Commission, 2013; KMK, 2004; Organisation for Economic Co-operation and Development, 2005; U.S. Department of Education, 2007) and research (Bauer & Prenzel, 2012; Bromme, Prenzel & Jäger, 2014; Hefter et al., 2017), teachers still base most of their arguments and decisions on the principle of common practice (Davies, 1999; Hargreaves, 2000). An example for an argument that is based on common practice would be *"We set regular homework tasks, because that is a well-established tradition"* versus an example for an evidence-based argument would be *"We set regular homework tasks, because study xy and study z showed a significant increase of learning"*.

Evidence-based practice refers to the integration of individual expertise with best available research findings (Davies, 1999; Petty, 2009; Sackett, Rosenberg, Gray, Haynes & Richardson 1996; Slavin, 2002). Individual expertise is the knowledge and the practical experience teachers gain over the years, which can ideally be balanced with the latest results of empirical educational studies. Consequently, decisions shouldn't be based solely on subjective convictions, but rather on the critical reflection of the best available evidence (Bromme et al., 2014; Sackett et al., 1996). It should be noted that evidence cannot replace individual expertise. On the contrary, it should serve as a useful addition to tradition and personal expertise, which has to be interpreted critically and with reference to the specific situation at hand.

Furthermore, international educational researchers (Niemi, 2008; Petty, 2009; Parr & Timperley, 2008;) and policies (KMK, 2004; OECD, 2005; U.S. Department of Education, 2007) state that it is necessary for teachers to have positive beliefs regarding appraising and applying evidence argumentatively because the "skill" on its own is not enough without the "will" to apply it (Hefter et al., 2014). Until quite recently, it was largely unclear what beliefs (pre-service) teachers held about using evidence (e.g. Walma van der Molen, van Aalderen-Smeets, 2013). However, it is very likely that there were negative tendencies because most teachers created their lessons based on others' templates or their own experiences (Hargreaves, 2000; Schildkamp & Kuiper, 2010). Furthermore, current initial teacher education (ITE) programs do not seem to effectively prepare future teachers for this task (Bauer et al., 2012; Lohmann, Seidel & Terhart, 2011) even though there is an increasing emphasis regarding the importance of research-based teacher education (Bauer et al., 2012; Weber & Achtenhagen, 2008).

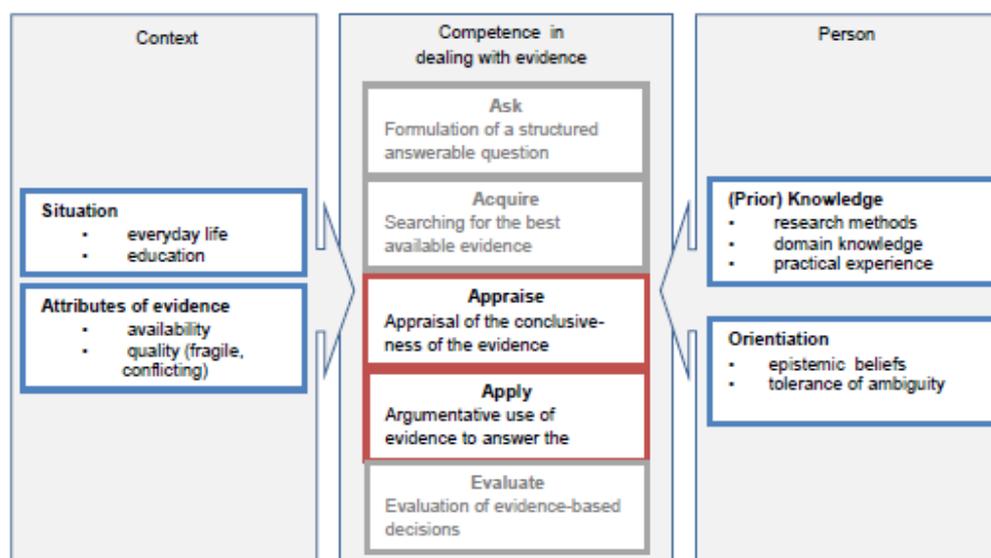
So far, there have been few intervention studies that deal with fostering preservice teachers' competences in using evidence-based practice (e.g. Csanadi, Kollar & Fischer, 2015; Erduran & Garcia-Mila, 2016; Iordanou & Constantinou, 2014). Trainings implemented in the existing research were either very short interventions of about two hours (e.g. Csanadi et al., 2015) or

focused on main topics from science education (e.g. Iordanou et al., 2014). But none of these trainings focused on the use of evidence in the professional educational field. That is why we developed a short training program aimed to foster pre-service teachers' competences in appraising and applying scientific evidence.

### 1.1. Which Competences are Required for Using Evidence in Professional Contexts?

To design a training program to foster evidence-based argumentation, it is a necessary first step to define the needed competences, as well as possible influential factors. Therefore, we designed a model of teachers' competences in dealing with evidence. This model is based on studies that have incorporated evidence-based practices and successful use of evidence (Ramos, Schafer & Tracz, 2003; Rubin, 2008; Shaneyfelt et al., 2006), as well as theories about evidence-based medicine (Finkel, Brown, Gerber & Supino, 2003; Fritsche, Greenhalgh, Falck-Ytter, Neumayer & Kunz, 2002; Malhotra et al., 2015; Sackett et al., 1996).

Figure 1. Competence Model on teachers' competences in dealing with evidence



In our study, we focused on two competence aspects that were part of a more comprehensive action model, as elaborated below: critically *Appraising* evidence and *Applying* evidence argumentatively (see Fig. 1). These competence aspects refer to the epistemic activities of *evidence generation* and *drawing conclusions* that are necessary for scientific reasoning and argumentation (Fischer et al., 2014). Former studies found that these two aspects are the ones students struggle with most (e.g. Fischer et al., 2014).

#### 1.1.1. Appraising evidence.

Appraising evidence, by definition, means "the process of assessing and interpreting evidence by systematically considering its validity, results, and relevance" (Burls, 2009). This is important because the quality of evidence varies and not every study is completely reliable. That is why the field of evidence-based

medicine already has several rating systems and appraisal tools to justify the reliability of evidence (e.g., Cochrane Collaboration, 2017; DNEbM, 2017). The highest quality level in these rating systems is the randomized control trial (RCT). In empirical educational studies, other research designs prevail, as RCTs are not suitable. Thus, we had to design a new rating system based on heuristics, as described below (1.2.1.). Particularly in the field of empirical educational research, the critical appraisal of evidence is challenging due to the fragile and conflicting aspects of research in this domain (Bromme, 2008; Kuhn & Crowell, 2011). Consequently, empirical evidence in this domain is frequently strongly context-bound, of limited generalizability, and, therefore, often leads to conflicting knowledge claims.

### 1.1.2. Applying evidence.

According to Toulmin's (1958) argumentation model, the primary purpose of evidence is to support the *claim*. Therefore, another core aspect of using evidence is being able to support one's claims with pertinent evidence, and putting it into a meaningful context. Especially in the field of medical education, evidence-based practice also has the function of a (legal) justification to support decisions (Chisholm, 1980; Curd & Cover, 1998; Sacket et al., 1996). Furthermore, evidence can serve as a *backing* to support a potential *warrant* (Booth, Colomb & Williams, 2008; Osbourne, 2010; Sandoval & Reiser, 2004).

It is assumed that the more one is able to critically appraise evidence, the better the quality of their arguments will be (Wilson & Sloane, 2000; Kuhn et al., 2011). In the field of science education, Wilson defined five different levels of argumentative quality (Kim, Anthony & Blades, 2014; Wilson, 2005; Wilson et al., 2000). The highest level is that *"the participant provides major objective reasons, supporting each with relevant and accurate evidence that goes beyond in some significant way, such as questioning or justifying the source, validity, and/or quantity of evidence."* (Wilson, 2005) Whereas the lowest level is that *"the participant offers no reasons and no evidence to support choice made"* (see Table 1).

Based on this concept we distinguished five increasing levels of competence in critically appraising evidence. The highest level is that the participant offers correct contents and critically assesses the source of evidence. Furthermore, he/she acknowledges the fragile and contradicting aspects of evidence in the argument. In contrast, the lowest level is that the participant offers irrelevant or wrong contents and doesn't critically assess the source of evidence. The clarity of an argument (Bell, 2010), as well as the position (one- or two-sided) (Kuhn et al., 2011), are both important indicators regarding the quality of evidence-based argumentation. An additional important element is the behavior in cases of disagreement. A low-quality level would be to just show *disagreement* without any evidence (for example "I disagree."). A high-quality level would be to add a Counter-Argument (*Counter-A*) or a well-founded criticism (*Counter-C*) (Kuhn et al., 2011).

### 1.1.2. Contextual and Personal Factors

Our competence model also includes contextual (1) and personal (2) factors, which might have an impact on *competence when dealing with evidence*. These factors need to be taken into account when designing trainings. The contextual factors include the situation (e.g., is it a personal everyday life problem or a professional one?) and attributes of the evidence. In particular, the *attributes of evidence* imply both availability and quality of the evidence (i.e., conflicting and fragile aspects; Bromme, 2008). The personal factors include prior knowledge (such as knowledge on *research methods*, *domain knowledge* and *practical experience*) and personal orientation (*epistemic beliefs* and *tolerance of ambiguity*). Specifically, *knowledge on research* means declarative prior knowledge on quantitative, scientific research methods, which is a key

part of evidence-based practice (Davies, 1999; Ramos et al., 2003; Shaneyfelt et al., 2006). *Domain knowledge* means procedural knowledge in the domain of natural sciences (Fischer et al., 2014). This might play an important role because teachers, including student teachers, have different backgrounds regarding the subjects they are teaching. Hence, their focus may include the natural sciences and mathematics (*STEM*), or on humanities, languages, or arts (*non-STEM*). Students that focus on STEM subjects while earning their degrees should be more familiar with the scientific method and the role of evidence than those who pursue non-STEM subjects. These various areas of concentration could be an influential factor.

Finally, *epistemic beliefs* and *tolerance of ambiguity* may relate to teachers' open-mindedness to deal with evidence. Epistemic beliefs are beliefs about the nature of knowledge and knowing (Hofer & Pintrich, 2002). Connected to that, the *tolerance of ambiguity* (Dalbert & Radant, 2010) may also have an impact on the ability to deal with fragile and conflicting aspects of evidence.

## 1.2. Fostering critical appraisal and the application of evidence in evidence-based argumentation

### 1.2.1. Focus on heuristics

To effectively train the competences mentioned above, it can be helpful to base the training on heuristics (Gigerenzer, 2008; Malhotra et al., 2015; Finkel et al., 2003; Ilic, Nordin, Glasziou, Tilson & Villanueva, 2015; Hutchinson & Gigerenzer, 2005). In this context, heuristics are rules of thumb and other strategies that enable people to make quick decisions despite limited background knowledge and bounded understanding (Bromme et al., 2014). These heuristics may help them generate a solution that is not guaranteed to be optimal, but may be "good enough" for the given purpose. From a curricular perspective, this focus on heuristics is necessary because teacher education curricula are often already full, making it difficult to include additional extensive methodological and statistical courses. Therefore, important questions to consider are: 1) To what degree can basic competences can be effectively built up in a relatively short time frame (Bauer, Diercks, Rösler, Möller & Prenzel 2012; Lohmann et al., 2011) How can respective trainings be designed to fit flexibly into existing curricula? Studies from the field of evidence-based medicine have already demonstrated the benefits of heuristics in several training programs for evidence-based practice; for example, use of hierarchical models with different levels of evidence can enable medical students to critically appraise evidence (Higgins & Green, 2011; Howick et al., 2011).

For the purpose of our study, we used basic review criteria for evaluating the conclusiveness of a study to derive heuristics for appraising evidence (e.g. Is the research question consistent with the empirical design? Are conclusions consistent with the findings (e.g. Abelson, 1995; Sternberg, 2005)? The heuristics for using evidence as *warrant* or *backing* came from research on scientific argumentation (Booth et al., 2008; Khine, 2012; Kuhn, Shaw & Felton, 1997; Wilson, 2005).

The objective of the training was to have an "Icebreaker" that enables student teachers to gain initial contact with evidence-based practice. That is why studies showed, that student teachers often struggle to understand empirical research and have negative beliefs about its relevance for teaching (Hetmanek et al., 2015; Hargreaves, 2000; Kühle & Peek, 2007; Parr et al., 2008; Schildkamp et al., 2010; Williams & Coles, 2008; Zembal-Saul, Munford, Crawford, Friedrichsen & Land, 2003). Similar problems are typical in medical education (Donner-Banzhoff, 2007; Johnston, Leung, Fielding, Tin & Ho 2003; Young & Ward, 2001).

### 1.2.2. Design principles of the training

The design principles of our training were derived from the following sources: 1) existing programs on evidence-based medicine (Finkel, 2003; Khan & Coomarasamy, 2006); 2) programs specifically designed to foster evidence-based argumentation from the field of science education (Fischer et al., 2014; Kopp & Mandl, 2011; Kuhn, 2010; Osbourne, 2010); 3) programs that have implemented evidence-based argumentation into the curriculum for their student teachers (Iordanou et al., 2014; Zembal-Saul et al., 2003; Zohar, 2008).

Specifically, we employed three design principles in order to increase the participants' motivation and the effectiveness and sustainability of the training: 1) *Case-based learning* with authentic case vignettes relevant for professional and personal problems in order to create an atmosphere that supports motivational aspects (e.g. individual interest and positive beliefs) and cognitive processing (Kahn et al., 2006; Fischer et al., 2014). To enhance students' motivation, we started with a problem-oriented beginning. That means we started with a daily life issue before we created the transfer to the professional issue; 2) *Collaborative and dialogic learning* in order to support the development of argumentation competences (Fischer et al., 2014; Iordanou et al., 2014; Kopp et al., 2011; Kuhn, 2010; Osbourne, 2010). Specifically, our aim was to foster critical engagement with cognitive conflicts within the scope of evidence-based practice. Furthermore, we also wanted to foster the participants' social development by providing tasks that required them to adopt different perspectives (Vygotsky, 1978); 3) *Cognitive modeling and scaffolding* to facilitate the learning of complex epistemic processes of scientific thinking (Collins, Brown & Newman, 1989; Iordanou et al., 2014; Kopp et al., 2011; Schworm & Renkl, 2007; Zembal-Saul et al., 2003). To achieve high quality cooperation, we used collaboration scripts (Fischer et al., 2014; Kollar, Fischer & Hesse, 2006). These scripts support group tasks by providing a specific framework and a clear distribution of roles. Thus, it helps to increase the learning effect and to avoid typical problems such as unequal distribution of tasks and diverse prior knowledge (Ertl & Mandl, 2004).

In sum, we used heuristics to foster the competences of applying and appraising evidence and facilitated these competences by providing problem-oriented case-vignettes and tutorial modeling while gradually decreasing support in cooperative settings.

### 1.3. Research Questions

In the present study, we wanted to extend prior research with knowledge about preconditions for evidence-based practice and effective support options for pre-service teachers' evidence-based argumentation (Iordanou et al., 2014; Parr et al., 2008; Zohar, 2008). Thus, the present study addressed the following two research questions:

(1) Does the developed training foster competences of appraising and applying evidence in group discussions during the training and in a subsequent individual written test? In which ways do students benefit from the training compared to a control group? Finally, is there a difference between students with experience in at least one STEM subject versus students without a STEM background?

We hypothesized that the training group would demonstrate higher levels in appraising and applying evidence than the control group that didn't receive special training beyond their regular teacher education program. Furthermore, we predicted that students with a STEM background might find it easier to use empirical educational evidence because they may be more familiar with these types of studies.

(2) Do epistemic beliefs, beliefs in using evidence, prior knowledge, domain knowledge and tolerance of ambiguity, influence training success?

We hypothesized that context as well as individual background may influence training success.

## 2. Methods

### 2.1. Design

The study was designed as a 2 x 2 experimental design (randomizing factor: experimental versus control group; observed factor: STEM versus non-STEM background). We used batch randomization to assign the participants to the experimental ( $N = 84$ ) or control group ( $N = 83$ ). Participants in both conditions received an individual, paper-based questionnaire and a short test on science knowledge. Afterwards, the experimental group started the training program. At the end of the training, they were split up into groups of 3-5 persons to conduct two videotaped group discussions. Subsequently, the participants worked on the assessment instrument. The wait-list control group worked on the assessment instrument first, before they started with the training program and the group discussions. At the end, all participants answered a post questionnaire.

### 2.2. Sample

Data were collected from  $N = 167$  student teachers from 10 different locations in Germany over 21 training dates. Sample recruitment was carried out by promoting the training within several educational and psychological courses for student teachers. Since participation was voluntary, we aimed to mitigate motivational selection bias by offering students 40-Euro vouchers for a big online store as incentive. The participants were 57.1% ( $N = 96$ ) female. Their mean age was  $M = 24.8$  years ( $SD = 5.90$ , range  $18 < 50$ ) and the average semester at the university  $M = 4.68$  ( $SD = 2.67$ ). The majority, 56% ( $N = 95$ ), of participants studied at least one subject from a STEM domain, opposed to the remaining participants with a background in humanities. Prior experience with empirical educational texts was somewhat uncommon, with 54.2% of the sample claiming no or almost no familiarity with these texts. Due to the batch randomization assignment, we controlled and compared the participants' background variables for the control and experimental groups. As a result, we found that the overall differences between these two groups were small and not statistically significant (all values  $p > .14$ ).

### 2.3. Training program

The training started with a problem-based introduction (misuses of statistics in political debates). The purpose of this introduction was to catch the participants' interest and motivate them to learn more about statistics and research methods. We provided them with examples from their daily life because we assumed that they might be more motivated to deal with an issue that was relevant to them before learning to transfer their knowledge to the professional (empirical educational) field. The training focused on the heuristics described above and consisted of three main components: discovering evidence, appraising (critical assessment of) evidence, and applying (using) evidence (see Table 1).

Table 1. Scheme of the training program

Training program: "Competence in using evidence"					
Questionnaire	(1) Discovering evidence: Recognizing the importance of using evidence (day one)	(2) Appraising evidence: Heuristics to critically appraise evidence (day one)		(3) Applying evidence: Heuristics to apply evidence argumentatively (day two)	
	Problem-oriented introduction: Health-related and educational issues	Introduction to core principles of evidence-based practice	Working on cases with fading tutorial support (both domains)	Core principles of evidence-based argumentation. Dealing with fragile and conflicting evidence	Group discussions (both domains)
					Test

During the first segment of the training, (1) *discovering evidence*, we pointed out the importance of using evidence in private as well as in professional tasks by providing them with sources of evidence from two different domains: personal health and educational sciences. Furthermore, we used problem-based introductions to increase motivation and to demonstrate the relevance of evidence-based practice. The second segment, (2) *appraising evidence*, could be considered a methodological crash-course. We taught participants the minimum amount of statistical and methodological knowledge they needed to understand and evaluate an empirical study. Therefore, we used heuristics based on reviewers' criteria. Didactically we used mainly cognitive apprenticeship (Collins et al., 1989) with decreasing tutorial support, to demonstrate to the participants how to critically appraise evidence. On the second day, in the third segment of the training, *applying evidence* (3), we focused on the argumentative use of evidence. After introducing argumentation models based on Toulmin (1978), Kuhn (2011), Bell et al. (2010), and Wilson (2005), we provided several sources of evidence to practice the argumentative use of evidence within two group discussions. The participants were divided into groups of 3-5 people. We provided case vignettes and different real empirical studies that contained conflicting and fragile information. The first case vignette was about a health-related issue (the impact of high salt intake on hypertension). The second case vignette was to transfer their knowledge to their professional field (gender differences in dealing with a computer). The participants were instructed by means of collaboration scripts (Kollar et al., 2006) to read their paper and determine the relevant results, including consideration of whether there was any fragile and/or conflicting information. After that, they were assigned to present the main results and a critical appraisal of their study to their group and discuss how to decide.

All participants completed two group discussions. During the first discussion, they had the chance to practice what they have learned during the training program by discussing a health-related issue (the impact of high salt intake on hypertension). In the second discussion, they had to transfer their knowledge to their professional field (gender differences in dealing with a computer).

#### 2.4. Research Instruments

Questions on *Socio-economic background* included all relevant information on the participant, such as sex, age, first language, domain (STEM background or not), study program, and prior studies (Bauer, Möller & Prenzel, 2010).

Questions on *domain knowledge* tested the participants' understanding of statistical terms such as evidence, experiments, effect sizes, correlations, and so on. The scale has 4 items with responses on a 1-4 Likert scale; the reliability was  $\alpha = .81$ . An example statement is: "How familiar are you with the term 'correlation'? 1= I have never heard about that; 2 = I have heard about that; 3 = I know what it is but I couldn't explain it; 4= I know what it is and I can explain it".

*Epistemic beliefs* were tested with the test instrument of Connotative Aspects of Epistemological Beliefs (Stahl & Bromme, 2007). The items in this measure are subdivided into two factors: *texture* (beliefs about the structure of science, 10 items) and *variability* (beliefs on the stability and falsifiability of science, 7 items). This scale is a semantic differential with 17 items and a 7-point response scale; the reliability was  $\alpha = .78$ .

Items assessing *beliefs in using evidence* were adapted from existing scales (Jette et al., 2003; Johnston et al., 2003; Young et al., 2001). They measure beliefs and attitudes towards empirical educational evidence and whether the participants think that evidence-based practices would be beneficial for their profession. The scale has 4 items with responses on a 1-4 Likert scale; the reliability was  $\alpha = .57$ .

*Questions on reading habits* assesses how often the participants read for pleasure and for their studies (scientific texts). Furthermore, we asked how many scientific papers they read per month. This scale has 6 items with responses on a 1-4 Likert scale; reliability was  $\alpha = .64$ .

*Tolerance of ambiguity* measures how much a participant can tolerate unexpected events and ambiguous situations (Dalbert et al., 2010). This scale has 6 items with responses on a 1-4 Likert scale; reliability was  $\alpha = .72$ .

*Science knowledge*. The prior knowledge test took approximately 15 minutes and was based on questions gauging knowledge about the nature of science (Weinert, 2011).

*Videotaped group discussions*. The two group discussions described in the training program section were videotaped and rated with the same rating instruments that were used for the written statements, as described below.

*Written assessment*. The individual written assessment instrument was analogous to the group discussions. The participants received a problem-based case-vignette and three abridged empirical studies (about one page each) of different quality with conflicting and fragile pieces of information (Heininger et al., 2017). The participants were instructed to read those sources of evidence, critically assess them, and write a short statement (about 500 words).

*Coding Rubric*. We developed a coding rubric to analyze the group discussions as well as the written statements. We used two high-inferent global ratings and seven low-inferent ratings to map their competences in assessing and applying evidence (see Table 2). All codings were rated by two independent and trained raters. To analyze their consistency, we calculated intraclass correlations for the individual indicators (mean ICC = .87; range .70 < .97).

Table 2. Scale intervals of the aspects of argumentative quality

Indicator	Levels/categories	References
High-inferent ratings		
Appraising evidence	<p>Level 5: the participant offers all relevant information and assesses the source of evidence critically. Furthermore he/she realizes the fragile and conflicting content of the domain</p> <p>Level 4: the participant offers all relevant information and assesses the source of evidence critically.</p> <p>Level 3: the participant offers all relevant information and assesses the source of evidence to some extent critically.</p> <p>Level 2: the participant offers all relevant information and doesn't assess the source of evidence critically.</p> <p>Level 1: The participant offers irrelevant or wrong content(s) and doesn't assess the source of evidence critically.</p>	Wilson (2005), Bromme (2010)
Applying evidence	<p>Level 5: the participant provides major objective reasons, supports each with relevant and accurate evidence, and goes beyond in some significant way, such as questioning or justifying the source, validity, and/or quantity of evidence.</p> <p>Level 4: the participant provides major objective reasons and supports each with relevant and accurate evidence.</p> <p>Level 3: the participant provides some objective reasons and some supporting evidence, but at least one reason is missing and/or part of the evidence is incomplete.</p> <p>Level 2: the participant provides only subjective reasons (opinions) for choice and/or uses inaccurate or irrelevant evidence.</p> <p>Level 1: the participant offers no reasons and no evidence to support the choice he/she made.</p>	Proprietary development modeled after Wilson (2005), Kuhn & Crowell (2011)
Low-inferent ratings:		
Frequency of arguments	<p>1 = Claims</p> <p>2 = Claims supported by evidence</p> <p>3 = Claims supported by evidence including a critical appraisal of the evidence</p>	Toulmin (1958), Kuhn & Crowell (2011)
Length	1 = short (max. 1 sentence); 2 = medium (1-2 sentences); 3 = long (3 or more sentences)	Bell (2010)
Warrant	1 = none; 2 = irrelevant; 3 = describing; 4 = single-warrant; 5 = multiple-warrant	Bell (2010), Sandoval & Reiser (2004)
Backing	1 = none; 2 = personal; 3 = individual; 4 = instructed; 5 = individual and instructed	Bell (2010)
Clarity	1 = very unclear; 2 = unclear; 3 = clear; 4 = very clear	Bell (2010)
Position	1 = one-sided; 2 = two-sided	Kuhn & Crowell (2011)
Disagreement	1 = disagreement; 2 = Counter-A; 3 = Counter-C; 4 = Counter-A and Counter-C	Kuhn & Crowell (2011)

## 2.5. Analyses

To answer the research questions, we computed two hierarchical regression models for the high inference ratings. Regarding research question 1, model 1 analyzed the treatment and domain effects along with their interaction while controlling for prior knowledge and reading habits (Research Question 1). To better understand how treatment and control group differed in their argumentation, we conducted further group comparisons (one way ANOVA) for the low inference ratings. Regarding research question 2, regression model 2 additionally controlled for epistemological beliefs and beliefs about the use of evidence. Statistical analyses were done in Mplus 7 (Muthén & Muthén, 1998-2015), with a correction of the standard errors to account for the multilevel structure of the data (i.e., training dates, locations, and groups within each training). An alpha level of .05 was used for all statistical tests.

## 3. Results

### 3.1. Research Question 1: Training and Domain effects

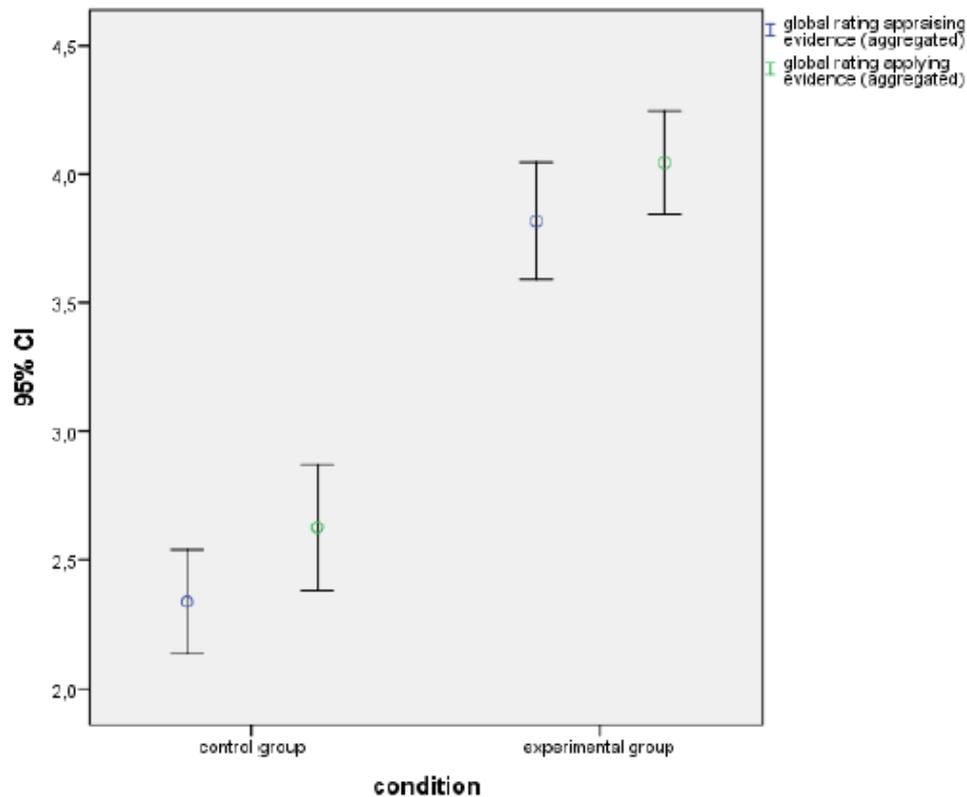
Our first goal was to investigate what argumentation quality the training instigates in group discussions during the training and in a subsequent individual written test. In addition, we aimed to explore whether and how the trained students' argumentation differed from the control group. The high-inference ratings showed that the training had a significant large effect on both competences (see Table 3, model 1). To point out the differences between the groups more explicitly, we also used Cohen's *d* to show the significant effects (appraising evidence:  $d = 1.55$ ;  $p < .001$ ; control group:  $M = 2.34$ ;  $SD = .90$  vs. training group:  $M = 3.82$ ;  $SD = 1.00$ ; applying evidence:  $d = 1.36$ ;  $p < .001$ ; control group:  $M = 2.63$ ;  $SD = 1.10$  vs. experimental group:  $M = 4.05$ ;  $SD = .88$ ; see Figure 2).

Table 3. Influences on the competences of appraising and applying evidence.

	Appraising evidence		Applying evidence					
	Modell I		Modell II		Modell I		Modell II	
	$\beta$	(SE)	$\beta$	(SE)	$\beta$	(SE)	$\beta$	(SE)
condition	<b>.56<sup>†</sup></b>	(.23)	<b>.57</b>	(.23)	<b>.51</b>	(.24)	<b>.51</b>	(.24)
STEM	.04	(.22)	.06	(.23)	-.02	(.23)	-.02	(.24)
Cond. x STEM	.08	(.31)	.08	(.31)	.11	(.33)	.12	(.33)
science	.01	(.02)	-.03	(.02)	.01	(.02)	-.02	(.02)
knowledge								
reading habits	-.02	(.15)	-.05	(.16)	-.07	(.15)	-.10	(.16)
CAEB texture			-.01	(.14)			-.03	(.14)
CAEB			.01	(.12)			.09	(.12)
variability								
tolerance of			-.03	(.13)			-.09	(.14)
ambiguity								
beliefs in using			.11	(.18)			.07	(.19)
evidence								
R <sup>2</sup>	.38		.40		.34		.36	

<sup>†</sup> Significant effects in bold

Figure 2. Error bar chart: high-inference ratings on appraising and applying evidence for both written statements (aggregated for both statements).



Beyond that, the low-inference ratings on the written statements also showed significant differences between the experimental and control groups (see Table 4). The participants in the experimental group had a higher number of arguments that were based on evidence, and a higher number of evidence-based arguments that included a critical appraisal. The experimental group participants also used fewer irrelevant sources of evidence, and they used their source of evidence more often with a single warrant. Some participants in the control group stated multiple warrants, while the experimental group participants did not. The clarity of the trained students' arguments was significantly higher. They also argued more from a dual perspective (*two-sided*). Moreover, the experimental group participants used bare disagreement (without any counter argument or critique) less than control group participants.

Table 4. Results of the low-inference ratings

Variable	Experimental group <i>M</i>	Control group <i>M</i>	Exp vs. ctrl. <i>p</i> <sup>2</sup>
<b>Arguments</b>			
without evidence	28.41%	50.83%	<.001
with evidence	49.24%	34.63%	.001
with evidence and critical appraisal	22.36%	14.54%	=.342
<b>Type of evidence</b>			
none	16.45%	35.75%	=.001
irrelevant	4.45%	15.05%	
describing	7.15%	3.80%	
single warrant	71.95%	42.40%	
multiple warrant	0.00%	3.00%	
<b>Source of evidence</b>			
none	16.45%	36.60%	=.001
personal	4.80%	12.75%	
individual	6.80%	4.60%	
instructed	71.95%	46.10%	
<b>Length</b> <sup>3</sup>	2.01	2.14	=.329
<b>Clarity</b> <sup>4</sup>	3.32	2.51	<.001
<b>Position</b>			
one-sided	15.15%	28.70%	<.001
two-sided	84.85%	71.35%	
<b>Disagreement</b>			
disagree	4.15%	8.90%	<.001
Counter-A	21.65%	35.40%	
Counter-C	68.45%	53.45%	
Counter-A and Counter-C	5.75%	2.25%	<.001

### 3.2. Research Question 2: The Role of Epistemological Beliefs and Beliefs about Evidence

Regression model 2 showed (see table 3, model II) that the treatment effect remained significant and large when controlling for the participants' beliefs and background variables. Contrary to our expectation, the domain (STEM vs. non-STEM subject) had no significant influence on the test results, neither in the group

<sup>2</sup> Significance of the scale

<sup>3</sup> 1 = short, 3 = long

<sup>4</sup> 1 = very unclear, 4 = very clear

discussions nor in the statements. Moreover, none of the other investigated variables had an impact on the test results beyond the training effect.

#### 4. Discussion

Evidence-based argumentation and the ability to critically assess evidence are core competences that need to be central in teacher education (e.g. EC, 2013; KMK, 2004/2014; Parr et al., 2008). However, in-service teachers rarely use evidence to justify their decisions (William et al., 2007). While educational policy and researchers emphasize the importance of more evidence-based practices in education (Bauer et al., 2012; EC, 2013), including teachers' work, current ITE programs do not seem to effectively prepare future teachers for this task. Furthermore, many teachers and student teachers have negative beliefs about using evidence in their daily profession (Hetmanek et al., 2015; Hargreaves, 2000; Kühle et al., 2007; Parr et al., 2008; Schildkamp et al., 2010; Williams et al., 2008; Zembal-Saul et al., 2003). Hence, there are many barriers and challenges for evidence-based practice in teacher education. In our study, we focused on training heuristics to enable the participants to develop solutions that were valid and efficient in regards to time and prior knowledge. Our findings demonstrated that by the use of heuristics, student teachers' competences on appraising and applying evidence can be effectively trained.

The large effect of the training (Research Question 1) showed that the competences of appraising and applying evidence could be fostered using a short time intervention based on heuristics. Even though the participants in the experimental group had a much higher competence-level compared to the students in the control group, very few students reached the highest level in both competences. The highest level requires more than just using the learned techniques, it requires use in a self-reflective and elaborative way. This likely reflects a limitation of a short, two-day intervention. Presumably, it would take more practice and experience to reach these highest levels in competences.

The results also demonstrated that students' domain, as well as their beliefs and background variables, did not contribute to explaining their performance beyond the effect of the training (Research Question 2). Though this finding was unexpected and warrants further investigation, it may suggest that training basic competences for using evidence can be done effectively for learners with different backgrounds.

These results are encouraging, however, there are several limitations that should be taken into account. First, the individuals participated in the training voluntarily and thus, may constitute a positive selection, with high interest and motivation. Consequently, it would be interesting to replicate this study with samples from universities where the training is a mandatory part of the curriculum for all students. Second, an additional posttest would be helpful to check the sustainability of training effects. Third, for the reasons discussed above, we focused on only two relevant competence aspects of dealing with evidence (cf. Fig 1). Therefore, as a further step, it would be interesting include other aspects of this competence, such as searching and selecting relevant evidence.

Despite the discussed limitations, we believe that our study advances current research by showing that even small scale trainings focusing on heuristics can be quite effective in preparing basic competences for using evidence in teacher education. Even though many challenges and complexities in moving towards an evidence-based practice in teaching still exist, our results indicate that it may not be necessary for teachers to have profound expert knowledge in statistics and research methods to be able to critically appraise empirical evidence. If teacher education adopts such a small scale intervention to foster both the skill and will (Hefter et al., 2014) to apply evidence argumentatively, evidence-based practice might come a bit more into reach.

#### Keypoints

- Despite policy calls, teachers don't utilize evidence for their professional reasoning.
- Studies from the field of evidence-based medicine showed the benefits of heuristics for practitioners.
- We developed and implemented a heuristics-based training for student teachers.
- The training had large positive effects on the necessary core competencies of appraising and applying evidence.
- This program can enable teachers to make quick but also evidence-based decisions using heuristics.
- Further, this program contributes to a better understanding of processes in learning evidence-based argumentation.

#### Acknowledgments

This research project was supported by a grant from the German Science Foundation within the framework of the Special Priority Program 1409 (Science and the public) [PR 473(8-1)].

## References

- Abelson, R. P. (1995). *Statistics as principled argument*. Mahwah, NJ: LEA.
- Bauer, J., Möller, J. & Prenzel, M. (2010). PaLea: Panel zum Lehramtsstudium. Professionsbezogene Entwicklung von Lehramtsstudierenden in den neuen Studienstrukturen. In B. Schwarz, P. Nenniger & R. S. Jäger (Hrsg.), *Erziehungswissenschaftliche Forschung – nachhaltige Bildung. Beiträge zur 5. DGfE-Sektionstagung "Empirische Bildungsforschung"/AEPF-KBBB im Frühjahr 2009* (S. 178-183). Landau: VEP.
- Bauer, J., Diercks, U., Rösler, L., Möller, J. & Prenzel, M. (2012). Lehramtsausbildung in Deutschland. *Unterrichtswissenschaft*, 40, 101-120.
- Bauer, J. & Prenzel, M. (2012). European teacher training reforms. *Science*, 336, 1642-1643.
- Bell, P. (2010). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education*, 22, 797-817.
- Booth, W. C., Colomb, G. G. & Williams, J. M. (2008). *The craft of research*. Chicago: University of Chicago Press.
- Bromme, R. (2008). *Wissenschaft und Öffentlichkeit: Das Verständnis fragiler und konfligierender wissenschaftlicher Evidenz* [Antrag an die DFG]. Retrieved from: <http://wissenschaftundoeffentlichkeit.de/DFG-SPP1409>
- Bromme, R., Prenzel, M. & Jäger, M. (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. *Zeitschrift für Erziehungswissenschaft (Sonderheft 27)*, 3-54.
- Burls, A. (2009). What is critical appraisal? *Evidence-based medicine: What is ...? series*. NPR09/1113, 1-8.
- Chisholm, R. M. (1980). Evidenz. In J. Speck (Hrsg.), *Handbuch wissenschaftstheoretischer Begriffe* (Bd. 1, S. 196-197). Göttingen: Vandenhoeck.
- Cochrane Collaboration (2017). GRADE approach to evaluating the quality of evidence: a pathway. Retrieved from: <http://training.cochrane.org/path/grade-approach-evaluating-quality-evidence-pathway>

- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Hrsg.), *Knowing, learning and instruction* (S. 453–494). Hillsdale, NJ: LEA.
- Csanadi, A., Kollar, I., & Fischer, F. (2015). *Teacher students' scientific reasoning: how can social context and internal scripts as predictors inform instructional design?*. Paper presented at the 19th Annual Conference for Junior Researchers of the European Association for Research on Learning and Instruction (JURE), Limassol, Cyprus.
- Curd, M. & Cover, J. A. (Hrsg.). (1998). *Philosophy of science*. New York, NY: Norton.
- Dalbert, C. & Radant, M. (2010). Ungewissheitstoleranz bei Lehrkräften. *Journal für LehrerInnenbildung*, 10, 53-57.
- Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, 47, 108-121.
- Deutsches Netzwerk für evidenzbasierte Medizin (2009). EbM Basics. Retrieved from: <http://www.ebm-netzwerk.de/ebm-events/kalender/ebm2017>
- Donner-Banzhoff, N. (2007). Die Evidenzbasierte Medizin: Gescheitert oder einfach nur anders? *Zeitschrift für ärztliche Fortbildung und Qualität im Gesundheitswesen*, 101, 441-444.
- Engelmann, K., Neuhaus, B. & Fischer, F. (2016). Fostering scientific reasoning in education – meta-analytic evidence from intervention studies. *Educational research and evaluation*.
- Erduran, S. & Garcia-Mila, M. (2015). Epistemic practices and thinking in science: Fostering teachers' development in scientific argumentation.

- Ertl, B., & Mandl, H. (2004). *Kooperationskripts als Lernstrategie*. (Forschungsbericht Nr. 172). München: Ludwig-Maximilians-Universität, Lehrstuhl für empirische Pädagogik und Pädagogische Psychologie.
- European Commission [EC] (2007). *Towards more knowledge-based policy and practice in education and training*. Retrieved from <http://www.eipee.eu/Default.aspx?tabid=2533&language=en-US>
- European Commission [EC] (2013). *Supporting teacher competence development for better learning outcomes*. Retrieved from [ec.europa.eu/education/policy/school/doc/teachercomp\\_en.pdf](http://ec.europa.eu/education/policy/school/doc/teachercomp_en.pdf)
- Finkel, M.; Brown, H.; Gerber, L.; Supino, P. (2003): Teaching evidence-based medicine to medical students. In: *Medical teacher* 25 (2), S. 202–204. DOI: 10.1080/0142159031000092634.
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., ... & Eberle, J. (2014). Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Frontline Learning Research*, 3, 1-18.
- Fritsche, L.; Greenhalgh, T.; Falck-Ytter, Y.; Neumayer, H.-H.; Kunz, 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. In: *British Medical Journal* 325, S. 1338–1341.
- Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, 3, 20-29.
- Hargreaves, D. (2000). Teaching as a research-based profession. In B. Moon (Hrsg.), *Leading professional development in education* (S. 200-210). London: Routledge.

- Hefter, M. H., Berthold, K., Renkl, A., Riess, W., Schmid, S., & Fries, S. (2014). Effects of a training intervention to foster argumentation skills while processing conflicting scientific positions. *Instructional Science*, 42(6), 929-947.
- Hetmanek, A., Wecker, C., Gräsel, C., Kiesewetter, J., Trempler, K. Fischer, M.R., & Fischer, F. (2015). Ressourcen nutzen Lehrkräfte wann? Eine Interviewstudie zur Schnittstelle zwischen Unterrichtsalltag und Wissenschaft, *Unterrichtswissenschaft*, 43 (3).
- Higgins, J. P. T. & Green, S. (Hrsg.). (2011). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 5.1.0). Retrieved from: [www.cochrane-handbook.org](http://www.cochrane-handbook.org)
- Howick, J., Chalmers, I., Glasziou, P., Greenhalgh, T., Heneghan, C., Liberati, A., ... & Thornton, H. (2011). *The 2011 Oxford CEBM Levels of Evidence: Introductory Document*. Retrieved from: <http://www.cebm.net/index.aspx?o=5653>
- Hutchinson, J. C.; Gigerenzer, G. (2005): Simple heuristics and rules of thumb: where psychologists and behavioural biologists might meet. In: *Behavioural processes* 69 (2), S. 97–124. DOI: 10.1016/j.beproc.2005.02.019.
- Ilic, D., Nordin, R., Glasziou, P., Tilson, J., Villanueva, E. (2015): A randomised controlled trial of a blended learning education intervention for teaching evidence-based medicine. In: *BMC medical education* 15, S. 39.
- Iordanou, K. & Constantinou, C. P. (2014). Developing pre-service teachers' evidence-based argumentation skills on socio-scientific issues. *Learning and Instruction*, 34, 42-57.
- Jette D, Bacon K, Batty C, Carlson M, Ferland A, Hemingway RD, et al. Evidence-based practice: beliefs, attitudes, knowledge, and behaviors of physical therapists'. *Physical Therapy* 2003; 83(9):786-805.

- Johnston, J. M., Leung, G. M., Fielding, R., Tin, Y. K. T. & Ho, L.-M. (2003). The development and validation of a knowledge, attitude and behaviour questionnaire to assess undergraduate evidencebased practice teaching and learning. *Medical Education*, 37, 992-1000.
- Khan, K. S. & Coomarasamy, A. (2006). A hierarchy of effective teaching and learning to acquire competence in evidenced-based medicine. *BMC Medical Education*, 6, 59.
- Khine, M. (2012). *Perspectives on scientific argumentation*. Dordrecht: Springer.
- Kim, M., Anthony, R., Blades, D. (2014). Decision making through dialogur: a case study of analyzing preservice teachers' argumentation on socioscientific issues. *Research in Science Education*. Online publication.
- KMK (2004/2014). *Gemeinsamer Rahmen der Länder für die frühe Bildung in Kindertageseinrichtungen*. Beschluss der Kultusministerkonferenz. Berlin: KMK.
- Kollar, I., Fischer, F. & Hesse, F. W. (2006). Collaboration scripts - a conceptual analysis. *Educational Psychology Review*, 18, 159-185.
- Kopp, B. & Mandl, H. (2011). Fostering argument justification using collaboration scripts and content schemes. *Learning and Instruction*, 21, 636-649.
- Kühle, B. & Peek, R. (2007). Lernstandserhebungen in Nordrhein-Westalen. Evaluationsbefunde zur Rezeption und zum Umgang mit Ergebnisrückmeldungen in Schulen. *Empirische Pädagogik*, 21, 428-447.
- Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentive reasoning. *Cognition and Instruction*, 15, 287-315.
- Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, 94, 810-824.

- Kuhn, D. & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science*, 22, 545–552.
- Lohmann, V., Seidel, V. & Terhart, E. (2011). Bildungswissenschaften in der universitären Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, 4, 271–302.
- Malhotra, A.; Maughan, D.; Ansell, J.; Lehman, R.; Henderson, A.; Gray, M. et al. (2015): Choosing Wisely in the UK: the Academy of Medical Royal Colleges' initiative to reduce the harms of too much medicine. In: *BMJ (Clinical research ed.)* 350, S. h2308.
- Muthén, L., & Muthén, B. (1998-2015). Mplus User's Guide (8<sup>th</sup> ed.) Los Angeles, CA: Muthén & Muthén.
- Niemi, H. (2008). Research-based teacher education for teachers' lifelong learning. *Lifelong Learning in Europe*, 13, 61–69.
- OECD (2005). *Teachers Matter. Attracting, developing and retaining effective teachers.* Retrieved from OECD Publishing website: <http://www.oecd.org/edu/school/34990905.pdf>
- Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, 328, 463–466.
- Parr, J.-M. & Timperley, H. S. (2008). Teachers, schools and using evidence: Considerations of preparedness. *Assessment in Education: Principles, Policy & Practice*, 15, 57-71.
- Petty, G. (2009). *Evidence-based teaching*. Cheltenham: Nelson Thornes.
- Hofer, B. & Pintrich, P. (2002). *Personal Epistemology: The Psychology of Beliefs About Knowledge and Knowing*. Mahwah, NJ: Lawrence Erlbaum.
- Ramos, K. D., Schafer, S. & Tracz, S. M. (2003). Validation of the Fresno Test of competence in evidence based medicine. *BMJ*, 326, 319-321.

- Rubin, A. (2008). *Practitioner's guide to using research for evidence-based practice*. Hoboken, N.J.: Wiley.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B. & Richardson, W. S. (1996). Evidence-based medicine: What it is and what it is not. *BMJ*, *312*, 71-72.
- Sandoval, W. A. & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, *88*, 345-372.
- Schildkamp, K. & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, *26*, 482-496.
- Schworm, S. & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, *99*, 285-296.
- Shaneyfelt, T., Baum, K. D., Bell, D., Feldstein, D., Houston, T. K., Kaatz, S. ... & Green, M. (2006). Instruments for evaluating education in evidence-based practice: a systematic review. *JAMA*, *296*, 1116-1127.
- Slavin, R. E. (2002). Evidence-based education policies: transforming educational practice and research. *Educational Researcher*, *31*, 15-21.
- Stahl, E. & Bromme, R. (2007). The CAEB: An instrument for measuring connotative aspects of epistemological beliefs. *Learning and Instruction*, *17*(6), 773-785.
- Sternberg, R. J. (Hrsg.). (2005). *Reviewing scientific works in psychology*. Washington, DC: APA.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- U.S. Department of Education. (2007). *Strategic Plan for Fiscal Years 2007–12*. Washington, D.C.

- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Walma van der Molen, J., van Aalderen-Smeets, S. (2013): Investigating and stimulating primary teachers' attitudes towards science: Summary of a large-scale research project. In: *Frontline Learning Research* 1, S. 1–9.
- Weber, S. & Achtenhagen, F. (2008). Forschungs- und evidenzbasierte Lehrerbildung. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität* (S. 477-487). Weinheim: Beltz.
- Williams, D. & Coles, L. (2007). Evidence-based practice in teaching: an information perspective. *Journal of Documentation*, 63, 812-835.
- Wilson, M. (2005). *Constructing measures*. New York: Taylor & Francis.
- Wilson, M. & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education*, 13, 181-208.
- Wiseman, A. W. (2010). The uses of evidence for educational policymaking: Global contexts and international trends. *Review of Research in Education*, 34, 1-24.
- Young, J. M. & Ward, J. E. (2001). Evidence-based medicine in general practice: beliefs and barriers among Australian GPs. *Journal of Evaluation in Clinical Practice*, 7, 201-210.
- Zemal-Saul, C., Munford, D., Crawford, B., Friedrichsen, P., & Land, S. (2003). Scaffolding preservice science teachers' evidence-based arguments during an investigation of natural selection. *Research in Science Education*, 32, 437-463.
- Zohar, A. (2008). Science teacher education and professional development in argumentation. In S. Erduran & M. P. Jimenez-Aleixandre (Hrsg.), *Argumentation in science education* (S. 245-268). Dordrecht: Springer.

## B. Testinstrumente

### B.1. Feedbackheft

# -Feedbackheft-



## Forschung Verstehen

„Ein Training im Rahmen des DFG-Projektes E4teach –  
Evidence for teachers“



Liebe Lehramtsstudierende,

Vielen Dank, dass Sie an der Pilotierung unseres Trainings teilnehmen.

Da es heute zum ersten Mal stattfindet, wünschen wir uns von Ihnen Rückmeldung zu den einzelnen Bereichen, damit diese gegebenenfalls verbessert und optimal auf die Bedürfnisse der Studierenden angepasst werden können.

Als kleines Dankeschön für Ihre Unterstützung erhalten Sie im Anschluss an das Training einen Amazon-Gutschein im Wert von 15 Euro.

Herzlichen Dank für Ihre Hilfe,

Sandra Wenglein

## 1. Evidenz kennenlernen

Haben Sie das Gefühl nun zu wissen warum es wichtig ist, dass sich zukünftige Lehrerinnen und Lehrer mit Evidenz befassen? Bitte begründen Sie Ihre Antwort.

War die Lerneinheit ansprechend und motivierend gestaltet? Bitte begründen Sie Ihre Antwort.

Waren die Inhalte für Sie verständlich aufbereitet? Bitte begründen Sie Ihre Antwort.

Fanden Sie die gewählten Lern- und Lehrmethoden ansprechend und angemessen? Bitte begründen Sie Ihre Antwort.

Fanden Sie die ausgegebenen Materialien (Handouts, Beispiele, ...) ansprechend und hilfreich?

Was hat Ihnen besonders gut gefallen?

Was hätten Sie sich anders gewünscht?

Gibt es noch etwas, das Sie uns gerne mitteilen würden?

## 2. Evidenz bewerten

Wie schätzen Sie die Teileinheit über die Fragestellung eines Forschungsbefundes ein?

Wie bewerten Sie die Teileinheit über die verschiedenen Forschungsmethoden?

Wie beurteilen Sie die Untereinheit über die Wahl und Größe einer Stichprobe?

Wie fanden Sie die Besprechung der verschiedenen Instrumente zur Gewinnung von Daten?

Wie beurteilen Sie die Untereinheit über den Ergebnisteil eines Forschungsbefundes?

War die Lerneinheit „Evidenz bewerten“ ansprechend und motivierend gestaltet? Bitte begründen Sie Ihre Antwort.

Waren die Inhalte für Sie verständlich aufbereitet? Bitte begründen Sie Ihre Antwort.

Fanden Sie die gewählten Lern- und Lehrmethoden ansprechend und angemessen? Bitte begründen Sie Ihre Antwort.

Fanden Sie die ausgegebenen Materialien (Handouts, Beispiele, ...) ansprechend und hilfreich?

Was hat Ihnen besonders gut gefallen?

Was hätten Sie sich anders gewünscht?

Gibt es noch etwas, das Sie uns gerne mitteilen würden?

***Vielen Dank für Ihre Mitarbeit und bis morgen!***

### 3. Evidenz nutzen

Wie schätzen Sie die Teileinheit über das Argumentieren ein?

Wie fanden Sie die Bearbeitung des Fallbeispiels „Salz“?

Wie beurteilen Sie die Bearbeitung des Fallbeispiels  
„Geschlechterunterschiede“?

Wie bewerten Sie die Bearbeitung des Fallbeispiels „Lehramtsstudium“  
in Einzelarbeit?

Wie beurteilen Sie insgesamt die gewählten Fallbeispiele?

Gibt es noch etwas, das Sie uns gerne mitteilen würden?

***Vielen Dank für Ihre Mitarbeit!***

B.2. Fragebogen (prä und post)

## **-Fragebogen-**



### **Forschung Verstehen**

**Ein Training im Rahmen des DFG-Projektes**

**E4teach – Evidence for teachers“**



Liebe Lehramtsstudierende,

Lehrkräfte sind heute zunehmend mit der Anforderung konfrontiert, Forschungsbefunde aus den Bildungswissenschaften kritisch lesen und interpretieren zu können. Wissenschaftliches Wissen wird auch als „Evidenz“ bezeichnet. Bildungswissenschaftliche Evidenz kann Ihnen helfen informierte Entscheidungen zu treffen und wissenschaftliche Erkenntnisse für ihre Unterrichtspraxis nutzbar zu machen.

Hierfür bieten wir Ihnen ein 2-tägiges Training an.

Das Training ist neben dem Erwerb wichtigen Wissens für die Berufspraxis als Lehrer auch eine gute Vorbereitung auf das erziehungswissenschaftliche Staatsexamen (EWS) und das Referendariat.

Da es sich um eine Pilotierung handelt, wünschen wir uns von Ihnen Feedback und Rückmeldungen zu unserem Training.

Hinweise zum Datenschutz finden Sie auf der nächsten Seite.

Im Voraus sehr herzlichen Dank für Ihre Mitarbeit!

Ihr E4teach-Team

## Hinweise zum Ausfüllen des Fragebogens

Bitte beachten Sie beim Ausfüllen des Fragebogens die folgenden Hinweise:

- Machen Sie bitte **keine Kreuze zwischen die Kästchen!**
- **Sie haben sich beim Ankreuzen vertan?** Kein Problem: Einfach Kästchen ganz ausmalen und ein neues Kreuz setzen.
- Wenn Sie **etwas aufschreiben sollen**, sind diese Stellen mit folgendem Symbol markiert: (✍). Bei diesen Freitextfeldern verwenden Sie bitte Druckbuchstaben!
- Manche Fragen klingen aus methodischen Gründen ähnlich, sie sind jedoch immer ein wenig unterschiedlich. Auch wenn Ihnen der Fragebogen dadurch manchmal etwas lästig vorkommt, bitten wir Sie, jeweils alle Fragen zu beantworten.
- Wenn Sie einzelne Fragen nicht beantworten wollen, können Sie diese überspringen. Wir bitten Sie trotzdem, den Fragebogen **vollständig** zu bearbeiten.

## Hinweise zum Datenschutz

Wir garantieren Ihnen, dass die von erfassten personenbezogenen Daten streng vertraulich behandelt werden. Alle Mitarbeiter/innen des Projekts, die zum Zweck der Auswertung Zugriff auf die erfassten Daten haben, werden auf die Einhaltung der Ihnen hier vorgelegten Datenschutzrichtlinien verpflichtet. Die Daten werden von Mitarbeiter/innen des Projekts nach wissenschaftlichen Kriterien ausgewertet und ggf. wissenschaftlich publiziert. In Publikationen werden die Daten stets in einer Form dargestellt, die keinerlei Rückschlüsse auf Einzelpersonen zulässt.

## Teil 1: Meine Vorerfahrungen im Umgang mit Evidenz

In den Standards zur Lehrerbildung fordert die Kultusministerkonferenz, dass Lehrerinnen und Lehrer bildungswissenschaftliche Forschungsbefunde - u.a. aus der Erziehungswissenschaft und pädagogischen Psychologie - wahrnehmen, interpretieren und als Grundlage für ihr professionelles Handeln nutzen sollen („evidenzbasierte Praxis“). Die folgenden Fragen drehen sich um Ihre Vorerfahrungen und Ansichten zu diesem Bereich.

Bitte kreuzen Sie pro Frage eine Antwortalternative an.

<b>1. Haben Sie Zugang zu folgenden Fachinformationsquellen?</b>	<b>1 nein</b>	<b>2 ja, an meiner Universität</b>	<b>3 ja, anderweitig</b>	<b>4 weiß nicht</b>
Forschungsartikel in wissenschaftlichen Fachzeitschriften (z.B. Zeitschrift für Erziehungswissenschaft, Zeitschrift für Pädagogik, Zeitschrift für Pädagogische Psychologie, Unterrichtswissenschaft, etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Artikel in anwendungsbezogenen pädagogische Fachzeitschriften (z.B. Pädagogik, Praxis der Naturwissenschaften – Physik in der Schule, etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pädagogische/psychologische Datenbanken zur Recherche von Fachartikeln (z.B. FIS-Bildung, ERIC, PSYINDEX, etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>2. Wie häufig nutzen Sie im Allgemeinen folgende Informationsquellen?</b>	<b>1 nie</b>	<b>2 weniger als sechs Mal im Jahr</b>	<b>3 mehr als sechs Mal im Jahr</b>	<b>4 jeden Monat oder öfter</b>
Forschungsartikel in wissenschaftlichen Fachzeitschriften (z.B. Zeitschrift für Erziehungswissenschaft, Zeitschrift für Pädagogik, Zeitschrift für Pädagogische Psychologie, Unterrichtswissenschaft, etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Artikel in anwendungsbezogenen pädagogische Fachzeitschriften (z.B. Pädagogik, Praxis der Naturwissenschaften – Physik in der Schule, etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pädagogische/psychologische Datenbanken zur Recherche von Fachartikeln (z.B. FIS-Bildung, ERIC, PSYINDEX, etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>3. Ich lese...</b>	<b>1 nie oder fast nie</b>	<b>2 ein- bis zweimal im Monat</b>	<b>3 ein- bis zweimal pro Woche</b>	<b>4 jeden Tag oder fast jeden Tag</b>
... zu meinem Vergnügen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... wissenschaftliche Texte	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... bildungswissenschaftliche Literatur (z.B. Zeitschrift für Erziehungswissenschaft, Zeitschrift für Pädagogik, Zeitschrift für Pädagogische Psychologie, Unterrichtswissenschaft, etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

... praxisorientierte pädagogische Literatur (z.B. Pädagogik, Praxis der Naturwissenschaften – Physik in der Schule, etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Zeitungen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Zeitschriften/ Magazine (z.B. Sport, Musik, Mode,...)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>4. Wie viele wissenschaftliche Artikel lesen Sie durchschnittlich pro Monat?</b>	<b>1</b> <input type="checkbox"/> keine	<b>2</b> <input type="checkbox"/> 1-5	<b>3</b> <input type="checkbox"/> 5-10	<b>4</b> <input type="checkbox"/> > 10
---	---	--	---	---

<b>5. Wie gut stimmen die folgenden Aussagen mit Ihrer persönlichen Sichtweise zu einer möglichst evidenzbasierten - im Sinn einer an bildungswissenschaftlichen Forschungsbefunden orientierten - Unterrichtspraxis überein?</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
Ich finde eine stärkere Evidenzbasierung im Bildungsbereich wichtig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die meisten meiner Mitstudierenden finden eine stärkere Evidenzbasierung im Bildungsbereich wichtig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die Unterrichtsqualität könnte erheblich verbessert werden, wenn Lehrerinnen und Lehrer sich stärker an Unterrichtsmethoden orientieren würden, deren Wirksamkeit wissenschaftlich überprüft ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die Forderung nach einer evidenzbasierten Unterrichtspraxis überfordert die meisten Lehrerinnen und Lehrer.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Versuchen Sie sich bitte für die folgende Frage an Ihre Erfahrungen in der Schulpraxis (z.B. Block- bzw. studienbegleitende Praktika) zu erinnern.

<b>6. Was sind aus Ihrer Sicht Hinderungsgründe dafür, erziehungswissenschaftliche Forschungsbefunde als Grundlage für Ihre Unterrichtsgestaltung zu nutzen?</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
Ich habe keinen/nur begrenzten Zugang zu Forschungsergebnissen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich weiß zu wenig darüber, wie ich für mich relevante Forschungsergebnisse finden kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zu vielen mich interessierenden Fragen gibt es kaum relevante Forschungsbefunde.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mir fehlt die Zeit, Forschungsbefunde zu suchen und zu lesen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Es fällt mir schwer, forschungsmethodische oder statistische Informationen zu verstehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die meisten erziehungswissenschaftlichen Forschungsbefunde sind für meine Tätigkeit als Lehrer/in nicht relevant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Erziehungswissenschaftlichen Forschungsbefunde geben in der Regel zu wenig konkrete Handlungsanweisungen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich interessiere mich nicht für	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

erziehungswissenschaftliche Forschungsbefunde.				
Ich verlasse mich lieber auf meine eigene Erfahrung als auf Forschungsbefunde.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>7. Wenn Sie einen bildungswissenschaftlichen Forschungsbericht lesen, könnten darin unter anderem folgende Fachbegriffe auftauchen. Wie gut schätzen Sie Ihr Verständnis dieser Fachbegriffe ein?</b>	<b>1</b> Kenne den Begriff nicht	<b>2</b> Habe ein grobes Verständnis des Begriffs	<b>3</b> Verstehe den Begriff, aber könnte ihn anderen nicht erklären	<b>4</b> Verstehe den Begriff und könnte ihn anderen erklären
Evidenz	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Randomisiertes Kontrollgruppendesign	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quasi-Experiment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Meta-Analyse	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Repräsentative Stichprobe	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Statistische Signifikanz	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Effektstärke	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Korrelation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Regression	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Validität	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>8. Bezüglich der Beurteilung von wissenschaftlichen Artikeln fühle ich mich ...</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
... sicher.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... kompetent.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... überfordert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... unerfahren.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## **Das denke ich über Wissenschaft und Forschung in den Gesundheits- bzw. den Bildungswissenschaften**

Im Folgenden interessieren wir uns für Ihre Wahrnehmung von Wissenschaft und Forschung in den beiden Domänen der Gesundheitswissenschaften (Medizin, Pflegewissenschaft, Ernährungswissenschaft) sowie der Bildungswissenschaften (Erziehungswissenschaft/Pädagogik, Psychologie, Bildungssoziologie, Fachdidaktik). Aus diesem Grund finden Sie im Folgenden zweimal annähernd die gleichen Fragen, die jeweils auf eine der beiden Domänen zugeschnitten sind.

## Gesundheitswissenschaften (Medizin, Pflegewissenschaft, Ernährungswissenschaft)

9. Wie sehr stimmen Sie mit den folgenden Aussagen überein?	1 stimme gar nicht zu	2 stimme eher nicht zu	3 stimme eher zu	4 stimme ganz zu
Im Allgemeinen macht es mir Spaß, mich mit gesundheitswissenschaftlichen Themen zu befassen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich eigne mir gerne neues Wissen in den Gesundheitswissenschaften an.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich bin interessiert, Neues in den Gesundheitswissenschaften zu lernen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

10. Ich finde, dass...	1 stimmt überhaupt nicht	2 stimmt eher nicht	3 stimmt eher	4 stimmt ganz genau
... sich gesundheitswissenschaftliche Forschungsergebnisse gut auf die Realität übertragen lassen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... aus der Forschung in den Gesundheitswissenschaften brauchbare Schlüsse für das richtige Leben gezogen werden können.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Gesundheitswissenschaften im Großen und Ganzen wichtige Erkenntnisse erbringt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Gesundheitswissenschaften im Großen und Ganzen verlässliche Ergebnisse produziert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Gesundheitswissenschaften in der Regel <u>keine</u> eindeutigen Ergebnisse hervorbringt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Wissenschaftlerinnen und Wissenschaftler in den Gesundheitswissenschaften manchmal nicht besonders kompetent sind.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Wissenschaftlerinnen und Wissenschaftler in den Gesundheitswissenschaften nur die Ergebnisse finden, die sie finden wollen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Wissenschaftlerinnen und Wissenschaftler in den Gesundheitswissenschaften oft voreingenommen sind.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Gesundheitswissenschaften im Allgemeinen wenig sinnvoll ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Mit den folgenden Fragen möchten wir etwas über Ihre Vorstellung von Gesundheitswissenschaften als Wissensgebiet erfahren. Damit meinen wir ihre **Ansichten zum Wissen in den Gesundheitswissenschaften im Allgemeinen und nicht Ihre persönlichen Kenntnisse** in diesem Gebiet. Dazu stellen wir Ihnen eine Liste mit verschiedenen Adjektivpaaren zur Verfügung. Die Adjektive umschreiben jeweils die Enden einer Skala. Zum Beispiel kann man sich Wissen als „einfach“ bis „komplex“ vorstellen. Die Skalen sind jeweils in sieben Stufen gegliedert.

### 11. Wissen im Bereich der Gesundheitswissenschaften ist:

	1	2	3	4	5	6	7	
stabil	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	instabil
objektiv	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	subjektiv
beweisbar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unbeweisbar
dynamisch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	statisch
oberflächlich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	tief
vergänglich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unvergänglich
exakt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	diffus
absolut	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	relativ
geordnet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ungeordnet
genau	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ungenau
flexibel	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	inflexibel
eindeutig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	mehrdeutig
ausgehandelt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	entdeckt
strukturiert	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unstrukturiert
fertig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unvollständig
widerlegbar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unwiderlegbar
offen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	abgeschlossen

## Bildungswissenschaften

12. Wie sehr stimmen Sie mit den folgenden Aussagen überein?	1 stimme gar nicht zu	2 stimme eher nicht zu	3 stimme eher zu	4 stimme ganz zu
Ich beschäftige mich gerne mit bildungswissenschaftlichen Problemen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich eigne mir gerne neues Wissen in den Bildungswissenschaften an.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich bin interessiert, Neues in den Bildungswissenschaften zu lernen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

13. Ich finde, dass...	1 stimmt überhaupt nicht	2 stimmt eher nicht	3 stimmt eher	4 stimmt ganz genau
... sich bildungswissenschaftliche Forschungsergebnisse gut auf die Realität übertragen lassen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... aus der Forschung in den Bildungswissenschaften brauchbare Schlüsse für das richtige Leben gezogen werden können.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Bildungswissenschaften im Großen und Ganzen wichtige Erkenntnisse erbringt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Bildungswissenschaften im Großen und Ganzen verlässliche Ergebnisse produziert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Bildungswissenschaften in der Regel <u>keine</u> eindeutigen Ergebnisse hervorbringt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Wissenschaftlerinnen und Wissenschaftler in den Bildungswissenschaften manchmal nicht besonders kompetent sind.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Wissenschaftlerinnen und Wissenschaftler in den Bildungswissenschaften nur die Ergebnisse finden, die sie finden wollen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Wissenschaftlerinnen und Wissenschaftler in den Bildungswissenschaften oft voreingenommen sind.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Bildungswissenschaften im Allgemeinen wenig sinnvoll ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

#### 14. Wissen im Bereich der Bildungswissenschaften ist:

---

	1	2	3	4	5	6	7	
stabil	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	instabil
objektiv	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	subjektiv
beweisbar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unbeweisbar
dynamisch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	statisch
oberflächlich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	tief
vergänglich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unvergänglich
exakt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	diffus
absolut	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	relativ
geordnet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ungeordnet
genau	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ungenau

— — — —

---

flexibel	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		inflexibel
eindeutig	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		mehrdeutig
ausgehandelt	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		entdeckt
strukturiert	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		unstrukturiert
fertig	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		unvollständig
widerlegbar	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		unwiderlegbar
offen	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		abgeschlossen

---

## Fragen zu Ihrem Hintergrund

<b>15. Ich bin...</b>	
<input type="checkbox"/> weiblich	<input type="checkbox"/> männlich
() _____ Jahre alt	

<b>16. Folgende Sprache(n) spreche ich als Erstsprache(n) bzw. Muttersprache(n)</b> (Mehrfachantwort möglich)	
<input type="checkbox"/> Deutsch	<input type="checkbox"/> Eine andere Sprache (): _____

<b>17. Ich befinde mich im ...</b>	
<input type="checkbox"/> Bachelor-Studium	<input type="checkbox"/> Master-Studium

<b>18. Ich studiere folgende Fächer (bitte alle studierten Fächer ankreuzen)</b>	
<b>Mathematik, Naturwissenschaften &amp; Technik</b>	
Mathematik	<input type="checkbox"/>
Physik	<input type="checkbox"/>
Biologie	<input type="checkbox"/>
Chemie	<input type="checkbox"/>
Geographie	<input type="checkbox"/>
Informatik	<input type="checkbox"/>
Technik	<input type="checkbox"/>
<b>Sprachen</b>	
Germanistik	<input type="checkbox"/>
Deutsch als Fremdsprache/Zweitsprache	<input type="checkbox"/>
Anglistik/Amerikanistik	<input type="checkbox"/>
Romanistik (Französisch)	<input type="checkbox"/>

Romanistik (Spanisch)	<input type="checkbox"/>
Romanistik (Italienisch)	<input type="checkbox"/>
Slawistik	<input type="checkbox"/>
Altphilologie (Latein)	<input type="checkbox"/>
Altphilologie (Altgriechisch)	<input type="checkbox"/>
Sonstige Sprachen (✍): _____	
Sonstige Sprachen (✍): _____	
<b>Soziales, Gesellschaft &amp; Wirtschaft</b>	
Geschichtswissenschaften	<input type="checkbox"/>
Politikwissenschaft, Soziologie, Rechtswissenschaften, Wirtschaftswissenschaften	<input type="checkbox"/>
Pädagogik, Psychologie, Erziehungswissenschaft als Schulfach	<input type="checkbox"/>
<b>Kunst, Kultur &amp; Sport</b>	
Kunst	<input type="checkbox"/>
Musik	<input type="checkbox"/>
Sport	<input type="checkbox"/>
Philosophie	<input type="checkbox"/>
Religion	<input type="checkbox"/>
<b>Andere Fächer (bitte angeben)</b>	
(✍): _____	
(✍): _____	

**19. Bisherige Studienzeiten (✍)**

In meinem aktuellen Studiengang bin im \_\_\_\_ Semester eingeschrieben (Fachsemester)

**20. Vor meinem jetzigen Studiengang habe ich bereits einen anderen Studiengang studiert**

(Falls ja, nennen Sie bitte das Studienfach)

<input type="checkbox"/> nein	<input type="checkbox"/> ja
	und zwar in folgendem Fach (✍): _____
	Wie viele Semester? (✍) ____

**21. Haben Sie bereits an einem empirischen Projekt (mit)gearbeitet (z.B. Hiwi, Bachelor-/Masterarbeit)?**

<input type="checkbox"/> nein	<input type="checkbox"/> ja, und zwar (✍) _____
-------------------------------	---

**22. Haben Sie bereits Lehrveranstaltungen zu Forschungsmethoden oder Statistik besucht?**

<input type="checkbox"/> nein	<input type="checkbox"/> ja, in empirischen Forschungsmethoden <input type="checkbox"/> ja, in Statistik <input type="checkbox"/> ja, _____ (✍)
-------------------------------	---

<b>23. Ich habe vor meiner ersten Hochschuleinschreibung bereits eine Berufsausbildung erfolgreich abgeschlossen:</b> (Falls ja, nennen Sie bitte den Ausbildungsberuf)	
<input type="checkbox"/> nein	<input type="checkbox"/> ja
und zwar in folgendem Beruf (✍️): _____	

<b>24. Im Zeugnis der Hochschulreife habe ich folgende Noten erzielt:</b>	✍️
Gesamtnote (Durchschnittsnote, z.B. 2,5 )	____, ____
In Deutsch (Gesamtnote in Punkten, z.B. „9“)	_____

Im Folgenden finden Sie noch einige unterschiedliche Aussagen zu Ihnen selbst. Wahrscheinlich werden Sie einigen Aussagen sehr zustimmen und andere stark ablehnen. Bei wieder anderen sind Sie vielleicht unentschieden. **Nehmen Sie bitte zu jeder der folgenden Aussagen Stellung und entscheiden Sie, inwieweit jede auf Sie ganz persönlich zutrifft.**

25.	1 stimmt überhaupt nicht	2 stimmt eher nicht	3 stimmt eher	4 stimmt ganz genau
Ich probiere gerne Dinge aus, auch wenn nicht immer etwas dabei herauskommt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich beschäftige mich nur mit Aufgaben, die lösbar sind.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich mag es, wenn unverhofft Überraschungen auftreten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich lassen Dinge auf mich zukommen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe es gerne, wenn die Arbeit gleichmäßig verläuft.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich warte geradezu darauf, dass etwas Aufregendes passiert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wenn um mich herum alles drunter und drüber geht, fühle ich mich so richtig wohl.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich weiß gerne, was auf mich zukommt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

# STOP

## Teil 2: Fragen zu Ihrer Beurteilung des Trainings „Forschung verstehen“

<b>26. Wie haben Sie sich während des Trainings gefühlt?</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
Das Training hat mir Spaß gemacht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe alles um mich herum vergessen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich war voll und ganz bei der Sache.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe mich voll auf das Training konzentriert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich war verärgert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe mich gelangweilt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Es ist mir schwer gefallen, bei der Sache zu bleiben.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe vor mich hingeträumt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich war während des Trainings abgelenkt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe über alle möglichen Dinge nachgedacht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>27. Wie fanden Sie die Inhalte des ersten Trainingstages?</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
<b>Die im Training angeführten Beispiele waren...</b>				
... praxisnah.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... relevant für meinen zukünftigen Lehrerberuf.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... interessant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... motivierend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... klar und verständlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Die im Training angewandten Unterrichtsmethoden waren...</b>				
... abwechslungsreich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... interessant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... motivierend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... klar und verständlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Die im Training durchgeführten Übungen waren...</b>				
... abwechslungsreich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... interessant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... motivierend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... klar und verständlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Ich empfand die Trainerin als...</b>				
... kompetent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... motivierend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... unterstützend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>28. Wie empfanden Sie die Dauer des ersten Trainingstages?</b>	<b>1</b> viel zu kurz	<b>2</b> etwas zu kurz	<b>3</b> genau richtig	<b>4</b> etwas zu lang	<b>5</b> viel zu lang
Die Dauer des ersten Trainingstages war ...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die Dauer des Blocks „Evidenz kennenlernen“ war ...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die Dauer des Blocks „Evidenz bewerten“ war ...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>29. Wie motiviert waren Sie während der Teilnahme an dem Training?</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
Ich fand die Lernschritte zur kritischen Bewertung von Evidenz sehr interessant	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe das Gefühl, Studien jetzt besser verstehen zu können.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich hätte Lust, mehr zum Thema Bewertung von Studien zu erfahren	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe die Lernschritte zur kritischen Bewertung von Evidenz konzentriert bearbeitet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe die Lernschritte zur kritischen Bewertung von Evidenz sorgfältig bearbeitet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe mir Mühe bei der Bearbeitung der Lernschritte zur kritischen Bewertung von Evidenz gegeben.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>30. Wie bewerten Sie Ihren Lernerfolg?</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
Die durch die Lernziele geweckten Erwartungen an das Training werden insgesamt erfüllt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich konnte die Lernziele alles in allem erreichen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich schätze meinen Lernerfolg insgesamt hoch ein.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

# STOP

### Teil 3: Fragen zu Ihren Ansichten bezüglich des Umgangs mit Evidenz

<b>31. Wie gut stimmen die folgenden Aussagen mit Ihrer persönlichen Sichtweise zu einer möglichst evidenzbasierten - im Sinn einer an bildungswissenschaftlichen Forschungsbefunden orientierten - Unterrichtspraxis überein?</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
Ich finde eine stärkere Evidenzbasierung im Bildungsbereich wichtig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die meisten meiner Mitstudierenden finden eine stärkere Evidenzbasierung im Bildungsbereich wichtig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die Unterrichtsqualität könnte erheblich verbessert werden, wenn Lehrerinnen und Lehrer sich stärker an Unterrichtsmethoden orientieren würden, deren Wirksamkeit wissenschaftlich überprüft ist	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die Forderung nach einer evidenzbasierten Unterrichtspraxis überfordert die meisten Lehrerinnen und Lehrer.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>32. Was sind aus Ihrer Sicht Hinderungsgründe dafür, erziehungswissenschaftliche Forschungsbefunde als Grundlage für Ihre Unterrichtsgestaltung zu nutzen?</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
Das ist im Allgemeinen nicht üblich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe keinen/nur begrenzten Zugang zu Forschungsergebnissen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich weiß zu wenig darüber, wie ich für mich relevante Forschungsergebnisse finden kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zu vielen mich interessierenden Fragen gibt es kaum relevante Forschungsbefunde.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mir fehlt die Zeit, Forschungsbefunde zu suchen und zu lesen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Es fällt mir schwer, forschungsmethodische oder statistische Informationen zu verstehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die meisten erziehungswissenschaftlichen Forschungsbefunde sind für meine Tätigkeit als Lehrer/in nicht relevant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Erziehungswissenschaftlichen Forschungsbefunde geben in der Regel zu wenig konkrete Handlungsanweisungen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich interessiere mich nicht für erziehungswissenschaftliche Forschungsbefunde.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich verlasse mich lieber auf meine eigene Erfahrung als auf Forschungsbefunde.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>33. Wenn Sie einen bildungswissenschaftlichen Forschungsbericht lesen, könnten darin unter anderem folgende Fachbegriffe auftauchen. Wie gut schätzen Sie Ihr Verständnis dieser Fachbegriffe ein?</b>	<b>1</b> Kenne den Begriff nicht	<b>2</b> Habe ein grobes Verständnis des Begriffs	<b>3</b> Verstehe den Begriff, aber könnte ihn anderen nicht erklären	<b>4</b> Verstehe den Begriff und könnte ihn anderen erklären
Evidenz	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Randomisiertes Kontrollgruppendesign	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quasi-Experiment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Meta-Analyse	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Repräsentative Stichprobe	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Statistische Signifikanz	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Effektstärke	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Korrelation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Regression	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Validität	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>34. Bezüglich der Beurteilung von wissenschaftlichen Artikeln fühle ich mich ...</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
... sicher.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... kompetent.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... überfordert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... unerfahren.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Fragen zu Ihren Einstellungen zu den beiden Domänen Gesundheits- und Bildungswissenschaften

Anschließend würden wir gerne mehr über Ihre Einstellung zu den Gesundheits- sowie den Bildungswissenschaften erfahren.

Aus diesem Grund finden Sie im Folgenden zweimal annähernd die gleichen Fragen, die jeweils auf eine der beiden Domänen zugeschnitten sind.

### Gesundheitswissenschaften

<b>35. Wie sehr stimmen Sie mit den folgenden Aussagen überein?</b>	<b>1</b> stimme gar nicht zu	<b>2</b> stimme eher nicht zu	<b>3</b> stimme eher zu	<b>4</b> stimme ganz zu
Im Allgemeinen macht es mir Spaß, mich mit gesundheitswissenschaftlichen Themen zu befassen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich lese gerne etwas über Gesundheitswissenschaften.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich beschäftige mich gerne mit gesundheitswissenschaftlichen Problemen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich eigne mir gerne neues Wissen in den Gesundheitswissenschaften an.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich bin interessiert, Neues in den Gesundheitswissenschaften zu lernen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>36. Ich finde, dass...</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
... sich gesundheitswissenschaftliche Forschungsergebnisse gut auf die Realität übertragen lassen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Wissenschaftlerinnen und Wissenschaftler in den Gesundheitswissenschaften manchmal nicht besonders kompetent sind.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Gesundheitswissenschaften in der Regel keine eindeutigen Ergebnisse hervorbringt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Wissenschaftlerinnen und Wissenschaftler in den Gesundheitswissenschaften nur die Ergebnisse finden, die sie finden wollen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Gesundheitswissenschaften im Großen und Ganzen verlässliche Ergebnisse produziert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Gesundheitswissenschaften im Großen und Ganzen wichtige Erkenntnisse erbringt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Wissenschaftlerinnen und Wissenschaftler in den Gesundheitswissenschaften oft voreingenommen sind.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... aus der Forschung in den Gesundheitswissenschaften brauchbare Schlüsse für das richtige Leben gezogen werden können.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Gesundheitswissenschaften im Allgemeinen wenig sinnvoll ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**37. Wissen im Bereich der Gesundheitswissenschaften ist:**

---

	1	2	3	4	5	6	7	
stabil	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	instabil
objektiv	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	subjektiv
beweisbar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unbeweisbar
dynamisch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	statisch
oberflächlich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	tief
vergänglich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unvergänglich
exakt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	diffus
absolut	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	relativ
geordnet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ungeordnet
genau	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ungenau
flexibel	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	inflexibel
eindeutig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	mehrdeutig
ausgehandelt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	entdeckt
strukturiert	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unstrukturiert
fertig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unvollständig
widerlegbar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unwiderlegbar
offen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	abgeschlossen

---

## Bildungswissenschaften

<b>38. Wie sehr stimmen Sie mit den folgenden Aussagen überein?</b>	<b>1</b> stimme gar nicht zu	<b>2</b> stimme eher nicht zu	<b>3</b> stimme eher zu	<b>4</b> stimme ganz zu
Im Allgemeinen macht es mir Spaß, mich mit bildungswissenschaftlichen Themen zu befassen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich lese gerne etwas über Bildungswissenschaften.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich beschäftige mich gerne mit bildungswissenschaftlichen Problemen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich eigne mir gerne neues Wissen in den Bildungswissenschaften an.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich bin interessiert, Neues in den Bildungswissenschaften zu lernen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>39. Ich finde, dass...</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
... sich bildungswissenschaftliche Forschungsergebnisse gut auf die Realität übertragen lassen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Wissenschaftlerinnen und Wissenschaftler in den Bildungswissenschaften manchmal nicht besonders kompetent sind.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Bildungswissenschaften in der Regel keine eindeutigen Ergebnisse hervorbringt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Wissenschaftlerinnen und Wissenschaftler in den Bildungswissenschaften nur die Ergebnisse finden, die sie finden wollen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Bildungswissenschaften im Großen und Ganzen verlässliche Ergebnisse produziert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Bildungswissenschaften im Großen und Ganzen wichtige Erkenntnisse erbringt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Wissenschaftlerinnen und Wissenschaftler in den Bildungswissenschaften oft voreingenommen sind.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... aus der Forschung in den Bildungswissenschaften brauchbare Schlüsse für das richtige Leben gezogen werden können.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... Forschung in den Bildungswissenschaften im Allgemeinen wenig sinnvoll ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**40. Wissen im Bereich der Bildungswissenschaften ist:**

---

	1	2	3	4	5	6	7	
stabil	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	instabil
objektiv	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	subjektiv
beweisbar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unbeweisbar
dynamisch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	statisch
oberflächlich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	tief
vergänglich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unvergänglich
exakt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	diffus
absolut	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	relativ
geordnet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ungeordnet
genau	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ungenau
flexibel	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	inflexibel
eindeutig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	mehrdeutig
ausgehandelt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	entdeckt
strukturiert	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unstrukturiert
fertig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unvollständig
widerlegbar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unwiderlegbar
offen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	abgeschlossen

---

## Teil 4: Fragen zu Ihrer Beurteilung des Trainings „Forschung verstehen“

<b>41. Wie haben Sie sich während des Trainings gefühlt?</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
Das Training hat mir Spaß gemacht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe alles um mich herum vergessen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich war voll und ganz bei der Sache.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe mich voll auf das Training konzentriert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich war verärgert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe mich gelangweilt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Es ist mir schwer gefallen, bei der Sache zu bleiben.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe vor mich hingeträumt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich war während des Trainings abgelenkt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe über alle möglichen Dinge nachgedacht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>42. Wie fanden Sie die Inhalte des zweiten Trainingstages?</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
<b>Die im Training angeführten Beispiele waren...</b>				
... praxisnah.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... relevant für meinen zukünftigen Lehrerberuf.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... interessant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... motivierend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... klar und verständlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Die im Training angewandten Unterrichtsmethoden waren...</b>				
... abwechslungsreich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... interessant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... motivierend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... klar und verständlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Die im Training durchgeführten Übungen waren...</b>				
... abwechslungsreich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... interessant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... motivierend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... klar und verständlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Ich empfand die Trainerin als...</b>				
... kompetent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... motivierend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... unterstützend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>43. Wie empfanden Sie die Dauer des zweiten Trainingstages?</b>	<b>1</b> viel zu kurz	<b>2</b> etwas zu kurz	<b>3</b> genau richtig	<b>4</b> etwas zu lang	<b>5</b> viel zu lang
Die Dauer des zweiten Trainingstages war ...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die Dauer des Blocks „Evidenz nutzen“ war ...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>44. Wie motiviert waren Sie während der Teilnahme an dem Training?</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
Ich fand die Lernschritte zur argumentativen Nutzung von Evidenz sehr interessant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe das Gefühl, jetzt besser auf Basis von Studien argumentieren zu können.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich hätte Lust, mehr zum Thema Bewertung von Studien zu erfahren.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe die Lernschritte zur argumentativen Nutzung von Evidenz konzentriert bearbeitet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe die Lernschritte zur argumentativen Nutzung von Evidenz sorgfältig bearbeitet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe mir Mühe bei der Bearbeitung der Lernschritte zur argumentativen Nutzung von Evidenz gegeben.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>45. Wie bewerten Sie Ihren Lernerfolg?</b>	<b>1</b> stimmt überhaupt nicht	<b>2</b> stimmt eher nicht	<b>3</b> stimmt eher	<b>4</b> stimmt ganz genau
Die durch die Lernziele geweckten Erwartungen an das Training werden insgesamt erfüllt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich konnte die Lernziele alles in allem erreichen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich schätze meinen Lernerfolg insgesamt hoch ein.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**46. Ich würde dem gesamten Training folgende Schulnote geben: \_\_\_\_\_**

**Herzlichen Dank für Ihre Rückmeldungen!**

## B.4. Kodierschema Studie 1

### Kategorienschema

TN: \_\_\_\_\_

Thema: HA / KG

#### 1. Globalrating (nach Wilson und Kuhn)

Welche der folgenden Stufen charakterisiert den Studierenden insgesamt am besten?  
(Hinweis: höher liegende Stufen sollten die darunterliegenden implizieren.)

5	<input type="checkbox"/>	Der Studierende begründet durchgängig evidenzbasiert und hinterfragt die Evidenz kritisch und berücksichtigt die Alternativ-Erklärungen der anderen Studien.
4	<input type="checkbox"/>	Der Studierende begründet durchgängig evidenzbasiert und verwendet dabei relevante und angemessene Evidenz.
3	<input type="checkbox"/>	Der Studierende begründet teilweise evidenzbasiert, aber nicht in Bezug auf alle Elemente.
2	<input type="checkbox"/>	Der Studierende zitiert irrelevante Evidenz oder macht inhaltliche Fehler.
1	<input type="checkbox"/>	Der Studierende nennt durchgängig keine Begründungen und führt keine Evidenz auf.

#### 2. Vorhandensein der Argumentations-Heuristiken

Argument	_____
Argument + Beleg	_____
Argument + Beleg + Kritische Bewertung	=====

#### 3. Evidenz innerhalb der Argumente (nach Bell und Toulmin)

Häufigkeiten zählen, daraus resultierend: Mittelwert ankreuzen

Länge	Argument 1	Argument 2	Argument 3	Argument 4	Argument 5
keine	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
kurz (nur Quelle)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
mittel (1-2 Sätze)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
lang (mehr als 3 Sätze)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Klarheit	Argument 1	Argument 2	Argument 3	Argument 4	Argument 5
sehr unklar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
eher unklar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
eher klar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
sehr klar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Typ	Argument 1	Argument 2	Argument 3	Argument 4	Argument 5
keine (benutzt keine Evidenz)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
irrelevant (benutzt irrelevant Evidenz)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
beschreibend (nacherzählen des Inhalts der Evidenzquelle)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Single-Warrant (enthält eine wissenschaftliche Behauptung)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Multiple-Warrant (enthält mehr als eine wissenschaftliche Behauptung)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Warrant / Backing	Argument 1	Argument 2	Argument 3	Argument 4	Argument 5
keine (zitiert keine Quellen)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
persönlich (zitiert Situationen aus dem Alltag)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
individuell (benutzt eigene Evidenzquelle)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
instruiert (benutzt die ihm vorgelegte Evidenzquelle)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
beides (benutzt vorgegebene und eigene Evidenzquellen)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

#### 4. Art der Argumente (nach Kuhn)

##### Position

- one-sided (aus der eigenen Perspektive betrachtet)
- two-sided (von zwei Seiten aus betrachtet)

#### 5. Verhalten bei Uneinigkeit

Wie verhält sich der Studierende bei Uneinigkeit (Häufigkeiten)?

##### Uneinigkeit

- keine
- Disagreement: Zeigt dass er /sie dem Text nicht zustimmt, allerdings ohne ein weiteres Argument weitere Einzelheiten zu nennen
- Counter-A: Zeigt dass er /sie dem Text nicht zustimmt und äußert ein Gegenargument
- Counter-C: Zeigt dass er /sie den Äußerungen dem Text nicht zustimmt und äußert Kritik

## B.5. Kooperationskript

Schritt	Auftrag
1. <i>Lesen und bewerten der vorgelegten Studie</i>	1.1. Bitte lesen Sie die Ihnen vorgelegte Studie. 1.2. Verwenden Sie einen Textmarker um wichtige Stellen zu markieren. 1.3. Wenden Sie dabei das Schema „Evidenz bewerten“ an, um die Belastbarkeit ihrer Studie zu überprüfen. 1.4. Versuchen Sie zunächst objektiv an die Studie heranzugehen um es anschließend mit Ihrer eigenen Meinung in Einklang zu bringen.
2. <i>Vorstellung der Studien</i>	2.1. Stellen Sie den anderen Mitgliedern Ihrer Gruppe die wichtigsten Eckpunkte Ihrer Studie vor. 2.2. Gehen Sie dabei kritisch auf Fragestellung, Methode und Design, Ergebnisse und Schlussfolgerungen ein. 2.3. Geben Sie anschließend Ihre Einschätzung zu der Studie ab und erfragen Sie gegebenenfalls die Meinung der anderen Gruppenmitglieder. 2.4. Hören Sie den Ausführungen der anderen genau zu und überlegen Sie, ob Sie mit deren Meinung übereinstimmen.
3. <i>Diskussion</i>	3.1. Diskutieren Sie nun mit Ihrer Gruppe über die Studienergebnisse und die Qualität Ihrer Studien. 3.2. Achten Sie bei der Diskussion darauf, dass Sie überwiegend auf Basis von Forschungsergebnissen argumentieren. 3.3. Versuchen Sie prinzipiell, jedes Ihrer Argumente zu belegen (siehe Toulmin Schema). 3.4. Überlegen Sie, ob Ihnen eventuell noch weitere Studien zu diesem Thema bekannt sind. 3.5. Stellen Sie gegebenenfalls Rückfragen an Ihre Gruppenmitglieder. 3.6. Sollten Sie einem Argument Ihrer Gruppenmitglieder nicht zustimmen, antworten Sie mit einem Gegenargument, welches Sie, wenn möglich, belegen können, oder mit einer kritischen Bewertung.
4. <i>Entscheidungsfindung</i>	4.1. Versuchen Sie nun gemeinsam in Ihrer Gruppe eine Entscheidung zu treffen. 4.2. Bringen Sie hierzu die eben diskutierten Forschungsergebnisse in Zusammenhang mit Ihren Erfahrungen und Ihrer Expertise. 4.3. Kommen Sie auf Basis dessen zu einer Entscheidung.

## B.6. Kodierschema Studie 2: schriftliche Statements

### Kategoriensystem Texte

TN: \_\_\_\_\_/ Standort: \_\_\_\_\_/ Bedingung: \_\_\_\_\_

Hausaufgaben (\_1)/ Kassengröße (\_2)/ Grippeimpfung (\_3)/ NEM (\_4)

#### I. Bewertungskompetenz

##### 6. Globalrating (in Anlehnung an Wilson): Bewertungskompetenz

*Welche der folgenden Stufen charakterisiert den Studierenden insgesamt am besten?*

*Kritisch bewerten bedeutet die Qualität des Artikels zu bewerten (positiv wie negativ)*

<b>5</b>	<input type="checkbox"/>	Der Studierende nennt die relevanten und wichtigsten Inhalte und bewertet diese kritisch. Darüber hinaus erkennt er die Fragilität und Konflikthaftigkeit des Forschungsfeldes.
<b>4</b>	<input type="checkbox"/>	Der Studierende nennt die relevanten und wichtigsten Inhalte und bewertet diese durchgängig kritisch.
<b>3</b>	<input type="checkbox"/>	Der Studierende nennt die relevanten Inhalte und bewertet diese teilweise kritisch.
<b>2</b>	<input type="checkbox"/>	Der Studierende nennt Inhalte der Studie ohne diese kritisch zu bewerten.
<b>1</b>	<input type="checkbox"/>	Der Studierende nennt irrelevante oder falsche Inhalte der Studie. Zudem bewertet er diese nicht kritisch oder sogar falsch.

#### II. Argumentationskompetenz

##### 1. Globalrating (nach Wilson): Argumentationskompetenz

*Welche der folgenden Stufen charakterisiert den Studierenden insgesamt am besten? (Hinweis: höher liegende Stufen sollten die darunterliegenden implizieren.)*

<b>5</b>	<input type="checkbox"/>	Der Studierende begründet durchgängig evidenzbasiert und hinterfragt die Evidenz kritisch.
<b>4</b>	<input type="checkbox"/>	Der Studierende begründet durchgängig evidenzbasiert und verwendet dabei relevante und angemessene Evidenz.
<b>3</b>	<input type="checkbox"/>	Der Studierende begründet teilweise evidenzbasiert, aber nicht in Bezug auf alle Elemente.
<b>2</b>	<input type="checkbox"/>	Der Studierende zitiert irrelevante Evidenz oder macht inhaltliche Fehler.
<b>1</b>	<input type="checkbox"/>	Der Studierende nennt durchgängig keine Begründungen und führt keine Evidenz auf.

## 2. Vorhandensein der Argumentations-Heuristiken

Häufigkeiten zählen:

<input type="checkbox"/>	Argument ( <i>Claim steht alleine</i> )
<input type="checkbox"/>	Argument + Beleg
<input type="checkbox"/>	Argument + Beleg + Kritische Bewertung

## 3. Evidenz innerhalb der Argumente (nach Bell und Toulmin)

*Die Evidenz innerhalb der Argumente ist überwiegend:*

<b>Länge</b>	
<input type="checkbox"/>	kurz (ca. ein Satz)
<input type="checkbox"/>	mittel (ca. 2 Sätze)
<input type="checkbox"/>	lang (mehr als 3 Sätze)

### Typ der Evidenzquelle

*Am häufigsten:*

<b>Typ</b>	
<input type="checkbox"/>	Keine (benutzt keine Evidenz)
<input type="checkbox"/>	irrelevant (benutzt irrelevante Evidenz)
<input type="checkbox"/>	beschreibend (nacherzählen des Inhalts der Evidenzquelle)
<input type="checkbox"/>	Single-Warrant (enthält eine wissenschaftliche Behauptung über die Evidenzquelle)
<input type="checkbox"/>	Multiple-Warrant (enthält mehr als eine wissenschaftliche Behauptung über die Evidenzquelle)

*Am zweit-häufigsten:*

<b>Typ</b>	
<input type="checkbox"/>	Keine (benutzt keine Evidenz)
<input type="checkbox"/>	irrelevant (benutzt irrelevante Evidenz)
<input type="checkbox"/>	beschreibend (nacherzählen des Inhalts der Evidenzquelle)
<input type="checkbox"/>	Single-Warrant (enthält eine wissenschaftliche Behauptung über die Evidenzquelle)
<input type="checkbox"/>	Multiple-Warrant (enthält mehr als eine wissenschaftliche Behauptung über die Evidenzquelle)

### Backing

*Am häufigsten:*

<b>Warrant-Quelle / Backing</b>	
<input type="checkbox"/>	Keine (benutzt keine Evidenzquelle)
<input type="checkbox"/>	persönlich (zitiert Situationen aus dem Alltag)
<input type="checkbox"/>	individuell (benutzt eigene Evidenzquelle)
<input type="checkbox"/>	instruiert (benutzt die ihm vorgelegte Evidenzquelle)
<input type="checkbox"/>	beides (benutzt vorgegebene und eigene Evidenzquellen)

Am zweit-häufigsten:

---

**Warrant-Quelle / Backing**

---

- Keine (benutzt keine Evidenzquelle)
  - persönlich (zitiert Situationen aus dem Alltag)
  - individuell (benutzt eigene Evidenzquelle)
  - instruiert (benutzt die ihm vorgelegte Evidenzquelle)
  - beides (benutzt vorgegebene und eigene Evidenzquellen)
- 

**4. Klarheit (nach Bell)**

*Die Klarheit der Argumente ist überwiegend:*

---

**Klarheit**

- sehr unklar
  - eher unklar
  - eher klar
  - sehr klar
- 

**5. Art der Argumente (nach Kuhn)**

---

**Position**

- one-sided (nur aus der eigenen Perspektive betrachtet)
  - two-sided (von zwei Seiten aus betrachtet)
- 

**6. Verhalten bei Uneinigkeit (nach Kuhn)**

*Der Studierende benutzt bei Uneinigkeit mit der Evidenzquelle überwiegend:*

---

**Uneinigkeit**

- Disagreement: Zeigt, dass er/sie den Äußerungen des Artikels/der Artikel nicht zustimmt, allerdings ohne ein weiteres Argument/weitere Einzelheiten zu nennen
  - Counter-A: Zeigt, dass er /sie den Äußerungen des Artikels/der Artikel nicht zustimmt und äußert ein Gegenargument
  - Counter-C: Zeigt, dass er /sie den Äußerungen des Artikels/der Artikel nicht zustimmt und äußert Kritik
  - Counter-A und Counter-C
-

## B.7. Kodierschema Studie 2: videografierte Gruppendiskussionen

### Kategoriensystem Videos

TN: \_\_\_\_\_/ Gruppe: \_\_\_\_\_/ Standort: \_\_\_\_\_/ Bedingung: \_\_\_\_\_

GeWi (\_1)/ BiWi (\_2)

#### I. Bewertungskompetenz

##### 7. Globalrating (in Anlehnung an Wilson): Bewertungskompetenz

*Welche der folgenden Stufen charakterisiert den Studierenden insgesamt am besten?*

*Kritisch bewerten bedeutet die Qualität des Artikels zu bewerten (positiv wie negativ)*

5	<input type="checkbox"/>	Der Studierende nennt die relevanten und wichtigsten Inhalte und bewertet diese kritisch. Darüber hinaus erkennt er die Fragilität und Konflikthaftigkeit des Forschungsfeldes.
4	<input type="checkbox"/>	Der Studierende nennt die relevanten und wichtigsten Inhalte und bewertet diese durchgängig kritisch.
3	<input type="checkbox"/>	Der Studierende nennt die relevanten Inhalte und bewertet diese teilweise kritisch.
2	<input type="checkbox"/>	Der Studierende nennt Inhalte der Studie ohne diese kritisch zu bewerten.
1	<input type="checkbox"/>	Der Studierende nennt irrelevante oder falsche Inhalte der Studie. Zudem bewertet er diese nicht kritisch oder sogar falsch.

##### 8. Vorhandensein der Bewertungs-Heuristiken

(Ja = 1/ Nein = 0)

	Ja	Nein
<b>Fragestellung...</b>		
...genannt	<input type="checkbox"/>	<input type="checkbox"/>
...richtig erkannt	<input type="checkbox"/>	<input type="checkbox"/>
...kritisch bewertet	<input type="checkbox"/>	<input type="checkbox"/>
<b>Methoden und Design...</b>		
...genannt	<input type="checkbox"/>	<input type="checkbox"/>
...richtig erkannt	<input type="checkbox"/>	<input type="checkbox"/>
...kritisch bewertet	<input type="checkbox"/>	<input type="checkbox"/>
<b>Ergebnisse...</b>		
...genannt	<input type="checkbox"/>	<input type="checkbox"/>
...richtig erkannt	<input type="checkbox"/>	<input type="checkbox"/>
...kritisch bewertet	<input type="checkbox"/>	<input type="checkbox"/>
<b>Schlussfolgerungen...</b>		
...genannt	<input type="checkbox"/>	<input type="checkbox"/>
...richtig erkannt	<input type="checkbox"/>	<input type="checkbox"/>
...kritisch bewertet	<input type="checkbox"/>	<input type="checkbox"/>

## II. Argumentationskompetenz

### 7. Globalrating (nach Wilson und Kuhn): Argumentationskompetenz

Welche der folgenden Stufen charakterisiert den Studierenden insgesamt am besten?  
(Hinweis: höher liegende Stufen sollten die darunterliegenden implizieren.)

5	<input type="checkbox"/>	Der Studierende begründet durchgängig evidenzbasiert und hinterfragt die Evidenz kritisch und berücksichtigt die Alternativ-Erklärungen der anderen Studierenden.
4	<input type="checkbox"/>	Der Studierende begründet durchgängig evidenzbasiert und verwendet dabei relevante und angemessene Evidenz.
3	<input type="checkbox"/>	Der Studierende begründet teilweise evidenzbasiert, aber nicht in Bezug auf alle Elemente.
2	<input type="checkbox"/>	Der Studierende zitiert irrelevante Evidenz oder macht inhaltliche Fehler.
1	<input type="checkbox"/>	Der Studierende nennt durchgängig keine Begründungen und führt keine Evidenz auf.

### 8. Vorhandensein der Argumentations-Heuristiken

Häufigkeiten zählen:

**Argument (*Claim steht alleine*)**

**Argument + Beleg**

**Argument + Beleg + Kritische Bewertung**

### 9. Evidenz innerhalb der Argumente (nach Bell und Toulmin)

Die Evidenz innerhalb der Argumente ist überwiegend:

**Länge**

kurz (ca. ein Satz)

mittel (ca. 2 Sätze)

lang (mehr als 3 Sätze)

**Typ der Evidenzquelle**

Am häufigsten:

**Typ**

Keine (benutzt keine Evidenz)

irrelevant (benutzt irrelevante Evidenz)

beschreibend (nacherzählen des Inhalts der Evidenzquelle)

Single-Warrant (enthält eine wissenschaftliche Behauptung über die Evidenzquelle)

Multiple-Warrant (enthält mehr als eine wissenschaftliche Behauptung über die Evidenzquelle)

Am zweit-häufigsten:

**Typ**

Keine (benutzt keine Evidenz)

irrelevant (benutzt irrelevante Evidenz)

beschreibend (nacherzählen des Inhalts der Evidenzquelle)

Single-Warrant (enthält eine wissenschaftliche Behauptung über die Evidenzquelle)

Multiple-Warrant (enthält mehr als eine wissenschaftliche Behauptung über die Evidenzquelle)

**Backing***Am häufigsten:***Warrant-Quelle / Backing**

- 
- Keine (benutzt keine Evidenzquelle)
  - persönlich (zitiert Situationen aus dem Alltag)
  - individuell (benutzt eigene Evidenzquelle)
  - instruiert (benutzt die ihm vorgelegte Evidenzquelle)
  - beides (benutzt vorgegebene und eigene Evidenzquellen)
- 

*Am zweit-häufigsten:***Warrant-Quelle / Backing**

- 
- Keine (benutzt keine Evidenzquelle)
  - persönlich (zitiert Situationen aus dem Alltag)
  - individuell (benutzt eigene Evidenzquelle)
  - instruiert (benutzt die ihm vorgelegte Evidenzquelle)
  - beides (benutzt vorgegebene und eigene Evidenzquellen)
- 

**10. Klarheit (nach Bell)***Die Klarheit der Argumente ist überwiegend:***Klarheit**

- 
- sehr unklar
  - eher unklar
  - eher klar
  - sehr klar
- 

**11. Art der Argumente (nach Kuhn)****Position**

- 
- one-sided (nur aus der eigenen Perspektive betrachtet)
  - two-sided (von zwei Seiten aus betrachtet)
- 

**III. Interaktion**

1. Redeanteil des Studierenden (in %) \_\_\_\_\_ %

**2. Interaktionsanalyse (nach Bales und Kuhn)***Welche der folgenden Wahlmöglichkeiten charakterisiert den Studierenden insgesamt am besten?(1/2)***Sozioemotionaler Bereich: Integration**

- 
- Zeigt Antagonismus, setzt andere herab, verteidigt oder behauptet sich
  - Zeigt Solidarität, bestärkt den anderen, hilft, belohnt
-

---

**Sozioemotionaler Bereich: Entscheidung**


---

- Stimmt nicht zu, zeigt passive Ablehnung, Förmlichkeit, gibt keine Hilfe
- Stimmt zu, versteht, stimmt überein, gibt nach
- 

---

**Aufgabenbereich: Orientierung (Versuche der Beantwortung)**


---

- bleibt passiv
- Orientiert, informiert, wiederholt, klärt, bestätigt
- 

---

**Aufgabenbereich: Orientierung (Fragen)**


---

- bleibt passiv
- Erfragt Orientierung, Information, Wiederholung, Bestätigung
- 

**3. Verhalten bei Uneinigkeit (nach Kuhn)**

*Der Studierende benutzt bei Uneinigkeit mit der Evidenzquelle überwiegend:*

---

**Uneinigkeit**


---

- Disagreement: Zeigt, dass er/sie den Äußerungen des Artikels/der Artikel nicht zustimmt, allerdings ohne ein weiteres Argument/weitere Einzelheiten zu nennen
- Counter-A: Zeigt, dass er /sie den Äußerungen des Artikels/der Artikel nicht zustimmt und äußert ein Gegenargument
- Counter-C: Zeigt, dass er /sie den Äußerungen des Artikels/der Artikel nicht zustimmt und äußert Kritik
- Counter-A und Counter-C
- 

**4. Einigung**

*Gab es am Ende der Diskussion eine Einigung der Gruppe auf einen konkreten Lösungsansatz?*

---

**Einigung**


---

- nein
- ja
- 

**5. Dauer der Diskussion (gesamt; in Minuten) \_\_\_\_\_ min**

## B.8. Kodierschema Studie 3

### Auswertungsschema – auf Statementebene

TN: \_\_\_\_\_ / Standort: \_\_\_\_\_ / Bedingung: \_\_\_\_\_

#### Evidenz:

\_\_\_ Ja

\_\_\_ Nein

#### Transaktiv

\_\_\_ Ja

\_\_\_ Nein

#### Level der Evidenznutzung

Level	Beschreibung
1	<i>Erfolgreiches Nutzen von Evidenz um eine Behauptung zu untermauern</i>
1.1a	...um die eigene Position zu bestärken
1.1b	...um die Position anderer zu schwächen
1.1c	...um die Position anderer zu unterstützen
1.1d	...um die eigene Position zu schwächen
1.2.	<i>Erfolgreiches Nutzen von Evidenz um eine Behauptung zu untermauern, teilt aber die Meinung der Evidenzquelle nicht</i>
1.2a	...um die eigene Position zu bestärken
1.2b	...um die Position anderer zu schwächen
1.2c	...um die Position anderer zu unterstützen
1.2d	...um die eigene Position zu schwächen
2.	<i>Versuch aber dennoch kein erfolgreiches Nutzen von Evidenz um eine Behauptung zu untermauern</i>
2.1.	Es ist unklar warum die Evidenzquelle zitiert wird (Passung zur Behauptung)
2.2.	Die Evidenzquelle wird fehlinterpretiert bzw. eine falsche Schlussfolgerung gezogen
3	Bloßes zitieren einer Evidenzquelle (ohne eine Behauptung)
4	Falsches zitieren oder fehlinterpretieren der Evidenzquelle (ohne eine Behauptung)

1 und 2: Behauptung plus Evidenz

3 und 4: keine Behauptung, sondern bloßes zitieren

**Funktion**

<b>Transactive questions (Fragen)</b>		
1.1	Agree-?	Nachfrage, ob die Diskussionspartner der eigenen Behauptung zustimmen
1.2	Case-?	Bitte eine bestimmte Position in einem bestimmten Szenario oder Fallbeispiel einzunehmen
1.3	Clarify-?	Bitte um Klärung, nachfrage
1.4	Justify-?	Frage nach Evidenz/Beleg für die getätigte Aussage
1.5	Meta-?	Eine Frage die die Diskussion selbst betrifft (nicht deren Inhalt)
1.6	Position-?	Bitte eine Position zu beziehen, Frage nach Position/ Meinung
1.7	Question-?	Informelle Frage die sich nicht auf die vorherige Äußerung der anderen bezieht
1.8	Respond-?	Bitte um Antwort
<b>Transactive statements (Aussagen, Behauptungen, Argumente)</b>		
2.1	Counter-A	Zeigt ‚nicht-zustimmen‘ durch Gegenargument
2.2	Counter-C	Zeigt ‚nicht-zustimmen‘ durch Kritik an dem Argument des anderen
2.3	Disagree	Zeigt ‚nicht-zustimmen‘ allerdings ohne Gegenargument oder Kritik (z.B. „Finde ich nicht.“)
2.4	Dismiss	Gibt an dass man die zuvor getätigte Aussage des anderen für irrelevant hält
2.5	Clarify	Klärung der eigenen Position, Klärung von Fragen, klarstellen
2.6	Agree	Zustimmung
2.7	Advance	Ergänzung zu einer zuvor von anderen angeführten Arguments
2.8	Add	Ergänzung zu einer zuvor von anderen getroffenen Aussage
2.9	Aside	Kommentar, der nicht die Aussage anderer ergänzt
2.10	Coopt	Behauptung dass die zuvor getätigte Aussage der anderen entgegen der eigentlichen Position ist
2.11	Interpret	Paraphrase
2.12	Meta	Äußerung über die Diskussion selbst (nicht deren Inhalt)
2.13	Null	Äußerung die nicht zum Thema gehört
2.14	Refuse	Explizites Ablehnen auf eine Frage zu antworten
2.15	Substantiate	Eine Äußerung die anbietet die zuvor getätigte Aussage der anderen zu unterstützen
<b>Nontransactive statements</b>		
3.1	Counter-A	Gegenargument
3.2	Continue	Weiterführen der zuvor getätigten eigenen Aussage ohne darauf zu achten was andere in der Zwischenzeit gesagt haben
3.3	Unconnected	Eine Aussage die sich nicht auf die zuvor getätigten Äußerungen der anderen bezieht

## C. Trainingsmaterialien:

### C.1. Ablauf

#### Tag 1

##### Teil 1: Motivation und Verständnis der Relevanz evidenzbasierter Praxis

1. Problembasierter Einstieg: Diverse Beispiele aus den Medien (z.B. „Digitale Demenz“, „Supersize me“)
2. Was ist Evidenz?
3. Wichtigkeit von Evidenz
  - a. *Evidenz im Alltag: Fehlinterpretationen von Studien in den Medien*
  - b. Evidenz im professionellen Umfeld: Forderungen in Bildungsstandards und EWS Examen
4. Kennenlernen des Ablaufs einer empirischen Studie
  - a. Modell idealisierter Forschungsablauf
  - b. Anschauungsbeispiele zweier empirischer Studien: BiWi und GeWi
5. Lernziele des Trainings (→Einklang von Expertise und Evidenz)

##### Teil 2: Evidenz kritisch hinterfragen und bewerten

1. Problembasierter Einstieg: Anschauliche Bilder und 1 Video warum Wissen über Statistik und Forschungsmethoden wichtig sein kann → Relevanz
2. *Über den gesamten Teil 2 hinweg:*  
*Fallbeispiele:* 2 empirische Studien (je eine aus den Bildungs- und eine aus den Gesundheitswissenschaften)  
*Methoden:* Problem- und Fallbasiertes Lernen, Cognitive Apprenticeship & Kooperatives Lernen
3. **Fragestellung**
  - a. Was ist der Sinn einer Forschungsfrage?
  - b. Was ist eine gute Forschungsfrage?
  - c. Welche Typen von Forschungsfragen gibt es?
  - d. Überblick über die Heuristiken zur kritischen Bewertung der Forschungsfragen
  - e. Kritisches Bewerten und anschließende Diskussion der beiden vorliegenden empirischen Studien hinsichtlich der Forschungsfragen
4. **Methoden und Design**
  - a. **Design**
    - i. Was ist ein Design?
    - ii. (GA): *Poster gestalten zu 3 zentralen Designs:* Experiment und Quasi-Experiment; Längsschnittstudie; Querschnittstudie
  - b. **Stichprobe**

- i. Einstiegsbeispiel (problemorientierter Einstieg) warum ist es so wichtig die Stichprobe korrekt zu definieren und zu wählen?
- ii. Was ist eine Grundgesamtheit? Erarbeiten mehrerer Fallbeispiele
- iii. Wann ist eine Stichprobe repräsentativ?
- iv. Was versteht man unter Verzerrung oder Bias?
- v. Was ist die ideale Stichprobengröße
- vi. Überblick über die Heuristiken zur kritischen Bewertung der Forschungsdesigns und Methoden
- vii. Kritisches Bewerten und anschließende Diskussion der beiden vorliegenden empirischen Studien hinsichtlich der Forschungsdesigns und Methoden

#### c. Instrumente

- i. Wie kann man etwas Messen?
- ii. Welche Art von Daten können gemessen werden: Selbstauskunfts- versus Leistungsdaten / Verhaltensdaten
- iii. Passung zur Fragestellung: Ist die Wahl der Instrumente plausibel? Kann man die Fragestellung damit beantworten?
- iv. Testgütekriterien
- v. Welche Fehler und Probleme können auftreten?
- vi. Überblick über die Heuristiken zur kritischen Bewertung der Instrumente
- vii. Kritisches Bewerten und anschließende Diskussion der beiden vorliegenden empirischen Studien hinsichtlich der Instrumente

### 5. Ergebnisse

- a. Was ist wichtig beim Ergebnisteil?
- b. Wichtige Begrifflichkeiten:
  - i. Mittelwert, Median, Standardabweichung
  - ii. Statistische Signifikanz
  - iii. Korrelation
  - iv. Effektstärke
- c. Transparenz
- d. Häufige Fehler und Probleme, Videobeispiel über „das Storchproblem“
- e. Überblick über die Heuristiken zur kritischen Bewertung der Ergebnisse
- f. Kritisches Bewerten und anschließende Diskussion der beiden vorliegenden empirischen Studien hinsichtlich der Ergebnisse

### 6. Schlussfolgerungen/ Diskussion

- a. Was ist wichtig beim Diskussionsteil?
  - i. Beantwortung der Fragestellung
  - ii. Limitationen
  - iii. Generalisierungen
- b. Wann sind Generalisierungen unzulässig?
- c. Überblick über die Heuristiken zur kritischen Bewertung der Diskussion

- d. Kritisches Bewerten und anschließende Diskussion der beiden vorliegenden empirischen Studien hinsichtlich der Diskussionen

**7. Diskussion und Vergleich** beider Studien im Ganzen hinsichtlich ihrer Belastbarkeit und praktischer Nutzbarkeit

**8. Zusammenfassung und Feedback**

### Tag 2

1. Eventuell aufgetauchte Fragen?
2. Problemorientiertes Beispiel zum Einstieg (z.B. „Demand Evidence“)

### Teil 3: Evidenz argumentativ nutzen

1. Was ist gutes wissenschaftliches Argumentieren?
2. Toulmin Schema
3. Ansätze nach Kuhn
4. Gruppendiskussionen (kooperatives lernen):
  - a. *(Einverständniserklärung zur Videoaufzeichnung)*
  - b. Kooperationskripts
  - c. Einteilung der Gruppen
  - d. Diskussion 1: Überhöhte Salzzufuhr und Herz-Kreislauf Erkrankungen
    - i. Zeitungsartikel „Salz, tödliche Gefahr fürs Herz“
    - ii. Fallvignette
    - iii. Zuweisung der Artikel
    - iv. Zeit zur individuellen Lektüre und kritischen Bewertung
    - v. Start der Diskussion
  - e. Diskussion 2: Genderunterschiede bei der Computernutzung im Unterricht
    - i. Foto „Lehrerkonferenz“
    - ii. Fallvignette
    - iii. Zuweisung der Artikel
    - iv. Zeit zur individuellen Lektüre und kritischen Bewertung
    - v. Start der Diskussion
  - f. Gesamtdiskussion und Austausch zwischen den Gruppen: Ergebnisse, Belastbarkeit, praktischer Nutzen
5. Reflexion und Diskussion

## C.2. Checkliste (Heuristiken)

### Fragestellung

	Ja	Nein	k.a.
Ist die Fragestellung klar und präzise formuliert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ist sie relevant für ein spezielles Problem, für die Praxis, für die Theorieentwicklung, ...?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Um welchen Typ (oder welche Typen) Fragestellung(en) handelt es sich?			
- Deskriptiv	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- Zusammenhänge	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- Ursache-Wirkung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Wie lautet/lauten die konkrete(n) Fragestellung(en)?

---



---

### Ergebnisse

	Ja	Nein	k.a.
Gibt es mindestens einen Beleg /eine Evidenz pro Fragestellung?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sind die Ergebnisse signifikant?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Falls die Ergebnisse Signifikant sind, sind sie dies in erwarteter Richtung?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gibt es Aussagen über die Größe des Effektes?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Falls es Aussagen über die Größe des Effektes gibt, ist diese mittel bis groß?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Werden die Ergebnisse klar, objektiv und nachvollziehbar dargelegt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Wie lauten die konkreten Ergebnisse?

---



---

### Methode

	Ja	Nein	k.a.
Um welches Design handelt es sich?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- Längsschnittstudie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- Querschnittstudie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- (Quasi-)Experiment?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Passt das gewählte Design zur Fragestellung?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Welche Daten wurden erhoben?			
- Selbstauskunftsdaten	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- Leistungsdaten	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- Verhaltensdaten	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ist die Art der gesammelten Daten der Fragestellung angepasst? (Selbstauskunfts- (subjektiv) versus Leistungsdaten (objektiv))	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ist die Wahl der Instrumente plausibel?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Es gibt Informationen über die Validität des Instruments?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Es gibt Informationen über die Reliabilität des Instruments (z. B. in Form von Cronbachs Alpha)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ist die Stichprobengröße der Fragestellung angemessen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ist die Zusammensetzung der Stichprobe angemessen (z. B. Alter, Geschlecht, ...)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ist die Stichprobe repräsentativ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### Diskussion

	Ja	Nein	k.a.
Werden die Einschränkungen (Limitationen) adäquat benannt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Werden die Fragestellungen beantwortet?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ist die Antwort auf Basis der Ergebnisse angemessen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Es gibt keine unzulässigen Generalisierungen (bezüglich Kausalität (bezogen auf N), Population oder Situation)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Was ist die Kernaussage des Textes?

---



---

