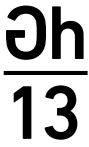


Waldis, Nitsche (Hrsg.)

Geschichtsdidaktisch intervenieren



Geschichtsdidaktik heute
www.geschichtsdidaktik-heute.ch

In der Reihe «Geschichtsdidaktik heute» werden Arbeiten veröffentlicht, die sich mit innovativen Ansätzen und methodischer Klarheit der empirischen Erforschung theoretischer Konzepte sowie der Rückbindung empirischer Befunde an die Theoriebildung im Feld der Geschichtsdidaktik widmen.

Herausgeberschaft der Reihe «Geschichtsdidaktik heute» im hep Verlag

Prof. em. Dr. Béatrice Ziegler

Politische Bildung und Geschichtsdidaktik, Pädagogische Hochschule FHNW am ZDA Aarau

Prof. Dr. Markus Furrer

Professor für Geschichte und Geschichtsdidaktik an der Pädagogischen Hochschule Luzern

Stephan Hediger

Bereichsleiter Geografie, Geschichte, Religion und Kultur, Pädagogische Hochschule Zürich

Dr. Nadine Ritzer

Dozentin am Institut Sekundarstufe I, Pädagogische Hochschule Bern

Monika Waldis, Martin Nitsche (Hrsg.)

Geschichtsdidaktisch intervenieren



Publiziert mit Unterstützung des Schweizerischen Nationalfonds zur Förderung
der wissenschaftlichen Forschung

Publiziert von
hep Verlag AG
Gutenbergstrasse 31
3011 Bern
hep-verlag.com

Text © Burghard Barte, Helene Bergmann, Christiane Bertram, Andrea Brait,
Christoph Bramann, Nicola Brauch, Daniel Fastlabend, Kristine Gollin,
Christian Heuer, Manuel S. Hubacher, Alexandra Krebs Hendrik Lohse-Bossenz,
Georg Marschnig, Philipp Marti, Johannes Meyer-Hamme, Marcel Mierwald,
Miriam Moser, Martin Nitsche, Christine Pauli, Mario Resch, Kurt Reusser,
Johanna Sachse, Ulrich Trautwein, Wolfgang Wagner, Julia Waldeyer, Monika
Waldis, Anke Wischgoll, Carina Wöhlke, Stefanie Zabold, Lisa Zachrich,
Matthias Zimmerman 2023

Umschlagbild: Debby Hudson/Unsplash

ISBN (Print): 978-3-0355-2191-7
ISBN (PDF): 978-3-0355-2192-4

DOI: <https://doi.org/10.36933/9783035521917>



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0
International Lizenz.

Inhaltsverzeichnis

Einleitung: Geschichtsdidaktisch intervenieren – eine Standortbestimmung	9
<i>Monika Waldis und Martin Nitsche</i>	
Quasiexperimentelle Interventionen	23
Was wirkt? Ein Forschungsprogramm zur Wirksamkeit von Zeitzeug*innen im Geschichtsunterricht	24
<i>Christiane Bertram</i>	
Leseförderung im Geschichtsunterricht? – Eine Interventionsstudie zur Wirksamkeit adaptiver Sprachunterstützung bei der Erschließung fachspezifischer Sachtexte	46
<i>Helene Bergmann</i>	
Welche Hilfen könnten helfen? – Einblick in eine Interventionsstudie zur Förderung des historischen Erklärens	69
<i>Christoph Bramann, Carina Wöhlke und Nicola Brauch</i>	
Narrative Kompetenz und epistemologische Beliefs im Geschichts- unterricht fördern? – Zur Wirksamkeit einer Schreibintervention an Deutschschweizer Gymnasien	97
<i>Martin Nitsche, Monika Waldis, Kristine Gollin, Philipp Marti und Manuel S. Hubacher</i>	
Eine Intervention zur Förderung geschichtsdidaktischen Wissens und Könnens zur Konstruktion von Aufgaben für den Geschichtsunterricht	128
<i>Mario Resch, Christian Heuer und Hendrik Lohse-Bossenz</i>	

Qualitative «Interventionen»	153
Multiperspektivität schriftlich erarbeiten – Mit Textprozeduren erzählen lernen	154
<i>Georg Marschnig</i>	
Von Stacheldraht bis Nivea-Dose – Wie die Durchführung einer archäologischen Ausgrabung bei Schüler*innen Konzeptveränderungen des historischen Denkens anstoßen kann	171
<i>Johanna Sachse</i>	
Dialogische Klassengespräche führen – Fallanalyse zur Entwicklung einer professionellen Kompetenz von Geschichtslehrpersonen	191
<i>Matthias Zimmermann, Miriam Moser, Anke Wischgoll, Kurt Reusser und Christine Pauli</i>	
Grundlagen für geschichtsdidaktische Interventionen	215
Concept-Mapping und konstruktivistische Grounded-Theory-Methodologie – Mit semantischen Netzen historisches Wissen erfassen	216
<i>Burghard Barte</i>	
HiTCH II – Was wir aus der Weiterentwicklung des HiTCH-Tests über historische Kompetenzen lernen	238
<i>Wolfgang Wagner, Christiane Bertram und das HiTCH-Konsortium</i>	
Historisches Lernen mit Originalen? – Zur Umsetzung einer Quellenorientierung in Museen	259
<i>Andrea Brait</i>	
«Fotos, die jede*r kennt?» – Eine empirische Untersuchung historischer Kenntnisse anhand von Bildikonen bei Studierenden des Lehramts Geschichte	279
<i>Daniel Fastlabend und Johannes Meyer-Hamme</i>	

Historisches Lernen mit der App in die Geschichte – Fallbeispiele einer Vorstudie zur Untersuchung historischer Lernprozesse und Narrationen von Studierenden im digitalen Medium 304
Alexandra Krebs

Kann man die Relevanz von Geschichte messen? – Erste Validierung der deutschsprachigen Version der Relevance of History Measurement Scale (RHMS-D) 331
Marcel Mierwald und Julia Waldeyer

GeDiKo-T – Zur Entwicklung und Implementation eines Kompetenztests für Geschichtslehrpersonen 351
Stefanie Zabold

Das «Aura-Erlebnis» in der Begegnung mit Zeitzeug*innen – Eine komplexe Lernerfahrung verstehen und standardisiert erfassen 375
Lisa Zachrich, Wolfgang Wagner, Christiane Bertram und Ulrich Trautwein

Abstracts 398

Biografische Angaben 406

Einleitung

Geschichtsdidaktisch intervenieren – eine Standortbestimmung

Monika Waldis und Martin Nitsche

1 Begriffsklärung

Geschichtsdidaktik beschäftigt sich unter anderem mit der Frage, wie Umgangsweisen mit Geschichte(n) und Vergangenen im Kontext von Schulen, Hochschulen und außerschulischen Lernorten gestaltet werden können. Lernende sollen vor dem Hintergrund aktueller gesellschaftlicher Entwicklungen und Diskurse wie beispielsweise Globalisierung und globaler Konflikte, Digitalisierung, Klimawandel sowie Migration historisch denken lernen, um gegenwärtige und zukünftige Herausforderungen bewältigen zu können. Sowohl formale Bildungsprozesse in der Schule als auch nonformale Bildungsprozesse an Orten wie Museen, Gedenkstätten, archäologischen Ausgrabungsstätten etc. haben die Förderung historischen Denkens zum Ziel. In einem breiten Sinn können sie demnach als «Intervention» verstanden werden; im Prinzip reichen aber auch Begriffe wie «historische Bildung», «Geschichtsunterricht», «Geschichtsvermittlung» oder «Geschichtsbegegnung» dafür aus. Fest steht, dass es dabei nicht nur um das Anstoßen individueller Lernprozesse geht, sondern ebenfalls um Enkulturation, also das Erlernen grundlegender kultureller Fertigkeiten und Verständnisformen von Welt sowie um gesellschaftliche Teilhabe beziehungsweise im Umkehrschluss um die Legitimation gesellschaftlicher Strukturen unter Einbeziehung historischer Perspektiven. Bildungsinstitutionen reproduzieren Normen, Werte und Weltansichten, die zur Stabilisierung der sozialen und politischen Verhältnisse dienen (Fend, 2011, 42). Vom Bildungssystem im Allgemeinen und von der Geschichtsvermittlung in schulischen und außerschulischen Settings im Speziellen werden demnach die Organisation von Lehr-Lern-Settings beziehungsweise der Begegnung mit dem Historischen mit Auswirkungen auf individueller,

institutioneller (z. B. Klassen, Schule) sowie gesamtgesellschaftlicher Ebene (z. B. gesellschaftliche Beteiligung) erwartet.

Bezüglich aller genannten Ebenen hat Lücke (2020) exemplarisch auf die Interventionsabsicht der «Holocaust Education» verwiesen.¹ Dieser liegt die Absicht zugrunde, gesellschaftlich zu intervenieren, um erstens Wissen über den Holocaust zu vermitteln, zweitens gesellschaftliche Solidarität zu sichern und drittens Identitätsbildung und historische Orientierung zu fördern (vgl. Lücke, 2020). Darüber hinaus soll etwas Unerwünschtes, nämlich vergleichbare Verbrechen, durch geeignete Maßnahmen zukünftig abgewendet werden. Erklärtes Ziel der «Holocaust Education» ist demnach *Prävention* (lat. *praevenire* für zuvorkommen, verhüten). Demgegenüber geht es beim Begriff «Intervention», der auf das lateinische Verb *intervenire* (dazwischentreten, sich einschalten) zurückgeht, generell um Vorgehensweisen, mit der in einen laufenden Prozess eingegriffen wird (Leutner, 2013, 17–18). In der Regel werden hierzu Bildungsmaßnahmen gewählt, von denen aus guten Gründen angenommen werden kann, dass sie die erwünschte Wirkung zeigen. Unter anderem beeinflussen hier (historische) Kontexte vor Ort, pädagogische Traditionen, Wissen und Erfahrungen von Lehrpersonen, Kurator*innen und Museumspädagog*innen sowie Reflexionen anhand von Forschungsbefunden, die punktuell in ähnlichen Schüler*innen- beziehungsweise Besucher*innengruppen erhoben wurden, die Ausrichtung pädagogischen Handelns. Eine adressatenspezifische Diagnostik entfällt.

Einem solch breitflächigen Interventionsbegriff stellt die pädagogische Interventionsforschung das Konzept einer gezielten Veränderung beziehungsweise eines gezielten Eingriffs ins System gegenüber. Im Idealfall besteht dieser darin, dass die Adressat*innen zuerst zu Vorwissen und Vorerfahrungen befragt und getestet werden, um darauf aufbauend Entwicklungsziele zu definieren und Maßnahmen zu ergreifen. Sodann wird in der Bildungsforschung von «Interventionsforschung» gesprochen, «wenn man sich wissenschaftlicher Methoden bedient, um die Wirkung einer Intervention zu evaluieren – um also festzustellen,

1 Der vorliegende Sammelband ist thematisch von der Tagung *Geschichtsdidaktik empirisch 20* inspiriert, die coronabedingt lediglich als «Light»-Version digital durchgeführt werden konnte. In diesem Rahmen gaben Martin Lücke, Christiane Bertram und Jannet van Drie am 3. September 2020 in je zwanzigminütigen Referaten Einblicke in ihre Forschungsarbeiten und Reflexionen zu Fragestellungen geschichtsdidaktischer Interventionen, auf die hier an verschiedenen Stellen Bezug genommen wird. Der Sammelband bündelt Beiträge von Kolleg*innen, die aufgrund des genannten Anlasses an einer thematischen Vertiefung interessiert waren.

ob die mit der Intervention verfolgten Ziele tatsächlich auch erreicht werden. Bei pädagogischer Interventionsforschung geht es um die Prüfung gezielter pädagogischer Maßnahmen auf deren erwartete pädagogische Wirkung» (Leutner, 2013, 19). Damit rückt die Notwendigkeit der Diagnose in den Vordergrund, welche genügend sensitiv ist, um Lernstände und -zuwächse, Interesse und Lernmotivation, Erweiterungen von Handlungsrountinen, Veränderungen der Einstellungen, epistemologischen Überzeugungen oder des moralischen Bewusstseins etc. festzustellen. Im Unterschied zu deskriptiven Forschungszugängen, welche die bereits etablierte Praxis beschreiben und analysieren, setzt Interventionsforschung «bei einer theoriebasierten Entwicklung und Erprobung innovativer Konzepte an, die zu einer Erweiterung und Erneuerung des Spektrums unterrichtlichen Handelns führen soll» (Souvignier, 2021, 2). Ziel pädagogischer Interventionsforschung ist somit Innovation, gekoppelt mit der Absicht, empirisch belastbare Erkenntnisse zu gewinnen, die für die zukünftige Gestaltung der Vermittlungspraxis herangezogen werden können. Geschichtsdidaktisch gewendet kann somit allgemein definiert werden: Geschichtsdidaktische Interventionsforschung untersucht intentionale Veränderungsmaßnahmen im Rahmen historischer Bildungskontexte auf deren erwartete Wirkung hinsichtlich geschichtsdidaktisch begründeter Zielvorstellungen. Eine solch eng gewählte Definition dient der Schärfung des Interventionsbegriffs beziehungsweise der Prävention von dessen Verwässerung mit deskriptiven Forschungsansätzen. Letztere werden in den Fachdidaktiken häufig gewählt, um didaktische Neuerungen und damit angestoßene Lernprozesse zu beschreiben. Sofern keine Vergleichsgruppe besteht, lassen sich auf diese Weise zwar Lernzuwächse beobachten, jedoch keine kausalen Schlüsse zur Wirksamkeit der gewählten Maßnahmen ziehen. Solche Studien erarbeiten idealerweise wissenschaftliche Grundlagen für eine erfolgreiche Intervention. Dazu gehört etwa die Bearbeitung der Frage, unter welchen Umständen die Anwendung einer bestimmten Maßnahme ratsam erscheint beziehungsweise gewinnbringend eingesetzt werden könnte (Leutner, 2013, 20).

2 Methodische Grundsätze der Interventionsforschung

Die pädagogische beziehungsweise geschichtsdidaktische Interventionsforschung nutzt für die Feststellung von Wirksamkeit methodische Werkzeuge der Sozialwissenschaften. Die eingesetzten Forschungsdesigns orientierten sich zu-

nächst an klinischen Studien der Psychologie und Humanmedizin. Allerdings werden experimentelle Interventionen im pädagogischen Bereich durch eine Vielzahl von Kontextfaktoren beeinflusst, die auf die eine oder andere Weise erfasst und mitberücksichtigt werden müssen. Als Standardvorgehen zur Prüfung der Wirksamkeit einer Intervention werden derzeit Studien mit Experimental- und Kontrollgruppenvergleichen durchgeführt. Erreicht eine mit der Kontrollgruppe vergleichbare Experimentalgruppe *nach* der Intervention den statistisch besseren Mittelwert in der Zielvariablen, so wird geschlossen, dass die Intervention wirkt (vgl. Gopalan, Rosinger & Ahn, 2020, 24). Grundsätzlich werden zwei methodische Vorgehensweisen der Interventionsforschung unterschieden. Erstens sind «experimentelle Studien» beziehungsweise «*randomized-controlled studies*» auszumachen, bei denen Proband*innen, Schulen oder Schuldistrikte den verschiedenen Interventionsbedingungen zufällig zugeordnet und somit überzufällige Effekte aufgrund bestimmter Gruppenmerkmale vermieden werden. Im Anschluss können kausale Zusammenhänge zwischen Maßnahme und Effekt ermittelt werden (Gaspard & Trautwein, 2022; Slavin, 2002). Eine zufällige Zuordnung von Individuen zu Interventionsbedingungen ist in Bildungsprozessen allerdings häufig nicht möglich. So müssen oftmals aus organisatorischen Gründen Interventionen mit Schüler*innen im angestammten Klassenverband oder in bestimmten Schulen durchgeführt werden. Damit verbunden sind Selektionsmechanismen wie die Studienteilnahmebereitschaft von Lehrpersonen, Schultyp, Steuerkraft der Gemeinde etc. Entfällt die Möglichkeit der Zuordnung zur Interventionsbedingung nach Zufall, so können zweitens «quasi-experimentelle Studien» durchgeführt werden. In diesen werden persönliche und gruppenbezogene Merkmale, die mit der durchgeführten Intervention in Zusammenhang stehen und demzufolge die abhängige Variable beeinflussen könnten, statistisch kontrolliert (Gopalan et al., 2020; Legewie, 2012).

Die Absicherung von standardisierten Bedingungen in experimentellen Studien geht mit hohen methodologischen Ansprüchen an die eingesetzten Forschungsinstrumente und Analysemethoden einher. Interventionsstudien benötigen 1) vorausgehend eine sorgfältige Diagnose des Ist-Zustandes des pädagogischen Handelns, 2) die Definition von Interventionszielen und die Entwicklung von innovativen fachdidaktischen sowie generischen Ansätzen zu deren Erreichung, 3) Forschungsinstrumente, die für die anzustrebenden Veränderungen sensitiv sind, 4) ein standardisiertes Vorgehen, das zumindest von Prä- zu Posttest durchgehalten wird und aus ethischen Gesichtspunkten gutgeheißen werden kann (Stichwort: Chancengerechtigkeit), 5) die zufällige Zuweisung zu den Bedingungen im Falle von «*randomized*

controlled studies» (Bertram, in diesem Band; Gaspard & Trautwein, 2021) oder die Einbeziehung von Kontrollvariablen zur Überprüfung der aufgrund der Logiken des Feldes entstandenen Zuordnung der Beteiligten (z.B. Klassengruppierung) zur Durchführung von «quasi-experimentellen Studien» (Waldis, 2022, 304), 6) eine dem Interventionsplan getreue Umsetzung der Intervention und deren sorgfältige Kontrolle (*treatment fidelity*, vgl. Sanetti, Cook & Cook, 2021), 7) dem Datenmaterial und der Stichprobe angemessene Analysen, sowie 8) eine Diskussion der Ergebnisse unter Einbeziehung inhaltlicher sowie methodischer Gütekriterien mitsamt der Reflexionen von Limitationen der Resultate und der Diskussion von Transfermöglichkeiten auf andere Bildungskontexte. Herausgehoben werden muss, dass eine angemessene Interpretation der Ergebnisse nur vor dem Hintergrund der realisierten Untersuchungsdesigns und weiteren Kontextinformationen sinnvoll ist (Souvignier, 2021, 6). Insbesondere die Einhaltung der Implementationstreue (*treatment fidelity*) entscheidet darüber, ob die gefundenen kausalen Zusammenhänge tatsächlich der durchgeführten Intervention zugeschrieben werden können. Fällt die Implementationstreue gering aus, ergeben sich uneindeutige Resultate, da möglicherweise weitere, unbeabsichtigte Faktoren eine Wirksamkeit entfalten (Sanetti et al., 2021). Forschungsüberblicke und Metastudien verweisen darüber hinaus darauf, dass die ermittelten Effekte in Abhängigkeit von den methodologischen Rahmenbedingungen (z.B. Stichprobenziehung, Anzahl Proband*innen) variieren, sodass die Beachtung von methodischen Standards für eine sinnvolle Interpretation von Interventionsstudien unabdingbar ist (Cheung & Slavin, 2016).

Experimentelle Studien, die auf eine Kontrolle des Kontextes und den Ausschluss von Einflussfaktoren ausgelegt sind, stehen oftmals vor der Herausforderung, sowohl den Beteiligten in deren Kontexten als auch den dargestellten methodologischen Ansprüchen gerecht werden zu müssen. Zum Beispiel wäre es aus experimenteller Sicht manchmal sinnvoll, eine längere Interventionsdauer einzuplanen und zusätzlich eine Nachfolgeerhebung (*follow-up*) einige Monate später durchzuführen, um Auskunft über die längerfristige Wirkung einer Intervention zu erhalten. In der Praxis steht die dafür benötigte Zeit jedoch oftmals kaum zur Verfügung. Dauert eine Intervention länger an, steigt zudem die Wahrscheinlichkeit, dass gesellschaftspolitische Kontexte und aktuelle Entwicklungen eine neue Komponente entfalten, man denke beispielsweise an die Coronapandemie der letzten zwei Jahre oder den Ukrainekrieg. Solche Entwicklungen können unter Umständen eine geplante Intervention beziehungsweise die darin eingesetzten Lern- und Testaufgaben beeinflussen, insbesondere dann, wenn offene Aufgabenstellungen vorliegen und lebensweltliche Bezüge erwünscht sind.

Dennoch scheint es in geschichtsdidaktischer Hinsicht sinnvoll und zielführend zu sein, mithilfe von Interventionsstudien genauer hinzuschauen, was mit welchen Inszenierungen, instruktionalen Aktivitäten und weiteren Förderstrategien ausgelöst und bewirkt werden kann. Zum Beispiel verdeutlicht der Blick auf den wachsenden Korpus von «quasi-experimentellen» Interventionsstudien zum historischen Lesen und Schreiben, dass die Strategievermittlung zum Umgang mit geschichtsspezifischen Herausforderungen (z. B. Analyse und Interpretation von Medien) anhand einer zentralen historischen Frage für Lernende mit und ohne Förderbedarf sowie auf unterschiedlichen Schulstufen wirksam ist (Waldis, 2022, 306). Geschichtsdidaktische Interventionsforschung unter dem «Paradigma der Evidenzbasierung» kann demzufolge trotz der beschriebenen «Komplexität» einen relevanten Beitrag zur Theoriebildung und zur Professionalisierung geschichtsdidaktischer Handelnder leisten.

Deutlich verschiedenen von diesen experimentellen Zugängen werden neuerdings auch an Designzirkeln orientierte Entwicklungen pädagogischer (Unterrichts-)Praxis vorgenommen. Bei einem solchen Ansatz im Sinn des «*design-based research*» geht es darum, zyklisch und adaptiv sogenannte «lokale» Lösungen für ein Praxisproblem zu entwickeln, welche nach Erprobungen und Anpassungen in einer späteren Phase allenfalls auf Effizienz und Wirksamkeit in einer breiteren Stichprobe überprüft werden können (Reinmann, 2005). *Design-based research* erfüllt in der Regel die Kriterien der Bestandesaufnahme (Diagnostik) und gezielten Veränderung. Unterschiede bestehen im Anspruch der Reichweite von Aussagen zur Wirksamkeit. Auch hier müssen Standards für die Evaluation und der Umgang mit differentiellen Effekten wie beispielsweise unterschiedliche Wirkungen in verschiedenen Adressat*innengruppen sowie Voraussetzungen für Implementation und Transfer diskutiert und geklärt werden.

3 Offene normative Fragen

Neben den oben dargestellten methodologischen und organisatorischen Anforderungen an Interventionsstudien stellen sich eine Reihe von theoretischen und normativen Fragen. Zum Beispiel fragte Lücke (2020) in seiner Auseinandersetzung mit Chancen und Grenzen geschichtsdidaktischer Interventionen unter dem Stichwort Kultur etwa, wo die normativen Grenzen geschichtsdidaktischer Interventionen liegen, wenn in deren Rahmen auf geschichtskulturell geprägtes Wissen zugegriffen und eine Auswahl der Interventionsthemen mit Blick auf le-

bensweltliche Relevanz vorgenommen wird. So müssen auf jeden Fall leitende Normen und Werte reflektiert werden. Ebenso sind normative Grenzen mit Blick auf Selbstbestimmung und Prozesse der Identitätsfindung und -behauptung einzuhalten. Für politisch-historische Lernprozesse in Gedenkstätten beispielsweise wurden hierzu die im Beutelsbacher Konsens festgehaltenen drei Prinzipien Überwältigungsverbot, Kontroversität und Schüler*innenorientierung ausdifferenziert (Müller & Ruppert-Kelly, 2016). Diesbezügliche Problematisierungen finden sich darüber hinaus in der geschichtsdidaktischen Theoriebildung etwa zu «Geschichtsbewusstsein», «historische Orientierung» und «Triftigkeit», indem der Eurozentrismus und die westliche Tradition der genannten Konzepte kritisiert wird, weshalb die Eignung zur Beschreibung und Normierung historischen Denkens in diversen Gesellschaften fraglich sei (z. B. Völkel, 2015).

Forschungspraktisch ergibt sich des Weiteren die Herausforderung, dass bei Interventionsmaßnahmen auf vorliegende Kontexte Rücksicht genommen werden muss. So sollten die geplanten Maßnahmen an die Lebenswelt aller Beteiligten sowie an die Handlungskompetenz der teilnehmenden Geschichtsmittler*innen sowie deren gängige Vermittlungspraxis anknüpfen. Diese kann sich unter Umständen von geschichtsdidaktischen Idealvorstellungen deutlich unterscheiden (z. B. positivistisch geprägte Wissensvermittlung). In der Folge müssen allenfalls Zwischenziele in den Blick genommen werden, die geschichtstheoretisch und -didaktisch weniger gut legitimiert werden können oder von denen unsicher ist, inwiefern sie mit den idealerweise zu erreichenden Zielen in Verbindung stehen. In diesem Zusammenhang machte Jannet van Drie (2020) auf die Rolle des *«first order knowledge»* als mögliche Voraussetzung von *«historical reasoning»* aufmerksam (vgl. van Boxtel & van Drie, 2018), dessen Rolle in der geschichtsdidaktischen Theoriebildung eher selten thematisiert wird, was es erschwert, diesbezügliche Interventionsmaßnahmen zu begründen. Ganz allgemein sei hier angemerkt, dass gerade in Geschichte theoriebasierte Lernziele, empirische Befunde zum Handeln informierter Laien sowie Beobachtungen zu Expert*innenhandeln durchwegs differieren können, was in Interventionsstudien die Frage nach der Legitimation wissenschaftsbasierten pädagogischen Handelns akzentuiert. Zu diskutieren ist darüber hinaus, inwiefern das Kriterium der bildungspraktischen Relevanz für Entscheidungen zu Interventionszielen und -ablauf ausschlagend sein sollten.

Schließlich ergeben sich weitere methodisch-normative Fragen. Eine wichtige Frage betrifft die Ausgestaltung der Kontrollbedingung, insbesondere bei unterrichtsbezogenen Interventionen. Werden Prä- und Posttests in Kontroll-

klassen durchgeführt, die ansonsten dem üblichen Unterricht folgen, kann dies die Testmotivation negativ beeinflussen, weil die Wiederholung eines identischen oder ähnlichen Tests für die Proband*innen sinnlos erscheint. Alternativ dazu besteht die Möglichkeit, in der Kontrollbedingung gewisse inhaltliche oder methodische Parallelen zu schaffen, die nahe am bisherigen pädagogischen Handeln orientiert sind und gleichzeitig eine zweite Messung begründungsfähiger machen. Allerdings muss auch dann aus forschungsethischen Gründen eine möglichst transparente Information der Beteiligten erfolgen. Darüber hinaus sollte die Ethik «guten Unterrichts» umgesetzt und eine gewisse Implementationsstreue garantiert sein, um mögliche Effekte einordnen und erklären zu können. Bei der Planung einer Interventionsstudie ist zudem darauf zu achten, dass zum Schluss explizite Hinweise zur Wirksamkeit einer fachdidaktischen Innovation vorliegen, die über den bloßen Aspekt der deutlich längeren «*time on task*» in der Experimentalgruppe hinausgehen. Es sind somit geschichtsdidaktische Impulse gefragt, welche eine qualitative Verbesserung historischen Denkens anstoßen können, respektive geschichtsdidaktische Kriterien und Erhebungsinstrumente, die entsprechende Veränderungen erfassen können. Noch wenig diskutiert ist derzeit, ob es neben niedrigen oder ausgebliebenen Interventionseffekten weitere Hinweise braucht, um eine didaktische Innovation als gescheitert oder unbrauchbar zu bewerten. Schließlich benötigt auch die Frage nach der hinreichenden Evidenz weitere Diskussionen, um zentrale Wirkmerkmale mit hoher Sicherheit benennen zu können. Wir hoffen, mit dem vorliegenden Sammelband, der Forschungsartikel zu aktuellen Interventionsstudien in der deutschsprachigen Geschichtsdidaktik enthält, Impulse für diesbezügliche Auseinandersetzungen zu geben.

4 Die Beiträge

Das vorliegende Buch versammelt Studien, in deren Rahmen Geschichtsdidaktiker*innen gezielt in historische Bildungsprozesse eingegriffen haben. Die Studien sind in den drei Teilen «Quasiexperimentelle Studien», «Qualitative Interventionen» und «Grundlagen für geschichtsdidaktische Interventionen» verortet, um eine Orientierung anhand ihrer Schwerpunkte zu ermöglichen. Den ersten Teil eröffnet *Christiane Bertrams* Beitrag, indem sie ihr Forschungsprogramm zur Wirksamkeit von Zeitzeug*innen im Geschichtsunterricht auf unterschiedlichen Klassenstufen an deutschen Schulen vorstellt, unter anderem auf

die theoretischen Grundlagen der genutzten «*randomized controlled field trials*» eingeht und vier geplante Forschungsprojekte genauer ausführt. Abschließend setzt sich die Autorin anhand ihrer Forschung mit Kritikpunkten an geschichts-didaktischer Interventionsforschung auseinander. Im zweiten Artikel stellt *Helene Bergmann* ihr Promotionsprojekt vor, in dessen Rahmen sie untersucht, welche Formen von Sprachunterstützung geeignet sind, um das Verständnis historischer Sachtexte bei deutschen Schüler*innen in der 5. Klassenstufe zu fördern. Dazu stellt die Autorin ihre Konzeption sowie erste Ergebnisse einer quasiexperimentellen Studie vor, in der die Effekte geschichtsspezifischer Scaffolding-Maßnahmen untersucht werden, welche anhand der Lern- und Sprachvoraussetzungen der Schüler*innen systematisch variiert wurden. *Christoph Bramann, Carina Wöhlke und Nicola Brauch* stellen in der dritten Studie erste Ergebnisse einer Intervention zur Förderung des historischen Erklärens vor. Sie analysieren unter anderem die Korrelationen zwischen drei Scaffoldingtools (fachlich, sprachlich, fachlich und sprachlich) und den Methoden der Überarbeitung von historischen Erklärungen in Drehbüchern zu Erklärvideos, die von einer Gruppe deutscher Schüler*innen der 9. und 11. Klassenstufe im Rahmen eines Lernlaborsetting entwickelt wurden. Die Autor*innen stellen zudem detaillierte Befunde zur Art der Nutzung der eingesetzten Scaffoldingtools dar. *Martin Nitsche, Monika Waldis, Kristine Gollin, Philipp Marti und Manuel S. Hubacher* thematisieren daraufhin, inwiefern eine geschichtsspezifische Schreibintervention an Deutschschweizer Gymnasien die narrative Kompetenz und die epistemologischen Beliefs von Schüler*innen auf der 10. und 11. Klassenstufe beeinflusste. Dazu illustrieren sie die eingesetzten Unterrichtsmethoden, Erhebungs- und Auswertungsverfahren und verorten die Ergebnisse im Kontext bisheriger Interventionsstudien für das Fach Geschichte. Schließlich stellen *Mario Resch, Christian Heuer und Hendrik Lohse-Bossenz* ihre Untersuchung zur Wirksamkeit geschichts-didaktischer Hochschullehre auf die Entwicklung geschichts-didaktischen Könnens zur Aufgabenkonstruktion sowie auf die Entwicklung von domänenspezifischem Fachwissen (Geschichtswissen und geschichts-didaktisches Wissen) vor. Interventionsteilnehmer*innen waren angehende deutsche Geschichtslehrpersonen. Die drei Autoren stellen das Interventions- und Studiendesign sowie die erreichten Effekte in der Versuchs- (mit Aufgabenfokus) und Kontrollgruppe (mit generellem geschichts-didaktischem Fokus) vor und diskutieren zukünftig notwendige Schritte der Forschung.

Den Teil «Qualitative Interventionen» eröffnet *Georg Marschnig*, indem er sein Habilitationsprojekt zur unterrichtspragmatischen Umsetzung des Konzeptes «Multiperspektivität» im Rahmen einer qualitativen Schreibintervention an österreichischen Gymnasien (Sekundarstufe II) vorstellt. Vertieft verdeutlicht er etwa das Aufgabendesign zur Erfassung des Umgangs mit Multiperspektivität oder das Unterrichtsvorgehen in der Versuchsgruppe mittels Trainings von Schreibprozeduren und stellt erste qualitative Resultate vor. Abschließend reflektiert der Autor seine Befunde und geht auf die Rolle der allgemeinen Sprachkompetenz ein. Nachfolgend stellt *Johanna Sachse* eine Teilstudie ihrer Dissertation vor, indem sie mittels Prä-Post-Interviews untersucht, inwiefern sich die Konzepte von deutschen Schüler*innen der 12. Klassenstufe unter anderem zu «Archäologie» aufgrund eines Ausgrabungsprojektes im KZ Neuengamme veränderten. Dazu knüpft die Autorin an die Conceptual-Change-Forschung an und stellt neben ersten Befunden ausführlich den Nutzen der «autofotografischen Methode» für ihre Forschung vor. *Matthias Zimmermann, Miriam Moser, Anke Wischgoll, Kurt Reusser und Christine Pauli* fokussieren in ihrer Fallanalyse Veränderungen der professionellen Kompetenz einer Schweizer Geschichtslehrperson aufgrund einer einjährigen Weiterbildung zur dialogischen Klassenführung. Mittels Videoanalysen des Geschichtsunterrichts, den die beteiligte Lehrperson auf der Sekundarstufe I («Grundanforderung») während der Studie durchführte, verdeutlichen die Autor*innen die Entwicklung des Gesprächsverhaltens der Lehrperson hinsichtlich Qualitätsmerkmalen dialogischer Gesprächsleitung sowie des argumentativen Verhaltens der Lernenden. Davon ausgehend leiten sie Schlussfolgerungen zur Wirksamkeit der Fortbildung ab.

Der Teil «Grundlagen für geschichtsdidaktische Interventionen» wird von *Burghard Barte* eröffnet, welcher anhand seines Promotionsprojektes die Methode des Concept-Mapping auf Grundlage der konstruktivistischen Grounded-Theory-Methodologie theoretisch und methodologisch begründet, um anhand von Daten von deutschen Studierenden (Lehramt und Fachmaster) vor und nach dem Abschluss von Mediävistik-Seminaren zu verdeutlichen, wie sich damit historisches Wissen und dessen Vernetzung sowie diesbezügliche Veränderungen untersuchen lassen. Abschließend diskutiert der Autor Chancen (z.B. Vernetzung) und Grenzen (z.B. Einfluss auf Erhebungspersonen) der Methode für die Erfassung des Wandels historischen Wissens. Anschließend stellen *Christiane Bertram und Wolfgang Wagner* ihre erweiterte Analyse der Validierungsstudie des HiTCH-Tests dar, indem sie anhand eines Samples von deutschsprachigen

Schüler*innen der 9. Klasse prüfen, inwieweit neu entwickelte Aufgaben aus der Validierungsstudie dieselben Aspekte der Kompetenzen historischen Denkens (mit-)erfassen wie jene aus der Pilotierung. Zudem untersuchten sie, inwieweit die Aufgaben andere Aspekte als die Pilotierungsaufgaben erfassen und ob sich diese anhand der Itemmerkmale erklären lassen. Abschließend diskutieren die Autor*innen unter anderem die Möglichkeiten des Einsatzes der Testaufgaben für Interventionsstudien und weitere empirische Analysen. Daraufhin stellt *Andrea Brait* Ergebnisse ihres österreichischen Projektes «Historisches Lernen zwischen Schule und Museum» vor, indem sie unter anderem anhand von Expert*inneninterviews exemplarisch auf die Überzeugungen von Geschichtslehrpersonen zu Museumsbesuchen eingeht und ausgehend von Beobachtungen von Museumsbesuchen von Sekundarschulklassen die Förderung des historischen Denkens im Museum thematisiert. Insbesondere fokussiert die Autorin den Wert der Quellenorientierung für Museumsbesuche und begründet diese aufgrund ihrer Analysen. *Daniel Fastlabend und Johannes Meyer-Hamme* überprüfen in ihrem Beitrag die verbreitete Annahme, es gebe historische Bild-Ikonen, die sich breiter Bekanntheit erfreuten. Anhand einer halboffenen Befragung sowie deren deskriptionsstatistischer Auswertung mit deutschen Geschichtslehramtsstudierenden differenzieren die Autoren die Hypothese und diskutieren die Reichweite der Untersuchung. *Alexandra Krebs* stellt die Pilotstudie ihres Promotionsprojektes zur «App in die Geschichte» vor. Anhand von Daten von deutschen Lehramtsstudierenden untersucht sie deren Nutzer*innenverhalten während der Auseinandersetzung um eine (fiktive) historische Straßenbenennung in deren Zusammenhang die Beteiligten ein digitales Archiv in der App zur Entscheidung konsultierten. Zudem thematisiert die Autorin den Zusammenhang zwischen den fassbaren historischen Lernprozessen und Narrationen und diskutiert die Potenziale des Vorgehens mit Schüler*innen. *Marcel Mierwald und Julia Waldemayer* stellen die Validierung eines Fragebogens zur Erhebung von historischen Relevanzüberzeugungen an einem Sample deutscher Geschichtsstudierender auf der Grundlage der Adaption eines niederländischen Fragebogens vor. Dazu entfalten sie die Bedeutung von Geschichtsrelevanz für historisches Denken, prüfen, inwiefern die Fragebogenantworten dieselben Dimensionen wie der Originalfragebogen aufweisen und untersuchen weitere Validitätsmerkmale. Abschließend verdeutlichen die Autor*innen unter anderem die Verwendungsmöglichkeiten des Fragebogens etwa im Rahmen von Interventionsstudien. *Stefanie Zabold* beschreibt darauffolgend die Entwicklung und Validierung eines Kompetenztests für Geschichtslehrpersonen. Anhand der Bereiche eines ge-

schichtsdidaktischen Kompetenzstrukturmodells (GeDiKo) stellt sie Aufgabenbeispiele und Ergebnisse aus Befragungen zum Testverständnis (Cognitive Labs) mit deutschen Geschichtslehrpersonen unterschiedlicher Schulformen vor. In der Diskussion thematisiert die Autorin den nachfolgenden Einsatz des Tests und dessen Grenzen. Zum Abschluss unseres Bandes stellen *Lisa Zachrich, Wolfgang Wagner, Christiane Bertram und Ulrich Trautwein* ihre Validierung eines Fragebogens zur Erfassung der pädagogisch-psychologischen Prozesse des «Aura-Erlebnisses» während der Begegnung mit Zeitzeug*innen sowie deren Zusammenhänge zu Situations- und Personenmerkmalen anhand der Daten von deutschen Schüler*innen der 8. und 11. Klassenstufe vor, die Zeitzeug*innen, welche die DDR erlebt hatten, begegnet waren. In ihren Analysen gehen sie auf die Bereiche der kognitiven und emotionalen Involviertheit sowie Zusammenhänge mit Personen- und Situationsmerkmalen ein und thematisieren abschließend empirische Einsatzmöglichkeiten.

Insgesamt verdeutlichen die Beiträge vielfältiges geschichtsdidaktisches Inter-venieren, das jedoch gegenwärtig seltener im Rahmen «klassischer» oder qualitativer Interventionsforschung, sondern häufig in Form von genereller empirischer Forschung erfolgt. Ein nicht unwesentlicher Teil der Studien dieses Bereiches thematisierte indes die Entwicklung von Erhebungsinstrumenten, die für nachfolgende Interventionsforschung vorgesehen ist oder in diesem Rahmen eingesetzt werden kann. Dies ist ermutigend, scheinen sich empirische Geschichtsdidaktiker*innen im deutschsprachigen Raum doch verstärkt mit der Frage zu beschäftigen, wie historisches Denken gezielt gefördert werden kann. Dabei werden neue Herangehensweisen und Konzepte erprobt, um Wandlungspotenziale der Praktiken historischen Lernens in formalen und nonformalen Kontexten auszuloten. Wir hoffen, dass wir mit diesem Band diesen eingeschlagenen Weg stützen und zur Durchführung weiterer diesbezüglicher Forschungsprojekte motivieren können.

Wir bedanken uns bei den Autor*innen für ihre Beiträge, bei den Gutachten für die wertvollen Impulse, bei Manuel S. Hubacher für seine Unterstützung bei der Fertigstellung des Bandes und bei den Mitarbeitenden des hep Verlags für die Unterstützung des Publikationsprozesses. Wir wünschen den Lesenden eine interessante Lektüre und spannende Auseinandersetzung mit dem Stand der geschichtsdidaktischen Interventionsforschung.

Literatur

- Boxtel, Carla van & Drie, Jannet van (2018). Historical reasoning: Conceptualizations and educational applications. In Scott Allan Metzger & Lauren McArthur Harris (Hrsg.), *The Wiley international handbook of history teaching and learning* (S. 149–176). New York: Wiley-Blackwell. <https://doi.org/10.1002/9781119100812.ch6>.
- Cheung, Alan C. K. & Slavin, Robert E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292. <https://doi.org/10.3102/0013189X16656615>.
- Drie, Jannet van (2020, 3. September). *Improving history education. The value of intervention studies*. Vortrag gehalten auf der Geschichtsdidaktik empirisch light.
- Fend, Helmut (2011). Die sozialen und individuellen Funktionen von Bildungssystemen: Enkulturation, Qualifikation, Allokation und Integration. In Stephanie Hellekamps/Wilfried Plöger & Wilhelm Wittenbruch (Hrsg.), *Handbuch der Erziehungswissenschaft* (Bd. 3, S. 41–53). Paderborn: Schöningh.
- Gaspard Hanna & Trautwein Ulrich (2022). Randomisierte kontrollierte Feldstudien. In Georg Weißeno & Béatrice Ziegler (Hrsg.), *Handbuch Geschichts- und Politikdidaktik* (S. 395–409). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-29668-1_33.
- Gopalan, Maithreyi/Rosinger, Kelly & Ahn, Jee Bin (2020). Use of quasi-experimental research designs in education research: Growth, promise, and challenges. *Review of Research in Education*, 44(1), 218–243. <https://doi.org/10.3102/0091732X20903302>.
- Legewie, Joscha (2012). Die Schätzung von kausalen Effekten: Überlegungen zu Methoden der Kausalanalyse anhand von Kontexteffekten in der Schule. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 64(1), 123–153. <https://doi.org/10.1007/s11577-012-0158-5>.
- Leutner, Detlev (2013). Perspektiven pädagogischer Interventionsforschung. In Eckart Severing & Reinhold Weiß (Hrsg.), *Qualitätsentwicklung in der Berufsbildungsforschung* (S. 17–28). Bielefeld: Bertelsmann.
- Lücke, Martin (2020, 3. September). *Intervention und Lebenswelt – Theoretisch-geschichtsdidaktische Perspektiven*. Vortrag gehalten auf der Geschichtsdidaktik empirisch light.
- Müller, Fabian & Ruppert-Kelly, Martina (2016). «Die Kinder sollen das ruhig mal nachempfinden können». Thesen zur Bedeutung des Beutelsbacher Konsenses in der Gedenkstättenpädagogik. In Benedikt Widmaier & Peter Zorn (Hrsg.), *Brauchen wir den Beutelsbacher Konsens? Eine Debatte der politischen Bildung* (S. 242–250). Bonn: Bundeszentrale für politische Bildung.
- Reinmann, Gabi (2005). Innovation ohne Forschung? Ein Plädoyer für den Design-Based Research-Ansatz in der Lehr-Lernforschung. *Unterrichtswissenschaft*, 33(1), 52–63. <https://doi.org/10.25656/01:5787>.
- Sanetti, Lisa M. H./Cook, Bryan G. & Cook, Lysandra (2021). Treatment fidelity: What it is and why it matters. Learning disabilities. *Research & Practice*, 36(1), 5–11. <https://doi.org/10.1111/ldrp.12238>.
- Slavin, Robert E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31(7), 15–21. <https://doi.org/10.3102/0013189X031007015>.

- Souvignier, Elmar (2021). Interventionsforschung im Kontext Schule. In Tina Hascher/Till-Sebastian Idel & Werner Helsper (Hrsg.), *Handbuch Schulforschung* (S. 1–17). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-24734-8_9-1.
- Völkel, Bärbel (2015). Subjektorientierte Geschichtsdidaktik. Überlegungen zu einem uneindeutigen Begriff. In Heinrich Ammerer, Thomas Hellmuth & Christoph Kühberger (Hrsg.), *Subjektorientierte Geschichtsdidaktik* (S. 73–91). Schwalbach/Ts.: Wochenschau.
- Waldis, Monika (2022). Interventionsstudien in Geschichte. In Georg Weißeno & Béatrice Ziegler (Hrsg.), *Handbuch Geschichts- und Politikdidaktik* (S. 299–313). https://doi.org/10.1007/978-3-658-29668-1_20.

Quasiexperimentelle Interventionen

Was wirkt? Ein Forschungsprogramm zur Wirksamkeit von Zeitzeug*innen im Geschichtsunterricht

Christiane Bertram

1 Was wirkt – und wie wirken Zeitzeug*innen?

«Es ist ein anderes Erlebnis, als wenn man 'nen Text liest», so kommentierte eine Schülerin die Begegnung mit einer Zeitzeugin, die im Rahmen einer Unterrichtseinheit zur DDR in ihre Klasse kam. Ähnliche Erfahrungen schildern Lehrkräfte und Schüler*innen auf dem Forum www.ddr-zeitzeuge.de: «So viel mehr als in Geschichtsbüchern steht», «schwer beeindruckt», «sehr eindringliche Erfahrung». Dass Schüler*innen Spaß an einer Abwechslung im Unterricht haben und dass sie insbesondere das Gespräch mit Zeitzeug*innen im Unterricht beeindruckt, kennen Praktiker*innen; hierauf setzen Didaktiker*innen (z. B. Barricelli & Lücke, 2012; Sauer, 2006) und Bildungsplanautor*innen (z. B. Bildungsplan 2016 Baden-Württemberg). Zur «Stärkung der Demokratieerziehung» wird von der Kultusministerkonferenz (KMK, 2009) empfohlen, Zeitzeug*innen in den Unterricht einzuladen. Gleichzeitig werden die Risiken der Zeitzeug*innenarbeit diskutiert, denn in der internationalen Studie «Youth and History» (Angvik & von Borries, 1997) vertrauten die Jugendlichen den Aussagen von Zeitzeug*innen mehr als den Schulbuchdarstellungen oder den Erzählungen von Lehrkräften. Daher wird befürchtet, dass mit der «Aura der Authentizität» (Sabrow, 2012, 52) die Gefahr einer «Überwältigung» der Schüler*innen einhergehe, die dem historischen Lernen abträglich sein könnte.

Hiervon ausgehend wurde in der vom Bildungsministerium für Bildung und Forschung (BMBF) finanzierten ersten Zeitzeug*innenstudie untersucht, ob die Befragung von Zeitzeug*innen, mit denen live im Geschichtsunterricht gearbeitet wird, eine andere Wirkung auf die Lernenden hat als die Arbeit mit Videos und Transkriptionen. Das Ergebnis dieser ersten randomisierten und kontrollierten Interventionsstudie (Bertram, 2017; Bertram, Wagner & Trautwein, 2017) führte zu weiteren Fragen und interdisziplinären Forschungsprojekten. In unserer Forschungsgruppe untersuchen wir mit qualitativen und quantitativen Methoden sowohl die Wirksamkeit der Arbeit mit Zeitzeug*innen

als auch das Potenzial der Oral History für Unterricht und Wissenschaft.¹ Ich beginne mit Überlegungen zu der Konzeptualisierung von «Zeitzeug*innen» und den Risiken und Chancen der Arbeit mit ihnen in einem didaktischen Lehr-Lern-Setting. Anhand zentraler statistischer Begriffe wird das Forschungsverständnis der «Randomized Controlled Field Trials» vorgestellt. Nachfolgend skizziere ich das Forschungsprogramm rings um die «Wirksamkeit von Zeitzeug*innen im Geschichtsunterricht» und greife abschließend die Kritik an der «evidenzbasierten Pädagogik» (Bellmann & Müller, 2011) auf, um die Chancen und Grenzen der evidenzbasierten Forschung zu diskutieren.

2 Was sind Zeitzeug*innen – und wie arbeiten wir mit ihnen in der Schule?

Der Begriff der «Zeitzeug*innen» ist spezifisch deutsch und lässt sich kaum ins Englische übersetzen, denn Begriffe wie «contemporary witnesses» oder «eyewitnesses of the past» treffen den Bedeutungsgehalt von «Zeitzeugen» nicht wirklich. Dies hängt zusammen mit der «Geburt des Zeitzeugen» (Sabrow & Frei, 2012) in den Gerichtssälen des Eichmann-Prozesses in Jerusalem. Seitdem ist der Begriff im deutschen Sprachraum so eng mit der Erinnerung an den Holocaust verbunden, dass die Begriffe «Zeitzeug*innen» und «Überlebende» häufig als Synonyme benutzt werden. Bei dem «Verschwinden der Zeitzeugen» (Skriebeleit, 2011) werden automatisch der Holocaust und die NS-Zeit assoziiert. In beiden deutschen Teilstaaten wurden Zeitzeug*innen zu einem wesentlichen Bestandteil der öffentlichen Erinnerung an den Nationalsozialismus. In der DDR waren es die Widerstandskämpfer, in der alten Bundesrepublik die Überlebenden, die in Gedenkstätten und Schulen ihre Lehren aus der gemeinsamen Vergangenheit zogen, die durchaus unterschiedlich waren (Weber, 2020, 1056–1059). Kritisiert wird häufig an Zeitzeug*innen im Kontext einer geschichtskulturellen Inszenie-

1 In unterschiedlicher Besetzung arbeite ich in dem interdisziplinären Forschungsteam mit den Kolleg*innen der Empirischen Bildungsforschung an der Universität Tübingen (Ulrich Trautwein, Wolfgang Wagner, Lisa Zachrich), mit der Stiftung Berliner Mauer (Bettina Effner, Sarah Bornhorst, Gerhard Sälter), dem «Archiv Deutsches Gedächtnis» (Almut Leh) und dem Zentrum für Schulqualität und Lehrerbildung (ZSL) Baden-Württemberg (Carsten Arbeiter und Stefan Schipperges) zusammen. Ich danke ihnen und den zahlreichen Hilfskräften an den Universitäten Konstanz und Tübingen für die konstruktive, produktive und inspirierende Arbeit in den letzten Jahren.

rung (z. B. bei einer Museumsausstellung oder einer Gedenkveranstaltung), dass sie lediglich einen bereits vorher bekannten und definierten Beitrag leisten (Classen, 2012), sodass sie von den «Interviewpartner*innen» in der Oral History deutlich abgegrenzt werden sollten (Wierling, 2014). In einer lexikalischen Argumentation hingegen wird das vorangestellte Lexem «Zeit» als Hinweis auf die «Zeitgeschichte» verstanden, die als «Epoche der Mitlebenden und ihre wissenschaftliche Behandlung» (Rothfels, 1957, 2) definiert wird. In diesem Verständnis gibt ein*e «Zeitzeug*in» Zeugnis von der Zeit ab, die selbst miterlebt wurde, sie sind mündliche Quellen, die zu vielfältigen Bereichen und Themen der Zeitgeschichte befragt werden können.

Mit der Erfindung mobiler Aufnahmegeräte Mitte des 20. Jahrhunderts wurde es möglich, mündliche Quellen – die Urform der geschichtlichen Erzählung (Thukydides, 411 v./1991) – zu konservieren. Da die Forschenden damit an der Entstehung der Quelle beteiligt (Geppert, 1994) sind, sollten in in einem methodisch kontrollierten Oral-History-Interview ergebnisoffene, lebensgeschichtliche Interviews geführt werden, in denen die Themen und Schwerpunktsetzungen von den Interviewpartner*innen kommen (Wierling, 2003). Ähnlich wie bei anderen Ego-Quellen (z. B. Tagebücher, Autobiografien, Briefe) werden die Lebensgeschichten und Alltagserfahrungen auf kollektive und individuelle Verarbeitungsweisen der vergangenen Ereignisse hin ausgewertet (Plato, 2000). So berichteten in dem ersten großen Oral-History-Projekt in der alten BRD, dem LUSIR-Projekt («Lebensgeschichte und Sozialkultur im Ruhrgebiet», Niethammer, 1983) Arbeiter*innen im Ruhrgebiet von ihren Faschismuserfahrungen 1930 bis 1950. Dorothee Wierling unternahm mit der Befragung des Geburtsjahrgangs 1949 in der DDR den «Versuch einer Kollektivbiographie» (Wierling, 2002). In der Interviewstudie «Generation 1975», die in diesem Beitrag noch eine Rolle spielen wird, wurden Menschen aus dem Osten und Westen, die bei dem Fall der Mauer 14 Jahre alt waren, zu ihrer Erinnerung an ein geteiltes und wiedervereinigtes Deutschland befragt (Bertram, 2020).

Im deutschen Sprachraum hat sich der Begriff der «Zeitzeug*in» in der Geschichtswissenschaft, Geschichtskultur und im Geschichtsunterricht gemeinhin etabliert (u. a. Erbar, 2012; Sauer, 2006). Sehr unterschiedliche Lehr-Lern-Szenarien sind denkbar, mit denen sehr unterschiedliche Zielsetzungen verfolgt werden. Wenn beispielsweise Schüler*innen mit Bewohner*innen eines Seniorenheims Interviews zu ihrer Jugend im «Wirtschaftswunderland» führen, arbeiten sie bei der methodisch kontrollierten Vorbereitung, Durchführung und Auswertung der Interviews wie Oral Historians. Wenn hingegen bei einer schulischen

Gedenkveranstaltung Zeitzeug*innen auf einem Podium befragt werden, spielen sie eine im Vorfeld abgesprochene Rolle in einem vorab definierten geschichtskulturellen Kontext. Dazwischen gibt es eine Vielzahl von Varianten, wie mit «Zeitzeug*innen» in der Schule, Wissenschaft und Erinnerungskultur gearbeitet wird, sodass die folgende Arbeitsdefinition notwendigerweise holzschnittartig bleibt und die Vielschichtigkeit der Zeitzeug*innenschaft nur andeuten kann.

Unter «Zeitzeug*innen» werden im Folgenden Menschen verstanden, die analog oder digital vermittelt von der Vergangenheit berichten. Ihre Zeugnisse sind sowohl eine *Quelle*, ein «Überbleibsel» der Vergangenheit, als auch eine *Darstellung*, eine Deutung und Narration aus späterer Sicht. Zeitzeug*innen unterscheiden sich auf verschiedenen *Dimensionen*, zum Beispiel hinsichtlich des Themas, zu dem sie befragt werden (als Teil einer Gewaltgeschichte, Demokratiegeschichte, Alltagsgeschichte ...), der *Rolle*, die sie in der Vergangenheit eingenommen haben (als Opfer, Täter*in, Held*in, Beobachter*in ...), nach der *Nähe* zu den berichteten Ereignissen (aktiv beteiligt, selbst beobachtend, informiert durch die Medien ...) und nach ihrem *lebensgeschichtlichen Abstand* zu den Ereignissen (jüngerer oder älterer Mensch; zeitliche Nähe und Ferne). Während sich die Zeugnisse von Überlebenden – auch Videos – eher nicht für «ein kritisches Hinterfragen einer Quelle und ihre gründliche Analyse und Interpretation» (Brüning, 2018, 372) eignen, können Zeugnisse von «Held*innen» oder «Alltagszeitzeug*innen» kritischer hinterfragt und methodisch kontrolliert ausgewertet werden. Die Ziele der Arbeit mit Zeitzeug*innen variieren dementsprechend, doch geht es im Kern darum, die Relevanz der Vergangenheit für die eigene Gegenwart sichtbar und spürbar zu machen. Die Arbeit mit Zeitzeug*innen könnte daher der insbesondere der Orientierungskompetenz und der Förderung eines reflektierten und (selbst-)reflexiven Geschichtsbewusstseins zuträglich sein.

Als theoretische Grundlage der hier vorgestellten Interventionsstudien dient das FUER-Kompetenzstrukturmodell historischen Denkens (Körper, Schreiber & Schöner, 2007). Unter Berufung auf Rüsen (1983) und Hasberg & Körper (2003) werden die «Kompetenzen historischen Denkens» innerhalb eines Denkprozesses modelliert, der von einer Verunsicherung ausgehend zu einer Fragestellung führt (*Fragekompetenz*), auf deren Grundlage methodisch kontrolliert und reflektiert Quellen und Darstellungen ausgewertet werden (*Methodenkompetenz*: Re- und De-Konstruktion), um zu einer Antwort auf die Frage und zu einer Orientierung in der Gegenwart zu kommen (*Orientierungskompetenz*). Die *Sachkompetenz*, unter der nicht vereinzelt Faktenwissen, sondern ein grundlegendes Verständnis der Begriffe, Strukturen und epistemologischen Prinzipien

von Geschichte verstanden wird, stellt den inneren Kern des historischen Denkprozesses dar.

Nach diesen theoretischen Überlegungen zu dem, was «Zeitzeug*innen» ausmacht und welchen Beitrag sie im Geschichtsunterricht leisten können, werden im nächsten Schritt einige methodische Voraussetzungen geklärt, die für das Verständnis von Wirksamkeitsstudien im Bildungskontext wichtig sind.

3 What Works Clearinghouse?

In der von der Pädagogischen Psychologie geprägten Empirischen Bildungsforschung wird der Begriff der «Wirksamkeit» häufig im Sinne der Initiative *What Works Clearinghouse* verstanden. Diese wurde von dem Institute of Education Sciences, einer Einrichtung des US-amerikanischen Bildungsministeriums, mit dem Ziel gegründet, Lehrkräften, Bildungspolitiker*innen und der Öffentlichkeit empirisch belastbare Informationen über Ansätze und Beispiele effektiver Praxis im Bildungsbereich zugänglich zu machen: «We focus on the results from high-quality research to answer the question ‘What works in education?’»² Randomisierte kontrollierte Interventionsstudien werden hier als «Goldstandard» für «evidence-based education policies» (Slavin, 2002) verstanden. In der Fachdidaktik wird das «Paradigma evidenzbasierter Pädagogik» häufig kritisch diskutiert, da es keine Methodenvielfalt biete und «komplexere Verhältnisbestimmungen» nicht zulasse (Bellmann & Müller, 2011, 15). In der Fachdidaktik wird die «Wirksamkeit von Unterricht» häufig nicht unbedingt empirisch, sondern eher theoretisch und pragmatisch verstanden. So sollen in der Reihe «Wirksamer Fachunterricht» die Interviews mit Expert*innen der Fachdidaktik (u. a. der Geschichtsdidaktik, Politikdidaktik, Geografiedidaktik) «zu einer verdichteten Beantwortung der Frage [...] [führen], was einen wirksamen Fachunterricht ausmacht» (Kuchler & Sommer, 2018). Um zunächst ein Grundverständnis des experimentellen Ansatzes und für das Verständnis von «Wirksamkeit» zu schaffen, sollen im Folgenden zentrale Fachbegriffe der Interventionsforschung geklärt werden.

Intervention: Der Begriff leitet sich von dem lateinischen Wort «intervenire» ab und bedeutet «dazwischentreten» oder «sich einschalten». Im Bildungskon-

² <https://ies.ed.gov/ncee/wwc/>.

text wird hierunter eine zielgerichtete Maßnahme verstanden, um erwünschte Verhaltensweisen oder Kompetenzen zu fördern oder hinderliche Verhaltensweisen abzubauen. Interventionen können hinsichtlich der Zielgruppe (von einzelnen Personen über Klassen bis hin zu Schulen oder Regionen) und der Dosierung und Komplexität der Maßnahmen (von kurzen Interventionen in einer Stunde bis hin zu komplexen Programmen mit mehreren Komponenten) variieren (Hager & Hasselhorn, 2000; Leutner, 2010).

Wirkung und Kausalität: Das Substantiv *Wirkung* leitet sich ab aus dem spätmittelhochdeutschen Verb *wirken* (in der Bedeutung von «tätig sein», «nützlich sein», «etwas schaffen» und «etwas vollbringen».³ In der Wissenschaftstheorie bezeichnet der Begriff «Wirkung» eine bestimmte Abfolge von Ereignissen und Zuständen, die aufeinander bezogen sind. A ist die Ursache für die Wirkung B, wenn B von A herbeigeführt wird. Es wird also eine Aussage über einen kausalen Zusammenhang (von lateinisch *causa*, «Ursache», und *causalis*, «ursächlich, kausal») getroffen, eine bestimmte Richtung des Zusammenhangs zwischen zwei Variablen wird angenommen (Spohn, 1983, 18). Zentral ist, dass der Nachweis korrelativer Zusammenhänge zwischen zwei Variablen keine Aussage über kausale Zusammenhänge und die Wirkrichtung zwischen den Variablen zulässt.

Unabhängige und abhängige Variablen: Bei der Feststellung eines statistisch signifikanten Zusammenhangs zwischen den Ereignissen A und B gibt es prinzipiell folgende Erklärungsmöglichkeiten: (1) A hat B verursacht, (2) B hat A verursacht, (3) A und B werden durch ein drittes Ereignis C verursacht oder (4) der Zusammenhang in den Daten ist fehlerhaft oder zufällig, das heißt in Wahrheit gar nicht vorhanden. *Unabhängig* wird die Variable genannt, die die Veränderung verursacht, *abhängig* ist die Variable, die die Folge darstellt. Die Festlegung des Wirkzusammenhangs erfolgt auf der Grundlage theoretischer Überlegungen und wird in den Hypothesen festgehalten, die inhaltlich-fachlich, nicht statistisch-methodisch begründet werden (Huber, 2005).

Statistische Testverfahren: Wenn aus den Ergebnissen in einer Stichprobe auf die Grundgesamtheit geschlossen werden soll, wenn also *inferenzstatistische* Aussagen über die Gesamtpopulation getroffen werden sollen – nicht nur *deskriptive* Aussagen über die Verhältnisse in der jeweiligen Stichprobe – sind die Aussagen zu einem gewissen Grade mit «Unsicherheit» behaftet. Daher geben statistische Tests Auskunft darüber, wie wahrscheinlich es ist, dass sich das ge-

3 Vgl. <https://de.wiktionary.org/wiki/wirken> [27.04.2022].

fundene Ergebnis, zum Beispiel zu den Zusammenhängen zwischen Variablen oder zu Unterschieden zwischen Gruppen, sich zwar in den Daten zeigt, aber eigentlich gar kein Zusammenhang oder Unterschied vorhanden ist, der Befund in den Daten also «zufällig» zustande gekommen sein könnte. Die Wahrscheinlichkeit, ein zufälliges Ergebnis als statistisch bedeutsam und überzufällig zu interpretieren, wird in der *statistischen Signifikanz* mathematisch bestimmt. Statistische Signifikanz bedeutet, dass beispielsweise Mittelwertunterschiede zwischen zwei Gruppen wie der gefundene beziehungsweise ein noch extremerer bei wiederholter Ziehung unter der Annahme eines bestimmten Populationsparameters (also z. B. keines Mittelwertunterschieds in der Grundgesamtheit) mit einer dem Signifikanzniveau entsprechenden relativen Häufigkeit zu erwarten ist (üblich ist hier 5 %). Dieses Signifikanzniveau wird mit dem *p*-Wert angegeben, das heißt: Je kleiner der *p*-Wert (also je «höher signifikant» der Test), desto geringer wird die Wahrscheinlichkeit, dass der Effekt «in Wirklichkeit», unter Annahme der Null-Hypothese, nicht existiert, das heißt, desto sicherer kann man sein, dass der Effekt existiert.

Stichprobe und Clusterstruktur: Die Teilnehmer*innen an einer schulischen Interventionsstudie, beispielsweise Lehrkräfte oder Schüler*innen, agieren im Kontext der Klasse beziehungsweise der Schule. Im statistischen Sinne sind die Daten «geschachtelt» in Schüler*innen innerhalb von Klassen, die sich innerhalb von Schulen befinden, und in Lehrkräfte innerhalb von Schulen. Daher sind im schulischen Kontext vergleichsweise große Stichproben notwendig und die Clusterstruktur der Daten muss in den Analysen berücksichtigt werden (Raudenbusch, 1997; Shadish, Cook & Campbell, 2002). Im Idealfall sollte bei der Studienplanung auf der Basis einer sogenannten *Power-Analyse* die Größe der benötigten Stichprobe berechnet werden. Die Stärke des erwarteten Effekts, der in pädagogischen Interventionen aufgrund zahlreicher möglicher Störfaktoren häufig nur klein oder mittel ist (Kraft, 2020), ist entscheidend für die Berechnung der für eine bestimmte *Teststärke* notwendigen Stichprobe (Brunner, Keller, Wenger, Fischbach & Lüdtke, 2018).

Randomized Controlled Field Trials kommen aus der medizinischen (und landwirtschaftlichen) Forschung. In den CONSORT-Richtlinien («Consolidated Statement of Reporting Trials») wurden 1996 in Form von 25 Punkten Qualitätsstandards definiert: von der Präregistrierung der Hypothesen über eine Power-Analyse, die Stichprobenakquise und Randomisierung bis hin zu der Erstellung von Berichten (Moher et al., 2010; Torgerson & Torgerson, 2013). Anhand der

Wortbestandteile soll das Prinzip experimenteller Interventionsstudien im Bildungskontext erklärt werden.

«*Trial*»: Der experimentelle Versuch zeichnet sich dadurch aus, dass in der *Experimentalgruppe* (in der «experimentellen Bedingung») die unabhängige Variable (d. h. die Intervention, z. B. ein Medikament) im Vergleich zu einer *Kontrollgruppe* (die kein Medikament, sondern ein Placebo erhält) variiert wird. Auf der Basis von statistischen Regressionsanalysen wird überprüft, ob die *unabhängige Variable* (Medikament versus Placebo) einen Einfluss hat auf die *abhängige Variable* (z. B. Gesundung der Patient*innen). Im Bildungskontext gibt es daher neben der Experimentalgruppe, die beispielsweise an einer Unterrichtsintervention teilnimmt, eine Kontrollgruppe, die keine Intervention erhält, aber an den Testungen teilnimmt. Als unabhängige Variable gilt die Teilnahme beziehungsweise Nichtteilnahme an der Intervention (Experimentalgruppe). Es kann mehrere Experimentalgruppen geben. Als abhängige Variablen werden die Effekte der Intervention verstanden, zum Beispiel die kognitiven oder motivationalen *Outcomes* bei den Schüler*innen.

«*Randomized*»: Die randomisierte (= zufällige) Zuweisung zu den Bedingungen (Experimental- oder Kontrollgruppe) ist entscheidend, um eine Stichprobenverzerrung zu vermeiden: «Randomised controlled trials (RDTs) are the best approach for demonstrating the effectiveness of a novel educational intervention. Most other approaches are susceptible to selection bias. Selection bias occurs because the method of selecting schools or students to receive an intervention is related to outcome» (Torgerson & Torgerson, 2013, 2). Indem die Gruppen den Bedingungen (Interventions- oder Kontrollgruppe) zufällig zugewiesen werden, werden potenzielle Störfaktoren wie die beobachteten Voraussetzungen der Teilnehmenden (z. B. Alter und Geschlecht), aber auch die nicht beobachteten Voraussetzungen (z. B. die Motivation zur Teilnahme an der Intervention) statistisch kontrollierbar, sodass alternative Erklärungen der beobachteten Effekte (z. B. ein Kompetenzerwerb in der Freizeit) zwischen den Bedingungen «ausbalanciert» werden können (Gaspard & Trautwein, 2021, 4).

«*Controlled*»: Das Vorhandensein einer Kontrollgruppe, das heißt, es gibt eine Gruppe, die zufällig zugewiesen keine «Behandlung», zum Beispiel Unterrichtsintervention, erfährt, ist zentral für ein experimentelles Design, denn es muss immer mit Effekten gerechnet werden (also auch ohne spezifische Intervention), beispielsweise aufgrund von Reifungsprozessen oder einfach wegen des wiederholten Testens in einem Pro-, Prä- und Follow-up-Design. Deshalb wäre eine positive Veränderung in der Interventionsgruppe noch kein hinreichender Beleg

für einen besonderen Nutzen der Intervention. Darüber hinaus bezieht sich die «Kontrolle» auf die Durchführung der Intervention beziehungsweise der Interventionsvarianten. Um beobachtete Effekte auf die Intervention zurückführen zu können, muss in einem experimentellen Design sichergestellt werden, dass die Intervention hinsichtlich der Inhalte, des zeitlichen Umfangs, des Ablaufs und der Erfassung der Ergebnisse in einem hohen Maß standardisiert durchgeführt wird. Dies ist im Schulkontext nicht leicht umzusetzen, da den individuell agierenden Lehrkräften im Unterrichtsgeschehen eine hohe Bedeutung zukommt (Hattie, 2009). Daher ist es wichtig, die *Durchführungstreue* (auch «program integrity», «treatment fidelity», «Implementationsgüte» genannt) der Intervention zu überprüfen (Herbein, 2020, 29).

«*Field*»: In der medizinischen Forschung wird zwischen den Experimenten im Labor (häufig Untersuchungen mit Tieren) und zwischen verschiedenen «klinischen Studien» im Feld (Untersuchungen mit Menschen) unterschieden. Hier wird zunächst das Wirkungsvermögen einer Maßnahme unter idealen Bedingungen überprüft (*efficacy*-Studie), dann folgt die Untersuchung der Wirksamkeit einer Maßnahme unter Alltagsbedingungen (*effectiveness*-Studie) (Shafrey, 2016). Übertragen auf die schulische Interventionsforschung bedeutet dies, dass zunächst die Wirksamkeit einer Unterrichtseinheit (= «Intervention») unter standardisierten Bedingungen und starker Kontrolle getestet wird, zum Beispiel indem nur eine Lehrkraft alle Interventionsgruppen unterrichtet, und dass im nächsten Schritt die Unterrichtseinheit von den eigenen geschulten Lehrkräften unterrichtet wird. Eine Intervention im Bildungskontext wird «dann als erfolgreich bezeichnet, wenn sie die intendierten Förderziele erreicht und gleichzeitig in der Praxis implementierbar ist» (Herbein, Golle, Nagengast & Trautwein, 2020, 84). Hierunter wird sowohl der flächendeckende Einsatz der Intervention verstanden, zum Beispiel die Aufnahme der Maßnahme ins Curriculum (Baker, 2006), als auch die Ausweitung des Implementationskontexts (Gottfredson et al., 2015).

Kritik an kontrollierten Interventionsstudien: Bellmann & Müller (2011) kritisieren in ihrem Herausgeberband «Wissen, was wirkt. Kritik evidenzbasierter Pädagogik» das «Paradigma» der «Evidenz» als «durch privilegierte wissenschaftliche Methoden, insbesondere randomisierte kontrollierte Studien generiertes Wissen» (14), bei dem das experimentelle Design zum «Königsweg zur Generierung eines Wissens über kausale Zusammenhänge erhoben» (14) werde. Da die «Komplexität als Merkmal pädagogischer Praxis» (Herzog, 135) in einem experimentellen Design nicht abgebildet werden könne, wird angezweifelt, ob das «Paradigma der Evidenzbasierung» überhaupt einen relevanten Beitrag für

die Theoriebildung oder Professionsforschung leisten könne. In der anglo-amerikanischen Forschung wird weniger grundsätzlich kritisiert, doch wird auf die Grenzen kontrollierter Interventionsstudien hingewiesen. Da in einem experimentellen Setting die Fragen notwendigerweise eng gestellt würden, führe dies zu «narrow answers» (Lareau, 2009, 145). Experimentelle Studien könnten daher trotz der Erkenntnisse über die – messbare – Wirksamkeit aufgrund ihres Designs nur einen sehr begrenzten Einblick in die «Black Box» der Prozesse liefern (Nelson et al., 2012). Daher sollten randomisierte kontrollierte Studien «als Bestandteil eines größeren Forschungsprozesses» verstanden werden, durch den «Brücken zwischen Theorie und Praxis gebaut werden können» (Gaspard & Trautwein, 2021, 9). Genau dies versuchen wir in dem nun vorzustellenden Forschungsprogramm zur Wirksamkeit von Zeitzeug*innen im Geschichtsunterricht.

4 Black Box «Zeitzeug*innen im Geschichtsunterricht»: ein Forschungsprogramm

4.1 Zeitzeug*innenstudie 1: Friedliche Revolution in der DDR

Anhand der ersten Interventionsstudie (Bertram, Wagner & Trautwein, 2017; Bertram, 2017) werden zentrale Prinzipien von Randomized Controlled Field Trials aufgezeigt.

Intervention: Die inhaltliche Schwerpunktsetzung der Unterrichtseinheit bezog sich auf den öffentlichen Diskurs zwanzig Jahre nach der Friedlichen Revolution. Die im Jahr 2009 massenmedial ausgetragene Kontroverse zwischen Erwin Sellering und Angela Merkel, ob die DDR ein «Unrechtsstaat» gewesen sei, bildete den Ausgangspunkt der 6½-stündigen Unterrichtseinheit, in der mit Zeitzeug*innen in unterschiedlicher «Form» (Live, Video, Text) gearbeitet wurde. Ziel der Intervention war es, die epistemologischen Voraussetzungen und Restriktionen der Arbeit mit Zeitzeug*innen als Ausgangspunkt für ein kritisches historisches Lernen zu nutzen und den Lernenden einen Einblick in die geschichtstheoretischen Grundlagen des Fachs Geschichte zu vermitteln.

Design: An der Studie nahmen 35 neunte Klassen an Gymnasien in Baden-Württemberg teil ($N = 962$). Die experimentelle Manipulation bestand in der medialen Repräsentation der Zeitzeug*innen: Diese kamen in jeweils zehn Klassen live in den Unterricht (Live-Bedingung) beziehungsweise kamen per Video (Vi-

deo-Bedingung) oder Text (Text-Bedingung) zur Sprache. Ansonsten war die Intervention von insgesamt 6½ Stunden identisch. Die fünf Kontrollklassen erhielten keine Intervention, doch sie nahmen an den drei Testungen teil. Die Klassen wurden den vier Bedingungen randomisiert zugewiesen. Eine Lehrkraft führte die Intervention in den dreißig Experimentalklassen durch. Die Implementationsgüte wurde in allen Interventionsstunden von jeweils zwei externen Rater*innen beurteilt.

Instrumente: In einem Vor- und Nachtest unmittelbar vor und nach der Intervention und in einem ca. 2 bis 3 Monate später stattfindenden Follow-up-Test wurden das themenbezogene Wissen und die historischen Sachkompetenzen der Schüler*innen erhoben. Die Themenkenntnisse der Lernenden wurden mit einem Lückentext erfasst, der in zwei Hälften aufgeteilt alternierend zu den ersten beiden Testzeitpunkten und im Follow-up-Test in ganzer Länge eingesetzt wurde. Die historischen Sachkompetenzen der Schüler*innen wurden in drei raschskalierten und auf die psychometrische Validität überprüfte Subtests erfasst: «Sachkompetenz Re-Konstruktion», «Sachkompetenz De-Konstruktion», «Sachkompetenz Eigenart von Zeitzeug*innen». Im Nachtest und Follow-up-Test schätzten die Lernenden die Effektivität der Unterrichtseinheit im Hinblick darauf ein, ob sie meinten, durch die Arbeit mit dem Zeitzeug*innenbericht inhaltlich beziehungsweise methodisch etwas gelernt zu haben und ob ihnen die Unterrichtseinheit Spaß gemacht hat. Um die individuellen Voraussetzungen der Schüler*innen kontrollieren zu können, wurden eine Vielzahl von Kovariaten erhoben (z. B. familiärer Hintergrund, Noten, kognitive Grundfähigkeiten, Leseschwindigkeit etc.).

Analysen: In multiplen Regressionsanalysen wurde unter Berücksichtigung der individuellen Voraussetzungen und der Leistung im Vortest auf statistisch signifikante Gruppenunterschiede im Nachtest beziehungsweise im Follow-up-Test getestet. In den Schätzparametern der Analysen wurde die Clusterstruktur der Daten berücksichtigt. Statistisch signifikante Mittelwert-Unterschiede zwischen den Gruppen beim Nachtest beziehungsweise Follow-up-Test – unter Kontrolle der Leistungen im Vortest und der individuellen Voraussetzungen – ließen Aussagen über die differenzielle Wirksamkeit der Intervention zu.

Ergebnisse: (1) Im Vergleich zur Kontrollgruppe schnitten die Interventionsgruppen (live, Video, Text) in zwei der drei Kompetenztests zur Einsicht in die epistemologischen Prinzipien besser ab, ein Effekt, der sich im Follow-up-Test noch steigerte. (2) Verglichen mit den Lernenden in der Video-/Textbedingung schnitten die Lernenden in der Live-Bedingung im Nachtest in zwei Kom-

petenztests statistisch signifikant schlechter ab. In dem Follow-up-Test wurde kein statistisch signifikanter Unterschied beobachtet. (3) Deutlich positiver als die Vergleichsgruppe schätzten die Schüler*innen in der Live-Bedingung die eigenen inhaltlichen und methodischen Lernfortschritte durch die Arbeit mit Zeitzeug*innen und das Motivationspotenzial der Methode ein. Die Schüler*innen in der Live-Gruppe hatten also deutlich mehr Spaß an der Unterrichtseinheit und hatten den Eindruck, viel gelernt zu haben, doch haben sie das zentrale Lernziel der Einheit – die Erkenntnis der Konstruktion von Geschichte – weniger gut verstanden als die Schüler*innen in der Video- und Textgruppe.

Implikationen: Der Befund, dass die Begegnung mit Live-Zeitzeug*innen einen Effekt auf das Lernen und die Motivation der Schüler*innen hat, führt zu der Frage, was die besondere Lernerfahrung mit Zeitzeug*innen ausmacht. In einem Dissertationsprojekt wird daher untersucht, wie diese Lernerfahrung konzeptuell beschrieben und empirisch erfasst werden kann.

4.2 «Aura der Authentizität» – eine komplexe Lernerfahrung erfassen

Was zeichnet die besondere Lernerfahrung mit Zeitzeug*innen aus? In dem «Framework for Encountering Complex Historical Sources» wurde die besondere Lernerfahrung mit «komplexen historischen Quellen», wie es historische Orte und Zeitzeug*innen sind, in Verbindung gebracht mit der kognitiven, emotionalen und physischen «Involviertheit» der Lernenden (Zachrich, Baron, Weller & Bertram, 2020). Ausgehend von diesen theoretischen Überlegungen wurde ein Fragebogen konzipiert und validiert, um die theoretisch angenommene Faktorenstruktur und die psychometrische Güte des Instruments zu überprüfen (siehe Zachrich et al. in diesem Band). Tatsächlich ließen sich die kognitive und emotionale Involviertheit der Lernenden (allerdings nicht die physische) mit dem standardisierten Fragebogen psychometrisch sauber erfassen. In der zweiten Zeitzeug*innenstudie soll mithilfe dieses Instruments untersucht werden, welchen Einfluss die individuelle Erfahrung der Begegnung mit Zeitzeug*innen auf die Wirksamkeit der Intervention hat.

4.3 Interviewprojekt «Generation 1975 – Mit 14 ins neue Deutschland»

Das Interviewprojekt «Generation 1975» ist ein Transferprojekt, das einen Beitrag für die Geschichtswissenschaft, die Geschichtskultur und den Geschichtsunterricht leisten will. In einer Kooperation mit der «Stiftung Berliner Mauer», dem «Archiv Deutsches Gedächtnis» und den Videokünstler*innen Ina Rommee und Stefan Krauss wurde eine Oral-History-Studie durchgeführt, bei der 30 Jahre nach dem Mauerfall 26 Zeitzeug*innen aus Baden-Württemberg, Brandenburg, Ost- und West-Berlin zu ihren Erfahrungen im geteilten und wiedervereinigten Deutschland befragt wurden. Zwei Videoinstallationen sind hieraus entstanden, die 2020 bis 2021 in Stuttgart und Berlin gezeigt wurden. Derzeit wird eine Videoinstallation für die offiziellen Feierlichkeiten zur Einheit in Erfurt vorbereitet. Im Haus der Geschichte von Brandenburg und Preußen in Potsdam wird das Material in der Dauerausstellung genutzt werden. Zudem arbeiten wir mit den Videos und den Zeitzeug*innen in der zweiten Zeitzeug*innenstudie.

Unter *Generation* wird – der Definition Karl Mannheims (1928/29) folgend – die spezifisch generationelle Vergemeinschaftung einer gleichen Alterskohorte durch einen gleichgelagerten kulturellen Kontext sowie die Wahrnehmung des Geschehens aus einer ähnlichen Lebens- und Bewusstseinsschichtung verstanden. Ein solcher Generationszusammenhang entsteht dadurch, dass bedeutende gesellschaftliche Ereignisse bei den Befragten in die gleichen oder vergleichbaren Lebens- und Entwicklungsphase fielen, sodass das gemeinsame Erleben die Zugehörigkeit zu einer «Generation» vermittelt (Ahbe & Gries, 2006; Wierling, 2002). Derzeit wird das Interviewmaterial, das im «Archiv Deutsches Gedächtnis» allen Interessierten zur Verfügung steht, an der Universität Konstanz ausgewertet. Das Design der Studie und erste Ergebnisse werden im Folgenden vorgestellt.

Stichprobe: Die Stichprobe zeichnet sich aus durch das jugendliche Lebensalter der befragten Zeitzeug*innen (Geburtsjahr 1975) und die wechselseitigen Perspektiven auf den eigenen und anderen deutschen Teilstaat. Um eine möglichst große gesellschaftliche Bandbreite abzubilden, wurden bei der Online-Anmeldung der Bildungshintergrund und die politische Orientierung erhoben («Sonntagsfrage: Welche Partei würden Sie am nächsten Sonntag bei einer Bundestagswahl wählen?»).

Fragestellung: In den lebensgeschichtlichen Interviews wurde nach den großen Linien gefragt: nach dem Aufwachsen in der DDR und alten BRD, den wechselseitigen Bezügen und Verflechtungen zwischen den Teilstaaten, nach den Erfahrungen 1989/1990, nach Veränderungen in der Nach-Wendezeit und nach der

heutigen Einschätzung der Einheit. Dabei hatten wir die langfristigen Veränderungen durch Globalisierung, Digitalisierung und Neoliberalisierung im Hinterkopf, die seit den 1970er-/1980er-Jahren die Entwicklung in beiden deutschen Teilstaaten prägten (Bösch, 2015; Großbölting, 2020; Kawalczyk, 2019; Kleßmann, 2005; Mau, 2020; Ther, 2014).

Method: Das Material wurde bereits in mehreren Abschlussarbeiten an der Universität Konstanz untersucht, zum Beispiel zu den Veränderungen der Bildungschancen im Osten durch die Wiedervereinigung oder zu der gegenseitigen Wahrnehmung im Osten und Westen Deutschlands. In einem induktiven und deduktiven Vorgehen werden die Interviews derzeit in Tandems mithilfe der Software MAXQDA codiert und inhaltsanalytisch ausgewertet (Kuckartz, 2010). Nach Abschluss der Codierung und Auswertung wird ein quantifizierender und qualitativer Vergleich der Interviews und ein qualitativer Zugriff auf die individuellen Erzählungen möglich sein.

Ergebnisse: Die bisherigen (noch vorläufigen) Ergebnisse (Bertram, 2020; Ullmann, 2021) beziehen sich auf 22 Interviews (jeweils 11 aus Ost und West): (1) Die Umbruchserfahrung im Osten und Westen ist fundamental anders gelagert. Während alle damals Jugendlichen im Osten eine komplette Veränderung ihres Alltags erlebt haben, die (mit zwei Ausnahmen) als persönliche Chance erlebt wurden, hat sich für die Jugendlichen im Westen gar nichts verändert. (2) Vorurteile und Stereotype gegenüber dem anderen Teilstaat wurden im Westen häufiger, ausführlicher und mit einer deutlich abwertenden Tendenz bedient, während die Zeitzeug*innen im Osten fast durchgehend pauschale Schuldzuweisungen mieden, den Handlungsdruck der Politik mit einbezogen und sich in die Situation der Westdeutschen eher einfühlen konnten, beispielsweise nachvollziehen konnten, dass die finanziellen Belastungen weniger erfreulich waren. (3) Es scheint kein gemeinsames «Generationsempfinden» zu geben, das den Osten und Westen umfasst. Ein Zeitzeuge aus Baden-Württemberg formuliert: «Meine Generation, wir in Deutschland, wir haben doch nie eine Krise erlebt» (21_BWm_3). Dieser Aussage würden die Altersgenossen im Osten nicht vorbehaltlos zustimmen.

Diskussion: Die Generation 1975 hat im Osten und Westen Deutschlands sehr unterschiedliche Erfahrungen gemacht und empfindet sich nicht als «eine» gemeinsame Generation. Die in unserer Stichprobe substanziiell höher ausgeprägte Empathie der Zeitzeug*innen aus dem Osten könnte mit dem unterschiedlichen Kenntnisstand zusammenhängen. Die Zeitzeug*innen im Osten hatten schon in DDR-Zeiten die «alte BRD» als einen relevanten Bezugs- und Referenzpunkt in den Medien, der Schule und der Familie wahrgenommen und leben seit 30 Jahren in der

Bundesrepublik Deutschland. Die West-Zeitzeug*innen hingegen – selbst die mit Verwandten in der DDR – hatten früher kaum Einblick in das Leben in der DDR und zeigten damals und zeigen heute meist wenig Interesse. Eine Zeitzeugin formuliert auf Nachfrage: «Die DDR? Das war kein Thema bei uns.» (14_WBw_2). Die herablassende Haltung der West-Zeitzeug*innen gegenüber der DDR und dem Osten scheint auf wenig Wissen und umso mehr Ignoranz zu beruhen, was von den Ostdeutschen in unserer Stichprobe durchaus registriert wurde. Daher könnte die Unzufriedenheit im Osten nicht nur mit dem «Verlust von Lebensgewissheit» (Merkel, 3. Oktober 2019 in Kiel), sondern auch mit einer mangelnden Wahrnehmung und Wertschätzung in Verbindung gebracht werden (Borsutzki, Machowecz & Wefing, 2019).

Implikationen: Aus geschichtsdidaktischer Sicht ist es daher wichtig, insbesondere im Westen Deutschlands Wissen über die DDR und über die Transformationszeit nach 1989/90 zu vermitteln. Daher hat es sich angeboten, in der zweiten Zeitzeug*innenstudie, die im Schuljahr 2021/22 in Baden-Württemberg durchgeführt wurde, mit dem Videomaterial und mit den Interviewpartner*innen der «Generation 1975» zu arbeiten.

4.4 Zeitzeug*innenstudie 2: Wechselseitige Perspektiven auf die Einheit

Forschungsfragen: In der von der Deutschen Forschungsgemeinschaft (DFG) geförderten Studie «Zeitzeugen⁴ im Geschichtsunterricht als Maßnahme zur Förderung historischer Kompetenzen: Eine cluster-randomisierte kontrollierte Interventionsstudie» gehen wir davon aus, dass der Einsatz von Zeitzeug*innen zu einer erhöhten kognitiven und emotionalen Anteilnahme der Schüler*innen (Zachrich et al., in diesem Band) und damit zu mehr Motivation und Interesse führen kann. Gleichzeitig kann dies aber zu einer Überwältigung führen, die mit einem weniger stark ausgeprägten Kompetenzzuwachs einhergehen kann (Bertram et al., 2017). Untersucht wird nun, ob dieser Effekt auch gefunden wird, wenn die Intervention fachdidaktisch überarbeitet und von den eigenen Lehrkräften durchgeführt wird. Insbesondere interessiert uns, welche (differenziellen) Effekte die Arbeit mit Live- versus Video-Zeitzeug*innen hinsichtlich der Kompetenzen, des Wissens und der Motivation der Schüler*innen bewirkt und wie die-

4 Erst seit 2020 nutze ich die Sternchen bei der Pluralbildung als Alternative zum generischen Maskulinum im Plural oder Singular, um Frauen, Männer und sich divers verstehende Menschen mit einzubeziehen.

se Effekte mit ihren individuellen Verarbeitungsprozessen (der besonderen Lernerfahrung mit Zeitzeug*innen) zusammenhängen.

Design: Drei Bedingungen werden miteinander verglichen: (1) zwei Zeitzeug*innne treten live auf (27 Klassen), (2) die Aussagen zweier Zeitzeug*innen werden per Video in den Unterricht eingebettet (27 Klassen), (3) das Thema wird materialbasiert, aber ohne Zeitzeug*innen behandelt (Wartekontrollbedingung, 18 Klassen). Der Stichprobenumfang wurde auf der Grundlage einer Power-Analyse berechnet. Die Fortbildung der Lehrkräfte (Titel: «Perspektiven auf die Transformationszeit 1990–2021») wurde zu zwei Zeitpunkten im Frühjahr 2022 zentral von dem Zentrum für Schulqualität und Lehrerbildung (ZSL) Baden-Württemberg angeboten. Nach ihrer Anmeldung wurden die Lehrkräfte zufällig einer der drei Gruppen (Live-, Video-, Wartekontrollgruppe) zugewiesen, wobei die Lehrkräfte in der Wartekontrollgruppe die Schulung erst ein Jahr später erhalten werden.

Intervention: In der Intervention wird mit den Zeitzeug*innen beziehungsweise mit den Videos der «Generation 1975» gearbeitet. Die Unterrichtseinheit von drei Doppelstunden in den Experimentalgruppen (live und Video) unterscheidet sich lediglich in der Doppelstunde, in der mit den Zeitzeug*innen gearbeitet wird. Ausgehend von aktuellen Umfragen und kontroversen Karikaturen werden die Lernenden im ersten Modul auf die Begegnung mit den Zeitzeug*innen der «Generation 1975» vorbereitet. Im zweiten Modul wird mit jeweils zwei von ihnen (aufgewachsen in West- bzw. Ostdeutschland) gearbeitet, entweder live oder mit professionell aufbereiteten Videoclips. Im dritten Modul werden die Aussagen der beiden Zeitzeug*innen mithilfe vertiefender Materialien (z. B. zur Arbeitsweise der Treuhand) vergleichend ausgewertet und kontextualisiert. Unter Berücksichtigung der fachdidaktischen Prinzipien der Multiperspektivität und Kontroversität (u. a. Bergmann, 2000; Lücke, 2012; Stradling, 2003; Wansink, Akkerman, Zuiker & Wubbels, 2018) werden Narrative zur Teilung und Wiedervereinigung Deutschlands kritisch hinterfragt werden.

Instrumente: Um die Kompetenzen der Lernenden (insbesondere die Methoden-, Reflexions- und Orientierungskompetenz, Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2016) und das themenbezogene Wissen zu erfassen, wurden Testaufgaben entwickelt, die im Winter 2021/22 bezogen auf die psychometrischen Gütekriterien validiert wurden. Die zentralen Outcomes (Kompetenzen, Wissen und Motivation) wurden in der Haupterhebung zu zwei Testzeitpunkten erfasst. Unmittelbar nach der Stunde mit den Zeitzeug*innen (live respektive Video) kam das «Aura-Instrument» zum Einsatz. Im Vortest wur-

den zudem Daten zu den individuellen Voraussetzungen der Schüler*innen erhoben (z. B. soziokultureller Hintergrund, kognitive Grundfähigkeiten, Lesekompetenz, allgemeiner historischer HiTCH-Kompetenztest, Noten).

Analysen: Wie in der ersten Zeitzeug*innenstudie werden in den statistischen Verfahren die Besonderheiten der Stichprobenziehung (Schachtelung der Daten auf Individualebene) berücksichtigt und fehlende Werte durch multiple Imputation ersetzt. Im Kontext der Frage nach dem Einfluss der «Aura» der*s Zeitzeug*in sollen sogenannte *Moderated-Mediation*-Modelle verwendet werden. Dazu werden Strukturgleichungsmodelle in Mplus mit der unabhängigen Variable Gruppenzugehörigkeit und dem Mediator wahrgenommene «Aura» berechnet, der wiederum als Moderator für den Zusammenhang zwischen Interventionsbedingung und dem Outcome betrachtet wird (Preacher, Rucker & Hayes, 2007). Mithilfe des Bootstrap-Verfahrens sollen zuverlässige Signifikanztests und Konfidenzintervalle für die Moderations- und Mediationseffekte ermittelt werden.

Zwischenstand: Aufgrund der pandemiebedingten Schulschließungen musste die Pilotierung mit sechs Expert*innenlehrkräfte und die Validierungsstudie für die Erprobung der Testinstrumente in das Schuljahr 2021/22 verschoben werden. Ende des Schuljahres 2021/22 wurde die Haupterhebung durchgeführt.

5 Diskussion

In unserem Forschungsprogramm zur Wirksamkeit von Zeitzeug*innen nutzen wir den empirischen Forschungsprozess in einem Wechselspiel von Theorie, Praxis und Empirie, indem Ziele definiert und operationalisiert, Hypothesen formuliert, Interventionen entwickelt und in ihrer Wirksamkeit untersucht werden. Hierbei arbeiten wir interdisziplinär, nutzen qualitative und quantitative Methoden und schreiben mit der groß angelegten Interventionsstudie zu den Perspektiven auf die Transformationszeit nach 1989/90 den Bildungsplan 2016 in Baden-Württemberg ein Kapitel weiter.

Mit der in der Erziehungswissenschaft und Fachdidaktik zuweilen vehement vorgetragenen Kritik an einem «exklusiveren Paradigma evidenzbasierter Pädagogik», das «weder Methodenvielfalt» einschlieÙe noch «komplexere Verhältnisbestimmungen» zulasse (Bellmann & Müller, 2011, 15), und mit dem Verdacht, dass es sich hierbei «um ein politisches Programm handelt, dessen Zweck in der besseren Kontrolle der pädagogischen Praxis liegt» (Herzog, 2011, 123), werden

die experimentellen Methoden der empirischen Bildungsforschung im Ganzen abgetan. Diese pauschalisierende Kritik verkennt die differenzierten Überlegungen zu theoretischen Bestimmungen und Wirkzusammenhängen, die hier bezogen auf die Forschung zu Zeitzeug*innen dargelegt wurden.

Drei Kritikpunkte an kontrollierten Interventionsstudien sollen abschließend genauer betrachtet werden. (1) Zur Kompetenzmessung: Können mit standardisierten Aufgabenformaten Kompetenzen historischen Denkens objektiv, reliabel und valide erfasst werden? Ist die Messung nicht viel zu ungenau und erfasst sie überhaupt das, was im Geschichtsunterricht vermittelt werden soll? (2) Zur Standardisierung: Bringt die Standardisierung der Unterrichtseinheit, das Verpflichten auf ein Manual und auf einen vorgegebenen Umgang mit dem Material nicht eine Vereinheitlichung mit sich, die im Widerspruch steht zu der Individualität der professionellen Lehrkraft und ihrem Wissen, wie die jeweilige Klasse am besten arbeiten kann? (3) Zu den Ergebnissen: Das haben wir doch schon vorher gewusst. Schüler*innen sind begeistert von Zeitzeug*innen im Unterricht – deswegen laden wir sie ja ein. Lohnt sich der Aufwand überhaupt? Von diesen drei Kritikpunkte ausgehend sollen abschließend die Chancen des Forschungsansatzes für die Fachdidaktik aufgezeigt werden.

Ad 1: Ob mit standardisierten oder offenen Aufgabenformaten: Die fachspezifischen Kompetenzen von Lernenden objektiv, reliabel und valide zu erfassen ist tatsächlich eine Herausforderung. Doch schärft die Konzeptualisierung der Zielkonstrukte und die Operationalisierung in Aufgaben und Items den Blick auf die Zielsetzung von Unterricht und leistet einen Beitrag zur Theoriebildung und Aufgabengestaltung. Ad 2: Das strenge experimentelle Design verlangt einen ungewöhnlichen Grad an Standardisierung, das der eigenständigen Gestaltung von Unterrichtsszenarien, eine zentrale Aufgabe von Lehrkräften, eigentlich widerspricht. In der Pilotierung entwickeln wir die Intervention mit den Lehrkräften weiter, erproben die Praxistauglichkeit und tarieren den notwendigen Grad zwischen Standardisierung und Offenheit aus, um Lehrkräfte optimal in ihrem Unterricht zu unterstützen. Ad 3: Die erste Studie hat das motivationale Potenzial der Methode, aber auch die Risiken der Arbeit mit Zeitzeug*innen bestätigt. Ziel ist es nun herauszufinden, wie dieses Potenzial einer besonderen Lernerfahrung für die Motivation und für das fachliche Lernen genutzt werden kann.

Mit empirischer fachdidaktischer Forschung soll keine Vermutung bestätigt oder letzte Antworten gegeben werden, sondern sie schärft den Blick für Chancen und Risiken und soll dabei helfen, den Geschichtsunterricht zu verbessern. Der vorliegende Beitrag sollte aufzeigen, wie mithilfe der evidenzbasierten Wirk-

samkeitsforschung die fachdidaktische Theorie und Pragmatik weiterentwickelt werden kann. Die geschichtsdidaktisch motivierte Forschung führte zu der Erschließung des Quellenkorpus der «Generation 1975», der in der Geschichtskultur und im Geschichtsunterricht genutzt wird, um Denkprozesse anzustoßen, die zu einem größeren Verständnis zwischen den Menschen in Ost- und Westdeutschland führen können, sodass die Geschichtsdidaktik im Sinne Jeismanns (1977) und Rüsens (1983) in die Gesellschaft hinein wirksam werden kann.

Literatur

- Ahbe, Thomas & Gries, Rainer (2006). Gesellschaftsgeschichte als Generationengeschichte: Theoretische und methodische Überlegungen am Beispiel der DDR. In Annegret Schüle, Thomas Ahbe & Rainer Gries (Hrsg.), *Die DDR aus generationengeschichtlicher Perspektive: Eine Inventur* (S. 475–571). Leipzig: Leipziger Universitätsverlag.
- Angvik, Magne & von Borries, Bodo (Hrsg.) (1997). *Youth and history: A comparative European survey on historical consciousness and political attitudes among adolescents*. Hamburg: Körber-Stiftung.
- Baker, Eva L. (2006). Principles for scaling up: Choosing, measuring effects, and promoting the widespread use of educational innovation. In Barbara Schneider & Sarah-Kathryn McDonald (Hrsg.), *Scale-up in education: Ideas in principle* (S. 37–54). Lanham: Rowman & Littlefield.
- Barricelli, Michele & Lücke, Martin (Hrsg.) (2012). *Handbuch Praxis des Geschichtsunterrichts*. Schwalbach/Ts.: Wochenschau.
- Bertram, Christiane (2020). «Generation 1975 – Mit 14 ins neue Deutschland»: Blick vom Osten und Westen auf die deutsche Teilungsgeschichte. *Bürger & Staat*, (1–2), 81–88.
- Bertram, Christiane (2017). *Zeitzeugen im Geschichtsunterricht. Chance oder Risiko für historisches Lernen? Eine randomisierte Interventionsstudie*. Schwalbach/Ts.: Wochenschau.
- Bertram, Christiane/Wagner, Wolfgang & Trautwein, Ulrich (2017). Learning historical thinking with oral history interviews: A cluster randomized controlled intervention study of oral history interviews in history lessons. *American Educational Research Journal*, 54(3), 444–484. <https://doi.org/10.3102/0002831217694833>.
- Bösch, Frank (2015). Geteilt und verbunden: Perspektiven auf die deutsche Geschichte seit den 1970er-Jahren. In Frank Bösch (Hrsg.), *Geteilte Geschichte: Ost- und Westdeutschland 1970–2000* (S. 7–37). Göttingen: Vandenhoeck & Ruprecht.
- Borsutzki, Doreen/Machowecz, Martin & Wefing, Heinrich (2019). Jetzt hört mal zu! *Die Zeit*, 41, 19.
- Brüning, Christina Isabel (2018). *Holocaust Education in der heterogenen Gesellschaft: Eine Studie zum Einsatz videographierter Zeugnisse von Überlebenden der nationalsozialistischen Genozide im Unterricht*. Schwalbach/Ts.: Wochenschau.
- Brunner, Martin/Keller, Ulrich/Wenger, Marina/Fischbach, Antoine & Lüdtke, Oliver (2018). Between-school variation in students' achievement, motivation, affect, and learning strategies: Results from 81 countries for planning group-randomized trials

- in education. *Journal of Research on Educational Effectiveness*, 11(3), 452–478. <https://doi.org/10.1080/19345747.2017.1375584>.
- Classen, Christoph (2012). Der Zeitzeuge als Artefakt der Medienkonsumgesellschaft. Zum Verhältnis von Medialisierung und Erinnerungskultur. In Martin Sabrow & Norbert Frei (Hrsg.), *Die Geburt des Zeitzeugen nach 1945* (S. 300–319). Göttingen: Wallstein.
- Geppert, Alexander C. T. (1994). Forschungstechnik oder historische Disziplin? Methodische Probleme der Oral History. *Geschichte in Wissenschaft und Unterricht*, 45(5), 303–323.
- Gottfredson, Denise C./Cook, Thomas D./Gardner, Frances E. M./Gorman-Smith, Deborah/Howe, George W./Sandler, Irwin N. & Zafft, Kathryn M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science: The Official Journal of the Society for Prevention Research*, 16(7), 893–926. <https://doi.org/10.1007/s11121-015-0555-x>.
- Großbölting, Thomas (2020). *Wiedervereinigungsgesellschaft: Aufbruch und Entgrenzung in Deutschland seit 1989/90*. Bonn: Bundeszentrale für politische Bildung.
- Hager, Willi & Hasselhorn, Marcus (2000). Psychologische Interventionsmaßnahmen: Was sollen sie bewirken können? In Willi Hager/Jean-Luc Patry & Hermann Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen* (S. 41–85). Bern: Huber.
- Hattie, John (2009). *Visible learning: A Synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Herbein, Evelyn/Golle, Jessica/Nagengast, Benjamin & Trautwein, Ulrich (2020). Förderung von Präsentationskompetenz: Schrittweise Implementation und Effektivitätsüberprüfung eines Präsentationstrainings für Grundschulkindern. *Zeitschrift für Erziehungswissenschaft*, 23(1), 83–120. <https://doi.org/10.1007/s11618-019-00923-y>.
- Huber, Oswald (2013). *Das psychologische Experiment: Eine Einführung*. Bern: Hans Huber.
- Kleßmann, Christoph (2006). Spaltung und Verflechtung – Ein Konzept zur integrierten Nachkriegsgeschichte 1945 bis 1990. In Christoph Kleßmann & Peter Lautzas (Hrsg.), *Teilung und Integration. Die doppelte deutsche Nachkriegsgeschichte als wissenschaftliches und didaktisches Problem* (S. 20–37). Schwalbach/Ts.: Wochenschau.
- Körber, Andreas/Schreiber, Waltraud & Schöner, Alexander (Hrsg.) (2007). *Kompetenzen historischen Denkens: Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik*. Neuried: ars una.
- Kowalczyk, Ilko-Sascha (2019). *Die Übernahme: Wie Ostdeutschland Teil der Bundesrepublik wurde* (4. Aufl.). München: C. H. Beck.
- Kraft, Matthew A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>.
- Kuchler, Christian & Sommer, Andreas (Hrsg.) (2018). *Wirksamer Geschichtsunterricht. Unterrichtspraxis: Perspektiven von Expertinnen und Experten*. Baltmannsweiler: Schneider Hohengehren.
- Kuckartz, Udo (2010). *Einführung in die computergestützte Inhaltsanalyse*. 3., aktualisierte Ausgabe. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kultusministerkonferenz (2009). *Demokratie als Ziel, Gegenstand und Praxis historisch-politischer Bildung und Erziehung in der Schule. Beschluss der Kultusministerkonferenz vom 06.03.2009 i. d. F. vom 11.10.2018*. <https://www.kmk.org/fileadmin/veroeffentlichungen>

- gen_beschluesse/2009/2009_03_06-Staerkung_Demokratieerziehung.pdf (21.04.2022).
- Lareau, Annette (2009). Narrow questions, narrow answers: The limited value of randomized controlled trials for education research. In Pamela Walters/Annette Lareau & Sheri Ranis (Hrsg.), *Education research on trial: Policy reform and the call for scientific rigor* (S. 145–161). New York: Routledge.
- Leutner, Detlef (2010). Perspektiven pädagogischer Interventionsforschung. In Tina Ha-scher & Bernhard Schmitz (Hrsg.), *Pädagogische Interventionsforschung: Theoretische Grundlagen und empirisches Handlungswissen* (S. 63–72). Weinheim: Juventa.
- Mannheim, Karl (1928/29). Das Problem der Generation. *Kölner Zeitschrift für Soziologie*, 7, 157–184.
- Mau, Stefan (2019). *Lütten Klein: Leben in der ostdeutschen Transformationsgesellschaft* (2. Aufl.). Berlin: Suhrkamp.
- Merkel, Angelika (2019). Rede von Bundeskanzlerin Merkel anlässlich des Festakts zum Tag der Deutschen Einheit am 3. Oktober 2019 in Kiel. Presse- und Informationsamt der Bundesregierung. <https://www.bundeskanzlerin.de/bkin-de/aktuelles/rede-von-bundeskanzlerin-merkel-anlaesslich-des-festakts-zum-tag-der-deutschen-einheit-am-3-oktober-2019-in-kiel-1678326> (18.03.2021).
- Moher, David et al. (2010). CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*, 340. <https://doi.org/10.1136/bmj.c869>.
- Nelson, Michael C./Cordray, David S./Hulleman, Chris S./Darrow, Catherine L. & Sommer, Evan C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *Journal of Behavioral Health Services and Research*, 39(4), 374–396. <https://doi.org/10.1007/s11414-012-9295-x>.
- Niethammer, Lutz (Hrsg.) (1983). «Die Jahre weiß man nicht, wo man die heute hinsetzen soll». *Faschismuserfahrungen im Ruhrgebiet. Lebensgeschichte und Sozialkultur im Ruhrgebiet*. Bonn: Dietz.
- Obertreis, Julia (Hrsg.) (2012). *Oral History*. Stuttgart: Franz Steiner.
- Plato, Alexander von (2000). Zeitzeugen und die historische Zunft: Erinnerung kommunikative Tradierung und kollektives Gedächtnis in der qualitativen Geschichtswissenschaft – ein Problemaufriss. *BIOS*, 13, 5–29.
- Rothfels, Hans (1953). Zeitgeschichte als Aufgabe. *Vierteljahresheft für Zeitgeschichte*, 1(1), 1–8.
- Rüsen, Jörn (1983). *Historische Vernunft: Die Grundlagen der Geschichtswissenschaft. Grundzüge einer Historik I*. Göttingen: Vandenhoeck & Ruprecht.
- Sabrow, Martin (2012). Der Zeitzeuge als Wanderer zwischen zwei Welten. In Martin Sabrow & Norbert Frei (Hrsg.), *Die Geburt des Zeitzeugen nach 1945* (S. 13–32). Göttingen: Wallstein.
- Sabrow, Martin & Frei, Norbert (Hrsg.) (2012). *Die Geburt des Zeitzeugen nach 1945*. Göttingen: Wallstein.
- Sauer, Michael (2006). *Geschichte unterrichten. Eine Einführung in die Didaktik und Methodik*. Seelze: Kallmeyer.

- Shadish, William R./Cook, Thomas D. & Campbell, Donald T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth.
- Shafrey, Jason (2016). Efficacy vs. effectiveness vs. efficiency. In *Healthcare Economist*. <https://www.healthcare-economist.com/2016/01/25/efficacy-vs-effectiveness-vs-efficiency/> (21.04.2022).
- Skriebeleit, Jörg (2011). *Das Verschwinden der Zeitzeugen. Metapher eines Übergangs*. Vortrag gehalten bei der Tagung «Zeitzeugen im Museum», Görlitz, 12.10. 2011 bis 14.10.2011. https://www.bkge.de/Downloads/Zeitzeugenberichte/Skriebeleit_Verschwinden_der_Zeitzeugen.pdf (07.04.2022).
- Slavin, Robert E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31(7), 15–21. <https://doi.org/10.1007/s11414-012-9295-x>.
- Spohn, Wolfgang (1983). *Eine Theorie der Kausalität*. Habilitation. München: Ludwig-Maximilians-Universität.
- Ther, Philipp (2014). *Die neue Ordnung auf dem alten Kontinent: Eine Geschichte des neoliberalen Europas*. Frankfurt a. M.: Suhrkamp.
- Thomson, Alistair (2007). Four paradigm transformations in oral history. *The Oral History Review*, 34(1), 49–70. <http://dx.doi.org/10.1525/ohr.2007.34.1.49>.
- Thukydides (411 v. Chr./1991). *Der Peloponnesische Krieg*. Übersetzt von G. P. Landmann. München: dtv.
- Torgerson, David & Torgerson, Carole (2013). *T1 – Randomised trials in education: An introductory handbook*. https://www.researchgate.net/publication/273421357_Randomised_trials_in_education_An_introduutory_handbook/citation/download (09.03.2021).
- Ullmann, Christine (2021). «Jammer-Ossi» und «Besser-Wessi»? *Gegenseitige Wahrnehmung in Ost- und Westdeutschland vor und nach 1990 am Beispiel der Interviewstudie «Generation 1975»*. Universität Konstanz (Masterarbeit).
- Weber, Petra (2020). *Getrennt und doch vereint: Deutsch-deutsche Geschichte 1945–1989/90* (2. Aufl.). Berlin: Monopol.
- Wierling, Dorothee (2002). *Geboren im Jahr Eins: Der Jahrgang 1949 in der DDR. Versuch einer Kollektivbiographie*. Berlin: Ch. Links.
- Wierling, Dorothee (2003). Oral History. In Michael Maurer (Hrsg.), *Neue Themen und Methoden der Geschichtswissenschaft* (S. 81–151). Stuttgart: Reclam.
- Wierling, Dorothee (2014). Oral History und Zeitzeugen in der politischen Bildung: Kommentar zu einem Spannungsverhältnis. In Christian Ernst (Hrsg.), *Geschichte im Dialog? DDR-Zeitzeugen in Geschichtskultur und Bildungspraxis* (S. 99–107). Schwalbach/Ts.: Wochenschau.
- Zachrich, Lisa/Baron, Chris/Weller, Allison & Bertram, Christiane (2020). Historical experiences: A framework for encountering complex historical sources. *History Education Research Journal (HERJ)*, 17(2), 243–275.

Leseförderung im Geschichtsunterricht?

Eine Interventionsstudie zur Wirksamkeit adaptiver Sprachunterstützung bei der Erschließung fachspezifischer Sachtexte

Helene Bergmann

1 Einleitung

Sprachbildung und -förderung im Geschichtsunterricht sind im Kontext der Diskussion über den «sprachsensiblen Fachunterricht» verstärkt ins Blickfeld der geschichtsdidaktischen Forschung gerückt. Aktuelle, oft qualitativ ausgerichtete Studien fokussieren etwa Sprachmerkmale fachspezifischer Texte, sprachbildende Aufgaben oder narrative Kompetenzen.¹ Rezeptive Sprachhandlungen wie fachspezifische Leseprozesse sind dagegen in der deutschsprachigen fachdidaktischen Forschung bislang eher unterbelichtet. Und selbst die in diesem Bereich weiter entwickelte angloamerikanische Forschung weist einen Mangel an quantitativen Wirksamkeitsstudien auf (etwa die Metaanalyse von McCully & Osman, 2015, 183–195).

Hier setzt das nachfolgend vorgestellte Promotionsprojekt an, das auf die Entwicklung und Evaluation eines binnendifferenzierenden und sprachsensiblen Unterrichtskonzepts für die Textarbeit im Fach Geschichte zielt. Dabei wird von einem engen Zusammenhang zwischen den Lesefähigkeiten der Lernenden und ihrem fachlichen Lernerfolg ausgegangen. Inwieweit letzterer durch fachspezifisch profilierte Lesefördermaßnahmen effektiviert werden kann, wird in einer (quasi-)experimentellen Studie mit mehreren Testzeitpunkten untersucht, in der die Leseunterstützung je nach Sprachvoraussetzung systematisch variiert wird. Damit soll ein praxisrelevanter Beitrag zu einer empirisch fundierten Weiterentwicklung eines adaptiven und sprachbewussten Geschichtsunterrichts geleistet werden.

¹ Vgl. dazu etwa die Themenschwerpunkte der letzten KGD-Tagung «Sprachen(n) des Geschichtsunterrichts» (Sandkühler & Bernhardt, 2019, 5–7).

2 Erkenntnisinteresse, Hypothesen und Theoriehintergründe

Das übergeordnete Ziel der Interventionsstudie besteht darin herauszufinden, welche Formen von Leseunterstützung sich eignen, um Schüler*innen mit unterschiedlichen Lern- und Sprachvoraussetzungen das Verständnis fachspezifischer Sachtexte zu erleichtern. Während die Voraussetzungen und Herausforderungen der Quelleninterpretation in der Geschichtsdidaktik schon länger diskutiert werden und zahlreiche methodische Ideen zum unterrichtlichen Umgang mit Quellen existieren (vgl. etwa die einschlägigen Handbücher von Hans-Jürgen Pandel und Michael Sauer), existieren vergleichsweise wenig Vorschläge zur Arbeit mit fachspezifischen Darstellungstexten (z. B. Rox-Helmer, 2010, 188; Mainzer-Murrenhoff & Mierwald, 2018, 46–55; Kühberger & Windischbauer, 2013, 102–105). Aus diesem Grund liegt der Fokus dieser Studie zu fachspezifischer Leseförderung auf den in der Praxis viel rezipierten Darstellungstexten aus Geschichtsschulbüchern (dazu Sauer, 2018a, 406–417). Um sie zu verstehen, bedarf es neben historischem Wissen allgemeiner und fachspezifischer Lesekompetenz.

Während man lange Zeit den Deutschunterricht als einzigen Ort zur Vermittlung und Einübung von Texterschließungsstrategien ansah, wird seit einigen Jahren auch auf die Verantwortung weiterer Schulfächer bei der Förderung fachspezifischer Lesekompetenz hingewiesen. So etwa im Bericht zur letzten PISA-Studie, wo auch der Geschichtsunterricht explizit angesprochen und der Stellenwert von Lesekompetenz beim historischen Lernen betont wird (Becker-Mrotzek et al., 2018, 22). Die nachfolgend vorgestellte Interventionsstudie greift dieses aktuelle Thema auf und erweitert die Frage nach wirksamer fachspezifischer Leseunterstützung um den Aspekt des binnendifferenzierenden historischen Lernens, welches angesichts gesellschaftlich-politischer Entwicklungen wie Migration, Inklusion und Schulstrukturreformen ebenso wie fachintegrierte Sprachförderung an Bedeutung gewinnt. Um Erkenntnisse dazu zu erzielen, welche Art von fachinte-

grierter Leseförderung² am geeignetsten ist, um Lernende mit unterschiedlichen Lern- und Sprachvoraussetzungen bei der Erschließung fachspezifischer Texte wirksam zu unterstützen, wird ein adaptives Unterrichtskonzept zum Umgang mit Sachtexten des Geschichtsunterrichts entwickelt und mit vorwiegend standardisierten Methoden beforscht. Neben Lesestrategieanwendungsaufgaben, die den Kern der Sprachfördermaßnahme bilden, beinhaltet diese gruppenabhängig weitere sprachliche Hilfen, mit denen die Lernenden abhängig von den zuvor erhobenen Fähigkeiten im Bereich Textverstehen differenziell gefördert werden. Diesem Vorgehen liegen folgende, aus der Unterrichtsforschung und der sprachwissenschaftlichen Forschung abgeleitete Hypothesen zugrunde.

Zum einen wird anknüpfend an Befunde aus den Bereichen Sprachdidaktik und DaZ angenommen, dass zwischen dem fachlichem Lernerfolg und den allgemeinen Sprach- und Lesefähigkeiten ein starker Zusammenhang besteht und dass demnach eine sprachensible Lernumgebung das fachliche Verstehen auch im Fach Geschichte erleichtert (etwa Ahrenholz, Jeuk, Lütke, Paetsch & Roll, 2019, 1–16; Beese & Benholz, 2013, 37–50; Artelt et al., 2002, 20). Zum anderen wird in Anknüpfung an die Aptitude-Treatment-Interaction-Forschung³ davon ausgegangen, dass die Effektivität unterrichtlicher Instruktionen (*treatment*) von den Merkmalen der Lernenden (*aptitude*) abhängt und Lehrmethoden daher an die

-
- 2 Wenn nachfolgend von Leseförderung gesprochen wird, so sind damit auch, aber nicht nur sprachliche Hilfen für Lernende mit diagnostizierten Lern- und Sprachschwächen gemeint, die häufig mit dem Begriff «Förderung» assoziiert werden (Achour & Sieberkrob, 2015, 18). Der Begriff «Leseförderung» wird hier stattdessen in einem weiteren Sinn genutzt, um differenzierende sprachliche Hilfestellungen zu kennzeichnen, die alle Schüler*innen einer Lerngruppe mit ihren jeweils unterschiedlichen Sprach- und Lernvoraussetzungen, Förderbedarfen, aber auch Begabungen möglichst optimal unterstützen. Der «sprachensible Geschichtsunterricht» wird demzufolge als ein Konzept begriffen, das sich aufgrund des konstitutiven Verhältnisses von historischem und sprachlichem Lernen grundsätzlich auf alle Lernenden einer Schulklasse bezieht (etwa Schrader, 2019, 125), welche jedoch im idealen Fall differenzielle sprachliche Förderung erhalten.
 - 3 Die Aptitude-Treatment-Interaction-Forschung, nachfolgend als ATI-Forschung abgekürzt, ist ein in den 70er Jahren von Cronbach und Snow begründeter Zweig der allgemeinen Lehr-Lern-Forschung (1989, 13–59). Der zentrale Begriff «Aptitude-Treatment-Interaktion» bezeichnet die Erkenntnis der pädagogisch-psychologischen Forschung, dass die Effektivität eines Lernarrangements von den Lernvoraussetzungen, Begabungen und Förderbedarfen der Lernenden beeinflusst wird. Für Unterrichtskonzeptionen, die diese unterschiedlichen Lernbedarfe innerhalb einer Lerngruppe berücksichtigen, haben sich Begriffe wie «Differenzieren», «Individualisierung», «individuelle Förderung» oder «adaptiver Unterricht» etabliert (Fischer, 2013). Von der allgemeinen und fachdidaktischen Lehr-Lern-Forschung wurden solche differenziellen Förderansätze bisher vor allem auf Individualebene und häufig unter Laborbedingungen erforscht, aber erst ansatzweise im Rahmen von binnendifferenzierendem realem Klassenunterricht (Friesen et al., 2021, 42–43).

Fähigkeiten der SchülerInnen angepasst werden müssen, um optimale Lernergebnisse zu erzielen (Snow, 1989, 13–59).

So besagt etwa eine relativ gut abgesicherte ATI-Hypothese, dass Lernende mit ungünstigeren Lernvoraussetzungen von stärkerer Strukturierung und zusätzlichen Instruktionen profitieren, Lernende mit günstigeren Lernvoraussetzungen hingegen von der Anregung selbstgesteuerten Lernens (Kalyuga, 2003, 23–31; Kalyuga & Renkl, 2009, 209–214). Auf diese und weitere ATI-Annahmen, wie etwa die Erkenntnisse zu Expertise-Umkehr-Effekten und Redundanzeffekten, die zeigen, dass unpassende Instruktionen das Lernen sogar stark beeinträchtigen können (etwa Wittwer & Renkel, 2008, 58–60), baut der beforschte adaptive Geschichtsunterricht auf. Da die genannten ATI-Effekte bisher vor allem unter Laborbedingungen untersucht wurden (Friesen et al., 2021, 42–43), ist dabei auch von Interesse, inwieweit sie sich in einem ökologisch validen Setting wie dem realen Klassenraum nachweisen lassen.

Neben den ATI-Prinzipien und den bereits genannten Befunden der sprachwissenschaftlichen Forschung zum Zusammenhang von sprachlichen Fähigkeiten und fachlichem Lernerfolg knüpft die Studie außerdem an Ergebnisse der Leseforschung an, die zeigen, dass Textverstehen besonders effektiv und nachhaltig durch Lesestrategievermittlung gefördert werden kann (die Metaanalyse von Schneider et al., 2013, 22–26; Philipp & Schilcher, 2012, 59–73). Dabei scheinen insbesondere diejenigen Schüler*innen von Strategiearbeit zu profitieren, die zwar flüssig lesen, aber Informationen nicht oder nur unzureichend verarbeiten können (Rosebrock, 2012, 8). Es ist anzunehmen, aber noch zu belegen, dass dies auch für die durch Strategieanwendungsaufgaben unterstützte Textarbeit im Geschichtsunterricht zutrifft, von der ein besseres Textverständnis und eine positive Wirkung auf den fachlichen Kompetenzaufbau erwartet wird. Da die auf sechs Unterrichtsstunden exklusive Prä- und Posttests angelegte Intervention zu einem Thema der altägyptischen Geschichte einen kürzeren Zeitraum umfasst als andere, auf längerfristige und systematische Lesekompetenzförderung ausgerichtete Förderprogramme wie etwa das KMK-Großprojekt «ProLesen. Auf dem Weg zur Leseschule», soll sie auch Hinweise dazu erbringen, welche Art von Leseunterstützung geeignet ist, um kurzfristig das Verständnis historischer Fachtexte zu verbessern. Erkenntnisse hierzu wären im Sinne niederschwelliger, unabhängig von längerfristigen gesamtschulischen Projekten realisierbarer fachspezifischer Lesefördermaßnahmen wünschenswert.

Denn ohne dass Texte des Geschichtsunterrichts verstanden werden, scheitern Lernende an den nachfolgenden Aufgaben und gelangen gar nicht erst zu

anspruchsvolleren fachlichen Denkkoperationen wie historischem Analysieren und Urteilen. Die Bedeutung rezeptiver Sprachhandlungen im Geschichtsunterricht darf also nicht unterschätzt werden. Sie bedingen die Aneignung von historischem Wissen als Grundlage eigener Rekonstruktionen und Narrationen. Wie gleich in einem kurzen Abriss der entsprechenden fachdidaktischen Forschung gezeigt wird, ist der Fokus der geschichtsdidaktischen Forschung derzeit aber überwiegend auf die aktiven Sprachhandlungen gerichtet, denen beim Aufbau historischer Kompetenzen unzweifelhaft eine große Bedeutung zukommt. Eine Ergänzung durch empirische, vor allem auch quantitative Studien zu fachspezifischen rezeptiven Fertigkeiten und Prozessen erscheint dennoch wichtig. Denn, wie Rox-Helmer es ausdrückt: «Geschichte-Lernen bedeutet Lesen» (2010, 183).

Mit Blick auf den realen Geschichtsunterricht scheint es allerdings, als fände eine systematische und explizite fachspezifische Leseförderung (noch) nicht statt. Die Anzahl der bislang erhältlichen, meist als Beihefte zu aktuellen Geschichtslehrwerken herausgegebenen fachspezifischen Sprachfördermaterialien ist entsprechend noch überschaubar (etwa Nartschick, 2012; Dzubiel & Harcom, 2018). Sie basieren meist auf theoretisch plausiblen oder intuitiven Wirksamkeitsannahmen (etwa Nartschick, 2012, 6–7), nicht aber auf systematischen Untersuchungen der Effektivität der jeweils empfohlenen Sprachfördermethoden. Hier besteht also weiterer Forschungsbedarf, um fachspezifisches Sprachfördermaterial künftig auf empirisch abgesicherter Grundlage entwickeln zu können.

Auch der Aspekt der Adaptivität, also der individuell oder gruppenweise angepassten Lernunterstützung ist dabei von großer Bedeutung, um allen Schülerinnen und Schülern fachlichen Lernerfolg zu ermöglichen.⁴ Für den Geschichtsunterricht werden binnendifferenzierende historische Lernsettings zwar hin und wieder in Handbüchern oder Praxiszeitschriften thematisiert und methodisch konkretisiert (z. B. Adamski, 2017, 2–11; Kühberger & Windischbauer, 2009, 22–29; Kühberger & Windischbauer, 2013, 102–105). In fachdidaktischen Studien finden sie dagegen eher selten Berücksichtigung, sondern eher in der Integrationsforschung (z. B. Tillmann & Wischer, 2006, 44–48) oder in der allgemeinen Unterrichtsforschung. Da deren Befunde zu Differenzierungsstrategien im regulären Klassenunterricht jedoch uneinheitlich sind (die Metaanalyse von Deunk

4 Vgl. zu Prinzipien, Hintergründen und Beispielen flexibler gruppenweiser Differenzierung im Fachunterricht etwa Friesen, Leuders & Loibl, 2021, 42–45.

et al., 2015, 47–54; van Geel et al., 2018, 12–15), gilt es, diese weiter empirisch abzusichern und im Hinblick auf die Erfordernisse des Geschichtsunterrichts zu spezifizieren.

3 Geschichtsdidaktischer und linguistischer Forschungskontext

Historisches Denken und Lernen ist elementar auf Sprache angewiesen. Diese Erkenntnis ist für Geschichtsdidaktiker*innen nicht neu. Seit den 1970er-Jahren befassen sich theoretische und empirische fachdidaktische Arbeiten etwa anlässlich historischen Begriffslernens, anlässlich von Quellenarbeit als fachspezifischer Erkenntnismethode oder anlässlich historischer Narrationskompetenz mit dem Zusammenhang von Sprache und historischem Lernen (Handro, 2016, 270–273). Angesichts zunehmend heterogener Lerngruppen und Forderungen der linguistischen Forschung nach einer Neubewertung der sprachlichen Anteile fachlichen Lernens (z. B. Beese & Benholz, 2013, 50) findet jedoch seit einigen Jahren eine «geschichtsdidaktische Neuakzentuierung sprachlichen Lernens» (Handro, 2016, 273) statt. Diese spiegelt sich in immer zahlreicher werdenden Publikationen zum sogenannten sprachsensiblen Geschichtsunterricht (etwa Buck, 2019, 5–11; Bernhard & Conrad, 2018, 2–9) und in einer stetig zunehmenden Zahl entsprechender Forschungsarbeiten.

Diese sind oft qualitativ ausgerichtet und untersuchen etwa Sprachanforderungen und Sprachhandlungsprozeduren in Schulbuch- oder Schülertexten (etwa Oleschko, 2015, 87–103; Oleschko & Moraitis, 2012, 114–116), sprachbildende Aufgaben (etwa Bramann & Brauch, 2020) und verschiedene Aspekte narrativer Kompetenz (etwa Brauch, Heine & Bramann, 2020; Husemann, 2020; Wickner, 2018, 38–45). Quantitative Wirksamkeitsstudien zu Effekten konkreter Sprachfördermaßnahmen im Geschichtsunterricht sind dagegen selten. Es scheint, als ob der von Eckerth & Resch 2016 in einem Tagungsrückblick festgestellte Mangel an fachspezifischen Studien zur Wirksamkeit von Scaffolding-Konzepten⁵ weiterhin fortbesteht, sieht man von noch unveröffentlichten und

5 Der Begriff «Scaffolding» bezeichnet nach Gibbons Unterstützungssysteme im (sprachsensiblen) Fachunterricht, die Schüler*innen, deren Muttersprache von der Unterrichtssprache abweicht, bei der Aneignung neuen fachlichen und sprachlichen Wissens helfen sollen (2009, 15). Laut Kniffka (2016) ist die Begriffsverwendung in der sprach- und erziehungswissenschaftlichen Literatur aber nicht einheitlich.

sprachwissenschaftlichen Studien ab, wie etwa die der Linguistin Marion Bönighausen (2019, 165–222) zur Wirksamkeit von Lesefördermaßnahmen in verschiedenen Schulfächern, darunter auch das Fach Geschichte.

Die angloamerikanische fachdidaktische Forschung zu Leseprozessen im Geschichtsunterricht (beziehungsweise im Schulfach *social studies*, welches auch die historische Bildung miteinschließt) ist dagegen weiter fortgeschritten und enthält Hinweise auf positive Effekte fachspezifischer Lesefördermaßnahmen. Darunter ist auch eine kleine Anzahl von als sehr valide geltenden experimentellen Studien, die signifikante Effekte von Leseförderung auf das Verständnis fachspezifischer Texte nachweisen (die Metaanalyse von McCully & Osman, 2015, 183–195). Jedoch bedarf es laut McCully und Osman (2015, 183–195) weiterer quantitativer Wirksamkeitsstudien und Replikationsstudien, um diese Effekte abzusichern. Zudem fehlen meist Hinweise zur differenziellen Wirkung von Lesefördermaßnahmen in leistungsheterogenen Klassen (McCully & Osman, 2015, 183–195; Reisman & McGrew, 2018; Kareva & Echevarria, 2013, 239–550; Short et al., 2011, IX).

Wie eingangs skizziert, sind für diese Studie auch empirische Befunde aus den Bereichen Sprachdidaktik und DaZ relevant, die zeigen, dass mangelnde Sprachkompetenz das Verstehen fachlicher Inhalte erschwert. Als Grund hierfür wird angenommen, dass die erfolgreiche Aneignung fachlichen Wissens die Beherrschung bildungs- und fachsprachlicher Normen voraussetzt. Lernende müssen also über Kenntnisse bezüglich fachspezifischer Sprachregister, -handlungen und Textgattungen verfügen, um in der Schule erfolgreich zu sein (Kniffka & Roelcke, 2016, 23; Beese & Benholz, 2013, 37–50).⁶ Aus der sprachwissenschaftlichen Forschung ist überdies die durch zahlreiche empirische Untersuchungen abgesicherte Erkenntnis relevant, wonach Lesestrategien einen der erfolgreichsten Ansätze zur nachhaltigen Förderung des Textverstehens und der Lesekompetenz darstellen (die Metaanalyse von Schneider et al., 2013,

6 Diese Annahme wird auch durch die mathematikdidaktische und die naturwissenschaftliche fachdidaktische Forschung gestützt, die ebenfalls Hinweise auf einen engen Zusammenhang zwischen der Sprach- und insbesondere der Lesekompetenz der Lernenden und deren fachlichem Lernerfolg liefert. So wurden für das mathematische Lernen bereits mehrfach signifikante Zusammenhänge zwischen der Lesekompetenz und der mathematischen Ausgangskompetenz sowie den fachlichen Lernzuwachsdaten der Lernenden nachgewiesen (etwa Decristan, 2015, 365–370; Prediger et al., 2014, 77–104).

22).⁷ Aus diesem Grund bilden unterschiedliche Varianten von Strategieanwendungsaufgaben den zentralen Förderansatz der hier vorgestellten Intervention.

Von geschichtsdidaktischen Autor*innen wurden Lesestrategien als sprachliches Unterstützungssystem bei der Textarbeit bislang vereinzelt, zuletzt von Handro & Kilimann (2019), ins Gespräch gebracht; sowie zuvor etwa auch von Sauer (2018b), Dietrich (2009) und Rox-Helmer (2010). Fachspezifische Konkretisierungen in Form von Strategieanwendungsaufgaben und konkreten Unterrichtsarrangements (z. B. Kühberger & Windischbauer, 2009, 22–29), sind jedoch eher selten. Aktuelle, als «sprachsensibel» und «differenzierend» betitelte Geschichtsschulbücher basieren häufig auf variabel nutzbaren historischen Lernaufgaben in Verbindung mit Textvereinfachungen und -kürzungen. Die nicht nur positiven Auswirkungen solcher Textvereinfachungen auf den fachlichen Gehalt historischer Darstellungstexte zeigen Degener und Kreutz (2021, 233–252) in einer Analyse von Verfasser*intexten verschiedener Geschichtslehrbücher. Bis dato erscheint es so, als ob dort eher defensive statt offensiver, auf nachhaltige Lesekompetenzförderung ausgerichtete sprachlicher Unterstützungsstrategien umgesetzt würden.

In gängigen Leseförderprogrammen für den Deutschunterricht, die zum Teil auch beanspruchen, fächerübergreifend anwendbar zu sein, und meist auf der Einübung von kognitiven und metakognitiven Lesestrategien in Verbindung mit reziprokem Lernen basieren, fehlen dagegen Aufgabenbeispiele für den Geschichtsunterricht (z. B. «Wir sind Lesedetektive» von Rühl & Souvignier, 2006; oder «Lesend lernen» von Bönninghausen & Winter, 2012). Ein Vorteil wiederum ist, dass diese und andere Programme bereits evaluiert und in ihrer Wirksamkeit bestätigt wurden (zur Effektivität einzelner Lesekompetenztrainings die Metaanalyse von Schneider et al., 2013). Es erscheint deshalb sinnvoll, bei der Konzeption fachspezifischer Lesefördermaßnahmen an bereits etablierte Lesetrainings

7 Im Unterschied zu Leseanimationen oder Lautleseverfahren sind Lesestrategien primär auf die Erschließung des inhaltlichen Textzusammenhangs ausgerichtet und tragen dadurch viel stärker zur Verbesserung des Leseprozesses und des Textverstehens bei (Rosebrock, 2012, 8). Sie können als «mentale Schritte, mit denen man sich einen Text gezielt und bewusst eigenständig erarbeitet» (ebd.) beschrieben werden und helfen der Leserin und dem Leser, die in einem Text enthaltenen Informationen zu erschließen und in ein Gesamtkonstrukt, ein mentales Repräsentationsmodell, einzubinden, die sogenannte globale Kohärenzbildung (Schneider et al., 2013, 22). Inzwischen existieren umfangreiche empirische Belege, wonach lesestrategisches Wissen und Können Textverstehen erheblich und nachhaltig unterstützt; vorausgesetzt, die Lesestrategien werden adäquat vermittelt, ausreichend geübt und in ihrer Funktion erklärt (Philipp, 2015; Philipp & Schilcher, 2012, 59–73).

anzuknüpfen und diese fachspezifisch zu profilieren, um statt allgemeiner Lesekompetenz fachbezogene Lesefähigkeiten und historisches Lernen zu fördern. Was dies bedeutet, kann an dieser Stelle nur kurz umrissen werden. Wichtig erscheinen unter anderem diese Aspekte:

- Ausrichtung der Lesemethoden an fachspezifischen Lernzielen, Erkenntnisweisen und historischen Kategorien wie Zeit und Wandel und weiteren sprachlichen Markern eines reflektierten Geschichtsbewusstseins (Schradler, 2019, 129). Hierzu gehören etwa die für Texte des Geschichtsunterrichts charakteristischen niedrigfrequenten Wörter, Zeitverhältniswörter oder auch die auf die prinzipielle Unsicherheit historischer Erkenntnisse verweisenden epistemisch modalen Wörter (etwa Siegmund, 2018, 133–143; Beese & Benholz, 2014, 113–114).⁸
- Gestaltung von Leseaufgaben, die historische Denkprozesse initiieren, indem sie neben der inhaltlichen Texterschließung eine kritische Auseinandersetzung mit dem Konstruktcharakter von Geschichte und dem Aussagewert verschiedener Textgattungen fördern (z. B. Wickner, 2018, 38–45; Hirsch, 2015; Rox-Helmer, 2010, 184–187) sowie zur Bewertung von textlichen Begründungszusammenhängen und Argumentationen im Sinne von Triftigkeitsprüfungen (Hodel, 2018, 338–342) und historischer Sach- und Werturteilsbildung anregen.

Es geht also bei einer fachspezifischen Nutzung von Lesestrategien um deren Gebrauch als sprachliches Hilfsmittel zur besseren Erreichung fachlicher Lernziele; wozu im Zusammenhang mit Textarbeit etwa die bereits angesprochene Gattungskompetenz (die im Amerikanischen als *historical literacy* bezeichnet wird) und historische Re- und Dekonstruktionsfähigkeiten zählen. Weitere domänenspezifische Leseziele, Textmerkmale, und Anforderungen an kompetente historische Leser*innen sind etwa bei Matthias Hirsch nachzulesen (2015), der Charakteristika und Operationen des Lesens historischer Fachtexte in Abgrenzung zu literarischer Lesekompetenz erörtert. Dabei wird deutlich, wie anspruchsvoll historische Lese- und Sinnbildungsprozesse sind, was aber nicht nur für Quellentexte gilt. Auch Darstellungstexte aus Schulbüchern stellen inhaltlich und sprachlich hohe Anforderungen an die Leser*innen, unter anderem auf-

8 Hinzu kommen übergreifende, allgemeine Fachtextmerkmale, die das Textverständnis erschweren können, wie z. B. fach- und bildungssprachliche Begriffe, Wortkomposita, Nominalisierungen, Funktionsverbgefüge und komplexe Sätze (Roelcke, 2020, 112–121).

grund ihres meist hohen Abstraktionsgrades und weiterer typischer Merkmale wie unbekannte Fachbegriffe, komplexe Verweisstrukturen, mehrteilige Konjunktoren und verwirrende rhetorische Strukturen (die Analyse von Kniffka & Linnemann, 2016, 163–178, zu Schwierigkeitsschwerpunkten fachspezifischer Sachtexte). Gleichzeitig kommt ihnen eine wichtige Rolle bei der kompakten Vermittlung historischen Sachwissens und der Vorbereitung von Quellenarbeit im Geschichtsunterricht zu.

Welche Formen von Leseunterstützung eignen sich also, um das Verständnis von Sachtexten des Geschichtsunterrichts und damit auch den historischen Kompetenzaufbau effektiv zu unterstützen? Die nachfolgend in ihrem forschungsmethodischen Aufbau beschriebene Interventionsstudie geht dieser Frage am Beispiel von Verfasser*innen zur altägyptischen Geschichte nach, die in ein adaptives, mit sprachlichen Hilfen zur Texterschließung versehenes Unterrichtskonzept eingebettet wurden.

4 Studiendesign, Methoden und Erhebungsinstrumente

Zur Untersuchung der Effekte des selbstentwickelten, an Fünftklässler*innen gerichteten sprachsensiblen Unterrichtskonzepts zu Aspekten von Herrschaft und Gesellschaft im Alten Ägypten wird eine (quasi-)experimentelle Interventionsstudie durchgeführt. Zuvor wird das eingesetzte adaptive Lernmaterial in mehreren Erprobungszyklen und vorwiegend in Einzelsettings im realen Unterricht getestet und anschließend überarbeitet. Darüber hinaus beinhaltet die Pilotierungsphase eine Erprobung sprachdiagnostischer Instrumente.

In der quantitativen Hauptstudie wird dann die Wirkung der sprachlich unterstützten Textarbeit im Rahmen von Prä-/Posterhebungen untersucht. Die Lesefähigkeiten der Schüler*innen, genauer deren Fähigkeiten im Bereich Textverstehen, werden dabei als zentrale Heterogenitätsvariable betrachtet⁹ und durch die Zuweisung zu verschiedenen Gruppen differenziell gefördert, um anschlie-

9 Selbstverständlich ist die individuelle Lesefähigkeit nur eine von mehreren Facetten von Heterogenität, die im Zusammenhang mit Binnendifferenzierung, Diversität und Inklusion im Geschichtsunterricht betrachtet werden können und müssen. In Bezug auf fachspezifische Leseprozesse erscheint es jedoch aus sachlichen und forschungsmethodischen Gründen sinnvoll, das Heterogenitätskriterium Lesekompetenz zu fokussieren, ohne damit andere Aspekte von Diversität zu negieren, die historische Lernprozesse beeinflussen und aktuell in der Geschichtsdidaktik diskutiert werden (dazu z. B. Barsch, Degener, Kühberger & Lücke, 2019).

ßend die Effekte der sprachlichen Hilfen auf das als abhängige Variable betrachtete fachspezifische Textverständnis zu untersuchen.

Den Kern der Intervention, welche ohne Prä- und Posttests sechs Unterrichtsstunden umfasst, bildet die strukturierte, strategiebasierte Erschließung von Verfassertexten zur Pharaonenherrschaft. Zu den Begleitmaterialien gehört unter anderem ein ausführliches Wortglossar, mit dessen Hilfe die Lernenden schwierige Fachbegriffe und -konzepte kooperativ erarbeiten. Während die Textgrundlage in allen Fördergruppen identisch ist, wird das Ausmaß der Sprachunterstützung in den einzelnen Gruppen systematisch variiert. Leseschwächere Schüler*innen erhalten bei der Textarbeit auf zwei Wegen sprachliche Unterstützung: zum einen durch Lesestrategien und zum anderen durch grafisch entlastete Texte. Die Strategiearbeit, die auch als solche reflektiert wird, zielt vor allem auf die sprachliche und inhaltliche Aufarbeitung bildungs- und fachsprachlicher Wörter sowie der dahinterstehenden historisch-politischen Fachkonzepte (wie etwa Herrschaft, Monarchie, Rechtsprechung, Gesetze erlassen, Steuern etc.). Neben einer intensiven Beschäftigung mit der fachsprachlichen Lexik leiten weitere Strategieanwendungsaufgaben etwa zur Auseinandersetzung mit fachspezifischen Textgenres sowie zur Strukturierung, eigenständigen Verbalisierung und Visualisierung von Textinhalten an, um schlussendlich die Entwicklung eines historischen Urteils zum Lerngegenstand Pharaonenherrschaft anzubahnen.

Die sogenannte grafische Textentlastung als Unterstützungsstrategie, welche nur die leseschwächeren Lernenden erhalten, beinhaltet zum Beispiel typografische Veränderungen der Schriftgröße und der Textstruktur (Krüger & Hemmer, 2019, 107–120), ohne dass Texteingriffe wie Vereinfachungen oder Kürzungen erfolgen; der *intrinsic load* bleibt also gleich. In Verbindung mit der Strategiearbeit wird dadurch eine Verbesserung des Textverständnisses und eine Erhöhung der Lesemotivation erwartet (Krüger & Hemmer, 2019, 107–120). Neben defensiver Textentlastung, die nur in der Gruppe der leseschwächeren Schüler*innen angewandt wird, erhalten beide Gruppen offensive Förderung durch Lesestrategievermittlung und -anwendung, welche die langfristige Lesekompetenzentwicklung unterstützt. Dabei variiert das Ausmaß der Selbststeuerung und der Strukturierungsgrad der Texte und Aufgaben in den einzelnen Gruppen. Der *extraneous load* für leseschwächere Lernende wird verringert und diese erhalten in Anknüpfung an die einführend dargestellten ATI-Prinzipien insgesamt mehr *prompts* (Hilfen) bei der Textarbeit.

Insgesamt werden drei Untersuchungsgruppen und eine Kontrollgruppe gebildet. Neben Gruppe A, in der die lesestärkeren Lernenden mit offener gehaltenen Lesestrategie-Anwendungsaufgaben an den Texten arbeiten, erhalten die lese-schwächeren Schüler*innen in Gruppe B (1) durch entsprechend enger geführte Aufgabenvarianten deutlich mehr Anleitung und Struktur. In der dritten Gruppe B (2) arbeiten leseschwächere Lernende ebenfalls mit dieser Variante der Strategie-Anwendungsaufgaben sowie zusätzlich mit grafisch entlasteten, sprachlich und inhaltlich aber identischen Texten. Die Gruppen A und B (1) nutzen stattdessen die originalen Schulbuchtexte. Hinzu kommt eine Kontrollgruppe, die keines der Treatments erhält und nur an den Testzeitpunkten teilnimmt.

Die Stichprobe umfasst 120 Schüler*innen der Klassenstufe fünf verschiedener Gemeinschaftsschulen, die auf der Basis der Ergebnisse vorab durchgeführter sprachdiagnostischer Tests den verschiedenen Förderbedingungen zugewiesen werden. Dabei wird ein Teil der Treatments auch unpassend zugeteilt, um die differenziellen Wirkungen der Fördermaßnahme zu erfassen. Die Auswertung erfolgt regressionsanalytisch. Zur Durchführung der Unterrichtseinheit werden die Lehrkräfte intensiv geschult, erhalten standardisierte Materialien sowie Unterrichtsskripte und es erfolgt zusätzlich eine Implementationskontrolle.

Zur Messung der Effekte der Lesefördermaßnahme auf das als abhängige Variable betrachtete Textverständnis wird vor und nach der Intervention ein Leseverstehens- und Fachwissenstest durchgeführt. Dieser prüft die wesentlichen Teile des textlich vermittelten historischen Sachwissens sowie darüber hinaus in kleinerem Umfang auch konzeptuelle und gattungsbezogene Kenntnisse, die durch die Begleitaufgaben zu den Texten vermittelt werden. Der Test erfragt also zum einen «klassisches» historisches Sachwissen zu altägyptischen Herrschaftspraxen, zum andere prüft er auch konzeptuelles Wissen im Sinne des Verfügung über historisch-politische Fachkonzepte wie etwa Herrschaft, Monarchie oder Gemeinwohl. Letzteres kann zu den Kompetenzen gezählt werden, da es sich auf fachspezifische Kategorien und Konzepte bezieht. Im Gegensatz zum themen- oder fallspezifischem Inhaltswissen können diese auf unterschiedliche Sachverhalte und Fragen angewendet und entsprechen damit der Kompetenzorientierung (Körper, 2017). Nach dem hier zugrunde gelegten FUER-Kompetenzmodell wird diesbezüglich von «historischer Sachkompetenz» (Körper, 2017, 12) gesprochen und hinsichtlich der erwähnten testrelevanten konzeptuellen und semantischen Kenntnisse von historischer «Begriffs- und Strukturierungs-kompetenz» (Schreiber, 2008, 206). Laut Körper, der sich mit der seit 2007 ge-

fürten Kompetenzdebatte, auf die hier nur in sehr kurzer Form eingegangen werden kann, und mit Kritik am FUER-Modell auseinandergesetzt hat, ist es nicht sinnvoll, historische Sachkompetenz und «klassisches» Sachwissen gegeneinander auszuspielen. Stattdessen besäße beide Wissenstypen unterschiedliche Funktionen im historischen Erkenntnisprozess und seien daher als einander ergänzend anzusehen (2017, 15–16).

Im Hinblick auf die Wirksamkeitsprüfung der hier vorgestellten Intervention bestand die Herausforderung darin, einen Test zu konstruieren, der es erlaubt das Textverständnis, den Aufbau von themenspezifischem historischem Sachwissen und historische konzeptuelle Kompetenz beziehungsweise Begriffskompetenz zu messen. Da aufgrund unterschiedlicher Inhalte und Adressaten kein direkter Rückgriff auf bereits validierte Items wie die des HiTCH-Tests möglich ist, (Trautwein et al., 2017), musste dieser Test selbst entwickelt werden, wobei jedoch Anleihen bei den Konstrukten und Aufgabenformaten bereits etablierter Instrumente genommen wurden. So wurden etwa die Items zur Prüfung des Textverständnisses eng an die Struktur und die Aufgabentypen des IGLU-Fragebogens zum Verständnis von Sachtexten angelehnt (Bos et al., 2005, 425–426), inhaltlich aber auf die spezifischen Inhalte der Intervention zugeschnitten. Eine Mischung aus offenen und geschlossenen Fragen, die drei Niveaus von Textverstehen repräsentieren, soll einerseits für die nötige Tiefe der Antworten sorgen (offene Frageformate) und andererseits den Einfluss sprachlicher Fähigkeiten, insbesondere der Schreibfähigkeit, auf das Testergebnis verringern (geschlossene Frageformate wie Mehrfachwahllantworten, Reihenfolgeaufgaben).¹⁰ Nachfolgend einige veranschaulichende Beispiele von Items zur Erfassung des Textverständnisses.

Tabelle 1: Beispiele von Items zur Erfassung des Textverständnisses

Reproduktion von Informationen	Geschlossene Frage (Mehrfachauswahl)	Warum galt der Pharao als gottähnlich?
Einfache Schlussfolgerungen	Offene Frage	Erkläre: Wie verhinderte der Pharao, dass es zu Hungersnöten im Volk kommt?
Komplexe Schlussfolgerungen	Offene Frage	Denke nach und erkläre: Warum erfahren wir aus dem Text nur so wenig über Pharaoninnen, also über Frauen, die herrschten?

¹⁰ Als qualitative Ergänzung der standardisierten Daten wird außerdem ein Leseprodukt der Schüler*innen ausgewertet, welches während der Intervention entsteht.

Um die Itemschwierigkeit und -verständlichkeit zu prüfen, wurde der Fragebogen bereits einmal in einer größeren Stichprobe im Feld getestet ($n = 75$), die Auswertung wird derzeit vorgenommen. Zusätzlich wurde das Verständnis der Fragen im Rahmen von kognitiven Interviews mit vier Schülern und Schülerinnen überprüft («Probing», dazu Pohontsch & Meyer, 2015, 53). Zur Erhöhung der Qualität und Validität wurde außerdem ein Expertenreview durchgeführt, bei dem unter anderem Einschätzungen dazu erbeten wurden, wie gut die relevanten Konstrukte durch die Items abgebildet werden, welche Anzahl richtiger Antworten erwartet wird und welche Relevanz die Fragen für die historisch-politische Bildung in Klasse fünf besitzen.

Weitere Testinstrumente, mit denen vor der Intervention die Fähigkeiten der Lernenden im Bereich des Textverstehens diagnostiziert werden sollen, auf deren Basis die Zuteilung zu den einzelnen Treatments erfolgt, liegen bereits vor. Hierzu zählt der Lernstand-5-Test, der Lesegeschwindigkeit und -verständnis misst und jeweils zu Schuljahresbeginn in allen fünften Klassen Baden-Württembergs durchgeführt wird (IBBW Baden-Württemberg, 2019). Ergänzend wird ein C-Test durchgeführt, der als zeitökonomisches und reliables Instrument zur Erfassung allgemeiner Sprachfähigkeiten gilt, welche wiederum stark mit Leseverstehen korrelieren (Grotjahn, 2019, 586–609; Linnemann, 2017, 149–168). In Entsprechung zu diesen Befunden aus der Forschungsliteratur hat der C-Test in einem Vortest ($n = 23$) eine hohe Übereinstimmung mit den Ergebnissen des Lernstand-5-Tests gezeigt.

Überdies zeigte sich eine breite Streuung der Sprach- und Lesefähigkeiten innerhalb der untersuchten Lerngruppe. Die Bandbreite reicht dabei von sehr guten Lesern und Leserinnen, die ganze Textabschnitte flüssig lesen und mühelos verstehen, bis hin zu sehr leseschwachen Lernenden, die sich einen Text stockend, Wort für Wort zu erschließen versuchen, wie sich in parallel geführten Einzelinterviews zeigte. Nachstehend ein kurzer Auszug aus einem solchen Interview,¹¹ in dem sich zwei Fünftklässler über die Bedeutung des Fachworts «Handelsexpedition» unterhalten.

11 Auszug aus einem während der Pilotierungsphase der nachfolgend vorgestellten Studie geführten Interview mit zwei Schülern, die im Team einen Sachtext aus dem Geschichtsschulbuch bearbeiteten. Insgesamt wurden vier Interviews mit jeweils zwei Schüler*innen geführt ($n = 8$), um das Text- und Aufgabenverständnis zu überprüfen und das in der Intervention eingesetzte Lesefördermaterial daraufhin zu optimieren.

S1: «Das hier ist das letzte (...) Handelsexpedition.» (#00:32:24-4#)

B1: «Was ist das?» (#00:32:28-4#)

S2: «Ich schreib' noch.» (#00:32:28-4#)

S1: «Ähm, ist die Ableitung, dass man in fremden Ländern Tempel und Paläste bauen kann.» (#00:32:58-8#)

S2: «(Liest vor) (...) schützte das Land auch gegen Feinde, führte Kriege, schickte Handelsexpeditionen in fremde Länder (...) ja.» (#00:33:15-4#)

S1: «Ja.» (#00:33:15-4#)

S2: «Das sind, glaube ich, damit ist gemeint, dass halt der (...) die suchen halt Schätze und geben das dem Pharao.» (#00:33:36-2#)

S1: «Da steht ja, dass sie in fremden Ländern Tempel, Paläste und Pyramiden bauen.» (#00:33:44-1#)

I: «Wie verstehst du **HANDELSEXPEDITIONEN**?» (#00:33:46-2#)

S2: «Also, dass halt (...) eine Expedition ist, glaub ich, sowas, dass sie was suchen, Schätze zum Beispiel, und das dann verkaufen, halt handeln.» (#00:34:04-8#)

S1: «Dass sie in fremden Ländern, Tempel, Paläste und Pyramiden bauen dürfen.» (#00:34:22-5#)

S2: «Der Pharao schickt Leute, die Schätze suchen und damit handeln.» (#00:35:56-4#)

Während einer der Schüler (S1) bis zum Schluss der Gesprächssequenz auf seiner Fehldeutung des Begriffs «Handelsexpedition» beharrt, die vermutlich auf einem falschen Analogieschluss beruht (er interpretiert eine Aufzählung in einem komplexen Satz fälschlicherweise als Definition), nähert sich sein Partner (S2) der Wortbedeutung sukzessive an, wenngleich diese bis zum Schluss noch verfeinerungsbedürftig bleibt. Es gelingt S2 jedoch, sich der Bedeutung des schwierigen Fachworts und Nominalkompositums «Handelsexpedition» ungefähr anzunähern, indem er die Strategie des «morphologischen Dekodierens» (Fuhrhop & Olthoff, 2019, 53) anwendet, das Wort also zunächst in seine einzelnen bedeutungstragenden Einheiten zerlegt (Fuhrhop & Olthoff, 2019, 53).

Während starke Leser und Leserinnen diese und andere Erschließungsstrategien erfolgreich anwenden, übergehen schwächere Leser und Leserinnen schwierige Begriffe häufig unverstanden, weil ihnen passende Rezeptionsstrategien fehlen. Für diese Gruppe von Lernenden sind adäquate sprachliche Unterstützungsangebote bei der Textarbeit im Geschichtsunterricht demnach besonders wichtig, weil es entscheidend vom Verlauf des Leseprozesses abhängt, ob

anschließende hierarchiehöhere Verstehens- und Sinnbildungsprozesse gelingen. Die Notwendigkeit, im Fachunterricht auf unterschiedliche Sprachvoraussetzungen von Lernenden zu reagieren, denen ansonsten fachliche Inhalte möglicherweise verschlossen bleiben, zeigen auch die Ergebnisse des in der Pilotklasse durchgeführten C-Tests.¹²

Dessen Auswertung ergab, dass rund ein Drittel der 23 beteiligten Fünftklässler*innen einer Gemeinschaftsschule einen deutlichen sprachlichen Förderbedarf sowie besonderen Unterstützungsbedarf im Bereich der Texterschließung besitzt. Die Fähigkeiten der Testteilnehmer*innen waren sowohl im Bereich Leseverstehen sowie auch bei den formalsprachlichen Fähigkeiten im Vergleich zur Referenzgruppe gering ausgeprägt.

Neben unterschiedlichen Fähigkeiten in den Bereichen Leseflüssigkeit und Textverstehen wirken, so ist anzunehmen, noch weitere Faktoren auf den Leseprozess und das historische Lernen mit Texten ein. Sie müssen als mögliche Hintergrundvariablen mitbedacht werden. Hier ist zum einen an individuelle Lernvoraussetzungen beim Lesen zu denken, wie etwa das Leseselbstkonzept, die Lesegewohnheiten bei Sachtexten sowie das lesestrategische Wissen, welche daher mithilfe von IGLU- und PISA-Items miterhoben werden. Zum anderen werden die individuellen Voraussetzungen beim historischen Lernen, wie das Geschichtsinteresse und die fachliche Motivation und Leistung, als mögliche Moderatorvariablen miterfasst.

12 Ein C-Test beruht auf mehreren Lückentexten, welche in einer bestimmten Zeit rekonstruiert werden müssen. Er gilt als äußerst valides und sehr zeitökonomisches Instrument zur Messung von allgemeiner Sprach- und Lesekompetenz und kann auch im binnendifferenzierenden Geschichtsunterricht als Grundlage zur Einteilung flexibler Fördergruppen genutzt werden (dazu etwa Friesen, Leuders & Loibl, 2021, 42–45).

5 Fazit

Bei aller theoretischen und normativen Plausibilität, die Konzepte fachintegrierter Sprachbildung und -förderung angesichts aktueller gesellschaftlicher Entwicklungen wie Migration und Inklusion besitzen, welche den Geschichtsunterricht als Sprachfach besonders betreffen, bedarf es zusätzlich auch empirischer Wirksamkeitsnachweise. Zumal nicht nur von praktizierenden Lehrkräften die Frage gestellt wird, inwiefern es sinnvoll ist, einem ohnehin marginalisierten Unterrichtsfach wie Geschichte mit der Aufgabe der Sprachförderung noch eine zusätzliche Bürde aufzuladen. Gibt es denn, so könnte man fragen, überhaupt einen realen fachlichen Mehrwert fachintegrierter Lesefördermaßnahmen, der die unbestreitbar nötigen zusätzlichen zeitlichen Ressourcen und Bemühungen seitens der Lehrkraft aufwiegt? Denn dass für die Konzeption fachlich sinnvoller sprachlicher Unterstützungssysteme zusätzliche Unterrichts- und Vorbereitungszeit benötigt wird, wurde auch bei der Entwicklung des hier vorgestellten, fachspezifischen Leseförderkonzepts deutlich.

Expert*innen weisen ebenfalls darauf hin, dass es für einen sprachbildenden und -fördernden Fachunterricht entsprechend aus- oder weitergebildeter (Fach-)Lehrkräfte bedarf (Kniffka & Roelcke, 2016, 116). Dies lässt sich wohl am ehesten dann rechtfertigen, wenn sich auch empirisch signifikante Vorteile fachintegrierter Sprachfördermaßnahmen, im Sinne eines «besseren» und nachhaltigeren historischen Lernens nachweisen lassen. Die vorgestellte Studie will hierzu einen Beitrag leisten, die zentrale quantitative Hauptstudie zur Wirksamkeit der sprachlich unterstützen Textarbeit steht allerdings noch aus. Abgeschlossen sind dagegen die Erprobung der sprachdiagnostischen Instrumente sowie die Erprobung des Fördermaterials in der Praxis. Aufgrund der Praxisanbindung, der mehrmaligen Prüfung der entwickelten Lernumgebung im Anwendungskontext und des «zyklischen, iterativen Vorgehens der Untersuchung» (Waldis, 2019, 15) erfüllt das Projekt wesentliche Kriterien, die von Waldis als Kennzeichen von Design-based-Research genannt werden (ebd.). Hierzu zählt auch die enge Zusammenarbeit mit Expert*innen aus der Schulpraxis, welche für HeLPS-Projekte¹³ charakteristisch ist.

13 Die Abkürzung «HeLPS» steht für «Heterogenität: effektive Lernsettings und Professionalität an Schulen». HeLPS ist ein interdisziplinärer Forschungsverbund der Pädagogischen Hochschule Freiburg, in dem in sieben Teilprojekten die Lernwirksamkeit binnendifferenzierender Unterrichtsssettings aus der Perspektive verschiedener Fächer und mit vorwiegend quantitativen Methoden beforscht wird.

Im Hinblick auf die für diese Studie zentrale Frage nach den Effekten fachintegrierter Leseförderung auf das historische Lernen hat die Pilotierungsphase sowohl erste Hinweise auf verbessertes Textverstehen und zunehmendes historisches Sachwissen als auch auf die ebenfalls interessierenden ATI-Effekte ergeben, wie etwa eine abnehmende Lese- und Lernmotivation infolge von «unpassend» zugeteilten Fördermaterialien. Nach einer Erprobung des adaptiven Lernsettings im Klassenunterricht schätzten zudem 85 Prozent der Befragten ($n = 23$) die Lesestrategie-Anwendungsaufgaben als lernförderlich ein, was zunächst erfreulich erscheint. Da jedoch unklar ist, inwiefern sich hier weitere Faktoren der speziellen Klassensituation auf die subjektiven Einschätzungen auswirkten, wie etwa Effekte sozialer Erwünschtheit oder auch die auffallend gute Beziehung zwischen der verantwortlichen Lehrkraft und der Lerngruppe, können belastbare Ergebnisse wohl erst aus einer größeren Stichprobe und unter Kontrolle weiterer Einflussfaktoren abgeleitet werden. Es bleibt also abzuwarten, ob sich die erwarteten positiven Effekte geschichtsdidaktisch profilierter Lesehilfen auch in größerer Zahl und als «objektive», anhand des fachlichen Lernzuwachses erhobene *hard facts* in der quantitativen Hauptstudie zeigen.

Literatur

- Achour, Sabine & Sieberkrob, Matthias (2015). Sprachbildung im Politikunterricht und Sprachbildung konkret: Heterogenität. *Wochenschau*, 15s, 18–37.
- Adamski, Peter (2017). Aufgaben, Materialien, Lernhilfen. Der Mikroprozess binnendifferenzierten Lernens. *Geschichte lernen*, (178), 2–11.
- Ahrenholz, Bernt/Jeuk, Stefan/Lütke, Beate/Paetsch, Jennifer & Roll, Heike (2019). Sprache im fachlichen Unterricht. In Bernt Ahrenholz/Stefan Jeuk/Beate Lütke/Jennifer Paetsch & Heike Roll (Hrsg.), *Fachunterricht, Sprachbildung und Sprachkompetenzen* (S. 1–16). Berlin, Boston: De Gruyter.
- Artelt, Cordula/Schiefele, Ulrich/Schneider, Wolfgang & Stanat, Petra (2002). Leseleistungen deutscher Schülerinnen und Schüler im internationalen Vergleich (PISA): Ergebnisse und Erklärungsansätze. *Zeitschrift für Erziehungswissenschaft*, 9(1), 6–27.
- Barsch, Sebastian/Degener, Bettina/Kühberger, Christoph & Lücke, Martin (2019). *Handbuch Diversität im Geschichtsunterricht. Inklusive Geschichtsdidaktik*. Schwabach/Ts.: Wochenschau.
- Becker-Mrotzek, Michael/Lindauer, Thomas/Pfost, Maximilian/Weis, Mirjam/Strohmaier, Anselm & Reiss, Kristina (2019). Lesekompetenz heute – eine Schlüsselqualifikation im Wandel. In Kristina Reiss/Mirjam Weis/Eckhard Klieme & Olaf Köller (Hrsg.), *PISA 2018: Grundbildung im internationalen Vergleich* (S. 21–46). Münster, New York: Waxmann.
- Beese, Melanie & Benholz, Claudia (2013). Sprachförderung im Fachunterricht: Voraussetzungen, Konzepte und empirische Befunde. In Charlotte Röhner & Britta Hövelbrinks

- (Hrsg.), *Fachbezogene Sprachförderung in Deutsch als Zweitsprache: Theoretische Konzepte und empirische Befunde zum Erwerb bildungssprachlicher Kompetenzen* (S. 37–56). Weinheim, Basel: Beltz.
- Bernhardt, Markus & Conrad, Franziska (2018). Sprachsensibler Geschichtsunterricht: Sprachliche Bildung als Aufgabe des Fachs Geschichte. *Geschichte Lernen*, 31(182), 2–9.
- Bönnighausen, Marion (2019). *Praxisprojekte in Kooperationsschulen. Fachdidaktische Modellierung von Lehrkonzepten zur Förderung strategiebasierten Textverstehens in den Fächern Deutsch, Geographie, Geschichte und Mathematik*. Münster: WTM.
- Bönnighausen, Marion & Winter, Katja (2012). *Lesend lernen! Texte besser verstehen: Ein Trainingsprogramm*. Bottrop: Henselowsky Boschmann.
- Bos, Wilfried et al. (2005). *IGLU: Skalenhandbuch zur Dokumentation der Erhebungsinstrumente*. Münster, New York: Waxmann.
- Buck, Thomas Martin (2019). Geschichte und Sprache: Vom Sach- zum Sprachfach. *Geschichte für heute*, 1(12), 5–12.
- Cronbach, Lee Joseph (1975). Wie kann Unterricht an individuelle Unterschiede angepasst werden? In Ralf Schwarzer & Klaus Steinhausen (Hrsg.), *Adaptiver Unterricht: zur Wechselwirkung von Schülermerkmalen und Unterrichtsmethoden* (S. 42–58). München: Kösel.
- Degener, Bettina & Kreutz, Jessica (2020). Darstellungstexte in differenzierenden Geschichtsschulbüchern im Spannungsverhältnis von Textverständlichkeit und Ansprüchen an historisches Lernen. In Thomas Sandkühler & Markus Bernhardt (Hrsg.), *Sprache(n) des Geschichtsunterrichts: Sprachliche Vielfalt und historisches Lernen* (S. 233–252). Göttingen: Vandenhoeck & Ruprecht.
- Deunk, Marjolein/Doolaard, Simone/Smale-Jacobse, Annemieke & Bosker, Roel (2015). *Differentiation within and across classrooms: A systematic review of studies into the cognitive effects of differentiation practices*. Groningen: Rijksuniversiteit.
- Dietrich, Tobias (2009). Sachtexte lesen im Fach Geschichte. In Staatliches Studienseminar für das Lehramt an Gymnasien (Hrsg.), *Sachtexte lesen im Fachunterricht der Sekundarstufe* (S. 178–188). Seelze: Kallmeyer.
- Dzubieli, Christine & Harcom, Stephanie (2018). *Zeitreise 1. Arbeitsheft Sprachförderung*. Stuttgart, Leipzig: Klett.
- Eckerth, Christiane & Resch, Mario (2016). Sprachsensibler Geschichtsunterricht: Von der geschichtsdidaktischen Theorie über die Empirie zur Unterrichtspraxis. Tagungsbericht. <https://www.hsozkult.de/conferencereport/id/tagungsberichte-6810> (21.04.2022).
- Fuhrhop, Nanna & Olthoff, Sarah (2019). Komposita als Herausforderung in Schulbuchtexten. In Martin Butler & Juliana Goschler (Hrsg.), *Sprachsensibler Fachunterricht. Chancen und Herausforderungen aus interdisziplinärer Perspektive* (S. 35–53), Wiesbaden: Springer.
- Friesen, Marita/Leuders, Timo & Loibl, Katharina (2021). Flexibles Gruppieren nach lernzielrelevanten Voraussetzungen. *Pädagogik*, (3), 42–45.
- Geel, Marieke van/Keuning, Trynke/Frèrejean, Jimmy/Dolmans, Diana/van Merriënboer, Jeroen & Visscher, Adrie (2018). Capturing the complexity of differentiated instruction. *School effectiveness and school improvement*, 30(1), 51–67.
- Gibbons, Pauline (2009). *English learners, academic literacy, and thinking*. Portsmouth: Heinemann.

- Grotjahn, Rüdiger (2019). C-Tests. In Stefan Jeuk & Julia Settinieri (Hrsg.), *Sprachdiagnostik Deutsch als Fremdsprache: Ein Handbuch* (S. 586–609). Berlin, Boston: De Gruyter.
- Handro, Saskia (2016). «Sprachsensibler Geschichtsunterricht»: Systematisierende Überlegungen zu einer überfälligen Debatte. In Wolfgang Hasberg & Holger Thünemann (Hrsg.), *Geschichtsdidaktik in der Diskussion: Grundlagen und Perspektiven* (S. 265–296). Frankfurt a. M.: Peter Lang.
- Handro, Saskia & Kiliman, Vanessa (2019). Textverstehen im Geschichtsunterricht. Ein Projekt zur Professionalisierung historischer Leseförderung (ProLeGu). In Marion Bönnighausen (Hrsg.), *Praxisprojekte in Kooperationsschulen. Fachdidaktische Modellierung von Lehrkonzepten zur Förderung strategiebasierten Textverstehens in den Fächern Deutsch, Geographie, Geschichte und Mathematik* (S. 162–222). Münster: WTM.
- Hirsch, Matthias (2015). Geschichte (er-)lesen. Überlegungen zu domänenspezifischen Lesemodi und -prozessen. *Zeitschrift für Geschichtsdidaktik*, 14, 136–153.
- Institut für Bildungsanalysen Baden-Württemberg (2019). *Lernstand 5 – Verfahren: Lernstandsanalyse*. https://ibbw.kultusbw.de/Lde/Startseite/Kompetenzmessung/Verfahren+_Lernstands-analyse+_Ebene+_ (21.04.2022).
- Kalyuga, Slava (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31.
- Kalyuga, Slava & Renkl, Alexander (2010). Expertise reversal effect and its instructional implications: introduction to the special issue. *Instructional Science*, 38(3), 209–215.
- Kareva, Veronika & Echevarria, Jana (2013). Using the SIOP model for effektive content teaching with second and foreign language learners. *Journal of Education and Training Studies*, 2, 239–248.
- Kniffka, Gabriele & Linnemann, Markus (2016). Zur Ermittlung der Schwierigkeit (fach-)sprachlicher Strukturen mittels fachlicher C-Tests: Eine explorative Studie. In Erwin Tschirner/Olaf Bärenfänger & Jupp Möhring (Hrsg.), *Deutsch als fremde Bildungssprachliche. Das Spannungsfeld von Fachwissen, sprachlicher Kompetenz, Diagnostik und Didaktik* (S. 163–178). Tübingen: Stauffenberg.
- Kniffka, Gabriele & Roelcke, Thorsten (2016). *Fachsprachenvermittlung im Unterricht*. Paderborn: Schöningh.
- Körper, Andreas (2017). *Kompetenzen historischen Denkens revisited. Das FUER-Modell historischer Kompetenzen vor dem Hintergrund von zehn Jahren Kompetenzdebatte in der Geschichtsdidaktik*. https://www.gdt-eichstaett.de/wp-content/uploads/2017/12/Körper_Keynote_2017.12.05..pdf (07.04.2022).
- Körper, Andreas (2017). Ärgerliches aus der Kompetenzdebatte – mit einem Exkurs zur Didaktik von Planspielen. <https://historischdenkenlernen.blogs.uni-hamburg.de/aergerliches-von-der-kompetenzdiskussion-mit-einem-exkurs-zur-didaktik-von-planspielen/>.
- Krüger, Sebastian & Hemmer, Michael (2019). Karten lesen – (k)eine Kunst. Theoretische Grundlagen und Konzeption des Praxisprojekts Geographie. In Marion Bönnighausen (Hrsg.), *Praxisprojekte in Kooperationsschulen. Fachdidaktische Modellierung von Lehrkonzepten zur Förderung strategiebasierten Textverstehens in den Fächern Deutsch, Geographie, Geschichte und Mathematik* (S. 107–165). Münster: WTM.
- Kühberger, Christoph & Windischbauer, Elfriede (2009). Literacy als Auftrag? Zur Förderung bei Leseschwächen im Geschichtsunterricht. *Geschichte lernen*, (131), 22–29.

- Kühberger, Christoph & Windischbauer, Elfriede (2013). *Individualisierung und Differenzierung im Geschichtsunterricht. Offenes Lernen in Theorie und Praxis*. Schwalbach/Ts.: Woenschau.
- Linnemann, Markus (2017). Sprachdiagnose im Unterricht: Grundlagen, Zugänge und Beispiele. In Hartmut Günther/Gabriele Kniffka/Gabriele Knoop & Thomas Riecke-Baulicke (Hrsg.), *Basiswissen Lehrerbildung: DaZ unterrichten* (S. 149–168). Seelze: Klett-Kallmeyer.
- Mainzer-Murrenhoff, Mirka & Mierwald, Marcel (2018). Wenn Historikerinnen und Historiker argumentieren... – Historisches Lesen von geschichtswissenschaftlichen Darstellungstexten in der gymnasialen Oberstufe. *Geschichte Lernen*, 31, 46–55.
- McCulley, Lisa & Osman, David (2015). Effekts of reading instruction on learning outcomes in social studies: A synthesis of quantitative research. *Journal of Social Studies Research*, 39, 183–195.
- Nartschik, Silvia (2012). *Entdecken und Verstehen. Lesetraining Geschichte 3: Vom Zeitalter des Absolutismus bis zum Ersten Weltkrieg*. Berlin: Cornelsen.
- Oleschko, Sven (2015). Herausforderungen einer domänenspezifischen Sprachdiagnostik im Kontext historischen Lernens. *Zeitschrift für Geschichtsdidaktik*, 14, 47–62.
- Oleschko, Sven & Moraitis, Anastasia (2012). Die Sprache im Schulbuch. Erste Überlegungen zur Entwicklung von Geschichts- und Politikschulbüchern unter Berücksichtigung sprachlicher Besonderheiten. *Bildungsforschung*, 9(1), 11–46.
- Philipp, Maik (2012). Elemente einer wirksamen Lese- und Schreibförderung. In Maik Philipp & Anita Schilcher (Hrsg.), *Selbstreguliertes Lesen: ein Überblick über wirksame Leseförderansätze* (S. 59–73). Seelze: Klett-Kallmeyer.
- Philipp, Maik (2015). *Lesestrategien. Bedeutung, Formen und Vermittlung*. Weinheim, Basel: Beltz.
- Pohontsch, Nadine & Meyer, Thorsten (2015). Das kognitive Interview. Ein Instrument zur Entwicklung und Validierung von Erhebungsinstrumenten. *Rehabilitation*, 54(1), 53–59.
- Prediger, Susanne/Wilhelm, Nadine/Büchter, Andreas/Gürsoy, Erkan & Benholz, Claudia (2015). Sprachkompetenz und Mathematikleistung – empirische Untersuchung sprachlich bedingter Hürden in den Zentralen Prüfungen 10. *Journal für Mathematik-Didaktik*, 36(1), 77–104. <https://doi.org/10.1007/s13138-015-0074-0>.
- Reisman, Abby & McGrew, Sarah (2018). Reading in history education: Text, sources and evidence. In Scott Alan Metzger & Lauren McArthur Harris (Hrsg.), *The Wiley international handbook of history teaching and learning* (S. 529–550). New York: Wiley-Blackwell.
- Roelcke, Thorsten (2020). *Fachsprachen*. Berlin: Erich Schmidt.
- Rosebrock, Cornelia (2012). Was ist Lesekompetenz, und wie kann sie gefördert werden? *leseforum.ch*, (3), 1–12. https://www leseforum.ch/myUploadData/files/2012_3_Rosebrock.pdf (21.04.2022).
- Rox-Helmer, Monika (2010). Lesen im Geschichtsunterricht: Notwendigkeit oder Chance? In *Pro Lesen. Auf dem Weg zur Leseschule – Leseförderung in den gesellschaftswissenschaftlichen Fächern* (S. 183–199). Donauwörth: Auer.
- Sandkühler, Thomas & Bernhardt, Markus (Hrsg.) (2020). *Sprache(n) des Geschichtsunterrichts: Sprachliche Vielfalt und historisches Lernen*. Göttingen: Vandenhoeck & Ruprecht.

- Sauer, Michael (2018a). Wie verwenden Geschichtslehrkräfte Schulbücher? Ergebnisse einer Lehrerbefragung. *Geschichte in Wissenschaft und Unterricht*, (7/8), 406–417.
- Sauer, Michael (2018b). *Textquellen im Geschichtsunterricht: Konzepte, Gattungen, Methoden*. Stuttgart: Klett.
- Sauer, Michael (2019). *Begriffslernen und Begriffsarbeit im Geschichtsunterricht*. Schwalbach/Ts.: Wochenschau.
- Schneider, Hansjakob/Becker-Mrotzek, Michael/Sturm, Afra/Jambor-Fahlen, Simone/Neugebauer, Uwe/Efing, Christian & Kernen, Nora (2013). *Expertise: Wirksamkeit von Sprachförderung*. Zürich: Bildungsdirektion des Kantons Zürich.
- Schrader, Viola (2019). Zur Integration historischen und sprachlichen Lernens im Geschichtsunterricht. Theoretische Leitgedanken und praktische Umsetzung. In Yauheniya Danilovch & Galina Putjata (Hrsg.), *Sprachliche Vielfalt im Unterricht: Fachdidaktische Perspektiven auf Lehre und Forschung im DaZ-Modul* (S. 123–139). Wiesbaden: Springer VS.
- Schreiber, Waltraud (2008). Ein Kompetenz-Strukturmodell historischen Denkens. *Zeitschrift für Pädagogik*, 54, 198–212.
- Short, Deborah/Vogt, Mary-Ellen & Echevarria, Jana (2011). *The SIOP model for teaching history-social studies to english learners*. Boston: Pearson.
- Sieberkrob, Matthias & Lücke, Martin (2017). Narrativität und sprachlich bildender Geschichtsunterricht – Wege zum generischen Geschichtslernen. In Brigitte Jostes/Daniela Caspari & Beate Lütke (Hrsg.), *Sprachen – Bilden – Chancen. Sprachbildung in Didaktik und Lehrkräftebildung* (S. 217–229). Münster, New York: Waxmann.
- Siegmund, Benjamin (2018). Womöglich/vermutlich könnte/muss es so gewesen sein. Epistemische Modalität im Geschichtsunterricht. In Kristina Peuschel & Anne Burkard (Hrsg.), *Sprachliche Bildung und Deutsch als Zweitsprache in den geistes- und gesellschaftswissenschaftlichen Fächern* (S. 133–143). Tübingen: Gunter Narr.
- Snow, Richard (1989). Aptitude-treatment interaction as a framework of research in individual differences in learning. In Phillip L. Ackerman, Robert J. Sternberg, & Robert Glaser (Hrsg.), *Learning and individual differences* (S. 13–59). New York: Freeman.
- Souvignier, Elmar & Rühl, Katja (2006). *Wir werden Lesedetektive*. Göttingen: Vandenhoeck & Ruprecht.
- Trautwein, Ulrich et al. (2017). *Kompetenzen historischen Denkens erfassen. Konzeption, Operationalisierung und Befunde des Projekts «Historical Thinking – Competencies in History» (HiTCH)*. Münster, New York: Waxmann.
- Waldis, Monika/Nitsche, Martin & Gollin, Kristine (2020). «Schülerinnen und Schüler schreiben Geschichte» – Eine Interventionsstudie an Deutschschweizer Gymnasien. *Zeitschrift für Geschichtsdidaktik*, 19, 90–108.
- Waldis, Monika & Ziegler, Béatrice (2019). Einleitung. In Monika Waldis & Béatrice Ziegler (Hrsg.), *Forschungswerkstatt Geschichtsdidaktik 17. Beiträge zur Tagung «geschichtsdidaktik empirisch 17»* (S. 15–16). Bern: hep.
- Wickner, Mareike-Cathrine (2018). So schließt sich der Kreis. Textsortenspezifische Schreibförderung im Geschichtsunterricht mit dem «Genre Cycle». *Geschichte lernen*, 31(182), 38–45.

Wittwer, Jörg & Renkel, Alexander (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist*, 43(1), 49–64. <https://doi.org/10.1080/00461520701756420>.

Welche Hilfen könnten helfen?

Einblick in eine Interventionsstudie zur Förderung des historischen Erklärens¹

Christoph Bramann, Carina Wöhlke und Nicola Brauch

1 Problemstellung und Ziel des Beitrags

Historisches Erklären ist – wie fachspezifisches Erklären im Allgemeinen – eine komplexe Sprachhandlung, die Schüler*innen vor vielfältige Herausforderungen stellt (Chapman, 2017; Lee & Shemilt, 2009; Mierwald, 2020; Mierwald & Brauch, 2019; Stoel, van Drie & van Boxtel, 2017; Stoel, 2017). Ausgehend von einer explorativen quasiexperimentellen Studie in einem Schülerlabor-Setting wird im Beitrag auf ein bereits an anderer Stelle (Brauch, Heine & Bramann, 2020) vorgestelltes Modell des historischen Erklärens (*Causal Reasoning in History* – CRH) Bezug genommen, das für den vorliegenden Zugang in sieben aus der Geschichtstheorie abgeleiteten epistemologischen Merkmalen (z. B. Perspektivität, Kausalität, Temporalität etc.) operationalisiert wird. Diese Merkmale des historischen Erklärens dienten als Basis zur Entwicklung von drei Scaffoldingtools (fachlich – sprachlich – integrativ fachlich und sprachlich) mit jeweils sieben Scaffolds (= «Hilfen») für jedes Merkmal (Gibbons, 2002). Ziel des Beitrags ist es, die Nutzung der drei Scaffoldingtools zur Überarbeitung selbst erstellter Drehbuchtexte für Erklärvideos durch Schüler*innen differenziert darzustellen.

Dabei kommen sowohl Methoden der qualitativen Inhaltsanalyse als auch Methoden der deskriptiven und korrelativen Statistik zur Datenanalyse zum Einsatz. Die Analysen ermöglichen damit eine integrative Sicht auf die drei Interventionsgruppen mit den unterschiedlichen Scaffoldingtools. Hierfür werden

¹ Die hier vorgestellte Studie wurde im Rahmen der interdisziplinären Projektgruppe «Sprache im Fach» an der Ruhr-Universität Bochum durchgeführt (Geschichtsdidaktik: Nicola Brauch, Christoph Bramann, Kira Busse; Physikdidaktik: Heiko Krabbe, Carina Wöhlke; Sprachbildungsforschung: Lena Heine, Lisa Otto).

die Daten auf Gruppenebene analysiert und zwar zum einen hinsichtlich der Nutzung der sieben Scaffolds als Hilfestellung zur Berücksichtigung der sieben Merkmale des historischen Erklärens in den Drehbuchtexten (= Forschungsfrage 1), und zum anderen hinsichtlich der angewendeten Methoden der Überarbeitung (= Forschungsfrage 2). Da es zu dieser Thematik bislang kaum vergleichbare Forschung gibt, bewegt sich die Studie im explorativen Bereich und die Ergebnisse erlauben erste Einblicke in die Entwicklung empirischer Scaffold-basierter Unterrichtsforschung. Dazu werden im nächsten Kapitel zunächst Einblicke in den theoretischen Hintergrund der Studie gegeben und daraus die Forschungsfragen abgeleitet (Kapitel 2), bevor anschließend der Aufbau der Interventionsstudie näher beschrieben (Kapitel 3) und die Ergebnisse der Studie vorgestellt (Kapitel 4) werden. Der Beitrag schließt mit der Interpretation der Befunde, einschließlich einer Reflexion von Limitationen der Studie (Kapitel 5).

2 Theoretischer Hintergrund und Forschungsfragen

2.1 Historisches Erklären erforschen

Die hier vorgestellte Studie wurde als Vorstudie im Rahmen einer interdisziplinären Projektgruppe «Sprache im Fach» (Geschichtsdidaktik, Physikdidaktik, Sprachbildungsforschung) durchgeführt, in der die verschiedenen Zugänge zum jeweiligen facheigenen «Erklären» in ihren theoretischen und konzeptionellen Grundlagen geschärft und hinsichtlich möglicher Umsetzungsstrategien für die Entwicklung fachlicher Lernhilfen adaptiert werden sollen.² Dabei wird davon ausgegangen, dass die epistemologischen Prinzipien eines Faches je fachspezifische Sprachmuster bedingen, die für eine fachlich valide Erklärung vonnöten sind (Vollmer & Thürmann, 2010, 116–118). So lautet ein möglicher Ansatz für Erklären im Fach Geschichte: «Etwas geschichtlich zu erklären heißt, ein historisches Phänomen auf seine Ursachen und Bedingungen zurückzuführen.» (Welskopp, 2002/2019, 81). Historisches Erklären beinhaltet demnach die Identifizierung kausaler Beziehungen zwischen Ursachen und Folgen (ebd.). Darüber hinaus stellt sich historisches Erklären im Licht der narrativistischen Ge-

2 Eine theoretische Reflexion zu Unterschieden und Gemeinsamkeiten der beiden Fachdisziplinen steht jedoch nicht im Fokus dieses Beitrages, sondern wird zukünftig an anderer Stelle fokussiert.

schichtstheorie als eine komplexe Angelegenheit dar (u. a. Trautwein et al., 2017; Berger, Lorenz & Brauch, 2021). Es würde den Rahmen dieses Beitrages mit seiner forschungspragmatischen Ausrichtung sprengen, die Verästelungen dieses Diskurses der Narrativitätstheorien im Einzelnen nachzuvollziehen (vgl. dazu Berger, Lorenz & Brauch, 2021). Im Folgenden werden daher lediglich die theoretischen Eckpfeiler für die Entwicklung der hier im Zentrum stehenden Unterstützungsmaßnahmen für die Förderung schriftlichen historischen Erklärens skizziert, wie sie im Konstrukt *Causal Reasoning in History* (CRH) modelliert und für das Schülerlaborprojekt exemplarisch operationalisiert wurden (Brauch et al., 2020). Als Orientierung diente der Begriff der «narrativen Erklärung» von Jörn Rüsen. Rüsen's Begriff der narrativen Erklärung beinhaltet die Annahme, «dass das Erzählen einer Geschichte selber einen erklärenden Charakter» habe (Rüsen, 2013, 162), und löst damit in der Domäne Geschichte den vermeintlichen Widerspruch zwischen Erklärung und Erzählung auf (Süßmann, 2002/2019, 85–87; vgl. auch Danto, 1965). Fachspezifisch geht es nach Rüsen in der narrativen Erklärung daher darum, «zeitliche Veränderungen dadurch zu erklären, dass man den Veränderungsvorgang erzählt» (Rüsen, 2013, 165; vgl. ähnlich auch Pandel, 2017, 407–408). In diesem Sinn ist historisches Erklären dadurch gekennzeichnet, dass grundsätzlich die Beantwortung einer historischen Leitfrage über einen zeitlichen Veränderungsvorgang angestrebt wird und die Erklärung durch den transparenten Bezug auf historisch valides Material als empirisch plausibel begründet werden kann (Brauch, 2019b; Mierwald, 2020; Seixas & Morton, 2013). In jüngerer Zeit hat vor allem Saskia Handro diese Aspekte des leitfragengestützten fachspezifischen Erklärens in die historischen Basiskategorien von Zeit und Wandel, Kausalität, Raum und Akteur*innen operationalisiert und für die forschungspragmatische Entwicklung von Scaffolds für die sprachbildende Förderung des fachspezifischen Lesens zugrunde gelegt (Handro & Kilimann, 2019).

Für den hier vorgestellten Zugriff wurde dieses Konstrukt geschichtstheoretisch erweitert und insbesondere das Prinzip der Evidenz integriert (Seixas & Morton, 2013, 40–73; vgl. korrespondierend dazu auch die empirische Plausibilität nach Rüsen, 2013, 60–61). Im Folgenden wird unter historischem Erklären daher eine empirisch reflektierte Antwort auf eine historische Leitfrage verstanden, die dabei die epistemologischen Prinzipien der narrativistischen Geschichtstheorie berücksichtigt: Narrativität, Perspektivität, Retrospektivität, Partikularität/Selektivität, Konstruktivität (Trautwein et al., 2017, 15–16). Operationalisiert man das so fundierte Konstrukt des historischen Erklärens in exemplarische (nicht umfassende) Bestandteile eines Erkenntnisprozesses mit dem Ziel einer

schriftlichen historischen Erklärung, so könnten diese forschungspragmatisch³ wie folgt beschrieben werden: Im Anschluss an eine historische Fragestellung zur Erklärung eines zeitlichen Veränderungsvorganges (historische Frage) werden materialbasiert und unter Berücksichtigung der Notwendigkeit der begründeten Auswahl (Selektivität) mögliche Einflussfaktoren auf das zu erklärende zeitliche Phänomen identifiziert. Dabei können sowohl Akteur*innen als auch Strukturen eine Rolle spielen, respektive interagieren, wobei zu klären ist, wie diese multiplen Faktoren kausal (Kausalität) und temporal (Temporalität) zusammenhängen. Es ist zudem darüber zu befinden, in welchem Ausmaß sich die multiplen Faktoren auf das zu erklärende historische Phänomen ausgewirkt haben (Konstruktivität, Gewichtung) und inwiefern die Material- beziehungsweise Datengrundlage für diese Gewichtung empirisch plausibel zu bewerten ist (Evidenz).

Dabei zeigen bisherige Untersuchungen, dass Schüler*innen mit der Umsetzung dieser Prinzipien Probleme haben (Seixas & Morton, 2013; Carretero & Lee, 2014; Günther-Arndt, 2014; Stoel, 2017; Handro & Kilimann, 2019). Auch in Geschichtsschulbüchern als zentralen Unterrichtsmedien werden diese häufig nur unzureichend explizit gemacht (Schrader, 2013) und zur aktiven Auseinandersetzung durch Schüler*innen aufbereitet (Bramann, 2020; Brauch, 2019a; Brauch & Jahn, 2018). Konkrete Ansätze zur Entwicklung fachsprachunterstützender Hilfen zur Förderung des historischen Erklärens existieren jedoch erst in Ansätzen (Mierwald & Brauch, 2015; Seixas & Morton, 2013; Stoel, van Drie & van Boxtel, 2015; Woodcock, 2005). Besonders vielversprechend scheint der Ansatz des «*explicit teaching*» zu sein, den Gerhard Stoel (2017) und Marcel Mierwald (2020) in quasiexperimentellen Untersuchungsdesigns mit Vergleichsgruppen anwandten und dabei positive Effekte sowohl von expliziten sprachlich-formalen als auch epistemologischen Unterstützungen nachweisen konnten, weshalb auch die hier vorgestellte Studie einem solchen Ansatz – in reduzierter Form – folgt.

3 Das sehr viel größere Universum der Spielarten des historischen Erklärens ist damit sicher nicht abgedeckt, vgl. dazu auch Lorenz, 1997.

2.2 Die Scaffolds für historisches Erklären und Forschungsfragen

Im Modell *Causal Reasoning in History* (CRH) wurden verschiedene Wissenstypen im Sinne konventioneller Niveaustufen (faktual, konzeptionell, prozedural; vgl. geschichtsdidaktisch profiliert auch bei Günther-Arndt, 2014; Kühberger, 2012) bestimmten fachlichen Konventionen (z. B. die Re-Konstruktion im Sinne einer «Historischen Erklärung»; das Verfügen über inhaltsbezogene Kategorien und Konzepte; epistemologische Prinzipien) und fachsprachlichen Konventionen (z. B. die fachlich und sprachlich angemessene Umsetzung des Genres einer «Historischen Erklärung» sowie der dafür notwendigen epistemologischen Prinzipien) zugeordnet (Brauch et al., 2020, 145–153). Aus dieser zunächst abstrakten Systematisierung wurden für die hier vorgestellte Interventionsstudie im Sinne einer forschungspragmatischen Operationalisierung sieben fachliche Merkmale historischen Erklärens abgeleitet und sieben fachliche und fachsprachliche Scaffoldgruppen entwickelt, die Schüler*innen bei der Erstellung und Überarbeitung von Drehbuchtexten für ein Erklärvideo unterstützen sollen (Brauch et al., 2020, 153–157, siehe weiter unten). Aus der Struktur der abgeleiteten Merkmale geht hervor, dass die entwickelten Scaffoldingtools vor allem Scaffolds zu epistemologischen Aspekten historischer Erklärungen beinhalten (Scaffolds 2 bis 6, explizite Förderung der Anwendung konzeptuellen Wissens), die durch zwei strukturelle Merkmale des Genres einer «Historischen Erklärung» ergänzt wurden (Scaffolds 1 und 7, explizite Förderung der Anwendung prozeduralen Wissens) (vgl. auch Mierwald & Brauch, 2015).

(1) Die zu klärende historische Frage	→	Merkmal <i>Frage</i>
(2) Selektion der Einflussfaktoren	→	Merkmal <i>Selektivität</i>
(3) Kausalitäten	→	Merkmal <i>Kausalität</i>
(4) Temporalitäten	→	Merkmal <i>Temporalität</i>
(5) Gewichtung der Einflussfaktoren	→	Merkmal <i>Gewicht</i>
(6) Empirische Kausalität	→	Merkmal <i>Evidenz</i>
(7) Zusammenfassende Erklärung	→	Merkmal <i>Antwort</i>

Abbildung 1: Sieben epistemologische Merkmale des historischen Erklärens, abgeleitet aus dem CRH-Modell (Brauch et al., 2020, 145–157)

Da das Projekt in einem interdisziplinären Forschungsverbund entwickelt wurde, war zusätzlich ein übergeordnetes Forschungsinteresse auf die Frage gerichtet, welche Art Scaffoldingtool (fachlich, sprachlich, integrativ) Schüler*innen bei der Sprachhandlung des jeweiligen fachspezifischen Erklärens am besten unterstützen kann. Die Arbeitshypothese zielte darauf, dass Schüler*innen von einem integrativen Tool – das heißt einer Kombination aus der expliziten Erläuterung der epistemologischen Funktion der Merkmale fachlichen Erklärens bei gleichzeitiger Unterstützung durch jeweils passende Wortbeispiele – mehr profitieren als von ausschließlich fachlichen oder sprachlichen Scaffolds. Für die quasiexperimentelle Intervention wurden daher drei verschiedene Scaffoldingtools entwickelt, die jeweils die sieben epistemologischen Merkmale einer historischen Erklärung aufgreifen:

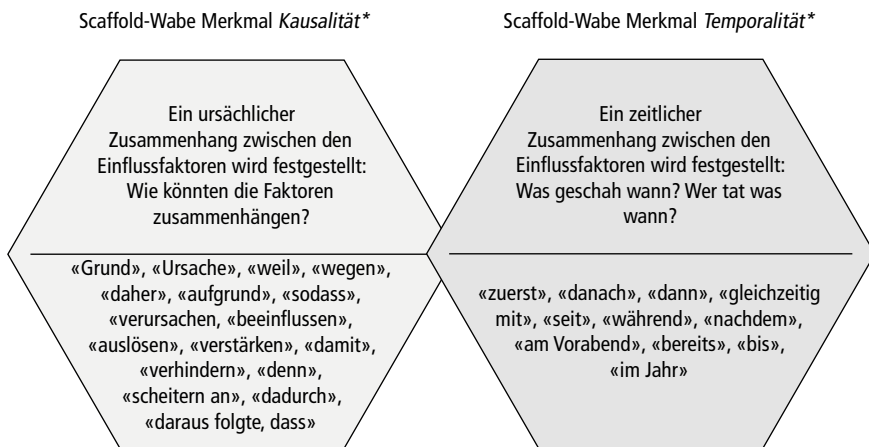
1. **Scaffoldingtool** für die **Interventionsgruppe Fach**
(= Scaffolds mit fachlichen Erläuterungen zu jedem epistemologischen Merkmal);
2. **Scaffoldingtool** für die **Interventionsgruppe Sprache**
(= Scaffolds mit sprachlichen Beispielen);
3. **Scaffoldingtool** für die **Interventionsgruppe Fach+Sprache**
(= Scaffolds mit fachlichen Erläuterungen und sprachlichen Beispielen).

Während für die Formulierung der fachlichen Scaffolds die vorgestellten Teilprinzipien historischen Erklärens jeweils in Form einer knappen Erläuterung mit anschließender Frage operationalisiert wurden, erfolgte die Definition der sprachlichen Scaffolds nach einer exemplarischen Auswahl verschiedener Konventionen der Versprachlichung dieser Bestandteile in der historiografischen Praxis (Lorenz, 1997). Eine besondere Herausforderung bildete hierbei die trennscharfe Unterscheidung zwischen einzelnen Worthilfen zu den Merkmalen *Kausalität* und *Temporalität*, auf die bereits in anderen Forschungszusammenhängen hingewiesen wurde (vgl. Schrader, 2013; siehe Abbildung 2 sowie auch Abbildung 4).⁴

Die Unterscheidung der sieben Merkmale einer historischen Erklärung in den Scaffoldingtools erfolgte über die Darstellung in «Waben», welche die jeweiligen Scaffolds zu den einzelnen Merkmalen enthalten (siehe Abbildung 2). Die

4 Insbesondere unter dem Merkmal Kausalität wurden daher auch Konnektoren zusammengefasst, die auf konsekutive und finale Verbindungen im Sinne von «Ursache und Folge» schließen lassen (Welskopp, 2002/2019, 81).

Visualisierung in Wabenform wurde gewählt, um die Waben in zwei Hälften unterteilen zu können, sodass die Interventionsgruppe *Fach+Sprache* sowohl die obere Hälfte zur fachlichen Erläuterung als auch die untere Hälfte für sprachliche Beispiele verwenden konnte, während die Waben der Interventionsgruppe *Fach* und der Interventionsgruppe *Sprache* jeweils nur den oberen fachlichen beziehungsweise unteren sprachlichen Teil eines Merkmals beinhalteten. Da es sich um ein laufendes Forschungsprojekt handelt, in das verschiedene Disziplinen involviert sind – und eine ausführliche Darstellung und Diskussion aller sieben Scaffolds zudem den hier verfügbaren Raum übersteigen würde –, wird im vorliegenden Beitrag der Fokus exemplarisch auf die im Kontext historischer Fachsprache besonders hervorzuhebenden Merkmale der *Kausalität* und *Temporalität* gelegt (siehe Abbildung 2).



*Die Zuordnung wird im eigentlichen Scaffoldingtool nicht vorgegeben.

Abbildung 2: Ein Beispiel aus dem Scaffoldingtool der Interventionsgruppe *Fach+Sprache*, hier: die Scaffolds zu den Merkmalen *Kausalität* (links) und *Temporalität* (rechts)

Die Nutzung der Scaffoldingtools durch die Schüler*innen wurde in der im Folgenden beschriebenen explorativen Interventionsstudie empirisch geprüft. Dabei wurden folgende Forschungsfragen adressiert, die in Abschnitt 3.2 hinsichtlich der gewählten Auswertungsstrategien weiter spezifiziert werden:

- *Forschungsfrage 1 (FF1):*
Inwiefern wurden welche Scaffolds in den drei Interventionsgruppen genutzt?
- *Forschungsfrage 2 (FF2):*
Welche Auswirkungen haben die unterschiedlichen Versionen des Scaffoldingtools (sprachlich – fachlich – integrativ) auf die Methode der Überarbeitung der Drehbücher?

3 Aufbau der Interventionsstudie

3.1 Aufbau und Durchführung der Schülerlaborstudie

Die explorative Interventionsstudie wurde im Alfried-Krupp-Schülerlabor (AKS) an der Ruhr-Universität Bochum durchgeführt (zur empirischen Forschung im AKS, vgl. Behrendt & Brauch, 2019). An den vier Projekttagen nahm jeweils eine Schulklasse inklusive der unterrichtenden Geschichtslehrpersonen teil.⁵ Inhaltlich ging es um einen «Klassiker» des Geschichts-Curriculums: Die Ursachen der Französischen Revolution. In der Vorbereitung der Intervention wurde deutlich, dass man sich auf eine von verschiedenen möglichen Ursachen beschränken musste, da das Arbeiten mit den Scaffolds und die dafür notwendigen Einführungen inhaltlicher, aber auch technischer Art sehr viel Zeit in Anspruch nahmen. Die Gefahr, damit den Hang der Schüler*innen zur Monokausalität zu verstärken, wurde versucht zu entschärfen, indem die fokussierte Ursache «Wirtschaft» als eine unter vielen verschiedenen in der Forschung diskutierten Ursachen herausgestellt wurde. Die Einführung in das Textgenre der historischen Erklärung erfolgte anhand der weiteren Ursache «Absolutismus-Kritik». Die Schülerlabortage waren auf 6,5 Zeitstunden angelegt und bestanden aus fünf Phasen, während derer die Schüler*innen selbstständig (also ohne weitere Unterstützung durch Lehrkräfte oder Testleitung) mit den Scaffolds arbeiteten:

5 Testleiter war Christoph Bramann (Erstautor dieses Beitrags). Besonderer Dank gilt zudem Kira Busse, langjährige Hilfskraft am Arbeitsbereich Geschichtsdidaktik, für ihr umfangreiches Engagement für diese Studie, auch bei der Vor- und Nachbereitung der Schülerlabortage.

- (1) **Einführungsphase:** Vorstellung eines prototypischen Erklärvideos
- (2) **Arbeitsphase I** (*Einzelarbeit am Interventionsgruppentisch*):
Erstellung eines Erklärvideo-Drehbuchs (= **Auswertungsdaten: «Drehbuch A»**)
- (3) Übungsphase (*gemeinsame Arbeit am Interventionsgruppentisch*):
Übung zur Anwendung der Scaffolds (Vorbereitung Interventionsphase)
- (4) **Arbeitsphase II** (*Einzelarbeit am Interventionsgruppentisch*):
Überarbeitung des Drehbuchs (= **Auswertungsdaten: «Drehbuch B»**)
- (5) **Arbeitsphase III** (*Arbeit in beliebigen Gruppen*):
Produktion eines animierten Erklärvideos auf Basis der erstellten Drehbücher (Daten wurden nicht ausgewertet)

Die für die Studie relevanten Forschungsdaten wurden damit in den Arbeitsphasen I und II generiert.⁶ An den Erhebungen nahmen insgesamt $N = 80$ Schüler*innen im Alter zwischen 14 und 18 Jahren teil ($M = 16.39$; $SD = 1.0$; 40 weiblich). Neben Schüler*innen der 9. Klasse eines Gymnasiums (Gruppe 1: $n = 23$) beteiligten sich Schüler*innen der 11. Klasse (Einführungsphase G9) zweier Gesamtschulen (Gruppe 2: $n = 27$; Gruppe 3: $n = 11$; Gruppe 4: $n = 19$) aus Nordrhein-Westfalen an der Studie. Bei der Durchführung wurde durch die zufällige Anordnung von Gruppentischen, an die jeweils eines der Scaffoldingtools ausgegeben wurde sowie eine ausgeglichene Aufteilung der drei Scaffoldingtools innerhalb jeder Lerngruppe geachtet, sodass sich die Verteilung der Interventionen insgesamt wie folgt darstellt: $n_{\text{Fach}} = 26$, $n_{\text{Sprache}} = 27$ und $n_{\text{Fach + Sprache}} = 27$ (siehe Tabelle 1).

6 Die nach einer Mittagspause durchgeführte *Arbeitsphase III*, in der die Schüler*innen auf Basis ihrer Drehbücher in beliebigen Gruppenkonstellationen ein animiertes und mit einer Audiospur unterlegtes Erklärvideo gestalten durften, war nicht zur Auswertung im Projekt bestimmt, sondern diente in erster Linie der Motivation für die Arbeitsphasen I und II.

Tabelle 1: Überblick über die Teilnehmer*innen der Interventionsstudie, inklusive soziodemografischer Daten und Verteilung der drei Scaffoldingtools

Lerngruppe	Jgst.	Schulform	Anzahl Schüler*innen	Intervention mit Scaffoldingtool			Familiärer Hintergrund ⁷	Kontakt zur deutschen Sprache ⁸	Schulnote Geschichte
				Fach	Sprache	Fach und Sprache			
G1	9	GYM	23	7	8	8	M = 2.09 SD = 0.85	M = 1.55 SD = 0.80	M = 3.22 SD = 0.80
G2	11	GS	27	9	9	9	M = 2.22 SD = 1.12	M = 1.11 SD = 0.32	M = 2.15 SD = 0.91
G3	11	GS	11	4	3	4	M = 3.18 SD = 0.98	M = 1.09 SD = 0.30	M = 2.82 SD = 0.75
G4	11	GS	19	6	7	6	M = 2.53 SD = 1.02	M = 1.16 SD = 0.69	M = 2.47 SD = 0.61
Insgesamt			80	26	27	27	M = 2.39 SD = 1.05	M = 1.24 SD = 0.60	M = 2.63 SD = 0.89

Die Studie folgte einem quasiexperimentellen Aufbau. In der *Einführungsphase* wurden soziodemografische Daten, insbesondere zur sprachlichen Sozialisation der Lernenden erhoben, um die Ergebnisse auch in größeren interdisziplinären Forschungen zur Fachsprache einordnen und interpretieren zu können. Es folgte die Präsentation eines vom Projektteam entwickelten Erklärvideos⁹ zur Kritik an der Herrschaftsform des Absolutismus als einer möglichen Ursache für die eskalierende Situation vor dem Sturm auf die Bastille, das alle wesentlichen Faktoren einer prototypischen historischen Erklärung beinhaltet. Dadurch sollten die Schüler*innen im Sinne eines «Sprachbads» (Leisen, 2010, 76) passiv rezeptiv mit der Textsorte der historischen Erklärung in ihrer spezifischen Form als «Erklärvideo» konfrontiert werden.

Anschließend bekamen alle Schüler*innen in *Arbeitsphase I* ein 3-seitiges Material- und Aufgabenblatt mit dem Titel «Ursachenforschung in der Geschichtswissenschaft», das neben der Aufgabenstellung («Erkläre die schlechte

7 Erhoben über eine sechsstufige Skala zur Frage «Wie viele Bücher gibt es bei dir zu Hause?» 1 = «keine»; 2 = «1–10»; 3 = «11–50»; 4 = «51–100»; 5 = «101–250», 6 = «mehr als 250».

8 Erhoben über eine fünfstufige Skala zur Frage «Seit wann hast du Kontakt zur deutschen Sprache? 1 = «seit meiner Geburt»; 2 = «seit dem Kindergarten»; 3 = «seit der Grundschule»; 4 = «seit ...».

9 Besonderer Dank gilt Kira Busse, die dem Erklärvideo ihre Stimme lieh.

Wirtschaftslage in Frankreich am Vorabend der Französischen Revolution: (1) Entwirf eine Visualisierung Deiner Erklärung in Form einer Erklärmatrix, (2) Schreibe einen Drehbuchtext für ein Erklärvideo») und weiteren technischen Hinweisen auch den Textauszug «Frankreich am Vorabend der Revolution» aus dem dtv Atlas Weltgeschichte (Kinder, Hilgemann & Hergt, 2017, 295) als Materialgrundlage beinhaltete.¹⁰ Die Schüler*innen bekamen daraufhin 60 Minuten Zeit, um in Einzelarbeit ein Erklärvideo-Drehbuch zu schreiben (= Drehbuch A) sowie eine dazugehörige Visualisierung im Sinne eines temporal-kausalen Concept-Mapping (Chapman, 2017) zu erstellen, um zusätzlich zur text-sprachlichen Ebene parallel auch eine visualisierende Darstellungsform historischer Erklärungen zu erhalten.¹¹ Im nächsten Schritt wurden die Scaffoldingtools an die jeweiligen Interventionsgruppen (getrennt nach Gruppentischen) verteilt. Nach einer kleinen Übungsphase, in der die Schüler*innen in die Nutzung der Scaffoldingtools eingeführt wurden, bekamen die Schüler*innen in *Arbeitsphase II* noch einmal 30 Minuten Zeit, um ihre zuvor erstellten Drehbücher (= Drehbuch A) mithilfe der Scaffolds zu überarbeiten (= Drehbuch B). Die ersten Drehbücher (A) wurden vor der Überarbeitung digital gesichert, sodass die Daten für die Auswertung vergleichend gegenübergestellt werden konnten (siehe Abbildung 3).

Die in Abbildung 3 kursiv gesetzten Stellen gehen (vermutlich) auf die Verwendung des hier zur Verfügung gestellten Scaffoldingtools *Fach+Sprache* zurück. Dabei scheinen insbesondere die Scaffolds der Merkmale *Kausalität* (fünf Anwendungen: «damit», «große Auswirkungen», «Doch», «negativ beeinflusst», «Durch») und *Temporalität* (vier Anwendungen: «als», «Gleichzeitig», «Während», «Später») sowie *Frage* (eine Anwendung: «Im Folgenden wollen wir Euch erklären [...]») und *Antwort* (eine Anwendung: «Abschließend lässt sich sagen [...]») verwendet worden zu sein (siehe dazu auch die Scaffoldwabe in Abbildung 2).

10 Die dritte Seite beinhaltete ein kleines Glossar, in dem einige wichtige Begriffe und Personen erläutert wurden.

11 Eine Auswertung der so entstandenen Temporal-Causal-Maps steht noch aus und soll in einer späteren Projektphase erfolgen.

Drehbuchtexte für ein Erklärvideo zur schlechten Wirtschaftslage als möglicher Ursache für den Sturm auf die Bastille 1789	
Drehbuch A (vorher)	Drehbuch B (nachher)
1774: König Ludwig XVI beruft Anne Robert Turgot zum Finanz-Minister , diese geben den Getreidehandel frei was gegen die Erhöhung der Brotpreise deinen soll.	Hallo Leute, <i>Im folgenden wollen wir euch erklären wie es zu dem Staatsbankrott und der daraus folgenden Revolution gekommen ist</i>
Dann wird eine allgemeine Grundsteuer eingeführt die aber durch Marie Antoinette und ihrer Hofpartei zum fall gebracht wird.	1774 als König Ludwig XVI regierte beruft er Anne Robert Turgot zum Finanz-Minister , diese geben danach den Getreidehandel frei was gegen die Erhöhung der Brotpreise deinen soll.
Es werden Freiwillige unter La Favette Nach Amerika gesendet.	<i>Gleichzeitig damit</i> wird eine allgemeine Grundsteuer eingeführt die aber durch Marie Antoinette und ihrer Hofpartei zum fall gebracht wird.
1786: Amerikanischer Freiheitserfolg wirkt sich positiv auf den Handelsvertrag mit England aus.	<i>Während</i> der Amerikanischen Unabhängigkeitskriegs werden Freiwillige und La Favette nach Amerika gesendet.
Der Handelsvertrag mit England wirkt sich schlecht auf die Hungersnöte, Missernten und auf die Industriekrise aus.	Später 1786 hat der Erfolg der Amerikanischen Unabhängigkeit <i>große Auswirkungen</i> auf den Handelsvertrag zwischen Frankreich und England.
1788: Daraus ergibt sich der Staatsbankrott.	Doch der Handelsvertrag sorgt dafür, dass die Hungersnöte, Missernten und die daraus folgende Industriekrise, <i>negativ beeinflusst</i> werden. <i>Durch</i> die zu fall gebrachte Grundsteuer, die niedrigen Brotpreise und der Handelsvertrag mit England, kommt es zu einem Staatsbankrott 1788. <i>Abschließend lässt sich sagen</i> , dass der Erfolg der Amerikanischen Unabhängigkeit eine große Auswirkung auf den Staatsbankrott hatte und damit auch auf dir Französische Revolution.
ID G_G2_F+S_T2-23, Schüler, 17. Jahre, 11. Klasse, Gesamtschule; Deutschnote: 4, Geschichtsnote: 3	

Abbildung 3: Beispiel für ein Erklärvideo-Drehbuch vor (links) und nach Überarbeitung mit dem Scaffoldingtool *Fach+Sprache*

3.2 Methode der Datenauswertung

Bei den 80 überarbeiteten Drehbüchern wurden anhand einer inhaltlich-strukturierenden Inhaltsanalyse (Schreier, 2014) diejenigen Überarbeitungen von Drehbuchtext A im Unterschied zu Drehbuchtext B codiert, die vermutlich auf die Verwendung mindestens eines der in den Scaffoldingtools zur Verfügung gestellten Scaffolds zurückzuführen sind. Die Probecodierungen zur Entwicklung und Schärfung des Codierleitfadens (Mayring, 2010, 62) erfolgte anhand von Drehbüchern, die zuvor in einem Pilot von $N = 13$ Schüler*innen einer 10. Klasse eines Gymnasiums erstellt worden waren. Codierung und Auswertung erfolgten anhand der Software MaxQDA.¹² Die im Codierleitfaden definierten Analysekategorien wurden theoriegeleitet (deduktiv) aus den sieben entwickelten Merkmalen einer historischen Erklärung abgeleitet: (1) *Frage*, (2) *Selektivität*, (3) *Kausalität*, (4) *Temporalität*, (5) *Gewicht*, (6) *Evidenz* und (7) *Antwort* (siehe auch Abbildung 1), und zusätzlich anhand eines Suffix («_F», «_F+S» oder «_S») jeweils einer der drei Interventionsgruppen zugeordnet (siehe Abbildung 4). Als Codiereinheit wurde ein ganzer Satz definiert (mit Ausnahme der Kategorien 1 (*Frage*) und 7 (*Antwort*), bei denen ein Code auch mehrere Sätze umfassen konnte), weil die Anwendung einiger Scaffolds zu Überarbeitungen führen konnte, die über die Wortebene hinausgingen. Auch für die Bestimmung der Interrater-Übereinstimmung erwies sich die Codierung ganzer Sätze forschungspragmatisch als praktikabler, da ein Satz auch mit verschiedenen Kategorien gleichzeitig codiert werden konnte. Dadurch, dass die Überarbeitung der Drehbuchtexte durch die Schüler*innen methodisch sowohl in Form von Änderungen bereits vorhandener Textbausteine (Wort- und [Neben-]Satzebene) als auch in Form von Ergänzungen mit zusätzlichen Textabschnitten erfolgte (siehe FF2 sowie die Abschnitte 4.2 und 5.2), wurden die sieben Analysekategorien noch einmal induktiv jeweils als «Änderung» oder «Ergänzung» (Codes mit dem Zusatz «_N», für «Neu») codiert (siehe Abbildung 4).

Insbesondere zur Erfassung der sprachlichen Scaffolds wurden dabei auch Begriffe wie «so», «als» oder «dabei» codiert, die im Kontext der jeweiligen Satzkonstruktion entweder im Sinne einer kausalen oder temporalen Bedeutung interpretiert wurden. Dies lässt sich am Beispiel des folgenden Satzes verdeutlichen: «Die Situation spitze sich zu, *als* die Bauern sich mithilfe eines

¹² MAXQDA Plus 2020 Release 20.4.0.

Reformprogrammes mehr Rechte zusichern wollten» (G2_F+S_T1-12). Hier kann der temporalen Bedeutung von «als» zugleich eine kausale Bedeutung (z. B. im Sinne eines «weil») beigemessen werden, die deutlich wird, wenn man den Satz sprachlich umstellt und dabei die zeitliche Reihenfolge der beiden Ereignisse (Zuspitzen der Situation; Bauern fordern mehr Rechte) berücksichtigt: «Weil die Bauern sich [...] mehr Rechte zusichern wollten, spitze sich die Situation zu». In anderen Zusammenhängen kann hingegen die temporale Bedeutung von «als» eindeutig überwiegen: «Welches die *Zeit* war *als* sich Frankreich mit den USA gegen England verbunden hatte» (G4_S_T2-23).

Merkmal des historischen Erzählens: Kausalität

Codierregeln (Auszug):

- Es werden immer ganze Sätze codiert, also Haupt- und (kausaler) Nebensatz, auch wenn sich nur ein Wort ändert.
- Auch «unscharfe» kausale Adverbien, Konjunktionen etc. («als», «so» ...) werden als kausal codiert, wenn der Satz eine kausale Bedeutung wahrscheinlich macht (z. B. durch ein Ersetzen durch «weil» oder «auf diese Weise»).
- [...], dass [...] wird nur codiert, wenn damit ein kausaler, konsekutiver oder finaler Nebensatz eingeleitet wird
- Um Doppelcodierungen mit Kategorie 2 SELEKTIVITÄT zu vermeiden, muss beim Codieren zwischen zwei Arten von Kausalzusammenhang unterschieden werden:
 - Kausalzusammenhänge, die sich auf die Begründung der Faktorenwahl beziehen, werden mit Codes der Kategorie 2 SELEKTIVITÄT codiert.
 - Kausalzusammenhänge, die sich auf kausale Beziehungen zwischen den Faktoren beziehen, werden mit Codes der Kategorie 3 KAUSALITÄT codiert.

Codes (6): Kausal_F; Kausal_F_N; Kausal_F+S; Kausal_F+S_N; Kausal_S; Kausal_S_N

Ankerbeispiele (Auszug):

Code: Kausal_F+S	Drehbuch A [G_G0_F+S_T1-14]: Dies gefiel der untersten Schicht «unter anderem den Bauern» überhaupt nicht.	→	Drehbuch B [G_G0_F+S_T1-14]: Dadurch entschärfte sich die Situation in der untersten Schicht «unter anderem auch bei den Bauern».
Code: Kausal_F+S_N	Drehbuch A [G_G0_F+S_T1-04]: (kein Beispiel)	→	Drehbuch B [G_G0_F+S_T1-04]: So kam es schließlich zum Staatsbankrott 1788 und die Revolution konnte Fahrt aufnehmen.

Abbildung 4: Ausschnitt aus dem Codierleitfaden zu den Scaffolds für das Merkmal Kausalität

Zu jedem Scaffold lag damit für jedes Drehbuch ein Zahlenwert über die Häufigkeit des Vorkommens von Überarbeitungen bezüglich des jeweiligen Merkmals vor. Das vollständige Material wurde vom Projektleiter und Erstautor des vorliegenden Beitrags codiert. Zur Absicherung der Reliabilität erfolgte die Berech-

nung eines bereinigten Kappa-Koeffizienten in MaxQDA. Dazu wurden 22,5 Prozent des Datenmaterials von einer zweiten Codiererin gegencodiert.¹³ Nach einer Diskussion im Sinne eines «konsensuellen Kodierens» (Kuckartz, 2016, 211–212) und minimaler Anpassung des Codierleitfadens ergaben sich für die Kategorien nach Landis und Koch (1977) insgesamt gute bis sehr gute Werte: (1) *Frage*: $\kappa = 0.92$, (2) *Selektivität*: $\kappa = 0.66$, (3) *Kausalität*: $\kappa = 0.80$, (4) *Temporalität*: $\kappa = 0.86$, (5) *Gewicht*: $\kappa = 0.85$, (6) *Evidenz*: $\kappa = 0.94$ und (7) *Antwort*: $\kappa = 0.88$.¹⁴

Die Daten wurden anschließend in eine SPSS-Umgebung überführt, um über weiterführende statistische Analysen Aussagen über die Verwendung der Scaffolds treffen zu können.¹⁵ Hierfür wurden deskriptive Statistiken zum Vergleich der Nutzung der Scaffolds in den drei Interventionsgruppen (Ebene der sieben Merkmale, Interventionsgruppenebene) angefertigt und anschließend die proportionalen Anteile von Änderungen und Ergänzungen in allen Überarbeitungen quantifiziert. Da die Daten nicht normalverteilt sind und Werte gehäuft vorkamen, wurden die Daten hinsichtlich der beiden Forschungsfragen durch die Anwendung von Kendalls τ analysiert (Wirtz & Casper, 2002).

Zur Beantwortung von FF1 (*Inwiefern wurden welche Scaffolds in den drei Interventionsgruppen genutzt?*) wurden dabei folgende Analysen durchgeführt:

- FF1a: Ausmaß der Nutzung der Scaffolds in allen Interventionsgruppen und im Vergleich
- FF1b: Korrelationen der Scaffoldnutzung zu den Interventionsgruppen
- FF1c: Kombinationscluster von Scaffoldnutzungen

Zur Beantwortung von FF2 (*Welche Auswirkungen haben die unterschiedlichen Versionen der fachlichen und sprachlichen Scaffolds auf die Methode der Überarbeitung der Drehbücher?*) wurden die proportionalen Anteile von Änderungen im Unterschied zu Ergänzungen bei allen Überarbeitungen quantifiziert. Bei signifikanten Korrelationen (mit $p \leq .05$)¹⁶ wurde anschließend mit einem Kruskal-Wallis-Test (Pett, 2015) überprüft, ob ein abhängiger Zusammenhang vorliegt.

13 Dank gilt hier erneut Kira Busse für die sorgfältige Durchführung der Zweitkodierung.

14 Der im Verhältnis zu den anderen Merkmalen niedrige Reliabilitätswert zum Merkmal der *Selektivität* ($\kappa = 0.66$), könnte auf die geringen Fallzahlen der in den Drehbüchern zu diesem Merkmal gefundenen Überarbeitungen zurückzuführen sein.

15 IBM SPSS Statistics Version 27.0.0.0 64-Bit-Version.

16 Korrelationen sind zwischen .1 und .3 schwach, zwischen .3 und .5 mittel und ab .5 stark (Cohen, 1992).

Ein Blick auf mögliche Korrelationen zwischen der Anwendung der Scaffolds und den soziodemografischen Daten der Schüler*innen zeigte zudem, dass die Anzahl der Überarbeitungen weder mit dem Bildungsgrad des familiären Hintergrundes (ausgedrückt durch die Anzahl der Bücher im elterlichen Haushalt, $\tau = -.056$, $p = .528$), noch mit dem Zeitpunkt des Erstkontakts zur deutschen Sprache ($\tau = -.009$, $p = .921$) oder mit der Geschichtsnote ($\tau = -.151$, $p = .091$) statistisch signifikant korreliert. Das ist ein Indiz dafür, dass die vorgenommenen Änderungen vermutlich auf die jeweiligen Scaffoldingtools – fachlich, sprachlich, integrativ – zurückgeführt werden kann.

4 Ergebnisse

4.1 Anwendung der Scaffolds nach Merkmalen historischen Erklärens (FF1)

Insgesamt wurden von den 80 Schüler*innen 409 Überarbeitungen an ihren Drehbuchtexten vorgenommen, die vermutlich auf die Verwendung der Scaffolds zurückgeführt werden können. Über die Hälfte der Überarbeitungen (232 von 409) fallen auf die Merkmale *Kausalität* (142) und *Temporalität* (90), während dem Merkmal *Selektivität* am wenigsten Überarbeitungen (21 von 409) zugeordnet werden konnten (siehe Abbildung 5, schwarze Balken). Die Scaffolds wurden demzufolge von den Schüler*innen grundsätzlich angenommen, wenn auch in sehr unterschiedlichen Ausmaßen (Analyse FF1a).

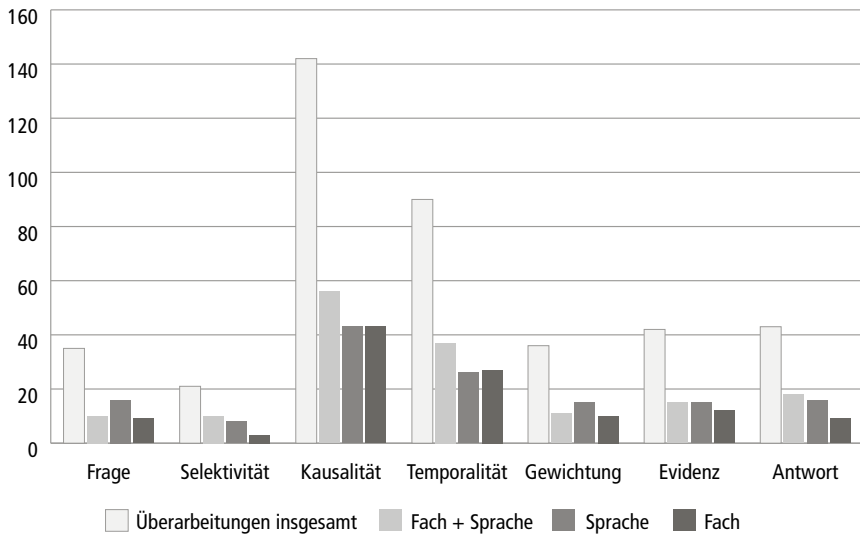


Abbildung 5: Verteilung der insgesamt 409 Überarbeitung auf die sieben Merkmale und aufgeteilt auf die Interventionsgruppen (*Fach* = 26; *Sprache* = 26; *Fach+Sprache* = 27)

Von den 409 Überarbeitungen fallen dabei die meisten, also 157 (38%) auf die Interventionsgruppe mit den sprachlichen und fachlichen Scaffolds, während die Interventionsgruppe mit rein sprachlichen Scaffolds 139 (34%) Überarbeitungen vorgenommen hat und die Interventionsgruppe mit dem rein fachlichen Scaffoldingtool die Scaffolds am seltensten verwendet hat (113 = 28%) (siehe Abbildung 5, Analyse FF1a). Das integrative Scaffoldingtool wird bei fünf von sieben Merkmalen ausgiebiger genutzt als die beiden einfachen Tools, obwohl ein aussagekräftiger Vorsprung lediglich bei den inhaltlichen Merkmalen *Kausalität* und *Temporalität* festgestellt werden kann (der Fokus auf diese beiden Merkmale zeigt sich jedoch auch in den anderen beiden Gruppen).

Bei einer explorativ festgelegten ordinalen Reihenfolge der drei Interventionsgruppen (1 = *Fach*; 2 = *Sprache*; 3 = *Fach+Sprache*) ergibt sich dabei eine schwache, aber signifikante Korrelation von $\tau = .181^*$ (*Sig. p* = .047) zwischen der Anzahl der Überarbeitungen und der Zugehörigkeit zu einer der drei Gruppen (FF1b). Das bedeutet, dass es einen positiven Zusammenhang (Korrelation) zwischen der Zugehörigkeit zu einer Interventionsgruppe und der Anzahl der Überarbeitungen gibt. Laut Kruskal-Wallis-Test hängt die Anzahl der Überarbeitung in den einzelnen Scaffolds jedoch nicht signifikant von der Interventionsgruppe ab, das

heißt, es kann daraus keine Kausalität abgeleitet werden. Dennoch lässt sich auf der Mikroebene zeigen, dass theoretisch – wenn auch nicht statistisch signifikant – das integrative Scaffoldingtool – zumindest bei den zentralen Merkmalen des historischen Erklärens (*Kausalität* und *Temporalität*) – zu mehr Überarbeitungen führte als das rein sprachliche oder das rein fachliche Scaffoldingtool (siehe Abbildung 5).

Neben dem quantitativen Vergleich des Nutzungsverhaltens in den unterschiedlichen Interventionsgruppen interessierte uns auf Ebene der Gesamtzahl der überarbeiteten Drehbücher ($N = 80$) auch die Quantität der Nutzung der sieben Scaffolds. Das Ergebnis der deskriptiven Analyse (siehe Tabelle 2) zeigt auch hier einen erhöhten Fokus auf Hilfen der inhaltlichen Merkmale *Kausalität* und *Temporalität* (diese kommen im Schnitt in jedem Drehbuch mindestens ein Mal vor) und unterstreicht damit das Ergebnis des Interventionsgruppenvergleichs (siehe Abbildung 5) noch einmal.

Tabelle 2: Deskriptive Statistik zur Nutzung der Scaffolds in den einzelnen Drehbüchern¹⁷

Scaffolds	N	Minimum	Maximum	M	SD
<i>Frage</i>	80	0	1	.45	.501
<i>Selektivität</i>	80	0	2	.25	.490
<i>Kausalität</i>	80	0	6	1.78	1.378
<i>Temporalität</i>	80	0	4	1.10	1.186
<i>Gewicht</i>	80	0	3	.49	.693
<i>Evidenz</i>	80	0	2	.51	.675
<i>Antwort</i>	80	0	1	.54	.502

Zusammenfassend lässt sich daher sagen, dass die Scaffolds zur fachspezifischen Versprachlichung der Merkmale *Temporalität* und *Kausalität* absolut gesehen mit Abstand am meisten genutzt wurden (siehe Abbildung 5). Alle weiteren Merkmale, sowohl die beiden strukturellen (*Frage* und *Antwort*) als auch die übrigen inhaltlichen Merkmale (*Selektivität*; *Gewicht*; *Evidenz*) fielen mit weniger als fünfzig Überarbeitungen weit dahinter zurück. Dennoch bleibt festzuhalten,

17 Am Beispiel der Scaffolds des Merkmals *Frage* kann diese Tabelle folgendermaßen gelesen werden: Pro Drehbuch wurden mindestens null Mal, maximal ein Mal und im Schnitt 0.45 Mal Überarbeitungen mit Scaffolds des Merkmals *Frage* vorgenommen.

dass im Schnitt etwa die Hälfte der Drehbücher auch Überarbeitungen in Bezug auf die Merkmale *Gewicht* und *Evidenz* sowie die Merkmale *Frage* und *Antwort* aufweist. Das Gesamtbild der Analyse zeigt daher auch, dass das Scaffold zum Merkmal der *Selektivität* in den Überarbeitungen am seltensten genutzt wurde (Überarbeitungen hierzu sind im Schnitt nur in rund einem Viertel der Drehbücher zu finden).

Eine anschließende Analyse, inwiefern sich eine gehäufte Clusterung verschiedener Scaffolds feststellen lässt (FF1c), ergab in der Nutzung der Scaffolds untereinander nur wenige Korrelationen, die zudem alle in einem schwachen Bereich ($\tau < .3$) lagen.¹⁸ Das bedeutet, dass die festgestellten korrelierenden Nutzungen bestimmter Merkmale gehäuft gemeinsam auftreten, dieser Zusammenhang aber nicht stark ist. Darüber hinaus kann jedoch festgestellt werden, dass Schüler*innen, die ihre Drehbücher durch Nutzung des Scaffolds des Merkmals *Kausalität* überarbeitet haben, auch mit den Scaffolds der Merkmale *Temporalität*, *Gewicht* und *Antwort* gearbeitet haben.

4.2 Anwendung der Scaffolds nach Methode der Überarbeitung (Änderungen und Ergänzungen; FF2)

Zur Beantwortung der zweiten Forschungsfrage wurden die proportionalen Anteile von Änderungen im Unterschied zu Ergänzungen bei allen Überarbeitungen quantifiziert. Die Forschungsfrage ergab sich aus der induktiven Datenanalyse, die ergeben hatte, dass die Schüler*innen in der Interventionsphase (Arbeitsphase II) ihre in Arbeitsphase I erstellten Texte nicht nur abgeändert (= Änderungen), sondern in tendenziell größeren Anteilen neue Textabschnitte (= Ergänzungen) verfasst hatten (siehe auch Kapitel 3). Da vermutet wurde, dass diese unterschiedlichen methodischen Herangehensweisen auf die unterschiedlichen sprachlichen Darstellungsweisen der Scaffolds zurückzuführen sein könnten, wurde der Beobachtung quantifizierend nachgegangen. Neue Textteile

18 Es korreliert die Nutzung der Scaffolds des Merkmals *Selektivität* mit denjenigen der *Evidenz* ($\tau = .254^*$) signifikant, die Nutzung der Scaffolds des Merkmals *Kausalität* mit der Nutzung der Scaffolds der Merkmale *Temporalität* ($\tau = .217^*$) und *Gewicht* ($\tau = .214^*$) signifikant und mit derjenigen des Scaffolds des Merkmals *Antwort* ($\tau = .274^{**}$) hoch signifikant. Außerdem korreliert die Nutzung der Scaffolds des Merkmals *Gewicht* hoch signifikant mit derjenigen der Scaffolds der Merkmale *Evidenz* ($\tau = .280^{**}$) und *Antwort* ($\tau = .369^{**}$). Korrelationen sind auf dem 0.05-Niveau signifikant (zweiseitig). Mit ** versehene Korrelationen sind auf dem 0.01-Niveau signifikant (zweiseitig). Korrelationen zwischen .1 und .3 sind schwach, zwischen .3 und .5 mittel und ab .5 stark (nach Cohen, 1992).

(«N») schrieben die Schüler*innen demnach insbesondere zu den Merkmalen der *Frage*, *Selektivität*, *Kausalität*, *Temporalität*, *Evidenz* und *Antwort*. Lediglich zum Merkmal *Gewicht* gab es (geringfügig) mehr Textänderungen (siehe Abbildung 6).

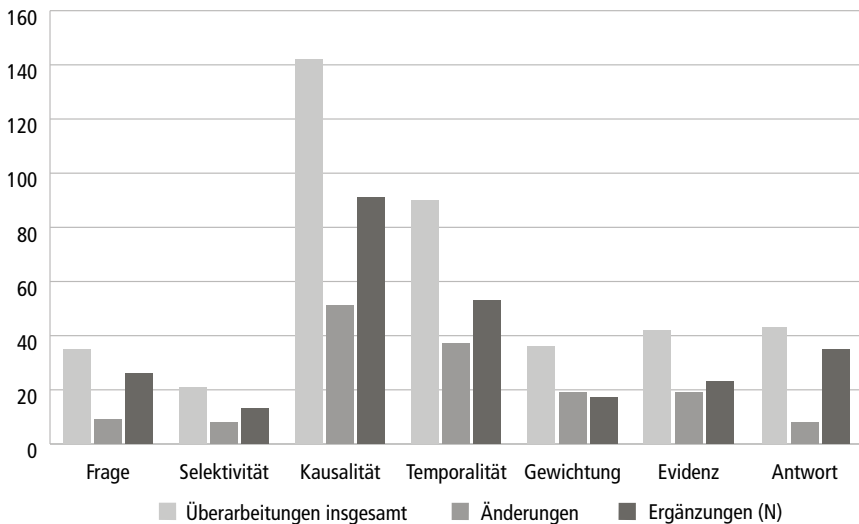


Abbildung 6: Verteilung der Überarbeitung je Merkmal in Bezug auf die verwendete Methode: (Text-)Änderung vs. (Text-)Ergänzung

Auf der Ebene der Interventionsgruppen zeigt sich jedoch ein differenzierteres Bild (siehe Abbildung 7), da insbesondere die Zuordnung zu einer Interventionsgruppe einen gewissen Einfluss auf die proportionalen Anteile von Änderungen und Ergänzungen in den Drehbüchern gehabt zu haben scheint. So überwiegen bei der Interventionsgruppe *Fach* bei jedem Merkmal die Ergänzungen gegenüber den Änderungen. Ein ähnliches Bild zeigt sich bei der Interventionsgruppe *Fach+Sprache*, in der sich zu fünf der sieben Merkmale mehr Ergänzungen als Änderungen finden. Auffällig ist außerdem die Diskrepanz zwischen Änderungen und Ergänzungen bei der Interventionsgruppe *Fach* bezüglich der Nutzung der Scaffolds zu den Merkmalen *Kausalität* und *Temporalität* zugunsten von Ergänzungen. Dagegen wurden in der Interventionsgruppe *Sprache* bezüglich der Merkmale *Kausalität*, *Temporalität* und *Evidenz* sogar mehr Änderungen als Ergänzungen vorgenommen. Es lässt sich demnach festhalten, dass die Schüler*innen mit sprachlichen Scaffolds das Grundgerüst ihres Textes weitgehend beibehielten

und vor allem Änderungen vorgenommen haben, während es in den anderen beiden Interventionsgruppen zu mehr Ergänzungen ganzer Textteile gekommen ist.

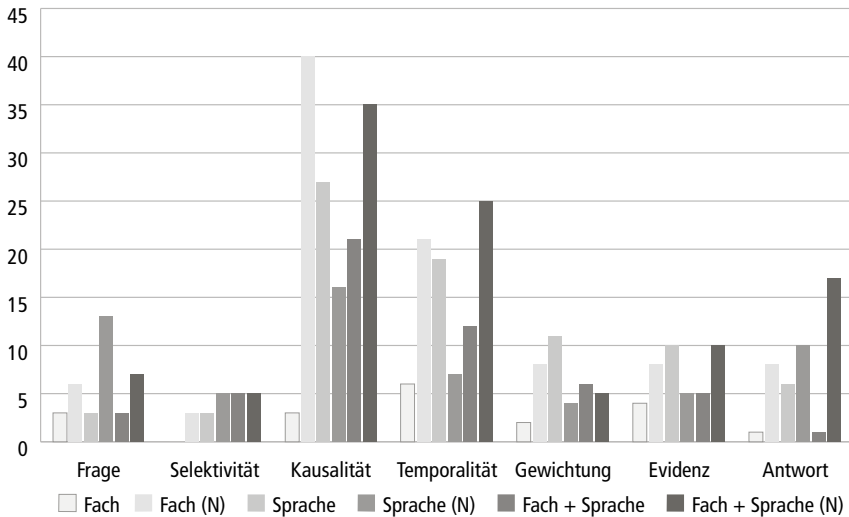


Abbildung 7: Verteilung der Änderungen und Ergänzungen auf die Scaffolds und Interventionsgruppen (Gruppenbezeichnung mit «N» [für «Neu»] stehen für Ergänzungen)

5 Interpretation der Befunde und Ausblick

Bereits die Entwicklung und Anwendung des Codierleitfadens zur Bearbeitung der ersten Forschungsfrage führte zu zwei Ergebnissen. Zum einen ist es bei der deduktiven Codierung gelungen, einen reliablen Codierleitfaden mit Erläuterungen zu jedem der sieben Merkmale historischen Erklärens zu entwickeln. Die quantifizierte Reliabilität der Beschreibungen der sieben Merkmale könnte darüber hinaus als ein Hinweis darauf gesehen werden, dass es sich bei den sieben definierten Merkmalen zudem um trennscharfe Konstrukte handeln könnte, die sich in weiteren Interventionen auch getrennt voneinander und im Verhältnis zueinander beforschen lassen könnten. Zum anderen zeigte sich bei der induktiven Analyse des Materials die Operationalisierung des Konstrukts der Überarbeitung in unterschiedlichen Methoden (Text[ab]änderung und Textergänzung). Aus dieser Beobachtung wurde eine zweite Forschungsfrage nach möglichen

Zusammenhängen zwischen der Überarbeitungsmethode und der Anwendung bestimmter Scaffolds abgeleitet.

5.1 Interpretation zur Nutzung der Scaffolds

Das Ergebnis, dass in den 80 Drehbüchern insgesamt 409 Überarbeitungen festgestellt werden konnten, zeigt, dass die Schüler*innen aller drei Interventionsgruppen die Scaffolds auch genutzt haben. Vor dem Hintergrund, dass hierbei keine statistisch signifikanten Korrelationen zum Bildungsgrad des familiären Hintergrunds, zum Zeitpunkt des Erstkontakts zur deutschen Sprache oder zur den Schulleistungen im Fach Geschichte festgestellt werden konnten, kann dieses Ergebnis dahingehend interpretiert werden, dass die Nutzung der entwickelten Scaffoldingtools weitgehend unabhängig vom soziodemografischen Hintergrund erfolgte und die Hilfen sich daher grundsätzlich zur Anwendung in heterogenen Schulklassen eignen könnten. Zudem konnte festgestellt werden, dass mehr als die Hälfte der Überarbeitungen durch die Nutzung von Scaffolds der Merkmale *Kausalität* und *Temporalität* erfolgte und damit in zwei Kernaspekten historischen Erklärens. Auch wenn in der integrativen Interventionsgruppe *Fach+Sprache* insgesamt etwas mehr Überarbeitungen codiert werden konnten als bei den beiden anderen Interventionsgruppen, kann zusammenfassend festgehalten werden, dass alle drei Scaffoldingtools die Schüler*innen motivierten, insbesondere mit Scaffolds zu den Merkmalen *Kausalität* und *Temporalität* zu arbeiten.

Die Stärke der Scaffoldingtools in Bezug auf die Überarbeitung der Merkmale *Kausalität* und *Temporalität* zeigte sich auch darin, dass sich in 71 der 80 Drehbücher Überarbeitungen zu mindestens einem der beiden Merkmale finden, während im Schnitt die Hälfte der Drehbücher auch Überarbeitungen zu den strukturellen Merkmalen *Frage* und *Antwort* sowie den inhaltlichen Merkmalen *Gewicht* und *Evidenz* enthielten (lediglich das Merkmal *Selektivität* wurde nur wenig berücksichtigt). Dieser Befund wird in Bezug auf die Fokussierung der Schüler*innen auf die Überarbeitung der Merkmale *Kausalität* und *Temporalität* ergänzt durch das Ergebnis der Analyse nach Kombinationsclustern in der Überarbeitung, die zeigte, dass Schüler*innen, die Überarbeitungen zum Merkmal *Kausalität* vorgenommen haben, auch dazu tendierten, die Merkmale *Temporalität*, *Gewicht* und *Antwort* zu überarbeiten.

Das Ergebnis, dass Schüler*innen die entwickelten Scaffolds grundsätzlich annehmen und auch nutzen, lässt sich als Bestätigung bisheriger Forschungen

zum Einsatz expliziter Scaffolds zur Förderung des historischen Erklärens einordnen (Stoel et al., 2017). Dabei ist die vermehrte Nutzung der Scaffolds zu den Merkmalen *Kausalität* und *Temporalität* im Sinne der Kompetenzförderung des historischen Denkens durch Scaffolding aus geschichtsdidaktischer Perspektive durchaus ermutigend, weil gerade diese beiden Aspekte im Geschichtsunterricht wenig explizit repräsentiert zu sein scheinen. Nimmt man als Maßstab beispielsweise die im Unterricht häufig verwendeten Verfasser*texte aus Schulbüchern, zeigt sich, dass das Merkmal *Temporalität* in vielen Büchern zwar als Zeitstrahl visuell präsent ist, der konkreten temporalen Abfolge historischer Veränderungsvorgänge in der Narrativierung des Verfasser*textes häufig jedoch wenig Beachtung geschenkt wird. Gleichzeitig wird in den Texten durch eher vage temporale Zusammenhänge eine implizite Kausalität vorgegeben, die nur selten explizit dargelegt wird (Schrader, 2013).

Darüber hinaus könnten die Ergebnisse darauf hinweisen, dass die anderen inhaltlichen Merkmale historischen Erklärens (*Selektivität*, *Gewicht*, *Evidenz*) für Schüler*innen eine größere kognitive Herausforderung darstellen (und daher auch ebenfalls in Geschichtsschulbüchern unterrepräsentiert sein könnten). Zur Überprüfung dieser Vermutung wäre künftig jedoch eine Abfrage der Wahrnehmung der Schüler*innen zu den verschiedenen Merkmalen nötig, zum Beispiel über entsprechende nachgelagerte Interviews oder Cognitive Labs (Alavi, 2005). Trotzdem ist diese Beobachtung bereits auffallend – insbesondere in Bezug auf das Merkmal der *Evidenz*, da das Quellenparadigma eines der zentralen geschichtswissenschaftlichen Prinzipien darstellt. Demgegenüber steht auch hier die Einsicht, dass Verfasser*texte von Geschichtsschulbüchern nur selten auf die ihnen zugrundeliegenden Quellen und Darstellungen verweisen und dadurch (durch ihre Anonymität) sogar weiterführend eine «Objektivität» suggerieren (Brauch, 2015, 103–117, v. a. 111), deren negative Auswirkungen auf das historische Denken jüngst auch empirisch bestätigt wurde (Mierwald, 2020, 183). Vor allem das Merkmal der *Selektivität* scheint für Schüler*innen eine große Hürde darzustellen. Auch dieser Befund könnte darauf zurückzuführen sein, dass es sich hierbei um ein in historischen Darstellungen der Geschichtskultur kaum explizit anzutreffendes Merkmal handelt, über das selbst Historiker*innen selten sprachlich reflektieren. Gleichwohl ist es ein wichtiges Merkmal, denn ohne offenzulegen, dass für eigene historische Erklärungen immer nur eine geringe Auswahl von Informationen berücksichtigt werden können, läuft man schnell Gefahr, zu behaupten, die «ganze, wahre» Geschichte zu erklären.

5.2 Interpretation zu Zusammenhängen zwischen der Art der Scaffolds und der Methode der Überarbeitung.

Die Befunde zu den Zusammenhängen zwischen der Art der zur Verfügung gestellten Scaffolds (fachlich, sprachlich, integrativ) und der Methode der Überarbeitung lassen auch Interpretationen zu, die für eine konkrete Weiterentwicklung von Scaffolds zum historischen Erklären wertvolle Hinweise liefern können. Hierfür gerät das Konstruktionsprinzip der drei verschiedenen Scaffoldingtools in den Blick. Die Unterscheidung in fachliche (Erläuterung) und sprachliche Scaffolds (Wortebene) bewirkte offenbar, dass die Interventionsgruppe mit rein sprachlichen Scaffolds eher die bestehenden Texte auf Wortebene abänderte, anstatt Textteile zu ergänzen. Auf der anderen Seite scheinen das rein fachliche Tool und mehr noch insbesondere das integrative Scaffoldingtool, die Schüler*innen vor allem zu ergänzenden Überarbeitungen motiviert zu haben. Das könnte als Hinweis dafür interpretiert werden, dass auch fachliche Erläuterungen, wie hier zu den Merkmalen für das historische Erklären, die Förderung historischen Denkens und seiner fachsprachlichen Umsetzung unterstützen können.

5.3 Limitationen und Ausblick

Die Möglichkeiten zur Verallgemeinerung der vorgestellten Ergebnisse sind aufgrund des quasiexperimentellen und explorativen Ansatzes der Studie sowie vor dem Hintergrund der geringen Stichprobe begrenzt. Die vorgestellte Studie kann daher nur einen ausschnitthaften Einblick in eine Möglichkeit zur Entwicklung und Anwendung fachsprachlicher Lernhilfen zur Unterstützung historischen Denkens im Fach Geschichte bieten, auf die weitere Forschungen aufbauen können und müssen. Dies betrifft selbstredend auch die Interpretation der Ergebnisse. So muss zusätzlich zur Annahme, dass die geringe Nutzung verschiedener Merkmale darauf zurückzuführen sein könnte, dass diese in den meisten Geschichtsschulbüchern – und damit voraussichtlich auch im Geschichtsunterricht generell – unterrepräsentiert sind, auch eine Überforderung durch die Anlage der Studie in Betracht gezogen werden, zum Beispiel durch die Anwendung von sieben Merkmalen des historischen Erklärens gleichzeitig. Es besteht also die Möglichkeit, dass Schüler*innen bei ihren Überarbeitungen vor allem auf Hilfen derjenigen Merkmale zurückgegriffen haben, die ihnen am ehesten aus anderen historischen Erklärungen aus dem Alltag bekannt waren und ihnen da-

her zum eigenen Verstehen des historischen Sachverhalts am hilfreichsten erschienen. In künftigen Studien könnte die Anzahl der durch integrative Scaffoldingtools fokussierten Aspekte daher reduziert und dafür auf dieser Ebene weiter ausdifferenziert werden, um einen tieferen Einblick in die konkreten sprachlichen Ausgestaltungen zu erhalten. Gleichzeitig sollte die Stichprobe vergrößert werden, um zu einer größeren Aussagekraft der generierten Daten zu gelangen. Auch könnte die Konstruktion eines zufälligen Samplings der Schulklassen und innerhalb dieser wiederum eine zufällige Verteilung der Interventionen sinnvoll sein, um Gruppeneffekte zu minimieren.

Darüber hinaus und auch jenseits des explorativen Charakters der Studie fehlt bislang die fachliche Validierung der Drehbücher hinsichtlich einer Qualitätssteigerung. Denn die Überarbeitung an sich – also die pure Nutzung von Scaffolds – bringt nicht zwangsläufig auch die gesteigerte fachliche Qualität einer historischen Erklärung mit sich. So haben bereits frühere Forschungen gezeigt, dass die Schüler*innen zwar mit Worthilfen arbeiten, diese jedoch häufig an «unsinnigen» Stellen im Text einsetzen (Stoel et al., 2017), sodass sich eine Qualitätssteigerung des historischen Denkens als Folge des Einsatzes von Wortbeispielen nicht zwangsläufig einstellen wird. Auch im vorliegenden Sample ließen sich bei einer Durchschau vereinzelt solche Fälle feststellen, sodass als nächster Schritt die qualitative Bewertung der überarbeiteten Drehbücher im Verhältnis zu den verwendeten Scaffolds angestrebt wird. Die Annäherung an die Frage, welche Hilfen wirklich helfen könn(t)en, das historische Denken von Schüler*innen zu fördern, bleibt daher auch für zukünftige Studien ein lohnenswertes Forschungsfeld, wobei gerade hier auch interdisziplinäre Ansätze als Chance verstanden werden können, die eigene Fachsprachlichkeit in Abgrenzung zu anderen Zugängen theoretisch und konzeptionell zu schärfen.

Literatur

- Alavi, Sayyed Mohammad (2005). On the adequacy of verbal protocols in examining an underlying construct of a test. *Studies in Educational Evaluation*, 31(1), 1–26. <https://doi.org/10.1016/j.stueduc.2005.02.004>.
- Behrendt, Lena & Brauch, Nicola (2019). Auf die Frage kommt es an! Zur Veränderung epistemologischer Überzeugungen durch Förderung historischer Fragestellungskompetenz im Schülerlabor. In Katrin Sommer/Joachim Wirth & Marie Vanderbeke (Hrsg.), *Handbuch Forschen im Schülerlabor* (S. 91–97). Münster, New York: Waxmann.
- Berger, Stefan/Brauch, Nicola & Lorenz, Chris (2021). *Analysing historical narratives: On academic, popular and educational framings of the past*. New York: Berghahn Books.

- Bramann, Christoph (2020). (Fach)sprachbildende Schulbuchaufgaben im Kontext einer Aufgabenkultur historischen Lernens erforschen. In Thomas Sandkühler & Markus Bernhardt (Hrsg.), *Sprache(n) des Geschichtsunterrichts* (S. 165–187). Göttingen: Vandenhoeck & Ruprecht. <https://doi.org/10.14220/9783737012058.165>.
- Brauch, Nicola (2015). *Geschichtsdidaktik*. Berlin: De Gruyter. <http://dx.doi.org/10.1515/9783110375244>.
- Brauch, Nicola (2019a). «Fasse den Verfasserstext zusammen!». Aufgaben im Geschichtsschulbuch – eine explorative Einzelfallanalyse. In Christoph Kühberger/Roland Bernhard & Christoph Bramann (Hrsg.), *Das Geschichtsschulbuch. Lehren – Lernen – Forschen* (S. 185–205). Münster, New York: Waxmann.
- Brauch, Nicola (2019b). Vom Umgang mit der Zeit im historischen Erzählen: Auf dem Weg zu Übungsaufgaben. In Waltraud Schreiber/Béatrice Ziegler & Christoph Kühberger (Hrsg.), *Geschichtsdidaktischer Zwischenhalt* (S. 187–193). Münster, New York: Waxmann.
- Brauch, Nicola & Jahn, Sarah J. (2018). Religiöse Vielfalt im Wandel? Historische Gegenwartsbezüge in Arbeitsaufträgen zu den abrahamitischen Religionen in Geschichts- und Gesellschaftslehreschulbüchern. In Sylvia Schütze & Eva Matthes (Hrsg.), *Religion und Bildungsmedien* (S. 281–295). Bad Heilbrunn: Klinkhardt.
- Brauch, Nicola/Heine, Lena & Bramann, Christoph (2020). Schriftliches Erklären im Fach Geschichte unterstützen: Ansätze eines sprachlich-epistemologischen Scaffolding-Tools. In Thomas Sandkühler & Markus Bernhardt (Hrsg.), *Sprache(n) des Geschichtsunterrichts* (S. 137–164). Göttingen: Vandenhoeck & Ruprecht. <https://doi.org/10.14220/9783737012058.137>.
- Carretero, Mario & Lee, Peter J. (2014). Learning Historical Concepts. In Robert K. Sawyer (Hrsg.), *The Cambridge handbook of the learning sciences* (S. 587–593). Cambridge: Cambridge University Press.
- Chapman, Arthur (2017). Causal explanation. In Ian Davies (Hrsg.), *Debates in history teaching* (2. Aufl., S. 130–143). London: Routledge.
- Cohen, Jacob (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>.
- Danto, Arthur Coleman (1965). *Analytical philosophy of history*. Cambridge: Cambridge University Press.
- Drie, Jannet van & Boxtel, Carla van (2008). Historical reasoning: towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review*, 20(2), 87–110. <https://doi.org/10.1007/s10648-007-9056-1>.
- Gibbons, Pauline (2002). *Scaffolding language, scaffolding learning: teaching second language learners in the mainstream classroom*. Portsmouth: Heinemann.
- Günther-Arndt, Hilke (2014). Historisches Lernen und Wissenserwerb. In Hilke Günther-Arndt & Meik Zülsdorf-Kersting (Hrsg.), *Geschichts-Didaktik. Praxishandbuch für die Sekundarstufe I und II* (6. Aufl.; S. 24–49). Berlin: Cornelsen.
- Handro, Saskia & Kilimann, Vanessa (2019). Textverstehen im Geschichtsunterricht: Ein Projekt zur Professionalisierung historischer Leseförderung (ProLeGu). In Marion Bönnighausen (Hrsg.), *Praxisprojekte in Kooperationsschulen: Fachdidaktische Modellie-*

- rung von Lehrkonzepten zur Förderung strategiebasierten Textverstehens in den Fächern Deutsch, Geographie, Geschichte und Mathematik* (S. 167–222). Münster: WTM.
- Kinder, Hermann/Hergt, Manfred & Hilgemann, Werner (2015). *dtv-Atlas Weltgeschichte* (42. Aufl.). München: dtv.
- Kuckartz, Udo (2016). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung* (3. Aufl.). Weinheim: Beltz Juventa.
- Kühberger, Christoph (2012). Konzeptionelles Wissen als besondere Grundlage für das historische Lernen. In Christoph Kühberger (Hrsg.), *Historisches Wissen. Geschichtsdidaktische Erkundungen über Art, Umfang und Tiefe für das historische Lernen* (S. 33–74). Schwalbach/Ts.: Wochenschau.
- Landis, J. Richard & Koch, Gary G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>.
- Lee, Peter & Shemilt, Denis (2009). Is any explanation better than none? *Teaching History*, 137, 42–49.
- Leisen, Josef (2010). *Handbuch Sprachförderung im Fach – Sprachsensibler Fachunterricht in der Praxis*. Bonn: Varus.
- Lorenz, Chris (1997). *Konstruktion der Vergangenheit: Eine Einführung in die Geschichtstheorie*. Köln: Böhlau.
- Mayring, Philipp (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (11. Aufl.). Weinheim: Beltz.
- Mierwald, Marcel (2020). Historisches Argumentieren und epistemologische Überzeugungen: Eine Interventionsstudie zur Wirkung von Lernmaterialien im Schülerlabor. Wiesbaden: Springer.
- Mierwald, Marcel & Brauch, Nicola (2015). Historisches Argumentieren als Ausdruck historischen Denkens. *Zeitschrift für Geschichtsdidaktik*, 14E(1), 104–120. <https://doi.org/10.13109/zfgd.2015.14e.1.104>.
- Mierwald, Marcel & Brauch, Nicola (2019). «Ein Argument für diese Behauptung lautet ...». Beschreibung, Graduierung und Bewertung der strukturellen Qualität argumentativer historischer Essays. In Christiane Bertram & Andrea Kolpatzik (Hrsg.), *Sprachsensibler Geschichtsunterricht: Von der Theorie über die Empirie zur Pragmatik* (S. 108–114). Frankfurt a. M.: Wochenschau.
- Pandel, Hans-Jürgen (2017). *Geschichtsdidaktik: Eine Theorie für die Praxis* (2. Aufl.). Schwalbach/Ts.: Wochenschau.
- Pett, Marjorie A. (2015). *Nonparametric statistics in health care research: Statistics for small samples and unusual distributions* (2. Aufl.). Los Angeles: SAGE.
- Rösen, Jörn (2013). *Historik: Theorie der Geschichtswissenschaft*. Köln: Böhlau.
- Schrader, Viola (2013). *Geschichte als narrative Konstruktion: Eine funktional-linguistische Analyse von Darstellungstexten in Geschichtsschulbüchern*. Berlin: Lit.
- Schreiber, Waltraud (2007). Kompetenzbereich historische Methodenkompetenz. In Andreas Körber/Waltraud Schreiber & Alexander Schöner (Hrsg.), *Kompetenzen historischen Denkens: Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik* (S. 194–235). Neuried: ars una.

- Schreier, Margrit (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten. *Varianten qualitativer Inhaltsanalyse*, 15(1). <https://doi.org/10.17169/FQS-15.1.2043>.
- Seixas, Peter & Morton, Tom (2013). *The big six: Historical thinking concepts*. Toronto: Nelson Education.
- Stoel, Gerhard L. (2017). *Teaching towards historical expertise. Developing students' ability to reason causally in history*. Dissertation. University of Amsterdam, Amsterdam.
- Stoel, Gerhard L./van Drie, Jannet & van Boxtel, Carla (2015). Teaching towards historical expertise. Developing a pedagogy for fostering causal reasoning in history. *Journal of Curriculum Studies*, 47(1), 49–76. <https://doi.org/10.1080/00220272.2014.968212>.
- Stoel, Gerhard L./van Drie, Jannet & van Boxtel, Carla (2017). The effects of explicit teaching of strategies, second-order concepts, and epistemological underpinnings on students' ability to reason causally in history. *Journal of Educational Psychology*, 109(3), 321–337. <https://doi.org/10.1037/edu0000143>.
- Süßmann, Johannes (2019). Erzählung. In Stefan Jordan (Hrsg.), *Lexikon Geschichtswissenschaft: Hundert Grundbegriffe* (S. 85–88). Stuttgart: Reclam. (Original erschienen 2002).
- Trautwein, Ulrich et al. (2017). *Kompetenzen historischen Denkens erfassen: Konzeption, Operationalisierung und Befunde des Projekts «Historical Thinking – Competencies in History» (HiTCH)*. Münster: Waxmann.
- Vollmer, Helmut J. & Thürmann, Eike (2010). Zur Sprachlichkeit des Fachlernens: Modellierung eines Referenzrahmens für Deutsch als Zweitsprache. In Bernt Ahrenholz (Hrsg.), *Fachunterricht und Deutsch als Zweitsprache* (S. 107–132). Tübingen: Narr.
- Weber, Max (1921). *Wirtschaft und Gesellschaft: Die Wirtschaft und die gesellschaftlichen Ordnungen und Mächte*. Tübingen: Mohr.
- Welskopp, Thomas (2019). Erklären. In Stefan Jordan (Hrsg.), *Lexikon Geschichtswissenschaft: Hundert Grundbegriffe* (S. 81–84). Stuttgart: Reclam. (Original erschienen 2002).
- Wirtz, Markus Antonius & Caspar, Franz (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.
- Woodcock, James (2005). Does the linguistic release the conceptual? Helping year 10 to improve their causal reasoning. *Teaching History*, (119), 5–14.

Narrative Kompetenz *und* epistemologische Beliefs im Geschichtsunterricht fördern?

Zur Wirksamkeit einer Schreibintervention an Deutschschweizer Gymnasien

*Martin Nitsche, Monika Waldis, Kristine Gollin, Philipp Marti und
Manuel S. Hubacher*

1 Ausgangssituation

Es besteht ein breiter Konsens darin, dass sich Schüler*innen im Geschichtsunterricht mit Quellen und Darstellungen auseinandersetzen und davon ausgehend Geschichte(n) entwickeln sollen, um historisch Denken zu lernen (vgl. Körber, 2022). Seit etwa zehn Jahren wird zudem betont, die damit verbundenen Lernprozesse vollzögen sich sprachgebunden etwa während des Schreibens (z. B. Handro, 2013; Schrader, 2019). Auch wenn die Einbeziehung von mündlichen Formen für Lernende mit diversen Voraussetzungen ebenfalls fruchtbar sein dürfte (Barsch, 2020), ist es daher naheliegend, das historische Denken von Schüler*innen mittels sprachbasierter Schreibförderung anzubahnen. Diese gilt im Rahmen des *writing-to-learn*-Ansatzes als wirkungsvolles Hilfsmittel zur Förderung des allgemeinen Schreibens und Lernens sowie des historischen Denkens und Wissenserwerbs von Schüler*innen auf diversen Schulstufen (z. B. van Drie, van Boxtel & Braaksma, 2014).

Dabei wurde ebenfalls die epistemische Funktion des Schreibens für historisches Denken thematisiert (Nokes & De La Paz, 2018). Qualitative Befunde weisen darauf hin, dass epistemische Überlegungen (*epistemic cognition*) und die Auseinandersetzung mit Quellen und Darstellungen zusammenhängen (vgl. VanSledright & Maggioni, 2016). Zudem verdeutlichen quantitative Studien, dass *epistemologische Beliefs* von Schüler*innen mit ihren historischen Schreibleistungen verbunden sind, deren Förderung wiederum die Ausbildung *wissenschaftsförmiger Beliefs* zu unterstützen scheint (z. B. Mierwald, 2020).

Deskriptive Studien, in denen untersucht wurde, wie Lernende mit Quellen und Darstellungen umgehen (z. B. Wolfrum & Sauer, 2007), welche epistemologischen Konzepte dabei involviert sind (z. B. Martens, 2015) und wie die genannten Aktivitäten im Geschichtsunterricht adressiert werden (z. B. Marti, Gollin & Waldis, 2019; Spieß, 2014) verdeutlichen, dass geschichtsdidaktische Ziele selten erreicht werden. Zudem scheint die Wirksamkeit von Fördermaßnahmen mit zahlreichen Aspekten verbunden, da etwa die damit verbundenen Aufgaben (z. B. Hartung, 2013), das Thema (Hodel, Waldis, Zülsdorf-Kersting & Thünemann, 2013), eingeforderte Textgenres (Waldis, 2016) und die Struktur der historischen Medien (Mierwald, 2020) als Einflussfaktoren identifiziert wurden. Vielleicht liegen deshalb kaum quasiexperimentelle deutschsprachige Untersuchungen zur Förderung historischen Schreibens vor (z. B. Wickner, 2019). Weitere weisen explorative (z. B. van Norden, 2011) oder laborähnliche Merkmale auf (Mierwald, 2020). Für den regulären Geschichtsunterricht auf Sekundarstufe II fehlen Studien, in denen der Zusammenhang zwischen den epistemologischen Beliefs der Schüler*innen und ihrer narrativen Kompetenz sowie die Wirksamkeit von Schreibfördermaßnahmen im Hinblick auf beide Konstrukte untersucht wurden.

Um diese Lücke zu adressieren, stellen wir Analysen der genannten Aspekte anhand eines Teilsamples einer quasiexperimentellen Studie vor (Waldis, Nitsche & Gollin, 2020), in deren Rahmen die Wirksamkeit einer Schreibintervention untersucht wurde, die von Geschichtslehrpersonen im Schuljahr 2017–2018 in ihrem Geschichtsunterricht an Deutschschweizer Gymnasien auf den Klassenstufen zehn und elf im Umfang von acht Doppelkationen (90 min) mittels argumentativer Schreibaufgaben umgesetzt wurde, um die narrative Kompetenz der beteiligten Schüler*innen zu fördern. Wir fragen *erstens*, inwiefern ein Zusammenhang zwischen narrativer Kompetenz und epistemologischen Beliefs vorliegt. Wir klären *zweitens*, inwiefern die Vorgehensweisen in der Versuchsgruppe die narrative Kompetenz der beteiligten Schüler*innen förderte und *drittens*, ob sich die epistemologischen Beliefs der Schüler*innen in der Versuchsgruppe veränderten.

Zunächst stellen wir einen Zusammenhang zwischen historischen Argumentationen und narrativer Kompetenz her, um die Verbindung zwischen diesem Genre, Narrativität und narrativer Kompetenz zu verdeutlichen. Anschließend setzen wir das Konstrukt zu epistemologischen Beliefs sowie zu geeigneten Fördermaßnahmen in Beziehung und stellen unsere Studie vor.

2 Narrative Kompetenz und Schreibförderung

2.1 Historische Argumentationen und narrative Kompetenz

Auf der Grundlage einer narrativen Geschichtstheorie wird Geschichte oft eher mit Erzählen als Argumentieren assoziiert. Geschichte sei von einer narrativen Struktur gekennzeichnet, deren wesentliches Distinktionsmerkmal darin bestehe, die Differenz von Aspekten zu zwei Zeitpunkten zu überbrücken (z. B. Pandel, 2010, 39; Rüsen, 2013, 44). Diese Klassifizierung sorgte geschichtstheoretisch und -didaktisch für Kontroversen, da etwa infrage stand, inwiefern das narrative Paradigma für die Beschreibung von Strukturen gilt (z. B. Daniel, 2006) und nicht explizit erzählende Aufgaben historisches Erzählen fördern (Barricelli, 2016).

Demgegenüber lassen sich alle Aussagen («narrative Sätze») und Formen (z. B. Beschreibungen, Argumentationen) sowie darauf bezogene Aufgaben dem narrativen Paradigma zuordnen, wenn sie aufgrund der verbalen Zeitformen oder geschichtlichen Begriffe auf Vergangenes und Geschichte(n) verweisen, da sie die Narrativität des dahinterliegenden Denkens und Wissens voraussetzen. Um formal als historische Narrationen zu gelten, ist indes ein Anfang und Schluss nötig (Ricoeur, 2007, 217–246). Wissenschaftliche Formen der Geschichtsschreibung sollten zudem Qualitätskriterien (z. B. empirische Plausibilität) genügen (Rüsen, 2013). Folglich lassen sich jene Aussagen als narrative und damit historische auffassen, die erkennbar auf zeitliche Aspekte menschlichen Lebens verweisen. Als wissenschaftsorientierte historische Narration können Formen gelten, die historische Aussagen enthalten, einen Anfang und Schluss aufweisen sowie überprüfbar sind.

Geschichtstheoretisch und sprachwissenschaftlich fehlen eindeutige Klassifikationsschemata für Textformen (Habscheid, 2011; Kansteiner, 2018). Geschichtsdidaktisch bietet die Unterscheidung von Coffin (2006, 28) eine Möglichkeit, Genres strukturell zu differenzieren. Die Autorin definiert Genres als Resultat eines zielorientierten Konstruktionsprozesses, der das Ergebnis des historischen Denkens von Personen und ihre Absicht verdeutlichen soll (Coffin, 2006, 10). Sie unterscheidet Berichte (*recording genre*)¹ von Erklärungen (*explaining genre*) und Argumentationen (*arguing genre*) und differenziert Letztere in drei Unterformen aus (Exposition, Diskussion, Herausforderung).

1 Hier werden historische Erzählungen (*historical accounts*) im engeren Sinn (Wandel in der Zeit) subsumiert (Coffin, 2006, 64).

Die vorliegende Interventionsstudie adressierte die «Diskussion», da diese aufgrund ihres abwägenden Charakters eine Passung zur geschichtsdidaktischen Kompetenzdiskussion aufweist, in deren Rahmen das Abwägen von geschichtlichen Perspektiven als wichtiger Teilaspekt historischer Kompetenz angesehen wird (z. B. Schreiber et al., 2006). Dabei wurde anhand der Literatur eine idealtypische Struktur der «Diskussion» zugrunde gelegt (Coffin, 2006, 81–82; De La Paz et al., 2017, 24; Mierwald & Brauch, 2015, 116), die in Tabelle 1 (siehe S. 102) dargestellt und auf die Operationen narrativer Kompetenz bezogen ist.

Das Vermögen derartige und weitere Genres zu entwickeln, kann mithilfe des Konzepts *narrative Kompetenz* beschrieben werden. Deutschsprachige Geschichtsdidaktiker*innen bezeichnen damit entweder den synthetisierenden Anteil historischer Kompetenz (van Norden, 2011; Pandel, 2010) oder eine grundlegende, mentale Kapazität historischen Denkens, die ebenfalls analytische Aspekte umfasst (z. B. Hodelet al., 2013; Rüsen, 2013). Englischsprachige Autor*innen haben beide Anteile anhand empirischer Studien differenziert (z. B. Voss & Wiley, 1997; Wineburg, 1991) und etwa als *historical reasoning* (van Boxtel & van Drie, 2018) oder *understanding* systematisiert (VanSledright, 2014). Insgesamt stimmen sie darin überein, dass historische Fragen, der Umgang mit Quellen und Darstellungen, das Entwickeln von Geschichte(n) sowie Bezüge zur Lebenswelt der historisch Denkenden beim Umgang mit Geschichte(n) involviert sind. Zudem betonen einige Autor*innen die Relevanz historischen Wissens erster (z. B. Ereignisse) und zweiter (z. B. Dauer und Wandel) Ordnung (van Boxtel & van Drie, 2018; VanSledright, 2014).

Daran anknüpfend und in Abwandlung des Prozessmodells historischen Denkens (Schreiber et al., 2006) konzipieren wir (z. B. Nitsche & Gollin, 2020; Waldis et al., 2020) *narrative Kompetenz* als Ausdruck historischen Denkens (z. B. Rüsen, 2013), das als mentaler Interaktionsprozess zwischen involvierten Operationen beschreibbar und vom soziohistorischen Kontext gerahmt ist (z. B. VanSledright, 2014). Es dient der Lösung von Gegenwartsproblemen (z. B. gesellschaftliche Krisen, Identitätsverunsicherungen) mittels Zeitbezügen (Schreiber et al., 2006) oder von geschichtsspezifischen Aufgaben in Lehr-Lern-Kontexten (Nitsche, Gollin & Waldis, 2017). Die Operationen der narrativen Kompetenz lassen sich in vier Teilbereiche differenzieren:

1. Gegenwartsprobleme oder Aufgaben werden in Erkenntnisinteressen überführt, indem eigene historische Fragen oder Hypothesen formuliert oder fremde erkannt werden (z. B. Schreiber et al., 2006; VanSledright, 2014).

2. Diese werden zwecks Dekonstruktion auf historische Medien (Quellen und Darstellungen) bezogen, um Aussagen über Vergangenes zu identifizieren und formal zu vergleichen (z. B. Voss & Wiley, 1997).
3. Während der Rekonstruktion werden Aussagen hinsichtlich ihrer historischen Plausibilität mittels Medienkritik oder -vergleich geprüft und anhand des historischen (Vor-)Wissens kontextualisiert sowie in die Konstruktion von Geschichte(n) einbezogen (Schreiber et al., 2006; Wineburg, 1991). Dazu wird Zeitlichkeit konstruiert, indem die Aussagen chronologisch angeordnet werden (z. B. Pandel, 2010). Historische Bedeutung wird mittels Klassifikation der Aussagen etwa als Dauer und Wandel sowie Ursache und Wirkung (van Boxtel & van Drie, 2018; VanSledright, 2014) oder mithilfe von geschichtsbezogenen Begriffen konstituiert (Schreiber et al., 2006). Zur Strukturierung von Geschichte(n) erfolgt zum Beispiel die Entwicklung und die Anordnung von Argumenten (Coffin, 2006; De La Paz et al., 2017).
4. Der Vorgang dient der Orientierung, wenn seine Resultate in die Begründung zukünftigen Handelns mittels Sinnbildungsmustern (Rüsen, 2013) oder in die Beurteilung der Gegenwartsrelevanz anhand gültiger Werte und Normen einfließen (Hodel et al., 2013).

Daneben sind sprachliche Fähigkeiten wie der Umgang mit grammatikalischen und lexikalischen Mitteln oder zielgruppenorientierten Registern (z. B. Handro, 2013), historisches Wissen erster (z. B. Ereignisse) und zweiter Ordnung (z. B. Dauer und Wandel) sowie epistemologische Beliefs involviert (van Boxtel & van Drie, 2018; VanSledright, 2014).

Um von historischen Argumentationen auf Operationen narrativer Kompetenz zu schließen, sind in Tabelle 1 die Bezüge zwischen beiden Aspekten zusammengefasst. Die Gegenüberstellung verdeutlicht, an welcher Stelle der Textgliederung spezifische Operationen narrativer Kompetenz hauptsächlich dienlich sein dürften. Sie hat nicht die Funktion, eine Schrittfolge der Textentwicklung zu lancieren.

Tabelle 1: Historische Argumentation («Diskussion») und narrative Kompetenz

Textteil und Aspekte	Operationen narrativer Kompetenz
<i>Ganzer Text</i>	Argumentation strukturieren;
<i>Einleitung</i> Historische Frage(n) und Hypothese(n); Aussageabsicht;	Fragen, Hypothesen formulieren; Deutungs-, Sinnbildungsmuster, geschichtliche Begriffe, Wertvorstellungen einbeziehen; Kontextualisieren
Einbettung in historischen Kontext	
<i>Hauptteil</i> Unterstützende Argumente bestehend aus histori- scher Aussage; Belege(n) mittels Medien- oder Kontextbezügen; Erläuterung der Belege; Gegenargument(e) zusammengesetzt wie unter- stützende Argumente; Entkräftigung unter Bezug auf Glaubwürdigkeit, vergangene, gegenwärtige oder zukünftige Bedeu- tung	Historische Argumente entwickeln; Medienbezüge herstellen; Kontextualisieren; Medienkritik nutzen; Medienkritik nutzen; Kontextualisieren; Deutungs-, Sinnbildungsmuster oder Wertvorstellungen einbe- ziehen
<i>Schluss</i> Zusammenfassung der Positionierung hinsichtlich Frage(n) bzw. Hypothese(n); Begründung für die Lesenden	Historische Bedeutung; Orientierung, Wertung for- mulieren

2.2 Epistemologische Beliefs und narrative Kompetenz

Epistemologische Beliefs lassen sich als subjektive mentale Konstrukte definieren, die auf (historische) Erkenntnismöglichkeiten und -weisen ausgerichtet sind. Sie werden etwa hinsichtlich ihrer Entwicklungslogik, Dimensionalität, Kontext- und Domänenspezifität sowie bezüglich ihrer kognitiven Funktion untersucht (Nitsche, 2019).

Geschichtsspezifisch haben Maggioni et al. Entwicklungstypen unterschieden. *Copier* assoziierten Geschichte mit der Abbildung der Vergangenheit. *Borrower* fassten Geschichte als Meinungssache auf, während *Criticalists* die Begründbarkeit von Geschichte anhand von Kontextbezügen (z. B. durch Quellen) voraussetzten. Die Autor*innen gehen indes nicht von einer linearen Entwicklung aus, da das Konstrukt zu komplex sei (VanSledright & Maggioni, 2016). Daher wählten Stoel, Logtenberg et al. (2017) einen dimensionalen Ansatz und differenzierten epistemologische Beliefs zur *Sicherheit des historischen Wissens* in naive (z. B. Geschichtswissen lässt sich nicht absichern) und nuancierte Beliefs

(z. B. Geschichtswissen kann kontextuell begründet werden). Nitsche (2019) verknüpfte den entwicklungslogischen mit dem dimensional Ansatz, indem er geschichtstheoretische Beliefs in drei Perspektiven («Positivismus», «Skeptizismus», «narrativer Konstruktivismus») zu je sechs Dimensionen (z.B. «Geschichtsbegriff») ausdifferenzierte.

Geschichtsdidaktisch wird seit längerem angenommen, dass die Einsicht in epistemologische Prinzipien (z. B. Konstruktcharakter von Geschichte) mit der narrativen Kompetenz verbunden ist (z. B. Pandel, 2010). Empirisch wurden Verbindungen zwischen epistemologischen Konzepten und analytischen Aspekten historischen Denkens zunächst in qualitativen Arbeiten aufgezeigt (z. B. Martens, 2015; Wineburg, 1991). Inzwischen liegen erste quantitative Studien für den Zusammenhang zwischen epistemologischen Beliefs und historischem Denken am Beispiel der Auseinandersetzung mit historischen Medien (Iordanou, Kendou & Zembylas, 2020) sowie der Verschriftlichung ihrer Ergebnisse vor (Wiley, Griffin, Steffens & Britt, 2020). Mierwald (2020, 192–193) weist zudem Korrelationen zwischen der Qualität historischer Argumentationen von deutschen Oberstufenschüler*innen und ihren elaborierten Beliefs (*criteria*list) in seiner Lernlaborstudie nach. Insgesamt verdichten sich somit die Hinweise für die Verbindungen zwischen epistemologischen Beliefs und historischem Denken, so dass das Konstrukt im Rahmen der vorliegenden Studie in den Blick geriet.

2.3 Geeignete Interventionsmaßnahmen

Anhand von im Jahr 2017 bereits vorliegenden Interventionsstudien, die auf die Förderung analytischer und synthetisierender Aspekte historischen Denkens abzielten, identifizierten wir geeignete Fördermaßnahmen. Mit Blick auf analytische Aspekte historischen Denkens verglichen Britt et al. (2000) in ihrer Studie, an der US-amerikanischen Elftklässler*innen aus vier Klassen teilnahmen, die Wirksamkeit eines inhaltsbasierten Vorgehens mit einem dreitägigen, forschungsbasierten Training, das am *cognitive apprenticeship* orientiert war (Collins, Brown & Newman, 1989). Die Autor*innen verdeutlichen, dass das forschungsbasierte Vorgehen, bei dem der Umgang mit Quellen angelehnt am Agieren von Historiker*innen explizit modelliert, mittels begleitender Materialien (z.B. Tutorials über Schritte des Vorgehens) unterstützt und auf unterschiedliche Themen transferiert wurde, wirksamer ausfiel als die Inhaltsvermittlung in der Kontrollgruppe. Die Studie von Nokes et al. (2007), unterstreicht die Wirksamkeit des Vorzeigens von analytischen Aspekten historischen Den-

kens an einem größeren Sample. Reismans (2012) Langzeitintervention mit US-amerikanischen Elftklässler*innen illustriert zudem, dass es über das Modellieren hinaus auf explizite historische Fragen, vielfältige Quellen und Darstellungen sowie auf den Austausch zwischen den Schüler*innen ankommt.

Um synthetisierende Aspekte historischen Denkens zu fördern, stellten De La Paz und Felton (2010) eine mehrwöchige Interventionsstudie mit US-amerikanischen Elftklässler*innen verteilt auf zehn Unterrichtsimpulse vor, in der die Wirksamkeit eines expliziten Lese- und (argumentativen) Schreibtrainings im Vergleich zu einem impliziten eher auf Inhalte ausgerichteten Vorgehen getestet wurde. Dabei zeigte sich, dass das explizite Training, in dessen Rahmen nicht nur analytische Aspekte historischen Denkens (z. B. Quellenkritik) vorgezeigt, sondern auch historische Beispielargumentationen untersucht, der Textaufbau mittels Gliederungshilfen unterstützt, die Möglichkeit für Peer-Feedback zwischen den Schüler*innen eingeräumt und das selbstständige Üben mittels Transfer des Gelernten auf andere Themen eingefordert wurde, zur Verbesserung der allgemeinen und historischen Textqualität (z. B. Perspektivität, Glaubwürdigkeit der Quellen) beitrug. Eine achtzehntägige Intervention, in der außerdem Formulierungshilfen für das Verfassen historischer Argumentationen eingesetzt wurden, bestätigte das Resultat an einem Sample mit US-amerikanischen Achtklässler*innen (De La Paz et al., 2017). Im deutschsprachigen Kontext findet sich zum Untersuchungszeitpunkt lediglich bei van Norden und Kolleg*innen ein vergleichbares Vorgehen (z. B. Ausrichtung auf historische Fragen, Zeigen fachlicher Methoden) im Rahmen von explorativen Studien, die unter anderem die narrative Kompetenz von Schüler*innen adressierten. Allerdings fehlt die inferenzstatistische Absicherung (z. B. van Norden & Schürenberg, 2019).

Zur Förderung historischen Denkens und epistemologischer Beliefs existierten zu Beginn unserer Untersuchung lediglich zwei Interventionsstudien. So erprobten Stoel et al. (z. B. 2015) mit niederländischen Schüler*innen aus elften Klassen die explizite Förderung von schriftlichem *causal reasoning* und epistemologischen Beliefs und verglichen es mit einem impliziten Unterrichtsvorgehen. In der Hauptstudie von 2017 orientierte sich ihr explizites Vorgehen unter anderem am Cognitive-Apprenticeship-Ansatz und war auf drei Lektionen beschränkt: Nach einer zweistündigen inhaltlichen Einführung in beiden Gruppen modellierten die Lehrpersonen in der expliziten Gruppe zunächst Strategien zur Konstruktion von Erklärungen und verdeutlichten den Lernenden anhand eines Modelltextes, wie multikausale Erklärungen strukturiert werden können. Anschließend wurden die Lernenden aufgefordert, das Gelernte auf ein histori-

sches Thema anzuwenden, bevor sie zum Austausch der Ergebnisse und zur Reflexion ihrer damit verbundenen epistemologischen Beliefs aufgefordert wurden. Dabei zeigte sich, dass es beiden Gruppen gleich gut gelang, historische Erklärungen zu schreiben, jedoch ließ sich lediglich in der expliziten Kondition eine Veränderung der epistemologischen Beliefs, allerdings sowohl der subjektivistischen (analog zu *borrower*) als auch der reflektierten (*criticalist*), aufzeigen. Die Autor*innen vermuten einerseits, dass ihre explizite Intervention zu kurz ausfiel, um *causal reasoning* bedeutsamer zu fördern, und andererseits, dass sich epistemologische Beliefs eher situativ in einem Kontinuum bewegen, statt sich linear zu entwickeln (Stoel, van Drie & van Boxtel, 2017).

Die Mehrzahl der dargestellten Studien nutzte, angelehnt an den Cognitive-Apprenticeship-Ansatz, ein Unterrichtsvorgehen, das in der Abfolge aus Zeigen, Begleiten und selbstständigem Anwenden besteht. Insgesamt lassen sich acht Designprinzipien wirksamer Interventionen identifizieren (z. B. Beispieltex-te untersuchen), die zur Förderung der narrativen Kompetenz und der epistemologischen Beliefs geeignet erscheinen und die in die Konzeption der Intervention eingeflossen sind (siehe Abschnitt 3.4). In Tabelle 2 ist der Ertrag der Literatursichtung systematisiert und auf die Operationen der narrativen Kompetenz sowie auf die epistemologischen Beliefs bezogen.

Tabelle 2: Aus der Literatur abgeleitete Designprinzipien und zu fördernde Zielkonstrukte

Literatur	Designprinzip	Zielkonstrukte
Alle ausgewerteten Studien: Britt et al. (2000), Nokes et al. (2007), van Norden (2011), Reisman (2012), De La Paz und Felton (2010), De La Paz et al. (2017), Stoel, van Drie et al. (2017);	Historische Probleme, Fragen und damit verbundene Aufgaben ins Zentrum stellen;	Fragen, Hypothesen formulieren;
Alle ausgewerteten Studien;	Operationen historischen Denkens modellieren;	Medienbezüge herstellen und vergleichen; Kontextualisieren; Medienkritik nutzen;
De La Paz und Felton (2010), De La Paz et al. (2017), Stoel, van Drie et al. (2017);	Historische Beispieltexthe untersuchen;	Argumentation strukturieren; historische Argumente strukturieren;
Reisman (2012), De La Paz und Felton (2010), De La Paz et al. (2017), Stoel, van Drie et al. (2017);	Austausch und Peer-Feedback bzgl. Ergebnissen und Vorgehensweisen ermöglichen;	Alle Operationen;
van Norden (2011), Reisman (2012), De La Paz et al. (2017);	Selbstständigen Transfer einfordern;	Alle Operationen;
Alle ausgewerteten Studien;	Multiperspektivische Medien anbieten;	Medienbezüge herstellen und vergleichen; Kontextualisieren; Medienkritik nutzen;
Reisman (2012), De La Paz und Felton (2010), De La Paz et al. (2017), Stoel, van Drie et al. (2017);	Unterstützende Materialien (z. B. Merkblätter) anbieten;	Medienbezüge herstellen und vergleichen; Kontextualisieren; Medienkritik nutzen; Argumentation strukturieren; Historische Argumente strukturieren;
Stoel, van Drie et al. (2017)	Epistemologische (Vor)Annahmen reflektieren	Epistemologische Beliefs

3 Die vorliegende Studie

3.1 Hypothesen

Die folgenden Hypothesen wurden für die vorliegende Studie getestet:

- (1) Es können statistisch signifikante Zusammenhänge zwischen der narrativen Kompetenz und den epistemologischen Beliefs der Schüler*innen aufgezeigt werden (H1).
- (2) Für die Versuchsgruppe lässt sich im Vergleich zur Kontrollgruppe ein statistisch signifikanter Zuwachs der narrativen Kompetenz nachweisen (H2).
- (3) Für die Versuchsgruppe lässt sich im Vergleich zur Kontrollgruppe eine statistisch signifikante Veränderung der epistemologischen Beliefs nachweisen (H3).

3.2 Studiendesign

Die Intervention wurde während des Schuljahr 2017–2018 an Deutschschweizer Gymnasien im Rahmen des Projektes «Schülerinnen und Schüler schreiben Geschichte» durchgeführt.² Sie war im Versuchs-Kontrollgruppen- sowie Prä-Post-Design strukturiert. Die beteiligten Lehrpersonen und ihre Klassen nahmen als Versuchs- oder Kontrollgruppe an der Studie teil. In allen Klassen erfolgten ein bis zwei Wochen vor (t₁) und nach (t₂) der Intervention Erhebungen der narrativen Kompetenz und der epistemologischen Beliefs. Die Personenmerkmale wurden zu t₁ erhoben. Nach t₁ führten die Lehrpersonen der Versuchsgruppe die schulische Intervention im Rahmen der regulären Geschichtslektionen in ihren Klassen durch, während der Unterricht der Kontrollgruppen nach den Planungen ihrer Lehrer*innen gestaltet war.

3.3 Teilnehmende

Insgesamt wurden 73 Schulen auf Gymnasialstufe per E-Mail kontaktiert. Sechzehn Lehrpersonen und ihre einundzwanzig Klassen der 10. und 11. Jahrgangsstufe erklärten sich zur Teilnahme bereit. Für die vorliegende Analyse wurde das Gesamtsample (Waldis et al., 2020) aus diversen Gründen (z. B. Schreibaufgabe

² Die Studie wurde vom Schweizer Nationalfond (SNF) finanziert (Fördernr.: 100019_160359/1).

zu t_2 in sechs Klassen nicht rotiert [siehe Abschnitt 3.5]) auf zehn Klassen reduziert. Um die Auswirkungen der Reduktion auf die Analysen zu überprüfen, wurden Vergleiche der Personenmerkmale der Schüler*innen der einbezogenen und ausgeschlossenen Klassen statistisch abgesichert. Die Ergebnisse weisen indes lediglich für das Geschlecht der Lernenden einen signifikanten Zusammenhang zur Gruppenzugehörigkeit (ein- vs. ausgeschlossene Klassen) auf. Allerdings fällt dieser eher gering aus ($\phi_c = -.23$), sodass aufgrund der Ausschlüsse keine Auswirkungen auf die Schätzungen der Interventionseffekte zu erwarten sind.

Das vorliegende Sample besteht aus acht Lehrpersonen und zehn Klassen. Zwei Lehrpersonen beteiligten sich jeweils mit zwei Klassen an der Interventions- oder Kontrollgruppe, während alle anderen Lehrpersonen mit je einer Klasse teilnahmen. Insgesamt wurden sechs Klassen der Interventions- und vier Klassen der Kontrollgruppe in die Analysen einbezogen. Ein Vergleich der Personenmerkmale der Schüler*innen der Versuchs- und Kontrollgruppe zu t_1 weist für den sozioökonomischen Status der Eltern (SES), die letzte Zeugnisnote in Deutsch und die skeptischen epistemologischen Beliefs signifikante Unterschiede auf, weshalb die Personenmerkmale in den Analysen berücksichtigt wurden (siehe Abschnitt 3.6). Weitere Sampleinformationen sind in Tabelle 3 zu finden. Die Erhebung der Merkmale ist in Abschnitt 3.5 ausführlich beschrieben.

Tabelle 3: Personenmerkmale und Unterschiede zwischen Versuchs- und Kontrollgruppe sowie Ausschluss (t_1)

Variable	Sample ($n = 173$)		Versuchsgruppe ($n = 100$)		Kontrollgruppe ($n = 73$)		Ausschluss ($n = 240$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Geschlecht (m)	26.04 %***		27.53 %		25.53 %		50.00 %***	
Klassenst. (10)	43.93 %		45.00 %		42.47 %		52.92 %	
Alter	16.76	1.11	16.68	0.97	16.86	1.27	16.56	1.28
SES	-0.02	0.64	0.12**	0.64	-0.22**	0.59	-0.04	0.96
Geschichts- note	4.65	0.61	4.58	0.66	4.76	0.52	4.73	0.57
Deutsch- note	4.73	0.44	4.63**	0.46	4.86**	0.37	4.71	0.52

Variable	Sample (n = 173)		Versuchsgruppe (n = 100)		Kontrollgruppe (n = 73)		Ausschluss (n = 240)	
	M	SD	M	SD	M	SD	M	SD
Narrative Kompetenz	4.47	2.53	4.74	2.58	4.08	2.43	a	a
Positivismus	-0.01	0.54	0.03	0.56	-0.06	0.53	a	a
Skeptizismus	-0.01	0.72	-0.20***	0.68	0.29***	0.67	a	a

Anm. Geschlecht, Klassenstufe = χ^2 -Test; alle anderen Variablen = *t*-Test; *n* = Fälle; *M* = Mittelwert; *SD* = Standardabweichung; m = männlich; ^a = aufgrund fehlender Testheftrotation zu t2 nicht einbezogen; **p* < .05; ***p* < .01; ****p* < .001.

3.4 Die Schreibintervention

Der Interventionsablauf und die zugehörigen Lernaufgaben samt Materialien wurden von uns auf der Grundlage der dargestellten Literatur (siehe Kapitel 2) entwickelt und von drei Geschichtslehrkräften hinsichtlich der schulischen Eignung geprüft. Nach Rücksprache mit den drei Lehrer*innen legten wir den Umfang auf acht Doppellektion (90 min) fest. Als Themen konnten die Lehrpersonen zwischen «Migration» und «Nationalismus» wählen. Dieses Angebot sollte den Lehrpersonen eine Wahlmöglichkeit geben und ihre Teilnahmebereitschaft erhöhen.

Der Unterrichtsablauf war für beide Themen auf der Grundlage des Cognitive-Apprenticeship-Ansatzes und anhand der aus der Literatur abgeleiteten Interventionsprinzipien strukturiert (siehe Abschnitt 2.3). Drei historische Fragen sowie die zugehörigen historischen Medien sorgten für geschichtsspezifische Schwerpunktsetzungen. Das Aufwerfen einer neuen historischen Frage unter Einbeziehung jeweils neuer Medien nach den Lektionen drei und sechs zielte auf den Transfer des Gelernten. Die Lektionen eins bis drei waren hauptsächlich dem Modellieren und der Lernbegleitung verpflichtet. Die Lektionen vier bis sechs sollten mittels Lernbegleitung und (Peer-)Feedback die Textüberarbeitung anstoßen, während die beiden Abschlusslektionen die selbstständige Entwicklung und Überarbeitung einer historischen Argumentation adressierten. Diese wurde von den Lehrpersonen benotet und die Schüler*innen erhielten ein Feedback. Auf die Reflexion epistemologischer (Vor-)Annahmen wurde aus Zeitgründen verzichtet.

Die Lernaufgaben samt den zugehörigen Materialien konnten von den Beteiligten über die Projektwebseite writinghistory.ch heruntergeladen werden. Die Aufgabenbearbeitung erfolgte am PC. In Tabelle 4 ist der Interventionsablauf zusammengefasst.

Tabelle 4: Interventionsdesign zu den Themen Migration (1.) und Nationalismus (2.)

Lektion (je 90 min) und historische Fragen zum 1. und 2. Thema	Materialien	Designprinzip und Operationen narrativer Kompetenz
<p>Eins:</p> <p>1. <i>Migration – Chance oder Gefahr?</i> 2. <i>Nationalismus – Chance oder Gefahr?</i></p>	<p>Videoclip «Medienkritik»; Merkblätter «Materialien untersuchen und vergleichen», «Analyseraster»; Auszüge aus Darstellungen zu 1. (Huntington, 2001; Münkler, 2015) oder 2. (Friedman, 2017; Roß, 2015)</p>	<p>Modellieren; Lernbegleitung; <i>Medienbezüge herstellen und vergleichen;</i> <i>Medienkritik nutzen;</i> Kontextualisieren</p>
<p>Zwei:</p> <p>1. und 2. (siehe Lektion eins)</p>	<p>Merkblätter «Analyseraster», «Eine historische Argumentation aufbauen»; Auszüge aus Darstellungen zu 1. (Conrad, 2016; Olmert, 2016) oder 2. (Günther-Arndt, 1995; Schümer, 2015; Thuri, 2011)</p>	<p>Lernbegleitung; <i>Historische Argumente entwickeln;</i> <i>Argumentation strukturieren</i></p>
<p>Drei:</p> <p>1. und 2. (siehe Lektion eins)</p>	<p>Modelltext einer historischen Argumentation; Merkblatt «Gliederung für meine Argumentation»; Medien aus Lektion zwei, drei</p>	<p>Modellieren; Lernbegleitung; <i>Operationen siehe Lektion zwei</i></p>
<p>Vier:</p> <p>1. <i>Auswanderung im 19. Jahrhundert – Aufbruch oder Flucht?</i> 2. <i>Der frühe Nationalismus – Freiheit für die Menschen oder Selbstbestimmung der Nation?</i></p>	<p>Merkblätter aus Lektion eins bis drei; Auszüge aus Darstellungen zu 1. (Olmert, 2016, 11–12; Ziegler, 2017) oder 2. (Westermann, 2010, 21); Auszüge aus zwei Quellen zu 1. (Wessendorf, 1973, 319–320; «Bericht, die Kolonisation in Brasilien betreffend», 1853) oder 2. (Arndt, 2008; Pfizer, 1995)</p>	<p>Transfer auf neue Frage; Lernbegleitung; <i>Alle Operationen</i></p>

Lektion (je 90 min) und <i>historische Fragen</i> zum 1. und 2. Thema	Materialien	Designprinzip und <i>Operationen narrativer Kompetenz</i>
Fünf: 1. und 2. (siehe Lektion vier)	Merkblätter aus Lektion eins bis drei; Medien aus Lektion vier; historische Argumentationen der Schüler*innen	Lernbegleitung mittels (Peer-) Feedbacks; <i>Alle Operationen</i>
Sechs: 1. und 2. (siehe Lektion vier)	Merkblätter aus Lektion eins bis drei; «Hilfreiche Formulierungen für meine historische Argumentation»; Medien aus Lektion vier; historische Argumentationen der Schüler*innen	Lernbegleitung mittels Überarbeitung (v. a. hinsichtlich Kohärenz); <i>Alle Operationen</i>
Sieben: 1. <i>Die Schweizer Überfremdungsdebatte: Ausdruck von Fremdenfeindlichkeit oder von gesellschaftlicher Verunsicherung?</i> 2. <i>Die Entwicklung des Schweizer Nationalstaats – Wunschprojekt oder Machtspiel?</i>	Merkblätter aus Lektion eins bis sechs; Auszug aus Darstellungen zu 1. (Skenderovic, 2015) oder 2. (Notz, 2010, 116; Meyer, 2002, 53); Auszüge aus Quellen zu 1. (Braun, 1970, 395; Schwarzenbach, 1974, 66–67, 77) oder 2. (Bitzius, 2001; Pfyffer, 2001; von Tavel, 2001)	Transfer auf neue Frage; Selbstständige Entwicklung einer historischen Argumentation; <i>Alle Operationen</i>
Acht: 1. und 2. (siehe Lektion sieben)	Merkblätter aus Lektion eins bis sechs; Argumentation aus Lektion sieben; Medien aus Lektion sieben	Selbstständige Überarbeitung; einer historischen Argumentation; <i>Alle Operationen</i>

Die Geschichtslehrer*innen der Versuchsgruppe wurden im Rahmen von je zwei sechsstündigen Workshops mit der Intervention vertraut gemacht. Der erste Workshop fand vor Interventionsbeginn statt. Er diente dazu, die Lehrpersonen in die Zielsetzung anhand unseres Modells narrativer Kompetenz, in den Interventionsablauf, die zugehörigen Materialien und die Projektwebseite einzuführen. Zudem wurden organisatorische Fragen geklärt. Das zweite Treffen fand etwa in der Mitte der Intervention statt und diente dazu, mit den Lehrpersonen Beurteilungs- und Feedbackstrategien anhand von Beispieltexten aus der Intervention zu besprechen.

3.5 Erhebungsmethoden

Die narrative Kompetenz der Schüler*innen wurde auf der Grundlage eigener Vorarbeiten (z. B. Marti & Waldis, 2017; Nitsche & Waldis, 2016; Waldis, Marti & Nitsche, 2015) mittels zwei historischer Argumentationsaufgaben erhoben, die zuvor in Cognitive Labs erprobt worden waren (z. B. Nitsche & Gollin, 2020). Eine Aufgabe adressierte die «Militärische Neutralität» und eine die «Wirtschaftliche Neutralität» der Schweiz im Ersten Weltkrieg. Die Aufgaben unterschieden sich voneinander sowie von den Interventionsthemen, um dem Kompetenzparadigma gerecht zu werden, da dieses Leistungsausprägungen in variablen Situationen fokussiert (Körber, 2022). Allerdings waren die Aufgaben vergleichbar konzipiert (Tabelle 5). Sie wurden zu t₁ in allen Klassen randomisiert auf die Hälfte der Schüler*innen verteilt und zu t₂ so rotiert, dass alle Schüler*innen beide Aufgaben entweder vor oder nach der Intervention bearbeiteten.

Tabelle 5: Aufgaben «Schweizer Neutralität im ersten Weltkrieg»

Auftrag & Thema	Materialien
«Lesen Sie die folgenden Texte (Quellen und Darstellungen). Bearbeiten Sie die Frage: Verhielt sich die Schweiz im 1. Weltkrieg neutral? Schreiben Sie dazu eine Behauptung auf (eine «These»), die Ihnen wichtig erscheint. Untersuchen Sie diese mithilfe von Argumenten aus den Quellen und Darstellungen. Verfassen Sie einen zusammenhängenden Text (ca. 400 Wörter), in dem Sie Ihre Überlegungen verarbeiten. Zeitbudget: 80 Minuten».	Auszug Lehrmitteldarstellung (Fuchs, Utz & Gautschi, 2014); Bundesbeschluss vom 3. August 1914 (Hardegger, 1989, 9)
<i>Militärische Neutralität</i>	Brief von Lengie (1915); Bericht von Sprecher von Bernegg (1917)
<i>Wirtschaftliche Neutralität</i>	Bundesratsprotokoll vom 5. Dezember 1914 (Bundesrat, 1914a); Geheimes Protokoll des Bundesrates vom 4. Dezember 1914 (Bundesrat, 1914b)

Die historischen Argumentationen wurden hinsichtlich der Operationen narrativer Kompetenz mittels Kategorien zwischen 0 und 3 Punkten von zwei wissenschaftlichen Assistent*innen beurteilt (siehe Tabelle 6), die jeweils die Hälfte aller Texte randomisiert zugewiesen bekamen. Die Texte waren zuvor hinsichtlich Erhebungszeitpunkt und so weiter anonymisiert worden. Die Interrater-Reliabilität zwischen dem Erstautor und den beiden Ratern wurde vor und in der

Mitte des Auswertungsprozesses geprüft. Fünf Kategorien wiesen eine befriedigende bis sehr gute Übereinstimmung auf (siehe Tabelle 7). Anhand dieser Ratings bestätigte eine Konfirmatorische Faktorenanalyse (CFA)³, ein Verfahren zur Überprüfung der Dimensionalität von nicht direkt zu beobachtenden Konstrukten wie zum Beispiel narrative Kompetenz (Weiber & Mühlhaus, 2014), die Eindimensionalität ($CFI_{t_1} = 0.99$, $RMSEA_{t_1} = .10$, $SRMR_{t_1} = .27$; $CFI_{t_2} = 1.00$, $RMSEA_{t_2} = .07$, $SRMR_{t_2} = .09$), sodass ein Summenscore «narrative Kompetenz» gebildet wurde ($M_{t_1} = 4.47$, $SD_{t_1} = 2.53$; $M_{t_2} = 7.12$, $SD_{t_2} = 3.78$; $Min = 0$, $Max = 15$).

Tabelle 6: Ratingkategorien «Narrative Kompetenz» und Beurteilungskategorien

Punkte und Aspekte	0	1	2	3
Medienbezug	<i>Fehlt</i>	<i>Implizit:</i> Aussagen aus angebotenen Medien werden nicht belegt.	<i>Teilweise vorhanden:</i> Ein eindeutiger Beleg wird verwendet.	<i>Transparent:</i> Mehr als ein eindeutiger Beleg wie «Material A», «Autor*in ...» wird verwendet.
Medienkritik	<i>Fehlt</i>	<i>Unbegründet:</i> Glaubwürdigkeit wird berücksichtigt, aber nicht begründet	<i>Teilweise begründet:</i> Glaubwürdigkeit wird durch einen Aspekt begründet.	<i>Umfangreich begründet:</i> Glaubwürdigkeit wird durch mind. zwei Aspekte wie die «Aussageabsicht», «Ereignisnähe» begründet.
Begriffsverwendung «Neutralität»	<i>Fehlt</i> oder <i>nicht plausibel anhand des Medienangebots</i> oder <i>nicht zeitlich verortet.</i>	<i>Zeitliche Verortung:</i> Der Begriff wird zeitlich verortet, aber den Lesenden nicht erklärt.	<i>Zeitbezogene Definition:</i> Der Begriff wird zeitlich verortet und den Lesenden erklärt, ohne dass zwei Zeitpunkte deutlich werden.	<i>Historische Definition:</i> Der Begriff wird den Lesenden so erklärt, dass Zeitdifferenz deutlich wird.

3 MLR-Schätzer in Stata 15.1 nach Robustheitscheck mittels DWLS-Schätzer in «laavaan» (Rosseel, 2012).

Punkte und Aspekte	0	1	2	3
Argumentationsstruktur	<i>Unstrukturiert:</i> Historische Aussagen sind unverbunden.	<i>Teilweise arg. strukturiert:</i> Die Struktur einer «Diskussion» ist erkennbar. Die Aussagen im Hauptteil sind nicht aufeinander bezogen.	<i>Linear arg. strukturiert:</i> Die Struktur einer «Diskussion» ist erkennbar. Im Hauptteil wird für oder gegen die Frage/These argumentiert.	<i>Abwägend arg. strukturiert:</i> Die Struktur einer «Diskussion» ist erkennbar (Tabelle 1). Im Hauptteil wird für oder gegen die Frage/These argumentiert, jedoch auch wenigstens ein Gegenargument, das entkräftet wird, einbezogen.
Aufbau der Argumente	<i>Fehlt</i>	<i>Selten vorhanden:</i> Ein Argument besteht aus allen Aspekten <i>oder</i> zwei Argumente bestehen aus zwei Aspekten.	<i>Teilweise vorhanden:</i> Zwei Argumente bestehen aus allen Aspekten <i>oder</i> mehr als zwei Argumente bestehen aus zwei Aspekten.	<i>Umfänglich vorhanden:</i> Mehr als zwei Argumente bestehen aus allen Aspekten (Tabelle 1)

Die epistemologischen Beliefs der Schüler*innen wurden zu t_1 und t_2 mittels einer gekürzten Version des Fragebogens von Nitsche (2019) erhoben, der positivistische, skeptische und narrativ-konstruktivistische Annahmen enthielt. Die CFA legte jedoch eine zwei- statt dreidimensionale Struktur nahe ($CFI_{t_1} = 0.94$, $RMSEA_{t_1} = .07$, $SRMR_{t_1} = .08$; $CFI_{t_2} = 1.00$, $RMSEA_{t_2} = .00$, $SRMR_{t_2} = .05$). Vier Items (z. B. «Geschichte ist grundsätzlich neutral und objektiv») repräsentieren den Positivismus ($M_{t_1} = -0.01$, $SD_{t_1} = 0.53$, $Min_{t_1} = -1.40$, $Max_{t_1} = 1.20$, $\alpha_{t_1} = .58$; $M_{t_2} = 0.01$, $SD_{t_2} = 0.55$, $Min_{t_2} = -1.23$, $Max_{t_2} = 1.17$, $\alpha_{t_2} = .65$), drei Items (z. B. «Jeder begründet seine Sicht auf die Vergangenheit individuell») den Skeptizismus ($M_{t_1} = -0.00$, $SD_{t_1} = 0.69$, $Min_{t_1} = -1.43$, $Max_{t_1} = 1.49$, $\alpha_{t_1} = .66$; $M_{t_2} = -0.01$, $SD_{t_2} = 0.72$, $Min_{t_2} = -1.37$, $Max_{t_2} = 1.54$, $\alpha_{t_2} = .66$). Für die Folgeanalysen wurden die Faktorscores verwendet.

Tabelle 7: Interrater-Reliabilität der Einzelkategorien narrativer Kompetenz

Kategorie	Krippendorfs α (Prozent) t1	Krippendorfs α (Prozent) t2
Medienbezug	.90 (95 %)	.89 (98 %)
Medienkritik	1.0 (100 %)	1.0 (100 %)
Begriffsverwendung «Neutralität»	.80 (93 %)	.74 (90 %)
Argumentationsstruktur	.74 (94 %)	.69 (92 %)
Aufbau Argumente	.73 (92 %)	.74 (95 %)

Die Personenmerkmale der Schüler*innen wurden zu t₁ erfasst. Das Alter wurde über die Differenz zwischen dem Datum der Erhebung und dem Geburtsdatum, das Geschlecht anhand einer binären Variable (weiblich (0), männlich (1)) und die letzten Zeugnisnoten in Deutsch und Geschichte mithilfe einer offenen Frage erfasst. Der sozioökonomische Status der Eltern (SES) wurde mittels drei geschlossener Teilfragen erhoben (Bücher im Haushalt: 0 bis 100 (0) – über 500 (6); Abschluss der Mutter beziehungsweise des Vaters: kein Abschluss (0) – Hochschulabschluss (7)), deren Beantwortung in die Schätzung eines eindimensionalen SES-Faktorscores einfluss ($CFI_{t_1} = 1.00$, $TLI_{t_1} = 1.00$, $RMSEA_{t_1} = .00$).

3.6 Auswertung

Die Analyse der Missings verdeutlicht, dass durchschnittlich 15.43 Prozent aller Werte fehlen. Für die CFA sowie die Gruppenvergleiche wurde mit den vorliegenden Werten gearbeitet. Vor der Schätzung der Korrelationen und Effekte erfolgte eine Imputation mittels *MICE-Algorithmus* (10 Imputationen). Dabei wurden die fehlenden Werte mittels Regressionsarten geschätzt, die für die jeweiligen Variablenarten (z. B. logistische Regressionen für kategoriale Variablen wie Geschlecht) spezifiziert wurde (Azur, Stuart, Frangakis & Leaf, 2011).

Mit Blick auf H₁ und die Kriteriumsvalidität der beiden Zielkonstrukte wurden Korrelationen zwischen allen Variablen mittels linearer Regressionen, einem Verfahren zur Bestimmung linearer Zusammenhänge (Fahrmeir, Kneib & Lang, 2009), anhand der z-standardisierten Variablen ($M = 0$, $SD = 1$) geschätzt, deren Koeffizienten sich analog zu *Pearsons-r* interpretieren lassen (Strohmer, 2018). Dazu wurden die Daten zu t₁ genutzt, da zu diesem Zeitpunkt alle Variablen erhoben wurden. Der Zusammenhang zwischen narrativer Kompetenz und epistemologischen Beliefs wurde zudem für t₂ analysiert. H₂ wurde mittels zwei

multipler linearer Regressionen mit narrativer Kompetenz als abhängiger Variable geschätzt. Die erste Schätzung diente der Überprüfung von Gruppenunterschieden hinsichtlich der gewählten Interventionsthemen innerhalb der Versuchsgruppe. Die zweite Analyse wurde anhand von Versuchs- und Kontrollgruppe vorgenommen. Die Testung von H₃ erfolgte mithilfe von zwei multiplen linearen Regressionen mittels *GLS-Schätzer* (Stein & Bekalarczyk, 2017), bei denen die Standardfehler für die Messwiederholung korrigiert wurden, mit Positivismus und Skeptizismus als abhängigen Variablen.

Um die Effekte der Schreibintervention zu t₂ zu bestimmen wurde der *difference-in-difference*-Ansatz gewählt. Bei diesem wird davon ausgegangen, dass sich das interessierende Merkmal der Personen in den Gruppen ohne Intervention vergleichbar entwickeln würde, sodass sich mittels Regressionsanalyse schätzen lässt, ob die Intervention das Merkmal der Beteiligten in der Versuchsgruppe im Vergleich zur Kontrollgruppe veränderte. Ist dies der Fall, wird die Veränderung als kausaler Effekt interpretiert (Lechner, 2011). Dazu schätzten wir den Interventionseffekt auf die narrative Kompetenz der Proband*innen (abhängige Variable) anhand eines Interaktionsterms (unabhängige Variable) zwischen den zwei Dummy-Variablen «Gruppe» (Kontrollgr. = 0, Versuchsgr. = 1) und «Zeitpunkt» (t₁ = 0, t₂ = 1) unter Kontrolle von «Gruppe» und «Zeitpunkt» (Kontrollvariablen). Alle Modelle wurden zudem für die Kontrollvariablen Alter, Geschlecht, SES sowie Geschichts- und Deutschnote korrigiert. Die Effekte auf die narrative Kompetenz wurden zudem mit Blick auf die Interaktion zwischen «Schreibaufgabe» (Wirtschaft = 0, Militär = 1) und «Zeitpunkt» unter Kontrolle von «Schreibaufgabe» und «Zeitpunkt» sowie für die Beurteilenden der Texte (Rater₁ = 0, Rater₂ = 1) analysiert, um die Einflüsse des Themas der Schreibaufgaben zu t₁, 2 und der Rater zu kontrollieren. Bei der Überprüfung von H₁ bis H₃ erfolgte die Korrektur der Standardfehler für die Gruppierung der Schüler*innen anhand ihrer Zuordnung zu Lehrpersonen mittels Sandwich-Schätzer (White, 1980).

Zur Interpretation der Effekte wurde der Summenscore «narrative Kompetenz» standardisiert, indem der M_{t_1} von jedem Score pro Fall subtrahiert und durch die SD_{t_1} für die Gesamtstichprobe dividiert wurde. Dadurch lassen sich die geschätzten Effekte als durchschnittliche Zuwächse in SD_{t_1} und damit bezogen auf die Leistungen des vorliegenden Samples vor der Intervention interpretieren (Bertram, Wagner & Trautwein, 2017). Für die Effektgrößen werden die Kennwerte von Kraft (2020) herangezogen (klein: $SD < 0.05$; mittel: $SD = 0.05-0.20$; groß: $SD > 0.20$). Um die Interventionseffekte ebenfalls qualitativ zu interpretieren (Lind, 2012), wurden die Unterschiede der Rohmittelwerte der Versuchs-

gruppe für die einzelnen Ratingkategorien zwischen t1 und t2 herangezogen und als Zuwachs hinsichtlich der Qualitätsabstufungen der Ratingkategorien interpretiert (siehe Tabelle 10).

3.7 Ergebnisse

Die Korrelationsanalyse zur Testung von H1 sowie der Kriteriumsvalidität verdeutlichen, dass zu t1 keine signifikanten Korrelationen zwischen narrativer Kompetenz und epistemologischen Beliefs vorliegen. Dies gilt ebenfalls für t2 ($r_{\text{narr. Komp. vs. Pos.}} = .12$, $r_{\text{narr. Komp. vs. Skept.}} = -.26$, $p > .05$), sodass die erste Annahme unter dem Vorbehalt verworfen werden muss, dass narrativ-konstruktivistische Beliefs nicht untersucht werden konnten. Korrelationen der narrativen Kompetenz und der epistemologischen Beliefs zu einem Außenkriterium etwa zur letzten Geschichtsnote im Zeugnis als schulischem Leistungsindikator liegen nicht vor. Weitere Korrelationen sind in Tabelle 8 dargestellt.

Tabelle 8: Korrelationen zwischen allen interessierenden Variablen zu t1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Narrative Kompetenz	1								
(2) Positivismus	-.03	1							
(3) Skeptizismus	-.07	-.40**	1						
(4) Geschichtsnote	.04	-.11	-.10	1					
(5) Deutschnote	.04	-.08	.03	.42**	1				
(6) Alter	.01	.10	-.03	.01	-.01	1			
(7) Geschlecht	-.13	-.07	.02	-.02	-.24*	.17	1		
(8) SES	.04	.06	-.10	.16	-.04	-.06	.01	1	
(9) Thema Schreibaufgabe	.24*	-.12	-.08	.07	.02	-.04	-.06	.00	1

Anm. * $p < .05$; ** $p < .01$.

Hinsichtlich H2 wurden zwei multiple Regressionen geschätzt (siehe Tabelle 9). Die erste Analyse für die Versuchsgruppe verdeutlicht, dass für das Thema der Schreibintervention («Nationalismus», «Migration») kein signifikanter Effekt auf die narrative Kompetenz der Schüler*innen nachweisbar ist. Hingegen lässt sich anhand der zweiten Modellanalyse unter Einbeziehung der Kontrollgruppe ein signifikanter Effekt der Schreibintervention auf die narrative Kompetenz der

Schüler*innen in der Versuchsgruppe feststellen, sodass die zweite Annahme nicht verworfen werden muss. Die Schätzung weist einen starken Effekt auf, da der Zuwachs um mehr als zwei SD_{t_1} ausfällt.

Tabelle 9: Effekte auf die narrative Kompetenz (an SD_{t_1} standardisierter Score) und die epistemologischen Beliefs (Faktorscore)

Variablen	Effekte	SE	M_{t_1} (SD_{t_1})	M_{t_2} (SD_{t_2})
Interventionseffekt auf narrative Kompetenz	2.13***	0.26		
Narrative Kompetenz d. Versuchsgruppe ($n = 100$)			0.07 (1.05)	2.30 (1.24)
Narrative Kompetenz d. Kontrollgruppe ($n = 73$)			-0.06 (0.95)	-0.19 (0.97)
Effekt d. Interventionsthemas	.08	0.33		
Narrative Kompetenz d. «Nationalismus»-Gruppe ($n = 45$)			-0.00 (1.05)	2.25 (0.94)
Narrative Kompetenz d. «Migration»-Gruppe ($n = 55$)			0.16 (1.05)	2.37 (1.54)
Positivistische Beliefs	0.15	0.11	-0.02 (0.53)	0.05 (0.58)
Skeptische Beliefs	0.12	0.11	0.11 (0.77)	-0.03 (0.83)

Anm. M = Mittelwert; SD = Standardabweichung; SE = Standardfehler; *** $p < .001$. Für die Methode der Effektschätzung siehe Abschnitt 3.6.

Eine Analyse der Veränderungen der Beurteilungen anhand der Versuchsgruppe illustriert (siehe Tabelle 10), dass die Operationen der narrativen Kompetenz durch die Intervention in unterschiedlich starkem Maß gefördert wurden. So zeigt etwa der Zuwachs im Bereich der Argumentationsstruktur, dass es den Lernenden vor der Intervention im Mittel lediglich gelang, ihre historische Argumentation teilweise zu strukturieren, während ihre Texte zu t_2 im Mittel zwischen linear und abwägend strukturiert ausfielen. Demgegenüber nahm die Qualität der historischen Begriffsverwendung am Beispiel «Neutralität» nur mäßig zu und die Schüler*innen kamen im Mittel kaum über eine zeitliche Verortung des Begriffes hinaus. Insgesamt veränderten sich jedoch vier der fünf Kategorien um annähernd eine Qualitätsstufe unseres Ratings (siehe Tabelle 6), was als Ergebnis einer acht Doppellektionen umfassenden Intervention sehr beachtlich ist.

Mit Blick auf H3 verdeutlichen die Analysen keinen signifikanten Einfluss der Schreibintervention auf positivistische und skeptische Beliefs (siehe Tabelle 9), sodass die Annahme wiederum unter dem Vorbehalt verworfen werden muss, dass narrativ-konstruktivistische Beliefs nicht untersucht werden konnten.

Tabelle 10: Rohmittelwerte (M_{t_2, t_1}) und Veränderungen ($M_{t_2}-M_{t_1}$) anhand der Einzelratings der Operationen narrativer Kompetenz für die Versuchsgruppe

Einzelrating	$M_{t_2}-M_{t_1}$	M_{t_1} (SD_{t_1})	M_{t_2} (SD_{t_2})
Medienbezug	0.83	1.99 (0.94)	2.82 (0.53)
Medienkritik	0.98	0.00 (0.00)	0.98 (1.28)
Begriffsverwendung «Neutralität»	0.28	0.87 (0.57)	1.15 (0.61)
Argumentationsstruktur	1.33	1.10 (1.00)	2.43 (0.79)
Aufbau der Argumente	1.16	0.78 (0.93)	1.94 (0.90)

3.8 Diskussion und Limitationen der Ergebnisse

Aufgrund des Fehlens entsprechender deutschsprachiger Literatur zu Beginn unserer Intervention 2017 fragten wir erstens nach dem Zusammenhang zwischen der narrativen Kompetenz und den epistemologischen Beliefs von Schüler*innen auf Klassenstufe zehn und elf in der Deutschschweiz. Allerdings waren derartige Zusammenhänge weder für t_1 noch t_2 nachweisbar, sodass H1 zu verwerfen ist. Der Befund mag damit zusammenhängen, dass narrativ-konstruktivistische Items und damit solche, die elaborierte Beliefs repräsentierten, aus statistischen Gründen ausgeschlossen werden mussten. Demgegenüber verdeutlichen inzwischen vorliegende Studien gerade zwischen elaborierten Beliefs und Aspekten historischen Denkens statistische Verbindungen (Iordanou et al., 2020; Mierwald, 2020). Zudem waren die eingesetzten Items ursprünglich für Lehrpersonen entwickelt worden, weshalb die Schüler*innen die narrativ-konstruktivistischen Items möglicherweise nicht wie intendiert verstanden. Zukünftige Analysen sollten daher auf inzwischen vorliegende Fragebögen für Schüler*innen zurückgreifen (z. B. Mierwald, 2020). Zudem zeichnet sich ab, dass spezifische epistemologische Beliefs etwa zur Integration historischen Wissens mit Aspekten historischen Denkens wie Kontextualisieren verbunden sind (Wiley et al., 2020), sodass es zukünftig fruchtbar erscheint, Beliefs erhe-

bungen zu konzipieren, die konzeptuell eng auf die zu erfassenden historischen Leistung bezogen sind.

Zweitens untersuchten wir die Wirksamkeit einer Schreibintervention auf die narrative Kompetenz von Schüler*innen in einer Versuchsgruppe. Unsere Analyse verdeutlicht den Einfluss der Schreibintervention auf das genannte Konstrukt, sodass H2 nicht verworfen werden muss. Es handelt sich um einen starken Effekt, da die historischen Argumentationen der Schüler*innen in der Interventionsgruppe zu t2 um mehr als zwei SD_{t_1} besser ausfielen im Verhältnis zu den Bewertungen der Texte von allen Schüler*innen vor der Intervention. Gemessen an der maximal erreichbaren Punktzahl von 15 Punkten bedeutet dies einen Leistungszuwachs um etwa ein Drittel. Allerdings dürfte der starke Effekt wohl auch dadurch begründet sein, dass er im Vergleich zu den Leistungen aller Lernende zu t1 sowie einer Kontrollgruppe geschätzt wurde, die auf alltäglichem Geschichtsunterricht basieren. Studien zum alltäglichen Geschichtsunterricht zeigen jedoch, dass darin historische Medien eher inhaltserschließend verwendet werden und zudem wohl eher selten als Schreibanlass dienen (z.B. Marti et al., 2019; Spieß, 2014). Zudem verdeutlicht die Analyse der Teilaspekte narrativer Kompetenz, dass unsere Intervention historisches Schreiben (z.B. Aufbau historischer Argumente) stärker förderte als den Umgang mit Medien, was nicht verwundert, da mehr Unterrichtszeit für das Schreiben aufgewendet wurde. Zukünftige Interventionsstudien könnten daher ähnlich wie für andere Konstrukte (z.B. Bertram et al., 2017) die Wirkung unterschiedlicher Konditionen auf die narrative Kompetenz und ihre Teilaspekte etwa mittels Variation der Medien, Lernaufgaben oder der zeitlichen Umfänge des Modellierens, Begleitens und selbstständigen Anwendens prüfen, um geeignete Methoden für den Geschichtsunterricht zu identifizieren.

Schließlich untersuchten wir, ob unsere Schreibintervention die epistemologischen Beliefs der Lernenden ebenfalls förderte. Allerdings konnten keine statistisch signifikanten Effekte festgestellt werden, sodass H3 verworfen werden muss. Vermutlich war der Ausschluss narrativ-konstruktivistischer Beliefs wiederum ursächlich. Allerdings könnte das Resultat ebenfalls darin begründet liegen, dass wir in der Intervention aus Zeitgründen auf eine explizite Reflexion der epistemologischen (Vor-)Annahmen der Schüler*innen verzichteten, die sich in der Studie von Stoel, van Drie et al. (2017) als wirksam erwies. Zukünftig dürfte es daher darauf ankommen, Lernaufgaben in Schreibinterventionen zu integrieren, die bereits während der Bearbeitung zur epistemologischen Refle-

xion anregen, um epistemologische Beliefs für den Unterricht in zeitlich angemessener Weise zu adressieren.

Aufgrund organisatorischer Belange (z. B. relative Autonomie der Schulen) sind unsere statistischen Analysen dadurch limitiert, dass die Klassen nicht randomisiert auf die Konditionen verteilt werden konnten. Zudem verzichteten wir auf die Kontrolle der Analysen für die Unterrichtsumsetzung unserer Intervention (*treatment fidelity*), um die beteiligten Lehrpersonen und Klassen nicht durch weitere Maßnahmen wie Unterrichtsbeobachtungen zu belasten. Zwar bestätigten uns die Lehrpersonen der Versuchsgruppe auf einem Treffen zum Projektabschluss, sich an den Ablauf gehalten zu haben. Allerdings wurde ebenfalls deutlich, dass sie die Lernenden bei der Aufgabenlösung in unterschiedlichem Maße unterstützt hatten.

Trotz der Limitationen halten wir unsere Analysen für plausibel, da zentrale Einflussfaktoren auf die Effekte statistisch kontrolliert wurden und unsere Resultate mit bereits vorliegenden Ergebnissen vergleichbar sind (De La Paz & Felton, 2010; Nokes & De La Paz, 2018). Sie verdeutlichen, dass es engagierten Geschichtslehrpersonen gelingt, wesentliche Aspekte der narrativen Kompetenz ihrer Schüler*innen zu fördern, wenn sie Fördermaßnahmen in ihrem regulären Geschichtsunterricht einsetzen, die geschichtsdidaktisch sowie evidenzbasiert begründet sind und deren Abfolge am Cognitive-Apprenticeship-Ansatz (Modellieren, Begleiten, selbstständig Anwenden) orientiert ist.

Literatur

- Arndt, Ernst Moritz (2008). Über den Volkshass. In *Histoire/Geschichte. Europa und die Welt. Vom Wiener Kongress bis 1945*. Leipzig: Klett.
- Azur, Melissa J./Stuart, Elizabeth A./Frangakis, Constantine & Leaf, Philip J. (2011). Multiple imputation by chained equations: what is it and how does it work? Multiple imputation by chained equations. *International Journal of Methods in Psychiatric Research*, 20(1), 40–49. <https://doi.org/10.1002/mp.329>.
- Barricelli, Michele (2016). Historisches Erzählen als Kern historischen Lernens. In Martin Buchsteiner & Martin Nitsche (Hrsg.), *Historisches Erzählen und Lernen* (S. 45–68). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-11078-9_3.
- Barsch, Sebastian (2020). *Inklusiven Geschichtsunterricht planen*. Frankfurt a. M.: Wochenschau.
- Bertram, Christiane/Wagner, Wolfgang & Trautwein, Ulrich (2017). Learning historical thinking with oral history interviews: a cluster randomized controlled intervention study of oral history interviews in history lessons. *American Educational Research Journal*, 54(3), 444–484. <https://doi.org/10.3102/0002831217694833>.

- Bitzium, Albert (2001). Die Ziele der Zofingia 1819. In William E. Simmat (Hrsg.), *Das Werden der modernen Schweiz. Vom Ancien Régime zum Ersten Weltkrieg (1798–1914)* (S. 59). Basel: Lehrmittelverlag. <https://doi.org/10.1002/9781119100812.ch6>.
- Boxtel, Carla van & Drie, Jannet van (2018). Historical reasoning: conceptualizations and educational applications. In Scott Alan Metzger & Lauren McArthur Harris (Hrsg.), *The Wiley international handbook of history teaching and learning* (S. 149–176). New York: Wiley-Blackwell.
- Braun, Rudolf (1970). Sozio-kulturelle Probleme der Eingliederung italienischer Arbeitskräfte in der Schweiz. Zürich: Rentsch.
- Britt, M. Anne/Perfetti, Charles A./Van Dyke, Julie & Gabrys, Gareth (2000). The sourcer's apprentice. A tool for document-supported history instruction. In Peter N. Stearns/Peter Seixas & Sam Wineburg (Hrsg.), *Knowing, teaching and learning history. National and international perspectives* (S. 437–470). New York, London: New York University Press.
- Bundesrat (1914a). *Bundesratsprotokoll vom 5. Dezember 1914, den Handel mit den Alliierten betreffend (Übersetzung aus dem Französischen)*. dodis.ch/43355.
- Bundesrat (1914b). *Geheimes Protokoll des Bundesrates vom 4. Dezember 1914*. dodis.ch/43565.
- Coffin, Caroline (2006). Historical discourse: the language of time, cause, and evaluation. London: Continuum.
- Collins, Allan/Brown, John Seely & Newman, Susan E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In Lauren Resnick (Hrsg.), *Knowing, learning, and instruction. Essays in honor of Robert Glaser* (S. 453–494). Hillsdale: Erlbaum.
- Conrad, Sebastian (2016, 8. Januar). Die Welt von gestern. *Frankfurter Allgemeine Zeitung*. <https://www.faz.net/aktuell/politik/die-gegenwart/migration-und-die-industrialisierung-der-verkehrsmittel-die-welt-von-gestern-13995291.html> (07.04.2022).
- Daniel, Ute (2006). *Kompendium Kulturgeschichte*. Frankfurt a. M.: Suhrkamp.
- De La Paz, Susan/ Monte-Sano, Chauncey/Felton, Mark/Croninger, Robert/Jackson, Cara & Piantedosi, Kelly Worland (2017). A historical writing apprenticeship for adolescents: integrating disciplinary learning with cognitive strategies. *Reading Research Quarterly*, 52(1), 31–52. <https://doi.org/10.1002/rrq.147>.
- De La Paz, Susan & Felton, Mark (2010). Reading and writing from multiple source documents in history: Effects of strategy instruction with low to average high school writers. *Contemporary Educational Psychology*, 35(3), 174–192. <https://doi.org/10.1016/j.cedpsych.2010.03.001>.
- Drie, Jannet van/Boxtel, Carla van & Braaksmä, Martine (2014). Writing to engage students in historical reasoning. In Perry Klein/Pietro Boscolo/Lori Kirkpatrick & Carmen Gelati (Hrsg.), *Writing as a learning activity* (S. 94–119). Leiden: Brill. <https://doi.org/10.1007/s10212-014-0219-4>.
- Fahrmeir, Ludwig/Kneib, Thomas & Lang, Stefan (2009). *Regression: Modelle, Methoden und Anwendungen* (2. Aufl.). Berlin: Springer.
- Friedman, Max Paul (2017, 18. April). Muss Nationalismus immer eine Gefahr sein? *Deutschlandradio Kultur, Politisches Feuilleton*. <https://www.deutschlandfunkkultur.de/patrio->

- tismus-muss-nationalismus-immer-eine-gefahr-sein.1005.de.html?dram:article_id=383962 (07.04.2022).
- Fuchs, Karin/Utz, Hans & Gautschi, Peter (2014). 14/18. *Die Schweiz und der Grosse Krieg. Dossier Sek I*. <https://www.ersterweltkrieg.ch/publikationen/> (07.04.2022).
- Günther-Arndt, Hilke/Hoffmann, Dirk & Zwölfer, Norbert (Hrsg.) (1995). *Geschichtsbuch Oberstufe. Von der Antike bis zum Ende des 19. Jahrhunderts* (Bd. 1). Berlin: Cornelsen.
- Habscheid, Stephan (Hrsg.) (2011). *Textsorten, Handlungsmuster, Oberflächen: linguistische Typologien der Kommunikation*. Berlin: De Gruyter.
- Handro, Saskia (2013). Sprache und historisches Lernen. Dimensionen eines Schlüsselproblems des Geschichtsunterrichts. In Michael Becker-Mrotzek, Karen Schramm, Eike Thürmann & Helmut J. Vollmer (Hrsg.), *Sprache im Fach: Sprachlichkeit und fachliches Lernen* (S. 317–333). Münster: Waxmann.
- Hardegger, Joseph (1989). *Das Werden der modernen Schweiz* (Bd. 2). Basel: Lehrmittelverlag.
- Hartung, Olaf (2013). *Geschichte Schreiben Lernen: empirische Erkundungen zum konzeptionellen Schreibhandeln im Geschichtsunterricht*. Berlin: Lit.
- Hodel, Jan/Waldis, Monika/Zülsdorf-Kersting, Meik & Thünemann, Holger (2013). Schülernarrationen als Ausdruck historischer Kompetenz. *Zeitschrift für Didaktik der Gesellschaftswissenschaften*, 4(2), 121–145.
- Huntington, Samuel P. (2001, 11. Mai). Die Gefahren der neuen Migration. *Die Welt*. <https://www.welt.de/print-welt/article450407/Die-Gefahren-der-neuen-Migration.html> (11.04.2022).
- Iordanou, Kalypso/Kendeou, Panayiota & Zembylas, Michalinos (2020). Examining my-side bias during and after reading controversial historical accounts. *Metacognition and Learning*, 15(3), 319–342. <https://doi.org/10.1007/s11409-020-09240-w>.
- Kansteiner, Wulf (2018). Argumentation, Beschreibung und Erzählung in der wissenschaftlichen Historiographie. In Thomas Sandkühler & Horst Walter Blanke (Hrsg.), *Historisierung der Historik. Jörn Rüsen zum 80. Geburtstag* (S. 151–168). Wien: Böhlau.
- Körper, Andreas (2022). Kompetenzmodelle in der Geschichtsdidaktik. In Georg Weißeno & Béatrice Ziegler (Hrsg.), *Handbuch Geschichts- und Politikdidaktik* (S. 3–16). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-29668-1_1.
- Kraft, Matthew A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>.
- Lechner, Michael (2011). The estimation of causal effects by difference-in-difference methods estimation of spatial panels. *Foundations and Trends in Econometrics*, 4(3), 165–224. <https://doi.org/10.1561/0800000014>.
- Lenje, André (1915). *Brief an den Vorsteher des Schweizer Militärdepartements vom 08.12.1915 (Übersetzung aus dem Französischen)*. <https://dodis.ch/43435>.
- Lind, Georg (2012). *Effektstärken: Statistische, praktische und theoretische Bedeutsamkeit empirischer Studien*. <http://nbn-resolving.de/urn:nbn:de:bsz:352-217760> (11.04.2022).
- Martens, Matthias (2015). Understanding the nature of history. Students' tacit epistemology in dealing with conflicting historical narratives. In Arthur Chapman & Arie Wil-schut (Hrsg.), *Joined-up history. New directions in history education* (S. 211–230). Charlotte: Information Age.

- Marti, Philipp/Gollin, Kristine & Waldis, Monika (2019). Zur Praxis und Reflexion historischen Schreibens im Geschichtsunterricht: Fallbasierte Ergebnisse einer Interviewstudie mit Lehrkräften an Deutschschweizer Gymnasien. *Zeitschrift für Geschichtsdidaktik*, 18, 168–189. <https://doi.org/10.13109/zfgd.2019.18.1.166>.
- Marti, Philipp & Waldis, Monika (2017). Zur Entwicklung und Bewertung von offenen Aufgabenformaten für den Geschichtsunterricht. *Erziehung und Unterricht. Österreichische Pädagogische Zeitschrift*, 167(9–10), 963–972.
- Meyer, Helmut (2002). *Die Geschichte der Schweiz*. Berlin: Cornelsen.
- Mierwald, Marcel (2020). *Historisches Argumentieren und epistemologische Überzeugungen: Eine Interventionsstudie zur Wirkung von Lernmaterialien im Schülerlabor*. Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-658-29955-2>.
- Mierwald, Marcel & Brauch, Nicola (2015). Historisches Argumentieren als Ausdruck historischen Denkens. *Zeitschrift für Geschichtsdidaktik*, 14, 104–120.
- Münkler, Herfried (2015, 5. September). Die Geschichte menschlicher Wanderungsbewegungen. *Neue Zürcher Zeitung*. <https://www.nzz.ch/feuilleton/ein-blick-in-die-geschichte-menschlicher-wanderungsbewegungen-1.18607660> (11.04.2022).
- Nitsche, Martin (2019). *Beliefs von Geschichtslehrpersonen – eine Triangulationsstudie*. Bern: hep. <https://doi.org/10.36933/9783035516005>.
- Nitsche, Martin & Gollin, Kristine (2020). Zeitlichkeit und narrative Kompetenz – zur kategorialen Erfassung des Umgangs mit Zeit. In Markus Bernhardt & Thomas Sandkühler (Hrsg.), *Sprache(n) des Geschichtsunterrichts. Sprachliche Vielfalt und Historisches Lernen* (S. 311–327). Göttingen: V&R unipress.
- Nitsche, Martin/Gollin, Kristine & Waldis, Monika (2017). Zur Konstruktion von offenen Testaufgaben für die Erfassung narrativer Kompetenz – Kriterienentwicklung und Studienvergleiche. *Zeitschrift für Geschichtsdidaktik*, 16(1), 235–249.
- Nitsche, Martin & Waldis, Monika (2016). Narrative Kompetenz von Studierenden erfassen. Zur Annäherung an formative und summative Vorgehensweisen im Fach Geschichte. *Zeitschrift für Didaktik der Gesellschaftswissenschaften*, 7(1), 17–35.
- Nokes, Jeffery D. & De La Paz, Susan (2018). Writing and argumentation in history education. In Scott Alan Metzger & Lauren McArthur Harris (Hrsg.), *The Wiley international handbook of history teaching and learning* (S. 551–578). New York: Wiley-Blackwell.
- Nokes, Jeffery D./Dole, Janice A. & Hacker, Douglas J. (2007). Teaching high school students to use heuristics while reading historical texts. *Journal of Educational Psychology*, 99(3), 492–504. <http://dx.doi.org/10.1037/0022-0663.99.3.492>.
- Norden, Jörg van (2011). *Was machst Du für Geschichten? Didaktik eines narrativen Konstruktivismus*. Freiburg: Centaurus.
- Norden, Jörg van & Schürenberg, Wanda (Hrsg.) (2019). *Lernprogression narrativer Kompetenz im Geschichtsunterricht: Ein Vergleich von Waldorf- und Regelschule*. Schwalbach/Ts.: Wochenschau.
- Notz, Thomas (Hrsg.) (2010). *Schweizer Geschichtsbuch 2. Vom Absolutismus bis zum Ende des Ersten Weltkrieges*. Berlin: Cornelsen.
- Olmert, Jochen (2016). *Globale Migration. Geschichte und Gegenwart*. München: C. H. Beck.
- o. A. (1853, 31. Juli). Bericht, die Kolonisation in Brasilien betreffend. *Der Kolonist*, 3(30). <https://kbaargau.visual-library.de/periodical/pageview/9590> (21.04.2022).

- Pandel, Hans-Jürgen (2010). *Historisches Erzählen. Narrativität im Geschichtsunterricht*. Schwalbach/Ts.: Wochenschau.
- Pfizer, Paul Achatius (1995). Gedanken über das Ziel und die Aufgabe des deutschen Liberalismus. In Norbert Zwölfer & Elisabeth Zwölfer (Hrsg.), *Geschichtsbuch Oberstufe. Von der Antike bis zum Ende des 19. Jahrhunderts* (Bd. 1, S. 372–373). Berlin: Cornelsen.
- Pfyffer, Kasimir (2001). Rede zur Bundesverfassung 1848 von Kasimir Pfyffer. In William E. Simmat (Hrsg.), *Das Werden der modernen Schweiz. Vom Ancien régime zum Ersten Weltkrieg (1798-1914)* (S. 65–66). Basel: Lehrmittelverlag.
- Reisman, Avishag (2012). Reading like a historian. A document-based history curriculum intervention in urban high schools. *Cognition and Instruction*, 30(1), 86–112. <https://doi.org/10.1080/07370008.2011.634081>.
- Ricoeur, Paul (2007). *Zeit und Erzählung. Band I: Zeit und historische Erzählung*. Paderborn: Fink.
- Roß, Jan (2015, 30. September). Dunkler Stolz von Pegida bis Peking: Die unheimliche neue Karriere des Nationalismus. *Die Zeit*. <https://www.zeit.de/2015/36/nationalismus-deutschland-china> (21.04.2022).
- Rosseel, Yves (2012). Lavaan: An R package for structural equation modeling. *Journal of statistical software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>.
- Rüsen, Jörn (2013). *Historik. Theorie der Geschichtswissenschaft*. Köln: Böhlau.
- Schrader, Viola (2019). Schülerinnen und Schüler vergleichen kontroverse Historikerpositionen. Zum Zusammenhang historischen Denkens und sprachlichen Handelns. In Christine Pflüger (Hrsg.), *Die Komplexität des kompetenzorientierten Geschichtsunterrichts. Aktuelle geschichtsdidaktische Forschungen* (S. 57–74). Göttingen: V&R unipress.
- Schreiber, Waltraud et al. (2006). *Historisches Denken. Ein Kompetenz-Strukturmodell*. Neuried: ars una.
- Schümer, Dirk (2015, 30. Dezember). Der gute alte Nationalismus. *Die Welt*. https://www.welt.de/print/welt_kompakt/debatte/article150449830/Der-gute-alte-Nationalismus.html (11.04.2022).
- Schwarzenbach, James (1974). *Die Überfremdung der Schweiz – wie ich sie sehe*. Zürich: Verlag der Republikaner.
- Skenderovic, Damir (2015). Fremdenfeindlichkeit. Xenophobie. In *Historisches Lexikon der Schweiz*. <https://hls-dhs-dss.ch/de/articles/016529/2015-05-05/> (21.04.2022).
- Spieß, Christian (2014). *Quellenarbeit im Geschichtsunterricht*. Göttingen: V&R unipress.
- Sprecher von Bernegg, Theophil (1917). Geheimer Bericht von Theophil von Sprecher, Generalstabschef der Schweizer Armee, an den Bundespräsidenten Arthur Hoffmann vom 7. April 1917. <https://dodis.ch/43565>.
- Stein, Petra & Bekalarczyk, Dawid (2017). Modelle mit variablen Regressionskonstanten – Random- und Fixed-Effects-Modelle. https://www.uni-due.de/imperia/md/content/soziologie/modelle_mit_variablen_regressionskonstanten.pdf (07.04.2022).
- Stoel, Gerhard/van Drie, Jannet & van Boxtel, Carla (2015). Teaching towards historical expertise. Developing a pedagogy for fostering causal reasoning in history. *Journal of Curriculum Studies*, 47(1), 49–76. <https://doi.org/10.1080/00220272.2014.968212>.

- Stoel, Gerhard/van Drie, Jannet & van Boxtel, Carla (2017). The effects of explicit teaching of strategies, second-order concepts, and epistemological underpinnings on students' ability to reason causally in history. *Journal of Educational Psychology*, 109(3), 321–337. <https://doi.org/10.1037/edu0000143>.
- Stoel, Gerhard/Logtenberg, Albert/Wansink, Bjorn/Huijgen, Tim/van Boxtel, Carla & van Drie, Jannet (2017). Measuring epistemological beliefs in history education: An exploration of naïve and nuanced beliefs. *International Journal of Educational Research*, 83, 120–134. <https://doi.org/10.1016/j.ijer.2017.03.003>.
- Strohmer, Janina (2018). Regressionsgewichte. In Markus Antonius Wirtz (Hrsg.), *Dorsch. Lexikon der Psychologie* (S. 1313). Bern: Hogrefe. <https://portal.hogrefe.com/dorsch/regressionsgewichte/>.
- Tavel, Emmanuel Rudolf von (2001). Entwurf eines Bundes der Gerechten zum Schutz der Religion, der Gerechtigkeit und der wahren Freiheit 1833. In William E. Simmat (Hrsg.), *Das Werden der modernen Schweiz. Vom Ancien régime zum Ersten Weltkrieg (1798-1914)* (S. 65–66). Basel: Lehrmittelverlag.
- Thurich, Eckart (2011). Nationalismus. In *pocket politik. Demokratie in Deutschland*. Bonn: Bundeszentrale für politische Bildung. <https://www.bpb.de/nachschlagen/lexika/pocket-politik/16503/nationalismus> (07.04.2022).
- VanSledright, Bruce A. (2014). *Assessing historical thinking and understanding: Innovative designs for new standards*. New York: Routledge.
- VanSledright, Bruce A. & Maggioni, Liliana (2016). Epistemic cognition in history. In Jeffrey Alan Greene, William A. Sandoval & Ivar Braten (Hrsg.), *Handbook of epistemic cognition* (S. 128–146). New York: Routledge.
- Voss, James F. & Wiley, Jennifer (1997). Developing understanding while writing essays in history. *International Journal of Educational Research*, 27(3), 255–265. [https://doi.org/10.1016/S0883-0355\(97\)89733-9](https://doi.org/10.1016/S0883-0355(97)89733-9).
- Waldis, Monika (2016). Erzählung oder Argumentation. In Stefan Keller & Christian Reintjes (Hrsg.), *Aufgaben als Schlüssel zur Kompetenz. Didaktische Herausforderungen, wissenschaftliche Zugänge und empirische Befunde* (S. 237–258). Münster: Waxmann.
- Waldis, Monika/Marti, Philipp & Nitsche, Martin (2015). Angehende Geschichtslehrpersonen schreiben Geschichte(n). Zur Kontextabhängigkeit historischer Narrationen. *Zeitschrift für Geschichtsdidaktik*, 14, 63–86.
- Waldis, Monika/Nitsche, Martin & Gollin, Kristine (2020). «Schülerinnen und Schüler schreiben Geschichte» – Eine Interventionsstudie an Deutschschweizer Gymnasien. *Zeitschrift für Geschichtsdidaktik*, 19, 90–108. <https://doi.org/10.13109/zfgd.2020.19.1.90>.
- Weiber, Rolf & Mühlhaus, Daniel (2014). *Strukturgleichungsmodellierung. Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS* (2. Aufl.). Berlin, Heidelberg: Springer Gabler.
- Wessendorf, Berthold (1973). *Die überseeische Auswanderung aus dem Kanton Aargau im 19. Jahrhundert*. Aarau: Sauerländer.
- Westermann (Hrsg.) (2010). *Horizonte 12, Geschichte Gymnasium Bayern*. Braunschweig: Westermann.

- White, Halbert (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817. <https://doi.org/10.2307/1912934>.
- Wickner, Mareike-Cathrin (2019). Über die Vorzüge einer textsortenbasierten Schreibförderung im Geschichtsunterricht. Forschungsergebnisse aus dem SchriFT-Projekt. In Heike Roll et al. (Hrsg.), *Schreiben im Fachunterricht der Sekundarstufe I unter Einbeziehung des Türkischen. Empirische Befunde aus den Fächern Geschichte, Physik, Technik, Politik, Deutsch und Türkisch* (S. 129–147). Münster, New York: Waxmann. <https://www.waxmann.com/index.php?eID=download&buchnr=4088> (07.04.2022).
- Wiley, Jennifer/Griffin, Thomas D./Steffens, Brent & Britt, M. Anne (2020). Epistemic beliefs about the value of integrating information across multiple documents in history. *Learning and Instruction*, 65. <https://doi.org/10.1016/j.learninstruc.2019.101266>.
- Wineburg, Samuel S. (1991). Historical problem solving: a study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83(1), 73–87. <http://dx.doi.org/10.1037/0022-0663.83.1.73>.
- Wolfrum, Birte & Sauer, Michael (2007). Zum Bildverständnis von Schülern. Ergebnisse einer empirischen Studie. *Geschichte in Wissenschaft und Unterricht*, 7(8), 400–416.
- Ziegler, Béatrice (2017). *Geschichte der schweizerischen Auswanderung nach Brasilien*. Unveröffentlichtes Manuskript.

Eine Intervention zur Förderung geschichtsdidaktischen Wissens und Könnens zur Konstruktion von Aufgaben für den Geschichtsunterricht¹

Mario Resch, Christian Heuer und Hendrik Lohse-Bossenz

1 Einleitung

In der Planung, Durchführung und Reflexion von historischen Lehr-Lern-Prozessen spielen Aufgaben eine bedeutende Rolle, denn sie bilden die Schnittstelle zwischen Unterrichtsangebot und -nutzung (Heuer, 2021). Durch Aufgaben ermöglichen Geschichtslehrer*innen Beziehungen zwischen Zielen, Inhalten, Medien und Schüler*innen (Köster, 2021, 7). In aktuellen Kompetenzmodellierungen werden Fähigkeiten, Fertigkeiten und Bereitschaften zur Konstruktion von Aufgaben als bedeutsame Lehrkompetenz hervorgehoben (Resch, 2021) und im geschichtsdidaktischen Diskurs besteht ein Konsens über das spezifische Aufgabenwissen, das unter anderem Kenntnisse über fachspezifische Gütekriterien und Analysekategorien zur Bestimmung der Aufgabenschwierigkeit sowie Wissen über den zielgerichteten Einsatz von handlungsinitierenden Operatoren vorsieht. Gleichzeitig zeigen empirische Befunde, dass die Aufgabenkonstruktion und das Erkennen geeigneter Aufgaben eine Herausforderung für (angehende) Geschichtslehrkräfte darstellt (z. B. Resch & Heuer, 2019).

Um Fähigkeiten, Fertigkeiten und Bereitschaften zur Konstruktion von Lernaufgaben bei angehenden Geschichtslehrkräften fördern zu können, wurde an der Pädagogischen Hochschule Heidelberg eine Seminarveranstaltung konzipiert und durchgeführt, die im Rahmen der hier vorzustellenden Interventionsstudie evaluiert wurde. Neben den Entwicklungen im Bereich des geschichtsdidaktischen Könnens für die Aufgabenkonstruktion fragen wir auch nach

¹ Diese Veröffentlichung wurde durch Sachbeihilfen des Landes Baden-Württemberg sowie der Pädagogischen Hochschule Heidelberg und Ludwigsburg im Rahmen des Forschungs- und Nachwuchskollegs «Effektive Kompetenzdiagnose in der Lehrerbildung (EKoL)» ermöglicht.

Veränderungen im Bereich des Fachwissens, denn es ist davon auszugehen, dass Reflexionen über die Konstruktion von Aufgaben den Aufbau von Wissen in den Bereichen des Geschichtswissens und des geschichtsdidaktischen Wissens (Heuer, Resch & Seidenfuß, 2017) unterstützen.

2 Theoretischer Hintergrund

2.1 Aufgaben konstruieren können als Facette geschichtsdidaktischen Könnens

Fähigkeiten, Fertigkeiten und Bereitschaften zur Beurteilung und Konstruktion von Aufgaben stellen zwei zentrale Bereiche der professionellen Kompetenz von Geschichtslehrkräften dar (Heuer, 2019; Thünemann, 2016). In aktuellen Modellierungen von Geschichtslehrer*innen-Kompetenzen werden Fähigkeiten zur Beurteilung und Konstruktion von Aufgaben übereinstimmend hervorgehoben. So wird im Modell der schweizerischen VisuHist-Studie das Wissen über kompetenzorientierte Aufgabenformate im «Wissen über instruktionale Aktivitäten zur Förderung des historischen Denkens und Lernens» gefasst (Waldis, 2019). Bei Lüke, Fenn und Seider (2018) wird dieser Wissensaspekt in die Facette «Wissen über das Potential fachspezifischer Lernmaterialien» integriert und im Heidelberger Modell wird das «Aufgaben formulieren können» als Facette des «Wissens über das Potential von Quellen und Darstellungen» postuliert (Heuer et al., 2017). In den kompetenztheoretischen Überlegungen bei Monte-Sano & Budona (2013) wird der Einsatz von angemessenen Aufgabenstellungen im Kompetenzbereich *Represent History* aufgeführt und auch für das geschichtsdidaktische Planungswissen werden Wissensaspekte für die Konstruktion von Aufgaben und Arbeitsaufträgen operationalisiert (z. B. Wolf, 2021).

In Bezug auf die theoriegeleitete Konzeptualisierung von Wissensaspekten zum Aufgaben-Konstruieren-Können ist mittlerweile ein recht klar definiertes Aufgabenwissen begründet, das unter anderem Wissen über textsortenspezifische Unterschiede zwischen Aufgaben und Arbeitsaufträgen oder Wissen über die Konstruktion von Aufgabenformaten für die Aufgabentypen Lern-, Leistungs- und Diagnoseaufgaben umfasst (Heuer & Resch, 2018).

2.2 Gütekriterien zur Beurteilung der Qualität von Aufgaben

Zur Beurteilung der Qualität von Aufgaben für den Geschichtsunterricht können geschichtsdidaktisch profilierte Gütekriterien für Lernaufgaben herangezogen werden. Auf den fachunspezifischen Kriterien von Blömeke et al. (2006) aufbauend, sollten nach Heuer (2011) gute Aufgaben für den Geschichtsunterricht «verständlich sein», «Operatoren beinhalten», «Offenheit aufweisen», für die Lernenden «herausfordernd sein», «differenziert sein», «zur Kooperation anregen» und als zentrales fachspezifisches Kriterium «zum historischen Erzählen» auffordern. Thünemann (2013) konkretisierte das von Heuer vorgeschlagene Kriterium des historischen Erzählens für den Lernprozess und ergänzte die Kriterien «Historische Leitfrage als Lernimpuls», «Historische Werturteilsbildung als Lernziel» und «Historische Reflexion als Lernevaluation». In einer aktualisierten Fassung des Katalogs präzisiert Köster (2021, 58–63) die Gütekriterien in Bezug auf den historischen Erkenntnisprozess und fügt dem Modell die zu beachtenden Kriterien «Heuristik als Ausgangspunkt», «Prozess- und Produktorientierung» und «Analyse und Synthese» hinzu. Bei allen Vorschlägen ist zu berücksichtigen, dass nicht alle Kriterien für jede Aufgabe gleichzeitig erfüllt werden können.

2.3 Kategoriensysteme und Operatoren zur Bestimmung der Aufgabenschwierigkeit

Lehrkräfte müssen ihre Lernangebote an unterschiedliche Wissens- und Kompetenzstände der Lernenden anpassen können. Hierfür müssen differenzierte, sich in ihrem Schwierigkeitsgrad unterscheidende Aufgaben angeboten werden. Um die Schwierigkeit von Aufgaben und die damit verbundenen kognitiven Operationen der Schüler*innen steuern zu können, müssen Geschichtslehrkräfte Kenntnisse über die fachspezifische Kategorisierung von Aufgaben für unterschiedliche Anforderungsbereiche anwenden und dazu passende Operatoren auswählen können.

Für den historischen Lernprozess liegen geschichtsdidaktische Konkretisierungen für die Niveaus der Anforderungsbereiche I bis III (Reproduktion, Reorganisation und Transfer, Transformation als Reflexion und Problemlösen) aus den «Einheitlichen Prüfungsanforderungen in der Abiturprüfung» (EPA, Kultusministerkonferenz, 2005) vor, die sich für den Bereich I als Aufgaben zur Quellenarbeit, für den Bereich II als Aufgaben zur Dekonstruktion historischer Erzählungen und für den Bereich III als Aufgaben zur Konstruktion und Reflexion

eigener Narrationen bestimmen lassen (Kühberger, 2011; Heuer, 2012; Waldis, 2013). Außerdem wurden mittlerweile auf der Grundlage des fachunspezifischen Kategoriensystems zur überfachlichen Aufgabenanalyse (Maier et al., 2013) unter anderem für die Konkretisierung der Wissensarten fachspezifische Analysekategorien vorgeschlagen. Aufgaben können in diesem Zusammenhang auf den Ebenen «inhaltlich», «methodisch», «sinnbildend» und «über Sinnbildungen reflektieren» entworfen werden (Resch & Seidenfuß, 2017).

In den fachlichen und überfachlichen Konkretisierungen wird übereinstimmend die Bedeutung der handlungsleitenden Operatoren hervorgehoben. Dabei besteht nach wie vor das Problem der «doppelten Operationalisierung» (Heuer, 2011), denn die genaue fachliche Handlungsaufforderung für die Schüler*innen ist in der Regel nicht definiert, was anhand von Untersuchungen zum Umgang mit dem Operator *beschreiben* empirisch nachgewiesen werden konnte (Oleschko, 2015). Auch wenn der Einsatz von Operatoren ohne fachliche Profilierung problematisch ist, wird im gemeinsamen Bildungsplan der Sekundarstufe I in Baden-Württemberg ein Operatorenkatalog vorgeschlagen, um die entsprechenden Anforderungsbereiche fokussieren zu können (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2016, 42–43).

2.4 Entwicklung geschichtsdidaktischen Könnens zur Aufgabenkonstruktionen

Obwohl die Konstruktion von Aufgaben als zentrale Lehrkompetenz postuliert wird und ein umfangreiches Aufgabenwissen theoriegeleitet entworfen wurde, konnten auf der Performanzebene die erwarteten Kompetenzentwicklungen für den zielgerichteten Einsatz von Aufgaben im Laufe des Professionalisierungsprozesses empirisch noch nicht hinreichend belegt werden (Resch & Heuer, 2019; Resch, Heuer & Lohse-Bossenz, 2019). Tendenziell scheinen Lehrkräfte vor allem mit dem Einsatz von Aufgaben auf höheren Niveaustufen überfordert zu sein. Eine schweizerische Videostudie zeigte, dass Lehrkräfte über die Konstruktion von Aufgaben kaum in der Lage waren, die Schüler*innen im eigenständigen historischen Denken zu unterstützen, sondern im Wesentlichen die Reproduktion und Festigung von deklarativem Wissen zum Ziel hatten (Waldis, Hodel & Fink, 2012).

3 Intervention zur Förderung des geschichtsdidaktischen Könnens zur Aufgabenkonstruktion

Zur Förderung des geschichtsdidaktischen Könnens und domänenspezifischen Fachwissens für die Konstruktion von Aufgaben wurde eine Seminarveranstaltung (Interventionsmaßnahme) konzipiert, die als Blockveranstaltung an sechs Terminen im Sommersemester 2018 an der Pädagogischen Hochschule Heidelberg durchgeführt wurde. Inhaltlich fokussierte die Intervention theoretische, empirische und pragmatische Aspekte zur Konstruktion qualitativvoller Aufgaben. Die Veranstaltung nutzte eine Kombination aus direkter Instruktion und Phasen des eigenständigen Erarbeitens und Übens, gemäß den empirisch geprägten Interventionsbausteinen *content focus* und *active learning* (Ingvarson, Meiers & Beavis, 2005).

Zu Beginn der Veranstaltung wurden die übergeordneten Ziele der Seminarveranstaltung bekannt gegeben, die folgendermaßen zusammengefasst werden können:

Die Studierenden ...

- ... kennen die Bedeutung der Formulierung und Beurteilung von Aufgaben für das kompetenzorientierte Unterrichten von Geschichte.
- ... kennen den aktuellen Stand der geschichtsdidaktischen Forschung zum Einsatz von Aufgaben und können empirische Befunde zur Konstruktion und Wirkung von Aufgaben auf die eigene Entwicklung und Beurteilung von Aufgabenformaten beziehen.
- ... können geschichtsdidaktisch profilierte Aufgabenformate (Diagnose-, Lern- und Leistungsaufgaben) zu ausgewählten Themenbereichen entwickeln.
- ... können vorliegende Aufgabenformate (Diagnose-, Lern- und Leistungsaufgaben) hinsichtlich ihrer Qualität beurteilen.
- ... können unterschiedliche Aufgabenformate in den Kontext einer umfassenden Lernumgebung integrieren.

Zu den einzelnen Terminen der Blockveranstaltung waren für die Seminar Teilnehmer*innen Lernaufgaben und Lernschritte zur Erlangung von Teilzielen vorgesehen. Tabelle 1 zeigt einen Überblick über die Gestaltung der einzelnen Veranstaltungen mit den entsprechenden Literaturbezügen. Auf der Grundlage von theoretischen Überlegungen, empirischen Befunden und Beispielen aus der Unterrichtspraxis zum Aufgabenwissen, zu allgemeinen und fachspezifischen

Gütekriterien für Aufgaben sowie zur Aufgabenkultur (Block 1 bis 3) stand die Konstruktion und Reflexion von Lernaufgaben (Block 4 bis 6) im Mittelpunkt der Lehrveranstaltung.

Tabelle 1: Überblick über den Interventionsablauf

Teilziele (pro Blockveranstaltung)	Lernaufgaben und Lernschritte	Literaturgrundlage
1. Organisation und Ziele des Seminars; Erhebung (t1); Einbettung des Themas in das Heidelberger Geschichtslehrerkompetenzmodell (HeiGeKo)	<ul style="list-style-type: none"> • Lektüre und Diskussion eines Einführungstextes 	Köster, Bernhard & Thünemann, 2016; Heuer et al., 2017
2. Aufgabentext und Kontext – Zur geschichtsdidaktischen Aufgabentheorie; Die Bedeutung von Aufgaben im kompetenzorientierten Geschichtsunterricht – Empirische Erkundungen;	<ul style="list-style-type: none"> • Diskussion «Aufgaben im kompetenzorientierten Geschichtsunterricht» • Diskussion zentraler empirischer Befunde • Zusammenführung: Zur Pragmatik der Aufgabenkonstruktion und Aufgabenbeurteilung 	Bürgler, Gautschi & Hediger, 2016 Bramann, 2017; Heuer, 2011; Heuer, 2012; Heuer, 2017; Heuer & Resch, 2018; Mägdefrau & Michler, 2014; Waldis, Hodel & Fink, 2012; Waldis, 2013
3. Gütekriterien zur Beurteilung und Konstruktion von Lernaufgaben; Zusammenführung: Zur Pragmatik der Aufgabenkonstruktion und Aufgabenbeurteilung; Konkretisierung der Seminaufgabe	<ul style="list-style-type: none"> • Entwicklung von Gütekriterien (Kleingruppen) • Präsentation und Diskussion zentraler Gütekriterien (Plenum) • Vorbereitungen für die Entwicklung von Lernaufgaben zu selbst gewählten Themen und Fragestellungen (Kleingruppen) 	Heuer, 2011; Thünemann, 2013; Köster, Bernhard & Thünemann, 2016; Resch & Seidenfuß, 2017; Dietrich 2018
4. Selbstständige Aufgabenentwicklung in Kleingruppen	<ul style="list-style-type: none"> • Entwicklung von Lernaufgaben zu selbst gewählten Themen und Fragestellungen (Kleingruppen) 	
5. Präsentation und Diskussion der selbst entwickelten Aufgabenformate	<ul style="list-style-type: none"> • Vorstellung, Analyse und Diskussion der selbst konstruierten Aufgaben anhand der Gütekriterien 	
6. Erhebung (t2); Roundtable und Abschluss der Veranstaltung		

4 Förderung von domänenspezifischem Fachwissen (Geschichtswissen und geschichtsdidaktisches Wissen) durch die Intervention

Für den Einsatz qualitätsvoller Aufgaben mit passenden Operatoren zur Bestimmung der Aufgabenschwierigkeit sind neben dem spezifischen Aufgabenwissen weitere Wissensaspekte nötig, die sich auf fachspezifische Operationen wie zum Beispiel die Re- oder De-Konstruktion von Quellen und Darstellungen beziehen. In diesem Zusammenhang wurden für die hier vorzustellende Studie Fachwissensfacetten aus dem Heidelberger Geschichtslehrerkompetenzmodell operationalisiert, die sich im Modell den Wissensbereichen Geschichtswissen (GW) und geschichtsdidaktisches Wissen (GDW) zuordnen lassen (Heuer et al., 2017). Ohne Kenntnisse über das Wesen und die Eigenheiten des Fachs (Geschichtswissen) und den Rückgriff auf etablierte geschichtsdidaktische Kategorien und Prinzipien (geschichtsdidaktisches Wissen) sind Reflexionen über fachspezifische Gütekriterien und fachlich profilierte Operatoren kaum zu realisieren, denn die eingesetzten Aufgaben zielen auf die von den Lernenden auszuführenden fachlichen Operationen mit dem Ziel, bestimmte historische (Teil-)Kompetenzen zu fördern. Mägdefrau und Michler (2014) untersuchten diesen Zusammenhang und berichten, dass Lehrkräfte das in Aufgaben angelegte Potenzial zur historischen Urteilsfähigkeit und Multiperspektivität nicht erkennen.

5 Zielstellung der vorliegenden Studie

Obwohl die Bedeutung von Aufgaben im Geschichtsunterricht unbestritten ist und Fähigkeiten, Fertigkeiten und Bereitschaften zur Konstruktion von qualitativollen Aufgaben als zentrale Lehrkompetenz hervorgehoben werden, deuten empirische Befunde darauf hin, dass (angehende) Geschichtslehrkräfte auf der Handlungsebene Schwierigkeiten mit dem Einsatz geeigneter Aufgaben haben. Um das geschichtsdidaktische Können für die Konstruktion von Aufgaben bei Studierenden fördern zu können, wurde die dargestellte Intervention mit dem Fokus auf Aufgaben im Geschichtsunterricht entwickelt, und an der Pädagogischen Hochschule Heidelberg durchgeführt.

Im Rahmen einer Interventionsstudie gilt es nun, die Wirksamkeit der Hochschulveranstaltung hinsichtlich des geschichtsdidaktischen Könnens zur Konstruktion von Aufgaben nachzuweisen. Dabei gehen wir davon aus, dass sich

bei den Proband*innen der Interventionsgruppe das geschichtsdidaktische Können zur Aufgabenkonstruktion entwickelt und in der Auseinandersetzung mit Aufgaben auch Geschichtswissen und geschichtsdidaktisches Wissen aufgebaut werden kann. Für die Proband*innen einer Kontrollgruppe, die ein geschichtsdidaktisches Seminar ohne Aufgabenbezug besuchen², wird erwartet, dass sich bedeutsame Veränderungen nur im Bereich des Fachwissens zeigen.

Weil Lerngelegenheiten von Lernenden vor dem Hintergrund individueller Voraussetzungen (u. a. Vorwissen und kognitive Voraussetzungen) unterschiedlich genutzt werden (Helmke, 2009), werden das Alter der Proband*innen, die Abiturnote als Indikator für die kognitiven Voraussetzungen, das Fachsemester, die bisher besuchten Seminare mit geschichtsdidaktischem Schwerpunkt sowie Praxiserfahrung als Außenkriterien für die Entwicklung des geschichtsdidaktischen Könnens sowie des Fachwissens untersucht. Bedeutsame Zusammenhänge werden für die Anzahl der besuchten Seminare mit geschichtsdidaktischen Schwerpunkten erwartet, denn in bisherigen Studien konnten für diese Variable positive Effekte auf Entwicklungen im Bereich des geschichtsdidaktischen Könnens berichtet werden, wohingegen für den Einfluss kognitiver Voraussetzungen negative und eher kleine Zusammenhänge berechnet wurden (Resch, Vollmer & Seidenfuß, 2017; Wolf et al., 2018). Universitären Praxisphasen wird insgesamt ein positiver Einfluss auf die Entwicklung fachdidaktischer Fähigkeiten bei angehenden Lehrpersonen zugeschrieben. Die Befunde aus der geschichtsdidaktischen Forschung zur Wirkung universitärer Praxisphasen divergieren allerdings und bilden vor allem in Bezug auf fachspezifische Kompetenzzuwächse ein Dilemma (Barsch & Plessow, 2020).

2 Die hier als Kontrollgruppe benannte Gruppe von Studierenden besuchte ebenfalls eine fachdidaktische Seminarveranstaltung (Blockseminar) mit dem Titel «Guter Geschichtsunterricht? Theoretische, methodische und empirische Annäherungen». Ausgehend von einer Wiederholung und Vertiefung zentraler geschichtsdidaktischer Konzepte wurde im Rahmen der Veranstaltung anhand theoretischer Überlegungen und empirischer Befunde der geschichtsdidaktischen Unterrichtsforschung die Planung, Durchführung und Reflexion von Geschichtsunterricht diskutiert. Die Konstruktion und Reflexion geeigneter Aufgaben spielten in den Seminarveranstaltungen eine untergeordnete Rolle.

6 Methoden

6.1 Stichprobe und Design

Um die Wirksamkeit der konzipierten Hochschulveranstaltung zu untersuchen, wurde ein Prä-Post-Vergleichsgruppendesign realisiert. Die Gesamtstichprobe setzt sich zusammen aus der Experimentalgruppe mit 22 Studierenden (63,6% weiblich; Alter: $M = 21.7$ Jahre), die an der Interventionsmaßnahme zur Förderung von Lehrkompetenzen für die Konstruktion von Aufgaben teilnahmen, und der Kontrollgruppe von 15 Studierenden (33,3% weiblich; Alter: $M = 23.0$ Jahre), die die Seminarveranstaltung mit fachdidaktischem Schwerpunkt ohne Aufgabenbezug belegte. Die Lehrveranstaltungen in der Experimental- und Kontrollgruppe wurden im Sommersemester 2018 als Blockseminare mit jeweils sechs Terminen à vier beziehungsweise sechs Stunden durchgeführt.

Die Befragung erfolgte für beide Gruppen jeweils in der ersten und letzten Sitzung der Seminarveranstaltungen. Für die Experimentalgruppe liegen zudem Daten einer eher explorativ angelegten Follow-up-Erhebung ($n = 5$, alle weiblich, Alter $M = 21.4$, Studiensemester $M = 5$) sechs Monate nach der Intervention vor. Prinzipiell wurden alle teilnehmenden Personen gebeten, sich auch an der Follow-up-Erhebung zu beteiligen. Die hohe Zahl an *Drop Outs* kann darauf zurückgeführt werden, dass die Befragung zum dritten Messzeitpunkt nicht im Rahmen der Seminarveranstaltungen durchgeführt wurde, sondern die Studierenden für einen zusätzlichen Termin an die Hochschule gebeten wurden. Außerdem kann die Verringerung der Stichprobe durch die hohe Beanspruchung bei der Bearbeitung des Fragebogens erklärt werden, die bereits in früheren Testungen abgefragt wurde (Resch, 2018, 132–133).

6.2 Testinstrumente

Die Wirksamkeit der Intervention wurde in Bezug auf das geschichtsdidaktische Können beim Erkennen geeigneter Aufgaben mit einem Vignettentest (Resch, 2018) erfasst und das dafür unter anderem benötigte geschichtsdidaktisch pro-

filierte Fachwissen wurde über einen klassischen Fachwissenstest (Resch & Heuer, 2019) abgefragt.³

6.2.1 Geschichtsdidaktisches Können beim Erkennen geeigneter Aufgaben

Im Vignettentest wurden den Proband*innen sieben fiktive Unterrichtssituationen präsentiert, die jeweils anschließend die Bewertung von sechs möglichen Handlungsalternativen im Sinne alternativer Aufgabenstellungen erfordern. Gefragt wurde, inwiefern die vorgeschlagenen Aufgabenvarianten geeignet sind, um bei den Lernenden einen Kompetenzzuwachs im Bereich der historischen Methodenkompetenz nach dem FUER-Kompetenzmodell (Schreiber, 2007) anbahnen zu können. Für den Test wurden Aufgabenvarianten wissensorientiert auf den Ebenen «inhaltlich», «methodisch», «sinnbildend» oder «über Sinnbildungen reflektieren» konstruiert (Resch & Seidenfuß, 2017). Die Einschätzung der Aufgabenvarianten erfolgte anhand einer sechsstufigen Likertskala («überhaupt nicht geeignet» bis «sehr gut geeignet»). Die Antworten der Proband*innen wurden mit einer Referenznorm (Expert*innennorm; $n = 9$) verglichen, als «übereinstimmend» beziehungsweise «nicht übereinstimmend» codiert und zu einem Summenwert zusammengefasst. Proband*innen, die den genauen, aus der Expert*innennorm generierten Skalenwert wählten, erhielten 1 Punkt, während Abweichungen um ± 1 vom Skalenwert mit 0.5 Punkten codiert wurden. Abweichende Beurteilungen von $> \pm 1$ Skalenwerte erhielten 0 Punkte. Weil in beiden Testteilen alle Missings unter 10 Prozent lagen, wurden fehlende Werte mit 0 Punkten codiert.⁴

-
- 3 Aufgrund der geringen Stichprobengröße können für die eingesetzten Instrumente keine verlässlichen Reliabilitätsmaße angegeben werden. Für eine Längsschnitterhebung mit größerer Stichprobe ($n = 171$) konnte die Modellgüte und starke partielle Messinvarianz über die Zeit für beide Testinstrumente nachgewiesen werden (Resch, Heuer & Lohse-Bossenz, 2019).
 - 4 Für eine ausführliche Darstellung der Testentwicklung und Testvalidierung anhand einer Beispielvignette vgl. Resch, 2018, 137–165.

6.2.2 Fachwissen – Geschichtswissen (GW) und Geschichtsdidaktisches Wissen (GDW)

Zur Erfassung des Fachwissens wurde ein Test mit insgesamt 57 Items eingesetzt. Für das Testinstrument wurden Wissensfacetten für das Geschichtswissen und das geschichtsdidaktische Wissen aus dem Heidelberger Geschichtslehrerkompetenzmodell (Heuer et al., 2017) operationalisiert. Im Bereich des Geschichtswissens (GW) wurden die Facetten «nature of history» (6 Items, z. B. «Quellen sind in der Regel das Ergebnis der Geschichtsschreibung») und «Geschichtstheoretisches Wissen» (16 Items, z. B. «Historisches Erzählen meint die sinnbildende Verknüpfung zeitdifferenter Ereignisse in Form einer Erzählung») erhoben. Für das geschichtsdidaktische Wissen (GDW) wurden die Facetten «nature of historical learning» (26 Items, z. B. «Rekonstruktionszeichnungen zeigen den Schüler*innen auf spannende und kindgerechte Art und Weise, wie es früher war») und «Geschichtskulturelles Wissen» (5 Items, z. B. «Geschichte begegnet uns heute in so vielen verschiedenen Medien und Gattungen. Dieses Ensemble ist als Geschichtskultur Teil der Lebenswelt») berücksichtigt. Die Antworten der Proband*innen wurden als «richtig» beziehungsweise «falsch» codiert und zu einem Summenwert zusammengefasst.

6.2.3 Selbsteinschätzung zum geschichtsdidaktischen Können

Neben dem Vignetten- und Fachwissenstest wurde außerdem eine Selbsteinschätzungsskala eingesetzt, über die auf einer vierstufigen Skala («kann ich überhaupt nicht» bis «kann ich sehr gut») der subjektiv erfahrene Kompetenzstand für die Konstruktion von Aufgaben eingeschätzt werden sollte. Die Skala setzt sich aus vier Items zusammen, die sich theoriebasiert auf zentrale Aspekte des Aufgabenwissens beziehen (siehe Abbildung 1).

Aufgrund meines bisherigen Geschichtsstudiums kann ich ...	kann ich überhaupt nicht			kann ich sehr gut
Aufgabenformate auf unterschiedlichen Niveaustufen formulieren.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Aufgaben einsetzen, die die Schüler*innen zur Formulierung eigener Narrationen anleiten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Operatoren für Aufgaben auswählen, die den Lernenden die Reflexion ihrer eigenen Sinnbildung ermöglichen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Aufgaben nach Diagnose, Leistungs- und Lernaufgaben unterscheiden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Abbildung 1: Selbsteinschätzung zum Aufgabenwissen

6.3 Datenanalyse

Um Analysen der Teildatensätze über mehrere Messzeitpunkte durchführen zu können, wurden die Teildatensätze anhand des persönlichen Codes, der zur Anonymisierung der Proband*innen generiert wurde, zusammengeführt. Entwicklungen über die Messzeitpunkte konnten zu diesen *Matched Data* mittels *t*-Tests für verbundene Stichproben im Statistikprogramm SPSS (IBM Corporation, 2013) für jede der beiden Gruppen untersucht werden (Welch-Test). Für die Vergleichbarkeit der Interventionseffekte zwischen den erhobenen Merkmalen wurden Effektstärken (Cohens *d*) berechnet.

Zur weiteren Untersuchung der Interventionseffekte wurde eine Varianzanalyse mit Messwiederholung (ANOVA) gerechnet. Das geschichtsdidaktische Können beim Erkennen geeigneter Aufgaben, das Fachwissen sowie die Selbsteinschätzung bildeten mit den zwei Messzeitpunkten den Innersubjektfaktor. Als Zwischensubjektfaktor wird die Gruppenzugehörigkeit aufgenommen. Ein bedeutsamer Effekt der Intervention zeigt sich in diesem Analyseansatz anhand einer signifikanten Interaktion zwischen Innersubjektfaktor und Zwischensubjektfaktor.

Ergänzend erfolgte für die betrachteten Merkmale die Bildung von Differenzwerten zwischen Prä- und Posterhebung: höhere Werte bedeuten dabei eine Steigerung des jeweiligen Merkmals und negative eine Verringerung im Verlaufe der jeweiligen Veranstaltungen. Diese Differenzwerte werden anschließend korrelativ in Zusammenhang mit dem Alter der Studierenden, der Abiturnote, des Fachsemesters, der vor der Intervention besuchten Hochschulveranstaltungen mit geschichtsdidaktischem Schwerpunkt und der Praxiserfahrungen (Außenkriterien) gebracht.

Auf einer explorativen Ebene wird außerdem untersucht, inwieweit sich verschiedene Facetten des geschichtlichen Fachwissens durch die Intervention unterschiedlich verändern. Hierfür werden für die vier betrachteten Facetten *t*-Tests für verbundene Stichproben innerhalb der Untersuchungsgruppen berechnet.

Zuletzt werden mit Blick auf langfristige Effekte individuelle Verläufe von fünf Personen betrachtet, die auch an der Follow-up-Erhebung sechs Monate nach der Veranstaltung teilgenommen haben. Hier wird auf deskriptiver Ebene untersucht, wie sich geschichtsdidaktisches Können und geschichtliches Fachwissen individuell entwickelt.

7 Ergebnisse

7.1 Vergleich Experimental- und Kontrollgruppe zu Beginn der Intervention

Um auszuschließen, dass sich Experimental- und Kontrollgruppe systematisch zu Beginn der Intervention unterscheiden, wurden diese hinsichtlich demografischer Variablen sowie der zentralen Zielkriterien verglichen. *t*-Tests für unabhängige Stichproben zeigten dabei für die demografischen Variablen Alter und Abiturnote keine signifikanten Unterschiede. Gleiches gilt hinsichtlich der studienspezifischen Vorerfahrungen (Anzahl Semester, Anzahl geschichtsdidaktischer Seminare), des geschichtsdidaktischen Wissens und Könnens sowie der Selbsteinschätzung zur Aufgabenkonstruktion ($.074 < p < .751$).

7.2 Effekte auf die Selbsteinschätzung geschichtsdidaktischen Könnens

Die Interventionsmaßnahme zeigt bei der Experimentalgruppe einen mittleren Effekt auf die Selbsteinschätzung bezüglich der Fähigkeit zur Konstruktion von Aufgaben ($M_s = 8.91$ vs. 10.55 , $SD_s = 2.04$ vs. 2.04 , $t(21) = -2.56$, $p = .018$, $d = .55$). Die Prüfung des Zusammenhangs zwischen Außenkriterien und Selbsteinschätzung ergab für die Experimentalgruppe einen schwachen und statistisch nicht signifikanten Einfluss der Abiturdurchschnittsnote ($r = .014$, $p = .949$), der Zahl der Studiensemester ($r = -.330$, $p = .133$), der Anzahl besuchter Seminare mit geschichtsdidaktischem Schwerpunkt ($r = -.312$, $p = .157$) sowie der schulpraktischen Erfahrung ($r = .082$, $p = .716$). Ein signifikanter Zusammenhang zeigte sich lediglich für das Alter der Proband*innen ($r = -4.75$, $p = .026$). Die Kontrollgruppe zeigte zwischen den beiden Messzeitpunkten keinen statistisch signifikanten Zuwachs in der Selbsteinschätzung bezüglich des geschichtsdidaktischen Könnens ($M_s = 9.64$ vs. 11.43 , $SD_s = 1.74$ vs. 2.41 , $t(14) = -2.76$, $p = .740$, $d = .45$). Hinsichtlich des Einflusses externer Variablen auf die Selbsteinschätzungen der Kontrollgruppe konnte eine statistisch signifikante Korrelation für die Schulpraxis ($r = .703$, $p = .005$) berechnet werden. Schwache und statistisch nicht signifikante Zusammenhänge zeigten sich für das Alter der Studierenden ($r = .246$, $p = .398$), die Abiturdurchschnittsnote ($r = -.271$, $p = .370$), das Studiensemester ($r = .083$, $p = .778$) und die Anzahl besuchter Seminare mit geschichtsdidaktischem Schwerpunkt ($r = .203$, $p = .486$).

7.3 Effekte auf das geschichtsdidaktische Können

Im Vignettentest konnten zum zweiten Erhebungszeitpunkt jeweils höhere Punktwerte für die Experimentalgruppe ($M_s = 32.64$ vs. 35.64 , $SD_s = 4.01$ vs. 4.20) und die Kontrollgruppe ($M_s = 33.20$ vs. 34.07 , $SD_s = 5.86$ vs. 4.13) festgestellt werden (siehe Abbildung 2). Für die Experimentalgruppe zeigte sich zwischen dem Prä- und Posttest ein mittlerer statistisch signifikanter Anstieg der Gesamtpunktwerte ($t(21) = -3.11$, $p = .005$, $d = .68$), während für die Kontrollgruppe keine signifikante Veränderung zwischen den beiden Messzeitpunkten zu verzeichnen war ($t(14) = -0.55$, $p = .590$, $d = .13$). Das von der Kontrollgruppe besuchte geschichtsdidaktische Seminar ohne konkreten Aufgabenbezug wirkte sich somit erwartungskonform nicht auf das geschichtsdidaktische Können aus. Der multivariate Ansatz über die Varianzanalyse mit Messwiederholung erbrachte keinen signifikanten Interaktionseffekt für die Entwicklung des geschichtsdidaktischen Könnens ($F(1) = 1.505$, $p = .23$, $\eta_p^2 = .041$). Hinsichtlich der Vorhersage der Differenzwerte zeigten sich für die Experimentalgruppe keine signifikanten Zusammenhänge mit dem Alter der Proband*innen ($r = .263$, $p = .237$), der Abiturdurchschnittsnote ($r = .236$, $p = .290$), dem Studiensemester ($r = .147$, $p = .515$), der Anzahl besuchter Seminare mit geschichtsdidaktischem Schwerpunkt ($r = .210$, $p = .349$) sowie den Erfahrungen in der Schulpraxis ($r = .124$, $p = .583$). Auch für die Kontrollgruppe ergab die Prüfung der Außenkriterien keine statistisch signifikanten Zusammenhänge für das Alter der Studierenden ($r = .183$, $p = .514$), die Abiturdurchschnittsnote ($r = -.077$, $p = .794$), das Studiensemester ($r = .191$, $p = .495$), die Anzahl besuchter Seminare mit geschichtsdidaktischem Schwerpunkt ($r = .181$, $p = .518$) sowie Erfahrungen in der Schulpraxis ($r = -.093$, $p = .742$).

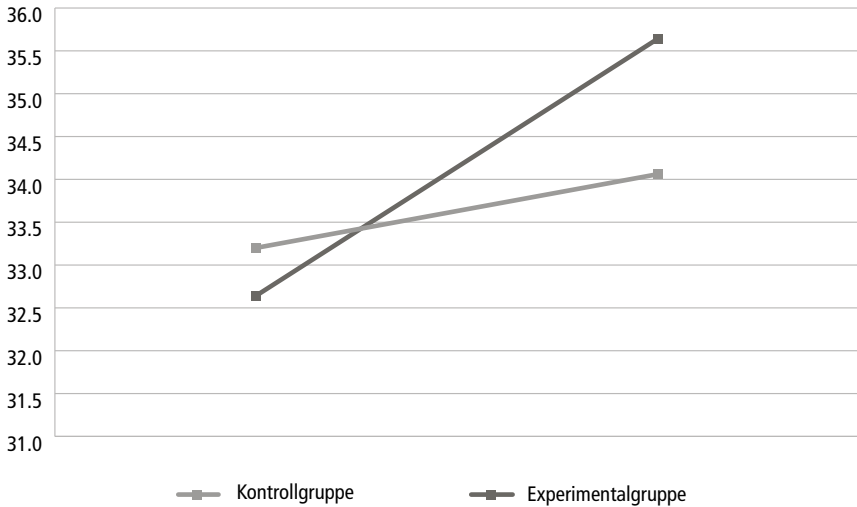


Abbildung 2: Geschichtsdidaktisches Können für die Konstruktion von Aufgaben im Prä-/Posttest der Experimental- und Kontrollgruppe

7.4 Effekte auf das geschichtliche Fachwissen

Im Bereich des Fachwissens kann zwischen dem ersten und zweiten Messzeitpunkt ebenfalls ein Anstieg der Punktwerte für die Experimentalgruppe ($M_s = 38.33$ vs. 44.09 , $SD_s = 4.62$ vs. 3.33) und die Kontrollgruppe ($M_s = 37.47$ vs. 41.60 , $SD_s = 4.98$ vs. 3.89) berichtet werden (siehe Abbildung 3). Im Vergleich der Mittelwerte sind die Veränderungen im Fachwissen für die Experimentalgruppe ($t(21) = -5.94$, $p = <.001$, $d = 1.15$) und die Kontrollgruppe ($t(14) = -4.08$, $p = .001$, $d = .91$) statistisch signifikant. Starke Effekte können sowohl für die Teilnehmer*innen der Seminarveranstaltung zur Konstruktion von Aufgaben (Interventionsmaßnahme), als auch für die Teilnehmer*innen der Veranstaltung ohne Aufgabenschwerpunkt (Kontrollgruppe) gezeigt werden. Die Varianzanalyse bestätigt die gleichsame Veränderung in der Experimental- und Kontrollgruppe und weist keinen signifikanten Interaktionseffekt zwischen Inner- und Zwischensubjektfaktor auf ($F(1) = 1.290$, $p = .264$, $\eta^2_p = .037$). Hinsichtlich der Zusammenhänge zwischen den Differenzwerten des Fachwissens und den Außenkriterien ergaben sich für die Experimentalgruppe keine signifikanten Effekte des Alters der Studierenden ($r = -.072$, $p = .756$), der Abiturdurchschnittsnote ($r = .060$, $p = .797$),

des Studiensemesters ($r = -.106, p = .647$), der Anzahl besuchter Seminare mit geschichtsdidaktischem Schwerpunkt ($r = .205, p = .374$) sowie der schulpraktischen Erfahrung ($r = .030, p = .896$). Auch für die Kontrollgruppe konnten keine statistisch signifikanten Zusammenhänge für das Alter der Studierenden ($r = -.409, p = .130$), die Abiturdurchschnittsnote ($r = -.223, p = .444$), das Studiensemester ($r = -.385, p = .157$), die Anzahl besuchter Seminare mit geschichtsdidaktischem Schwerpunkt ($r = -.459, p = .085$) sowie die Erfahrungen in der Schulpraxis ($r = -.280, p = .313$) festgestellt werden.

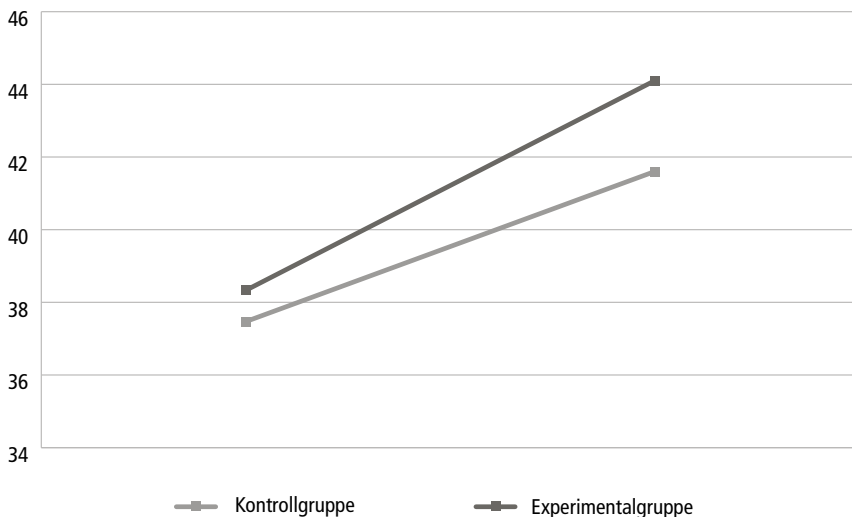


Abbildung 3: Fachwissen im Prä-/Posttest der Experimental- und Kontrollgruppe

7.5 Effekte auf einzelne Facetten des Fachwissens

Tabelle 2 zeigt die Entwicklung in den einzelnen Fachwissensfacetten zwischen den beiden Messzeitpunkten. Für die Experimentalgruppe können signifikante Veränderungen in den Bereichen «nature of historical learning», «Geschichtstheoretisches Wissen» und «Geschichtskulturelles Wissen» berichtet werden. Keine statistisch signifikanten Entwicklungen zeigten sich für die Facette «nature of history». Für die Kontrollgruppe konnten signifikante Veränderungen für den Bereich «nature of historical learning» berechnet werden. Demgegenüber gab es

keine signifikanten Entwicklungen in den Facetten «nature of history», «Geschichtstheoretisches Wissen» und «Geschichtskulturelles Wissen».

Tabelle 2: Fachwissensfacetten im Prä-/Posttest der Experimental- und Kontrollgruppe

Wissensfacette	Gruppe	Prätest		Posttest		t(21)	p	d
		M	SD	M	SD			
«nature of historical learning»	EG	17.68	2.68	20.55	2.72	-4.77	<.001	1.03
	KG	16.67	3.85	19.27	2.34	-3.10	.008	.71
«Geschichtstheoretisches Wissen»	EG	9.77	1.85	12.0	1.63	-4.99	<.001	1.01
	KG	9.87	1.36	10.93	2.34	-1.79	.096	.66
«Geschichtskulturelles Wissen»	EG	3.57	.98	4.14	.96	-2.17	.042	.47
	KG	3.6	.74	3.6	.63	.00	1.0	.00
«nature of history»	EG	4.36	.85	4.36	.95	.00	1.0	.00
	KG	4.27	.96	4.53	.92	-1.0	.334	.25

Anm. EG = Experimentalgruppe, KG = Kontrollgruppe; Erreichbare Punktzahl: «nature of historical learning»: 0–26; «Geschichtstheoretisches Wissen»: 0–16; «Geschichtskulturelles Wissen»: 0–5; «nature of history»: 0–6.

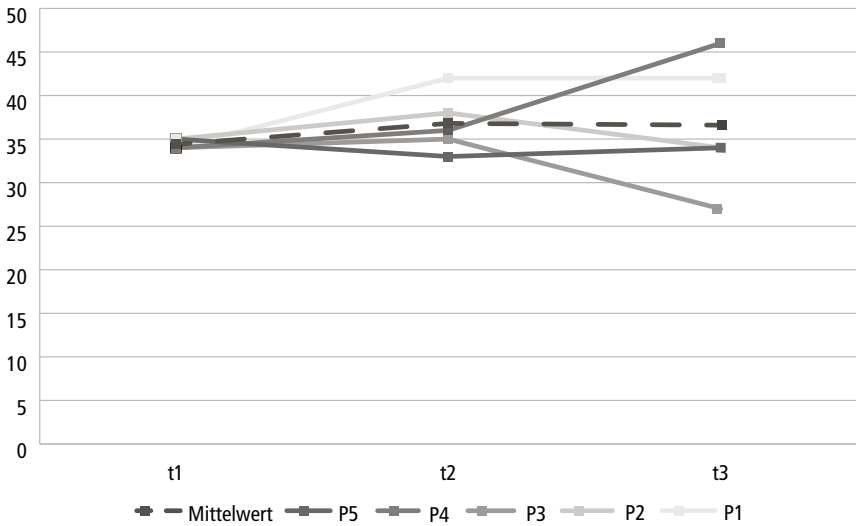
7.6 Langfristige Wirkungen der Intervention

Die Proband*innen der Follow-up-Erhebung ($n = 5$) erreichten zwischen dem zweiten und dritten Messzeitpunkt nahezu die gleichen Punktwerte im Vignettest ($M_s = 36.8$ vs. 36.6 , $SD_s = 3.42$ vs. 7.47) und im Fachwissenstest ($M_s = 44.0$ vs. 44.0 , $SD_s = 3.24$ vs. 4.81). Im Vergleich der Mittelwerte wurden keine statistisch bedeutsamen Veränderungen für das geschichtsdidaktische Können ($t(4) = .07$, $p = .95$, $d = -.055$) und das Fachwissen ($t(4) = .00$, $p = 1.0$, $d = .0$) festgestellt. Die Proband*innen halten in beiden Bereichen das zum zweiten Messzeitpunkt erreichte Niveau.

Allerdings zeigen sich bei der Entwicklung im geschichtsdidaktischen Können und im Fachwissen große individuelle Unterschiede innerhalb der befragten Gruppe. In Abbildung 4 sind neben den Mittelwerten (gestrichelte Linie) die erreichten Summenwerte für die fünf Proband*innen der Follow-up-Erhebung für alle drei Messzeitpunkte (t_1 , t_2 , t_3) als graue Linien abgebildet. Während die berechneten Ausgangswerte im Vignettest zum ersten Erhebungszeitpunkt recht ausgeglichen sind, unterscheiden sich die Ausgangswerte im Fachwissenstest zum ersten Messzeitpunkt um bis zu 15 Punkte. Die erwarteten progressiven

Verläufe über die drei Messzeitpunkte sind nur vereinzelt für das Fachwissen bei den Proband*innen P5 ($M_s = 32$ vs. 41 vs. 41) und P2 ($M_s = 43$ vs. 48 vs. 53) und für das geschichtsdidaktische Können bei P1 ($M_s = 34$ vs. 42 vs. 42) und P4 ($M_s = 34$ vs. 36 vs. 46) zu verzeichnen. Demgegenüber sind stagnierende Entwicklungen im Fachwissen bei den Studierenden P1 ($M_s = 40$ vs. 42 vs. 41), P3 ($M_s = 47$ vs. 47 vs. 47) und P4 ($M_s = 42$ vs. 42 vs. 40) zu erkennen. Im geschichtsdidaktischen Können zeigen sich zu vernachlässigende Veränderungen bei den Studierenden P2 ($M_s = 35$ vs. 38 vs. 34) und P5 ($M_s = 35$ vs. 33 vs. 34). Dass Proband*innen zum dritten Messzeitpunkt weniger Punktwerte erreichen als zum ersten Messzeitpunkt zeigt sich für das geschichtsdidaktische Können zum Erkennen geeigneter Aufgaben bei den Proband*innen P2 ($M_s = 35$ vs. 38 vs. 34), P3 ($M_s = 34$ vs. 35 vs. 27) und P5 ($M_s = 35$ vs. 33 vs. 34). Der Wissenszuwachs im Bereich des Fachwissens scheint nachhaltiger zu sein, denn hier ist zum dritten Erhebungszeitpunkt kein «Rückfall» unter den in der ersten Befragung erreichten Ausgangswert zu verzeichnen.

Geschichtsdidaktisches Können



Geschichtsdidaktisches Fachwissen

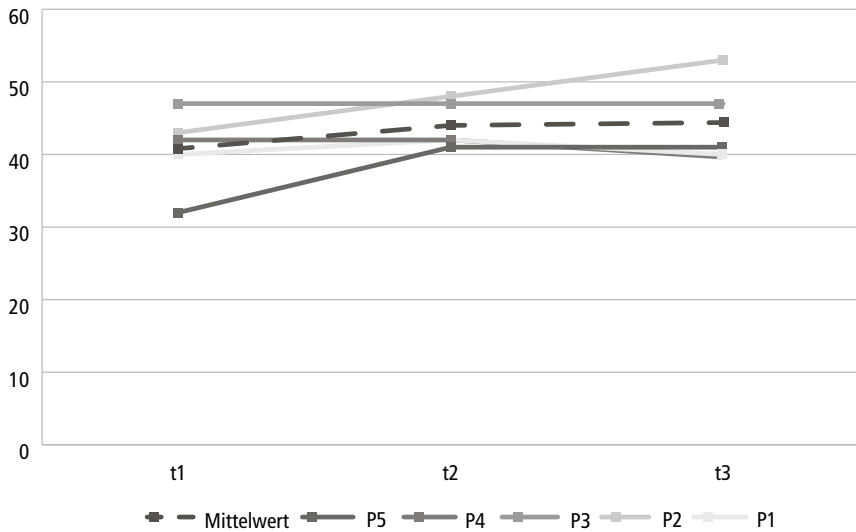


Abbildung 4: Punktwerte der Follow-up-Kohorte zu drei Messzeitpunkten (Anmerkung: gestrichelte Linie = Mittelwertverlauf)

8 Diskussion

Ausgehend von der Annahme, dass die Konstruktion von Aufgaben für angehende Lehrkräfte eine Herausforderung darstellt, wurde eine Seminarveranstaltung zur Förderung geschichtsdidaktischer Kompetenzen für die Konstruktion von Aufgaben entwickelt und auf ihre Wirksamkeit hin untersucht. Wir sind neben den Zuwächsen im geschichtsdidaktischen Können außerdem davon ausgegangen, dass die Auseinandersetzung mit Aufgaben auch zu Zuwächsen im Fachwissen führt. Auch wenn die Befunde aufgrund der geringen Stichprobengröße in ihrer Generalisierbarkeit eingeschränkt sind, zeigen die Ergebnisse, dass sich die konkrete Interventionsmaßnahme spezifisch auf das geschichtsdidaktische Können auswirkte: Nicht nur, dass die Studierenden einen signifikanten Zuwachs an Fähigkeiten, Fertigkeiten und Bereitschaften zur Konstruktion von Aufgaben *wahrnehmen*, sondern dass diese darüber hinaus auch eine signifikante Entwicklung des geschichtsdidaktischen Könnens für das Erkennen geeigneter Aufgaben in einem standardisierten Test *zeigen*, belegt die Wirksamkeit der universitären Lerngelegenheit.

Ebenfalls den Erwartungen entsprechend sind für das Fachwissen statistisch signifikante Wissenszuwächse bei der Experimentalgruppe und der Kontrollgruppe festgestellt worden. Der bei der Experimentalgruppe festgestellte Wissenszuwachs vor allem in den Bereichen «Geschichtstheoretisches Wissen» und «nature of historical learning» deutet darauf hin, dass die in der Intervention angelegte Auseinandersetzung mit Aufgabenkonstruktionen den Aufbau von Wissen in diesen Bereichen des Geschichtswissens und des Geschichtsdidaktischen Wissens in besonderem Maße fördert. Der Befund kann unter Umständen darauf zurückgeführt werden, dass die Teilnehmer*innen an der Seminarveranstaltung eigene Aufgaben entwickelten, dazu passende Quellen und Darstellungen auswählten und dieses Arrangement aus Lernaufgabe und Materialien im historischen Lernprozess verorteten. Dabei spielten auch Überlegungen zur Aufgabenkultur (Heuer, 2012; Köster, 2021, 85–101), die neben dem Aufgabentext auch die Einführung, Bearbeitung und Reflexion der Aufgabe im historischen Lernprozess als Aufgabenkontext vorsieht, eine entscheidende Rolle. Die in diesem Setting durchgeführten Reflexionen über den Einsatz von Aufgaben scheinen bestimmte Fachwissensbereiche über den Aufgabenzusammenhang hinaus fördern zu können.

Die Nachhaltigkeit der Interventionsmaßnahme wurde durch eine Follow-up-Erhebung zu einem dritten Messzeitpunkt ca. sechs Monate nach der Inter-

vention bei fünf Proband*innen betrachtet. Aufgrund der geringen Stichprobengröße hat die Follow-up-Erhebung allerdings eher explorativen Charakter. Betrachtet man die erreichten Punktwerte dieser Kohorte über alle drei Messzeitpunkte, sind für das geschichtsdidaktische Können und das Fachwissen individuell sehr unterschiedliche Entwicklungsverläufe zu erkennen. Dieser Befund entspricht der Annahme, dass sich Wissen und Kompetenzen nicht linear entwickeln und Lerngelegenheiten zur Professionalisierung individuell genutzt werden (Hascher, 2014; Lohse-Bossenz & Unger, 2021; Heuer, Körber, Schreiber & Waldis, 2019). Die Nutzung der Intervention als Lerngelegenheit auf individueller Ebene kann anhand von Einzelfallstudien weiter untersucht werden, denn neben den Daten aus den geschlossenen Antwortformaten liegen Kommentare der Proband*innen zu den im Vignettentest eingesetzten Aufgaben (Items) vor. Durch die Auswertung der Kommentare könnten Veränderungen im «geschichtsdidaktischen Blick» (Seidenfuß, 2018) auf Aufgaben zu unterschiedlichen Zeitpunkten der Interventionsstudie analysiert und Rückschlüsse über die Wirksamkeit bestimmter Lernangebote auf individueller Ebene generiert werden.

Zusammenfassend zeigt die vorgestellte Studie, dass das geschichtsdidaktische Können zur Konstruktion von Aufgaben gefördert werden kann, ohne dass hierdurch Einbußen hinsichtlich des Aufbaus fachlichen Wissens zu verzeichnen sind. Neben der Replikation der Befunde erscheint es zielführend, individuelle Lernprozesse zu beschreiben und weiterführend zu untersuchen, wie sich dieses geschichtsdidaktische Können auf die tatsächliche Orchestrierung von Aufgaben im geschichtsunterrichtlichen Handeln auswirkt.

Literatur

- Barsch, Sebastian & Plessow, Oliver (2020). Einleitung. Geschichtsdidaktische Perspektiven auf universitäre Praxisphasen im Lehramtsstudium. In Sebastian Barsch & Oliver Plessow (Hrsg.), *Universitäre Praxisphasen im Fach Geschichte: Wege zu einer Verbesserung der Lehramtsausbildung?* (S. 1–8). Münster: LIT
- Blömeke, Sigrid/Risse, Jana/Müller, Christine/Eichler, Dana & Schulz, Wolfgang (2006). Analyse der Qualität von Aufgaben aus didaktischer und fachlicher Sicht. *Unterrichtswissenschaft*, 34(4), 330–357.
- Bramann, Christoph (2017). Historisch Denken lernen mit Schulbuchaufgaben?: Medienspezifische Analyse von Arbeitsaufträgen in österreichischen Geschichtsschulbüchern. In Christoph Bramann/Christoph Kühberger & Roland Bernhard (Hrsg.), *Historisch Denken lernen mit Schulbüchern* (S. 181–214). Schwalbach/Ts.: Wochenschau.
- Bürgler, Beatrice/Gautschi, Peter & Hediger, Stephan (2016). Kompetenzorientiert arbeiten im Geschichtsunterricht. In Marcel Naas (Hrsg.), *Kompetenzorientierter Unterricht auf*

- der Sekundarstufe I: Erziehungswissenschaftliche und fachdidaktische Perspektiven (S. 257–279). Bern: hep.
- Dietrich, Tobias (2018). Lernaufgaben. *Praxis Geschichte*, 2, I–IV.
- Fenn, Monika (2013). Konstruktivistisches Geschichtsverständnis im Unterricht fördern: Studierende ändern ihr Lehrverhalten von einseitig instruktional in problemorientiert. In Jan Hodel/Monika Waldis & Béatrice Ziegler (Hrsg.), *Forschungswerkstatt Geschichtsdidaktik 12. Beiträge zur Tagung «geschichtsdidaktik empirisch 12»* (S. 60–71). Bern: hep.
- Hascher, Tina (2014). Forschung zur Wirksamkeit der Lehrerbildung. In Ewald Terhart/Hedda Bennewitz & Martin Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 542–571). Münster: Waxmann.
- Helmke, Andreas (2009). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts*. Stuttgart: Klett.
- Heuer, Christian (2011). Gütekriterien für kompetenzorientierte Lernaufgaben im Fach Geschichte. *Geschichte in Wissenschaft und Unterricht*, 62, 443–458.
- Heuer, Christian (2012). Zur Aufgabenkultur im Geschichtsunterricht. In Ute Bender & Stefan Keller (Hrsg.), *Aufgabenkulturen: Fachliche Lernprozesse herausfordern, begleiten, reflektieren* (S. 100–112). Seelze: Kallmeyer.
- Heuer, Christian (2017). Vorschläge, Aufträge, Aufgaben? Zum Wandel von Aufgaben im Schulgeschichtsbuch. *Erziehung und Unterricht. Österreichische Zeitschrift für Pädagogik*, 9(10), 935–944.
- Heuer, Christian (2019). Gute Aufgaben?! Plädoyer für einen geschichtsdidaktischen Perspektivenwechsel. In Christoph Kühberger, Roland Bernhard & Christoph Bramann (Hrsg.), *Das Geschichtsschulbuch: Lernen – Lehren – Forschen* (S. 147–160). Münster: Waxmann.
- Heuer, Christian (2021). Kompetenzorientierte Planungsmodelle für den Geschichtsunterricht. In Georg Weißeno & Béatrice Ziegler (Hrsg.), *Handbuch Geschichts- und Politikdidaktik*. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-29673-5_13-1.
- Heuer, Christian/Körper, Andreas/Schreiber, Waltraud & Waldis, Monika (2019). GeDiKo – Professionstheoretische Überlegungen zur Modellierung geschichtsdidaktischer Kompetenzen. *Zeitschrift für Geschichtsdidaktik*, 17, 99–113.
- Heuer, Christian & Resch, Mario (2018). Aufgaben im Geschichtsunterricht: Stand, Desiderate, Perspektiven. *Geschichte für heute*, 11(3), 5–23.
- Heuer, Christian/Resch, Mario & Seidenfuß, Manfred (2017). Geschichtslehrerkompetenzen? Wissen und Können geschichtsdidaktisch. *Zeitschrift für Didaktik der Gesellschaftswissenschaften*, 7(2), 158–176.
- IBM Corporation (2013). *IBM SPSS Statistics for Windows*. Armonk: IBM Corp.
- Ingvarson, Lawrence/Meiers, Marion & Beavis, Adrian (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes & efficacy. *Education Policy Analysis Archives*, 13(10). <https://doi.org/10.14507/epaa.v13n10.2005>.
- Köster, Manuel (2021). *Aufgabenkultur im Geschichtsunterricht*. Frankfurt a. M.: Wochenschau.

- Köster, Manuel/Bernhardt, Markus & Thünemann, Holger (2016). Aufgaben im Geschichtsunterricht: Typen, Gütekriterien und Konstruktionsprinzipien. *Geschichte lernen*, 29, 2–11.
- Kühberger, Christoph (2011). Operatoren als strukturierende Elemente von Aufgabenstellungen für Geschichte und Sozialkunde, Politische Bildung. In Bundesministerium für Frauen und Bildung (Hrsg.), *Geschichte und Sozialkunde, Politische Bildung: Richtlinien und Beispiele für Themenpool und Prüfungsaufgaben* (S. 15–20). Wien: Bundesministerium für Bildung und Frauen.
- Kultusministerkonferenz (2005). Einheitliche Prüfungsanforderungen in der Abiturprüfung. Geschichte. Beschluss der Kultusministerkonferenz vom 01.12.1989 i. d. F. vom 10.02.2005. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1989_12_01-EPA-Geschichte.pdf (07.04.2022).
- Lohse-Bossenz, Hendrik & Unger, Valentin (2021). Veränderung professionellen Wissens über verschiedene Phasen der Lehrerbildung hinweg – Metaanalytische Integration der Befunde aus dem Forschungsprogramm «Effektive Kompetenzdiagnose in der Lehrerbildung». In Hendrik Lohse-Bossenz/Markus Rehm/Marita Eva Friesen/Mannfried Seidenfuß/Juliane Rutsch/Markus Vogel & Tobias Dörfler (Hrsg.), *Professionalisierung in der Lehrerbildung – Erkenntnisse und Perspektiven des interdisziplinären Forschungsprogramms «Effektive Kompetenzdiagnose in der Lehrerbildung»* (S. 198–215). Münster: Waxmann.
- Lüke, Nicole/Seider, Jessica & Fenn, Monika (2018). Struktur und Inhalt des fachbezogenen Professionswissens angehender Lehrkräfte in den Geisteswissenschaften. *heEDUCATION Journal*, (1/2), 75–98.
- Mägdefrau, Jutta & Michler, Andreas. (2014). Arbeitsaufträge im Geschichtsunterricht: Diskrepanz zwischen Lehrerintention und didaktischem Potenzial? In Bernd Ralle, Susanne Prediger, Marcus Hammann & Martin Rothgangel (Hrsg.), *Lernaufgaben entwickeln, bearbeiten und überprüfen: Ergebnisse und Perspektiven fachdidaktischer Forschung* (S. 105–119). Münster: Waxmann.
- Maier, Uwe/Bohl, Thorsten/Kleinknecht, Marc & Metz, Kerstin (2013). Allgemeindidaktische Kriterien für die Analyse von Aufgaben. In Marc Kleinknecht/Thorsten Bohl/Uwe Maier & Kerstin Metz (Hrsg.), *Lern- und Leistungsaufgaben im Unterricht: Fächerübergreifende Kriterien zur Auswahl und Analyse* (S. 9–46). Bad Heilbrunn: Klinkhardt.
- Ministerium für Kultus, Jugend und Sport (Hrsg.) (2016). *Gemeinsamer Bildungsplan der Sekundarstufe 1: Geschichte*. Stuttgart: Neckar.
- Oleschko, Sven (2015). Herausforderungen einer domänenspezifischen Sprachdiagnostik. *Zeitschrift für Geschichtsdidaktik*, 14, 87–103.
- Resch, Mario (2018). *Aufgaben formulieren können: Entwicklung und Validierung eines Vignetentests zur Erfassung professioneller Kompetenz für historisches Lehren*. Frankfurt a. M.: Wochenschau.
- Resch, Mario (2021). Professionelle Kompetenz von Geschichtslehrkräften. In Georg Weiseno & Béatrice Ziegler (Hrsg.), *Handbuch Geschichts- und Politikdidaktik*. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-29673-5_18-1.
- Resch, Mario & Heuer, Christian (2019). Fachwissen und geschichtsdidaktisches Wissen und Können bei Lehramtsanwärterinnen und Lehramtsanwärtern zu Beginn der

- zweiten Ausbildungsphase. In Monika Waldis & Béatrice Ziegler (Hrsg.), *Forschungswerkstatt Geschichtsdidaktik 17. Beiträge zu Tagung «geschichtsdidaktik empirisch 17»* (S. 107–118). Bern: hep.
- Resch, Mario/Heuer, Christian & Lohse-Bossenz, Hendrik (2019). Zur Entwicklung von Fachwissen und geschichtsdidaktischem Können während des Referendariats: Ergebnisse einer Längsschnittstudie zum Professionalisierungsprozess. *Zeitschrift für Geschichtsdidaktik*, 18, 61–77.
- Resch, Mario & Seidenfuß, Manfred (2017). Fachdidaktische Kompetenzen angehender Geschichtslehrkräfte beim Formulieren von Lernaufgaben: Theoretische Beschreibung und empirische Erfassung mit einem Vignettest. In Juliane Rutsch/Markus Rehm/Markus Vogel/Manfred Seidenfuß & Tobias Dörfler (Hrsg.), *Effektive Kompetenzdiagnose in der Lehrerbildung: Professionalisierungsprozesse angehender Lehrkräfte untersuchen* (S. 115–128). Wiesbaden: Springer VS.
- Resch, Mario/Vollmer, Christian & Seidenfuß, Manfred (2017). Ein Vignettest mit geschlossenem Antwortformat zur Erfassung fachdidaktischer Kompetenzen bei angehenden Geschichtslehrkräften. In Béatrice Ziegler & Monika Waldis (Hrsg.), *Forschungswerkstatt Geschichtsdidaktik 15. Beiträge zu Tagung «geschichtsdidaktik empirisch 15»* (S. 163–176). Bern: hep.
- Schreiber, Waltraud (2007). Kompetenzbereich historische Methodenkompetenz. In Andreas Körber/Waltraud Schreiber & Alexander Schöner (Hrsg.), *Kompetenzen historischen Denkens: Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik* (S. 194–235). Neuried: ars una.
- Seidenfuß, Manfred (2018). Adressatenorientierung im geschichtsdidaktischen Blick. In Christian Kuchler & Andreas Sommer (Hrsg.), *Wirksamer Geschichtsunterricht* (S. 173–183). Baltmannsweiler: Schneider Hohengehren.
- Thünemann, Holger (2013). Historische Lernaufgaben: Theoretische Überlegungen, empirische Befunde und forschungspragmatische Perspektiven. *Zeitschrift für Geschichtsdidaktik*, 12, 141–155.
- Thünemann, Holger (2016). Probleme und Perspektiven der geschichtsdidaktischen Kompetenzdebatte. In Saskia Handro & Bernd Schönemann (Hrsg.), *Aus der Geschichte lernen? Weiße Flecken der Kompetenzdebatte* (S. 37–51). Berlin: Lit.
- Waldis, Monika (2013). Fachdidaktische Analyse von Aufgaben in Geschichte. In Marc Kleinknecht/Thorsten Bohl/Uwe Maier & Kerstin Metz (Hrsg.), *Lern- und Leistungsaufgaben im Unterricht: Fächerübergreifende Kriterien zur Auswahl und Analyse* (S. 145–162). Bad Heilbrunn: Klinkhardt.
- Waldis, Monika (2019). Professionelle Unterrichtswahrnehmung als Annäherung an fachdidaktisches Wissen von angehenden Geschichtslehrpersonen. In Waltraud Schreiber/Béatrice Ziegler & Christoph Kühberger (Hrsg.), *Geschichtsdidaktischer Zwischenhalt: Beiträge aus der Tagung «Kompetent machen für ein Leben in, mit und durch Geschichte» in Eichstätt vom November 2017* (S. 223–241). Münster, New York: Waxmann.
- Waldis, Monika/Hodel, Jan & Fink, Nadine (2012). Lernaufgaben im Geschichtsunterricht und ihr Potential zur Förderung historischer Kompetenzen. *Zeitschrift für Didaktik der Gesellschaftswissenschaften*, 3(1), 142–157.

- Wolf, Jörgen (2021). Geschichtsunterricht planen können. Ein Vorschlag zur handlungsnahen fachspezifischen Kompetenz und Wissensmodellierung am Beispiel der Unterrichtsplanungskompetenz und des Unterrichtsplanungswissens von Geschichtslehrpersonen. *Geschichte in Wissenschaft und Unterricht*, 72(1/2), 58–74.
- Wolf, Jörgen/Seiffert, Joana/Seifert, Andreas/Rothland, Martin & Brauch, Nicola (2018). Das geschichtsdidaktische Planungswissen von angehenden Geschichtslehrer/innen: Entwicklung und Pilotierung eines Testinstruments zur Messung des fachdidaktischen Planungswissens von Lehramtsstudierenden im Fach Geschichte. *Geschichte in Wissenschaft und Unterricht*, 69(7/8), 373–392.

Qualitative «Interventionen»

Multiperspektivität schriftlich erarbeiten

Mit Textprozeduren erzählen lernen

Georg Marschnig

1 Einleitung: «Geschichte ist nicht ‹an sich› schon da»

«Geschichte ist nicht ‹an sich› schon da, wir müssen sie vielmehr erst selbst (aktiv) zu einem für uns bedeutungsvollen Gewebe (Text) verknüpfen.» Olaf Hartung (2013, 9) spricht im Eröffnungssatz seiner 2013 veröffentlichten Habilitationsschrift eine ebenso simple, wie fundamentale Wahrheit der Geschichtsforschung aus: Geschichte als «bedeutungsvolles Gewebe» ist das (vorläufige) Ergebnis eines aktiven und un abgeschlossenen (Erzähl-)Prozesses. Geschichte wird geschrieben, sie wird von jemandem produziert. Dadurch wird eine weitere zentrale geschichtstheoretische Wahrheit eröffnet, die auch für den vorliegenden Text und die ihm zugrundeliegende Studie von großer Relevanz ist: «Jede Geschichte ist perspektivisch» (Rüsen, 1983, 100). Da dieses Produkt aktiv erstellt, ist es an die Ideen, Absichten und Hintergründe der Erzählenden gebunden. Und wenn Geschichte als «unendliches, niemals abreißendes Nachdenken über vergangenes menschliches Handeln und Leiden, das sich Orientierungsbedürfnissen einer jeweiligen Gegenwart gegenüber einer erwarteten oder gefürchteten Zukunft verdankt» (Bergmann, 2007, 28) eben immer an diejenigen geknüpft ist, die sie erzählen, ist sie auch grundlegend von deren Fähigkeiten abhängig. Die deutschsprachige Geschichtsdidaktik bezeichnet sie als historische Kompetenzen – «inextricably linked with the concept of historical consciousness» (Kölbl & Konrad, 2015, 17) – und hat sie in verschiedenen Modellen formuliert und breit diskutiert (Gautschi, 2009; Körber, Schreiber & Schöner, 2007; Pandel, 2005). Die Aufmerksamkeit für die sprachlichen Performanzen dieser Kompetenzen war zwar von Beginn an in der Debatte angelegt, entwickelte sich aber besonders intensiv erst im Laufe der 2010er-Jahre. Wurde noch 2010 die systematische «Marginalisierung dieses Grundlagenproblems» (Handro, 2010, 10) kritisiert und verstärkte Forschungstätigkeit eingefordert, so lässt sich heute ein verstärktes Bewusstsein für Sprache als «Maß aller Dinge» (Grannemann & Kuchler, 2019, 31)

im Geschichtsunterricht feststellen, das sich auch an der wachsenden Zahl von Untersuchungen zum Thema ablesen lässt. (Sandkühler & Bernhardt, 2020)

Eine dieser Forschungsarbeiten, das Habilitationsprojekt «Multiperspektivität als sprachliche Herausforderung», soll im vorliegenden Beitrag präsentiert werden. Dafür wird zunächst das theoretische Fundament der Studie dargelegt, um danach das empirische Setting sowie erste Einblicke in Teilergebnisse darzustellen und zu diskutieren.

2 Theoriediskurse

2.1 «Schon immer ein Sprachfach»

Die Reflexion der sprachlichen Bedingtheit historischen Denkens begleitet die historischen Wissenschaften schon lange und ist insofern nichts Neues. Der von den Schriften Hayden Whites' angestoßenen «Grundlagenstreit» (Sarasin, 2002, 80) führte zu einer tiefgreifenden Reflexion der eigenen Forschungspraktiken und schließlich dazu, dass Narrativität heute unzweifelhaft als Wesensmerkmal historischer Forschung verstanden wird (Marschnig, 2019a, 64–67).

Auch im geschichtsdidaktischen Theoriediskurs hat sich – vornehmlich von den Schriften Jörn Rüsen's verantwortet – längst eine «narrativen Wende», vollzogen. Historisches Lernen beschrieb Rüsen als Bildung von Geschichtsbewusstsein durch Erzählen, wodurch sich sein populäres Diktum vom «Geschichtsunterricht als Erzählveranstaltung» erklärte. Historisches Erzählen beschreibt er als die Vereinigung unterschiedlicher sprachlicher Handlungstypen: «Verhandeln, Aushandeln, Arrangieren, Umgang mit Widersprüchen», die zur «Erzeugung von Sinn in der Form zeitspezifischer Deutungsmuster» führen. Dieser «Modus von Denken, Erkennen, Erklären und Einsehen» ist von «Zukunftserwartungen und Orientierungsabsichten im Blick auf gegenwärtige Lebensverhältnisse» (Rüsen, 2001, 10) geprägt. Historisches Lernen als «Erzählenlernen» ist also mitnichten ausschließlich in der Vergangenheit verankert, sondern verknüpft diese narrativ mit den Gegenwartserfahrung der Lernenden, mit all ihren Fragen und Orientierungsproblemen. Michelle Barricelli verlangt daher, Geschichtsunterricht so zu planen, dass Narrativität bewusst und erfahrbar gemacht wird und die «Handlungsfähigkeit in diesem Feld» (Barricelli, 2012, 255) – also narrative Kompetenz – als «höchstes Lernziel des historischen Lernens» (Barricelli, 2005, 8) erworben werden kann. Das Narrativieren beschreibt er als

synthetisierende Sprachhandlung, die das Beschreiben, Analysieren, Interpretieren miteinschließt und somit als fachspezifisches Erzählverfahren historischen Sinn produziert. Der Erwerb und die Weiterentwicklung derartiger Fähigkeiten stellt Barricelli (2015) folgend eine große Anstrengung dar: «Narrative Kompetenz ist mindestens Sache einer Schul- und Berufskarriere» (30). Geschichte sei «schon immer ein Sprachfach» (Barricelli, 2015, 25) gewesen, in dem die Entwicklung narrativer Kompetenz als «linguistische und kognitive Kompetenz» (Barricelli, 2008, 143) die zentrale Unterrichtsleitlinie zu sein hat.

2.2 Das «Schlüsselproblem»

Historisches Lernen als «Erzählenlernen» ist somit eng verknüpft mit der Entwicklung fachsprachlicher Kompetenz. Bereits 2010 identifizierte Saskia Handro «Sprache» als ein «Schlüsselproblem» (5) des kompetenzorientierten Geschichtsunterrichts und verglich fachliches und fachsprachliches Lernen später mit den «zwei Seiten einer Medaille, die bei der Planung und Strukturierung historischer Lehr- und Lernprozesse integral betrachtet werden müssen» (Handro, 2015, 12). Gerade vor dem Hintergrund der Kompetenzorientierung könne fachsprachliches und fachliches Lernen nicht voneinander getrennt werden. (Marschnig, 2019b, 126–128) Auch Hilke Günter-Arndt (2010) beschrieb den Geschichtsunterricht nicht zu Unrecht wenig schmeichelhaft als «Abrichtungsinstitution», in der die Sprache der Geschichte und das Sprechen über Geschichte, also eine «Grammatik des historischen Denkens und Kommunizierens» gelehrt und gelernt werden» (36).

Auch im angloamerikanischen Diskurs wird der Begriff der «Grammatik» für die Beherrschung des historischen Erzählens, also gewissermaßen die Alphabetisierung (engl. *literacy*) im Feld des Historischen, verwendet:

«Reading, writing and thinking are all interconnected in the study of history. Because learning a new discipline involves mastering its oral and written language, it is often difficult to distinguish between content learning and content literacy learning. [...] Rather than being discrete skills, these practices overlap and mutually reinforce one another.» (Monte-Sano et al., 2014, 8–9)

Wie Handro sprechen auch Monte-Sano et al. das Ineinanderfließen von fachsprachlichem und fachlichem Lernen an und fordern die systematische Integration von sprachaufmerksamen Arbeiten in den Geschichtsunterricht: «Students

won't pick up historical thinking or literacy practices through occasional assignments; the effort must be strategic and sustained» (Monte-Sano et al., 2014, 4). Folgt man dem US-amerikanischen Diskurs (Wissinger et al., 2020; Nokes & De La Paz, 2020), der sich weitaus ausgiebiger mit Historical Literacy befasst hat, so sind es gerade sprachlich benachteiligte Lernende, die von einem stärker fachsprachlich orientierten Geschichtsunterricht profitieren würden: «We believe that it is our struggling readers who most need instruction in learning how to read like a historian» (Wineburg et al., 2013, 12). Schwächeren Lernenden durch präparierte Texte entgegenzukommen, lehnen Wineburg et al. (2013) allerdings ab, ihrer Meinung nach ist das genaue Gegenteil von Nöten: «It is this rich diet, not the thin gruel of textbooks, that our students most need» (12).

Fachsprachliche Bildung im Rahmen des historischen Lernens zu fokussieren, ist allerdings nicht unumstritten. Handro (2015) warnt etwa davor, historisches Denken auf «erlernbare sprachliche Mittel und Diskursfunktionen» (20) zu reduzieren. In einem kritischen Zwischenruf auf *Public History Weekly* verdeutlichte Christoph Kühberger 2017, dass die Entwicklung des historischen Denkens keinesfalls durch eine sprachdidaktische Vereinnahmung korrumpiert werden sollte. Unzulängliche Aufgabendesigns, wie etwa die von Gattungsfragen und historischen Bezügen vollends befreite Erfindung einer römischen Gründungssage als «pure Fiktion», würden zwar vorgeben, «fachspezifische Kompetenzen anzubahnen, wenngleich ein Blick auf die Arbeitsaufgaben offenbart, dass man davon weit entfernt ist» (Kühberger, 2017). Derlei «Irrwege» (Kühberger, 2017) müssen zwar mitgedacht werden, will man Markus Bernhardt und Franziska Conrad (2018, 4) folgen, die 2018 dazu aufriefen, die allzu theorielastige «Höhenkamm»-Diskussion über narrative Kompetenz zu beenden und sich endlich mit der Umsetzung derselben im Klassenraum zu beschäftigen. Dennoch stellt das aktive Herangehen an sprachliche Hürden eine wesentliche Voraussetzung für den Kompetenzaufbau im Geschichtsunterricht dar, was auch Handro (2020) bestätigt: «Unstrittig ist dagegen, dass die Berücksichtigung der sprachlichen Dimension historischen Lehrens und Lernens für eine lernerorientierte Operationalisierung historischer Kompetenzentwicklung unabdingbar ist» (343). Als eine jedenfalls sehr lohnenswerte Perspektive wird dabei seit einiger Zeit die Rolle des Schreibens als produktiver Fertigkeit im Geschichtsunterricht beforscht.

2.3 Schreiben im Geschichtsunterricht

Obwohl im bundesdeutschen Kontext eine lange Tradition schriftlicher Textproduktion im Geschichtsunterricht feststellbar ist, wird dieser oft als mündliches Unterrichtsfach (Thürmann et Al., 2015, 20) beschrieben. Dies gilt insbesondere für den österreichischen Kontext, wo Text(e) hauptsächlich als Schulbuchtexte in Erscheinung treten und das Schreiben meist reproduktiven Charakter aufweist, also entweder in Testsituationen oder zur Memorierung von Informationen zur Anwendung kommt. Dies ist umso bedauerlicher, als gerade das «Distanzmedium der Schrift» das «Erkenntnisverfahren der Historischen Methode» (Hartung & Memminger, 2018, 4) weit besser unterstützen kann als das mündliche Gespräch. Das «*emplotment*» (Coffin, 1997, 199), also die produktive Verknüpfung von Quellenpassagen, Historikerurteilen und persönlichen Ansichten, die letztlich zu einer eigenen Narration führt, wird durch die «Reflexivität des Schreibprozesses» gefördert (Hartung & Memminger, 2018, 3). Außerdem – und das erscheint gerade vor dem Hintergrund der Kompetenzorientierung als besonders bedeutsam – nähert sich ein stärker dem Schreiben verpflichteter Geschichtsunterricht dem geschichtswissenschaftlichen Forschungsprozess sowohl formal als auch inhaltlich an, verfügt er doch über das Potenzial, den Prozess der historischen Wissensgenerierung zu simulieren. Wie in der Arbeit der Historiker*innen könnte das «Schreiben zu einem mächtigen Werkzeug des selbstständigen historischen Denkens [werden], mit dem Schüler*innen die geschichtlichen Erkenntnisprozesse nachvollziehen und – was besonders wichtig ist – diese in elementarisierter Form selbst vollziehen und bewerten können» (Hartung & Memminger, 2018, 6).

Derartige Zugänge sind in der Unterrichtsrealität noch nicht angekommen, was sich auch in verschiedenen Studien zur Textproduktion im Geschichtsunterricht ablesen lässt. Bereits Schönemann et al. (2010, 124) haben auf die mannigfaltigen Herausforderungen aufmerksam gemacht, die jugendliche Schreibende beim Verfassen von Narrationen zu bewältigen haben. Besonders die Distanzierung von Quellenaussagen, die Produktion von kohärenten Texten, aber auch die sprachliche Verfertigung von Sach- und Werturteilen bereitete große Schwierigkeiten. Auch Josef Memminger (2009, 209) und Barricelli (2005, 273) stellten zum Teil erhebliche sprachliche Probleme in der Textproduktion fest. Demgegenüber konnte Hartung (2013; 2015) nachweisen, dass die Schüler*innen stark davon profitieren, wenn ihnen bestimmte Genres vorgegeben werden und sie innerhalb dieser Genretypen auf unterschiedliche Weise historischen Sinn produzieren. Daher definierte Hartung (2019, 73) eine «performative Gattungskompetenz» als

hervorragende Aufgabe des Geschichtsunterrichts, der so «produktionsorientiertes formbewusstes Geschichtslernen» bereitstellen sollte. Dabei gehe es ihm freilich nicht darum, schlichtes deklaratives Wissen über Gattungen zu erwerben, sondern vielmehr, dieses in prozedurales Wissen zu überführen und damit das historische Denken der Schüler*innen zu unterstützen.

Die stärkere Integration des Schreibens in das historische Lernen, das zeigen Hartungs Studien, aber auch andere Forschungsarbeiten (Nitsche et al., 2019; Hodel, 2013; Mierwald & Brauch, 2015), ist zweifelsohne ein Gebot der Stunde, lässt sich damit doch weit zielgerichteter und feingliedriger an der narrativen Kompetenz der Geschichtslernenden arbeiten. So verdeutlichten auch Nitsche et al., dass ihre Intervention zwar die narrative Kompetenz der Jugendlichen zwar stark verbesserte, nicht aber ihre Kompetenzen zum historischen Denken (Nitsche et al., 2019, 3). Hier setzt das Habilitationsprojekt «Multiperspektivität als sprachliche Herausforderung» an, das derzeit am Institut für Geschichte der Karl-Franzens-Universität Graz durchgeführt wird und im Folgenden dargestellt werden soll.

3 Projekt «Multiperspektivität als sprachliche Herausforderung»

3.1 Theoretische Basis

3.1.1 Multiperspektivität

Die Hinwendung der Geschichtsdidaktik zur Sprache lässt sich also theoretisch schlüssig begründen und wurde auch bereits empirisch in Angriff genommen. Dabei stand aber zumeist eine als Gesamttext zu schreibende Narration («verpackt» in verschiedene didaktische Genres, Argumentationen oder Erörterungen) im Fokus, während die fachsprachliche Förderung einzelner didaktischer Prinzipien des historischen Denkens bisher außer Acht gelassen wurde. Darauf zielt die gegenständliche Studie am Beispiel des geschichtsdidaktischen Grundprinzips der Multiperspektivität ab, das im Folgenden knapp beschrieben werden soll.

«Multiperspektivität [ist vielleicht] der beliebteste Terminus technicus unseres Faches, zumindest einer der einflussreichsten» (Lücke, 2012, 281). Martin Lücke formuliert diese These mit dem Verweis auf die Rahmenlehrpläne der deutschen Bundesländer, wo Multiperspektivität durchweg prominent und nor-

mativ wirksam erscheint, was auch für die österreichischen Lehrpläne gilt. Das von Klaus Bergmann in die geschichtsdidaktische Diskussion eingebrachte Konzept ist Lücke zufolge «keinesfalls revisionsbedürftig» (Lücke, 2012, 281), auch Hans-Jürgen Pandel (2017) beschreibt Multiperspektivität treffend als «eine didaktische Regel, die verlangt, ein Ereignis aus mehreren Perspektiven darzustellen» (31). Die von Bergmann geforderte, möglichst durchgängige Integration unterschiedlicher Perspektiven in das historische Lernen erscheint unzweifelhaft als zentrales Gütekriterium für einen qualitätsvollen Geschichtsunterricht. Aus gutem Grund hat auch der kanadische Geschichtsdidaktiker Peter Seixas das Verstehen von Perspektivität als eine von sechs grundlegenden Konzepten des historischen Denkens definiert (Seixas & Morton, 2012). Damit ist auch bereits die Bergmanns Konzept zugrundeliegende geschichtstheoretische Grundkonstante angesprochen. Geschichte, die ja die Vergangenheit über Quellen erschließt, ist immer standortgebunden und deshalb perspektivisch. «Alle Aussagen [...] sind deshalb nicht ›reine‹ Geschichte, sondern sie sind schon durch die Sichtweise der jeweiligen Person interpretierte und gewertete Aussagen» (Pandel, 2000, 102).

Allerdings verlangt Multiperspektivität im Sinne Bergmanns nicht, wahllos unterschiedliche Meinungen oder Eindrücke von historischen Akteuren zu präsentieren. Vielmehr geht es ihm darum, soziale Standorte deutlich zu machen. Pandel nennt dafür beispielhaft «soziologische, religiöse/konfessionelle, ökonomische, politische, ideologische, ethnische Standorte», die er als «natürliche» zusammenfasst und zusätzlich auch «situative Standorte» benennt, die von der jeweiligen historischen Begebenheit abhängen, also etwa «Täter und Opfer, Kläger und Angeklagter, Angreifer und Verteidiger» (Pandel, 2013, 350). Bergmann forderte die Einbeziehung ebendieser unterschiedlichen Perspektiven in den Geschichtsunterricht, was letztlich der Integration von historiografischen Gütekriterien – der Berücksichtigung möglichst vieler Zugänge zu einem Thema – in die Geschichtsdidaktik bedeutete:

«Multiperspektivität ist ein Prinzip historischen Lernens, bei dem historische Sachverhalte aus den Perspektiven verschiedener beteiligter und betroffener Menschen dargestellt und betrachtet werden. Die in einer historischen Konstellation handelnden und leidenden Menschen haben ihre Situation unterschiedlich wahrgenommen und mit unterschiedlichen Absichten auf die Situation einzuwirken versucht. [...] Schülerinnen und Schüler werden mit diesen unterschiedlichen Perspektiven konfrontiert, um sich mit ihnen in Akten historischen Denkens auseinander zu setzen.» (Bergmann, 2007, 69)

Mit dem Grundprinzip der Multiperspektivität wird also eine geschichtstheoretische Prämisse – nämlich die der unumgänglichen Perspektivität jeder historischen Erkenntnis – in das historische Lernen integriert. Schüler*innen zu befähigen, ebendiese Perspektivität wahrzunehmen, und noch viel schwieriger, in ihren eigenen historischen Erzählungen deutlich zu machen, stellt Pädagog*innen ohne Zweifel vor große Herausforderung. Klaus Bergmann hat selbst festgestellt, dass eine multiperspektivische Lernumgebung den Lernenden sehr viel abverlangt, wird ihnen doch zugemutet, «anhand unterschiedlicher Zeugnisse über vergangenes menschliches Handeln und Leiden nachzudenken und dabei gegebenenfalls auch zu höchst unterschiedlichen, voneinander abweichenden und auch von der Auffassung von Lehrerin und Lehrer abweichenden Urteilen zu kommen [...]» (Bergmann, 2007, 164). Dass in der Inszenierung eines solches Lernsettings auch für Lehrende eine große Herausforderung zu sehen ist, liegt auf der Hand.

3.1.2 Textprozeduren

Die Schwierigkeiten, die Jugendliche dabei haben, multiperspektivisch zu erzählen, zu verdeutlichen und außerdem ein Instrument zu entwickeln, mit dem man diesen Herausforderungen begegnen kann, ist eines der zentralen Ziele der dargestellten Studie. Dazu wurde ein Schreibsetting entwickelt, in dem das Ziel verfolgt wird, die Textkompetenz (Marschnig, 2017, 778–780) und somit auch Schreibleistungen von Lernenden durch die Bereitstellung von Textprozeduren zu verbessern. Textprozeduren können als textkonstituierende Elemente beschrieben werden, die sich an der Textoberfläche durch sogenannte «Prozedurenausdrücke» (Feilke & Bachmann, 2014, 7–9) manifestieren. Damit sind konventionalisierte Ausdrücke gemeint, die ein Handlungsschema sprachlich abbilden. Textprozeduren sind also «Form-Funktions-Gefüge», die für die kommunikativ-funktionale Gestaltung von Texten verantwortlich sind. Am Beispiel festgemacht verdeutlichen die Prozedurenausdrücke «nach Jeismann» oder «laut Schörken» etwa das wissenschaftliche Referieren. Ohne derartige Ausdrücke zu beherrschen, lässt sich ein wissenschaftlicher Text nicht schreiben. Man benötigt derartige «domänenspezifische Werkzeuge» (Feilke, 2014, 11), um den inhaltlichen und formalen Anforderungen gerecht zu werden (Bushati et al., 2018, 9). Je besser diese Werkzeuge beherrscht werden, desto souveräner wird die Textproduktion. Bushati et al. (2018, 10) verdeutlichen das Potenzial des Ansatzes für die Schreibdidaktik:

«Das Vorhandensein – oder Nichtvorhandensein – bzw. die Varianten im Gebrauch von Textprozeduren in einem Text können somit sowohl als Kompetenzindikatoren als auch als Ansatzpunkte für deren Förderung betrachtet werden. Textprozeduren sind daher nicht nur aus textlinguistischer Sicht von Interesse, sondern auch aus Perspektive der Schreibentwicklungsforschung und der Schreibdidaktik.»

Je früher die Auseinandersetzung mit Textprozeduren beginnt, desto nachhaltiger ist ihre Integration in die eigene Sprachkompetenz. Eine Möglichkeit hierfür stellt das «Kontroversenreferats» (Feilke & Lehnen, 2011, 40) dar. Dabei handelt es sich um eine materialgestützte Schreibaufgabe, die Schüler*innen dabei unterstützen soll, widersprüchliche Positionen in einem Text deutlich zu machen. Werden diese Positionen als unterschiedliche Perspektiven auf ein Ereignis verstanden, so liegt es nahe, das «Kontroversenreferat» als einen möglichen Zugang zur Förderung von multiperspektivischem Schreiben im Geschichtsunterricht zu verstehen.

3.2 Empirisches Setting

Im Zuge der Studie «Multiperspektivität als sprachliche Herausforderung» wurden Schüler*innen der 11. Schulstufe – insgesamt 186 ihrer Altersgenossen aus vier Gymnasien (insgesamt acht Schulklassen) – angelehnt an das «Kontroversenreferat» dazu aufgefordert, zwei Quellentexte gegenüberzustellen, deren Hauptaussagen zu präsentieren, diese mit den dahinterliegenden politischen Überzeugungen zu verknüpfen und dadurch ihre Perspektivität zu erklären. Das inhaltlich-fachliche Fundament dieses Unterfangens stellte eine vorgelagerte Unterrichtsstunde dar, in der ein grober Überblick über die politische und gesellschaftliche Situation in der österreichischen Zwischenkriegszeit, in welche die zu verarbeitende historische Situation eingebettet war, gegeben wurde. Dabei erhielten die Jugendlichen in einem standardisierten Lehrereinput Informationen zu den politischen Lagern, der zunehmend «auf der Straße» ausgetragenen Differenzen zwischen diesen und schließlich auch zur konkreten Situation, die von den Quellentexten verarbeitet wurde. Zusätzlich wurde den Jugendlichen ein knapper, halbseitiger Einführungstext zu dem historischen Ereignis bereitgestellt. Die Schüler*innen waren auf mehreren Ebenen gefordert: Sie mussten sich mit einem unbekanntem Kontext auseinandersetzen, achtzig Jahre alte Quellentexte lesen, die darin enthaltenen Argumentationsstrategien freilegen, darü-

ber nachdenken, die verschiedenen Sichtweisen auf ein und dasselbe Ereignis darstellen und das alles schließlich in einen eigenen Text gießen, den sie – den Vorgaben der Literalen Didaktik (Schmölzer-Eibinger, 2007) folgend – kollaborativ, also zu zweit verfassten. Für all das stand nur eine Unterrichtsstunde zur Verfügung. Insgesamt beschäftigten sich die Jugendlichen also nur zwei Stunden mit der zu verhandelnden Thematik. Dieses enge Zeitkonzept wurde sehr bewusst gewählt, denn das Unterrichtskonzept sollte möglichst nah an der Unterrichtsrealität der Jugendlichen liegen. Diese sieht für das gewählte Unterrichtsthema, die gewaltsamen Zusammenstöße in Schattendorf 1927, keinesfalls mehr Unterrichtszeit vor. Das bestätigten auch die beteiligten Lehrkräfte, die zudem angaben, wenig bis gar keine Erfahrungen mit derartigen Schreibaufgaben im Geschichtsunterricht zu haben. Ihre Schüler*innen stimmten dieser Schilderung zu. Die Anforderung, einen Text über ein historisches Ereignis zu schreiben – und das auch noch auf der Grundlage von gegenläufigen Quellaussagen – ist den Geschichtslernenden ziemlich fremd.

Auch der thematische Rahmen war bewusst gewählt. Kaum ein Ereignis der österreichischen Geschichte des 20. Jahrhunderts ist derart umstritten wie die Zusammenstöße von Schattendorf und die darauffolgende Radikalisierung der Innenpolitik, die schließlich zum Brand des Justizpalastes am 15. Juli 1927 führten. Diese Polarisierung lässt sich bereits in der zeitgenössischen Berichterstattung ablesen, die das Ereignis im burgenländischen Schattendorf, bei dem zwei Menschen erschossen wurden, aus klar ideologisch gefärbten Perspektiven darstellte. Zwei dieser Perspektiven, jene der sozialdemokratischen «Arbeiterzeitung» und jene der rechtsgerichteten «Reichspost», wurden auch für das Schreibsetting verwendet.

Während der Unterrichtsstunden konnte man gut beobachten, wie sehr die Jugendlichen von der Aufgabe gefordert waren. Eine Hälfte von ihnen (also je eine Klasse pro Schule) wurde als Interventionsgruppe in der einführenden Unterrichtsstunde mit Informationen zu Textprozeduren versorgt, um sie auf die Intervention in der darauffolgenden Stunde vorzubereiten. In dieser zweiten Stunde erhielten dann alle Lernenden den Auftrag, in Partnerarbeit einen Text¹ über das Ereignis von Schattendorf zu verfassen, in dem sie sowohl die Perspek-

1 Im Gegensatz zu Hartungs Studien wurde bewusst auf die Einführung eines spezifischen Genres verzichtet, um den Fokus der Jugendlichen nicht von den Teilaufträgen abzulenken. Im Zuge der acht Durchführungen fragten zwei Schüler*innen nach der Textgattung, konnten aber mit dem Hinweis, dass kein spezifisches Genre verlangt sei, gut weiterarbeiten.

tive der «Arbeiterzeitung», als auch jene der «Reichspost» sichtbar machen und mit deren politischer Grundierung verknüpfen sollten. Die Formulierung des Auftrags «*Verfasst einen multiperspektivischen Text (unter Zuhilfenahme der Prozedurausdrücke)!*» war für die Jugendlichen klar nachvollziehbar und wurde durch folgende vier Zusatzaufträge ergänzt: «*Gebt einen nicht wertenden Überblick über das Ereignis in Schattendorf 1927. Stellt die Perspektiven, die in den Zeitungstexten deutlich werden, gegenüber! Verwendet auch Beispiele aus den Texten. Setzt die Darstellungen der Zeitungen zu den jeweiligen politischen Perspektiven in Bezug. Begründet dies mit der Wortwahl und der Schilderung des Ereignisses sowie der Darstellungen des politischen Gegners. Fasst abschließend den Einfluss der Quellenperspektiven auf die Wahrnehmung von Ereignissen zusammen!*»²

Während die Vergleichsgruppe diesen Auftrag ohne Hilfsmittel erledigte, erhielten die Jugendlichen der Interventionsgruppe eine Liste mit Textprozeduren-Ausdrücken als *scaffold*³ vorgelegt, die ihnen beim Schreiben des Textes helfen sollten. Abgerundet wurde die Studie durch eine Fragebogenerhebung, in der die Vorerfahrungen mit dem Schreiben im Geschichtsunterricht und dem Quelleneinsatz und die generelle Haltung zum schulischen Geschichtsunterricht erfragt wurden.

-
- 2 Jedem dieser Teilaufträge wurde als zentrale Intervention ein Set von Prozedurausdrücken zugeordnet, das die Jugendlichen frei verwenden und variieren konnten. Beispielsweise standen für den Teilauftrag zur Gegenüberstellung der Prozedurausdruck «Betrachtet man die zwei Texte, fällt auf, dass...» oder für den Teilauftrag zur Begründung der Perspektiven die Prozedurausdrücke «weil», «deshalb», «aus diesen Gründen» oder «infolgedessen» zur Verfügung.
 - 3 Das Konzept des Scaffolding stammt aus der Fremdsprachendidaktik und versteht sprachliche Hilfestellungen als Gerüst, das im Zuge des Lernfortschritts Schritt für Schritt abgebaut werden kann, vgl. dazu Gibbons, 2002. Zur geschichtsdidaktischen Sicht vgl. Körber et al., 2020.

3.3 Erste Ergebnisse

Anschließend wurden die Texte einer qualitativen Inhaltsanalyse nach Philipp Mayring (Mayring & Fenzl, 2014) unterzogen. Dafür wurden die folgenden fünf Kategorien entwickelt, auf deren Basis die Texte analysiert wurden: Überblick, Gegenüberstellung der Quellenperspektiven, In-Bezug-Setzen, Begründung der Perspektiven, Schlussfolgerung. Die korrespondierenden Textabschnitte wurden innerhalb der Kategorien den Zuschreibungen «ansprechend umgesetzt», «deutlich umgesetzt», «kaum umgesetzt» oder «nicht umgesetzt» zugeordnet. Eine weitere Kategorie mit dem Titel «Gesamtumsetzung der Aufgabenstellung» wurde eingeführt, um den Gesamteindruck der Texte zu beschreiben. Zusätzlich wurden alle Unterrichtsstunden, die von den jeweiligen Klassenlehrer*innen einem strikt vorgegebenen Plan entsprechend unterrichtet wurden, dicht beschrieben. Des Weiteren wurden Vorerfahrungen aller Teilnehmer*innen mittels Fragebogen erhoben.

Bisher wurde etwa die Hälfte der Texte analysiert.⁴ Anhand der Texte von vier Klassen wird bereits klar ersichtlich, dass die Interventionsgruppe die Anforderungen deutlich besser lösen konnte als die Vergleichsgruppe. Die Qualität der Texte stieg durch die vermehrte Verwendung von Prozedurausdrücken deutlich an. Insbesondere in der Unterkategorie «Gegenüberstellung» (Verbindung bzw. Konfrontation kontroverser Quellenperspektiven) wurde der Unterschied zwischen Interventionsgruppe und Vergleichsgruppe manifest. In der Interventionsgruppe Schule 1 ($n = 11$) weisen 82 Prozent der Texte die Zuschreibung «ansprechend umgesetzt» und 18 Prozent die Zuschreibung «deutlich umgesetzt» auf, in der Vergleichsgruppe ($n = 11$) erreichten nur 55 Prozent die Zuschreibung «ansprechend umgesetzt», 18 Prozent «deutlich umgesetzt» und 27 Prozent «kaum umgesetzt». In der Interventionsgruppe der Schule 2 ($n = 12$) wiesen 42 Prozent der Texte diese Aufgabe «ansprechend» auf und 50 Prozent «deutlich umgesetzt», während in einem Text diese Teilaufgabe gar nicht gelöst worden war. Die Texte der Vergleichsgruppe ($n = 9$) fielen vergleichsweise stark ab («ansprechend» = 0 %, «deutlich» = 56 %, «kaum» = 33 %, «nicht» = 11 %).

Die Elaborate aus der Interventionsgruppe in Schule 1 waren besonders auffällig. Die Jugendlichen in dieser Gruppe hatten während ihrer gesamten Schullaufbahn sehr intensive sprachliche Förderung durchlaufen. Um diese Klas-

4 Bisher wurden 43, also etwa die Hälfte der insgesamt erhobenen Texte analysiert, ein allfälliges Interrating steht noch aus.

se – in der betreffenden Schule wird sie als «Bili» (Bilinguale Klasse) adressiert – besuchen zu dürfen, werden die Kinder einem «Eignungsgespräch» unterzogen. Danach wird der Unterricht acht Jahre hindurch in den Fächern Biologie, Geografie, Geschichte und Musikerziehung auf Englisch durchgeführt. Diese durchgängige Sprachförderung zeigt sich im Umgang der Jugendlichen mit den Aufgabenstellungen sehr deutlich. Die Texte dieser Schüler*innen heben sich deutlich von den übrigen ab und weisen einen sehr reflektierten und gewandten Umgang mit den Aufgabenstellungen auf. Der Vergleich mit den Texten der Interventionsgruppe von Schule 2 macht dies deutlich. Die Texte in Schule 1 wurden in der Kategorie «Gesamtumsetzung» zu 73 Prozent mit «ansprechend umgesetzt» und zu 27 Prozent «deutlich umgesetzt» beurteilt. Demgegenüber schnitt die Interventionsgruppe von Schule 2 deutlich schlechter ab (16 % erreichten die Zuschreibung «ansprechend umgesetzt» und 42 % «deutlich umgesetzt», «kaum» erreichten 42 %, «nicht» jeweils 0). Am viel schwächeren Abschneiden dieser Klasse lässt sich außerdem die große Bedeutung des Lehrerhandels ablesen. In der betreffenden Klasse erfolgte der Input in der ersten der beiden Stunden sehr mangelhaft. Besonders die Einführung der Textprozeduren war augenscheinlich nicht sinnstiftend, da diese teilweise wahllos von Schüler*innen im Text eingesetzt wurden und oft nicht zur Strukturierung beitragen. Dies bestätigt außerdem den Focus-on-Form-Ansatz, der ja davon ausgeht, dass formale Strukturen besonders dann gut angewendet werden können, wenn Lernende explizit auf deren Funktion hingewiesen werden.

4 Conclusio

Es wird also deutlich, dass Scaffolds bei der Produktion von Texten im Geschichtsunterricht klar positive Auswirkungen haben. Die in den beiden Interventionsgruppen produzierten Texte konnten, aufgrund sinnvoll eingesetzter Prozedurenausdrücke, viel klarer darstellen, was bei den Zusammenstößen in Schattendorf geschehen war.

Diese positiven Effekte werden größtenteils auch von den Schüler*innen bestätigt. Zwei Drittel der Befragten in der Interventionsgruppe 2 ($n = 27$) stimmten der Aussage zu, dass ihnen die Prozedurenausdrücke die Arbeit sehr erleichtert hätten und fast drei Viertel fanden den Einsatz von Scaffolds in der Schreibearbeit sinnvoll. Überdies bestätigten in beiden Interventionsgruppen

mehr als die Hälfte der Befragten ($n = 48$), dass die Scaffolds das Herausarbeiten der Quellenperspektiven unterstützt hätten.

Interessanterweise konnten in der Interventionsgruppe 1 nur knapp 48 Prozent der Befragten der Arbeit mit Prozedurenausdrücken sehr viel abgewinnen. Möglicherweise erfahren diese – sprachlich bereits sehr gut ausgebildeten – Jugendlichen die Arbeit mit Scaffolds als Einschränkung ihrer Kreativität. Anhand ihrer Texte wird deutlich, dass eine kontinuierliche, jahrelange sprachliche Förderung zu klaren Vorteilen in der schriftlichen Verarbeitung von Multiperspektivität führt. Die sehr unterschiedlichen Sichtweisen der beiden Quellentexte wurden von ihnen weitaus klarer dargestellt und mit den dahinter liegenden ideologischen Auffassungen verknüpft.

Sprachliche Kompetenz beeinflusst die historische Sinnbildung jedenfalls. Diese lässt sich über sprachensible Zugänge auch verbessern, das wird durch diese ersten Ergebnisse sichtbar. Je kompetenter Lernende mit Sprache umgehen, desto besser können sie das Grundprinzip der Multiperspektivität in ihren Texten umsetzen. Dass ein bewusster Einsatz von Textprozeduren als sprachlicher Hilfe dem historischen Erzählen sehr zuträglich sein kann – insbesondere dann, wenn deren Potenziale und Wirkweisen der Scaffolds offengelegt werden – ist jedenfalls klar ersichtlich. Textprozeduren können jedenfalls dabei helfen, «die ‹narrative Kompetenz› vom Kopf auf die Füße zu stellen» (Bernhardt & Conrad, 2018, 8), und können eine sinnvolle Ergänzung für einen an narrativer Kompetenz ausgerichteten Geschichtsunterricht sein.

Literatur

- Barricelli, Michele (2005). *Schüler erzählen Geschichte: Narrative Kompetenz im Geschichtsunterricht*. Schwalbach/Ts.: Wochenschau.
- Barricelli, Michele (2008). «The story we're going to try and tell»: Zur andauernden Relevanz der narrativen Kompetenz für das historische Lernen. *Zeitschrift für Geschichtsdidaktik*, 7, 140–153.
- Barricelli, Michele (2012). Narrativität. In Michelle Barricelli & Martin Lücke (Hrsg.), *Handbuch Praxis des Geschichtsunterrichts* (Bd. 1, S. 281–288). Schwalbach/Ts.: Wochenschau.
- Barricelli, Michele (2015). Worte zur Zeit: Historische Sprache und narrative Sinnbildung im Geschichtsunterricht. *Zeitschrift für Geschichtsdidaktik*, 14, 25–46.
- Bergmann, Klaus (2007). Multiperspektivität. In Ulrich Mayer/Hans-Jürgen Pandel & Gerhard Schneider (Hrsg.), *Handbuch Methoden im Geschichtsunterricht* (S. 66–77). Schwalbach/Ts.: Wochenschau.
- Bernhardt, Markus & Conrad, Franziska (2018). Sprachsensibler Geschichtsunterricht: Sprachliche Bildung als Aufgabe des Faches Geschichte. *Geschichte Lernen*, (182), 2–9.

- Coffin, Caroline (1997). Constructing and giving value to the past: An investigation into secondary school education. In Frances Christie & James Robert Martin (Hrsg.), *Genres and institutions: Social processes in the workplace and school* (S. 196–230). London: Open Linguistics Series.
- Feilke, Helmuth (2014). Argumente für eine Didaktik der Textprozeduren. In Thomas Bachmann & Helmuth Feilke (Hrsg.), *Werkzeuge des Schreibens: Beiträge zu einer Didaktik der Textprozeduren* (S. 11–34). Stuttgart: Fillibach.
- Feilke, Helmuth & Bachmann, Thomas (2014). Werkzeuge des Schreibens: Zur Einleitung. In Thomas Bachmann & Helmuth Feilke (Hrsg.), *Werkzeuge des Schreibens: Beiträge zu einer Didaktik der Textprozeduren* (S. 7–10). Stuttgart: Fillibach.
- Feilke, Helmuth & Lehnen, Kathrin (2011). Wissenschaftlich referieren – Positionen wiedergeben und konstruieren. *Der Deutschunterricht*, 5, 34–44.
- Gautschi, Peter (2009). *Guter Geschichtsunterricht, Grundlagen, Erkenntnisse, Hinweise*. Schwalbach/Ts.: Wochenschau.
- Gibbons, Pauline (2002). *Scaffolding language, scaffolding learning. teaching second language learners in the mainstream classroom*. Portsmouth: Heinemann.
- Grannemann, Katharina & Kuchler, Christian (2019). (Die) Sprache als Maß aller Dinge: Perspektive für einen sprachsensiblen Geschichtsunterricht. *geschichte für heute*, 12(1), 31–40.
- Günter-Arndt, Hilke (2010). Hinwendung zur Sprache in der Geschichtsdidaktik: Alte Fragen und neue Antworten. In Saskia Handro & Bernd Schönemann (Hrsg.), *Geschichte und Sprache* (S. 17–48). Münster: Lit.
- Handro, Saskia (2015). Sprache(n) und historisches Lernen: Zur Einführung. *Zeitschrift für Geschichtsdidaktik*, 14, 5–24.
- Handro, Saskia & Schönemann, Bernd (2010). Geschichte und Sprache: Eine Einführung. In Saskia Handro & Bernd Schönemann (Hrsg.), *Geschichte und Sprache* (S. 3–16). Münster: Lit.
- Hartung, Olaf (2013). *Geschichte Schreiben Lernen. Empirische Erkundungen zum konzeptionellen Schreibhandeln im Geschichtsunterricht*. Münster: Lit.
- Hartung, Olaf (2015). Generisches Geschichtslernen. Drei Aufgabentypen im Vergleich. *Zeitschrift für Geschichtsdidaktik*, 14, 47–62.
- Hartung, Olaf (2019). «Gattungskompetenz 3.0». Zur Performativität formbewussten Geschichtslernens. In Christiane Betram & Andrea Kolpatzik (Hrsg.), *Sprachsensibler Geschichtsunterricht: Von Theorie über die Empirie zur Pragmatik* (S. 64–75). Schwalbach/Ts.: Wochenschau.
- Hartung, Olaf & Memminger, Josef (2018). Schreibend Geschichte lernen. Mehr als ein «mündliches Fach» – die Vielfalt des Schreibens im Geschichtsunterricht entdecken. *Geschichte Lernen*, 17(6), 2–11.
- Hodel, Jan (2013). *Verkürzen und verknüpfen. Geschichte als Netz narrativer Fragmente, Wie Jugendliche digitale Netzmedien für die Erstellung von Referaten im Geschichtsunterricht nutzen*. Bern: hep.
- Kölbl, Carlos & Konrad, Lisa (2015). Historical consciousness in germany, concept, implementation, assessment. In Kadriye Ercikan & Peter Seixas (Hrsg.), *New directions in assessing historical thinking* (S. 17–28). New York: Routledge.

- Körber, Andreas/Schreiber, Waltraud & Schöner, Andreas (Hrsg.) (2007). *Kompetenzen historischen Denkens: Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik*. Neuried: ars una.
- Körber, Andreas/Seidl, Patrizia/Witt, Dirk & Bormuth, Heike (2020). Inklusives Geschichtslernen via Scaffolding von Aufgaben. In Sebastian Barsch, Bettina Degner, Christoph Kühberger & Martin Lücke (Hrsg.), *Handbuch Diversität im Geschichtsunterricht. Inklusive Geschichtsdidaktik* (S. 405–423). Schwalbach/Ts.: Wochenschau.
- Kühberger, Christoph (2017). Irrwege sprachsensiblen Geschichtsunterrichts. *Public History Weekly*, 5(3). <https://doi.org/dx.doi.org/10.1515/phw-2017-8205>.
- Lücke, Martin (2012). Multiperspektivität, Kontroversität, Pluralität. In Michelle Barricelli & Martin Lücke (Hrsg.), *Handbuch Praxis des Geschichtsunterrichts* (Bd. 1, S. 281–288). Schwalbach/Ts.: Wochenschau.
- Marschnig, Georg (2017). Die Grenzen meiner Sprache... Literale Didaktik im Geschichtsunterricht. *Erziehung & Unterricht*, (1), 777–783.
- Marschnig, Georg (2019a). «Das Zeug, das im Schulbuch steht, verstehen sie nicht». Sprachaufmerksamkeit als demokratiepolitische Notwendigkeit. In Daniela Vladu, Laura Laza & Veronika Zwing (Hrsg.), *Werte – Zeiten – Orte. Die Kraft der Multikulturalität in Literatur und Sprache* (S. 63–78). Cluj-Napoca: Casa Cărții de Știință.
- Marschnig, Georg (2019b). Language matters. The hidden curriculum of historical thinking as a challenge in teacher training. *International Journal of Research on History Didactics, History Education, and History Culture*, 40, 115–128.
- Mayring, Philipp & Fenzl, Thomas (2014). Qualitative Inhaltsanalyse. In Nina Baur & Jörg Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 543–558.). Wiesbaden: Springer.
- Memminger, Josef (2009). Schulung von historischem Denken oder bloß fiktionale Spielelei? Über kreative Schreibformen im Geschichtsunterricht. *Geschichte in Wissenschaft und Unterricht*, 60(4), 204–221.
- Mierwald, Marcel & Brauch, Nicola (2015). Historisches Argumentieren als Ausdruck historischen Denkens. Theoretische Fundierung und empirische Annäherungen, *Zeitschrift für Geschichtsdidaktik*, 14, 104–120.
- Monte-Sano, Chauncey/De La Paz, Susan & Felton, Mark (2014). *Reading, thinking, and writing about history: Teaching argument writing to diverse learners in the common core classroom, grades 6–12*. New York: Teachers College Press.
- Nitsche, Martin/Waldis, Monika & Gollin, Kristine (2019). *Fostering historical writing in upper secondary school students in Switzerland: Results of an intervention study*. Paper präsentiert im Rahmen des Symposiums «Investigating and fostering historical learning activities at lower and upper secondary school levels» auf der 18. EARLI Conference, Aachen.
- Nokes, Geoffrey & De La Paz, Susan (2018). Writing and argumentation in historical education. In Scott Alan Metzger & Lauren McArthur Harris (Hrsg.), *The Wiley international handbook of history teaching and learning*. New York: Wiley-Blackwell. <https://doi.org/10.1002/9781119100812.ch21>.
- Pandel, Hans-Jürgen (2000). *Quelleninterpretation: Die schriftliche Quelle im Geschichtsunterricht*. Schwalbach/Ts.: Wochenschau.

- Pandel, Hans-Jürgen (2005). *Geschichtsunterricht nach PISA. Kompetenzen, Bildungsstandards und Kerncurricula*. Schwalbach/Ts.: Wochenschau.
- Pandel, Hans-Jürgen (2013). *Geschichtsdidaktik: Eine Theorie für die Praxis*. Schwalbach/Ts.: Wochenschau.
- Pandel, Hans-Jürgen (2017). *Geschichtstheorie. Eine Historik für Schülerinnen und Schüler – aber auch für ihre Lehrer*. Schwalbach/Ts.: Wochenschau.
- Rüsen, Jörn (1983). *Historische Vernunft. Grundzüge einer Historik*. Göttingen: Vandenhoeck & Ruprecht.
- Rüsen, Jörn (2001). Einleitung. In Jörn Rüsen (Hrsg.), *Geschichtsbewusstsein: Psychologische Grundlagen, Entwicklungskonzepte, empirische Befunde*. Wien: Böhlau.
- Sandkühler, Thomas & Bernhardt, Markus (Hrsg.) (2020). *Sprache(n) des Geschichtsunterrichts. Sprachliche Vielfalt und Historisches Lernen*. Göttingen: Vandenhoeck & Ruprecht.
- Sarasin, Philipp (2002). *Geschichtswissenschaft und Diskursanalyse*. Frankfurt a. M.: Suhrkamp.
- Schmölzer-Eibinger, Sabine (2007). Auf dem Weg zu einer Literalen Didaktik. In Sabine Schmölzer-Eibinger & Georg Weidacher (Hrsg.), *Textkompetenz: Eine Schlüsselkompetenz und ihre Vermittlung, Festschrift für Prof. Paul R. Portmann-Tselinkas zum 60. Geburtstag* (S. 207–222). Tübingen: Narr.
- Schönemann, Bernd/Thünemann, Holger & Zülsdorf-Kersting, Meik (2010). *Was können Abiturienten? Zugleich ein Beitrag zur Debatte über Kompetenzen und Standards im Fach Geschichte*. Münster: Lit.
- Seixas, Peter & Morton, Tom (2012). *The big six historical thinking concepts*. Toronto: Nelson.
- Thürmann, Eike/Pertzel, Eva & Schütte, Anna Ulrike (2015). Der schlafende Riese. Versuch eines Weckrufs zum Schreiben im Fachunterricht. In Sabine Schmölzer-Eibinger & Eike Thürmann (Hrsg.), *Schreiben als Medium des Lernens. Kompetenzorientierung durch Schreiben im Fachunterricht* (S. 17–46). Münster: Waxmann.
- Wineburg, Sam/Martin, Daisy & Monte-Sano, Chauncey (2013). *Reading like a historian: Teaching literacy in middle and high school history classrooms*. New York: Teachers College Press.
- Wissinger, Daniel/De La Paz, Susan & Jackson, Carla (2021). The effects of historical reading and writing strategy instruction: With fourth- through sixth-grade students. *Journal of Educational Psychology*, 113(19), 49–67. <https://doi.org/10.1037/edu0000463>.

Von Stacheldraht bis Nivea-Dose

Wie die Durchführung einer archäologischen Ausgrabung bei Schüler*innen Konzeptveränderungen des historischen Denkens anstoßen kann

Johanna Sachse

1 Archäologie geht immer?

Unbestreitbar besitzt die Archäologie seit jeher ein Faszinosum in der Gesellschaft, was die konstant hohe Zahl von Besucher*innen aller Altersgruppen der archäologischen Museen (Deutscher Museumsbund, 2020, 12–13) sowie die relativ hohe Frequenz von Dokumentationen mit archäologischem Bezug im Fernsehen oder auch auf YouTube bestätigen. Auch in Videospiele sind Menschen häufig (in)direkt mit (verzerrten) Bildern von Archäologie konfrontiert (Tomb Raider, No Man's Sky, Assassin's Creed u. a. [Reinhard, 2018]). Aus geschichtsdidaktischer Perspektive existieren somit viele Anknüpfungspunkte, die in diesem Themenfeld unterschiedliche Forschungen hinsichtlich der Begegnung zwischen den Menschen und archäologisch aufgearbeiteter Geschichte ermöglichen. Gerade im schulischen Bereich könnte diese intrinsische Motivation theoretisch einen großen Beitrag zur Förderung des historischen Denkens leisten. Dieses definiert Günther-Arndt als Erinnerung, mit der sich der Mensch Erfahrungen aus der Vergangenheit deutend aneignet, die nicht nur privater Natur sind, sondern den Menschen im Allgemeinen betreffen (Günther-Arndt, 2020, 25). Von Borries (2008) betont hier noch einmal die «komplexe Verbindung von Vergangenheit und Zukunft» (9). Bei diesem Prozess spielt die Methodik eine wichtige Rolle, denn mithilfe dieser können unter anderem Quellen fachgerecht ausgewertet und beurteilt werden, während die dadurch entstehenden reflektierten Erkenntnisse in den Modus des Erzählens münden: «Das Narrativieren schließt jeden historischen Interpretationsprozess (als Teil der Historischen Methode) ab. Es ordnet durch sprachliche Verknüpfung die Menge der aus Quellen ermittelten Tatsachen» (von Borries, 2008, 26). Insgesamt zeichnet sich so-

mit das historische Denken bei Schüler*innen durch das Vorhandensein und Zusammenspiel der grundlegenden verschiedenen Kompetenzbereiche aus (Sach-, Methoden-, Deutungs- und Reflexions- sowie Orientierungskompetenz) (Baumgärtner, 2019, 89–85). Vor allem im Bereich der zu verwendenden Methoden kann die Archäologie je nach inhaltlichem Gegenstand einen wichtigen Beitrag leisten.

In den letzten Jahren sind auf internationaler Ebene einige wichtige empirische Beiträge veröffentlicht worden, die den Bildungswert von Archäologie für die Studienteilnehmer*innen herausstellen, der zum Beispiel in der Förderung eigener Interpretationsansätze von Geschichte sowie in einem physischen und emotionalen Bezug zur lokalen Geschichte bestehe (Kasvikis, 2020, 87; vgl. auch Trškan & Bežjak, 2020; Arias-Ferrer & Egea-Vivancos, 2017). Hierbei wurden bei unterschiedlichen Lernarrangements mit archäologischem Bezug (u. a. Ausgrabungen, Fundinterpretation) im Rahmen von pädagogischen Angeboten Daten von meist jüngeren Schüler*innen erhoben. Was mit Blick auf die Forschungslandschaft jedoch noch fehlt, sind Daten, die einerseits auf der Grundlage einer authentischen Ausgrabung mit Schüler*innen entstanden sind und andererseits sowohl vor als auch nach der Intervention erhoben werden, sodass eine kognitive Entwicklung bei den Proband*innen erkennbar werden kann. Zudem stehen in derartigen Studien zum Großteil jüngere Schüler*innen im Fokus der Untersuchung. Jugendliche und junge Erwachsene werden in der aktuellen Forschungslandschaft tatsächlich wenig beachtet, wenn es um die Potenziale der Archäologie geht. Trotz der Verankerung archäologischer Themen oder auch der Archäologie als Methode in allen gängigen deutschen Lehrplänen existiert im deutschsprachigen Raum bisher keine empirische Studie, die sich diesem Potenzial für den Geschichtsunterricht widmet. Dafür mangelt es auf theoretischer Ebene keineswegs an fachdidaktischen Beiträgen, die der Archäologie attestieren, wie sinnvoll und gewinnbringend ihr Einsatz im Geschichtsunterricht sei. In Bezug auf die dem historischen Denken zugrunde liegenden Kompetenzen bietet der Einsatz der Archäologie in der Schule eine große Bandbreite an Möglichkeiten, diese zu fördern. Allgemein gesehen schaffe die Behandlung archäologischer Themen «Identifikation und Integration innerhalb der Gesellschaft» (Schöbel & Wemhoff, 2016, 236), wenn zum Beispiel im Umgang mit früheren Kulturen ein Diversitätsbewusstsein im Rahmen der heutigen Einwanderungsgesellschaft gefördert würde (Orientierungskompetenz) (Schöbel, 2016, 243). Zudem könnten grundlegende historisch-epistemologische Prinzipien der Erkenntnisgewinnung erlernt werden, die bei der Re- und Dekonstruktion von

Geschichte zur Anwendung kommen, indem Objekte multiperspektivisch interpretiert werden (Methodenkompetenz) (Fenn, 2016, 263–264; Hasberg, 2012, 130). Darüber hinaus bietet die Archäologie viele Möglichkeiten, regionalgeschichtlich zu arbeiten, sodass die Schüler*innen einen persönlicheren Bezug zum Untersuchungsgegenstand bekommen, weil sie in ihrer Lebenswelt verortet sind (Fenn, 2016, 265). Hierbei besteht ein weiterer Vorteil darin, dass alle historischen Zeiten archäologisch untersucht werden können und somit keine Einschränkung beispielsweise nur auf die Ur- und Frühgeschichte des Heimatortes besteht. Weiterhin bekämen die Schüler*innen zusätzlich einen haptischen Zugang zur Vergangenheit, wodurch der in den Lehrplänen verankerten Forderung nachgekommen würde, dass alle Quellengattungen im Geschichtsunterricht Berücksichtigung finden sollen (Hasberg, 2012, 130; Erdmann, 2004, 775), zumal ein historisches Lernen «nicht allein kognitiv, sondern auch sinnlich-affektiv bestimmt» (Erdmann, 2004, 778) werde. Um diese Vorgänge des historischen Denkens beim Arbeiten mit archäologischen Quellen gezielt zu fördern, ist es jedoch unerlässlich, mit den Schüler*innen Gespräche auf einer Metaebene zu führen. Bereits 2004 wurde in der unter anderem von Archäolog*innen und Bildungsvertreter*innen verabschiedeten Weißenburger Erklärung eine umfangreiche Liste veröffentlicht, die noch weiterführende positive Aspekte beim Einsatz von Archäologie in der Schule herausstellt (Bakker et al., 2004). Die zugeschriebenen Potenziale sind somit sehr weitreichend.

Aus fachwissenschaftlicher Perspektive gibt es viele Studien, die sich mit der Archäologie in der Schule beschäftigen. Hierzu zählt die Dissertation von Sénécheau (2006) über die Darstellung von Alt-, Mittel- und Jungsteinzeit unter anderem in Schulbüchern. Des Weiteren sind zahlreiche Beiträge von Samida (2010) zu nennen, die sich vehement für eine Etablierung einer «Archäologiedidaktik» einsetzt, welche zum Ziel hat, «die Vermittlung und die Wirkungsweise ihrer Wissenschaft und der von ihr erforschten Vergangenheit in der Gesellschaft zur Aufgabe [zu] machen und eine fachübergreifende archäologisch-didaktische Methodik [zu] entwickeln» (12). Dieser Ansatz würde ebenso ein Wirken im schulischen Bereich implizieren.

Die zuletzt genannten Forschungen und Aufrufe stammen ausschließlich aus der Feder von Archäologinnen. Dementsprechend verwundert es nicht, dass Samida die Geschichtsdidaktiker*innen explizit dazu aufruft, ihr «offensichtliches Desinteresse an dieser Thematik» (Samida, 2010, 10) abzulegen und sich mit Forschungen auf diesem Gebiet am Diskurs zu beteiligen. Aus geschichtsdidaktischer Sicht spricht vor allem Hasberg die Potenziale der Archäologie für den

Geschichtsunterricht an, konstatiert aber gleichzeitig: «Eine systematische Befassung mit dem Potenzial der Archäologie zur Beförderung des historischen Lernens existiert bislang nicht» (Hasberg, 2012, 129). Somit ist das Forschungsdesiderat klar formuliert, welches aus geschichtsdidaktischer Perspektive den Vorteil mit sich bringt, epochenungebunden zu sein, obwohl Archäologie häufig ausschließlich mit der Ur- und Frühgeschichte in Verbindung gebracht wird. Dabei hat sich diese Wissenschaft im Laufe der Jahrhunderte in ca. acht Fächer aufgeteilt, die zeitlich und räumlich gesehen unterschiedliche Schwerpunkte verfolgen (Eggert, 2006).¹ Vor allem in den letzten Jahren hat sich mit dem Aufkommen der Zeitgeschichtlichen Archäologie ein neues Fach etabliert, in dem die Suche nach beziehungsweise Aufdeckung von unter der Oberfläche noch befindlichen Relikten und Spuren an nationalsozialistischen Tatorten im Zentrum steht, die in Vergessenheit geraten waren und wieder in das kollektive Gedächtnis zurückgeholt werden (Theune, 2016, 8–9). Aus didaktischer Sicht birgt dieser neue fachwissenschaftliche Ansatz eine Menge neuer Möglichkeiten. Um der Forschungslücke zu begegnen, beschäftige ich mich im Rahmen meiner qualitativ angelegten Dissertation mit der Frage, inwiefern bei Schüler*innen aufgrund der Durchführung einer archäologischen Ausgrabung Konzeptveränderungen in Bezug auf Archäologie als Wissenschaft sowie des historischen Denkens initiiert werden können. Mit Blick auf den oben genannten Forschungsstand sollten diesmal ältere Schüler*innen der 9. und 11./12. Klasse archäologische Ausgrabungen durchführen, wobei sich die insgesamt vier Lerngruppen auf eine authentische (Bremen) und eine nachgestellte (Kalkriese) Ausgrabung mit jeweils unterschiedlichen historischen Kontexten (NS-Zwangsarbeit und «Varusschlacht») aufteilen, um einen Vergleich herstellen zu können. Die Datenerhebung fand – anders als in bisherigen Studien üblich, bei denen ausschließlich Präkonzepte erhoben wurden (u. a. Mathis, 2015) – als zweischrittiges Verfahren statt, indem die Schüler*innen sowohl vor als auch nach der Ausgrabung befragt wurden, um auf diese Weise Entwicklungen betrachten zu können (siehe Kapitel 3). Im vorliegenden Artikel sollen das Forschungsdesign sowie Ergebnisse einer ersten Sichtung des Teilmaterials der authentischen Ausgrabung mit einer 12.

1 Die gängige Einteilung besteht aus der Prähistorischen Archäologie, der Vorderasiatischen Archäologie, der Biblischen Archäologie, der Klassischen Archäologie, der Provinzialrömischen Archäologie, der Christlichen Archäologie, der Archäologie des Mittelalters und der Neuzeit sowie der Zeitgeschichtlichen Archäologie. Allerdings sind die Trennlinien nicht immer scharf zu ziehen, denn in konkreten Fällen kommt es durchaus auch zu Überschneidungen der Fächer.

Klasse vorgestellt werden. Hierbei stellt die Conceptual-Change-Forschung die geschichtsdidaktische Rahmentheorie der Datenerhebung und -auswertung dar.

2 Konzeptionelle Verankerung

Der Conceptual-Change-Ansatz hat seinen Ursprung in der naturwissenschaftlichen Didaktik. Er erforscht, inwiefern sich vorhandene Präkonzepte durch den Erhalt neuer Informationen oder das Sammeln neuer Erfahrungen von einem alltagsweltlichen hin zu einem wissenschaftlicheren Verständnis entwickeln können. Wurde zu Beginn der Forschung Ende des 20. Jahrhunderts noch von einem harten Konzeptwechsel gesprochen, hat sich inzwischen die Auffassung etabliert, dass es sich vielmehr um Konzeptveränderungen handelt, da nicht immer ein komplettes Konzept ersetzt wird (Stark, 2003, 133).

Im Kontext der Geschichtsdidaktik wurde der theoretische Ansatz zur Erforschung in den letzten Jahren zunehmend aufgegriffen (u. a. Mathis, 2015; Fenn, 2018). Die Herausforderung in Bezug auf historische Inhalte besteht allerdings darin, dass das Fach – anders als die naturwissenschaftlichen Fächer – interpretativ angelegt ist und somit eine offene Struktur für individuelle Vorstellungen impliziert. Zudem kann bei Präkonzepten nicht von richtigen oder falschen Vorstellungen gesprochen werden, gerade weil Geschichte subjektiv geprägt ist. Demzufolge vermutet Limón, dass Präkonzepte im geschichtlichen Bereich resistenter gegenüber Veränderungen sind, weil sie oft mit individuellen Identitäten, Ansichten und Werten im Zusammenhang stehen (Limón, 2002, 282–283). Die grundlegende Ausrichtung dieses Forschungszweiges in der Geschichtsdidaktik zielt somit auf die Frage hin, «wie ein Alltagsverständnis [...] [bzw. Präkonzept, J. S.] von Schüler*innen zu einem bestimmten Sachverhalt oder einer bestimmten Methode der Erkenntnisgewinnung durch wissenschaftlich angemessene Begriffe, Erklärungen und Methoden «erweitert» oder «ersetzt» werden kann» (Günther-Arndt, 2004, 258–259).

Mathis hat im Rahmen seiner Dissertation hierfür Schüler*innenvorstellungen in vier Ebenen unterteilt, die jeweils entweder den alltagsweltlichen Vorstellungen oder der geschichtswissenschaftlichen Vorstellung gleichen. *Begriffe* bezeichnen vergangene oder gegenwärtige Objekte, Sachverhalte oder auch Fachbegriffe, die definiert sind (z. B. «Ausgrabung»). *Konzepte* sind sprachlich verknüpfte Begriffe, die Aussagen über Geschichte darstellen (z. B. «Die Menschen jüdischen Glaubens wurden durch die Etablierung der NS-Ideologie diskrimi-

niert»). *Erklärungsmuster* nutzen im weiteren Schritt Begriffe und Konzepte zur Erklärung komplexerer inhaltlicher Zusammenhänge. *Erklärungsmuster der ersten Ordnung* beziehen sich dabei auf die Erklärung historischer Sachverhalte auf der Basis von alltagsweltlichen oder wissenschaftsadäquaten Konstrukten, wohingegen *Erklärungsmuster der zweiten Ordnung* den Fokus auf erkenntnistheoretische Aspekte legen, das heißt darauf, wie Wissen über historische Ereignisse konstruiert wird (Mathis, 2015, 33–37). Diese von verschiedenen internationalen Geschichtsdidaktiker*innen aufgeführten «Meta concepts» (Limón, 2002, 271) beinhalten ca. acht verschiedene Erklärungsmuster: Zeit/Raum, Wandel, Bedeutung, Beweis/Quelle, Ursache, Erklärung/Beschreibung/Narration, historische Handlungsspielräume und Empathie. In jüngster Forschung wurde dieser Liste noch der Aspekt «Account» hinzugefügt, welcher als «representations of events in the past, in written and in other forms, produced subsequent to the events/people/states of affairs that they describe» (Chapman & Georgiou, 2021, 78) definiert und in insgesamt sechs Progressionsstufen eingeteilt wird (Chapman & Georgiou, 2021, 80). Alle vier Ebenen spielen laut aktueller Forschung bei der Konstruktion von Geschichte auf kognitiver Ebene eine Rolle, wobei Mathis auch anmerkt, dass die Konzepte keine starren Konstrukte sind, sondern je nach Situation abgerufen werden und damit stets dynamischer Natur sind (Mathis, 2015, 40). Trotzdem oder gerade deshalb können sie bei der Aufdeckung von Konzeptveränderung bei Schüler*innen hilfreich sein. Die beschriebenen Konzepte bilden einen wichtigen Teil des historischen Denkprozesses, wenn es darum geht, Geschichte im Rahmen des *historical reasoning* als einen wichtigen Bereich des historischen Denkens zu organisieren und zu analysieren:

«We define historical reasoning in context of history education as an activity in which a person organizes information about the past in order to describe, compare, and/or explain historical phenomena. In doing this, he or she asks historical questions, contextualizes, makes use of substantive and meta-concepts of history, and supports proposed claims with arguments based on evidence from sources that give information about the past.» (van Drie & van Boxtel, 2008, 89)

An dieser Stelle wird gleichzeitig noch einmal die Relevanz der historischen Kompetenzen deutlich (Baumgärtner, 2019, 79–85). Die Schüler*innen sollen dazu befähigt werden, eben nicht mit alltagsweltlichen, sondern vielmehr mit wissenschaftlichen Vorstellungen beziehungsweise Kompetenzen an Geschichte heranzutreten und diese zu untersuchen (Mathis, 2015, 37).

3 Studiendesign

3.1 Sampling und Aufbau der Studie

Die Forschungsfrage der Dissertation, inwiefern bei Schüler*innen mit der Durchführung einer archäologischen Ausgrabung Konzeptveränderungen in Bezug auf Archäologie als Wissenschaft sowie des historischen Denkens initiiert werden können, soll mithilfe einer qualitativ-rekonstruktiven Interventionsstudie (Flick, von Kardorff & Steinke, 2017, 21–22.) beantwortet werden. Im Rahmen der vorliegenden Teilstudie der Dissertation hat eine Schulklasse des 12. Jahrgangs eines Gymnasiums archäologische Ausgrabungen durchgeführt.² Es wurden bewusst junge Erwachsene für die Studie gewählt. Dies hatte vor allem den Grund, dass der häufig (u. a. in Schulbüchern und Lehrplänen) präsentierten Auffassung, Archäologie beziehungsweise Sachquellen seien ausschließlich in jüngeren Jahrgangsstufen gut einsetzbar, entgegengewirkt werden sollte, denn mit Blick auf die obigen Ausführungen, bestehen auch vor allem für ältere Schüler*innen eine Menge Möglichkeiten, mit der Methode das historische Denken zu fördern. Außerdem haben ältere Schüler*innen schon relativ viel Erfahrung mit dem gängigen Geschichtsunterricht gesammelt und können damit einen Einblick in den Unterrichtsalltag geben.

Eine Pilotstudie diente dazu, das methodische Vorgehen noch einmal zu überprüfen und für weitere Durchgänge anzupassen (Sachse, 2019). In der hier im Zentrum stehenden Teilhauptstudie wurde schließlich mit Schüler*innen eine Ausgrabung durchgeführt und sowohl davor als auch danach wurden in teilstandardisierten und fokussierten Einzelinterviews Gespräche geführt, damit die persönlichen Einschätzungen zum Gegenstand isoliert erhoben werden konnten (Hopf, 2017, 354). Das Sampling bestand aus vier Schülerinnen und zwei Schülern im Alter von 17 Jahren, die alle den gleichen Oberstufen-Grundkurs im Fach Geschichte besuchen.

Im Präinterview, das heißt dem Gespräch vor der Teilnahme an der Ausgrabung, wurden die Vorstellungen der Proband*innen von Archäologie (Assoziationen, Inhalt, Arbeit von Archäolog*innen etc.) sowie der anstehenden Ausgrabung (historischer Kontext, Ablauf der Ausgrabung etc.) ausführlich und mithilfe unterschiedlicher Frageimpulse aus verschiedenen Blickwinkeln er-

2 Vgl. Sachse (2019) für eine Übersicht der Gesamtstudie im Rahmen des Dissertationsprojekts.

fragt (Beispiele: «Was fällt dir spontan alles ein, wenn du an Archäologie denkst?», «Womit beschäftigen sich Archäolog*innen deiner Meinung nach?», «Was weißt du über den Zielort des Ausflugs?»). In den folgenden Tagen ging die Schulklasse zur Ausgrabung, wobei die Studienteilnehmer*innen Digitalkameras mit dem Auftrag ausgehändigt bekamen, alles zu fotografieren, was sie interessant, spannend oder auch merkwürdig während der Exkursion fanden (maximal 15 Fotografien). Die Idee, die Probanden eigene Fotografien während der Intervention aufnehmen zu lassen, geht auf die Methode der Autofotografie zurück. Das Verfahren kommt aus der erziehungswissenschaftlichen Forschung und soll erstmals im geschichtsdidaktischen Kontext Anwendung finden (Kirsch, 2009, 115–117). Diese Bilder hatten vor allem eine erzählgenerierende Funktion. Sie dienten zusammen mit der erfahrenen Intervention selbst als «Stimulus-Situation» (Kirsch, 2009, 115–117) und waren Teil des fokussierten Postinterviews, welches ein paar Tage nach der Ausgrabung stattfand. Mithilfe der Fotografien konnten sich die Schüler*innen besser an das persönlich Erlebte erinnern, zumal sie noch einmal erläutern sollten, warum sie das Abgebildete fotografiert hatten. Hierdurch sind persönliche Schwerpunktsetzungen und für die Schüler*innen besonders einprägsame Elemente erkennbar (Niesyto & Marotzki, 2006, 7). Im zweiten Interview nach der Ausgrabung wurden die Schüler*innen mit den im ersten Interview getätigten Aussagen zu ihren Vorstellungen von Archäologie und dem thematischen Kontext der Ausgrabung konfrontiert, um zu erfahren, was sich an ihren Sichtweisen verändert hat. Erst durch die Betrachtung der zwei Perspektiven auf dieselbe Thematik, das heißt vor und nach der Ausgrabung, konnten die Konzepte verglichen und mögliche Veränderungen aufgedeckt werden.

Da es sich um eine qualitative Forschung handelt, bei der aufgrund fehlender Studienergebnisse zu dieser Thematik im Vorfeld bewusst auf die Bildung von Hypothesen verzichtet wurde (vgl. Mayring, 2010, 225), sollte der Korpus auf der Basis der qualitativen strukturierenden Inhaltsanalyse nach Kuckartz (2018) beziehungsweise Schreier (2014) codiert und ein induktiver Codierleitfaden subsumtiv mit der Datenauswertung erstellt werden (Kuckartz, 2018, 200). Die Vorgehensweise nach Kuckartz beziehungsweise Schreier erscheint an dieser Stelle sinnvoll, weil diese mehr Flexibilität bei der Auswertung der Daten ermöglicht.

3.2 Ausgrabungsort

Am authentischen Ausgrabungsort suchten die Schüler*innen auf dem Gelände eines ehemaligen Zwangsarbeiterlagers des KZ Neuengamme nach Spuren des Außenlagers «Schützenhof» in Bremen-Gröpelingen. Die Aufarbeitung von Zwangsarbeiter*innenlagern ist der zeitgeschichtlichen Forschung zuzuordnen. In Bremen gab es während der Zeit des Nationalsozialismus über 200 Zwangsarbeiterlager, die oftmals mitten in Wohngebieten lagen. Im KZ-Außenlager «Schützenhof» waren in der Zeit zwischen September 1944 und Kriegsende über 700 Häftlinge untergebracht, die an der nahe gelegenen AG Weser Werft für die Rüstungsindustrie Zwangsarbeit verrichten mussten. Die hier entstandenen Teile sollten im U-Boot-Bunker «Valentin» in Bremen-Farge zusammengesetzt werden, doch befand sich dieser noch im Bau, ebenfalls durch Zwangsarbeiter*innen (Hemmer & Milbradt, 2005, 23). 2017 begannen die umfassenden Recherchen, die die Geschichte des «Schützenhofs» aufarbeiten sollten (Halle & Huhn, 2019). Im Rahmen der Zeitgeschichtlichen Archäologie an ehemaligen Tatorten nationalsozialistischen Terrors sind vor allem bauliche Relikte und Objekte von großer Relevanz, die Fragen zum Lageralltag und zu individuellen Schicksalen beantworten können (Schopper & Theune, 2016, 5). Dementsprechend wurde auf dem Gelände des ehemaligen KZ-Außenlagers in Bremen unter der Leitung der Landesarchäologin Prof. Dr. Uta Halle dort eine Ausgrabung angesetzt, wo die Grundrisse einer Häftlingsbaracke vermutet wurden.

Die Thematik Zwangsarbeit wurde für die Dissertation gewählt, weil sie zum einen in nahezu alle Lehrpläne aufgenommen wurde und damit für den konkreten Geschichtsunterricht Relevanz besitzt. Zum anderen soll mit der archäologischen Aufarbeitung dem Vorurteil entgegengewirkt werden, dass Archäologie ausschließlich bei der Erforschung der Ur- und Frühgeschichte eine Rolle spielt. Klar ist, dass es im Schulalltag nicht immer möglich ist, an einer authentischen Ausgrabung teilzunehmen, doch soll die Studie einen ersten Einblick in die Effekte bei Schüler*innen geben, die das Prinzip der archäologischen Arbeit nachvollziehen. Dem Ausgrabungsort kommt dementsprechend eine exemplarische Funktion zu.

3.3 Durchführung der Studie

Der in diesem Beitrag im Fokus stehende Teil der Datenerhebung fand im September 2018 in Bremen auf dem ehemaligen KZ-Außenlagergelände statt, auf dem eine 12. Klasse den zuvor getätigten Grabungsschnitt ein Stück erweiterte. Hierbei waren die Schüler*innen aus sicherheitstechnischen Gründen noch einmal in kleinere Gruppen aufgeteilt, die zu unterschiedlichen Zeiten das Gelände aufsuchten, sodass jede*r Schüler*in ca. 2 ½ Stunden bei der Ausgrabung aktiv gewesen ist. Da die Ausgrabung in Kooperation mit der Landesarchäologie Bremen stattfand, übernahm die Landesarchäologin die Leitung, wobei sie unter anderem von einem Grabungstechniker unterstützt wurde.

Die Arbeit bestand aus dem Abtragen der Grassode und der darunterliegenden Erde. Diese wurde gesiebt und die Objekte ihrem Material entsprechend in Tüten sortiert. Im späteren Verlauf putzten die jungen Erwachsenen noch das Planum sowie die Profile und führten gemeinsam mit dem Grabungstechniker die Einmessung durch. Sofern interessante oder außergewöhnliche Objekte gefunden wurden wie eine Stacheldrahtrolle oder auch eine Strumpfhose, führte die Landesarchäologin noch einmal ein Gespräch darüber im Plenum. Die Herausforderung bei der Bestimmung der Funde beziehungsweise der Grundrisse im Boden bestand darin, dass vor der Lagerzeit auf dem Gelände ein Schützenverein sein Gebäude stehen hatte, welches im Laufe des 2. Weltkriegs zerbombt wurde. Dementsprechend fanden die Schüler*innen mehrere Mauern vor. Generell musste jedes entdeckte Detail einer der drei Zeitschichten vor der Lagerzeit, während der Lagerzeit oder der Nachnutzung des Geländes zugeordnet werden. Da nicht alle Schüler*innen sämtliche Funde sehen konnten, besuchten die Bremer Landesarchäologin Uta Halle und ich die Klasse in der nächstmöglichen Geschichtsstunde noch einmal mit den Funden. Hier stellte die Landesarchäologin die verschiedenen, teilweise noch rätselhaften Fundgruppen vor und berichtete auch von bisherigen Untersuchungsergebnissen. Die Objekte wurden dabei durch die Reihen gegeben, sodass die Schüler*innen diese auch anfassen konnten und Vermutungen anstellten, worum es sich dabei handeln könnte. Zudem wurden sie im Sinne einer Anbahnung der Methoden- sowie Deutungs- und Reflexionskompetenz auch dazu aufgefordert, an derartige archäologische Quellen Fragen zu formulieren und Überlegungen anzustellen, wie diese beantwortet werden könnten beziehungsweise welchen Mehrwert diese für die Aufarbeitung des historischen Kontexts besitzen.

4 Erste Ergebnisse

Die grobe Sichtung des ersten Materials ergab bereits einige Hinweise auf mögliche Konzeptveränderungen bei den Schüler*innen der 12. Klasse, die die Ausgrabung am KZ-Außenlager «Schützenhof» durchgeführt haben.

4.1 Präinterviews

Mit Blick auf die Präinterviews stellte sich heraus, dass viele der Proband*innen bereits gute Ansätze historischen Denkens haben, wenn sie unter anderem betonten, dass Geschichte nie abgeschlossen und das Wissen über vergangene Ereignisse enorm wichtig sei, um heutige Fragen beantworten oder mögliche Entwicklungen in Zukunft verhindern zu können (Interview Greta³ I, Schülerin 12. Klasse, Gymnasium, 07.09.2018, Zeile 56–60, Interview Hanna I, Schülerin 12. Klasse, Gymnasium, 07.09.2018 Hanna, Zeile 45–50). Dabei weckt inhaltlich gesehen vor allem das Geschehen während des letzten Jahrhunderts in ihnen Interesse, denn «[m]an kann sich ein bisschen besser vorstellen, wie das Leben dort mehr unserem heutigen geähnelt hat als noch vor drei-, vierhundert Jahren» (Interview Iris I, Schülerin 12. Klasse, Gymnasium, 07.09.2018, Zeile 60–61).

In Bezug auf das Präkonzept von Archäologie unterscheiden sich die Konzepte der jungen Erwachsenen teilweise. Als erste Assoziationen zum Begriff werden unter anderem Ausgrabungen, das Alte Ägypten und das Sich-draußen-Aufhalten genannt sowie die Vorstellung, dass Sachen entdeckt und erforscht werden (Interview Hanna I, Zeile 71; Interview Iris I, Zeile 68). Während die einen konkretere Vorstellungen haben, ist bei anderen an der Wortwahl «Begrabungen» (Interview Jannis I, Schüler 12. Klasse, Gymnasium, 07.09.2018, Zeile 86) oder «buddeln» (Interview Iris I, Zeile 81) erkennbar, dass sie bisher kaum Kontakt mit Archäologie hatten und dieser nicht unbedingt einen (ernsthaften) wissenschaftlichen Charakter zuschreiben. Allgemein geben die Proband*innen an, dass archäologische Arbeit vor allem darin bestünde, durch das behutsame Abtragen von Bodenschichten tragbare Gegenstände zu finden (u. a. Interview Hanna I, Zeile 122). Befunde hingegen kommen nur indirekt zur Sprache, wenn die Freilegung ganzer Städte wie Pompeji Erwähnung findet (Interview Iris I, Zeile

3 Im Rahmen der Anonymisierung waren die Proband*innen dazu angehalten, sich einen eigenen Namen auszudenken, wobei ich bei dem Anfangsbuchstaben dem Alphabet folgend vorgegangen bin.

144). Die gefundenen Objekte würden dann von Archäolog*innen analysiert sowie interpretiert, indem Fragen nach der Herkunft und der Verwendung beantwortet werden (Interview Greta I, Zeile 101–104). Die Präkonzepte beruhen bei den Proband*innen auf unterschiedlichen Erfahrungen. Vor allem Medien spielen eine große Rolle bei der Gewinnung von Informationen über den Beruf. In frühen Jahren waren das zum Beispiel Bücher aus der «WAS IST WAS?»-Reihe, während heute vor allem das Fernsehen eine wichtige Rolle spielt, indem Dokumentationen nebenbei angeschaut werden oder auch in den Nachrichten von neuen archäologischen Entdeckungen berichtet wird. Interessanterweise erinnert sich Iris daran, dass ihr sogar in ihrem Spanischlehrwerk der Schule Archäologie begegnete: «Wir hatten mal [...] so einen Text über sowas, wo Jugendliche so ein Freiwilliges Soziales Jahr gemacht haben und da an einer Ausgrabungsstätte gearbeitet haben und da habe ich das ein bisschen das Bild jetzt, dass die da viel draußen sind und viel graben und dann sich viel austauschen und dann vergleichen, was die da so entdecken» (Interview Iris I, Zeile 70–72). Bei einer anderen Schülerin ist die Familie ein wichtiger Bezugspunkt, weil die Mutter Geschichte studiert hat und die Urlaubsreisen zu antiken Städten gingen, bei denen die Familienmitglieder eine Menge Informationen zu den Überresten bekamen (Interview Greta I, Zeile 153–161).

Das Interesse an Archäologie besteht bei den Schüler*innen vor allem in dem Finden von Objekten, die aus einer anderen Zeit stammen und damit «Beweise» (Interview Greta I, Zeile 178) der Geschichte repräsentieren. Sie ermöglichen einen Einblick in «eine ganz andere Welt» (Interview Hanna I, Zeile 164–165) und lassen die Verstorbenen auf eine gewisse Art weiterleben, indem sie Hinweise auf ihre Besitzer*innen geben (Interview Iris I, Zeile 168–172).

Die Präinterviews zeigen folglich, dass die Schüler*innen ein vages Bild von Archäologie haben, das auf zufällig erlebten Alltagserfahrungen basiert und das mit dem in der Schule erlernten Geschichtsverständnis verknüpft wird. Bei der Thematik zur Ausgrabung selbst lassen die Interviews hingegen erkennen, dass die Proband*innen aufgrund der mehrfachen Behandlung im Geschichtsunterricht ein umfangreiches Vorwissen zur Zeit des Nationalsozialismus haben. Dieses befindet sich allerdings auf einer allgemeinen Ebene, denn einen Bezug zu ihrer Heimatstadt Bremen konnten die Schüler*innen eher nicht herstellen. Mit diesen Präkonzepten gingen die jungen Erwachsenen zusammen mit ihrer Klasse die Ausgrabung auf dem ehemaligen Gelände des KZ-Außenlagers «Schützenhof» an.

Vor Ort bekamen die Proband*innen ihre persönliche, mit Namen gekennzeichnete Digitalkamera ausgeteilt, die sie am Ende des Präinterviews bereits ausprobieren konnten. Eine gekennzeichnete Stelle neben dem Grabungsschnitt diente zur Ablage der Kameras, damit die Schüler*innen die Hände frei hatten. Die Intervention in Form der Ausgrabung sowie des Besuchs im Unterricht verlief bei sehr sonnigem und warmem Wetter wie geplant. Im Nachhinein traf ich mich mit den Proband*innen noch einmal für das Postinterview, um über das Erlebte zu sprechen.

4.2 Postinterviews

Mithilfe der Gespräche nach der Ausgrabung ergaben sich einige Ansatzpunkte, die darauf hindeuten, dass bei den Schüler*innen an manchen Stellen durchaus Bewegung in ihren Konzepten stattgefunden hat. In Bezug auf Archäologie wurde erkennbar, dass die Proband*innen während des zweiten Interviews teilweise eine andere Ausdrucksweise wählten. Sprachten einige zuvor noch von «buddeln», nutzten sie nun auch Fachbegriffe wie «Stichgrabung» (Interview Klara II, Schülerin 12. Klasse, Gymnasium, 28.09.2018, Zeile 41), «datieren» oder «Grabungstechniker» (Interview Hanna II, Schülerin 12. Klasse, Gymnasium, 26.09.2018, Zeile 135 u. 278). Auf der Sachebene haben die Schüler*innen somit an Wissen hinzugewonnen, welches sie direkt anwenden konnten. Damit einhergehend ist auch durch die Antworten ersichtlich geworden, dass sich für die Proband*innen die Archäologie selbst als eine komplexe Wissenschaft herausgestellt hat, denn zuvor waren die Vorstellungen nur vage und die Suche nach Gegenständen zentral. Dass die archäologische Arbeit in mehrere Prozesse untergliedert ist, bei der unterschiedliche Experten und Methoden benötigt werden, wurde den Schüler*innen erst durch die aktive Auseinandersetzung mit den Arbeitsschritten ersichtlich (Interview Iris II, Schülerin 12. Klasse, Gymnasium, 28.09.2018, Zeile 310–315). Dabei war ein Objekt, welches sich auch auf vielen Fotografien der Schüler*innen wiederfindet, zentral und rief ein starkes Interesse hervor, denn dieses Stück konnte bis zum Ende der Datenerhebung noch nicht näher bestimmt werden (siehe Abbildung 1). Gerade weil es zu dem Zeitpunkt noch unklar war, welcher Zeitebene dieses zugeordnet werden kann, kamen in den Gesprächen entsprechend viele Nachfragen, ob die Frage schon geklärt werden konnte (Interview Greta II, Schülerin 12. Klasse, Gymnasium, 15.10.1018, Zeile 146–147). Dieses Objekt führte den jungen Erwachsenen vor Augen, dass Forschung ein Prozess ist, der stets neue Fragen aufwirft und eben nicht Objekte freigelegt und

direkt zugeordnet werden können. Vielmehr besteht Archäologie aus der Informationssuche und Interpretation der vorgefundenen Quellen, was in manchen Fällen auch sehr lange dauern kann. In diesem Zusammenhang bekam die ausgegrabene Nivea-Dose eine besondere Bedeutung, denn viele Probanden hätten gar nicht damit gerechnet, ein solches Objekt im Boden zu finden (Interview Jannis II, Schüler 12. Klasse, Gymnasium, 28.09.2018, Zeile 139–142). Gerade weil dieser Gegenstand so alltagsnah ist, konnte daran der Vorgang der Datierung von Fundstücken exemplarisch gut demonstriert werden.



Abbildung 1: Fotografie der Probandin Greta von dem zum Zeitpunkt der Ausgrabung noch nicht bestimmten Objekt

Bemerkenswert ist darüber hinaus, dass sich bei einigen Schüler*innen das Verhältnis von Archäologie und Geschichtswissenschaft stark verändert hat, wenn Hanna sagt: «Also klar, ich wusste, dass es Archäologie gibt und dass es auch wichtig ist für die Geschichte, aber nicht wirklich, wie viel das einem sagen kann über bestimmte Orte. Also zum Beispiel jetzt über dieses KZ-Außenlager, wie viel man da noch jetzt noch erfahren kann durch Archäologie» (Interview Hanna II, Zeile 294–297). An dieser Stelle wird deutlich, wie sich das Konzept der Schülerin von Geschichtsforschung erweitert hat, da sie zuvor insbesondere in Bezug auf die Zeitgeschichte davon ausgegangen ist, dass fast ausschließlich Text- und Bildquellen zur Erforschung dieser Zeit relevant sind. Mit der Durchführung der Ausgrabung hat sich in methodischer Hinsicht ihr Verständnis erweitert, dass

Archäologie und Geschichtswissenschaft eng miteinander verknüpft sind und sich gegenseitig befruchten können.

Letztendlich wurde von allen Studienteilnehmer*innen in den Interviews betont, dass ihnen die Ausgrabung auf thematischer Ebene sehr viel gebracht habe. Auf der einen Seite sagt eine Schülerin in Bezug auf den Nationalsozialismus und die Verbrechen, «dass es wirklich stimmt, was man im Geschichtsunterricht lernt» (Interview Klara II, Zeile 259). Auf der anderen Seite berichten die Schüler*innen über einen emotionaleren Zugang zur Thematik, was vor allem mit der Freilegung des Stacheldrahts zusammenhängt, welcher klar dem KZ-Außenlager zugeordnet werden konnte. Bereits während der Freilegung war eine emotionale Anspannung in Form von erhöhter Aufmerksamkeit innerhalb der Gruppe zu bemerken, da das Objekt die Grausamkeit der Nationalsozialist*innen «beweist» (Interview Iris II, Zeile 342): «Also gerade der Stacheldraht war so: «ok, ja». Also, dass das noch mal verdeutlicht hat, wie schlecht diese Bedingungen da waren» (Interview Hanna II, Zeile 249–250). Dabei haben sich die Schüler*innen auch Details über den charakteristischen Aufbau von militärischem Stacheldraht gemerkt, die die Grabungsleiterin der Gruppe erzählte, was noch einmal hervorhebt, wie einprägsam das Objekt für die Proband*innen war. Demzufolge konnte die archäologische Ausgrabung bewirken, dass die Schüler*innen durch die persönlichen Berührungspunkte die geschichtlichen Ereignisse auf dem Gelände reeller und damit auch emotionaler wahrnehmen, gerade weil ihnen damit verdeutlicht wurde, dass auch in ihrer Heimatstadt Bremen für alle Bewohner*innen sichtbar nationalsozialistische Verbrechen begangen wurden.

5 Reflexion der Datenerhebung und Ausblick

Die erste Sichtung des Materials konnte bereits gute Einblicke in die Wahrnehmung der Schüler*innen von ihrer archäologischen Tätigkeit geben. Vor allem das Nebeneinander von Nivea-Dose und Stacheldraht im selben Boden evokierte bei einigen Schüler*innen ein Bewusstsein von Historizität, wenn sie verwundert feststellten: «Das hat man also nicht wirklich erwartet, dass da im Boden so viel drin steckt» (Interview Iris II, Zeile 51–52); «[...] dass wir das nicht gedacht hätten [...] worauf man alles schließen kann, wenn man nur so Mauern ausgräbt» (Interview Greta II, Zeile 305–306).

Mit Blick auf die zu Beginn gestellte Forschungsfrage und damit die aktuelle Conceptual-Change-Forschung sind durchaus Verschiebungen bei den oben genannten Beispielen vor allem auf zwei Ebenen erkennbar. Auf der von Mathis definierten *Begriffsebene* (Mathis, 2015, 33) ist bei einigen Schüler*innen durchaus eine Konzeptveränderung festzustellen, weil in den Postinterviews im Vergleich zu den Präinterviews deutlich mehr Fach- beziehungsweise weniger Alltagsbegriffe verwendet wurden, wenn von der Archäologie und ihrer Methodik die Sprache war. Der andere Schwerpunkt bei Konzeptveränderungen ist im metakonzeptuellen Bereich zu finden, wenn die Schüler*innen den durchaus wichtigen Beitrag der Archäologie zur Erforschung von Geschichte erkannt haben (Limón, 2002, 271; van Drie & van Boxtel, 2008, 89). Insbesondere diese methodische Ebene historischer Forschung, das heißt die Bedeutung der Archäologie sowie der Quellen, wurde von den Schüler*innen in den Interviews häufig mit zielführenden Gedankengängen angesprochen. An dieser Stelle konnte somit das historische Denken der Schüler*innen hin zu einer wissenschaftlicheren Vorstellung gefördert werden, wenn auch bei den einzelnen Proband*innen unterschiedlich ausgeprägt (Mathis, 2015, 37). Generell ist allerdings festzustellen, dass sich die Präkonzepte der Proband*innen nicht in jedem Punkt unbedingt stark verändert, sondern eher konkretisiert beziehungsweise erweitert haben, was unter anderem darin begründet sein kann, dass sie bereits im Präinterview durchaus ein Bewusstsein für historisches Denken aufwiesen. Obwohl dies selbstverständlich kein abschließbarer Prozess ist (Günther-Arndt, 2020, 37), konnte die Teilnahme an der Ausgrabung dennoch hierzu fruchtbare Denkanstöße liefern. Insgesamt können mithilfe der Daten auch teils die oben genannten dem Einsatz von Archäologie zugeschriebenen Potenziale bestätigt werden, wenn die Schüler*innen beispielsweise durch den erwähnten Stacheldraht einen affektiveren Zugang zur NS-Thematik bekamen (Erdmann, 2004, 778). Es deutet

sich folglich an, dass mit der Durchführung einer archäologischen Ausgrabung durchaus Konzeptveränderungen initiiert werden können. Auf welcher Ebene diese stattfinden, ist jedoch immer individuell, weil jede*r Schüler*in die Ereignisse unterschiedlich wahrnimmt. In der weiterführenden Interpretation der Daten wird noch ein verstärkter Blick auf die Wahrnehmung der Schüler*innen von archäologischen Objekten als «Beweis» für die Geschichte erfolgen. Scheinbar wird reellen Funden mehr Wahrhaftigkeit als zum Beispiel Schriftquellen zugeschrieben. Im Sinne des epistemologischen Verständnisses von Geschichte sowie der im Conceptual-Change-Ansatz beinhalteten Erklärungsmuster soll dies noch einmal kritischer analysiert werden.

Die Konzeption des Studiendesigns, welches auf die autofotografische Methode gesetzt hat, ist trotz des Risikos aufgrund fehlender Erfahrungswerte beim Einsatz im geschichtsdidaktischen Forschungsbereich aufgegangen. Während der Ausgrabung haben die Kameras die Proband*innen bei der Arbeit nicht behindert und vor allem in den Postinterviews hat sich der Mehrwert der individuellen Fotografien gezeigt, indem sie als Gesprächsbasis und zugleich den Schüler*innen als Erinnerungsstütze dienten. Natürlich sind auch Spaß-Bilder entstanden und ein paar Fotografien sehr unscharf geworden, doch hat dies den Nutzen nicht beeinträchtigt, denn es ging bei den Bildern nicht primär um die fotografische Qualität, sondern vielmehr um die Inhalte und Perspektiven, die in jedem Fall ersichtlich waren.

Da es sich hier nur um eine erste Sichtung der Daten handelt, lässt eine intensivere Auswertung der Interviews im Rahmen meiner Dissertationsschrift auf weiterführende interessante Aspekte hoffen, wenn hier auch noch eine stärkere Verknüpfung mit der geschichtsdidaktischen Theorie stattfindet. Ziel ist es abschließend, auf dieser Basis dann Konsequenzen für den Einsatz von Archäologie im Geschichtsunterricht abzuleiten.

Schon jetzt kann jedoch gesagt werden, dass die Durchführung der hier im Fokus stehenden authentischen Ausgrabung als beiderseitiger Erfolg zu verbuchen ist. Während die Schüler*innen Geschichte und Archäologie hautnah erleben konnten, half die Unterstützung der Jugendlichen und jungen Erwachsenen bei der Erweiterung des Grabungsschnitts den Mitarbeitern der Landesarchäologie bei der weiteren Erschließung und Aufarbeitung der Geschichte des KZ-Außenlagers «Schützenhof». Vor allem im Hinblick auf die Opfer der vor Ort begangenen Verbrechen durch die Nationalsozialist*innen setzt das aktuelle umfangreiche und interdisziplinäre Forschungsvorhaben, von dem die Schüler*innen ein Teil sein konnten, ein wichtiges Zeichen.

Literatur

- Bakker, Lothar et al. (2004). *Weißener Erklärung: Memorandum Archäologie und Schule in Bayern*. http://www.bglv-ev.de/downloads/weissenburger_erklaerung.pdf (11.04.2022).
- Baumgärtner, Ulrich (2019). *Wegweiser Geschichtsdidaktik: Historisches Lernen in der Schule* (2. Aufl.). Paderborn: Schöningh.
- Borries, Bodo von (2008). *Historisch Denken Lernen – Welterschließung statt Epochenüberblick: Geschichte als Unterrichtsfach und Bildungsaufgabe*. Opladen: Barbara Budrich.
- Chapman, Arthur & Georgiou, Maria (2021). Powerful knowledge building and conceptual change research: learning from research on «historical accounts» in England and Cyprus. In Arthur Chapman (Hrsg.), *Knowing history in schools: powerful knowledge and powers of knowledge* (S. 72–96). London: UCL Press.
- Deutscher Museumsbund (2020). Zahlen & Fakten: Museumsstatistik. *Bulletin*, (1), 12–17. <https://www.museumsbund.de/wp-content/uploads/2020/02/bulletin20-1-online.pdf> (11.04.2022).
- Drie, Jannet van & Bostel, Carla van (2008). Historical reasoning: Towards a framework of analyzing students' reasoning about the past. *Educational Psychology Review*, 20(2), 87–110.
- Eggert, Manfred (2006). *Archäologie: Grundzüge einer historischen Kulturwissenschaft*. Tübingen: A. Francke.
- Erdmann, Elisabeth (2004). Archäologie II: Die Relevanz der Archäologie für den Geschichtsunterricht. In Waltraud Schreiber (Hrsg.), *Erste Begegnungen mit Geschichte: Grundlagen historischen Lernens* (2. Aufl., S. 773–784). Neuried: ars una.
- Fenn, Monika (2016). Schliemanns Erbe und die Medien: Von der Relevanz der Archäologie beim historischen Lernen. *Blickpunkt Archäologie*, 4, 263–267.
- Fenn, Monika (2018). Conceptual change von Vorstellungen über epistemologische Basis-konzepte bei Grundschülerinnen und -schülern fördern? Ergebnisse einer explorativen Interventionsstudie. In Monika Fenn (Hrsg.), *Frühes historisches Lernen: Projekte und Perspektiven empirischer Forschung* (S. 146–199). Frankfurt a. M.: Wochenschau.
- Flick, Uwe/von Kardorff, Ernst & Steinke, Ines (2017). Was ist qualitative Forschung? Einleitung und Überblick. In dies. (Hrsg.), *Qualitative Forschung: Ein Handbuch* (12. Aufl., S. 13–29). Reinbek: Rowohlt.
- Günther-Arndt, Hilke (2006). Conceptual Change-Forschung: Eine Aufgabe für die Geschichtsdidaktik? In Hilke Günther-Arndt & Michael Sauer (Hrsg.), *Geschichtsdidaktik empirisch: Untersuchungen zum historischen Denken und Lernen* (S. 251–277). Berlin: Lit.
- Günther-Arndt, Hilke (2020). Historisches Lernen und Wissenserwerb. In Hilke Günther-Arndt & Meik Zülsdorf-Kersting (Hrsg.), *Geschichtsdidaktik: Praxishandbuch* (S. 24–49). Berlin: Cornelsen.
- Halle, Uta & Huhn, Ulrike (2017). *Bremen-Gröpelingen, Bromberger Straße 117: Schützenhof – Internierungslager – Polenlager – KZ-Außenlager – Wohn- und Arbeitsort: Forschung und Erinnerung zur vielschichtigen Geschichte des Schützenhofs im 20. Jahrhundert*. Bremen: Edition Falkenberg.
- Hasberg, Wolfgang (2012). Kultur – Bildung – Archäologie: Anmerkungen zum Verhältnis von Archäologie und historischem Lernen. *Archäologische Informationen*, 35, 125–132.

- Hemmer, Eike & Milbradt, Robert (2005). *Bunker «Hornisse». KZ-Häftlinge in Bremen und die U-Boot-Werft der «AG Weser» 1944/45*. Bremen: Donat.
- Hopf, Christel (2017). Qualitative Interviews – ein Überblick. In Uwe Flick/Ernst von Kardorff /Ines Steinke (Hrsg.), *Qualitative Forschung. Ein Handbuch* (12. Aufl., S. 349–360). Reinbek: Rowohlt.
- Kasvikis, Kostas (2020). Developing historical thinking through archaeological heritage education: the minoan Palaikastro educational project. In Danijela Trškan&Spela Bezjak (Hrsg.), *Archaeological heritage and education: An international perspective on history education* (S. 69–90). Ljubljana: Slovenian National Commission for UNESCO.
- Kirsch, Richard (2009). *Sozialräumliche Methodik der Jugendarbeit. Aktivierende Zugänge und praxisleitende Verfahren*. Weinheim, München: VS Verlag für Sozialwissenschaften.
- Kuckartz, Udo (2018). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*. München: Beltz Juventa.
- Limón, Margarita (2002). Conceptual change in history. In Margarita Limón & Lucia Mason (Hrsg.), *Reconsidering conceptual change: Issues in theory and practice* (S. 259–297). Dordrecht, Boston, London: Kluwer.
- Mathis, Christian (2015). «Irgendwie ist doch da mal jemand geköpft worden»: Didaktische Rekonstruktion der Französischen Revolution und der historischen Kategorie Wandel. Baltmannsweiler: Schneider.
- Mayring, Philipp (2010). Design. In Günther Mey & Katja Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 225–237). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Niesyto, Horst & Marotzki, Winfried (2006). Einleitung. In Horst Niesyto & Winfried Marotzki (Hrsg.), *Bildinterpretation und Bildverstehen. Methodische Ansätze aus sozialwissenschaftlicher, kunst- und medienpädagogischer Perspektive* (S. 7–13). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Reinhard, Andrew (2018). *Archaeogaming: An introduction to archaeology in and of video games*. New York: Berghahn.
- Sachse, Johanna (2019). «Dann war das doch was ganz anderes, als wir dachten»: Konzeptveränderungen bei Schülerinnen und Schülern im Rahmen der Durchführung einer archäologischen Ausgrabung. In Christine Pflüger (Hrsg.), *Die Komplexität des kompetenzorientierten Geschichtsunterrichts* (S. 173–190). Göttingen: Vandenhoeck & Ruprecht.
- Samida, Stefanie (2010). Was ist und warum brauchen wir eine Archäologiedidaktik? Reflexionen über eine vernachlässigte Aufgabe archäologischer Forschung. *Zeitschrift für Geschichtsdidaktik*, 9, 215–226.
- Schopper, Franz & Theune, Claudia (2016). Vorwort. In: Thomas Kersting/Claudia Theune/ Axel Drieschner/ Astrid Ley Thomas Lutz (Hrsg.), *Archäologie und Gedächtnis: NS-Lagerstandorte erforschen – bewahren – vermitteln* (S. 5–6). Brandenburg, Berlin: Michael Imhof.
- Schöbel, Gunter (2016) Archäologie und Schule: Historische Entwicklung und gegenwärtige Herausforderung. *Blickpunkt Archäologie*, 4, 238–244.
- Schöbel, Gunter & Wemhoff, Matthias (2016). Archäologie und Schule: Einführung in das Thema. *Blickpunkt Archäologie*, 4, 236–237.

- Schreier, Margrit (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der der Begrifflichkeiten. *Forum: Qualitative Sozialforschung*, 15(1). <https://www.ssoar.info/ssoar/handle/document/37708> (11.04.2022).
- Sénécheau, Miriam (2006). *Archäologie im Schulbuch. Themen der Ur- und Frühgeschichte im Spannungsfeld zwischen Lehrplananforderungen, Fachdiskussion und populären Geschichtsvorstellungen. Schulbücher, Unterrichtsfilme, Kinder- und Jugendliteratur*. Dissertation. Universität Freiburg, Freiburg im Breisgau.
- Stark, Robin (2003). Conceptual Change: kognitiv oder situiert? *Zeitschrift für Pädagogische Psychologie*, 17(2), 133–144.
- Theune, Claudia (2016). Zeitgeschichtliche Archäologie in ehemaligen Konzentrationslagern: Erinnerungsort, Denkmalpflege, Forschung. In Thomas Kersting/Claudia Theune/Axel Drieschner/Astrid Ley & Thomas Lutz (Hrsg.), *Archäologie und Gedächtnis: NS-Lagerstandorte erforschen – bewahren – vermitteln* (S. 7–16). Brandenburg, Berlin: Michael Imhof.
- Trškan, Danijela & Bezjak, Špela (Hrsg.) (2020). *Archaeological heritage and education: An international perspective on history education*. Ljubljana: Slovenian National Commission for UNESCO.

Dialogische Klassengespräche führen

Fallanalyse zur Entwicklung einer professionellen Kompetenz von Geschichtslehrpersonen

Matthias Zimmermann, Miriam Moser, Anke Wischgoll, Kurt Reusser und Christine Pauli

1 Einleitung

Klassengespräche nehmen den größten Anteil an Unterrichtszeit im schweizerischen Geschichtsunterricht ein (Hodel & Waldis, 2007, 107, 113). Ihr Potenzial zur Förderung historischen Verstehens im Unterricht scheint unbestritten (u. a. Reisman, 2015; Spieß, 2015). Gleichzeitig zeigen die wenigen Untersuchungen zur Unterrichtspraxis im Geschichtsunterricht, dass Klassengespräche stark lehrpersonenzentriert und -dominiert verlaufen (z. B. Spieß, 2015). Ein Grund dafür liegt gemäß Spieß (2015, 168) in der vorherrschenden Wissensasymmetrie zwischen der Lehrperson und den Lernenden bezüglich der im Unterricht behandelten historischen Sachverhalte. Obschon eine explorative Fallanalyse von Eichner, Kaestner und von Reeken (2019, 258) darauf hinweist, dass Lehrpersonen durchaus über adaptive und lernförderliche Gesprächsleitungsstrategien verfügen, überwiegen vor allem kurzschrittige, mehrheitlich mit geschlossenen Fragen geprägte Muster der Gesprächsleitung durch die Lehrpersonen (Reisman, 2015, 35–37). Diese Erkenntnisse zur Praxis der Klassengespräche im Geschichtsunterricht decken sich mit denjenigen aus anderen Fächern (Heller & Morek, 2019; Pauli, 2010; Resnick, Asterhan, Clarke & Schantz, 2018).

Demgegenüber stehen Ansätze zur Etablierung einer dialogischen Gesprächskultur im Unterricht (Überblick: Kim & Wilkinson, 2019). Die daraus entwickelten praxisorientierten Ansätze bieten den Lehrpersonen konkrete Gesprächsleitungsstrategien, die sie in Klassengesprächen einsetzen können (u. a. Michaels, O'Connor, Hall & Resnick, 2010; Reznitskaya & Wilkinson, 2017). Forschungsergebnisse verschiedener Fächer belegen die Lernwirksamkeit dialogischer Klassengespräche für fachliche und diskursive Kompetenzen. Dialogische

Klassengespräche zeichnen sich durch die erhöhten Sprechzeitanteile sowie eine veränderte Rolle der Lernenden im Gespräch aus (Heller & Morek, 2019). Trotz der positiven Ergebnisse zeigen Interventionsstudien auch, wie herausfordernd die Etablierung einer dialogischen Gesprächskultur für die Lehrpersonen ist (Resnick et al., 2018; Weil, Seidel, Schindler & Gröschner, 2020). Dies lässt sich auch auf den Geschichtsunterricht übertragen. Dabei stellt sich die Frage, wie die Gesprächskompetenz von Geschichtslehrpersonen hinsichtlich einer dialogischeren Gesprächsleitung mittels einer Fortbildung weiterentwickelt werden kann.

Der vorliegende Beitrag setzt hier an. Er untersucht anhand einer Fallanalyse, wie sich langfristig angelegte Professionalisierungsmaßnahmen zu dialogischer Gesprächsführung auf die Sprechzeitverteilung, die Gesprächsleitung einer Geschichtslehrperson der Sekundarstufe 1 und den argumentativen Gehalt von Lernendenaussagen auswirkt.

2 Theoretischer Rahmen

2.1 Dialogische Klassengespräche: Lehr-lern-theoretische Grundlage

Trotz einer seit einigen Dekaden stattfindenden empirischen Auseinandersetzung mit Merkmalen lernwirksamer Klassengespräche scheinen sich die Erkenntnisse daraus bisher kaum auf die Unterrichtspraxis zu übertragen. Gegenüber den empirisch belegten Erkenntnissen steht die häufig beobachtete Praxis der asymmetrischen Unterrichtskommunikation mit der Lehrperson als einziger Wissensquelle im Unterrichtsalltag (u. a. Cazden, 2001; Resnick et al., 2018). Die Gesprächsstruktur solcher traditioneller Klassengespräche weist häufig das Muster «Initiation-Response-Evaluation» (I-R-E) auf (Mehan, 1979). Obwohl dieses Muster von Mehan vorurteilsfrei als Sequenzbeschreibung eingeführt wurde, gilt es heute als Inbegriff für das kleinschrittig geführte Lehr-Lern-Gespräch (Becker-Mrotzek & Vogt, 2009; Pauli, 2010). Der Hauptkritikpunkt an solchen Gesprächen ist die Tendenz, dass sie zu Ratespielen verkommen, bei denen die Lernenden die eine richtige Lösung herausfinden sollen und ihre Gesprächsbeteiligung auf die Rolle von Stichwortgebenden reduziert wird (Pauli, 2010, 145–146).

Aus Sicht einer sich etablierenden kognitivistischen und konstruktivistischen Lehr-Lern-Theorie stellen solche lehrpersonenorientierten, quasimonolo-

gischen Gespräche keine ausreichenden Lerngelegenheiten für die Lernenden dar (Alexander, 2020; Pauli, 2010; Resnick, Michaels & O'Connor, 2010). Im Gegensatz dazu stehen dialogische Klassengespräche, an denen sich alle beteiligen können und die geprägt sind durch Ko-Konstruktivität von Wissen sowie einer wertschätzenden wie auch kritischen Gesprächskultur (Pauli & Reusser, 2018). Auf lehr-lern-theoretischer Grundlage bedeutet dies für Klassengespräche im Unterricht, dass die Lernenden integraler Bestandteil einer *Wissensbildungs-gemeinschaft* sind, in der sie Verantwortung für das Lernen übernehmen und gegenseitig einfordern (Lave & Wenger, 1991). Die Klassengespräche sind zwar *lehrpersonengeleitet*, aber *lernendenorientiert* in dem Sinne, dass Lernende angeleitet durch die Lehrperson Raum für eigenständige Sinnbildungen erhalten und ihr Lernen als gemeinsames Ziel im Zentrum steht (Pauli & Reusser, 2018; Resnick et al., 2018). Kognitive Konflikte ausgelöst in sozialen Situationen stellen dabei ein zentrales Element im Lernprozess dar (Reusser & Pauli, 2015, 914). Diese bedingen einen Dialog, der Verstehen und Aushandlungen zwischen den Beteiligten notwendig macht und idealtypisch durch immer neue soziokognitive Klärungsbedürfnisse vorangetrieben wird (Reusser & Pauli, 2015, 914–915). Zudem sollen die Lernenden Unterricht als *cognitive apprenticeship* mit der Lehrperson als Modell für die Entwicklung eigener fachspezifischer Denkwege und -strategien erleben (Collins, Brown & Holum, 1991). Durch den Einsatz adaptiver kognitiver Lerngerüste (*scaffolds*) fördert die Lehrperson in der mündlichen Interaktion die Lernenden bei der selbstständigen Durchquerung ihrer Zone der proximalen Entwicklung (Wood, Bruner & Ross, 1976). Eine Voraussetzung dafür ist, dass die Lernenden vermehrt die Möglichkeit erhalten, ihre Denkwege und -prozesse zu artikulieren und zu reflektieren (Collins et al., 1991). Im dialogischen Klassengespräch erfolgt Lernen über den sozialen Austausch. Durch *Ko-Konstruktion* kann ein vertieftes fachliches Verstehen erreicht werden, wobei gleichzeitig soziokulturelle überfachliche Diskurspraktiken eingeübt werden (Heller & Morek, 2019; Reusser & Pauli, 2015).

Deutlich wird aus diesen drei Aspekten die entscheidende Rolle von Sprache für das verständnisorientierte Lernen. Einerseits sollen in dialogischen Klassengesprächen die eigenen kognitiven Prozesse verbalisiert werden, andererseits sind die Gesprächsteilnehmenden aufgefordert, Äußerungen aufzunehmen, zu verarbeiten, daran anzuknüpfen und damit zum gemeinsamen Wissensaufbau beizutragen.

2.2 Lernpotenzial dialogischer Klassengespräche im Geschichtsunterricht

Für das Fach Geschichte können dialogische Klassengespräche verschiedenen epistemische Funktionen in einem vollständigen Lernzyklus historischen Verstehens erfüllen (u. a. Handro, 2013; Hug, 1980; Luís & Rapanta, 2020; van Boxtel & van Drie, 2017). Durch *Konfrontation* mit einer zunächst fremden Lebenswelt können Vermutungen und Hypothesen zu einem gemeinsamen Problemraum gemacht werden (Fina, 1978; Reisman, 2015). Zu den daraus entstandenen Fragen oder Vermutungen gilt es durch die leitfragengeleitete *Erarbeitung* der notwendigen Informationen aus Dokumenten Antworten zu generieren oder einen Sachverhalt zu beschreiben. Die dazu benötigten Strategien der Dokumentanalyse können im Klassengespräch modelliert und deren Anwendung gemeinsam geübt werden (Reisman et al., 2018; Zülsdorf-Kersting, 2020). Gleichzeitig fördern dialogische Klassengespräche durch die mündliche Artikulation das Begriffslernen (Eichner et al., 2019), die Offenlegung von Denkprozessen (Fina, 1978, 39–40) und es dient als argumentativer Aushandlungsraum von plausiblen historischen Erzählsträngen (Hug, 1980; van Boxtel & van Drie, 2017). Zur *Konstruktion plausibler Narrative* als Ziel des Geschichtsunterrichts benötigen die Lernenden außerdem eine konzeptionelle Syntax (z. B. Ursachen–Wirkung; Wandel–Kontinuität; Perspektivität) zur fachlich adäquaten Strukturierung und Deutung der erarbeiteten Informationen (Günther-Arndt, 2019; Reisman et al., 2018). Dialogische Klassengespräche als Element eines kompetenzorientierten Geschichtsunterrichts fördern folglich den fachspezifischen Diskurs entlang von Teiloperationen historischen Verstehens zum Aufbau narrativer Kompetenz (schon vor der Kompetenzorientierung bei Fina, 1978, 40–42; Zülsdorf-Kersting, 2020). Zusätzlich erweitern dialogische Klassengespräche den individuellen Prozess historischen Denkens (u. a. Gautschi, 2009) um eine kollektiven und intersubjektive Aushandlung narrativer Sinnbildungen zu einem vertieften historischen Verstehen.

Obwohl Klassengespräche einen hohen Anteil der Unterrichtszeit ausmachen und das Lernpotenzial unbestritten ist, gibt es zurzeit wenige empirische Untersuchungen im Geschichtsunterricht. Einige explorative Studien betonen die zentrale Rolle der Lehrperson sowohl bei der Planung als auch der Durchführung von Klassengesprächen (Havekes, van Boxtel, Coppen & Luttenberg, 2017; Spieß, 2015). Des Weiteren zeigt die Untersuchung von van Drie und van de Ven (2017, 537–538), dass sich dialogisch geleitete Klassengespräche positiv auf die Qualität von schriftlichen Essays auswirkten. Die Lernenden waren fähig, die fragmentarischen Er-

zählstränge aus der Diskussion zu reproduzieren und in einen plausiblen Zusammenhang zu bringen. Reismans (2015) Untersuchung zu Klassengesprächen im Rahmen ihrer Interventionsstudie zum Einsatz von quellenbasierten Unterrichtseinheiten im Geschichtsunterricht verdeutlicht die Herausforderung dialogischer Gesprächsleitung (vgl. auch Havekes et al., 2017). Obwohl den Lehrpersonen Gesprächsleitungsstrategien in Form von fachunspezifischen und fachspezifischen *talk moves* sowie Aufgabenstellungen und Unterlagen zur Verfügung gestellt wurden, gelang es den Lehrpersonen kaum, fachlich gehaltvolle Klassengespräche durchzuführen. Diese Ergebnisse verdeutlichen die Wechselwirkung aus kognitiv aktivierenden Lernaufgaben mit entsprechenden Dokumenten, einem diskussionsförderlichen Unterrichtsetting und der Gesprächsleitungscompetenz der Lehrperson. Dass die Lernenden in Gruppengesprächen mit entsprechenden Lernaufgaben und Dokumenten auch ohne die Lehrperson fähig sind, fachlich zu argumentieren, deuten einige Untersuchungen bereits an (Monte-Sano & Allen, 2019; Pontecorvo & Girardet, 1993).

2.3 Ansatz dialogischer Gesprächsleitung im Geschichtsunterricht – Accountable Talk aus geschichtsdidaktischer Perspektive

Das unter 2.2 aufgeführte Lernpotenzial von Klassengesprächen bedingt, dass Lehrpersonen dialogische Gesprächskompetenzen entwickeln. Auf lehr-lern-theoretischer Grundlage (siehe Abschnitt 2.1) wurden seit der Jahrtausendwende praxisorientierte Handlungsanleitungen zur dialogischen Gesprächsführung für Lehrpersonen entwickelt (Resnick et al., 2018). Diese zeichnen sich dadurch aus, dass sie Gesprächsleitungsstrategien für die Lehrpersonen aus der Analyse von gelungenen, dialogischen Gesprächssequenzen identifizierten und für die Lehrpersonen als produktive fachunspezifische Gesprächsleitungsstrategien zugänglich machten (Kim & Wilkinson, 2019; Resnick et al., 2018).

Hinsichtlich deren Umsetzung im Geschichtsunterricht formulieren Reisman et al. (2018) in Anlehnung an den praxisorientierten Ansatz von *Accountable Talk* (Michaels et al., 2010) sowohl fachspezifische als auch transversale Qualitätsmerkmale für textbasierte Klassengespräche im Geschichtsunterricht. *Accountable Talk* unterscheidet drei Qualitäten der Verantwortlichkeit (*accountability*) als zentrale Dimensionen dialogischer Klassengespräche: die Verantwortlichkeit a) gegenüber der *Wissensbildungsgemeinschaft*; b) gegenüber dem *Wissen* und c) gegenüber dem *folgerichtigen Denken und Argumentieren* (Pauli & Reusser, 2018, 368). Um fachspezifische Aspekte dieses Ansatzes zu präzisieren, lässt sich

die Rolle der Geschichtslehrperson hinsichtlich dialogischer Gesprächsleitung anhand dieser drei Dimensionen weiter ausführen (Reisman et al., 2018). Damit eine *Wissensbildungsgemeinschaft* entstehen kann, sollen die Lernenden aufeinander ausgerichtet werden. Dies gilt als Basis für ein konstruktivistisches Lehr-Lern-Verständnis (siehe Abschnitt 2.1). Demnach soll eine dialogische Gesprächskultur entwickelt werden, in der die Lernenden einander ernsthaft zuhören, an geäußerten Ideen anknüpfen und diese respektvoll erweitern oder kritisieren (Resnick et al., 2010).

Klassengespräche im Geschichtsunterricht sind geprägt von der Auseinandersetzung mit historischen Sachverhalten oder Akteur*innen. In der unterrichtlichen Auseinandersetzung mit dem Sachverhalt sollen die Lernenden Verantwortlichkeit übernehmen gegenüber dem *Wissen*, das gemeinsam generiert wird. Deshalb ist es Aufgabe der Lehrperson, die Lernenden am fachspezifischen Wissen in Form von inhaltsbezogenen Begriffen und Strategien zu orientieren (Günther-Arndt, 2019). In der fachlichen Auseinandersetzung mit historischen Sachverhalten innerhalb der Wissensbildungsgemeinschaft kann gemeinsam ein begriffliches Verständnis eines Sachverhalts aufgebaut und ein strukturierendes Fachvokabular erarbeitet werden (Eichner et al., 2019; van Boxtel & van Drie, 2017). Zusätzlich benötigen die Lernenden Strategiewissen zur Analyse von unterschiedlichen Arten von Dokumenten als Informationsträger aus und über die Vergangenheit (Britt & Rouet, 2012). Geschichtslehrpersonen sollen in Klassengesprächen das analytische Vorgehen modellieren und die Anwendung der Strategien in der gemeinsamen Erarbeitung adaptiv unterstützen (Reisman et al., 2018; van Boxtel & van Drie, 2017).

Die Verantwortlichkeit gegenüber dem *folgerichtigen Denken und Argumentieren* besteht im Geschichtsunterricht darin, dass die Lernenden leitfragengeleitet in der Auseinandersetzung mit historischen Sachverhalten Sinn bilden oder bereits vollzogene Sinnbildungen kritisch hinterfragen (Hug, 1980; Luis & Rapanta, 2020). Dabei stützen sie sich auf die analysierten Dokumente und verknüpfen die daraus erarbeiteten Informationen mittels inhaltsstrukturierender Erkenntniskonzepte (*metaconcepts*) zu plausiblen narrativen Strängen (van Drie & van Boxtel, 2008). Die Plausibilität dieser narrativen Sinnbildungen kann durch die intersubjektive Überprüfung der Wissensbildungsgemeinschaft abgesichert werden (Barricelli, 2015; Reisman et al., 2018) und Prozesse historischen Verstehens im mündlichen Diskurs offengelegt werden (Zülsdorf-Kersting, 2020, 402–403). Dies erlaubt es der Lehrperson zudem, die Lernenden auf inhaltsstrukturierende Erkenntniskonzepte (z. B. Ursachen, Wirkungen, Entwicklung, Pers-

pektiven) auszurichten und sie aufzufordern, ihre Aussagen zu begründen sowie die Informationsbasis ihrer Folgerungen zu belegen (Reisman et al., 2018; van Boxtel & van Drie, 2017). Dadurch können sich die Lernenden als eigenständige Sinnbildende innerhalb einer Wissensbildungsgemeinschaft engagieren und zudem in ihrer narrativen Kompetenz gefördert werden. Eine Voraussetzung dafür ist die Kompetenz zu dialogischer Gesprächsleitung der Lehrperson.

2.4 Merkmale erfolgreicher Lehrpersonenfortbildungen

Gemäß einer Metaanalyse von Lipowsky und Rzejak (2019, 42–43) zeichnen sich erfolgreiche Lehrpersonenfortbildungen vor allem durch einen spezifischen Fokus auf einen unterrichtlichen Aspekt aus, wobei Fortbildungen häufig zu breit konzipiert werden (Desimone & Garet, 2015, 259). Des Weiteren konstituiert die Verschränkung aus Input-, Erprobungs-, Feedback und Reflexionsphasen ein Gelingensmerkmal. Dabei betonen Lipowsky und Rzejak (2019, 39–40), dass innerhalb dieser Phasen die soziale Eingebundenheit in einer professionellen Lerngemeinschaft, Handreichungen für die praktische Umsetzung der Fortbildungsinhalte, das Erleben der Wirksamkeit der Professionalisierungsmaßnahmen im eigenen Unterricht sowie die videobasierte Reflexion und Coaching weitere Merkmale für erfolgreiche Fortbildungen darstellen (Lipowsky & Rzejak, 2019, 46–57). Dabei scheinen unterrichtliche Verhaltensveränderungen nur zu gelingen, wenn die obengenannten Professionalisierungsmaßnahmen langfristige angelegt werden. Erste Annäherungen an die Erfassung der Entwicklung professioneller Kompetenz von Geschichtslehrpersonen werden vor allem mit angehenden Lehrpersonen durchgeführt (u. a. Heuer, Resch & Seidenfuß, 2017). Hinsichtlich Lehrpersonenfortbildungen zu dialogischer Gesprächsleitung im Geschichtsunterricht gibt es kaum Interventionsstudien, die den Aufbau einer professionellen Kompetenz von Geschichtslehrpersonen untersuchten (Reisman & Enumah, 2020; van de Pol, Brindley & Higham, 2017).

3 Fragestellungen

Die vorliegende Arbeit knüpft hier an. Mittels einer exemplarischen Fallanalyse wird die Entwicklung einer Lehrperson während einer einjährigen Fortbildung im Rahmen einer Interventionsstudie zu dialogischer Klassengesprächsleitung nachgezeichnet. Dazu wird untersucht, welche Effekte die Fortbildung auf das Verhalten der Lehrperson und der Lernenden in Klassengesprächen hat (Lipowsky & Rzejak, 2019, 36). Dazu wurden a) quantitativ die Sprechzeitanteile aller Beteiligten als Indikator für die mögliche veränderte Rolle der Lernenden erfasst; b) inhaltsanalytisch einerseits das Gesprächsverhalten der Lehrperson als Indikator für eine dialogische Gesprächsleitung; und andererseits c) der argumentative Gehalt der Lernendenäußerungen als Indikator für deren inhaltliche Qualität ausgewertet.

Im Rahmen der Interventionsstudie «Socrates 2.0» wurde eine Lehrpersonfortbildung mit dem Ziel durchgeführt, die Lehrpersonen beim Aufbau einer dialogischen Gesprächskompetenz zu fördern (Pauli & Reusser, 2018). Daraus leitet sich die erste Fragestellung ab, die sich auf beobachtbare Verhaltensveränderungen der Lehrperson im Klassengespräch bezieht.

Fragestellung 1: Wie entwickelte sich das Gesprächsleitungsverhalten einer Geschichtslehrperson im Verlauf der Fortbildung von Vor- zum Nachtest?

- a) Welche Veränderungen zeigen die *Sprechanteile* der Lehrperson im Vergleich mit der Gruppe von Schüler*innen?
- b) Welche Entwicklung zeigen die Lehrpersonen bezüglich einer *dialogischeren Gesprächsleitung* in Klassengesprächen?

Durch die komplementären Rollen von Lehrperson und Lernenden in dialogischen Klassengesprächen stellt sich auch die Frage nach den Auswirkungen der Fortbildung auf Merkmale der Lernendenbeteiligung an Klassengesprächen. Betrachtet man die Aussagen der Lernenden als *discursive nodes* (Binnenkade, 2015, 31–33), dann kann untersucht werden, inwiefern die Lernenden Informationen aus im Unterricht eingesetzten Dokumenten einbringen. Des Weiteren gilt es im Sinne der argumentativen Aushandlung von plausiblen Narrativen, die eigenen Aussagen zu begründen. Beides stellen Indikatoren für Effekte der veränderten Gesprächsleitung der Lehrperson hinsichtlich der Verantwortlichkeit gegenüber dem Wissen und dem folgerichtigen Denken auf den fachlich-argumentativen Gehalt der Lernendenäußerungen dar. Daraus ergeben sich folgende Fragestellungen:

Fragestellung 2: Wie entwickelt sich der argumentativ-fachliche Gehalt der Lernendenaussagen in Sinnbildungsgesprächen im Verlauf der Lehrpersonenfortbildung?

- a) Wie entwickelten sich während der Intervention die Anteile von begründeten Aussagen der Lernenden?
- b) Inwiefern veränderten sich die Anteile von Lernendenaussagen, die sich inhaltlich auf im Unterricht eingesetzte Dokumente beziehen?

4 Methodisches Vorgehen

4.1 Beschreibung der Lehrpersonenfortbildung und Forschungsdesign «Socrates 2.0»

Auf der Grundlage von Erkenntnissen erfolgreicher Lehrpersonenfortbildungen wurde die einjährige Intervention «Socrates 2.0»¹ zur Förderung dialogischer Gesprächskompetenzen von Lehrpersonen in den Fächern Mathematik ($N = 6$) und Geschichte ($N = 3$) entwickelt (vgl. Pauli & Reusser, 2018). Die einzelnen Elemente der Fortbildung für die Geschichtslehrpersonen werden nun ausführlicher beschrieben.

Die Fortbildung «Socrates 2.0» bestand aus zwei unterschiedlichen Bedingungen (siehe Abbildung 1). Die freiwillig teilnehmenden Lehrpersonen wurden vor der ersten Präsenzveranstaltung zufällig in zwei unterschiedliche Interventionsbedingungen aufgeteilt: Die *allgemeine Bedingung* bestand aus einem eintägigen Input zu *Accountable Talk* und dem Einsatz von Gesprächsleitungsstrategien im Unterricht. Dazu wurden Werkzeuge zu den Dimensionen von *Accountable Talk* mit konkreten vorformulierten Gesprächsleitungsstrategien in Form von Satzanfängen abgegeben. Anhand von Fremdvideos von Klassengesprächen aus Mathematik- und Geschichtslektionen wurde ein gemeinsames Verständnis von dialogischen Klassengesprächen aufgebaut sowie der konkrete Einsatz der Strategien modelliert und Einsatzmöglichkeiten diskutiert. Diesen allgemeinen Input absolvierten alle teilnehmenden Lehrpersonen gemeinsam.

1 Finanziert vom Schweizerischen Nationalfond SNF als Projekt-Nr. 100019-166070 / 1 «Lernwirksame Klassengespräche führen – eine Interventionsstudie zur Förderung der Gesprächskompetenz von Lehrpersonen» («Socrates 2.0»).

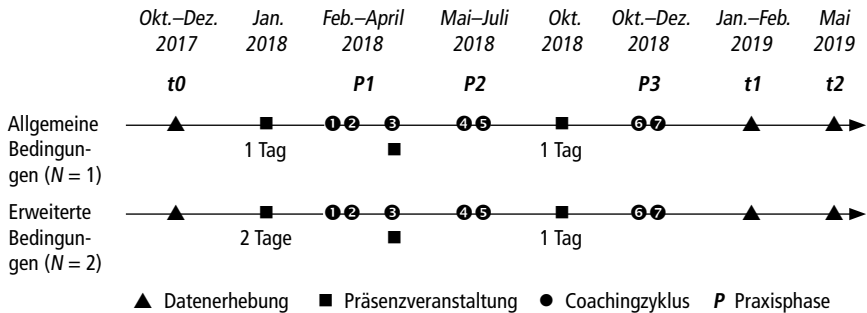


Abbildung 1: Elemente der Lehrpersonenfortbildung Geschichte und Forschungsdesign Socrates 2.0

Die *erweiterte Bedingung* bestand aus einem an die allgemeine Fortbildung anschließenden zweiten Tag mit Inputs zur fachspezifischen Gestaltung von Gesprächsanlässen unter Einbeziehung fachdidaktischer Aspekte wie der Wahl von Dokumenten und die Formulierung von Gesprächsleitfragen. Nach dieser ersten Präsenzveranstaltung folgten in Anlehnung an Allen, Pianta, Gregory, Mikami und Lun (2011) zwei Coachingphasen bestehend aus insgesamt fünf videobasierenden Coachingzyklen (C) verteilt über einen Zeitraum von sechs Monaten. Der Inhalt der Coachinggespräche wurde individuell gestaltet. Als Basis für die Gestaltung diente einerseits die Zugehörigkeit zu einer Interventionsbedingung und andererseits das gezeigte Unterrichtshandeln in den videografierten Lektionen der Lehrperson.

Am Ende der ersten Coachingphase (P1) bestehend aus drei Coachingzyklen erhielten die Geschichtslehrpersonen der *erweiterten Bedingung* einen zusätzlichen Werkzeugkasten zum «Historischen Denken» (vgl. Zimmermann, im Druck). Dieser lag schwerpunktmäßig auf der Gestaltung von Gesprächsanlässen und beinhaltete *Trigger Questions* entlang inhaltsstrukturierender Konzepte des Fachs Geschichte als Leitfragen für Gespräche. Anschließend an die zweite Coachingphase (P2) folgte eine zweite Präsenzveranstaltung getrennt nach Bedingung. Diese Inputveranstaltung bestand aus einer reflexiven Rückschau und einem Input zu produktiven Aufgabenstellungen (Reusser, 2014). Daran knüpfte eine letzte Coachingphase (P3) mit zwei Coachingzyklen an.

4.2 Fallbeschreibung Roland

Für den vorliegenden Artikel wurde der Lehrer Roland als exemplarische Fallanalyse ausgewählt (Trittel, 2010). Es handelt sich um einen Geschichtslehrer mit 25 Jahren Unterrichtserfahrung. Außerdem betreut er Praktikant*innen während der Lehrpersonenausbildung. Rolands Klasse befindet sich auf dem niedrigsten Leistungsniveau «Grundanforderungen» der schweizerischen Sekundarstufe 1 (Alter 13–15 Jahre). Deren Leistungsniveau schätzt Roland im Vergleich mit anderen Klassen desselben Niveaus als sehr heterogen ein. Des Weiteren betont er die mangelnde Sprachkompetenz einzelner Lernender, was sie an der aktiven Teilnahme an Klassengespräche hindere (Interview, to, ID 111, 11min 55s bis 12min 16s). Roland absolvierte im Rahmen der Fortbildung das Basistraining ohne fachdidaktische Erweiterung. Dadurch ermöglicht die Fallanalyse Erkenntnisse zu Effekten der allgemeinen Professionalisierungsmaßnahmen.

4.3 Datenauswahl und Datenanalyse

4.3.1 Datenauswahl

Im vorliegenden Beitrag wurden Videoanalysen anhand der Prozessdaten, videoografierte Lektionen der Coachingzyklen und Datenerhebungen vor und nach der Fortbildung (to und t₁) durchgeführt. Zum Messzeitpunkt t₂ konnten im Falle von Roland keine Datenerhebung stattfinden. Aus jeder Praxisphase und den Datenerhebungen wurde, wenn beobachtbar, ein «Sinnbildungsgespräch» des Lehrers ausgewertet. Diese wurden vom Autor induktiv definiert als «Synthesgespräche zu argumentativer Urteilsbildung, Einschätzung von Entwicklungen und Perspektiven oder zur reflexiven Auseinandersetzung mit der eigenen Gegenwart». Bei fehlenden Sinnbildungsgesprächen wurden «Erschließensgespräche»² untersucht, aber keine Einstiegsgespräche. Tabelle 2 bietet eine Übersicht über die sechs analysierten Klassengespräche von Roland. Handelt es sich bei den ausgewerteten Klassengesprächen nicht um ein Sinnbildungsgespräch, wurden dies grau hinterlegt.

2 Definiert als Klassengespräche, in denen es um die analytische Auseinandersetzung mit Dokumenten geht. Sie zeichnen sich vor allem durch das Sammeln von Informationen aus.

Tabelle 1: Übersicht der analysierten Klassengespräche von Roland

Fall	Zeitpunkt	Gesprächsdauer
Roland (N = 6)	t0	17 min 40 s
	C1	16min 58 s
	C4	12 min 03 s
	C6	16 min 45 s
	t1	11 min 56 s (KG1) 10 min 40 s (KG2)
	t2	Keine Daten

Bemerkung: In t1 wurden zwei Sinnbildungsgespräche geführt.

4.3.2 Forschungsdesign und Datenanalyse

Mittels eines sequenziellen *mixed-methods*-Forschungsdesigns wurde für jeden Analyseschritt die Datenbasis fokussiert (Kuckartz, 2014). Zunächst wurden alle Lektionen zur Erfassung der Sozialform «Klassengespräch» im *event-sampling*-Verfahren codiert (Hodel & Waldis, 2007). Anhand der erfassten Klassengespräche konnten induktive Kategorien zur Einteilung der Klassengespräche nach ihrer epistemischen Funktion gebildet und diese inhaltsanalytisch den entsprechenden epistemischen Funktionen im Lernprozess (siehe Abschnitt 2.1) zugeordnet werden.

Zur Beantwortung der ersten Fragestellung wurden die Sprechzeitanteile der Klassengespräche in Tabelle 2 mittels der Software MaxQda (Version 12.3.9) quantitativ erfasst. Zur Ermittlung der Sprechanteile wurde nur die tatsächlich gesprochene Zeit aller am Gespräch Beteiligten erfasst. Als Folge davon bleiben bei der Angabe der prozentualen Sprechanteile bestimmte Restanteile übrig. Diese beinhalten Sprechpausen, kurze Momente der Tafelanschrift oder Notizführung.

Anschließend wurden Rolands Sinnbildungsgespräche *turnweise* inhaltsanalytisch (Mayring, 2010) codiert zur Erfassung des *Gesprächsverhaltens der Lehrpersonen* anhand deduktiv gebildeter Kategorien. Dazu wurde das bereits bestehende Kategoriensystem von Kooloos, Oolbekkink-Marchand, Kaenders und Heckman (2020) mit Bezug auf die Qualitätsdimensionen von *Accountable Talk* (siehe Abschnitt 2.3) angepasst. Auf der obersten Ebene werden die Kategorien folgendermaßen unterschieden: a) *Regulierende* Handlungen beinhalten organisatorische Anweisungen und Rederechtserteilungen. b) *Konvergente* Handlungen umfassen alle Impulse der Lehrperson, die den Gesprächsverlauf in engen Bahnen halten und klar eingrenzen. c) *Divergente* Impulsen bestehen aus offenen

Fragen der Lehrperson und Gesprächsstrategien im Sinne von *Accountable Talk*. Dabei sollten unter Bezugnahme auf *Accountable Talk* (Abschnitt 2.3) alle Dimensionen dialogischer Gesprächsleitung anhand divergenter Handlungen erfasst werden.

Nachfolgend werden die drei Oberkategorien kurz zusammengefasst und dazugehörige Codekategorien kurz beschrieben (siehe Tabelle 3). Die *regulierenden* Handlungen (a) bestehen aus 1) disziplinarisch-regulierenden Interventionen und 2) Rederechtserteilungen, welche auch Redeermunterung und Wiederholungsaufforderungen beinhalten. Die *konvergenten* Handlungen (b) umfassen geschlossene Fragen oder Aufforderung zu Zusammenfassung (Kohärenz), Eigenleistungen der Lehrperson wie die Hervorhebung wichtiger Aspekte, Reformulierungen oder Inputs und Bewertungen von Lernendenaussagen ohne weitere Anschlussmöglichkeit und das Abschließen des Gesprächs. Die *divergenten* Impulse (c) bestehen aus offenen Fragen als Gesprächseröffnungen sowie aus an Lernendenaussagen anknüpfenden Impulsen der Dimensionen 1) *Wissensbildungsgemeinschaft*, dazu gehören allgemeine und gezielte Aufforderungen an die Lernenden zu einer Stellungnahme zu einer vorhergehenden Lernendenaussage und Äußerungseinladungen an alle Lernenden; 2) *Wissen*, das umfasst Aufforderungen zur eigenen gezielten oder allgemein anknüpfenden Begriffsklärung, Aussagenpräzisierung sowie Revoicing der Lehrperson; und 3) *Folgerichtiges Denken und Argumentieren*, welche Aufforderung zur Begründung, Erweiterung bestehender Denkwege durch Anschlussfrage oder Herausforderung von Lernendenaussagen beinhaltet.

Tabelle 2: Kategoriensystem zur Erfassung der Gesprächsleitung von Lehrpersonen

Oberkategorie	Regulierende Handlungen	Konvergente Handlungen	Divergente Impulse
Codes	Unterrichtsregulierung	Kohärenz Geschlossene Frage	Eröffnung
	Rederechtserteilung	Hervorhebung Input Reformulierung	<i>Wissensbildungsgemeinschaft</i> <ul style="list-style-type: none"> • Offene Einladung • Allgemeine Verknüpfung • Gezielte Verknüpfung
		Evaluation Schließen	<i>Wissen</i> <ul style="list-style-type: none"> • Allgemeine präzisierende Ausführungen • Gezielte präzisierende Ausführungen • Revoicing³
			<i>Folgerichtiges Denken und Argumentieren</i> <ul style="list-style-type: none"> • Begründungsaufforderung⁴ • Erweiternde Anschlussfrage⁵ • Herausforderung⁶

Die Codierung von einem Drittel des Datenmaterials zur ersten Fragestellung wurde vom Autor und einer Hilfskraft ohne Bezug zum Projekt durchgeführt (*Cohens Kappa* .72–.94). Aufgrund der Multifunktionalität mündlicher Diskurse als komplexe Analysedaten kann von sehr guten bis ausgezeichneten *Kappa*-Werten gesprochen werden (Hennessy, Howe, Mercer & Vrikkki, 2020).

Zur Beantwortung der zweiten Fragestellung wurden die Sinnbildungsgespräche anhand deduktiv entwickelter Kategorien zur Erfassung der Anteile an a) *begründeten* und b) *dokumentbezogenen* Aussagen der *Lernenden* wiederum *turnweise* inhaltsanalytisch codiert. Jeder Turn der Lernenden wurde in zwei Durchgängen codiert.

3 Ergänzung aus: Tytler und Aranda (2015, 432–433).

4 Siehe Fußnote 3

5 Siehe Fußnote 3

6 Siehe Fußnote 3

Tabelle 3: Codierung der Lernendenaussagen nach argumentativ-diskursivem Gehalt

1. Codierung	<i>Begründete Aussage</i> <u>Indikatoren:</u> weil, da, deshalb, deswegen ...	<i>Nicht begründete Aussage</i>	<i>Rest</i> Aussage ohne Inhalt
2. Codierung	<i>Belegte Aussage</i> nachweislicher Inhaltsbezug aus Dokumenten der Lektion	<i>Nicht belegte Aussage</i> Informationsbezug ist unklar	<i>Rest</i> Aussage ohne Inhalt

Wies eine inhaltliche Aussage weder eine Begründung noch einen nachweislichen Dokumentenbezug auf, wurde sie der Kategorie «Behauptung» zugeordnet (Megill, 2007, 96–98). Die Restanteile beinhalten Äußerungen der Lernenden ohne Inhalt, wie Kurzantworten («Ja»), organisatorische Rückfragen («Welches Arbeitsblatt ist das?») oder Zwischenbemerkungen («Ich wollte zu etwas anderem was sagen»). Insgesamt wurden wiederum zwei Sinnbildungsgespräche von zwei Codierern zur Gewährleistung der Reliabilität codiert und *Cohens-Kappa*-Werte von .79 bis .92 erzielt.

5 Ergebnisse

5.1 Fallanalyse Roland

5.1.1 Sprechzeitanteile

Zur Beantwortung der Fragestellung 1a) wurden die Sprechzeitanteile sowie die Anzahl Turns von Roland und seiner Klasse quantifiziert. Tabelle 6 zeigt die Entwicklung der Sprechzeitanteile und der Turns von t_0 bis t_1 .

Tabelle 4: Entwicklung Turnverteilung und Sprechzeitanteile Roland

Zeitpunkt	Turnverteilung		Sprechzeitanteile	
	LP	SuS	LP	SuS
t_0	72	73	43 %	44 %
C1 (KSR)	109	103	43 %	46 %
C4	34	33	29 %	59 %
C6	60	54	25 %	65 %
t_1 KG1	25	21	16 %	78 %
t_1 KG2	33	33	24 %	68 %

Legende: t: Datenerhebung; C: Coachingzyklus; KG: Klassengespräch; KSR: Kein Sinnbildungs- oder Reflexionsgespräch; LP: Lehrperson; SuS: Schülerinnen und Schüler

Die Verteilung der Turns bleibt während der gesamten Fortbildung anteilmäßig beim Lehrer und den Lernenden in etwa gleich. Dies gilt jedoch nicht für die Sprechzeit, die sich ab C4 deutlich zugunsten der Schüler*innen erhöht. Gleichzeitig ist die Sprechzeit von Roland ab C4 bei konstanter Turnverteilung stark rückläufig.

5.1.2 Entwicklung Gesprächsleitungsverhalten des Lehrers

Zur Beantwortung der Fragestellung 1b) wurde das Gesprächsverhalten Rolands entlang der Oberkategorien des Codiersystems quantifiziert. Abbildung 2 zeigt die Entwicklungskurven des Gesprächsverhaltens von Roland beginnend bei der Eingangsdatenerhebung t_0 . Die Datenerhebungen wurden mittels vertikaler Linien von den Praxisphasen getrennt. Von den weiter oben ausgeführten Kategorien wurden Codes in den jeweiligen Oberkategorien zusammengenommen, um die übergeordneten Entwicklungen nachzuzeichnen. Innerhalb der Oberkatego-

rie «divergente Handlungen» finden sich von der Fortbildung geförderten Gesprächsleitungsstrategien.

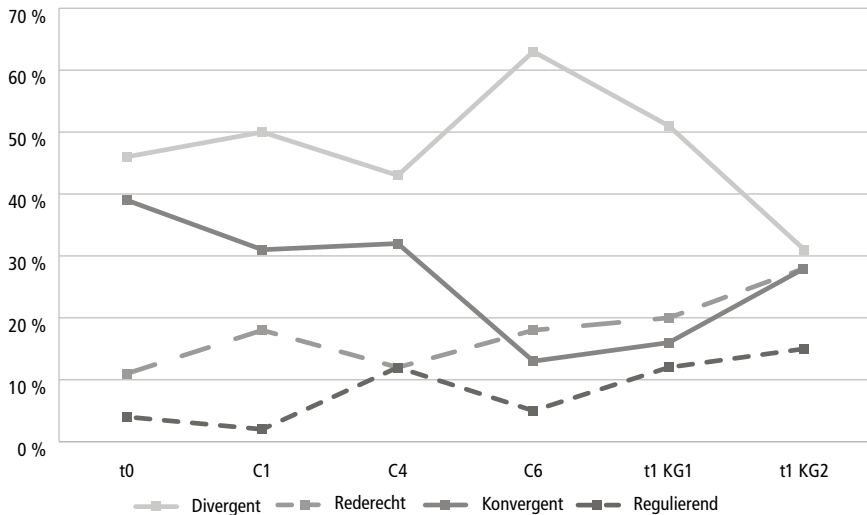


Abbildung 2: Entwicklung Gesprächsverhalten Roland

Rolands Gesprächsverhalten weist bereits bei t0 einen hohen Anteil an divergenten Impulsen in seinem Gesprächsverhalten auf. Diese Impulse treten aber zu einem ähnlich hohen Anteil wie konvergente Handlungen auf. Im Vergleich von t0 zu t1 KG1 nehmen die konvergenten Handlungen Rolands deutlich ab. Zudem zeigt sich ein gleichmäßiger Anstieg an Rederechtserteilungen im Verlauf der Fortbildung. Auffällig ist der starke Rückgang an konvergenten Handlungen von C4 zu C6, wobei anzumerken ist, dass die konvergenten Handlungen Rolands zu keinem Zeitpunkt den Hauptanteil seines Gesprächsverhaltens ausmachen. Auffallend ist die deutliche Zunahme divergenter Impulse zum Zeitpunkt C6. Diese fällt zusammen mit einem merklichen Rückgang konvergenter Handlungen.

5.1.3 Veränderungen der argumentativ-diskursiven Lernendenaussagen

Zur Beantwortung der zweiten Fragestellung wurden die Häufigkeiten der begründeten und der dokumentbezogenen Äußerungen ebenso zusammengefasst wie die Behauptungen und die Restanteile ohne Inhalt.

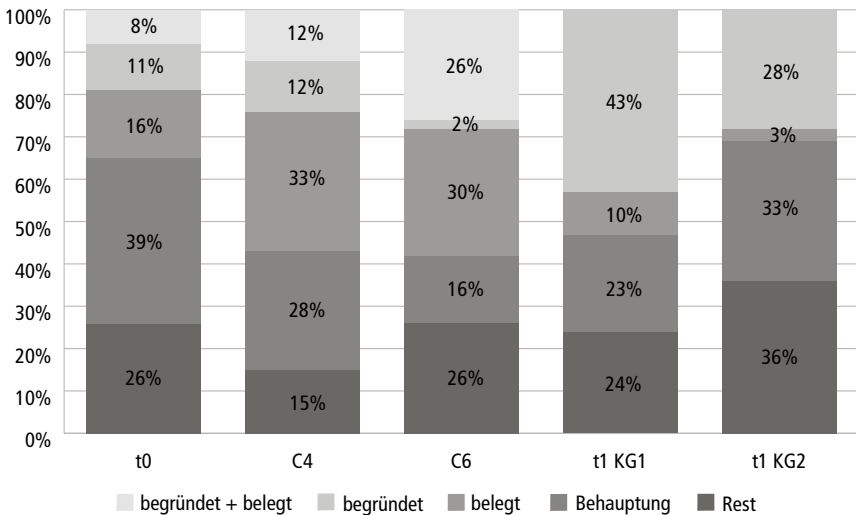


Abbildung 3: Entwicklung Anteile begründeter und belegter Lernendenaussagen von Rolands Klasse

Betrachtet man die Entwicklung der Lernenden in Abbildung 3, fallen bei t0 die hohen Anteile an Behauptungen und Aussagen ohne inhaltlichen Bezug auf (65%). Dies verändert sich bis C4 und bleibt konstant unter 50 Prozent bis t1 KG1. Besonders hohe Anteile an begründeten und belegten Aussagen finden sich bei C6. Betrachtet man nur den Anteil an begründeten Aussagen, dann lässt sich eine stetige Zunahme von t0 (19%) bis t1 KG1 (43%) feststellen. Sehr gering fallen die Anteile an belegten Aussagen mit Dokumentbezug zum Zeitpunkt t1 mit 10 Prozent respektive 3 Prozent aus. Gleichzeitig weist insbesondere t1 KG2 einen sehr hohen Anteil an Behauptungen und Äußerungen ohne Inhaltsbezug auf.

6 Diskussion

6.1 Entwicklungen

Die Ergebnisse zu Fragestellung 1a) zeigen eine Zunahme der Sprechanteile der Lernenden Rolands bei in etwa gleichbleibender Verteilung der Turns. Dies weist darauf hin, dass der Lehrer weniger Raum im Gespräch einnimmt. Außerdem kann die Verteilung der Sprechzeitanteile als ein Indiz für eine erhöhte inhaltli-

che Mitgestaltung der Gespräche durch Lernendenäußerungen betrachtet werden. Dies stützen die Ergebnisse bei Fragestellung 1b), die eine feststellbare Zunahme an divergenten Impulsen und Rederechtserteilungen seitens des Lehrers aufweisen. Dies gilt als Indikator für eine an Aussagen der Lernenden orientierte Gesprächsleitung durch Roland und somit als Hinweis auf eine dialogischere Gesprächsleitung.

Erste Veränderungen in Rolands Gesprächsleitung zeigen sich bereits in den ersten beiden Praxisphasen. Eine deutliche Veränderung im Gesprächsmuster lässt sich vor allem nach der *zweiten Präsenzveranstaltung* feststellen. Die deutliche Abnahme konvergenter Handlungen und die starke Zunahme divergenter Impulse (Zeitpunkt C6) weisen auf mögliche Effekte dieser zweiten Inputveranstaltung hin. Es gelingt Roland zwar schon ab der zweiten Praxisphase, die Sprechzeit der Lernenden zu erhöhen, seine Gesprächsleitung verändert sich aber erst in geringem Maße. Dies legt die Hypothese nahe, dass er durch seinen bereits zu Beginn hohen Anteil an divergenten Impulsen zunächst vor allem seine konvergenten Handlungen reduziert.

Die Ergebnisse zu Fragestellung 2 zeigen, dass die Klasse eine ähnliche Entwicklung aufweist wie Roland. Die Behauptungen der Lernenden nehmen insbesondere nach der zweiten Präsenzveranstaltung der Fortbildung stark ab. Außerdem nimmt der Anteil der begründeten Aussagen von t_0 bis t_1 stetig zu. Dies entspricht einer erwarteten Entwicklung, schließlich fordern das Prinzip der Lerngemeinschaft und die plausible historische Sinnbildung die Begründung einer Aussage. Angesichts des Leistungsniveaus der Klasse erscheint dies als ein positiver Effekt einer veränderten Gesprächsleitung. Der Anteil an begründeten und/oder belegten Aussagen beläuft sich ab C4 bis und mit t_1 KG1 auf über 50 Prozent aller Aussagen. Die Restanteile bei den Lernenden zum Zeitpunkt t_1 KG2 fallen aber sehr hoch aus, was auf unruhiges Klassenverhalten hindeutet und eine mögliche Erklärung für den hohen Anteil konvergenter und regulierender Lehrpersonenverhaltens (42 %) ist. Bereits in der zweiten Praxisphase übersteigt der nachweisliche Dokumentbezug die Aussagen ohne den selbigen. Die höchsten Werte zeigen sich zum Zeitpunkt C6.

Die Entwicklungen des argumentativen Gehalts der Lernendenaussagen von C6 zu t_1 lassen sich womöglich mit der Formulierung der Gesprächsleitfrage in Verbindung bringen. Vergleicht man die Leitfrage des Gesprächs t_1 KG1 mit t_1 KG2 fällt auf, dass es sich bei KG1 um eine explanative Fragestellung («Welche möglichen historischen Gründe gibt es für die Ungleichstellung unterschiedlicher Bevölkerungsgruppen?») handelt. Im Gegensatz dazu wird in KG2 die Frage

nach der reflexiven Einschätzung heutiger Entwicklungen gestellt («Haben diese Ursachen und Gründe noch heute einen Einfluss? Sind bei uns alle gleich oder sind manche gleicher?»), was mit einem Rückgang der divergenten Impulse des Lehrers sowie der Dokumentbezüge der Lernenden einhergeht. Wie die obigen Erkenntnisse andeuten, scheint die für die Lernenden erfassbare inhaltliche Kohärenz aus Leitfrage, gesprächsvorbereitenden Aufgabenstellungen und Dokumentinhalten eine wichtige Rolle zu spielen, ob überhaupt dialogische Klassengespräche geführt werden können. Außerdem könnten reflexive Fragen die Lernenden dazu verleiten, bei ihren Äußerungen im Klassengespräch eigenen Erfahrungsberichten den Vorzug gegenüber dokumentarischer Evidenz zu geben. Dazu sind weitere Analysen zu Gelingensbedingungen dialogischer Klassengespräche im Geschichtsunterricht notwendig (vgl. Zimmermann, im Druck).

Außerdem lässt die Einzelfallanalyse natürlich keine generalisierbaren Erkenntnisse für erfolgreiche Fortbildungen für Geschichtslehrpersonen zu. Zusätzlich müssten die beiden anderen Fälle der erweiterten Bedingung hinzugezogen werden, um idealtypische Erkenntnisse aus der Kontrastierung der Fälle zu gewinnen (vgl. Zimmermann, im Druck). Aus forschungsmethodischer Sicht gilt es anzumerken, dass sich die turnweise Codierung für die Quantifizierung des beobachtbaren Verhaltens eignet, jedoch keine finalen Aussagen zur Dialogizität des gesamten Gesprächsverlaufs zulässt. Dazu müsste das Gesprächsverhalten der Lernenden bezüglich dialogischer Elemente, zum Beispiel direkte Bezugnahme der Lernenden untereinander, analysiert werden.

6.2 Folgerungen

Für eine dialogischere Gesprächsleitung im Sinne von Abschnitt 2.3 benötigte Roland einen Input zur Aufgabengestaltung als Basis für ein Klassengespräch. Erst die gehaltvollere Aufgabenstellung ermöglichte es Roland, vermehrt divergente Impulse einzusetzen. Trotz leichter Zunahme konvergenter Handlungen bei t_1 setzt Roland weiterhin mehrheitlich divergente Impulse und Rederechtserteilungen zur Gesprächsleitung ein. Anhand der veränderten Sprechzeitanteile und den Anteilen an divergenten Impulsen im Vergleich zu konvergenten Handlungen kann insgesamt von einer dialogischeren Gesprächsleitung ausgegangen werden.

Der Wendepunkt zu mehr Begründungen in den Aussagen der Lernenden zwischen der zweiten und dritten Praxisphase weist erneut auf die Bedeutung des Aufgabeninputs hin. Erst die produktive Verbindung von Leitfragen mit der

Wahl von Dokumenten, begünstigt durch die zweite Präsenzveranstaltung, führt sowohl zum Rückgang unbegründeter Aussagen und zu einer dialogischeren Gesprächsleitung von Roland. Die dialogischere Gesprächsleitung Rolands bei Sinnbildungsgesprächen geht einher mit der Zunahme begründeter und dokumentbasierter Aussagen der Lernenden. Die langfristige angelegte Verschränkung aus Input-, Erprobungs- und Feedbackphasen scheint die erwarteten Effekte zu erzielen. Außerdem zeigt sich, wie bedeutend die fachdidaktische Qualität der gesprächsvorbereitenden Aufgabenstellung für die beobachtbare Dialogizität sowie für den fachlichen Gehalt des Gesprächs ist (vgl. Zimmermann, im Druck). Weitere Analysen dazu können den bisherigen Erkenntnisstand zu produktiven Aufgaben im Geschichtsunterricht als Merkmal guten Geschichtsunterrichts anreichern (v. a. Gautschi, 2009, 254).

Literatur

- Alexander, Robin (2020). *A dialogic teaching companion*. New York: Routledge.
- Allen, Joseph/Pianta, Robert/Gregory, Anne/Mikami, Amori Yee & Lun, Janetta (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, 333(6045), 1034–1037.
- Barricelli, Michele (2015). Wort zur Zeit. Historische Sprache und narrative Sinnbildung im Geschichtsunterricht. *Zeitschrift für Geschichtsdidaktik*, 14, 25–46.
- Becker-Mrotzek, Michael & Vogt, Rüdiger (2009). *Unterrichtskommunikation. Linguistische Analysemethoden und Forschungsergebnisse* (2., bearb. und aktualis. Aufl.). Tübingen: Niemeyer.
- Binnenkade, Alexandra (2015). Doing memory. Teaching as a discursive node. *Journal of Educational Media, Memory, and Society*, 7(2), 29–43.
- Britt, Anne & Rouet, Jean-François (2012). Learning with multiple documents: Component skills and their acquisition. In John Kirby & Michael Lawson (Hrsg.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes* (S. 276–314). New York: Cambridge University Press.
- Boxtel, Carla van & Drie, Jannet van (2017). Engaging students in historical reasoning: The need for dialogic history education. In Mario Carretero/Stefan Berger & Maria Grever (Hrsg.), *Palgrave handbook of research in historical culture and education* (S. 573–589). London: Springer.
- Cazden, Courtney (2001). *Classroom discourse: The language of teaching and learning* (2. Aufl.). Portsmouth: Heinemann.
- Collins, Allan/Brown, John & Holum, Anne (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, 6, 38–46.
- Desimone, Laura & Garet, Michael (2015). Best practices in teachers' professional development in the United States. *Psychology, Society and Education*, 7(3), 252–263.

- Drie, Jannet van & Boxtel, Carla van (2008). Historical reasoning: Towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review*, 20(2), 87–110.
- Drie, Jannet van & Ven, Piet-Hein van de (2017). Moving ideas: An exploration of students' use of dialogue for writing in history. *Language and Education*, 31(6), 526–542.
- Eichner, Sinje/Kaestner, Max-Simon & von Reeken, Dietmar (2019). «Ja, das ist auch so ein Begriff». Zum Potenzial von Scaffolding als Unterstützungsstrategie zur Begriffsbildung im Geschichtsunterricht. In Martin Butler & Juliana Goschler (Hrsg.), *Sprachensibler Fachunterricht: Chancen und Herausforderungen aus interdisziplinärer Perspektive* (S. 239–266). Wiesbaden: Springer.
- Fina, Kurt (1978). *Das Gespräch im historisch-politischen Unterricht. Ein Kurs für Studenten und Lehrer*. München: Wilhelm Fink.
- Gautschi, Peter (2009). *Guter Geschichtsunterricht. Grundlagen, Erkenntnisse, Hinweise*. Schwalbach/Ts.: Wochenschau.
- Günther-Arndt, Hilke (2019). Historisches Lernen und Wissenserwerb. In Hilke Günther-Arndt & Meik Zülsdorf-Kersting (Hrsg.), *Geschichtsdidaktik. Praxishandbuch Sekundarstufe I und II* (S. 24–49). Berlin: Cornelsen.
- Handro, Saskia (2013). Sprache und historisches Lernen. Dimensionen eines Schlüsselproblems des Geschichtsunterrichts. In Michael Becker-Mrotzek/Karen Schramm/Eike Thürmann & Helmut Vollmer (Hrsg.), *Sprache im Fach. Sprachlichkeit und fachliches Lernen* (S. 317–333). Münster: Waxmann.
- Havekes, Harry/van Boxtel, Carla/Coppen, Peter-Arno & Luttenberg, Johan (2017). Stimulating historical thinking in a classroom discussion: The role of the teacher. *Historical Encounters Journal*, 4(2), 71–93.
- Heller, Vivien & Morek, Miriam (2019). Fachliches und sprachliches Lernen durch diskurs(erwerbs)orientierte Unterrichtsgespräche. Empirische Evidenzen und Desiderata mit Blick auf inklusive Settings. *Didaktik Deutsch*, 24(46), 102–121.
- Hennessy, Sara/Howe, Christine/Mercer, Neil & Vrikki, Maria (2020). Coding classroom dialogue: Methodological considerations for researchers. *Learning, Culture and Social Interaction*, 25, 100404.
- Heuer, Christian/Resch, Mario & Seidenfuß, Manfred (2017). Geschichtslehrerkompetenz? Wissen und Können geschichtsdidaktisch. *Zeitschrift der Didaktik der Gesellschaftswissenschaften*, 8(2), 158–176.
- Hodel, Jan & Waldis, Monika (2007). Sichtstrukturen im Geschichtsunterricht – die Ergebnisse der Videoanalyse. In Peter Gautschi/Daniel Moser/Kurt Reusser & Pit Wiher (Hrsg.), *Geschichtsunterricht heute. Eine empirische Analyse ausgewählter Aspekte* (S. 91–142). Bern: hep.
- Hug, Wolfgang (1980). *Geschichtsunterricht in der Praxis der Sekundarstufe I: Befragungen, Analysen und Perspektiven* (2. Aufl.). Frankfurt a. M.: Moritz Diesterweg.
- Kim, Min-Young & Wilkinson, Ian (2019). What is dialogic teaching? Constructing, deconstructing, and reconstructing a pedagogy of classroom talk. *Learning, Culture and Social Interaction*, 21, 70–86.

- Kooloos, Chris/Oolbekkink-Marchand, Helma/Kaenders, Rainer & Heckman, Gert (2020). Orchestrating mathematical classroom discourse about various solution methods: Case study of a teacher's development. *Journal für Mathematik-Didaktik*, 41(2), 357–389.
- Kuckartz, Udo (2014). *Mixed Methods. Methodologie, Forschungsdesigns und Analyseverfahren*. Wiesbaden: Springer VS.
- Lave, Jean & Wenger, Etienne (1991). *Situated learning legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Lipowsky, Frank & Rzejak, Daniela (2019). Empirische Befunde zur Wirksamkeit von Fortbildungen für Lehrkräfte. In Paul Platzbecker & Botho Priebe (Hrsg.), *Zur Wirksamkeit und Nachhaltigkeit von Lehrerfortbildungen. Dokumentation der Fachtagung vom 26.–27. September 2018 in Wermelskirchen* (S. 34–74). Essen: Institut für Lehrerfortbildung.
- Luis, Rita & Rapanta, Chrysi (2020). Towards (re-)defining historical reasoning competence: A review of theoretical and empirical research. *Educational Research Review*, 31, 100336.
- Mayring, Philipp (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim: Beltz.
- Mehan, Hugh (1979). *Learning lessons: Social organization in the classroom*. Cambridge: Harvard University Press.
- Michaels, Sarah/O'Connor, Catherine/Hall, Megan & Resnick, Lauren (2010). *Accountable Talksourcebook: For classroom conversation that works*. Pittsburgh: University of Pittsburgh.
- Monte-Sano, Chauncey & Allen, Amina (2019). Historical argument writing: the role of interpretive work, argument type, and classroom instruction. *Reading and Writing*, 32(6), 1383–1410.
- Pauli, Christine (2010). Klassengespräche – Engführung des Denkens oder gemeinsame Wissenskonstruktion selbstbestimmt lernender Schülerinnen und Schüler? In Timo Bohl/Katja Kansteiner-Schänzlin/Marc Kleinknecht/Britta Kohler & Anja Nold (Hrsg.), *Selbstbestimmung und Classroom Management. Empirische Befunde und Entwicklungsstrategien zum guten Unterricht* (S. 145–161). Bad Heilbrunn: Klinkhardt.
- Pauli, Christine & Reusser, Kurt (2018). Unterrichtsgespräche führen – das Transversale und das Fachliche einer didaktischen Kernkompetenz. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 36(3), 365–377.
- Pol, Janneke van de/Brindley, Sue & Higham, Rupert (2017). Two secondary teachers' understanding and classroom practice of dialogic teaching: a case study. *Educational Studies*, 43(5), 497–515.
- Pontecorvo, Clotilde & Girardet, Hilda (1993). Arguing and reasoning in understanding historical topics. *Cognition and Instruction*, 11(3/4), 365–395.
- Reisman, Abby (2015). Entering the historical problem space: Whole-class, text-based discussion in history class. *Teachers College Record*, 117(2), 1–44.
- Reisman, Abby & Enumah, Lisette (2020). Using video to highlight curriculum-embedded opportunities for student discourse. *Journal of Teacher Education*, 71(5), 551–567.
- Reisman, Abby/Schneider Kavanagh, Sarah/Monte-Sano, Chauncey/Fogo, Brad/McGrew, Sarah/Cipparone, Peter & Simmons, Elizabeth (2018). Facilitating whole-class discus-

- sions in history: A framework for preparing teacher candidates. *Journal of Teacher Education*, 69(3), 278–293.
- Resnick, Lauren/Asterhan, Christa/Clarke, Sherice & Schantz, Faith (2018). Next generation research in dialogic learning. In Gene Hall/Linda Quinn & Donna Gollnick (Hrsg.), *The Wiley handbook of teaching and learning* (S. 323–338). Hoboken: Wiley-Blackwell.
- Resnick, Lauren/Michaels, Sarah & O'Connor, Catherine (2010). How (well structured) talk builds the mind. In David Preiss & Robert Sternberg (Hrsg.), *Innovations in educational psychology. Perspectives on learning, teaching, and human development* (S. 163–194). New York: Springer.
- Reusser, Kurt (2014). Aufgaben – Träger von Lerngelegenheiten und Lernprozesse im kompetenzorientierten Unterricht. *Seminar*, 4, 77–101.
- Reusser, Kurt & Pauli, Christine (2015). Co-constructivism in educational theory and practice. In James Wright (Hrsg.), *International encyclopedia of the social & behavioral sciences* (Bd. 3, 2. Aufl., S. 913–917). Oxford: Elsevier.
- Reznitskaya, Alina & Wilkinson, Ian (2017). *The most reasonable answer. Helping students build better arguments together*. Cambridge: Harvard Education Press.
- Spieß, Christian (2015). Das Unterrichtsgespräch als zeitgemäße Form der Geschichtserzählung. Asymmetrische Kommunikation im Geschichtsunterricht. *Zeitschrift für Geschichtsdidaktik*, 14, 154–168.
- Trittel, Monika (2010). Einzelfallanalysen und Studien mit kleinen Fallzahlen. In Tina Häscher & Bernhard Schmitz (Hrsg.), *Pädagogische Interventionsforschung. Theoretische Grundlagen und empirisches Handlungswissen* (S. 280–286). Weinheim: Juventa.
- Tytler, Russell & Aranda, Georg (2015). Expert teachers' discursive moves in science classroom interactive talk. *International Journal of Science and Mathematics Education*, 13(2), 425–446.
- Weil, Maralena/Seidel, Tina/Schindler, Ann-Kathrin & Gröschner, Alexander (2020). Opening «windows» for teachers to change classroom discourse. *Learning, Culture and Social Interaction*, (26), 100425.
- Wood, David/Bruner, Jerome & Ross, Gail (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100.
- Zülsdorf-Kersting, Meik (2020). Qualitätsmerkmale von Geschichtsunterricht. Zum Verhältnis generischer und fachspezifischer Merkmale. *Unterrichtswissenschaft*, 48, 385–407.
- Zimmermann, Matthias (im Druck). *Dialogische Klassengesprächführung im Geschichtsunterricht. Veränderungsprozesse einer fachlichen und transversalen Kompetenz von Lehrpersonen im Rahmen der Interventionsstudie Socrates 2.0*. Schwabach/Ts.: Wochenschau.

Grundlagen für geschichts- didaktische Interventionen

Concept-Mapping und konstruktivistische Grounded-Theory-Methodologie

Mit semantischen Netzen historisches Wissen erfassen

Burghard Barte

1 Zur Geschichte als Wissensdomäne

Im Hinblick auf das domänenspezifische Wissen der Geschichte(n) stehen Lernende vor der Herausforderung, dass «historisches Denken Welt und Mensch in umfassender Weise betrifft» (Schreiber, 2012, 122). Dies bedingt eine strukturelle Offenheit und zugleich inhaltliche Unabgeschlossenheit des Faches, die auch andere geisteswissenschaftliche Domänen auszeichnet (Jahr, 2009, 82–88). Zum Ausdruck kommen diese Auszeichnungsmerkmale beispielsweise darin, dass sich historische Begriffe, analytische Begriffe der historischen Betrachtung und solche aus der eigenen Lebenspraxis vielfach überlappen oder als Bezeichnung identisch sein können, wenngleich das mit ihnen jeweils Bezeichnete differiert. Ein konkretes Profil gewinnen Begriffe im Fach Geschichte durch ihre kontext- und fragestellungsbezogene «Justierung» innerhalb eines Begriffsfeldes (Alavi, 2004, 57; 2006, 28). So werden der spezifische Gehalt eines Begriffs und das mit ihm Bezeichnete für den historischen Denkprozess operationalisiert. Damit kann historisches Wissen erkenntnistheoretisch als eine «begrifflich formatierte Erfahrung [...] eingebunden in ein Netz begrifflicher Bestimmungen» (Schmidt, 2010, 140) betrachtet werden. Daraus folgt zugleich, dass Begriffe eine sprachliche Fassung komplexer mentaler Vorstellungen darstellen. So weisen sie eine prinzipielle Unschärfe auf und differieren interindividuell hinsichtlich dessen, was sie jeweils als Wissen repräsentieren (Sander, 2011, 37–38).

Im geschichtsdidaktischen Diskurs liegen drei prominente Modelle vor, die in unterschiedlicher Akzentuierung den beschriebenen Umgang mit Begriffen sowie die Entwicklung von Konzepten als Wissen theoretisch modellieren. Mit der Begriffs- und Strukturierungskompetenz als Teilkompetenzen des Sachkompetenzbereichs wird historisches Wissen im FUEP-Modell vordergründig in einem kategorienbezogenen Zugang performativ gefasst (Schöner, 2007). Das das

konzeptionelle Wissen ins Zentrum stellende Wissensmodell Kühbergers fokussiert mit dem Vorschlag von ausgewählten Basiskonzepten auf die Relationalität des Wissens und den Transfer von Konzepten (Kühberger, 2012, 39, 43). Eine eher funktionale Perspektive wird im *Framework for Historical Reasoning* gewählt, der von der *conceptual change*-Forschung ausgehend historisches Wissen vordergründig vom historischen Denken in einer narrativen Logik her greift (van Drie & van Boxel, 2008, 99–101). Aufbauend auf diesen Modellen sind auch auf pragmatischer Ebene für die unterrichtliche Begriffsarbeit mehrere didaktische Konzepte ausgearbeitet worden (exemplarisch: Alavi, 2004; Sauer, 2015; Ventzke, 2012).

Empirisch ist die Subjektperspektive, wie Lernende historisches Wissen strukturieren und wie sie welche Begriffe nutzen, um historische Kenntnisse kognitiv zu organisieren und neues Wissen zu integrieren, jedoch ein bisher wenig bearbeitetes Feld (Langer-Plän & Beilner, 2006, 248; Barricelli & Sauer, 2015, 189; Kühberger, 2016, 103; Schreiber, 2012, 129). Diesen Befund stützend schlägt Ziegler vor, in einem offenen Verständnis und Zugang zu historischem Wissen mit dem Konstrukt des *Konzepts* zu arbeiten, welches operationalisiert das fokussiert, «was Individuen über einzelne Themen wissen, wie sie diese begrifflich fassen und zwischen Komponenten Verknüpfungen herstellen, welche Sinngemungsmuster sie verwenden, welche Bedeutung sie ihnen für ihre Orientierung verleihen» (Ziegler, 2012, 143). Vorliegende empirische Studien, die hinsichtlich ihrer Ausrichtung diesem Zugang zum Wissen zugeordnet werden können, orientieren sich entweder an der *conceptual change*-Forschung und fokussieren in einem holistischen Ansatz fachliche Vorstellungen von Lernenden (exemplarisch dafür: Fenn, 2018; Mathis, 2015; Stöckle, 2011) oder widmen sich in engerem Zuschnitt gezielt dem Begriffsverständnis (Langer-Plän & Beilner, 2006; Zabold, 2020). Dabei wird in der Reflexion der Studien zweierlei deutlich: Zum einen ist die Organisation historischen Wissens vielfach ein mitadressierter Teilaspekt. Selbst stand sie jedoch bisher nicht explizit im Fokus. Zum anderen zeigt sich, dass die vorliegenden Studien der geschichtsdidaktischen Lehr-Lern-Forschung mit Fokus auf historischem Wissen sich bisher zumeist auf den Kontext Schule konzentrieren.

Ausgehend von diesen Befunden wird im Folgenden eine am qualitativen Forschungsparadigma orientierte Umsetzung des Concept-Mapping für die Erfassung historischen Wissens sowie der Begriffsbildung vorgestellt. Konkret fokussiert der Beitrag darauf, wie bei der Erhebung und Interpretation von Concept-Maps die konstruktivistische Grounded-Theory-Methodologie (im Weiteren: KGT-Methodologie) als Forschungsstil genutzt werden kann. Dafür werden

zunächst erhebungsbezogene Implikationen beleuchtet, bevor ein analytischer Perspektivrahmen für die datenzentrierte Auswertung vorgestellt wird, mit dem Concept-Maps als Texte betrachtet werden können. Das vorgeschlagene Verfahren für eine datenzentrierte Interpretation orientiert sich an den drei zentralen Codierphasen der KGT-Methodologie. Veranschaulicht wird es anhand ausgewählter Daten aus dem Promotionsprojekt des Verfassers und dabei aufgezeigt, wie eine Kategorie einer *grounded theory* entwickelt werden kann. Demnach liegt der Schwerpunkt des Beitrags darauf, das Potenzial einer nicht referenzbasierten Auswertung von Concept-Maps zu diskutieren und weniger darauf, zentrale Ergebnisse des Forschungsprojekts zu präsentieren.

2 Concept-Maps als Darstellung des subjektiven historischen Wissens

In einer methodologisch-methodischen Reflexion der vorliegenden Studien zu fachlichen Vorstellungen, dem historischen Begriffsverständnis sowie dem *conceptual change* zeigt sich, dass zumeist Gruppendiskussionen, leitfadengestützte Einzelinterviews, schriftliche Definitionsaufgaben oder Kombinationen dieser Methoden Anwendung finden. Zuletzt hat Zabold mit dem gruppenförmigen Mind-Mapping das Spektrum der Erhebungsmethoden ergänzt (2020, 98–99). Darüber hinaus hat sie mit der Struktur-lege-Technik im Rahmen von Interviews eine dem Concept-Mapping nahestehende Erhebungsmethode umgesetzt. Concept-Maps im engeren Sinne fanden in der geschichtsdidaktischen Lehr-Lern-Forschung im Vergleich dazu bisher wenig Aufmerksamkeit. Neben Lange (2011), die diese triangulierend zur Erforschung des Bildverstehens von Schüler*innen eingesetzt hat, haben zuletzt Fenn, Stebener & Urban (2019) einen qualitativ evaluierenden Ansatz für die Erfassung professionsbezogenen Fachwissens von Studierenden via Concept-Maps vorgestellt.

Das grundsätzliche Potenzial von Concept-Maps für die Erfassung von Wissen ergibt sich aus zwei Auszeichnungsmerkmalen: Zum einen sind sie hinsichtlich der strukturellen Gestaltungsmöglichkeiten nicht festgelegt, also nichtlinear und prinzipiell offen (Graf, 2014, 327). Zum anderen adressieren sie explizit die Strukturierung von Begriffen und bieten darin einen gezielten Zugang zur kognitiven Organisation des Wissens (Gebhardt, Mühling, Gartmeier & Tretter, 2015, 611). Dabei sind gestaltete Concept-Maps jedoch nicht als ein Abbild der kognitiven Struktur im Wortsinn zu verstehen, sondern als situative Darstellung des

aktiven Wissens (Barte, Landgraf & Mühling, 2020). Gleichwohl wird eine gewisse intraindividuelle Stabilität bezüglich der fachlichen Kenntnisse und ihrer Relationierung angenommen (so auch Kühberger, 2016, 96). So können Concept-Maps als sprachlich-begriffliche Darlegung des subjektiven historischen Wissens im Sinne eines Netzes aus «narrative[n] Abkürzungen» (Rüsen, Fröhlich, Horstkötter & Schmidt, 1991, 231) betrachtet werden.

Die dem Beitrag zugrunde liegenden Concept-Map-Daten entstammen einer multimethodischen Studie, die sich in ihrem Erkenntnisinteresse auf die Organisation und Strukturierung historischer Kenntnisse, den Umgang mit historischem Fachwissen im Bildungskontext Hochschule konzentriert. Dafür bilden neben Concept-Maps auch Beobachtungsprotokolle und leitfadengestützte Interviews die Datengrundlage. Deren Erhebung fand im Rahmen dreier verschiedener mediävistischer Master-Seminare an zwei verschiedenen Universitäten im norddeutschen Raum statt. In zwei Seminaren wurde jeweils in der Einführungs- sowie Abschlussitzung eine Concept-Map zum thematischen Fokus des entsprechenden Seminars gestaltet. In einem dritten Seminar überarbeiteten die Studierenden ihre am Beginn gestalteten Concept-Maps im Verlauf des Lehr-Lern-Kontextes. So sollte gezielt auch der Umgang mit und die Verarbeitung von erworbenen historischen Kenntnissen erhellt werden. Dieses dritte Seminar, das an einer zweiten Universität stattfand, dient auch im Hinblick auf seine Anlage als ein differenter Erhebungskontext. So liegen jeweils zwei Concept-Maps von zwölf Studierenden sowie jeweils eine Ausgangsmap und deren überarbeitete Version von sieben Studierenden vor. Von den insgesamt neunzehn Studienteilnehmenden studierten zum Erhebungszeitpunkt fünfzehn im *Master of Education*, drei im *Master of Arts* und eine Person im *Bachelor of Arts* das Fach Geschichte. Das Studiensemester, in dem sie sich befanden, variierte zwischen dem sechsten und dreizehnten Semester. Ihr Alter betrug im Durchschnitt 24,8 Jahre. Sieben der Studierenden ordneten sich dem männlichen und zwölf dem weiblichen Geschlecht zu.¹

1 Zur Erläuterung und Begründung des forschungspraktischen Vorgehens, zur Anlage der multimethodischen Studie und der Erhebung der verschiedenen Daten (Concept-Maps, Interviews sowie Beobachtungsprotokolle) vgl. Barte, Burghard (2022).

3 Concept-Mapping und konstruktivistische Grounded-Theory-Methodologie

In der KGT-Methodologie bildet die Annahme, dass «neither data nor theories are discovered, but researchers construct them as a result of their interactions with their participants and emerging analyses» (Thornberg & Charmaz, 2014, 154) das Forschungsparadigma. Demnach steht bei dieser methodologischen Variante im Vordergrund, durch die systematische Einbindung bestehender theoretischer Erkenntnisse sowie die produktive Nutzung der subjektiven Erfahrungen des*der Forschenden im Forschungsprozess eine abduktive Haltung zu entwickeln, die die «Entdeckung neuer Zusammenhänge» (Hohage, 2016, 115) ermöglicht. Zentral ist dabei, dass der disziplinäre Forschungsstand den Start- und nicht den Zielpunkt der Forschungsarbeit im Allgemeinen sowie der Analysepraxis im Konkreten bildet. So sollen neue Deutungsangebote entwickelt oder bestehende Theorien ergänzt respektive modifiziert werden (Bücker, 2020, Abs. 22).

Entsprechend steht die verstehende und theoriebildende Interpretation von Handlungen und Prozessen im Vordergrund (Charmaz, 2014, 15), worin sich die KGT-Methodologie von der Inhaltsanalyse mit ihrem Fokus auf Themen und Strukturen unterscheidet. Das dafür grundlegende analytische Vorgehen folgt dem Prinzip der Iteration, für dessen Umsetzung – ausgehend von ersten Interpretationen und der spezifischen Fokussierung auf herausgearbeitete Phänomene – im Sinne des theoretischen Samplings eine Rückkehr ins Feld oder zu bereits interpretierten Daten eingefordert wird (Strübing, 2018, 32).

Dieser Gleichzeitigkeit von Sampling und Auswertung liegt die Idee des fortwährenden Vergleichs von Fällen zugrunde, in dessen Zuge die datennahen Codes sukzessive zu komplexen Kategorien mit vielfältigen Eigenschaften abstrahiert werden (Bücker, 2020, Abs. 12). Ein Fall ist dabei nicht zwingend auf eine Person beschränkt. Fälle können ebenso Phänomene im Sinne von Ereignissen, Strategien oder Mustern sein, die miteinander verglichen und hinsichtlich gemeinsamer Merkmale gruppiert werden (Strübing, 2018, 41; Kelle & Kluge, 2010, 86). Der Abstraktions- und Theoretisierungsprozess ist durch ein dreischrittiges Codierverfahren systematisiert, bei dem die Phasen des *initial*, *focused* und *theoretical coding* aber nicht streng konsekutiv, sondern ineinanderfließend konzeptioniert sind (Charmaz, 2014).

3.1 Implikationen für die Erhebungspraxis

Beim Einsatz des Concept-Mapping im Rahmen von Studien, die sich an dem Forschungsstil der *Grounded Theory* orientieren, sind zwei Vorentscheidungen von Bedeutung, da sie die erhaltenen Concept-Maps nicht nur in ihrem Inhalt, sondern auch in der Struktur präformieren und so die Auswertungsmöglichkeiten und Interpretationsmöglichkeiten bestimmen: Zum einen die gestellte «focus question» (Cañas & Novak, 2012, 250) oder Themenstellung und zum anderen die Vorgabe von Begriffen und Relationen für das Mapping. Beide zusammen, wenn kombiniert, bestimmen den Freiheitsgrad bei der Gestaltung einer Concept-Map und damit implizit auch, welches Wissen erhoben wird (Cañas & Novak, 2006, 499). Bezüglich der Vorauswahl von Begriffen in der Konzeption der Erhebung werden im methodologischen Diskurs zum Concept-Mapping Mappingszenarien differenziert, die unterschiedlich restriktiv sind: Neben der Vorgabe eines Ausgangsbegriffes besteht die Möglichkeit, eine abgeschlossene oder nicht abgeschlossene Liste von Begriffen festzulegen. Dem folgt im Restriktionsgrad fortschreitend die Präsentation einer sogenannten «expert skeleton map» (Cañas & Novak, 2012, 252), bei der eine Grundstruktur zu erweitern ist. Die restriktivste Form des Mappings ist die Darbietung einer bereits ausgearbeiteten Map mit leeren Feldern, bei der sich die Aktivität der Gestaltenden auf das Einfügen von Begriffen beschränkt (Cañas & Novak, 2012).

Die Vorgabe eines Ausgangsbegriffes oder einer Liste von ausgewählten Begriffen als Basis für die Erstellung einer Map können wiederum in ihrer Restriktion variiert werden, wenn es den Gestaltenden obliegt, ob sie die Begriffe einbeziehen oder aber eigene verwenden (Cañas & Novak, 2012, 250–254). In der Kombination einer Themenstellung oder Fokusfrage mit einer Liste von möglichen oder zwingend zu verwendenden Begriffen eröffnen sich weitere Variationsmöglichkeiten hinsichtlich der gestalterischen Freiheit im Concept-Mapping. Im Projekt des Verfassers waren die im Seminar fokussierten Begriffe in die Themenstellung integriert, wurden aber nicht obligatorisch vorgegeben. Bei der Themenstellung waren die Studierenden zusätzlich dazu aufgefordert, sämtliche inhaltsbezogenen Assoziationen sowie aus ihrer Sicht anschlussfähige Kenntnisse aus dem bisherigen Studium mit aufzunehmen. Damit sollte das Vorwissen sowie der Umgang mit komplexen Begriffen zu Beginn des Seminars sowie die Integration der im Rahmen des Seminars erworbenen Kenntnisse in die bestehende Wissensstruktur beziehungsweise deren Entwicklung erfasst werden. Dieses restriktionsarme Vorgehen adressierte mit Blick auf die theoriebil-

dende Ausrichtung der Studie das themenbezogene, «vorhandene aktive Wissen» (Graf, 2014, 330) und die subjektiven Muster seiner Organisation. Zugleich wurde so vermieden, dass Begriffe integriert werden, die eigentlich nicht Bestandteil der Assoziationen und demnach kein bedeutsamer Bestandteil des Wissens sind (Graf, 2014, 331). Da dieses Vorgehen eine höhere Vertrautheit mit der Methode seitens der Teilnehmenden voraussetzt, um aussagekräftige Maps zu erhalten, wurde bei der Datenerhebung in der Studie des Verfassers dem Mapping eine Übungsphase zu einem vertrauten Thema vorangestellt.

3.2 Textualität als Perspektivrahmen für die datenzentrierte Auswertung

Der jeweiligen Eigenlogik der Daten und den herausgearbeiteten Phänomenen Rechnung zu tragen, bildet eine zentrale Prämisse der KGT-Methodologie. Dies erfordert einen Zugang zu Concept-Maps, mit dem deren spezifische Auszeichnungsmerkmale als ein semantisches Netz analytisch zugänglich werden. Im Folgenden wird der ausgearbeitete Ansatz skizziert, Concept-Maps als Daten in den Blick zu nehmen, indem diese aufbauend auf den linguistischen Textualitätskriterien im Anschluss an de Beaugrande und Dressler (1981/2011) als Text betrachtet werden. Dies impliziert zunächst als Grundannahme, dass eine Concept-Map eine kommunikative Handlung darstellt, die sich durch sieben Kriterien auszeichnet: 1) *Kohäsion*, 2) *Kohärenz*, 3) *Intentionalität*, 4) *Akzeptabilität*, 5) *Informativität*, 6) *Situationalität* sowie 7) *Intertextualität* (de Beaugrande & Dressler, 1981/2011).

Das Kriterium der *Kohäsion* bezieht sich auf die topologische Relationierung, also wie Begriffe und Propositionen strukturell verbunden und angeordnet werden. Dieses steht im weiten Verständnis auch bei grafentheoretischen Auswertungen von Concept-Maps (Mühling, 2016, 183) oder qualitativ-evaluierenden Strukturanalysen von Einzelmaps (Kinchin & Hay, 2000) im Vordergrund. Für die qualitative Auswertung von Concept-Maps auf Grundlage der KGT-Methodologie dient die Kohäsion in Verbindung mit dem zweiten Kriterium, der Kohärenz, dazu, zu analysieren, welche Begriffe eine Ordnungsfunktion haben, sowie zu reflektieren, in welchem Verhältnis die Struktur der Map zur inhaltlichen Verortung von Begriffen steht.

Die *Kohärenz* adressiert die expliziten und impliziten semantischen Zusammenhänge und somit die Inhaltsebene (de Beaugrande & Dressler, 1981/2011, 3–5). Mit ihr steht demnach die inhaltliche Bedeutung der Begriffe und Relationen im Vordergrund. In Concept-Mapping-Studien liegt bei einer qualitativen Auswer-

tung der Fokus in der Regel auf der Analyse der Propositionen als zwei durch eine Relationierung direkt miteinander verbundene Begriffe, die eine Aussage bilden. Mit dem Kohärenzkriterium kann dieser detailbezogene Blick erweitert und können ergänzend größere Teilbereiche oder die Map als Ganzes in den Blick genommen werden. So werden auch implizite Sinnzusammenhänge von Begriffen sowie ihre «Verankerung im semantischen Feld» (Sauer, 2015, 9) gezielt für die Interpretation zugänglich. Dafür können die Bedeutungen mehrerer verschiedener Relationen eines Begriffs sowie seine semantische Verortung in der Map interpretativ zusammengedacht werden. Darüber hinaus wird mit der Implikation, dass «Kohärenz nicht bloß ein Merkmal von Texten, sondern vielmehr das Ergebnis kognitiver Prozesse der Textverwender» (de Beaugrande & Dressler, 1981/2011, 7) ist, zugleich die konstruktive Rolle der Forschenden bei der Interpretation der Daten der Reflexion zugänglich.

Intentionalität fokussiert die «Einstellung des Textproduzenten» (de Beaugrande & Dressler, 1981/2011, 8) und die mit dem Text verfolgte Absicht im Sinne der kommunikativen Zielsetzung. Bezogen auf Concept-Maps, lässt sich mit der Intentionalität die Form des Wissens interpretieren, das dargestellt wird. So kann der Frage nachgegangen werden, ob beispielsweise Schlagworte gesammelt, Detailkenntnisse präsentiert oder Zusammenhänge, Prozesse oder eine spezifische Kombination aus diesen entfaltet werden.

Bei dem Kriterium der *Akzeptabilität* steht die Wahrnehmung und Aufnahme im Zentrum, indem sie auf die «Einstellung des Textrezipienten in der Kommunikation» abzielt. Dies meint im Konkreten die Fähigkeit, dem Text Sinn zu entnehmen (de Beaugrande & Dressler, 1981/2011, 135–136). So können in einer Selbstbeobachtungsschleife der Fokus der Interpretation und die eingenommene Perspektive offengelegt werden; wie Forschende durch «eigene Beiträge zum Textsinn die Kohärenz unterstützen» (de Beaugrande & Dressler, 1981/2011, 9). In der zugrunde liegenden Studie wurde dies Kriterium analytisch dafür genutzt, die erstellte Concept-Map in ihrer jeweiligen Verfasstheit explizit anzuerkennen und von dieser ausgehend Phänomene herauszuarbeiten.

Darauf, wie erwartbar und bekannt die präsentierten Elemente und Informationen des Textes sind, bezieht sich die *Informativität* (de Beaugrande & Dressler, 1981/2011, 10–11, 145). Beim Concept-Mapping lässt sich diese für die Inblicknahme der Begriffsauswahl sowie die inhaltliche Verbundenheit von verschiedenen Begriffen fruchtbar machen. Mit ihr kann analysiert werden, welche Begriffe und Aussagen in der Auswertung beispielsweise überraschen oder irritieren. Zugleich bietet sie in einem holistischen Zugang die Möglichkeit, größere

Sinnzusammenhänge und so die Informationsdichte der Map als Ganzes zu analysieren. Herausarbeiten lässt sich die Informativität beispielsweise in der Analyse von dichter vernetzten Teilbereichen oder Propositions-Pfaden. Letztere wurden im Projekt als semantische Verbindungslinien innerhalb einer Map betrachtet.

Die *Situationalität*, die sich auf «die Faktoren [bezieht], die einen Text für eine Kommunikations-SITUATION RELEVANT machen» (de Beaugrande & Dressler, 1981/2011, 12), birgt vor allem forschungs- und methodenreflexives Potenzial. In der Konzentration auf die Situationalität können Concept-Maps vor dem Hintergrund der erfolgten Einführung in das Mapping, der formulierten Instruktion, der vorgenommenen Restriktionen sowie des verfügbaren Zeitrahmens zur Gestaltung hinsichtlich ihrer Aussagekraft und ihres interpretativen Potenzials beleuchtet werden.

Ähnliches bietet die *Intertextualität*. Mit ihr kann die Erhebungssituation sowie die Rolle der Forschenden reflektiert werden. Darüber hinaus eröffnet das Auszeichnungsmerkmal für den Einsatz von Concept-Maps im Rahmen der Lehr-Lern-Forschung die Möglichkeit, den Aufgriff von Begriffen und die Bedeutung der Darbietung von Inhalten sowie ihrer Didaktisierung im Sinne intertextueller Bezüge zwischen Concept-Map und Lehr-Lern-Kontext zu analysieren.

4 Concept-Maps initial, fokussiert und theoretisch codieren

Die Anwendung des skizzierten Perspektivrahmens zur Auswertung von Concept-Maps wird im Folgenden im Fokus auf die textzentrierten Kriterien *Kohäsion*, *Kohärenz* sowie die *Informativität* konkret veranschaulicht. Mit dem Ziel, die domänenspezifische Wissensorganisation aus geschichtsdidaktischer Perspektive zu erfassen, wurden den Textualitätskriterien als sensibilisierende Konzepte die Verzeitlichung sowie der Umgang mit Zeit (Rüsen, 2008; Schreiber, 2012, 122) zur Seite gestellt. Diese wurden dabei – wie oben beschrieben – in einem explizit offenen Verständnis genutzt, wonach sie den Blick für fachspezifische Phänomene schärfen sollten, ohne jedoch den konkreten Inhalt der Kategorien vorab zu bestimmen (Charmaz, 2014, 117).

4.1 «Aufbrechen» des Netzes durch drei heuristische Betrachtungsebenen

Der erste Zugang zu den Daten im Rahmen des *initial coding* entspricht einer frei assoziierenden Auseinandersetzung mit dem Material (Muckel, 2011, 342). Im Vordergrund steht, das jeweilige Segment, das in seinem Umfang von einem Wort bis zu einer ganzen Aussage reichen kann, mit einem Code interpretativ aufzuladen, also mit Bedeutung zu versehen (Muckel, 2011, 338). Mit dem Fokus auf dem historische Wissen wurde die Strategie des «coding for actions» (Charmaz, 2014, 121) angewendet, um die Wissensorganisation auch analytisch als aktiven Prozess zu markieren. Aus der Operationalisierung der hier fokussierten Textualitätskriterien ergaben sich drei heuristische Betrachtungsebenen, von denen ausgehend generative Fragen formuliert wurden:

1. *Entfaltung*: Wie wird ein Begriff über direkt mit ihm verbundene Begriffe inhaltlich entfaltet?
2. *Relationierung*: In welchen semantischen Zusammenhang wird ein Begriff durch die spezifischen Relationen gestellt? Welche Aussagen ergeben sich daraus?
3. *Verweisungszusammenhänge*: Welche Aussagekraft und inhaltliche Qualität gewinnt ein Begriff, eine Proposition oder eine *Community* durch ihre Positionierung in der Concept-Map?

Wie davon ausgehend eine Kategorie entwickelt werden kann, wird im Folgenden am Beispiel von «Bedingungsbeziehungen flechten» aufgezeigt. Für verschiedene Concept-Map-Segmente – gemeint sind damit Elemente und Teilbereiche von Maps – wurden im Projekt zunächst vielfältige Codes vergeben, die sukzessive in einer Kategorie abstrahiert wurden. Im fortgeschrittenen Interpretationsprozess etablierte sich «Bedingungsbeziehungen flechten» dann als eine Eigenschaft der theorierelevanten Kategorie «den Kenntnisraum texturieren».

Die beiden nachstehend präsentierten Segmente entstammen Concept-Maps, bei denen das Seminarthema den Fokus der Instruktion bildete: «Erstellen Sie eine Concept-Map zum Thema »Raumbildung und Kommunikation im Spätmittelalter«». Diese war, wie bereits dargelegt, thematisch dahingehend geöffnet, alle themenbezogenen Assoziationen einzubeziehen. Das codierte Segment in Abbildung 1 ist ein Teil einer Concept-Map, die zu Beginn des Seminars vor dem thematischen Einstieg erstellt wurde.

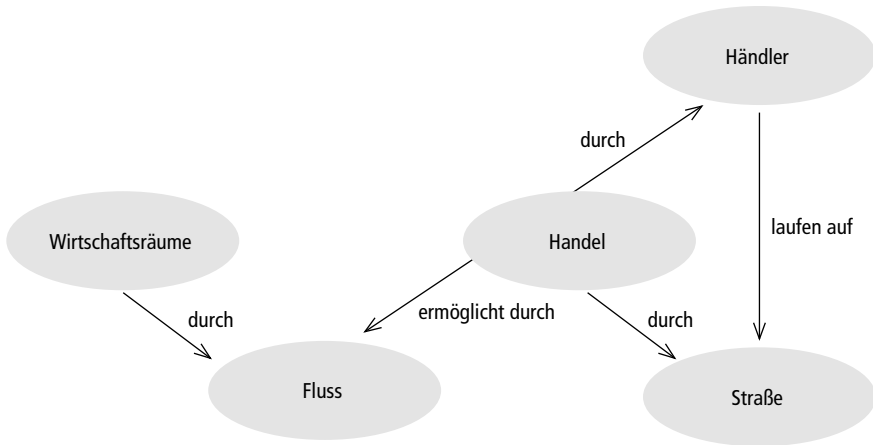


Abbildung 1: Stud3_K1_Z1_Map, Pos. 336/540–882/737

In der Analyse zeigt sich, dass dem Begriff «Handel» eine zentrale Bedeutung zugewiesen wird, indem alle seine direkten Relationen von ihm ausgehend gezogen werden. In ihrer interpretativen Zusammenführung kann die Lesart verfolgt werden, dass «Handel» explizit hinsichtlich der ihn ermöglichenden Voraussetzungen entfaltet wird. Zugleich wird «Handel» auf impliziter Ebene als eine aktive Tätigkeit markiert, die in der Proposition «Handel → durch → Händler» greifbar ist. Diese Interpretation kann damit gestützt werden, dass «Händler» als auf der «Straße» laufend bestimmt werden. Die «Straße» wird selbst wiederum als Voraussetzung für den Handel benannt. Hier kommt auf impliziter Ebene eine Historisierung des Handels zum Ausdruck, da die Fortbewegung der Händler als Fußreise benannt wird. Führt man das Laufen der Händler auf der Straße mit dem eingefügten Begriff «Fluss» zusammen, so kann daraus abgeleitet werden, dass «Handel» vordergründig als Bewegung und somit als Reisetätigkeit qualifiziert wird.

Mit Blick auf den Verweisungszusammenhang, in dem «Handel» steht, zeigt sich, dass dieser mittelbar an «Wirtschaftsräume» angebunden ist. Die Verbindung von «Handel» mit «Wirtschaftsräume[n]» wird allerdings semantisch nicht explizit vollzogen, was darin zum Ausdruck kommt, dass die Verbindungen von Wirtschaftsräumen und Handel zu dem Begriff «Fluss» jeweils auf «Fluss» zulaufen. Damit stellen sie kein Propositionspfad dar. Gleichwohl wird implizit eine semantische Verbundenheit dadurch erzeugt, dass Flüsse – hier als Kollektivsingular eingefügt – ebenso für Wirtschaftsräume wie für den Handel eine Voraus-

setzung darstellen. Aus dieser Interpretation heraus wurde der Code «Voraussetzungen der Handelstätigkeit» für das Segment vergeben.

Für den Nachvollzug, wie sich aus der Fortsetzung der datennahen Auseinandersetzung mit weiteren Concept-Maps das «Bedingungsbeziehungen» als ein bedeutsames Phänomen ergab, wird ein anderes Concept-Map-Segment aus demselben Erhebungskontext herangezogen. In dem Segment wird der Begriff «Krieg» in einer spezifischen Bedeutung mit dem Begriff «Kommunikation» verknüpft. Diese Proposition ist Bestandteil einer Map, die zum Abschluss des mediävistischen Seminars erstellt wurde. Die hier zum Ausdruck kommende gleichrangige Betrachtung der Maps vom Beginn sowie zum Abschluss begründet sich in der nichtevaluativen Ausrichtung der Studie (siehe Abbildung 2).



Abbildung 2: Stud8_K1_Z2_Map, Pos. 654/221–1115/268

Mit Blick auf die inhaltliche Aussage der Verbindung fällt auf, dass die formulierte Proposition semantisch zwei Ebenen enthält, die gegenläufig zueinander sind. Die Verwendung des Schrägstrichs zur Trennung von «erschwert» und «verstärkt» kann als Marker interpretiert werden, mit dem die beiden Implikationen des Kriegs für die Kommunikation als ambivalent qualifiziert werden: Einerseits erschwert Krieg die Kommunikation und andererseits verstärkt er sie. Damit wird Krieg als ein die Kommunikationsmöglichkeiten prägender Aspekt entfaltet. Der «Krieg» impliziert somit Bedingungen für «Kommunikation», worin im Unterschied zum vorangegangenen Fallbeispiel die Verbindung weniger eine voraussetzungsbezogene darstellt als vielmehr eine hinsichtlich der Folgen.

Neben diesen beiden Formen der semantischen Strukturierung wurden weitere doppelläufige Relationen herausgearbeitet, bei denen zwei Aspekte als wechselseitig füreinander bedeutsam markiert werden. Darüber hinaus konnten beispielsweise auch fundierende Verbindungen identifiziert werden, bei denen ein Aspekt aus dem anderen hervorgeht. Mit dem Ziel des Theoretisierens dieser verschiedenen Phänomene der Strukturierung von Kenntnissen gilt es nun, sie zu gruppieren und zu bündeln. Für diesen Schritt sieht die KGT-Methodologie das fokussierte Codieren als analytische Strategie vor.

4.2 Gruppieren und Abstrahieren im Rahmen des fokussierten Codierens

Das fokussierte Codieren konzentriert sich darauf, mittels des fortwährenden Vergleichs von Phänomenen, eine Entscheidung zu treffen, welche Codes zu Kategorien abstrahiert werden. Die Abstraktion (Charmaz, 2014, 146) wird in der Form vollzogen, dass Codes auf Basis eines gemeinsamen Kriteriums (Strübing, 2018, 44), also hinsichtlich Ähnlichkeit und Relationalität untereinander (Muckel, 2011, 336), gruppiert werden. Demnach zeichnen sich Kategorien dadurch aus, dass «sie eine Vielzahl von Codes zusammenfassen und dabei auch widersprüchliche Hypothesen und Varianten zu integrieren vermögen» (Muckel, 2011, 350). Darauf aufbauend erfolgt die «Ausdifferenzierung und Priorisierung von Kategorien» (Hohage, 2016, 118). So wird im fokussierten Codieren festgelegt, was als bedeutsam für theoriebildenden Überlegungen erachtet und in den Daten weiterverfolgt wird. Diese Codierphase ist entsprechend selektiver und grobmaschiger.

Wendet man sich in diesem Zugang den vorangegangenen Codierbeispielen zu, stellt sich die Frage nach dem gemeinsamen Kriterium der herausgearbeiteten Phänomene. Gleichwohl sich die Entfaltung von «Handel» und «Krieg» im Konkreten qualitativ unterscheidet, lässt sich auf einer abstrakteren Ebene eine erste Gemeinsamkeit feststellen: Bei beiden wird ein kausaler Zusammenhang dargelegt. Diese Form der kausalen Verknüpfung wird auch von Günther-Arndt als Beispiel für ein «Denkzeug» angeführt und unter anderem von «zeitlichen», «finalen» und «modalen» Relationen unterschieden (2005, 42). In der weitergehenden Auseinandersetzung mit dem Material zeigte sich, dass die von Günther-Arndt unterschiedenen Relationsformen in den gestalteten Concept-Maps, wenn man auch mittelbare Zusammenhänge betrachtet, in einer komplexen Form zusammengeführt werden. Vor diesem Hintergrund wurden die herausgearbeiteten Phänomene weiter verglichen und nach einer Gemeinsamkeit gesucht, die eine abstrahierende Gruppierung ermöglicht. So kristallisierte sich im Vergleich mit den anderen Relationierungsmustern heraus, dass sie alle sich dadurch auszeichnen, dass Bedingungen in dem Sinne dargelegt werden, dass verschiedene «historische Fokussierungen» (Nitsche, 2019, 132) – in Anlehnung an Nitsche hier als subjektive Inblicknahme von Geschichtlichem verstanden – auf ihre Voraussetzungen, wechselseitigen Abhängigkeiten, Folgen, füreinander bestimmende Gegebenheiten oder Funktionen verknüpft werden. Durch ihre Integration entstand so die Kategorie «Bedingungszusammenhänge flechten».

Mit dieser Kategorie erfolgte die Rückkehr zu den Daten, bei der größere Teilbereiche analysiert wurden. Die beiden nachstehenden Segmente stellen Bei-

spiele für fokussierte Codierungen von Segmenten dar, in denen fachliche Kenntnisse in ihren Bedingungsbeziehungen dargestellt werden. Das erste Segment entstammt einer Concept-Map, die ebenfalls im ersten Erhebungskontext und zu Beginn des Seminars zum Thema «Raumbildung und Kommunikation im Spätmittelalter» erstellt wurde (siehe Abbildung 3).

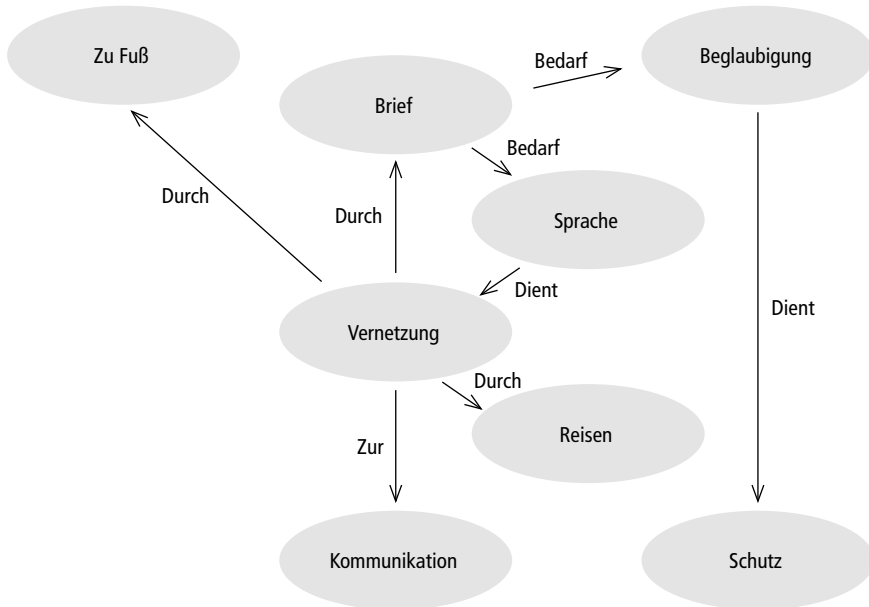


Abbildung 3: Stud2_K1_Z1_Map, Pos. 805/110–1192/430

In dem Segment zeigt sich, wie die Begriffe «Brief», «Sprache» und «Reisen» sowie «Kommunikation in unterschiedlichen Formen als Bedingungen für die «Vernetzung» qualifiziert werden. Ihre Relationiertheit bringt dabei zum Ausdruck, dass die Begriffe in ihrer wechselseitigen Bedeutsamkeit füreinander organisiert werden. Die Verarbeitung von Zeit ist in dem Segment mit der Proposition, dass Briefe der Beglaubigung bedürfen und diese dem Schutz dient, implizit greifbar. Explizit wird sie im nachstehenden Teilbereich einer Concept-Map, die im zweiten Erhebungskontext zum Abschluss des Seminars zum Thema «Ordnung und sozialer Wandel im Mittelalter» erstellt wurde.

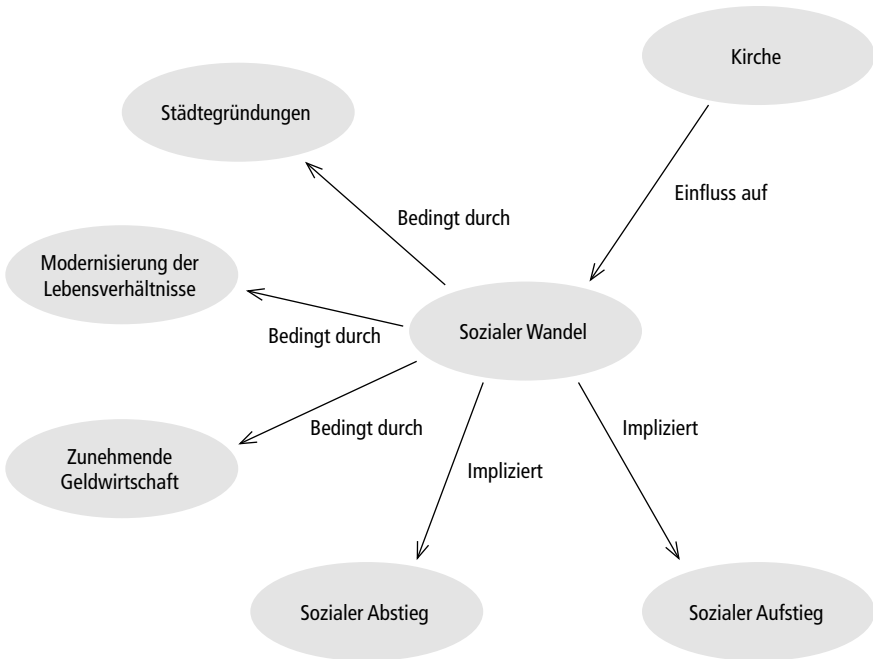


Abbildung 4: Stud11_K2_Z2_Map, Pos. 112/223–637/594

Wenngleich der Begriff «Sozialer Wandel» bereits selbst als Veränderung zum Ausdruck bringend interpretiert werden kann, zeigt sich, dass auch die mit ihm direkt verknüpften Begriffe explizit auf Bewegung in der Zeit verweisen. Zugleich werden die Begriffe wie «Städtegründungen» mit sozialem Wandel durch die Relation «bedingt durch» ebenso begrifflich in ein zeitlichkeitsbezogenes Bedingungsverhältnis gesetzt, das nicht näher bestimmt ist. «Sozialer Wandel» wird darüber hinaus mit dem sozialen Ab- und Aufstieg relationiert. Indem diese als dem sozialen Wandel inbegriffen markiert werden, kann die Lesart angelegt werden, dass die den sozialen Wandel präformierenden Bedingungen in den Kontext von sozialem Ab- und Aufstieg gestellt und als Implikationen qualifiziert werden. Durch die gezogenen Verbindungen wird «Sozialer Wandel» so als eine multirelationale Bewegung in der Zeit entfaltet und die hier präsentierten fachlichen Kenntnisse als ein interdependentes Bedingungsgefüge organisiert.

Im weiteren Prozess fokussierten Codierens ergaben sich als von «Bedingungsbeziehungen flechten» unterscheidbare Kategorien «Eingliedern», «Kategorisieren» sowie «Prozessualisieren». Diese wurden im Rahmen des theo-

retischen Codierens wiederum in der erwähnten Kategorie «den Kenntnisraum texturieren» zusammengeführt, vor dem Hintergrund ihrer Gemeinsamkeit, dass mit ihnen den Kenntnissen eine mehrdimensionale Struktur verliehen wird. So gewannen die Kategorien selbst wieder den Status von Subkategorien.

4.3 Anknüpfen an den Diskurs durch das Einbeziehen geschichtsdidaktischer Theorien

Der im Vorigen vollzogene Schritt der Abstraktion wird beim theoretischen Codieren dahingehend fortgesetzt, dass nun die herausgearbeiteten Kategorien sowie vorliegenden disziplinären Erkenntnisse wechselseitig aufeinander bezogen werden (Hohage, 2016, 119). Durch das Zusammenführen können die Kategorien theoretisch angereichert und mit dem disziplinären Diskurs verknüpft werden (Charmaz, 2014, 151). Diese bildet die Grundlage dafür, abduktiv bestehende Theorien und Kategorien zu ergänzen, weiterzuführen oder ein neues Deutungsangebot als eine «analytic story» (Charmaz, 2014, 150) auszuarbeiten. Entgegen der Codierstrategien, wie sie Strauss und Corbin (1996) mit dem Codierparadigma oder Glaser (1998) mit Codierfamilien vorschlagen, realisiert sich das theoretische Codieren und die Relationierung der Kategorien gemäß der KGT-Methodologie entlang eines «interdisziplinär offenen Blicks auf die vorhandene Theorielandschaft» (Hohage, 2016, 120).

In der Abstraktion der im Projekt herausgearbeiteten Kategorien «Eingliedern», «Kategorisieren», «Prozessualisieren» und des skizzierten «Bedingungs-zusammenhänge flechten» ergab sich als Oberkategorie «den Kenntnisraum texturieren», die das zentrale Phänomen greift, dass das aktive historische Wissen im Prozess seiner Organisation als eine komplexe innere Struktur und Plastizität gewinnt. Diese Kategorie weist eine Nähe zu der von der FUER-Gruppe modellierten Strukturierungskompetenz auf. Ihr zufolge hat als «historisch strukturierungskompetent [...] der zu gelten, der über zur Strukturierung genutzte historische Begriffe auf unterschiedlichen Abstraktionsniveaus verfügt und sich ihrer funktionalen Bedeutung für die Systematisierung der Domäne bewusst ist» (Schöner, 2007, 280).

In der Zusammenführung dieser Kompetenzmodellierung und der herausgearbeiteten Subkategorien zeigt sich, dass anhand des herausgearbeiteten Texturierens die Strukturierung historischen Wissens, für das im FUER-Modell das kategorienbezogene gedankliche Sichbewegen zwischen Konkretem und Abstraktem ins Zentrum gerückt wird (von Borries & Pflüger, 2007, 626), näher kon-

zeptualisiert werden kann. Im Einklang mit ihr zeigen die Daten, dass die Organisation historischen Fachwissens in Form einer Strukturbildung vollzogen wird, mit der der so konstituierte Kenntnisraum eine mehrdimensionale Struktur gewinnt. Bei dieser stehen individuell jeweils unterschiedliche Strukturmuster im Vordergrund. So kann das Strukturieren im Sinne der FUER-Gruppe aus empirischer Perspektive weiterführend konzeptualisiert werden, dass dem vertikal ausgerichteten Verfügen über zentrale Kategorien ein horizontales Zu- und Nebenordnen, das Knüpfen inhaltreicher semantischer Verbindungslinien zwischen historischen Fokussierungen als Organisationsmuster gleichrangig bedeutsam zur Seite gestellt ist. Dies ist anschlussfähig an das von Wineburg herausgearbeiteten Kontextualisieren als aktivem und konstruktivem Verweben der Assoziationen und Kenntnisse, das historisch Forschende bei der Interpretation von Quellen vollziehen (Wineburg, 1998, 339). Die Wahl der Begriffe als sprachliche Fassung von Konzepten, die von den Geschichtsstudierenden im Rahmen der Studie des Verfassers als bedeutsame und abstrakte Knotenpunkte genutzt wurden, folgt dabei keiner streng klassifizierbaren Systematik und orientiert sich nicht ausschließlich an geschichtswissenschaftlichen Grundbegriffen. Vielmehr ist es ebenso durch fachliche wie durch alltagsweltliche Interessenschwerpunkte und nicht zuletzt auch durch den Blick auf und das Verhältnis zu Geschichtlichem, in das man sich selbst setzt, bestimmt.

5 Reflexion und Ausblick

Für einen qualitativ ausgerichteten Zugang zu historischem Wissen wurden eine Adaption des Concept-Mapping sowie ein analytischer Perspektivrahmen vorgestellt. Dieser Ansatz ergänzt die bisher etablierten Methoden in der Form, als dass mit dem Concept-Mapping je nach Instruktion in unterschiedlicher Form nicht nur konkrete themenbezogene Kenntnisse und deren kognitive Organisation, sondern auch das «residue knowledge» (Kühberger, 2016, 95) interpretativ zugänglich wird. Jedoch wurde in der Durchführung der Studie auch deutlich, dass es sich bewährt, gerade bei thematisch freieren und/oder zeitlich begrenzten Mappingszenarien mit Methoden wie dem begleitenden Lauten Denken (so z. B. Fenn, Stebener & Urban, 2019) oder mit nachgelagerten Interviews, in denen auch die Concept-Maps zum Gesprächsgegenstand gemacht werden können, zu triangulieren. So kann zum einen der Zugang zum Forschungsgegenstand mit Blick auf das «Verstehen komplexer Zusammenhänge» (Flick, 2018, 26) erweitert

werden. Beispielsweise eröffnet das begleitende Laute Denken oder ein nachträgliches Erläutern der gestalteten Map im Rahmen von Interviews einen Zugang zur narrativen Funktion der eingefügten Begriffe als Abkürzungen im Hinblick auf die historische Sinnbildung. Zum anderen erscheint eine Triangulation im Rahmen von Concept-Mapping-Studien auch aus einer methodenkritischen Perspektive zielführend. Aus dieser heraus können je nach Fragestellung die Grenzen des Mapping beispielsweise durch Verbaldaten überschritten und so tiefere Erkenntnisse gewonnen werden. Bezüglich der Grenzen zeigte sich etwa, dass Concept-Maps als ausschließliches Datenmaterial nur bedingt evaluative Rückschlüsse auf das erworbene Wissen zulassen, da Lernende trotz intensiver Übungsphasen unterschiedlich gut mit der spezifischen Anforderung der Methode zurechtkommen, das eigene Wissen in Form von einzelnen Begriffen und mithilfe von Propositionen in kurzen Aussagefragmenten darzulegen. Letzteres bleibt auch mit Blick auf den geschichtsdidaktischen Konsens bezüglich der Narrativität historischen Wissens und seine damit verbundenen unterschiedlichen Ebenen ein limitierendes Moment des Concept-Mapping für die Analyse historischer Sinnbildungsprozesse.

Gleichwohl birgt das Concept-Mapping Potential für die geschichtsdidaktische Lehr-Lern-Forschung. So können beispielsweise restriktivere Formen der Erhebung in Kombination mit deduktiv-induktiven Auswertungsverfahren die themenspezifische Erfassung von «Strategien im Umgang mit verschiedenen Begriffsarten» (Langer-Plän & Beilner, 2006, 247) ermöglichen. In dieser Form kann das Concept-Mapping ein methodischer Zugang sein, das enge Verhältnis sowie die Wechselwirkung von Wissen und Kompetenz näher zu bestimmen (Schreiber, 2012). Für eine solche theoriesensible Erkundung und kritisches Weiterdenken kann auch die KGT-Methodologie ein dienlicher Forschungsstil sein.

Literatur

- Alavi, Bettina (2004). Begriffsbildung im Geschichtsunterricht: Problemstellungen und Befunde. In Uwe Uffelman & Manfred Seidenfuß (Hrsg.), *Verstehen und Vermitteln. Armin Reese zum 65. Geburtstag* (S. 21–38). Idstein: Schulz-Kirchner.
- Alavi, Bettina (2006). Begriffslernen. In Ulrich Mayer/Hans-Jürgen Pandel/Gerhard Schneider & Bernd Schönemann (Hrsg.), *Wörterbuch Geschichtsdidaktik* (S. 27–28). Schwalbach/Ts.: Wochenschau.
- Barricelli, Michele & Sauer, Michael (2015). Empirische Lehr-Lern-Forschung im Fach Geschichte. In Georg Weißeno & Carla Schelle (Hrsg.), *Empirische Forschung in den gesellschaftswissenschaftlichen Fachdidaktiken: Ergebnisse und Perspektiven* (S. 185–200). Wiesbaden: Springer.

- Barte, Burghard (2022). «Das Ganze mehr verdichten» – Theoriefragment eines datenbe-
gründeten Deutungsangebots zur Organisation historischen Fachwissens. In Meike
Hensel-Grobe & Heidrun Ochs (Hrsg.), *Geschichtsdidaktik Update. Aktuelle geschichts-
didaktische Forschungsansätze der Early Career Researchers* (S. 189–210). Göttingen: V&R
unipress.
- Barte, Burghard/Landgraf, Julia & Mühling, Andreas (2020). Concept-Mapping als ein Weg
zur Erfassung professionsbezogener Wissensanteile im Praxissemester. In Tobias
Heinz/Birgit Brouër/Margot Janzen & Jörg Kilian (Hrsg.), *Formen der (Re)Präsentation
fachlichen Wissens: Ansätze und Methoden für die Lehrerinnen- und Lehrerbildung in den
Fachdidaktiken und in den Bildungswissenschaften* (S. 79–94). Münster, New York: Wax-
mann.
- Beaugrande, Robert-Alain de & Dressler, Jürgen (2011). *Einführung in die Textlinguistik*. Tü-
bingen: Max Niemeyer. <https://doi.org/10.1515/9783111349305>. (Original erschienen
1981).
- Bücker, Nicola (2020). Kodieren – aber wie? Varianten der Grounded-Theory-Methodologie
und der qualitativen Inhaltsanalyse im Vergleich. *Forum Qualitative Sozialforschung/
Forum: Qualitative Social Research*, 21(1), Art. 2. <https://doi.org/10.17169/fqs-21.1.3389>.
- Borries, Bodo von & Pflüger, Christine (2007). Verfügen über überfachliche Kategorien als
Teil der «Historischen Sachkompetenz»: Das Beispiel «Herrschaft». In Andreas Körber/
Waltraud Schreiber & Andreas Schöner (Hrsg.), *Kompetenzen historischen Denkens: Ein
Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik* (Bd. 2,
S. 621–639). Neuried: ars Una.
- Cañas, Alberto J. & Novak, Joseph D. (2006). Re-examining the foundations for effective use
of concept maps. In Alberto J. Cañas & Joseph D. Novak (Hrsg.), *Concept maps: Theory,
methodology, technology. Proceedings of the second international conference on concept
mapping* (S. 494–502). San José: Universidad de Costa Rica. [https://cmc.ihmc.us/cmc-
proceedings/](https://cmc.ihmc.us/cmc-proceedings/) (11.04.2022).
- Cañas, Alberto J. & Novak, Joseph D. (2012). Freedom vs. restriction of content and structure
during concept mapping – Possibilities and limitations for construction and assess-
ment. In Alberto J. Cañas/Joseph D. Novak & Jacqueline Vanhear (Hrsg.), *Concept maps:
Theory, methodology, technology. Proceedings of the fifth International Conference on
Concept Mapping. Volume 2, Full papers, part two* (S. 247–257). Msida: Malta. [https://cmc.
ihmc.us/cmc-proceedings/](https://cmc.ihmc.us/cmc-proceedings/) (11.04.2022).
- Charmaz, Kathy (2014). *Constructing grounded theory* (2. Aufl.). Los Angeles: Sage.
- Drie, Jannet van & Boxel, Carla van (2008). Historical reasoning: Towards a framework for
analyzing students' reasoning about the past. *Educational Psychology Review*, 20, 87–
110. <https://doi.org/10.1007/s10648-007-9056-1>.
- Fenn, Monika (2018). Conceptual change von Vorstellungen über epistemologische Basis-
konzepte bei Grundschülerinnen und -schülern fördern? Ergebnisse einer explorati-
ven Interventionsstudie. In Monika Fenn (Hrsg.), *Frühes historisches Lernen: Projekte
und Perspektiven empirischer Forschung* (S. 146–199). Frankfurt a. M.: Wochenschau.
- Fenn, Monika/Stebener, Franziska & Urban, Stefanie (2019). Concept Maps zur Erfassung
von professionellem Fachwissen bei (angehenden) Geschichtslehrkräften. In Wal-
traud Schreiber/Béatrice Ziegler & Christoph Kühberger (Hrsg.), *Geschichtsdidakti-*

- scher Zwischenhalt: Beiträge aus der Tagung «Kompetent machen für ein Leben in, mit und durch Geschichte» in Eichstätt vom November 2017 (S. 264–271). Münster, New York: Waxmann.
- Flick, Uwe (2018). Triangulation: *State of the Art* der deutschen und internationalen Diskussion. In Jutta Ecarius & Ingrid Miethe (Hrsg.), *Methodentriangulation in der qualitativen Bildungsforschung* (2. Aufl., S. 21–44). Opladen: Barbara Budrich.
- Gebhardt, Markus/Mühling, Andreas/Gartmeier, Martin & Tretter, Tobias (2015). Wissen über Inklusion als gedankliches Netz. Vergleich zwischen Studierenden des beruflichen und des sonderpädagogischen Lehramts mithilfe von Concept Maps. *Zeitschrift für Heilpädagogik*, 66, 609–622.
- Glaser, Barney G. (1998). *Doing grounded theory: Issues and discussions*. Mill Valley: Sociology.
- Graf, Dittmar (2014). Concept Mapping als Diagnosewerkzeug. In Dirk Krüger/Ilka Parchmann & Horst Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 325–337). Berlin: Springer Spektrum.
- Günther-Arndt, Hilke (2014). Historisches Wissen und Wissenserwerb. In Hilke Günther-Arndt & Meik Zülsdorf-Kersting (Hrsg.), *Geschichtsdidaktik. Praxishandbuch für die Sekundarstufe I und II* (6. Aufl., S. 24–47). Berlin: Cornelsen Scriptor.
- Jahr, Silke (2009). Strukturelle Unterschiede des Wissens zwischen Naturwissenschaften und Geisteswissenschaften und deren Konsequenzen für den Wissenstransfer. In Tilo Weber & Gerd Antos (Hrsg.), *Typen von Wissen: Begriffliche Unterscheidung und Ausprägungen in der Praxis des Wissenstransfers* (S. 76–98). Frankfurt a. M.: Peter Lang.
- Hohage, Christoph (2016). Kathy Charmaz' konstruktivistische Erneuerung der Grounded Theory. In Claudia Equit & Christoph Hohage (Hrsg.), *Handbuch Grounded Theory. Von der Methodologie zur Forschungspraxis* (S. 108–127). Weinheim, Basel: Beltz Juventa.
- Kelle, Udo & Kluge, Susanne (2010). *Vom Einzelfall zum Typus. Fallvergleich und Fallkontrastierung in der qualitativen Sozialforschung* (2. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kinchin, Ian M. & Hay, David B. (2000). How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational Research*, 42(1), 43–57. <https://doi.org/10.1080/001318800363908>.
- Kühberger, Christoph (2012). Konzeptionelles Wissen als besondere Grundlage des historischen Lernens. In Christoph Kühberger (Hrsg.), *Historisches Wissen: Geschichtsdidaktische Erkundung zu Art, Tiefe und Umfang für das historische Lernen* (S. 33–74). Schwalbach/Ts.: Wochenschau.
- Kühberger, Christoph (2016). Historisches Wissen – verschiedene Formen seiner Strukturiertheit und der Wert von Basiskonzepten. In Wolfgang Hasberg & Holger Thüneemann (Hrsg.), *Geschichtsdidaktik in der Diskussion: Grundlagen und Perspektiven* (S. 91–108). Frankfurt a. M.: Peter Lang.
- Lange, Kristina (2011). *Historisches Bildverstehen oder Wie lernen Schüler mit Bildquellen? Ein Beitrag zur geschichtsdidaktischen Lehr-Lern-Forschung*. Berlin: Lit.
- Langer-Plän, Martina & Beilner, Hartmut (2006). Zum Problem historischer Begriffsbildung. In Hilke Günther-Arndt & Michael Sauer (Hrsg.), *Geschichtsdidaktik empirisch: Untersuchungen zum historischen Denken und Lernen* (S. 215–249). Berlin: Lit.

- Mathis, Christian (2015). «Irgendwie ist doch da mal jemand geköpft worden»: *Didaktische Rekonstruktion der Französischen Revolution und der historischen Kategorie Wandel*. Baltmannsweiler: Schneider.
- Muckel, Petra (2011). Die Entwicklung von Kategorien mit der Methode der Grounded Theory. In Günter Mey & Katja Mruck (Hrsg.), *Grounded Theory Reader* (2. Aufl., S. 333–352). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Mühling, Andreas (2016). Aggregating concept map data to investigate the knowledge of beginning CS students. *Computer Science Education*, 26(2–3), 176–191. <https://doi.org/10.1080/08993408.2016.1241340>.
- Nitsche, Martin (2019). *Beliefs von Geschichtslehrpersonen – Eine Triangulationsstudie*. Bern: hep. <https://doi.org/10.36933/9783035516005>.
- Rüsen, Jörn (2008). *Historisches Lernen: Grundlagen und Paradigmen. Mit einem Beitrag von Ingetraud Rüsen* (2. Aufl.). Schwalbach/Ts.: Wochenschau.
- Rüsen, Jörn/Fröhlich, Klaus/Horstkötter, Hubert & Schmidt, Hans Günter (1991). Untersuchungen zum Geschichtsbewusstsein von Abiturienten im Ruhrgebiet. In Bodo von Borries/Hans-Jürgen Pandel & Jörn Rüsen (Hrsg.), *Geschichtsbewusstsein empirisch* (S. 221–344). Pfaffenweiler: Centaurus.
- Sander, Wolfgang (2011). Konzepte und Kategorien in der politischen Bildung. In Thomas Goll (Hrsg.), *Politikdidaktische Basis- und Fachkonzepte* (S. 32–43). Schwalbach/Ts.: Wochenschau.
- Sauer, Michael (2015). Begriffsarbeit im Geschichtsunterricht. *Geschichte lernen*, 168, 2–11.
- Schmidt, Siegfried J. (2010). *Die Endgültigkeit der Vorläufigkeit: Prozessualität als Argumentationsstrategie*. Weilerswist: Velbrück Wissenschaft.
- Schöner, Alexander (2007). Kompetenzbereich historische Sachkompetenzen. In Andreas Körber/Waltraud Schreiber & Andreas Schöner (Hrsg.), *Kompetenzen historischen Denkens: Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik* (Bd. 2, S. 265–314). Neuried: ars una.
- Schreiber, Waltraud (2012). Zum Verhältnis von Wissen und Kompetenzen: Ein Essay. In Christoph Kühberger (Hrsg.), *Historisches Wissen: Geschichtsdidaktische Erkundung zu Art, Tiefe und Umfang für das historische Lernen* (S. 119–134). Schwalbach/Ts.: Wochenschau.
- Stöckle, Friederike (2011). «Die armen kleinen Bäuerlein ...»: *Schülervorstellungen zu mittelalterlichen Herrschaftsformen: Ein Beitrag zur didaktischen Rekonstruktion*. Oldenburg: Didaktisches Zentrum.
- Strauss, Anselm & Corbin, Janet (1996). *Grounded Theory. Grundlagen qualitativer Sozialforschung*. Weinheim: Beltz.
- Strübing, Jörg (2018). Grounded Theory: Methodische und methodologische Grundlagen. In Christian Pentzold/Andreas Bischof & Nele Heise (Hrsg.), *Praxis Grounded Theory: Theoriegenerierendes empirisches Forschen in medienbezogenen Lebenswelten. Ein Lehr- und Arbeitsbuch* (S. 27–52). Wiesbaden: Springer VS.
- Thornberg, Robert & Charmaz, Kathy (2014). Grounded theory and theoretical coding. In Uwe Flick (Hrsg.), *The Sage handbook of qualitative data analysis* (S. 153–169). Los Angeles: Sage.

- Ventzke, Marcus (2012). Begriffliches Arbeiten und «Geschichte denken» – theoretische Voraussetzungen und unterrichtliche Vorgehensweisen. In Christoph Kühberger (Hrsg.), *Historisches Wissen: Geschichtsdidaktische Erkundung zu Art, Tiefe und Umfang für das historische Lernen* (S. 75–102). Schwalbach/Ts.: Wochenschau.
- Wineburg, Sam (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science*, 22(3), 319–346. https://doi.org/10.1207/s15516709cog2203_3.
- Zabold, Stefanie (2020). *Vor dem ersten Geschichtsunterricht: zur empirischen Erschließung des historischen Denkens junger Lernerinnen und Lerner*. Frankfurt a. M.: Wochenschau.
- Ziegler, Béatrice (2012). Historische Konzepte und Kompetenzmodelle – ein empirischer Zugang zur Bedeutung von Wissen für Kompetenzen. In Christoph Kühberger (Hrsg.), *Historisches Wissen: Geschichtsdidaktische Erkundung zu Art, Tiefe und Umfang für das historische Lernen* (S. 135–152). Schwalbach/Ts.: Wochenschau.

HiTCH II

Was wir aus der Weiterentwicklung des HiTCH-Tests über historische Kompetenzen lernen

Wolfgang Wagner, Christiane Bertram und das HiTCH-Konsortium¹

1 Einleitung

Am 12. Juli 2021 wurde an das Postfach der HiTCH-Gruppe («HiTCH» für «Historical Thinking – Competencies in History») eine Mail geschickt:

«Sehr geehrte Frau Bertram,

mein Name ist N. N. und ich bin Lehrkraft am N. N. Gymnasium in Holstein. Mit großem Interesse habe ich das HiTCH-Projekt wahrgenommen und es in unserer Fachschaft für das Fach Geschichte vorgestellt, was positive Reaktionen nach sich zog und auch die Bereitschaft, dieses Instrument an unserer Schule einzusetzen. Wir hätten großes Interesse daran, mit Ihnen gemeinsam zu arbeiten und erachten dies, auch wegen der für Schleswig-Holstein geltenden Fachanforderungen, als sehr sinnvoll. Sehr gut vorstellen könnten wir uns den Einsatz des HiTCH-Projekts im kommenden Schuljahr in unserem 8. Jahrgang. Wir würden uns sehr freuen, wenn Sie diesbezüglich Kontakt zu uns aufnehmen würden.

Mit freundlichen Grüßen, N. N.»

¹ Das HiTCH-Konsortium bestand zum Zeitpunkt der Erhebung aus den durch das Programm des Bundesministeriums für Bildung und Forschung zur Förderung von Forschungsvorhaben in Anknüpfung an Large-Scale-Assessments unter der Fördernummer LSA006 finanziell unterstützten Beteiligten der Universitäten Tübingen, Eichstätt-Ingolstadt, Hamburg sowie des Leibniz-Instituts für Wissensmedien Tübingen. Weiterhin beteiligten waren Geschichtsdidaktiker*innen der Universität Bochum, Universität Paderborn, der Pädagogischen Hochschulen Salzburg und Nordwestschweiz. Für eine aktuelle Übersicht zu den am HiTCH-Konsortium Beteiligten sei auf die Website des HiTCH-Projekts verwiesen (<https://uni-tuebingen.de/fakultaeten/wirtschafts-und-sozialwissenschaftliche-fakultaet/faecher/fachbereich-sozialwissenschaften/hector-institut-fuer-empirische-bildungsforschung/forschung/aktuelle-studien/hitch/#c1510410>).

Nicht nur die Kolleg*innen in der HiTCH-Gruppe², auch die Kolleg*innen am Institut für Qualitätssicherung Schleswig-Holstein (IQSH) Benjamin Stello (Landesfachberater Geschichte), Karin Hülsen (Fachfortbildnerin Geschichte) und Nadja Einhaus (Testplattform, Online-Tests) nahmen die Mail mit großer Freude zur Kenntnis, denn eine Auswahl von HiTCH-Aufgaben in digitaler Form lag für Schleswig-Holstein bereits vor. Aus den Instrumenten in HiTCH I und HiTCH II hatten wir Aufgaben ausgesucht und einen digitalen Kompetenztest («HiTCH-Online») erstellt. Das Testinstrument war bereits programmiert. Geplant war, im Schuljahr 2020/21 mit Lehrkräften und Schüler*innen eine Forschungsstudie mit Aufgaben aus «HiTCH-Online» durchzuführen und das Instrument danach auf dem Bildungsserver des IQSH den Lehrkräften bereitzustellen. Doch dann kam Corona und die Studie musste auf das Schuljahr 2021/22 verschoben werden. Ab dem Schuljahr 22/23 sollen ausgewählte Aufgaben daraus interessierten Lehrkräften zur Verfügung gestellt werden. Die Lehrkräfte werden somit die Möglichkeit haben, die fachbezogenen Kompetenzen der Lernenden reliabel und valide mit einem einfach zu handhabenden standardisierten Leistungstest zu erfassen. Die Bearbeitung der Testaufgaben gibt den Lehrkräften einen Einblick in Kompetenzen historischen Denkens der Schüler*innen. Zudem können sie mit den Lernenden ins Gespräch darüber kommen, was das Besondere am «historischen Denken» ist, das mit den Aufgaben erfasst werden soll.

Die Entwicklung eines Kompetenztests ist aufwendig, doch es lohnt sich. Zum einen zwingt die Entwicklung eines standardisierten Tests dazu, die zentralen Ziele des Geschichtsunterrichts – Förderung eines reflexiven Geschichtsbewusstseins und des historischen Denkens – in Kompetenzaufgaben zu «übersetzen». Die «Operationalisierung» verlangt eine Definition der Zielkonstrukte und leistet damit einen Beitrag zur Theoriebildung in der Geschichtsdidaktik. Darüber hinaus sind standardisierte Instrumente, die reliabel und valide messen und einfach eingesetzt und ausgewertet werden können, unverzichtbar, um in groß angelegten Interventionsstudien über die Stichprobe hinaus Aussagen über die Grundgesamtheit treffen zu können. Und nicht zuletzt geben standardisierte und frei verfügbare Leistungstests den Lehrkräften einen unkomplizierten Einblick in den fachlichen Leistungsstand der Schüler*innen, wobei im Idealfall kri-

2 In dem interdisziplinären Konsortium ist die Empirische Bildungsforschung durch die Universität Tübingen und die Geschichtsdidaktik mit aktuell sieben Universitäten und Hochschulen in Aarau, Bochum, Eichstätt, Hamburg, Konstanz und Salzburg vertreten (Salzburg ist mit zwei Hochschulen, der Universität und der Pädagogischen Hochschule, beteiligt).

teriale (z. B. durch Expert*innen definierte Mindeststandards) Vergleiche klassenintern, aber auch klassen- und schulübergreifend möglich sind.

Im vorliegenden Beitrag zeigen wir einen Ausschnitt aus dem nun zehnjährigen Forschungsprojekt der HiTCH-Gruppe. In diesem Beitrag gehen wir einen Schritt zurück und schauen uns die psychometrische Güte des Tests an. Datengrundlage ist hier die Validierungsstudie von HiTCH 1 (Trautwein et al., 2017), in der zusätzliche Testhefte und Aufgaben eingesetzt wurden. Viele dieser in 2014 erstmals erprobten Aufgaben sind Bestandteil der digitalen HiTCH-Version.

Wir skizzieren im Folgenden in der gebotenen Kürze den theoretischen Rahmen, stellen an zwei Aufgaben die Konstruktionslogik des Tests vor und geben eine Übersicht über die Aufgaben in HiTCH I und HiTCH II. Die Forschungsfragen beziehen sich darauf, ob allgemeine Aufgabenmerkmale einen Zusammenhang aufweisen mit der Bearbeitung durch die Schüler*innen. Nach der Vorstellung der eingesetzten Methoden und Ergebnisse diskutieren wir diese abschließend.

2 Was sind «historische Kompetenzen»?

Was ist «historisch»? Was ist das Besondere an den Kompetenzen «historischen Denkens»? Wodurch unterscheidet sich «historisch zu denken» davon, einfach schlau zu sein, aufmerksam zu lesen, zu analysieren, zu vergleichen und Schlussfolgerungen zu ziehen? Um diese Fragen geht es in der Auseinandersetzung über die Definition von Kompetenzen und über die Konzeptualisierung von Kompetenzmodellen. Die Kompetenzmodelle unterscheiden sich hinsichtlich der Begrifflichkeit bei der Definition historischer Kompetenzen. So steht in der FUER-Gruppe (Körber, Schreiber & Schöner, 2007) und im Bildungsplan 2016 Baden-Württemberg (Ministerium für Kultus, Jugend und Sport, 2016) das «historische Denken» im Mittelpunkt: «historical thinking and understanding» findet sich in der US-amerikanischen Fachdidaktik (z. B. VanSledright, 2014) und in den US-Standards für historische Kompetenzen (UCLA, 2021). Die Niederländer nutzen den Begriff des «historical reasoning» (van Drie & van Boxtel, 2008), häufig wird im deutschsprachigen Raum die «narrative Kompetenz» in den Fokus gestellt (z. B. Pandel, 2013; van Norden, 2011). Gemeinsam ist allen Modellen die Erkenntnis der narrativen Geschichtstheorie (Danto, 1965; Ricoeur, 1983/1988), dass Geschichte ein vor allem sprachlich vermitteltes Konstrukt ist (Barricelli, Gautschi & Körber, 2012), bei dem durch die Herstellung von Verbindungen zwischen zeitlich differenten Ereignissen oder Strukturen mit den Frage-

stellungen und Interessen der jeweiligen Gegenwart und der Betrachter*in eine «Geschichte» rekonstruiert und erzählt wird, mit der ein historischer «Sinn» gebildet wird. Der US-amerikanische Fachdidaktiker Bruce VanSledright (2014) nennt dies in Anlehnung an Rösen (1983): «to make sense of the past». Ziel des Geschichtsunterrichts ist es, dass dies methodisch kontrolliert und reflektiert passiert.

Im FUER-Kompetenzstrukturmodell (Körper et al., 2007) wird hieraus die Notwendigkeit abgeleitet, bei der «Re-Konstruktion» der Vergangenheit methodengeleitet und quellenbasiert vorzugehen und im Gegenzug bei der «De-Konstruktion» vorliegender Darstellungen, die Argumentationsweise und «Gemachtheit» der wissenschaftlichen oder geschichtskulturellen Erzeugnisse zu erkennen und zu hinterfragen. Der «historische Denkprozess» (Hasberg & Körper, 2003) liegt dem FUER-Modell zu Grunde. Basierend auf einem grundlegenden Verständnis von Geschichte («Sachkompetenzen») und ausgehend von einer Verunsicherung in der Gegenwart werden Fragen an die «Vergangenheit» oder an die «Geschichte» gestellt («Fragekompetenzen»), Quellen und Darstellungen in re- und de-konstruktiven Arbeitsschritten ausgewertet («Methodenkompetenzen»), um Antworten auf die Fragen zu finden («Orientierungskompetenzen»). Um die Güte vorliegender (wissenschaftlicher oder geschichtskultureller) Darstellungen einzuschätzen, bieten sich die Triftigkeitskriterien von Rösen (1983) an. Diese unterscheiden in der De-Konstruktion zwischen der (empirischen) Basis, der (narrativen) Argumentation und der (normativen) Botschaft einer vorliegenden Darstellung, lassen sich aber auch nutzen als Qualitätskriterien für die Beurteilung von Schüler*innenantworten, die im Rahmen von Aufgabenstellungen im Geschichtsunterricht eigene Narrationen erstellen (Bertram et al., 2015).

Der HiTCH-Test soll mit standardisierten Instrumenten einen Einblick in die Verfügung über die hier nur sehr kurz skizzierten komplexen Kompetenzen historischen Denkens geben. Es ist offensichtlich, dass nur Teilkompetenzen und hier jeweils kleine Ausschnitte des Denkprozesses beziehungsweise die Resultate der Erwägungen mit stark formalisierten und standardisierten Aufgaben erfasst werden können. Umso spannender ist es darüber nachzudenken, was das spezifisch Historische an diesen Aufgaben ist und inwiefern die Aufgabengestaltung (z. B. motivierendes Bildmaterial) einen Einfluss auf die Bearbeitung der Aufgaben hat. Es geht hier um typische Fragen, die in der Psychometrie während der Testentwicklung gestellt und auf Basis der Testergebnisse bearbeitet werden.

3 HiTCH I und HiTCH II: Forschungsfragen zur Testentwicklung

Der «finale» HiTCH-Test (HiTCH I, 91 Items) war nicht final – und sollte es auch gar nicht sein. Bereits die HiTCH-I-Validierungsstudie in 2014 haben wir dazu genutzt, zusätzliche Testaufgaben auszuprobieren, um den HiTCH-Aufgabenpool zu erweitern. Erstes Ziel der Validierungsstudie mit 2853 Schüler*innen aller Schularten der 9. Klasse war die Überprüfung der psychometrischen Qualität des HiTCH-I-Tests, der bereits zweimal pilotiert worden war. Um die historischen Kompetenzen der Lernenden von der Lesekompetenz und der Intelligenz abgrenzen zu können, wurde im zweiten Teil der Testung (nach dem Einsatz von HiTCH I) in einem Teil der Stichprobe entweder ein Test zur Erfassung der kognitiven Grundfähigkeiten oder ein Lesekompetenz-Test bearbeitet. In dem Rest der Stichprobe wurden drei Erweiterungstesthefte mit neu konzipierten Aufgaben eingesetzt.

Tatsächlich erwies sich der HiTCH-I-Test mit 15 Aufgaben (insgesamt 91 Items), die die Kompetenzen historischen Denkens nach dem FUER-Modell (Körper, Schreiber & Schöner, 2007) erfassen sollten, als psychometrisch valide. Es ließ sich eine zentrale Dimension «Historisches Denken» (HiTCH-I-Faktor) bestimmen, die sich von der Lesekompetenz und von den kognitiven Grundfähigkeiten separieren ließ. Allerdings zeigten sich nur geringfügige spezifische Varianzanteile für die Teilkompetenzen Frage-, Methoden-, Orientierungs- und Sachkompetenzen (Trautwein et al., 2017). Der vorliegende Beitrag berichtet nun von der Überprüfung der HiTCH-Aufgaben aus den drei Erweiterungstestheften (EH), die wir unter *HiTCH II* zusammenfassen. Die Aufgaben in EH 1 und EH 2 waren in einer ähnlichen Logik wie der HiTCH-I-Test konzipiert. Mit und ohne zusätzliche Materialien wurden alle Kompetenzbereiche des historischen Denkens nach dem FUER-Modell mit unterschiedlichen Aufgabenformaten (häufig Zuordnungsaufgaben) adressiert. Das EH 3 stellt eine Besonderheit dar, da alle Aufgaben an der Pädagogischen Hochschule Salzburg entwickelt und ausschließlich in österreichischen Klassen eingesetzt wurden. Die 22 Aufgaben basieren auf kurzem Material aus der Geschichtskultur, die für die Lernenden ein besonders hohes Motivationspotenzial haben müssten: Karikaturen, Comicstrips, Computerspiele oder Zeitungsartikel stellen die Materialbasis dar, die in einem verhältnismäßig einfach zu erschließenden Multiple-Choice-Format de-konstruiert werden. Die Aufgaben haben sich vielfach bewährt, die psychometrische Qualität wurde bereits detailliert berichtet (Kühberger, Neureiter & Wagner, 2018).

In der vorliegenden Studie wurde überprüft, inwieweit die im Rahmen der HiTCH I Validierungsstudie (2014) eingesetzten Aufgaben aus den drei zusätzlichen Testheften dasselbe Konstrukt (mit-)erfassen und somit als HiTCH II zur Erweiterung des HiTCH-Aufgabenpools genutzt werden können. Konkret ging es dabei um folgende Fragestellung:

1. Welche Aufgaben aus den Erweiterungstestheften sind zur Erweiterung des HiTCH-Aufgabenpools geeignet? Unterscheiden sich dabei die Aufgaben aus EH 1 und 2 von denen aus EH 3 (das ausschließlich De-Konstruktionskompetenz fokussiert) hinsichtlich des Zusammenhangs mit dem HiTCH-I-Test? Sind die EH also unterschiedlich repräsentativ für das, was der HiTCH-I-Test misst?

Eine eigenständige psychometrische Prüfung des gesamten Itempools aus den Erweiterungstestheften war designbedingt nicht möglich, da keine Zusammenhänge zwischen Items aus verschiedenen Testheften geschätzt werden können, weil jeweils nur ein zusätzliches Testheft bearbeitet wurde. Als Bezugspunkt zur Beurteilung der psychometrischen Güte der Items wird daher der HiTCH-I-Test verwendet, wobei berücksichtigt werden muss, dass die nach den HiTCH-I-Aufgaben entstandenen Testhefte möglicherweise auch zusätzliche Aspekte erfassen, die theoretisch als valide im Sinne eines «breiteren» Konstrukts der historischen Kompetenz betrachtet werden können. Entsprechend bezog sich die zweite Fragestellung auf folgende Bereiche:

2. Lassen sich bedeutsame EH-spezifische (Kompetenz-)Facetten nachweisen, die über HiTCH I hinausgehen? Dies könnte insbesondere für das Heft 3 der Fall sein, das ausschließlich De-Konstruktionsaufgaben enthielt. Darüber hinaus sollte untersucht werden, inwieweit sich auf Einzelitemebene mehr oder weniger starke Zusammenhänge mit einer jeweils testheftspezifischen Facette ergeben.

Darüber hinaus wollten wir in der vorliegenden Studie herausfinden, ob sich gegebenenfalls über das in HiTCH I erfasste Konstrukt hinausgehende Kompetenzfacetten auf Itemmerkmale zurückführen lassen, die in den Erweiterungstestheften besonders ausgeprägt vertreten sind:

3. Da die Items aus den Erweiterungstestheften auch hinsichtlich der für die Beantwortung als relevant erachteten Itemmerkmale Lebensnähe, Konkretheit und Aufgabenmaterial (es wurden neben Texten auch Bilder als Grundlage verwendet) variierten, sollte geprüft werden, ob sich spezifische Zusammenhänge der Items je nach jeweiligem Merkmal ergeben.

4 Stichprobe

Die Stichprobe wurde in Deutschland, Österreich und der Schweiz erhoben und umfasste insgesamt 2851 Schülerinnen und Schüler der neunten Klassenstufe aus allen Bildungsgängen (49.3 % weiblich, Alter in Jahren $M = 14.4$, $SD = 0.71$). Der Test war so angelegt, dass der HiTCH-I-Test in sämtlichen sechs Testheftversionen enthalten war, aber nur in drei Testheftversionen jeweils eines der Erweiterungstesthefte (EH 1: $n = 444$; EH 2: $n = 423$; EH 3: $n = 587$). Die Teilstichprobe für EH 3 wurde wie oben bereits erwähnt ausschließlich in Österreich erhoben (Kühberger, Neureiter & Wagner, 2018).

5 Operationalisierung: die Instrumente

Die Instrumente der Erweiterungstesthefte 1 (7 Aufgaben, 55 dichotome Items sowie ein Summenscore einer Reihenfolgeaufgabe, siehe Tabelle 1) und 2 (7 Aufgaben, 61 dichotome Items, siehe Tabelle 2) enthielten Aufgaben und Items verschiedener Standorte, die in einer ähnlichen Logik wie der HiTCH-I-Test alle Kompetenzbereiche adressierten. Häufig sollten Materialien (Quellen und Darstellungen) bestimmten Aussagen oder Konzepten zugeordnet werden. Um die Konstruktionslogik der Aufgaben am Beispiel zu zeigen, stellen wir im Folgenden eine typische Aufgabe aus HiTCH I vor, die auf Textmaterial basiert. Die Aufgabe wurden nicht in dieser Form eingesetzt, sondern für die zentrale Publikation zu dem HiTCH-Test «nachgebaut» (Trautwein et al., 2017), um die Originalaufgaben in Studien nutzen zu können.

Beispiel: Hier werden die drei geschichtstheoretisch begründeten «Fokussierungen», wie sie in der Sechsfeldermatrix systematisiert sind (Hasberg & Körber, 2003), bei der Klassifikation von Fragen genutzt. Historische Fragen sind häufig dadurch gekennzeichnet, dass sie einen narrativen Zusammenhang zwischen

Prozessen in der Vergangenheit und ihrer Bedeutung für die Gegenwart herstellen («Bedeutung für heute»). Davon abzugrenzen sind solche Fragen, die schwerpunktmäßig auf «Zusammenhänge in der Vergangenheit» zielen, und solche, die nach «Einzelheiten in der Vergangenheit» fragen. Die Jugendlichen beantworten die Fragen nicht, sondern sie zeigen, ob sie in der Lage sind, die unterschiedlichen Reichweiten der Fragen zu erkennen (Trautwein et al., 2017, 68).

Aufgabenbeispiel

<p>Wenn man sich für die Hexenverfolgungen früher interessiert, kann man ganz unterschiedliche Fragen stellen. Die unten stehenden Fragen zum Thema «Hexen» sollt ihr nicht selbst beantworten. Ihr sollt lediglich einordnen, <u>um welchen Typ von Frage</u> es sich jeweils handelt: Handelt es sich um eine Frage, die nach Einzelheiten in der Vergangenheit fragt (z. B. Namen, Ereignisse)? Oder ist es eine Frage, die nach Zusammenhängen in der Vergangenheit fragt (z. B. Entwicklungen)? Oder handelt es sich um eine Frage, in der direkt nach der Bedeutung der Vergangenheit für heute gefragt wird? Bitte setze nur ein Kreuz pro Zeile – dort, wo der Schwerpunkt der Frage liegt!</p>			
	Die Frage zielt vor allem ab auf		
	Einzelheiten in der Vergangenheit	Zusammenhänge in der Vergangenheit	Bedeutung für heute
In welchem Jahr wurde die letzte Hexe in Deutschland verbrannt?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wie kam es dazu, dass in der Frühen Neuzeit so viele angebliche Hexen verfolgt wurden?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Welche Erkenntnisse über den Umgang mit Außenseitern können aus der Hexenverfolgung gezogen werden?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Inwiefern haben sich die Argumente, mit denen die «Hexen» verfolgt wurden, vom 14. bis zum 17. Jahrhundert verändert?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Sind die Gegner der Hexenverfolgung Helden, die uns ein Vorbild sein können?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Wer war der Autor des Buches «Der Hexenhammer»?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

In den folgenden drei Tabellen werden alle Aufgaben, die im Rahmen von HITCH II in den drei Erweiterungstheften eingesetzt wurden, im Überblick vorgestellt.

Tabelle 1: Übersicht über die Aufgaben im Erweiterungstestheft 1 (HiTCH II)

Aufgabe, Thema & Epoche	Material- grundlage	Art der Aufgabe	Fokus Kompetenz- bereich
Aufgabe 1: Jugendkrawalle in der Schweiz in den 1980er-Jahren			
Aufgabe 1.1	Kurzer Quellentext	Zuordnung der Items zum Material	Methodenkompetenz
Aufgabe 1.2	Kurzer Quellentext	Zuordnung der Items zum Material	Methodenkompetenz
Aufgabe 1.3	Konzepte (z. B. Rechtsstaat)	Zuordnung der Konzepte zum Material	Sachkompetenz
Aufgabe 2: Epochen und ihre Bezeichnungen	Zwei kurze Definitionen	Zuordnung von Items zu dem Material	Sachkompetenz
Aufgabe 3: Die Mayflower (1620), Gründungsdokument der Vereinigten Staaten			
Aufgabe 3.1	Textbausteine einer Quelle	Reihenfolge herstellen	Methodenkompetenz
Aufgabe 3.2	Textbausteine einer Darstellung	Reihenfolge herstellen	Methodenkompetenz
Aufgabe 4: Fragen an die Geschichte der USA	Drei Ebenen / Fokussierungen & Beispielfragen	Zuordnung der Fragen zu den drei Ebenen	Fragekompetenz

Die Aufgaben in den ersten beiden Erweiterungstestheften sind jedes Mal etwas anders aufgebaut. Häufig handelt es sich um Zuordnungsaufgaben, bei denen die Kriterien beziehungsweise Kategorien unterschiedlich benannt sind, sodass die Schüler*innen immer wieder etwas andere kognitive Operationen durchführen müssen. Die Reihenfolge-Aufgaben sind in ähnlicher Weise kognitiv anspruchsvoll.

Tabelle 2: Übersicht über die Aufgaben im Erweiterungstestheft 2 (HiTCH II)

Aufgabe, Thema & Epoche	Materialgrundlage	Art der Aufgabe	Kompetenzbereich
Aufgabe 5: Fragen an die Geschichte der DDR	Drei Ebenen Fokussierungen & Beispielfragen	Zuordnung der Fragen zu den drei Ebenen	Fragekompetenz
Aufgabe 6: Die Pest aus Sicht der Zeitgenossen (1349/1350/1563)			
Aufgabe 6.1	Quellentext & Tatsache vs. Interpretation	Zuordnung der Sinn- schritte der Quelle zu Kategorien	Methoden- kompetenz
Aufgabe 6.2	Quellentext & Tatsache vs. Interpretation	Zuordnung der Sinn- schritte der Quelle zu Kategorien	Methoden- kompetenz
Aufgabe 6.3	Quellentext mit historischen Begriffen & Begriffserklärungen	Zuordnung: Richtigkeit der angebotenen Erklärung	Sachkompetenz
Aufgabe 7: Die Mayflower (1620)	Zwei Texte: Quelle & Darstellung	Zuordnung von Items zu den beiden Texten	Methoden- kompetenz
Aufgabe 8: Epistemologie			
Aufgabe 8.1: Historische Quellen	Aussagen zu den Merkmalen einer Quelle	Richtigkeit der Aussage	Sachkompetenz
Aufgabe 8.2: Historische Darstellungen	Aussagen zu den Merkmalen einer Darstellung	Richtigkeit der Aussage	Sachkompetenz

Das Erweiterungstestheft 3 (22 Aufgaben, 92 dichotome Items; Kühberger et al., 2018) unterscheidet sich, da hier ausschließlich De-Konstruktionsaufgaben auf der Grundlage geschichtskultureller Phänomene angeboten werden, die sich durch eine hohe Varianz der Materialgrundlage auszeichnen. Die Epochen und Themen sind vielfältig: von der Ur- und Frühgeschichte (Ötzi), über die Antike (Sklaverei) bis hin zum Ersten Weltkrieg, Nationalsozialismus und zu aktuellen Auseinandersetzungen über das Erbe der Vergangenheit. Bei fast allen Aufgaben wurde ein Multiple-Choice-Antwortformat mit direktem Bezug zu den Aufgabenmaterialien verwendet.

Tabelle 3: Übersicht über die Aufgaben im Erweiterungstestheft 3 (HiTCH II)

Aufgabe, Thema & Epoche	Materialgrundlage	Art der Aufgabe
Aufgabe 1: Antike (Antike Sklaverei)	Karikatur	Zuordnungsaufgabe, Single Choice
Aufgabe 2: Neuzeit (Erinnerungskultur zum Ersten Weltkrieg)	Informationsbroschüre	Single Choice
Aufgabe 3: ohne Epoche (fiktive Familiengeschichte)	Comic	Reihenfolgeaufgabe
Aufgabe 4: Mittelalter (Ritter)	Bilder aus Computerspiel und bildliche Quelle/Miniatur	Zuordnungsaufgaben
Aufgabe 5: Neuzeit (Ausbruch des Ersten Weltkriegs)	Schriftliche Quelle/Zeitungsbericht	Forced Choice (richtig/falsch), Single Choice
Aufgabe 6: Antike (Alexander III. von Makedonien)	Zeitungsbericht	Single Choice
Aufgabe 7: Ur-/Frühgeschichte (menschliche Entwicklung)	Werbung	Single Choice
Aufgabe 8: Mittelalter (Burg)	Karikatur	Single Choice
Aufgabe 9: Neuzeit (Piraten)	Spielzeug	Single Choice, Forced Choice (richtig/falsch)
Aufgabe 10: Antike (römisches Alltagsleben)	Spielzeug, Spielfilm, Sachbuch, privater Internetblog, Internetseite eines Museums	Rating («sehr verlässlich» bis «gar nicht verlässlich»)
Aufgabe 11: Zeitgeschichte (Schwarzenegger-Biografie)	Buchrezension	Single Choice, Zuordnung
Aufgabe 12: Ur-/Frühgeschichte (Ötzi)	Rekonstruktionszeichnung/ Zeitungsbericht	Single Choice
Aufgabe 13: Zeitgeschichte (Iran/1979)	Interview mit Regisseur	Single Choice
Aufgabe 14: Zeitgeschichte (Obama)	Politische Rede	Single Choice
Aufgabe 15: Zeitgeschichte (Madonna)	Kurzbiografie/Musikzeitschrift	Zuordnung
Aufgabe 16: Neuzeit (Native Americans)	Blog-Eintrag	Single Choice

6 Statistische Analysen

Für die Analysen wurden *Nested-Factor-Modelle* (Gustafsson & Åberg-Bengtsson, 2010) mit Probit-Link für dichotome Items (IRT-Modelle) eingesetzt unter Verwendung eines Bayes-Schätzers (Mplus, Muthén & Muthén, 1998–2017). Dazu

wurde für jedes Erweiterungstestheft eine separate Analyse mit den Items des jeweiligen Hefts sowie den Items des HiTCH-I-Tests berechnet, wobei sämtliche Items auf einen Generalfaktor (HiTCH-I-Faktor) luden und die EH-Items zusätzlich auf einen mit dem Generalfaktor unkorrelierten EH-spezifischen Faktor. Der EH-spezifische Faktor kann dabei im Sinne einer zusätzlichen Facette betrachtet werden, auf der sich Personen – bei identischen Ausprägungen auf dem HiTCH-I-Faktor – unterscheiden und somit unterschiedliche Lösungswahrscheinlichkeiten für die EH-Items aufweisen. Bei durchweg positiven Ladungen (Itemdiskriminationen) auf den spezifischen Faktor (die latente Variable) könnte dies beispielsweise bedeuten, dass die Items eine gemeinsame Kompetenzfacette erfassen, die im HiTCH-I-Test unterrepräsentiert ist. Schüler*innen mit höheren Ausprägungen auf dieser Facette hätten dann bei identischen Fähigkeiten gemäß HiTCH-I-Test höhere Lösungshäufigkeiten auf den jeweiligen EH-Items.

In einem ersten Schritt wurden sämtliche Items der Erweiterungstesthefte ermittelt, die einen hinreichend hohen Zusammenhang mit dem HiTCH-I-Test aufwiesen (Kriterium: untere Grenze des 95 Prozent *Credibility Intervals* der standardisierten Faktorladung auf den HiTCH-I-Faktor größer .20) und somit potenziell zur Erweiterung des HiTCH-Aufgabenpools genutzt werden können. Alle weiteren Analysen basierten dann auf dieser jeweiligen Itemauswahl. Damit sollte gewährleistet werden, dass die Analyseergebnisse nicht verzerrt werden durch Items, die als nicht hinreichend valide betrachtet werden können. Mit dem verbliebenen Itempool haben wir untersucht, ob sich dabei die Aufgaben aus EH 1 und 2 von denen aus EH 3 (das ausschließlich De-Konstruktionskompetenz fokussiert) hinsichtlich des Zusammenhangs mit dem HiTCH-I-Test unterscheiden. Sind die EH also unterschiedlich repräsentativ für das, was der HiTCH-I-Test misst?

Bezüglich der zweiten Fragestellung wurden dann für die ausgewählten Items die Ladungen auf den jeweils EH-spezifischen Faktor untersucht. Durchgängig hohe positive Ladungen würden hier für eine gemeinsame zusätzlich erfasste Kompetenzfacette über HiTCH I hinaus sprechen. Dabei wäre es allerdings auch möglich, dass sich solche Zusammenhänge zumindest teilweise durch spezifische Itemmerkmale erklären lassen (unabhängig vom adressierten Kompetenzbereich). Um die Bedeutung solcher Itemmerkmale für die EH-spezifischen Faktoren zu ermitteln, wurden die Modelle erweitert um zwei zusätzliche unkorrelierte Faktoren, auf die die jeweiligen Items, für die das Merkmal zutrifft beziehungsweise nicht zutrifft, luden. Es wurde also beispielsweise ein zusätzlicher Faktor spezifiziert, auf den sämtliche Items des jeweiligen Erweiterungstest-

hefts mit Bildmaterial als Grundlage luden, und ein weiterer Faktor, auf den die verbleibenden EH-Items luden. Insgesamt wurden sechs solcher Modelle berechnet (zwei für EH 1, eins für EH 2 und drei für EH 3).

7 Ergebnisse

7.1 Forschungsfrage 1: Zusammenhänge der Items aus den drei Erweiterungstestheften mit HiTCH I

Zunächst wurden drei Modelle mit jeweils sämtlichen Items aus dem HiTCH-I-Test sowie den Items aus dem jeweiligen Erweiterungstestheft berechnet, wobei alle Ladungen frei geschätzt wurden mit Ausnahme der jeweils ersten Ladung auf einem Faktor, die zur Modellidentifikation auf den Wert 1 fixiert wurde. Dabei ergaben sich deutlich positive standardisierte Ladungen (untere Grenze des 95 %-Glaubwürdigkeitsintervalls $> .20$) auf dem HiTCH-I-Faktor für 33 von 56 Items in EH 1 ($.34 \leq \lambda_{\text{std}} \leq .72$, $Mdn = .52$), 40 von 61 Items in EH 2 ($.35 \leq \lambda_{\text{std}} \leq .74$, $Mdn = .52$) sowie für 63 von 93 Items in EH 3 ($.31 \leq \lambda_{\text{std}} \leq .79$, $Mdn = .53$). Somit zeigte sich für einen großen Teil der Items aus den Erweiterungstestheften, dass sie mit der anhand von HiTCH I gemessenen historischen Kompetenz deutlich assoziiert sind. Allerdings gab es keine Hinweise auf tendenziell unterschiedlich hohe Ladungen für das EH 3, das ausschließlich De-Konstruktionskompetenz fokussiert, verglichen mit den Ladungen für EH 1 beziehungsweise EH 2.

7.2 Forschungsfrage 2: Erfassen die Testhefte jeweils auch spezifische Kompetenzfacetten?

In der zweiten Forschungsfrage haben wir uns gefragt: Welche Ladungen ergeben sich auf der Einzelitemebene für die jeweilige Itemauswahl auf einen testheftspezifischen zusätzlichen Faktor, insbesondere des Hefts 3, neben den Ladungen auf den Generalfaktor HiTCH I? Die Modelle waren analog zu denen aus Forschungsfrage 1, bezogen sich allerdings nun jeweils nur auf die ausgewählten Items mit hinreichend großen standardisierten Ladungen auf den HiTCH-I-Faktor.

Dabei fanden sich auf den jeweiligen spezifischen Faktoren der Erweiterungstesthefte folgende statistisch signifikante Ladungen (Forschungsfrage 2): 26 positive von insgesamt 32 geschätzten Ladungen in EH 1, fünf positive und acht negative von insgesamt 39 geschätzten Ladungen in EH 2 sowie 40 positive

von insgesamt 62 geschätzten Ladungen in EH 3. Mit Ausnahme von EH 2 fanden sich also Hinweise darauf, dass die Items aus den Erweiterungstestheften großteils jeweils noch eine weitere, im HiTCH-I-Test nicht umfassend repräsentierte Kompetenzfacette erfassen, die bei EH 3 darauf zurückgehen könnte, dass sich sämtliche Items auf die De-Konstruktionskompetenz beziehen. Bei EH 1 ist eine Erklärung der positiven Zusammenhänge der Items unter Kontrolle des HiTCH-I-Faktors hinaus auf Basis der fokussierten Kompetenzbereiche weniger nahe liegend, da dort drei verschiedene Bereiche (siehe Tabelle 1) vertreten sind. Bei EH 2 gingen die statistisch signifikant negativen Ladungen fast durchgängig auf eine einzige Aufgabe (sieben der acht negativen Ladungen) zurück, was eher auf aufgabenspezifische Besonderheiten verweist, wobei die fünf positiven Ladungen auf drei verschiedene Aufgaben verteilt waren.

7.3 Forschungsfrage 3: Bedeutung der Itemmerkmale Lebensnähe, Konkretheit und Grundlage

Als dritte Forschungsfrage hatten wir formuliert: Lassen sich testheftspezifische Aspekte anhand der Itemmerkmale Lebensnähe (3a), Konkretheit (3b) und Itemgrundlage Text versus Bild (3c) erklären? Das dritte Kriterium lässt sich einfach ermitteln (wird in der Aufgabe ein Bild eingesetzt?). Das erste Kriterium (Lebensnähe) haben wir im Hinblick auf den Lebensalltag von Vierzehn- und Fünfzehnjährigen verstanden. Items zu den «Jugendkrawallen in der Schweiz» oder zum «Tagebuch von Anne Frank» wurden zum Beispiel wegen der lebensbiografischen Nähe als «lebensnah» codiert. «Konkretheit» bezieht sich auf die Klarheit der Aufgabenstellung. Wenn die Lösung einer Aufgabe möglichst wenig kognitive Anforderungen stellte, die nicht unmittelbar mit der angezielten historischen Kompetenz in Verbindung stehen (z. B. bei kurzen Textauszügen, prägnanter Aufgabenstellung und Beispiel-Items), dann wurde das jeweilige Item als konkret codiert.

Daher wurden in einem ersten Schritt alle Items in den drei Erweiterungstestheften von zwei Ratern codiert. Bei der Überprüfung der Interrater-Übereinstimmung fanden wir bezogen auf die Lebensnähe (LN) eine Übereinstimmung von 77 Prozent; bei der Konkretheit 76 Prozent, bei der Materialgrundlage 99 Prozent. In den Fällen, in denen die Urteile der beiden Rater nicht übereinstimmten, wurde anhand des dritten Beurteilers eine Entscheidung für eine der beiden Antwortalternativen getroffen. Nur in EH 3 gab es Unterschiede auf allen drei Merkmalen. In EH 1 und EH 2 wurden keine Bilder verwendet (die Materialgrundlage

war also konstant). Zudem variierte das Merkmal Lebensnähe nicht in EH 2 (durchgängig niedrige Lebensnähe).

Die Ergebnisse zu der dritten Forschungsfrage werden im Folgenden detailliert vorgestellt. Die Modelle bezogen auf EH 1 und EH 3 wurden nun jeweils um zwei zusätzliche Faktoren erweitert, auf die sämtliche Items mit niedriger beziehungsweise hoher Lebensnähe luden (Forschungsfrage 3a; bei EH 2 wurden sämtliche Items als wenig lebensnah eingestuft). Die Ergebnisse sind in Tabelle 4 dargestellt. Für den Faktor für niedrige Lebensnähe ergaben sich statistisch signifikante Ladungen jeweils nur in positiver Ausprägung für EH 1. In EH 3 hingegen fand sich für diesen Faktor keine statistisch signifikante Ladung. Bezogen auf den Faktor für hohe Lebensnähe zeigte sich ebenfalls kein konsistentes Bild. Für EH 1 ergaben sich drei statistisch signifikante Ladungen, für EH 3 eine statistisch signifikante positive Ladung. Auf den spezifischen Faktoren für die gesamten Erweiterungshefte verblieben weiterhin nahezu unverändert viele statistisch signifikante positive Ladungen. Die Ergebnisse sprechen also insgesamt nicht dafür, dass die eingeschätzte Lebensnähe der Items die für die Erweiterungstesthefte bestehenden spezifischen Assoziationen der Items von besonderer Relevanz sind. In den Modellen ohne die spezifischen Faktoren für Lebensnähe ergab sich kein Muster mit statistisch signifikant positiven sowie negativen Ladungen, die man gegebenenfalls auf Lebensnähe hätte zurückführen können. Und die Modellerweiterung um Faktoren für Lebensnähe führte zu einem inkonsistenten Ergebnismuster.

Tabelle 4: Standardisierte Ladungen auf Faktoren für Items mit niedriger bzw. hoher Lebensnähe

Faktor	statistisch signifikante standardisierte Ladungen						
	geschätzte Ladungen	positiv	Bereich	negativ	Bereich		
Faktor für Items mit niedriger Lebensnähe							
EH 1	24	11 (46 %)	.19 bis .42	0 (0 %)	—		
EH 3	26	0 (0 %)	—	0 (0 %)	—		
Faktor für Items mit hoher Lebensnähe							
EH 1	7	2 (29 %)	.39 bis .41	1 (14 %)	-.26		
EH 3	25	1 (4 %)	.41	0 (0 %)	—		
EH-spezifischer Faktor							
EH 1	32	22 (69 %)	.14 bis .48	0 (0 %)	—		
EH 3	62	40 (65 %)	.16 bis .41	0 (0 %)	—		

Eine analoge Erweiterung der Modelle um jeweils zwei zusätzliche Faktoren für Konkretheit (niedrig, hoch) ergab folgendes Bild (Forschungsfrage 3b; Ergebnis- se siehe Tabelle 5): Die Ladungen auf die Faktoren für niedrige Konkretheit in den drei Modellen waren nur in EH 1 teilweise statistisch signifikant, nicht jedoch in EH 2 und EH 3. Für die Faktoren, die spezifische Facetten für Items mit hoher Konkretheit repräsentierten, fand sich ein durchgängig konsistentes Muster statistisch signifikanter positiver Ladungen. Auf die spezifischen Faktoren für die gesamten Erweiterungshefte luden 15 (EH 1), 10 (EH 2) und 20 (EH 3) Items statistisch signifikant positiv, bei EH 2 zusätzlich noch ein Item statistisch signifikant negativ. Zusammenfassend lässt sich festhalten, dass die spezifischen Faktoren für hohe Konkretheit deutliche Assoziationen mit den jeweiligen Items aufwiesen. Dabei ist allerdings zu beachten, dass diese sich auch jeweils auf recht große Anteile der Items der Erweiterungshefte beziehen (67 % in EH 1, 88 % in EH 2 sowie 83 % in EH 3).

Tabelle 5: Standardisierte Ladungen auf Faktoren für Items mit niedriger bzw. hoher Konkretheit

Faktor	statistisch signifikante standardisierte Ladungen						
	geschätzte Ladungen	positiv		Bereich	negativ		Bereich
Faktor für Items mit niedriger Konkretheit							
EH 1	10	4	(40 %)	.24 bis .38	0	(0 %)	—
EH 2	4	0	(0 %)	—	0	(0 %)	—
EH 3	10	0	(0 %)	—	0	(0 %)	—
Faktor für Items mit hoher Konkretheit							
EH 1	21	11	(52 %)	.17 bis .45	0	(0 %)	—
EH 2	34	14	(41 %)	.18 bis .36	0	(0 %)	—
EH 3	51	15	(29 %)	.16 bis .47	0	(0 %)	—
EH-spezifischer Faktor							
EH 1	32	15	(47 %)	.17 bis .46	0	(0 %)	—
EH 2	39	10	(26 %)	.22 bis .46	1	(3 %)	-.21
EH 3	62	20	(32 %)	.21 bis .43	0	(0 %)	—

Bezogen auf eine Modellerweiterung um die zwei Faktoren für Items mit Text-beziehungswise Bildgrundlage (Forschungsfrage 3c), die ausschließlich in EH 3 möglich war, da nur dort auch Bildmaterial verwendet wurde, fanden sich jeweils

drei positive ($.23 \leq \lambda_{\text{std}} \leq .37$) beziehungsweise negative statistisch signifikante Ladungen ($-.44 \leq \lambda_{\text{std}} \leq -.22$) auf den textbezogenen Faktor (insgesamt 27 geschätzte Ladungen). Für die Items mit Bildgrundlage ergaben sich sechs statistisch signifikante Ladungen, davon fünf positive ($.30 \leq \lambda_{\text{std}} \leq .60$) und eine negative ($\lambda_{\text{std}} = -.23$) (insgesamt 34 geschätzte Ladungen). Statistisch signifikant positive Ladungen fanden sich weiterhin für 37 Items ($.18 \leq \lambda_{\text{std}} \leq .45$) auf den spezifischen Faktor für das gesamte Erweiterungstestheft. Diese durchgängig positiven Ladungen und die wenigen sowohl positiven wie auch negativen Ladungen auf die Text- beziehungsweise Bild-Faktoren sprechen insgesamt betrachtet für keinen systematischen Einfluss der Verwendung dieser unterschiedlichen Aufgabengrundlagen, was auf eine eingeschränkte Relevanz der allgemeinen Lesekompetenz bei der Bearbeitung der Fragen verweist.

8 Diskussion

Die vorliegende Studie verfolgte zwei Ziele: Erstens sollte geprüft werden, inwieweit die im Rahmen der HiTCH-Validierungsstudie miterhobenen Erweiterungstesthefte ebenfalls zur Messung der zentralen HiTCH-I-Dimension geeignet sind. Zweitens sollten potenzielle spezifische Facetten der damals zuletzt entwickelten Items aufgedeckt und gegebenenfalls anhand von Itemmerkmalen näher untersucht werden.

Die Ergebnisse zeigten, dass große Anteile der Items aus den Erweiterungstestheften bedeutsame Ladungen (Untergrenze des 95 % *Credibility Intervals* der standardisierten Ladung $> .20$) auf den HiTCH-I-Faktor aufweisen (59 % der Items in EH 1 bis 68 % der Items in EH 3). Diese Items können somit prinzipiell als relevante Erweiterungen des HiTCH-Aufgabenpools betrachtet werden. Somit kann der 91 Items umfassende HiTCH-I-Aufgabenpool aus psychometrischer Sicht um 136 zusätzliche Items erweitert werden, wobei selbstverständlich auch inhaltliche Gründe zu berücksichtigen wären, zum Beispiel um ein Übergewicht an Items zum Bereich De-Konstruktion zu vermeiden.

Im Hinblick auf die zweite Fragestellung fanden sich insbesondere bei EH 1 und EH 3 deutliche Hinweise auf eine aufgabenübergreifend erfasste Komponente, die über die mit den HiTCH-I-Items erfasste Kompetenz hinausgeht. Dabei ergaben sich für 81 Prozent (EH 1) beziehungsweise 65 Prozent (EH 3) der Items statistisch signifikant positive Ladungen auf einen jeweils zusätzlichen Faktor für das Erweiterungstheft. Für die als (wenig) lebensnah codierten Items fand

sich für beide Hefte kein Ergebnispattern, das auf eine besondere Relevanz dieses Aspekts hinweist (Fragestellung 3). Auch die Berücksichtigung des zugrundeliegenden Materials (Text versus Bild) erbrachte keine substanzielle Aufklärung der Zusammenhänge. Lediglich für die Items mit hoher Konkretheit ergaben sich bedeutsame Ladungen auf entsprechende spezifische Faktoren, und dies in allen drei Erweiterungstestheften, wobei allerdings mit 67 Prozent (EH 1) bis 88 Prozent (EH 2) recht große Anteile der Items jeweils als hoch in ihrer Konkretheit eingeschätzt wurden.

Insofern könnte man argumentieren, dass die empirische Trennung zwischen EH-spezifischen Faktoren und spezifischen Faktoren für Items mit hoher Konkretheit gewissen Einschränkungen unterlag. Es ist jedoch auch nicht unwahrscheinlich, dass gerade Items mit einer hohen Konkretheit zum Beispiel von Schülergruppen mit relativ geringer Leistungsmotivation in der Testsituation eher aufmerksam bearbeitet werden als Items mit geringer Konkretheit. Vielleicht erstaunlich mag der Befund erscheinen, dass die Aufgabengrundlage (Text versus Bild) kaum eine Rolle zu spielen scheint. Dies zeigte sich in ähnlicher Weise in Kühberger et al. (2018), wobei dort ein höherdimensionales Modell (insgesamt 10 Dimensionen), allerdings ohne frei geschätzte Ladungen, spezifiziert wurde. Zusammen betrachtet spricht dies dafür, dass – zumindest in den hier zugrundeliegenden Teilstichproben – die Schüler*innen einen vergleichbaren Zugang zu beiden Aufgabentypen hatten. Bezogen auf die Lebensnähe stellt sich die Frage, ob hier möglicherweise die jeweils individuelle Perspektive der Schüler*innen, die gegebenenfalls deutlich von den hier vorgenommenen Ratings abweichen könnte, eine größere Rolle spielt.

Zusammenfassend lässt sich festhalten, dass der HiTCH-Aufgabenpool um zahlreiche Items aus den Erweiterungstestheften ergänzt werden konnte. Um den Einfluss spezifischer Dimensionen (d. h. zusätzlicher Facetten der Aufgabe, die sich – bisher – nur eingeschränkt auf bestimmte Itemmerkmale zurückführen lassen) auf das Testergebnis zu minimieren, erscheint es empfehlenswert, ein möglichst breites Spektrum an unterschiedlichen Aufgaben zu verwenden. Diese Erweiterung des Aufgabenspektrums ist auch deswegen wichtig, um Aufgaben aus dem HiTCH-Test in Schulleistungs- und Interventionsstudien und in der Lehrer*innenausbildung einsetzen zu können. Tatsächlich wurden Aufgaben aus dem HiTCH-Aufgabenpool bereits in groß angelegten Interventionsstudien eingesetzt, zum Beispiel in dem Forschungsprojekt «QQM», in dem die Qualität und Quantität der Nutzung des mBooks Geschichte in Belgien untersucht wurde (Schreiber & Bertram, 2018). In der DFG-Zeitzeug*innenstudie (Bertram, in die-

sem Band) wurde eine Kurzversion des HiTCH-Tests eingesetzt. Aufgaben aus dem HiTCH-Test wurden bereits in mehreren experimentellen Studien und Abschlussarbeiten eingesetzt.³

Über den Einsatz in der empirischen Forschung hinaus leisten wir mit den HiTCH-Aufgaben einen Beitrag für die Diskussion über Aufgaben im Geschichtsunterricht. Denn Aufgaben, die in einem Verständnis von Unterricht als ein System von lehrerseitigem Angebot und schülerseitiger Nutzung (Helmke, 2012) an der «Schnittstelle von Angebot und Nutzung» (Thünemann, 2013, 144) stehen, stellen die Basis eines kompetenzorientierten Unterrichts dar (Barricelli, Gautschi & Körber, 2012), da durch die «Anordnung von Inhalten und Medien spezifisch historische Fragen» (Günther-Arndt, 2015, 18) evoziert und historische Denkprozesse angestoßen werden. Die im HiTCH-Projekt entwickelten «Leistungsaufgaben» für die Erfassung der Kompetenzen der Schüler*innen werden zwar nicht unmittelbar als «Lernaufgaben» im Unterricht eingesetzt, aber sie befruchten die Diskussion über die Spezifik kompetenzorientierter Aufgaben.

Vor allem aber bieten die in der HiTCH-Gruppe entwickelten Aufgaben das Potenzial für die Diagnostik der Lernergebnisse im Klassenkontext (Bertram & Trautwein, 2019). Dieses Potenzial erproben wir derzeit in einer Kooperation mit dem Institut für Qualitätssicherung in Schleswig-Holstein. Ein Teil der HiTCH-Aufgaben soll den Lehrkräften für diagnostische Zwecke zur Verfügung gestellt werden. Die eingangs zitierte Mail der Lehrkraft aus Schleswig-Holstein verweist auf den Bedarf an standardisierten Kompetenztests für die Diagnostik der Lernergebnisse in einem «ganz normalen» Geschichtsunterricht.

Literatur

- Barricelli, Michele/Gautschi, Peter & Körber, Andreas (2012). Historische Kompetenzen und Kompetenzmodelle. In Michele Barricelli & Martin Lücke (Hrsg.), *Handbuch Praxis des Geschichtsunterrichts* (S. 207–235). Schwalbach/Ts.: Wochenschau.
- Bertram, Christiane & Trautwein, Ulrich (2019). Diagnostizieren in den gesellschaftswissenschaftlichen Fächern (Fokus: Fach Geschichte). In Friederike Zimmermann, Jens Möller & Thomas Riecke-Baulecke (Hrsg.), *Basiswissen Lehrerbildung: Diagnostik im schulischen Kontext* (S. 128–145). Seelze: Klett-Kallmeyer.
- Bertram, Christiane/Wagner, Wolfgang/Kühberger, Christoph/Hülßen, Karin/Einhaus, Nadja/Stello, Benjamin & Trautwein, Ulrich (im Druck). *Wie gut können Schüler:innen*

3 Interessent*innen können über die HiTCH-Projektwebsite (www.hitch-projekt.de) Kontakt aufnehmen und dürfen nach Absprache und mit der Verpflichtung auf Geheimhaltung Items aus dem HiTCH-Aufgabenpool für eigene Studienvorhaben nutzen.

- historisch denken? Auf dem Weg zu empirisch verankerten Bildungsstandards im Fach Geschichte.* Kiel: Institut für Schulqualität in Schleswig-Holstein (IQSH).
- Bertram, Christiane/Wagner, Wolfgang & Schaser, Elisabeth (2015). Historische Kompetenzen mit offenen Antwortformaten messen – Eine Studie auf Basis der «Sechser-Matrix» des FUER-Modells. In Monika Waldis & Béatrice Ziegler (Hrsg.), *Forschungswerkstatt Geschichtsdidaktik 13. Beiträge zur Tagung «geschichtsdidaktik empirisch 13»* (S. 165–180). Bern: hep.
- Bertram, Christiane/Wagner, Wolfgang/Werner, Michael/Trautwein, Ulrich & Schreiber, Waltraud (2019). Vier Jahre Unterricht mit dem «mBook Belgien»: Zu den Kompetenz- und Interessensverläufen der Lernenden von der neunten bis zwölften Klasse. In Monika Waldis & Béatrice Ziegler (Hrsg.), *Forschungswerkstatt Geschichtsdidaktik 17. Beiträge zur Tagung «geschichtsdidaktik empirisch 17»* (S. 205–217). Bern: hep.
- Drie, Jannet van & Boxtel, Carla van (2008). Historical reasoning: Towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review*, 20(2), 87–110. <https://doi.org/10.1007/s10648-007-9056-1>.
- Günther-Arndt, Hilke (2015). Umriss einer Geschichtsmethodik. In Hilke Günther-Arndt & Saskia Handro (Hrsg.), *Geschichts-Methodik. Handbuch für die Sekundarstufe I und II* (5. überarb. Neuaufl., S. 9–23). Berlin: Cornelsen.
- Gustafsson, Jan-Erik & Åberg-Bengtsson, Lisbeth (2010). Unidimensionality and interpretability of psychological instruments. In Susan E. Embretson (Hrsg.), *Measuring psychological constructs: Advances in model-based approaches* (S. 97–121). Washington, DC: American Psychological Association. <https://doi.org/10.1037/12074-005>.
- Hasberg, Wolfgang & Körber, Andreas (2003). Geschichtsbewusstsein dynamisch. In A. Körber (Hrsg.), *Geschichte – Leben – Lernen. Bodo von Borries zum 60. Geburtstag* (S. 177–200). Schwalbach/Ts.: Wochenschau.
- Helmke, Andreas (2021). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (8. Aufl.). Seelze: Klett-Kallmeyer.
- Körber, Andreas/Schreiber, Waltraud & Schöner, Alexander (Hrsg.) (2007). *Kompetenzen historischen Denkens. Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik*. Neuried: ars una.
- Kühberger, Christoph/Neureiter, Herbert & Wagner, Wolfgang (2018). Umgang mit Darstellungen der Vergangenheit. Historische De-Konstruktion historisch empirisch messen. *Geschichte in Wissenschaft und Unterricht*, (7/8), 418–434.
- Ministerium für Kultus, Jugend und Sport (Hrsg.) (2016). *Bildungsplan Baden-Württemberg. Fach Geschichte am Gymnasium*. http://www.bildungsplaene-bw.de/Lde/Startseite/BP2016BW_ALLG/BP2016BW_ALLG_GYM_G (11.04.2022).
- Muthén, Linda K. & Muthén, Bengt O. (1998–2017). *Mplus user's guide* (8. Aufl.). Los Angeles: Muthén & Muthén.
- National Center for History in the Schools UCLA (1996). *National standards for history*. Basic edition. <https://phi.history.ucla.edu/nchs/history-standards/> (11.04.2022).
- Norden, Jörg van (2011). *Was machst Du für Geschichten? Didaktik eines narrativen Konstruktivismus*. Freiburg: Centaurus.
- Pandel, Hans-Jürgen (2010). *Historisches Erzählen. Narrativität im Geschichtsunterricht*. Schwalbach/ Ts.: Wochenschau.

- Rüsen, Jörn (1983). *Historische Vernunft. Die Grundlagen der Geschichtswissenschaft. Grundzüge einer Historik I*. Göttingen: Vandenhoeck & Ruprecht.
- Thünemann, Holger (2013). Historische Lernaufgaben – theoretische Überlegungen, empirische Befunde und forschungspragmatische Perspektiven. *Zeitschrift für Geschichtsdidaktik*, 12, 141–155.
- Trautwein, Ulrich et al. (2017). *Kompetenzen historischen Denkens erfassen – Konzeption, Operationalisierung und Befunde des Projekts «Historical Thinking – Competencies in History» (HiTCH)*. Münster: Waxmann.
- VanSledright, Bruce A. (2014). *Assessing historical thinking & understanding. Innovative designs for new standards*. New York: Routledge.

Historisches Lernen mit Originalen?

Zur Umsetzung einer Quellenorientierung in Museen

Andrea Brait

1 Vorbemerkungen

Im nachfolgenden Artikel werden ausgewählte Ergebnisse des Forschungsprojekts «Historisches Lernen zwischen Schule und Museum» dargestellt. Darin wurde untersucht, inwiefern Museumsbesuche im Rahmen des Geschichtsunterrichts dem historischen Lernen dienen (zu den Ergebnissen im Detail: Brait, 2020). Dies erfolgte anhand des österreichischen Geschichtsunterrichts in der Sekundarstufe (unter Berücksichtigung der Tatsache, dass dieser stets in einer Fächerkombination organisiert ist) und den neun österreichischen Landesmuseen. Damit wurden klassische Museen in den Blick genommen, die objektbasierte historische Ausstellungen unter anderem zur jeweiligen Landesgeschichte zeigen. Ausgehend von geschichtsdidaktischen Theorien und Forschungserkenntnissen zur Quellenorientierung einerseits und zum Lernort Museum andererseits wird in der Folge – als ein Teilergebnis des Projekts – auf (1) berufsbezogene Überzeugungen von Lehrkräften (die aus Platzgründen jedoch nur mit wenigen Beispielen veranschaulicht werden können) und (2) Umsetzungsvarianten in Bezug auf die von der Geschichtsdidaktik geforderte Quellenorientierung am außerschulischen Lernort Museum eingegangen.

2 Quellenorientierung in der Geschichtsdidaktik

Quellen sind, wie Hans-Jürgen Pandel (2000, 10) es ausdrückt, «Grundlage und Ausgangspunkt historischen Denkens». Die Quellenorientierung ist ihm zufolge in den letzten Jahrzehnten derart fundamental geworden, dass sie gar nicht mehr als geschichtsdidaktisches Prinzip genannt werden muss (Pandel, 2013, 331). Wie Bodo von Borries (2014, 451) nachgewiesen hat, war diese in bundesdeutschen Schulbüchern schon zu Beginn der 1970er-Jahre erkennbar, wenn-

gleich Textquellen früher als Bildquellen eine heuristische Quellenfunktion zugeschrieben wurde. Obwohl die Arbeit mit Quellen, vor allem jene mit Textquellen, nicht bei allen Schüler*innen sehr beliebt ist (Tischner, 2008, 51 f.), gehört sie heute «zu den wenigen fast völlig unumstrittenen Grundprinzipien schulischer Geschichtsvermittlung», wie Christian Spieß (2014, 13) betont. Allerdings beziehen sich diese Feststellungen in erster Linie auf die Bundesrepublik Deutschland, während Michael Sauer (2013, 176) zu Recht darauf verweist, dass in anderen Staaten der Geschichtsunterricht stärker inhaltsbezogen erfolgt und verschiedene Materialien hauptsächlich zum Beleg von Informationen eingesetzt werden.

Für Österreich stellen Ulrike Kipman und Christoph Kühberger (2019, 86) fest, dass die «Wende hin zur Nutzung von historischen Quellen im Geschichtsunterricht zur Initiierung von historischen Lernprozessen nur marginal bis nicht mitgetragen wurde». Dies ist verwunderlich, zumal sich bereits im 1983 publizierten Überblickswerk zum Geschichtsunterricht in Österreich von Heinz Strotzka umfangreiche Hinweise zum Arbeiten mit Quellen finden. Strotzka (1983, 105) ging davon aus, dass die Quellenarbeit «allgemein Anerkennung gefunden» habe. Dies ist nicht nur aufgrund der aktuellen Forschungsdaten anzuzweifeln, sondern auch aufgrund der curricularen Vorgaben. So wurde Quellen in den Lehrplänen für die Sekundarstufe I aus dem Jahr 1985 lediglich die Funktion eines didaktischen Hilfsmittels zugeschrieben, um den Lehrstoff «anschaulich und lebensnah zu gestalten» (BGBl. 34/1985; BGBl. 88/1985).

Ab den 2000er-Jahren wurde hingegen in den Lehrplänen zunehmend gefordert, dass die Schüler*innen eigenständig mit Quellen arbeiten. So heißt es beispielsweise in den Lehrplänen für die Sekundarstufe I aus dem Jahr 2000: «Es soll ein wichtiges Anliegen des Unterrichts sein, Eigenständigkeit in der Analyse und Interpretation von Textquellen, Bildern, Karten, Statistiken und Diagrammen zu entwickeln» (BGBl. II 133/2000; BGBl. II 134/2000). 2008 erfolgte erstmals die Implementierung der fachspezifischen Kompetenzorientierung (BGBl. II 290/2008) in Form des FUER-Modells (Körber, Schreiber & Schöner, 2007). Damit wurde vom Bildungsministerium das klare Ziel vorgegeben, dass die Lernenden im Zuge der Arbeit mit Quellen sowie Geschichtsdarstellungen historische Kompetenzen entwickeln. Auch die Reform der Reife- und Diplomprüfung (Pichler, 2020, 84–102) zielte auf die Implementierung der fachspezifischen Kompetenzorientierung und eine durchgängige Quellenorientierung ab (Bundesministerium für Bildung und Frauen, 2011). Die bis dahin üblichen Aufgabenformulierungen – vielfach ohne direkten Quellenbezug und auf die Abtes-

tung von historischem Faktenwissen abzielend, wie Philipp Mittnik (2014) bei der empirischen Untersuchung des Matura-Haupttermins 2013 an Allgemeinbildenden höheren Schulen in Wien nachweisen konnte – sind damit nicht mehr möglich.

Allerdings lassen die bisherigen empirischen Forschungen zum Geschichtsunterricht in Österreich Zweifel aufkommen, dass die Geschichtslehrkräfte ihren Unterricht rasch und vollständig an die neuen Vorgaben angepasst haben, auch wenn eine stetige Veränderung anzunehmen ist. Christian Pichler (2016, 26–27) konnte in seiner 2014 durchgeführten Befragung von 89 Geschichtslehrkräften, die an Gymnasien im Bundesland Kärnten unterrichten, unter anderem zeigen, dass diese die Arbeit mit Quellen und Geschichtsdarstellungen und einen gezielten Aufbau von Methodenkompetenz mehrheitlich als nicht zentral erachten, was mit fehlender Zeit begründet wurde.

Dies macht deutlich, dass die Grundvoraussetzung für einen auf die Förderung von fachspezifischen Kompetenzen ausgerichteten Geschichtsunterricht in Österreich lange Zeit nicht gegeben war. Die Studienergebnisse von Kipman und Kühberger (2019, 85) zeigen, dass österreichische Geschichtslehrkräfte ($n = 277$), wenn sie mit vorgegebenen Antworten befragt werden, durchaus ein Bewusstsein dafür zeigen, dass die Arbeit mit Quellen eine Bedeutung im Geschichtsunterricht haben sollte. Allerdings deuten die Ergebnisse dieser Untersuchung ebenso darauf hin, dass die Umsetzung im Unterricht mehrheitlich noch kein Standard ist (ebd., 120–127). Die Ausführungen von Roland Bernhard und Christoph Kühberger (2019) verdeutlichen darüber hinaus, dass das Kompetenzverständnis österreichischer Geschichtslehrkräfte als fachunspezifisch zu beschreiben ist; Claus Oberhauser (2018, 319–320) stellt in seiner Untersuchung zu Maturaprüfungen ebenfalls fest, dass die Geschichtslehrpersonen die theoretischen Grundlagen – konkret das FUER-Kompetenzmodell – nicht kennen und/oder eine Auseinandersetzung mit geschichtsdidaktischer Theorie grundsätzlich ablehnen.

3 Museen als Lernorte für den Geschichtsunterricht

Im Gegensatz zum Geschichtsunterricht ist der Stellenwert von Quellen für Museen trotz einiger Debatten im Zusammenhang mit den sich wandelnden gesellschaftlichen Erwartungen an Museen und dem Selbstverständnis der Institution (Thiemeyer, 2012; Schulze, 2017) international weitgehend unbestritten. Das

Narrativ einer Ausstellung wird zwar von verschiedenen Faktoren bestimmt, doch sind Originale der aktuellen ICOM-Museumsdefinition folgend konstitutiv für die Institution (International Council of Museums [ICOM], o. J.) und für das Lernen in dieser (Evans, Mull & Poling, 2010, 55).

Der geschichtsdidaktische Diskurs um den Lernort Museum schließt an die Grundüberlegungen des museumswissenschaftlichen an, indem auf die Besonderheiten der Institution im Vergleich zum Klassenzimmer verwiesen wird (Marcus, Levine & Grenier, 2012, 68). Olaf Hartung (2020, 26) zufolge ergeben sich Lernpotenziale insbesondere über diese beziehungsweise die «unmittelbare sinnliche Erfahrung» und für Thorsten Heese (2020, 259) stellt der Objektbezug das «Fundament historischen Lernens im Museum» dar. Folgt man von Borries (2009, 116) dann bieten die Originale mit ihrer Aura Möglichkeiten, die weit über jene von Schule hinausgehen, und auch Martin Schlutow (2012, 32) betont, dass die Authentizität der Exponate «das entscheidende Charakteristikum [ist], welches das Museum vom Geschichtsschulbuch oder anderen Medien, die sich mit Geschichte auseinandersetzen, unterscheidet». Somit ist das Museum ein Lernort für den Geschichtsunterricht, der sich zur Realisierung der Quellenorientierung besonders eignet, wenngleich diese auch hier durch entsprechende methodische Zugänge realisiert werden muss.

Das Betrachten von Objekten allein führt noch zu keinem historischen Lernen beziehungsweise zu keiner Ausbildung eines reflektierten Geschichtsbewusstseins (Hartung, 2015, 169); eine inhaltsorientierte Führung kann ebenso wie ein Lehrervortrag im Klassenzimmer dazu führen, dass Quellen bloß als Belege für ein historisches Narrativ herangezogen werden, statt diese als Ausgangspunkt einer Re-Konstruktion zu nutzen. Wie Roman Weindl (2019, 245–246) nachweisen konnte, wecken Originale auch nicht automatisch Interesse, wie dies etwa von Dietmar von Reeken (2014, 148) angenommen wurde. Die Möglichkeiten des Museums in Bezug auf das historische Lernen ergeben sich außerdem weniger über die vielzitierte «Aura» als durch den Prozess der Musealisierung (Heesen, 2015, 44), indem das Objekt zum Exponat wird. Um historisches Lernen im Museum zu ermöglichen, braucht es daher, wie Weindl (2019, 274) zu Recht betont,

«fundierte museumspädagogische Konzepte, die eine kritische Begegnung mit den in den Museen gesammelten, bewahrten und ausgestellten historischen Überresten ermöglicht, anstatt sie als reine <Beeindruckungsdinge> zum Gegenstand irrationaler, auratischer Verehrung zu machen.»

Aufgrund dieser besonderen Chancen, die für das historische Lernen in Museen angenommen werden, verwundert es nicht, dass sich die Geschichtsdidaktik auch schon der Frage zuwandte, inwiefern im Zuge von Museumsbesuchen historische Kompetenzen gefördert werden können, also welche «übertragbaren Problemlösefähigkeiten Geschichtsmuseen fördern können und sollen» (Hartung, 2020, 11) sowie in welcher Beziehung diese zu verschiedenen Kompetenzmodellen stehen (Popp & Schönemann, 2009). Angenommen wird, dass ein historischer Lernprozess im Museum nicht nur angeregt, sondern vollzogen und (vorläufig) abgeschlossen werden kann (Hasberg, 2009, 233).

Zum historischen Lernen von Kindern und Jugendlichen in historischen Ausstellungen und Museen wurden in den letzten Jahren verschiedene empirische Untersuchungen durchgeführt. Zu den Vorstellungen von Geschichtslehrkräften liegen Ergebnisse aus Polen und den USA vor. Die von Joanna Wojdon (2018, 98) befragten 19 polnischen Geschichtslehrpersonen nannten als wichtigste Vorteile von Museumsbesuchen: den Erwerb von mehr Wissen, Lernen auf eine attraktivere Art und Weise, außerhalb der Schule zu sein, das Kennenlernen des Museums als Institution sowie die Konfrontation des Schulbuchwissens mit Originalen. In Tiefeninterviews mit zwölf US-amerikanischen Lehrkräften wiesen alle Befragten darauf hin, dass Museumsbesuche die Vergangenheit lebendig machen und Ereignisse sowie Objekte der Vergangenheit mit Unterrichtsinhalten verknüpfen. Eine Lehrperson meinte beispielsweise:

«I think the greatest value [of museums] is for kids to see that it's real. To see artifacts, to see the material [...] once they've seen material it hits them in a way that a picture doesn't. It's a different kind of connection they make to the object.»
(Marcus, Levine & Grenier, 2012, 79)

In der Studie von Christian Kohler (2016, 117) zeigte sich, dass auch Jugendliche mit Museen Objekte assoziieren. Die Aussage «Wenn ich eine Ausstellung in einem historischen Museum besuche, betrachte ich am liebsten mir unbekannte und außergewöhnliche Dinge» erhielt unter den abgefragten Items zu den bisherigen Museumserfahrungen durchschnittlich die höchste Zustimmung. Zu ähnlichen Ergebnissen kam Berit Pleitner (2006, 97–98) auf der Grundlage von Interviews mit sechs Schüler*innen einer integrierten Gesamtschule. Bernd Wagners Studie (2013, 213) zufolge interessieren sich Kinder speziell für Objekte, die Alteritätserfahrungen ermöglichen, allerdings nicht völlig fremd erscheinen. In der Untersuchung von Hannah Röttele (2020) zeigte sich, dass die Aufmerk-

samkeit der Lernenden insbesondere auf Objekte gerichtet ist, die sie entweder als «bekannt» einstufen oder die eine gewisse Irritation auslösen. Darüber hinaus kann auf Basis ihrer Erkenntnisse geschlossen werden, dass Arbeitsaufträge eine Konkurrenz zur Objektwahrnehmung erzeugen, wenn diese auf Informationen abzielen, die unabhängig von den Exponaten gewonnen werden können (Röttele, 2017, 118–119). Dass die Arbeit mit Quellen im Museum bereits bei Kindern in der Primarstufe Prozesse des historischen Denkens initiieren kann, belegt die Studie von Megan Martinko und Jessica Luke (2018), in der 29 Kinder in Mitmachbereichen in Geschichtsmuseen videografiert wurden. Von den insgesamt 143 beobachtbaren Beispielen für historisches Denken erfolgten 49 Prozent mit Berücksichtigung einer Quelle.

In einem Widerspruch zu diesen feststellbaren Potentialen von Museen in Bezug auf die Initiierung historischen Lernens sind die Ergebnisse von Felicitas Iris Klingler (2018) zu sehen, die sich auf die Angebote der Museumspädagogik konzentrierte. Auf der Grundlage von Angebotsbeschreibungen, Interviews, ausgewählten Artikeln sowie vier videografierten Museumsbesuchen stellte sie fest, dass eine nähere Beschäftigung mit den Exponaten im Gegensatz zum historischen Wissen oft nicht im Zentrum steht. Dazu kommt der Anspruch mit möglichst vielen unterschiedlichen Materialien und Methoden abseits der Objekte zu arbeiten, womit Museen ihre Potentiale nicht ausschöpfen. Fehlende Objektbezüge bei musealen Vermittlungsprogrammen wurden auch in den Studien von Pandel (2006) und Dannemann (2010) kritisiert.

4 Erkenntnisinteresse und methodisches Vorgehen

Trotz der vielfältigen Studien zum historischen Lernen in Ausstellungen und Museen sind Forschungsdesiderate festzustellen: auf der einen Seite in Bezug auf die Formen des Arbeitens mit Objekten im Zuge von Museumsbesuchen im Rahmen des Geschichtsunterrichts und den diesbezüglichen Erwartungen von Geschichtslehrkräften und auf der anderen Seite zu Museumsbesuchen im Rahmen des Geschichtsunterrichts in Österreich generell vor dem Hintergrund der bisherigen Erkenntnisse zur Umsetzung der fachspezifischen Kompetenzorientierung. Insbesondere fällt auf, dass Museumsbesuche von Schulklassen in der Forschung meist isoliert vom Geschichtsunterricht betrachtet werden, wiewohl jegliches außerschulisches Lernen durch die Lehrkräfte und den schulischen

Kontext geprägt ist (Budde & Hummrich, 2016, 33). Obwohl es kritische Stimmen gibt, die betonen, dass Museen nicht gegründet wurden, um bei Schüler*innen Kompetenzen zu fördern (von Borries, 2009, 100), ist zu betonen, dass Ausstellungen und Museen curricular genutzt werden können und Stéphane Lévesque (2006, 46) zufolge auch sollen, auch wenn diese nicht an Lehrplänen ausgerichtet oder gar auf der Basis von Lehrplänen entstanden sind (Heese, 2014, 13). In der Folge wird daher untersucht, (1) welche Bedeutung Geschichtslehrkräfte den musealen Objekten bei Museumsbesuchen im Rahmen des Geschichtsunterrichts zuschreiben beziehungsweise welche Erwartungen sie in Bezug auf die Arbeit mit Objekten äußern und (2) auf welche Art und Weise im Zuge von solchen mit den Exponaten gearbeitet wird.

Im Rahmen des Projekts «Historisches Lernen zwischen Schule und Museum» kam eine Methodentriangulation zur Anwendung, deren Möglichkeiten in der geschichtsdidaktischen Forschung seit kurzem intensiv diskutiert werden (vgl. Kelle, Kühberger & Bernhard, 2019, 14–17). Diese erlaubte unterschiedliche Perspektiven auf den Forschungsgegenstand (vgl. Flick, 2011, 12). Konkret wurden Experteninterviews mit den Leiter*innen der Kulturvermittlungsabteilungen¹ der österreichischen Landesmuseen ($n = 9$) und Geschichtslehrkräften ($n = 85$) sowie Beobachtungen von Schulklassen ($n = 11$) vor, während und nach Museumsbesuchen durchgeführt. Außerdem wurden nach den beobachteten Museumsbesuchen Kurztexte der Lernenden ($n = 202$) erhoben. Die folgenden Ausführungen beziehen sich jedoch ausschließlich auf die Interviews mit den Geschichtslehrkräften sowie auf die Beobachtungen während der Museumsbesuche.

Bei der Auswahl der Lehrpersonen wurde auf Vielfalt geachtet, um eine theoretische Sättigung zu erreichen (Bogner, Littig & Menz, 2014, 37). Interviewt wurden daher Geschichtslehrkräfte aus allen österreichischen Bundesländern, aus städtischen und ländlichen Regionen, von verschiedenen Schultypen sowie Personen mit unterschiedlichen Dienstaltern. Auch «ungeprüfte» Lehrkräfte sind Teil des Samples, also Personen, die Geschichte unterrichten, dieses Fach aber nicht studiert haben. Wiewohl darauf geachtet wurde, das Feld in möglichst vielen Aspekten zu erfassen, handelt es sich um keine repräsentative Stichprobe, weshalb keine quantifizierenden Aussagen in Bezug auf die Grundgesamtheit der

1 Im Gegensatz zu Deutschland hat sich in Österreich der Ansatz der Kulturvermittlung durchgesetzt. Zum Selbstverständnis vgl. Österreichischer Verband der KulturvermittlerInnen (o. J.).

österreichischen Geschichtslehrkräfte möglich sind. Das Sampling ist am ehesten als Schneeballverfahren zu beschreiben (Krippendorff, 2019, 121 f.), das mit weiteren Samplingmethoden kombiniert wurde (Przyborski & Wohlrab-Sahr, 2014, 185).

Die Interviews² wurden zwischen Januar 2018 und Dezember 2019 teilstandardisiert durchgeführt; die durchschnittliche Interviewdauer beträgt 56 Minuten. Mit dem eingesetzten Leitfaden wurde das Ziel verfolgt, durch kurze Fragen relativ freie Erzählungen der Interviewten zu erreichen (Gläser & Laudel, 2010, 174–177). Bei der Formulierung wurde möglichst darauf geachtet, dass die Geschichtslehrkräfte keine Rückschlüsse daraus ziehen konnten, welche Antworten sozial erwünscht sind. Daher wurde hauptsächlich nach praktischen Erfahrungen gefragt und eine Abtestung von geschichtsdidaktischem Wissen vermieden. Zentral für die Interviewdurchführung war daher aktives Zuhören: Die persönlichen Sichtweisen der Geschichtslehrkräfte wurden nicht bewertet, auch wenn sie aus Sicht der Autorin problematisch erschienen (Helfferich, 2011, 91). Das methodische Vorgehen hat selbstverständlich den Nachteil, dass keine quantifizierenden Aussagen möglich sind. Allerdings erlaubt die gewählte Herangehensweise einen genaueren Blick auf die sogenannten *beliefs* der Lehrkräfte (Nitsche, 2019), also auf ihre berufsbezogenen Ansichten, Meinungen und Überzeugungen.

Die Auswertung der wörtlich (aber nicht lautsprachlich) transkribierten Interviews erfolgte mit der inhaltlich strukturierenden qualitativen Inhaltsanalyse. Diese ermöglicht die Identifizierung von Themen und Subthemen, deren Systematisierung und Analyse der wechselseitigen Relationen (Kuckartz, 2018, 123). Das Kategoriensystem wurde hauptsächlich – wenn auch nicht ausschließlich (Ruona, 2005, 238) – induktiv entwickelt. Dieses Vorgehen ermöglicht *«research findings to emerge from the frequent, dominant, or significant themes inherent in raw data, without the restraints imposed by structured methodologies»* (Thomas, 2006, 238). In die Kategorienbildung flossen gewisse Vorannahmen auf Basis der bisherigen Forschungen mit ein. So war beispielsweise anzunehmen, dass verschiedene strukturelle Rahmenbedingungen Museumsbesuche erschweren, woraus sich der Subcode *«Probleme – Hinderungsgründe»* zum Hauptcode *«Mu-*

2 Die Interviews wurden anonymisiert und werden im Folgenden im Stil *«Schultyp_Bundesland_Nr.»* angegeben. *«AHS_Wien_1»* steht also für das erste Interview, das in einer Allgemeinbildenden höheren Schule (AHS) in Wien geführt wurde. BHS steht für Berufsbildende höhere Schulen und NMS für Neue Mittelschulen.

seumsbesuche» ergab; die genannten Hinderungsgründe (Angebote fehlen, Erreichbarkeit, Finanzen etc.) wurden in einer weiteren Hierarchieebene erfasst. Das im Programm MAXQDA erstellte Kategoriensystem wurde schließlich anhand von rund 12 Prozent des Materials (Lombard, Snyder-Duch & Bracken, 2002, 601) auf Intercoderreliabilität (Brennan & Prediger, 1981) geprüft, die im Mittel den akzeptablen Wert von $\kappa_n = 0.79$ ergab.

Die Beobachtungen³ ($n = 11$) hatten das Ziel, eine größere Nähe zum beforchten Gegenstand zu erreichen (Robson, 2008, 310) sowie «die fachspezifischen Wahrnehmungen des Forschungsgegenstandes» zu weiten und zu vertiefen (Kühberger, 2021, 54). Beobachtungen sind auf verschiedene Arten möglich (Döring & Bortz, 2016, 330). Jene im Rahmen des Projekts «Historisches Lernen zwischen Schule und Museum» sind als direkte, nicht-teilnehmende, offene und teilstrukturierte Fremdbeobachtungen im Feld zu beschreiben. Wesentlich war die Festlegung von klar umrissenen Beobachtungszielen im Vorfeld, für die grobe Kategorien festgelegt wurden, ohne die Offenheit zu sehr einzuschränken (Lamnek & Krell, 2016, 536).

Unvermeidlich ist, dass solche direkte, nicht-teilnehmende Beobachtungen eine zweifache Selektion bedeuten: einerseits durch die begrenzte Wahrnehmungsfähigkeit und andererseits durch die Entscheidung, einen beobachteten Aspekt als relevant oder nicht relevant einzustufen und folglich in die Feldnotizen aufzunehmen oder nicht (Reh, 2012, 119). Außerdem sind Wahrnehmungsfehler (Döring & Bortz, 2016, 331–332) nicht auszuschließen. Um eine Prüfung der Güte der Beobachtungen zumindest ansatzweise sicherzustellen, wurden zwei Schulstunden und ein Museumsbesuch neben der Autorin von einer zweiten Person⁴ beobachtet.

Die Herstellung einer elektronischen Reinschrift der Beobachtungsprotokolle (Breidenstein, Hirschauer, Kalthoff & Nieswand, 2013, 94–95) erfolgte so zeitnah wie möglich zur Beobachtung; in diese flossen auch die nach den Beobachtungen erstellten Gedankenprotokolle ein, die halfen, das Geschehen möglichst detailreich festzuhalten. Klar ist jedoch, dass sich bei Beobachtungen «Da-

3 Die im Zuge der Beobachtungen erhobenen Daten wurden anonymisiert; zu erkennen sind aufgrund der Zitierweise wie bei den Interviews das Bundesland und der Schultyp. AHS_Wien_A_MB steht beispielsweise für das Beobachtungsprotokoll zum Museumsbesuch der ersten AHS-Klasse aus Wien, die beobachtet wurde.

4 Dabei handelt es sich um Jasmin Fischer, die am Institut für Fachdidaktik der Universität Innsbruck als studentische Mitarbeiterin tätig war.

tenerhebung und -analyse nicht scharf trennen lassen», wie Georg Breidenstein (2012, 32) zu Recht betont.

Die Daten wurden wie die Interviews mit der inhaltlich strukturierenden qualitativen Inhaltsanalyse mithilfe von MAXQDA ausgewertet, wobei die Kategorien teils induktiv und teils deduktiv gebildet wurden. Die Güte des Kategoriensystems wurde anhand der Codierung von vier Beobachtungsprotokollen von Schulstunden (zwei Stunden vor dem Museumsbesuch und zwei Stunden nach dem Museumsbesuch) sowie von zwei Museumsbesuchen geprüft, was 18,18 Prozent des Datensatzes entspricht. Die Intercoder-Reliabilität ergab im Mittel den guten Wert von $\kappa_n = 0.86$.

5 Begegnung mit Originalen bei Museumsbesuchen

5.1 Beliefs österreichischer Geschichtslehrkräfte

Zunächst ist festzustellen, dass die Begegnung mit Originalen für viele interviewte Geschichtslehrkräfte ein zentrales Ziel eines Museumsbesuchs darstellt. NMS_Wien_3 (Abs. 31) meint beispielsweise: «Also dass sie sich da in einer besonderen Weise mit Exponaten beschäftigen können, die wir ihnen nicht bieten können. Also oft auch mit einem Arbeitsplan, den sie dann erfüllen können». Auch für die Lehrkraft AHS_Wien_4 (Abs. 61) sind Originale ein wesentlicher Grund für Museumsbesuche: «Aber dieses wirkliche Anschauen, zum Beispiel im KHM [Kunsthistorischen Museum] die Mumien und so, ist ja was ganz anderes, wenn man sie wirklich sieht. Und dann eben auch ein Gefühl bekommt, wie das auch ausschaut. Und dann ein Bild kann das einfach nie transportieren.»

Allerdings wird von den interviewten Geschichtslehrkräften keine richtiggehende Analyse der Exponate erwartet – vielmehr geht es diesen um eine Veranschaulichung der im Geschichtsunterricht besprochenen Entwicklungen und Ereignisse mithilfe von Objekten, wie beispielsweise BHS_NÖ_1 (Abs. 28) ausführt: «Ja, einfach Unterstützung von dem, was wir im Unterricht gemacht haben, dass man das auch sieht, dass man hier Objekte sieht, die das unterstreichen [...]». Die Schüler*innen sollen nach Ansicht der interviewten Geschichtslehrkräfte bestimmte Objekte einmal im Original gesehen haben, wie unter anderem AHS_Wien_17 (Abs. 32) meint. AHS_Tirol_1 (Abs. 51) betont auch: «[D]as Objekt als solches hat natürlich so seine Wirkmächtigkeit.» BHS_Wien_5 (Abs. 52) führt hierzu aus:

«Im Klassenzimmer kann ich zwar mit Bildern arbeiten und mit Videos, aber dann haben sie etwas, was sie vor sich sehen, teilweise Exponate vielleicht auch anfassen können, also das Haptische noch mal ansprechen. Aber einfach auch dieses direkt vor sich sehen und das in 3D sehen sozusagen, ist noch einmal ganz, ganz was anderes, als einfach nur ein zweidimensionales Bild zu sehen. Das berührt dann viele noch anders und manche fangen dann ja an und lesen wirklich, was auf der Tafel steht, oder hören sich das Audio dazu an, damit sie dann zu den Exponaten genau wissen, um was geht es da.»

Sehr ähnlich argumentiert beispielsweise AHS_Sbg_2 (Abs. 110):

«Das heißt, angenommen, es geht um gegenständliche Quellen, das kann ich im Unterricht dann nur anhand von einem Foto zeigen oder so. Aber das dann wirklich zu sehen und betrachten zu können und da etwas bemerken zu können, geht eben nur im Museum, wenn auch meistens eingeschränkt. Und ... es ist, kann, also es ist nicht das Ziel, aber es ist doch auch unbestreitbar, dass ... dass das eine gewisse Wirkung hat auch auf die Schüler und Schülerinnen dann zu sehen quasi, wie das wirklich ausgeschaut hat diese Quelle, ja.»

5.2 Arbeit mit Museumsobjekten in der Praxis

Im Rahmen des Projekts konnten bei den Museumsbesuchen der beobachteten Schulklassen verschiedene museale Vermittlungsformen beobachtet werden:

- Drei Klassen (AHS_Tirol_B, AHS_Wien_A, AHS_Wien_D) nahmen an einem Vermittlungsprogramm teil, das als (dialogische) Führung einzuordnen ist.
- Drei Klassen (AHS_Wien_B, AHS_Wien_C, NMS_Wien_A) nahmen an einem Vermittlungsprogramm teil, das aus einem längeren Workshopteil ohne direkten Bezug zur Ausstellung und einer (dialogischen) Führung bestand.
- Zwei Klassen (AHS_Tirol_A, NMS_Tirol_A) nahmen an einem Vermittlungsprogramm teil, das aus einem kurzen Workshopteil und einer Führung durch die Lernenden bestand.
- Eine Klasse (AHS_NÖ_A) nahm an einer (dialogischen) Führung mit einer kurzen Übung teil und erhielt danach ein Arbeitsblatt von der Lehrkraft.
- Eine Klasse (AHS_Tirol_C) erkundete das Museum mit einem Arbeitsblatt der Lehrkraft.
- Eine Klasse (NMS_Wien_B) erkundete das Museum auf Basis eines kurzen, mündlichen Arbeitsauftrags der Lehrkraft.

In allen Fällen konnte eine Beschäftigung mit Exponaten beobachtet werden, doch erfolgte diese durchaus unterschiedlich. Konkret lassen sich drei Varianten unterscheiden.

5.2.1 Variante 1: Keine Objektanalyse

Diese Variante der Beschäftigung mit Quellen im Museum zeichnet sich dadurch aus, dass die Lernenden zwar zahlreiche Objekte sahen und auch vorgestellt bekamen, doch die Kulturvermittlerin beziehungsweise der Kulturvermittler die Lernenden nicht dazu animierte, zu eigenständigen Erkenntnissen zu gelangen. Stattdessen teilte die Kulturvermittlerin beziehungsweise der Kulturvermittler ihr beziehungsweise sein Wissen über eine Quelle mit der Klasse und leitete dann zum nächsten Objekt beziehungsweise zum nächsten Teil des Vermittlungsprogramms über. Dies zeigte sich bei allen drei beobachteten Workshops mit (dialogischen) Führungen, zumal sich der «Dialog» rein auf Faktenwissen bezog. So war beispielsweise beim Museumsbesuch der Klasse AHS_Wien_B (Abs. 11) zu beobachten:

«GD [Der Guide]⁵ verweist auf einen ‹Hängegleiter›, mit dem ein Tscheche über den Eisernen Vorhang geflogen ist. GD erklärt, wie der Tscheche diesen gebaut und den Motor von einem Trabi verbaut hat; er lag oben auf dem Flieger und hat von dort gesteuert; das war sehr waghalsig, da er es zuvor nie testen konnte; der Tscheche ist dann bei Nacht geflogen, unter dem Radar, und hat es erfolgreich nach Österreich geschafft; er ist im Herbst 1988 geflohen; er konnte natürlich nicht wissen, dass 1989 der Eiserner Vorhang abgebaut wird.»

Die Schüler*innen wurden damit zwar eingeladen, das Objekt zu bestaunen, doch erfolgte keine weitere Analyse oder Interpretation. Die Kulturvermittlerin beziehungsweise der Kulturvermittler ging in der Folge dazu über, mit den Lernenden zu besprechen, was mit Personen passieren konnte, deren Flucht scheiterte. Das eventuell vorhandene Potenzial zur Irritation, das historische Denken initiieren könnte, indem die Schüler*innen Fragen an das Objekt formulieren, wurde somit nicht genutzt.

Keine Aufforderung zu einer näheren Objektanalyse war auch in jener Klasse (NMS_Wien_B) zu beobachten, die das Museum ohne Vermittlungsprogramm

5 Die Abkürzung wurde aus praktischen Erwägungen gewählt, gemeint sind die Kulturvermittler*innen.

und auch ohne konkrete Arbeitsaufträge durch die Lehrkraft besichtigte. Die Lernenden wurden in diesem Fall lediglich dazu aufgefordert, sich im ganzen Museum umzusehen und die Ausstellungstexte zu lesen (Abs. 3). Die meisten Lernenden beschäftigten sich in der Folge nicht näher mit den Objekten. Manche Schüler*innen waren allerdings bei Ansätzen zu eigenständigen Objektanalysen (Variante 3) zu beobachten, wenngleich diese recht einfach gehalten waren und sich auf vormalige Verwendungszwecke von Exponaten beschränkten.

5.2.2 Variante 2: Einbeziehung der Lernenden in die Objektanalyse

Im Fall der beobachteten Vermittlungsprogramme, die durchgängig als (dialogische) Führung einzuordnen sind, versuchten die Kulturvermittlerinnen oder Kulturvermittler, die Schüler*innen – im Gegensatz zu Variante 1 – ab und zu (wenngleich nicht durchgängig) in die Analyse einzubeziehen. Dies soll anhand eines Beispiels vom Museumsbesuch der Klasse AHS_Tirol_B (Abs. 9) veranschaulicht werden:

*«GD fragt in der Folge, ob jemand die Überschrift des ausgestellten Buches lesen kann. Ein/e S [Schülerin bzw. Schüler] meint, «verderben» erkennen zu können, was GD bestätigt und «ein» ergänzt. Ein/e andere/r S ergänzt: «vier Ding» und «Bergwerk». GD fasst zusammen: «Vier Ding verderben ein Bergwerk». GD fordert die Klasse in der Folge auf, sich die Zeichnungen anzusehen und daraus abzuleiten, was mit den vier Dingen gemeint ist. Daraufhin nennen die S/S [Schüler*innen] Krieg, Tod, Handel und Faulheit. GD führt daraufhin aus, dass es sich beim Schwazer Bergbuch um ein wichtiges Dokument zur Bergwerksgeschichte handelt und davon nur einige Exemplare existieren. GD erklärt, dass es sich um ein Original handelt, weshalb es in dem Bereich so dunkel ist, damit das Papier nicht ausbleicht.»*

Die Aktivität der Lernenden beschränkte sich in diesem Fall darauf, die Schrift zu entziffern und zu den vier Abbildungen auf der aufgeschlagenen Buchseite je ein passendes Schlagwort zum Inhalt zu nennen. Eine weitere Analyse oder Interpretation blieb aus. Die Kulturvermittlerin oder der Kulturvermittler teilte rasch ihr oder sein Wissen über das Exponat mit den Lernenden und führte die Gruppe in der Folge zum nächsten. Eine Aufforderung, auf Basis der Quelle Geschichte zu re-konstruieren beziehungsweise eine historische Narration zu entwickeln, erfolgte nicht. Im Vergleich zu Variante 1 konnten sich die Lernenden aber zumindest teilweise selbst mit dem Exponat beschäftigen. Ähnliche Auffor-

derungen zu Beschreibungen waren bei allen drei (dialogischen) Führungen zu beobachten.

5.2.3 Variante 3: Eigenständiges Erforschen von Objekten

Eine eigenständige Erforschung von Quellen im Museum durch die Lernenden war bei den beiden Workshops mit Führung durch die Lernenden zu beobachten sowie bei den Arbeitsaufträgen der Lehrpersonen. Das Ergebnis einer solchen Objektanalyse zeigt beispielsweise der folgende Ausschnitt aus dem Beobachtungsprotokoll zum Museumsbesuch der Klasse AHS_Tirol_A (Abs. 12–13):

Die Gruppe begibt sich daraufhin zum nächsten Objekt; die S/S bitten die anderen, sich so hinzustellen, dass die ausgewählten Objekte (Malereien von Dörfern und ein Reisebericht) gut zu sehen sind.

13.00: Ein/e S erklärt, dass sie sich mit dem Begriff ›katholisch‹ beschäftigt haben; sie erklären, dass man anhand der Objekte gut erkennen kann, dass den Touristen scheinbar aufgefallen ist, dass Tirol sehr katholisch ist, da auf den Bildern viele Kirchen dargestellt sind [...]. Die/der S erklärt, dass es auch im Reisebericht um Kirchen geht; in der Zeit gab es der/dem S zufolge viele Reiseberichte, in denen auf Kirchen eingegangen wurde. GD fragt nach, um welche Zeit es geht. Die/der S meint, dass es um das 18./19. Jahrhundert geht. GD präzisiert, dass die Bilder aus der 1. Hälfte des 19. Jahrhunderts bis 1900 stammen und dass es zu der Zeit noch kaum Fotografie gab. Die/der zweite S des Teams führt aus, dass die Darstellung der Kirchen auf den Bildern sicher dazu geführt hat, dass sich die Touristen dafür interessieren. GD weist noch daraufhin, dass hier keine Städte, sondern ländliche Regionen gezeigt werden.»

In diesem Fall beschrieben, analysierten und interpretierten die Lernenden in Kleingruppen die Bild- und Textquellen (von wenigen Hilfestellungen der Kulturvermittlerin oder des Kulturvermittlers abgesehen) eigenständig und präsentierten ihre Ergebnisse der gesamten Klasse, womit Re-Konstruktionsprozesse erkennbar wurden, wie sie in den österreichischen Lehrplänen beschrieben sind. Die Kulturvermittlerin oder der Kulturvermittler beschränkte sich – im Gegensatz zu Variante 2 – darauf, gezielte Nachfragen zu stellen und historische Details zu ergänzen.

Die in zwei Klassen eingesetzten Arbeitsblätter von Lehrpersonen forderten von den Lernenden Objektanalysen unter bestimmten vorgegebenen Aspekten. Die Arbeitsaufträge ähnelten teilweise jenen aus Schulbüchern, deckten alle drei

Anforderungsbereiche ab und können als kompetenzorientiert bezeichnet werden. So wurden die Lernenden der Klasse AHS_NÖ_A unter anderem aufgefordert: «Beurteile, ob mit der Auswahl des Objekts beziehungsweise der Anordnung im Museum ein bestimmter Zweck erreicht werden soll.» Die Lernenden der Klasse AHS_Tirol_C erhielten unter anderem die folgende Aufgabenstellung: «Was könnte ein Historiker/eine Historikerin auf dem Bild mit der *zweitältesten Ansicht von Innsbruck* über das damalige Aussehen der Stadt herausfinden?»

6 Fazit

Museumsbesuche im Rahmen des Geschichtsunterrichts stellen – so die Grundannahme der Studie – eine spezifische Form von Besuchen dieser Institution dar. Ausstellungen und Museen werden dabei für schulische Zielsetzungen genutzt. Die Ergebnisse sind somit im Kontext der bisherigen Erkenntnisse zur Quellenorientierung im österreichischen Geschichtsunterricht und der Umsetzung der in den Lehrplänen verankerten fachspezifischen Kompetenzorientierung zu sehen.

Es ist daher nicht verwunderlich, dass die befragten österreichischen Geschichtslehrkräfte zwar vielfach auf den zentralen Wert von Objekten verweisen, womit sie der geschichtsdidaktischen Theorie zu Museen weitgehend folgen, jedoch keine Objektanalyse als Ausgangspunkt einer Re-Konstruktion erwarten. Dass im Zuge von Museumsbesuchen die im Lehrplan der Sekundarstufe I geforderte «Eigenständigkeit im kritischen Umgang mit historischen Quellen zum Aufbau von Vorstellungen und Erzählungen über die Vergangenheit» (BGBl. II 113/2016) gefördert werden könnte, spielt für die interviewten Lehrkräften keine Rolle. Vielmehr erwarten sich die meisten Lehrkräfte ein reines Betrachten, teilweise auch Bestaunen von Exponaten.

In der musealen Vermittlung lassen sich zwar Beispiele für quellenorientierte Ansätze finden, doch fungieren Quellen auch im Museum oft nur als Belege für mittels Erzählung präsentiertes Faktenwissen. Besonders auffällig ist hierbei, dass Lernende bei Vermittlungsprogrammen, die (fast) ausschließlich als (dialogische) Führung zu klassifizieren sind (AHS_Tirol_B, AHS_Wien_A, AHS_Wien_D), zumindest ein wenig in die Objektanalyse einbezogen werden, während dies bei jenen Vermittlungsprogrammen, bei denen auch ein längerer Workshop stattfand (AHS_Wien_B, AHS_Wien_C, NMS_Wien_A), nicht erfolgte.

Nicht ganz eindeutig zuordenbar ist jene Klasse (NMS_Wien_B), die sich im Museum frei bewegte und nur einen mündlichen Arbeitsauftrag der Lehrkraft

erhielt. Hierbei waren Schüler*innen zu beobachten, die sich Gedanken über die Bedeutung beziehungsweise die historische Nutzung von Gegenständen machten; gleichzeitig nutzten einige Lernende den Museumsbesuch auch für Aktivitäten abseits der Objekte und des Ausstellungsthemas (Verstecken spielen).

Die im Rahmen von Workshops von Kulturvermittler*innen initiierten Führungen durch die Lernenden (AHS_Tirol_A, NMS_Tirol_A) sowie die beiden Arbeitsblätter, die von Lehrpersonen für die Museumsbesuche ihrer Klassen (AHS_NÖ_A, AHS_Tirol_C) entwickelt wurden, zeigen allerdings deutlich die Potenziale von Museen für die Initiierung historischen Denkens. Diese Museumsbesuche dienten klar der Förderung von fachspezifischen Kompetenzen, wie sie in den österreichischen Lehrplänen verankert sind.

Für eine Einschätzung zur Verbreitung der unterschiedlichen Varianten des Objektbezugs bei Museumsbesuchen im Rahmen des Geschichtsunterrichts wären weitere Studien mit einem deutlich größeren Sample nötig. Die Untersuchung hat allerdings deutlich gezeigt, dass mit entsprechenden methodischen Ansätzen Quellenorientierung mit dem Ziel der Initiierung historischer Denkprozesse am Lernort Museum umgesetzt werden kann, wobei auffällt, dass dies in je zwei Fällen von Kulturvermittler*innen sowie von Lehrkräften angeleitet wurde. Dies erstaunt im Vergleich zu den Interviews, wobei jedoch offen bleiben muss, inwiefern es sich bei den Arbeitsblättern um Einzelfälle handelt, die vielleicht auch deshalb so positiv hervorstechen, weil den Lehrkräften wohl bewusst war, dass dieser Museumsbesuch von einer Forscherin von der Universität beobachtet wird.

Literatur

- Bernhard, Roland & Kühberger, Christoph (2019). Domänen(un)spezifisch – Empirische Befunde zum Kompetenzverständnis von Geschichtslehrpersonen. In Monika Waldis & Béatrice Ziegler (Hrsg.), *Forschungswerkstatt Geschichtsdidaktik 17. Beiträge zur Tagung «geschichtsdidaktik empirisch 17»* (S. 119–130). Bern: hep.
- Bogner, Alexander/Littig, Beate & Menz, Wolfgang (2014). *Interviews mit Experten: Eine praxisorientierte Einführung*. Wiesbaden: Springer VS.
- Brait, Andrea (2020). *Museumsbesuche im Geschichtsunterricht. Eine Studie zum historischen Lernen im Zuge von Besuchen der österreichischen Landesmuseen*. Habilitation. Universität Innsbruck, Innsbruck.
- Breidenstein, Georg (2012). Ethnographisches Beobachten. In Heike de Boer & Sabine Reh (Hrsg.), *Beobachtung in der Schule – Beobachten lernen* (S. 27–44). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Breidenstein, Georg/Hirschauer, Stefan/Kalthoff, Herbert & Nieswand, Boris (2013). *Ethnographie: Die Praxis der Feldforschung*. München: UTB.
- Brennan, Robert L. & Prediger, Dale J. (1981). Coefficient kappa: some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41, 687–699.
- Budde, Jürgen & Hummrich, Merle (2016). Die Bedeutung außerschulischer Lernorte im Kontext der Schule – eine erziehungswissenschaftliche Perspektive. In Jan Erhorn & Jürgen Schwier (Hrsg.), *Pädagogik außerschulischer Lernorte: Eine interdisziplinäre Annäherung* (S. 29–52). Bielefeld: transcript.
- Bundesministerium für Bildung und Frauen (Hrsg.) (2011). *Die kompetenzorientierte mündliche Reifeprüfung: Geschichte, Richtlinien und Beispiele für Themenpool und Prüfungsaufgaben*. Wien: BMBF.
- Dannemann, Sandra (2010). *Aktivblätter auf dem Prüfstand: Eine Evaluationsstudie zu einem museumspädagogischen Vermittlungsangebot für Schulen*. Saarbrücken: VDM.
- Döring, Nicola & Bortz, Jürgen (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin: Springer.
- Evans, Margaret E./Mull, Melinda S. & Poling, Devereaux A. (2010). The authentic object? A child's-eye view. In Scott G. Paris (Hrsg.), *Perspectives on object-centered learning in museums* (S. 55–77). New York: Routledge.
- Flick, Uwe (2011). *Triangulation: Eine Einführung* (3. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gläser, Jochen & Laudel, Grit (2010). *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen* (4. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hartung, Olaf (2015). Aktuelle Trends in der Museumsdidaktik und ihre Bedeutung für das historische Lernen. In Vadim Oswalt & Hans-Jürgen Pandel (Hrsg.), *Geschichtskultur: Die Anwesenheit von Vergangenheit in der Gegenwart* (S. 153–173). Schwalbach/Ts.: Wochenschau.
- Hartung, Olaf (2020). *Museen und Geschichtsunterricht*. Stuttgart: Kohlhammer.
- Hasberg, Wolfgang (2009). Vermittlung geschichtskultureller Kompetenzen in historischen Ausstellungen. In Susanne Popp & Bernd Schönemann (Hrsg.), *Historische Kompetenzen und Museen* (S. 211–236). Idstein: Schulz-Kirchner.
- Heese, Thorsten (2014). Außerschulische Lernorte im Geschichtsunterricht: Das Museum. In Bärbel Kuhn/Susanne Popp/Jutta Schumann & Astrid Windus (Hrsg.), *Geschichte erfahren im Museum* (S. 13–21). St. Ingbert: Werner J. Röhrig.
- Heese, Thorsten (2020). «3 D» – Der historische Lernort Museum zwischen Authentizität und Virtual Reality. In Sebastian Barsch & Jörg van Norden (Hrsg.), *Historisches Lernen und Materielle Kultur: Von Dingen und Objekten in der Geschichtsdidaktik* (S. 253–264). Bielefeld: transcript.
- Heesen, Anke te (2015). Exponat. In Heike Gfrereis/Thomas Thiemeyer & Bernhard Tschofen (Hrsg.), *Museen verstehen: Begriffe der Theorie und Praxis* (S. 33–44). Göttingen: Wallstein.
- Helfferich, Cornelia (2011). *Die Qualität qualitativer Daten: Manual für die Durchführung qualitativer Interviews* (4. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.

- International Council of Museums (o. J.). Museum definition. <https://icom.museum/en/resources/standards-guidelines/museum-definition/> (20.05.2020).
- Kelle, Udo/Kühberger, Christoph & Bernhard, Roland (2019). How to use mixed-methods and triangulation designs: An introduction to history education research. *History Education Research Journal*, 16(1), 5–23. <https://doi.org/10.18546/HERJ.16.1.02>.
- Kipman, Ulrike & Kühberger, Christoph (2019). *Einsatz und Nutzung des Geschichtsschulbuches: Eine Large-Scale-Untersuchung bei Schülern und Lehrern*. Wiesbaden: Springer.
- Klingler, Felicitas (2018). *Lernort Museum. Eine empirische Untersuchung der Gestaltung museumspädagogischer Angebote für Schulklassen*. Dissertation. Georg-August-Universität Göttingen, Göttingen.
- Kohler, Christian (2016). *Schülvorstellungen über die Präsentation von Geschichte im Museum: Eine empirische Studie zum historischen Lernen im Museum*. Berlin: Lit.
- Körper, Andreas/Schreiber, Waltraud & Schöner, Alexander (Hrsg.) (2007). *Kompetenzen historischen Denkens: Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik*. Neuried: ars una.
- Krippendorff, Klaus (2019). *Content analysis: An introduction to its methodology* (4. Aufl.). Thousand Oaks: SAGE.
- Kuckartz, Udo (2018). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung* (4. Aufl.). Weinheim, Basel: Beltz Juventa.
- Kühberger, Christoph (2021). Geschichtsdidaktik im Ruderboot? Ethnographische Zugänge für die empirische Forschung. In Christoph Kühberger (Hrsg.), *Ethnographie und Geschichtsdidaktik* (S. 11–63). Frankfurt a. M.: Wochenschau.
- Lamnek, Siegfried & Krell, Claudia (2016). *Qualitative Sozialforschung* (6. Aufl.). Weinheim: Beltz.
- Lévesque, Stéphane (2006). Integrating museum education and school history: Illustrations from the RCR Museum and London Museum of Archaeology. *History Education Research Journal*, 6(1), 40–47.
- Lombard, Matthew/Snyder-Duch, Jennifer & Bracken, Cheryl C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, 28(4), 587–604.
- Marcus, Alan S./Levine, Thomas H. & Grenier, Robin S. (2012). How secondary history teachers use and think about museums: Current practices and untapped promise for promoting historical understanding. *Theory & Research in Social Education*, 40(1), 66–97.
- Martinko, Megan & Luke, Jessica (2018). «They ate your laundry!»: Historical thinking in young history museum visitors. *Journal of Museum Education*, 43(3), 245–259.
- Mittnik, Philipp (2014). Zentrale Themen des Geschichtsunterrichts in Österreich: Analyse der Reifeprüfungsaufgaben am Wiener AHS aus dem Fach Geschichte und Sozialkunde/ Politische Bildung. Eine empirische Erhebung. *Historische Sozialkunde*, 44(4), 26–37.
- Nitsche, Martin (2019). *Beliefs von Geschichtslehrpersonen – eine Triangulationsstudie*. Bern: hep. <https://doi.org/10.36933/9783035516005>.
- Oberhauser, Claus (2018). «[I]ch sehe, dass es mehr mit dem zu tun hat, was Geschichte eigentlich ist»: Tiroler Gymnasiallehrpersonen im Fach Geschichte und Sozialkunde/ Politische Bildung und ihre Einstellung zur Neuen Mündlichen Reifeprüfung. In Waltraud Schreiber/Béatrice Ziegler & Christoph Kühberger (Hrsg.), *Geschichtsdidaktik*

- scher Zwischenhalt: Beiträge aus der Tagung «Kompetent machen für ein Leben in, mit und durch Geschichte» in Eichstätt vom November 2017 (S. 317–322). Münster, New York: Waxmann.
- Österreichischer Verband der KulturvermittlerInnen (o.J.). Kulturvermittlung. <http://www.kulturvermittlerinnen.at/kulturvermittlung/> (03.02.2020).
- Pandel, Hans-Jürgen (2000). *Quelleninterpretation: Die schriftliche Quelle im Geschichtsunterricht*. Schwalbach/Ts.: Wochenschau.
- Pandel, Hans-Jürgen (2006). Museumspädagogische Materialien in der Geschichtskultur. *Zeitschrift für Geschichtsdidaktik*, 5, 109–118.
- Pandel, Hans-Jürgen (2013). *Geschichtsdidaktik: Eine Theorie für die Praxis*. Schwalbach/Ts.: Wochenschau.
- Pichler, Christian (2016). Kompetenzorientierter Geschichtsunterricht und fachspezifisches Professionsverständnis, ein Dilemma. *Zeitschrift für Didaktik der Gesellschaftswissenschaften*, 7(2), 13–31.
- Pichler, Christian (2020). *Geschichtsmatura: Eine empirische Untersuchung zum kompetenzorientierten Prüfungsmodus*. Innsbruck: StudienVerlag.
- Pleitner, Berit (2006). «Da kann man so viel lernen, gerade für junge Leute ...»: Überlegungen zum Verhältnis von Jugendlichen und Museen. *Zeitschrift für Geschichtsdidaktik*, 5, 93–108.
- Popp, Susanne & Schönemann, Bernd (2009) (Hrsg.). *Historische Kompetenzen und Museen*. Idstein: Schulz-Kirchner.
- Przyborski, Aglaja & Wohlrab-Sahr, Monika (2014). *Qualitative Sozialforschung: Ein Arbeitsbuch* (4. Aufl.). München: Oldenbourg.
- Reeken, Dietmar von (2014). Gegenständliche Quellen und museale Darstellungen. In Hilke Günther-Arndt & Meik Zülsdorf-Kersting (Hrsg.), *Geschichts-Didaktik: Praxishandbuch für die Sekundarstufe I und II* (6. Aufl., S. 144–157). Berlin: Cornelsen Scriptor.
- Reh, Sabine (2012). Beobachtungen aufschreiben: Zwischen Beobachtungen, Notizen und «Re-writing». In Heike de Boer & Sabine Reh (Hrsg.), *Beobachtung in der Schule – Beobachten lernen* (S. 115–129). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Robson, Colin (2008). *Real World Research: A Resource for Users of Social Research Methods in Applied Settings* (2. Aufl.). Chichester: Wiley.
- Röttele, Hannah (2017). Mensch, Objekt, Objektbegegnung – Eine empirische Studie zum Wahrnehmungsverhalten von Schülerinnen und Schülern bei einem Besuch im Historischen Museum. In Monika Waldis & Béatrice Ziegler (Hrsg.), *Forschungswerkstatt Geschichtsdidaktik 15. Beiträge zur Tagung «geschichtsdidaktik empirisch 15»* (S. 111–123). Bern: hep.
- Röttele, Hannah (2020). «Objektbegegnungen» im historischen Museum: Eine empirische Studie zum Wahrnehmungs- und Rezeptionsverhalten von *Schüler_innen*. München: kopaed.
- Ruona, Wendy E. A. (2005). Analyzing qualitative data. In Richard A. Swanson & Elwood F. Holton (Hrsg.), *Research in organizations: Foundations and methods in inquiry* (S. 233–263). San Francisco: Berrett-Koehler.
- Sauer, Michael (2013). Quellenarbeit im Geschichtsunterricht – Empirische Befunde. *Zeitschrift für Geschichtsdidaktik*, 12, 176–197.

- Schlutow, Martin (2012). *Das Migrationsmuseum: Geschichtskulturelle Analyse eines neuen Museumstyps*. Berlin: Lit.
- Schulze, Mario (2017). *Wie die Dinge sprechen lernten: Eine Geschichte des Museumsobjektes 1968–2000*. Bielefeld: transcript.
- Spieß, Christian (2014). *Quellenarbeit im Geschichtsunterricht: Die empirische Rekonstruktion von Kompetenzerwerb im Umgang mit Quellen*. Göttingen: Vandenhoeck & Ruprecht.
- Strotzka, Heinz (1983). *Zur Praxis des Geschichtsunterrichts: Analysen, Beispiele, Materialien*. Wien: Österreichischer Bundesverlag.
- Thiemeyer, Thomas (2012). Die Sprache der Dinge: Museumsobjekte zwischen Zeichen und Erscheinung. In Gisela Staupe (Hrsg.), *Das Museum als Lern- und Erfahrungsraum: Grundlagen und Praxisbeispiele* (S. 51–60). Wien, Köln, Weimar: Böhlau.
- Thomas, David R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2), 237–246. <https://doi.org/10.1177/1098214005283748>.
- Tischner, Christian K. (2008). *Historische Reden im Geschichtsunterricht*. Schwalbach/Ts.: Wochenschau.
- Wagner, Bernd (2013). Informelles Sachlernen von Kindern im Museum der Dinge. *Zeitschrift für Erziehungswissenschaft*, 16, 203–218. <https://doi.org/10.1007/s11618-013-0422-0>.
- Weindl, Roman (2019). *Die «Aura» des Originals: Über den Zusammenhang von Authentizität und Besucherinteresse*. Bielefeld: transcript.
- Wojdon, Joanna (2018). Museum lessons and the teacher's role. *International Journal of Research on History Didactics, History Education and History Culture*, 39, 93–105.

«Fotos, die jede*r kennt?»

Eine empirische Untersuchung historischer Kenntnisse anhand von Bildikonen bei Studierenden des Lehramts Geschichte

Daniel Fastlabend und Johannes Meyer-Hamme

1 Einleitung

Ausgangspunkt dieser empirischen Studie ist die Beobachtung, dass zwischen Aussagen über die Bekanntheit von Bildikonen im Allgemeinen und den konkreten Kenntnissen von Studierenden des Lehramts Geschichte in vielen Fällen krasse Widersprüche bestehen. Das Foto von Ernesto «Che» Guevara, der «Junge im Warschauer Ghetto», der gefolterte «Kapuzenmann» im Gefängnis Abu Ghraib und die brennenden Twin Towers sind ganz sicher sehr häufig reproduzierte und auch in Geschichtskulturen präsente Foto-Ikonen, aber in Seminaren über historisches Lernen mit Bildern sind sie längst nicht allen Studierenden bekannt.

Dabei spiegelt sich eine boomende Geschichtskultur, in der vor allem visuelle Zeugnisse eine besondere Rolle einnehmen, auch konkret in Büchern wider, die die gegenwärtigen Bildikonen dokumentieren. Demnach war das 20. Jahrhundert ein «Jahrhundert der Bilder», wie es Gerhard Paul ausgedrückt hat (Paul, 2008; 2009), wobei sicher auch bewegte Bilder dazu zu zählen sind (Bunnenberg, 2020). Als Beschreibungen der Geschichtskultur sind diese Aussagen zwar überzeugend, allerdings stellte Bodo von Borries in einer kleinen Untersuchung mit Studierenden ($N = 10$) zu verschiedenen Bildikonen fest, dass «Bekanntheitsgrad und Wirkungstiefe der zeitgeschichtlichen Bildikonen heftig überschätzt [werden]» (von Borries, 2015, 124). Hier zeigt sich exemplarisch, dass angehende Geschichtslehrkräfte bisher kaum erforscht wurden, sich aus forschungspragmatischen Gründen für explorative Studien, wie die hier vorgestellte, jedoch besonders eignen (z. B. Resch, 2021; Nitsche, 2021).

Daraus resultieren die Fragen, inwiefern Studierenden die historischen Zusammenhänge von Bildikonen bekannt sind und welche Deutungen sie damit ver-

binden. Damit tangieren diese Forschungsfragen die grundlegende Debatte darum, welchen Stellenwert historische Kenntnisse bei der Erforschung und der Beeinflussung des «Geschichtsbewusstseins in der Gesellschaft» (Jeismann, 1977, 12) haben. Dass historische Kenntnisse in vielen Forschungen eine zentrale Rolle spielen, zeigt sich exemplarisch an den großen Studien zum Geschichtsbewusstsein Jugendlicher von Bodo von Borries, über die Studien von Klaus Schröder (Schroeder, Deutz-Schroeder, Quasten & Schulze-Heuling, 2012) bis hin zu den Studien zur Conceptual-Change-Forschung insbesondere der Arbeitsgruppe um Hilke Günther-Arndt (Günther-Arndt, 2006). Allerdings ist ebenso festzustellen, dass sich in diesen Forschungen die Definitionen historischen Wissens beziehungsweise historischer Kenntnisse deutlich unterscheiden und dementsprechend auch unterschiedliche Forschungsinstrumente eingesetzt und Bewertungen vorgenommen wurden (Meyer-Hamme, 2022). Vor diesem Hintergrund wird in diesem Aufsatz die Frage diskutiert, inwiefern sich anhand von Bildikonen historische Kenntnisse erheben lassen. Die Diskussion dieser Frage (Abschn. 5) setzt sowohl Begriffsklärungen (Abschn. 2), methodische Überlegungen (Abschn. 3) ebenso wie die Vorstellung von Ergebnissen der Befragung voraus (Abschn. 4).

2 Begriffsklärung: Historische Kenntnisse und Bildikonen

Oberflächlich betrachtet scheint klar, dass historisches Lernen neben einer Kompetenzentwicklung auch zu historischem Wissen führen soll. Dabei ändert sich das, was als historisch relevantes Wissen angesehen wird, mit der Zeit und es gehört zur gängigen Kritik älterer Menschen an der jüngeren Generation, dass sie nicht genügend über relevante historische Ereignisse informiert seien (z. B. Wineburg, 2018, 15). Bei genauerer Betrachtung wird allerdings deutlich, dass es keineswegs eine anerkannte theoretische Modellierung davon gibt, was unter historischem Wissen verstanden werden kann. So gibt es in den aktuellen geschichtsdidaktischen Handbüchern keine eigenständigen Beiträge zu diesem Begriff (z. B. Barricelli & Lücke, 2012; Mayer, 2006). Vielmehr gibt es einzelne Ansätze dazu: So schlägt Beatrice Ziegler vor, das «historische Konzept» als Arbeitsbegriff zu nutzen, um das «Wissen» zu einem spezifischen Thema zu erfassen, das die Lernenden sich eigenmächtig aneignen. Damit solle das Geschichtsbewusstsein der Individuen in seiner inhaltlichen und themenspezifischen Dimension erfasst werden (Ziegler, 2012). Wolfgang Hasberg unterscheidet mit Rückgriff auf die psychologischen Gedächtnisformen hingegen zwischen dem «prozeduralen Wis-

sen» (Erkenntnisstruktur der Disziplin), der «episodischen Vorstellung» (Wissen über Zustände und Ereignisse in der Vergangenheit) und der «semantischen Einstellung» (Prinzipien und Kategorien historischen Denkens), die zusammen das Geschichtswissen im Geschichtsbewusstsein bilden (Hasberg, 2012). Christoph Kühberger zieht hingegen die Unterscheidung zwischen «Faktenwissen», «konzeptionellem Wissen», «prozeduralem Wissen» und «metakognitivem Wissen» vor (Kühberger, 2012, 35). Eine durchaus ähnliche Modellierung, die verschiedenste Wissensbereiche umfasst, schlägt Arthur Chapman im Zuge des gegenwärtigen *knowledge turn* vor. Dazu unterscheidet er zwischen *first and second order knowledge and understanding* (Chapmann, 2021, 13), wobei dem ersten Bereich dann Vorstellungen und deklaratives Wissen wie Tatsachenbehauptungen und Konzeptvorstellungen von Vergangem zugeordnet und unter dem zweiten Bereich Aspekte des Metawissens über historisches Denken verstanden werden. Im Gegensatz dazu spricht von Borries von historischen Kenntnissen und fokussiert sich damit auf das Wissen über Vergangenheit und klammert das Wissen über die Logik historischen Denkens aus (von Borries, 2012).

Von einer allgemein anerkannten Terminologie und Konzeption kann also nicht gesprochen werden und auch die Begriffe «Wissen» und «Kenntnisse» können in diesem Beitrag nur problematisierend verwendet werden. Gemeinsam ist diesen Ansätzen allerdings, dass zum historischen Wissen immer auch die Nennung von Einzelheiten in Form eines deklarativen Wissens gehört, die mal unter dem Begriff des «Faktenwissens», mal unter dem «episodischen Gedächtnis» mal als «Konzept» oder «Kenntnis» bezeichnet werden. Im vorliegenden Aufsatz wird deshalb von Kenntnissen gesprochen. Als historische Kenntnisse werden hier solche Bilddeutungen interpretiert, die plausibel geschichtskulturell gängige Narrationen («narrative Abbreviaturen») reproduzieren können. Anhand der konkret diskutierten historischen Beispiele wird aber auch deutlich, wie sehr die in der Geschichtskultur relevanten historischen Kenntnisse zugleich auch in Kontroversen eingebettet sind (von Borries, 2012), beziehungsweise dass es sich um Erinnerungsorte handelt, wie Hasberg am Beispiel Canossa herausgearbeitet hat (Hasberg, 2012).

Auch Bildikonen können als Erinnerungsorte verstanden werden. Dabei stellt sich die Frage, wie Bilder zu Ikonen werden und was diese Zuschreibung über den Umgang mit Geschichte(n) in Gesellschaften aussagt. Ein möglicher Ausgangspunkt einer solchen Betrachtung ist die Unterscheidung von Merkmalen zwischen Bildern und Bildikonen. Letztere werden auch als «Schlüsselbilder» bezeichnet (Reiche, 2009, 13–14). Es sind Bilder von vergangenem Geschehen,

von denen angenommen wird, dass sie zum «festen Inventar unserer historischen Imagination» zählen (Krammer, 2008, 165). Somit ist ein erstes Merkmal, ob ein Bild als Ikone einzustufen ist oder nicht, die Frage danach, inwieweit es weitertradiert wird und dieser Ausschnitt von Vergangenen in der Gegenwart mehr oder weniger präsent beziehungsweise bekannt gehalten und nicht, wie unzählige weitere Bilder, dem Vergessen preisgegeben wird. Auf der Ebene der Gegenwart wird einer Ikone somit eine geschichtskulturelle Präsenz und damit ein gewisser Bekanntheitsgrad zugesprochen.

Daraus resultiert die Frage, welche Merkmale von Bildikonen überhaupt dazu führen können, dass diese in Gesellschaften präsent gehalten werden. Grundlegend zählt dazu die Bedeutungszuschreibung, ein Schlüsselbild für historische Ereignisse zu sein. Damit wird häufig verknüpft, dass Ikonen ebenfalls eine ästhetische, dramatisierende sowie simplifizierende Wirkung auf Betrachter*innen haben können (Hamann, 2019, 9). In diesem Sinne wird Ikonen nachgesagt, ebenfalls Einfluss auf das Denken und Handeln von Menschen zu nehmen beziehungsweise bewusstseins- sowie meinungsbildend wirken zu können, sodass darüber auch gewisse Vorstellungen von Vergangenen nicht nur tradiert, sondern auch erzeugt werden können (Reiche, 2009). Paul spricht daher davon, dass Ikonen auch «Geschichtsmotoren» seien, die nicht nur die Erinnerung an das Ereignis, sondern auch an die damit verbundene Narration präsent halten (Paul, 2006, 19). Somit wird Bildern, die die jeweils aus der Gegenwart getroffene Zuschreibung Ikone erhalten, zum einen ein gewisser Bekanntheitsgrad in Geschichtskulturen zugesprochen beziehungsweise eine Einbettung in geschichtskulturell präsente Narrationen. Zum anderen wird betont, dass diese Bilder eine z. T. auch ästhetische sowie emotionale Wirkung auf das Bewusstsein ihrer Rezipient*innen besitzen, wobei besonders Fotos und Filme demnach Einflüsse auf Vorgänge des Erinnerns haben (Hamann, 2019, 5). Andererseits ist diese Wirkung des Visuellen auf die Betrachter*innen nicht unumstritten und scheint z. T. sogar negative Effekte zu haben, wobei dies für Ikonenbilder noch dezidiert empirisch zu überprüfen ist (Ammerer, 2019).

Jedoch ist nicht nur die Ebene der gegenwärtigen Präsenz und Verwendung von Bildikonen in der Geschichtskultur zu beachten, sondern bereits die Entstehungsgeschichte des Bildes. So spielt zum Beispiel der unmittelbare Rezeptions- sowie Verbreitungskontext eine gewichtige Rolle im Hinblick auf die Zuschreibung als Ikone. Hier ist unter anderem darauf hinzuweisen, dass bereits bei der Produktion/Erschaffung die spätere Wirkung häufig mitbedacht wird, sodass nur wenige Ikonen auch als zufällige Schnappschüsse zu charakterisieren sind.

Vielmehr ist zu beachten, dass «historische Momente so in Szene gesetzt [werden], wie man sie später im Gedächtnis der Menschen verankern möchte» (Schreiber, 2004, 30–31). Deshalb ist das Objektiv der Kamera aufgrund der menschlichen Nutzung auch immer zugleich subjektiv (Hartmann, 2019, 6). Dennoch wird besonders Fotografien ein gewisser Beweis- beziehungsweise Belegcharakter zugesprochen (Krammer, 2008).

Nichtsdestotrotz ist die Zuschreibung, eine Ikone zu sein, eine aus der jeweiligen Gegenwart formulierte, und somit als eine rückblickende Bedeutungszuschreibung von Vergangenen zu verstehen. Dementsprechend handelt es sich bei Ikonen zwar um Bilder von Vergangenen, jedoch dienen sie vor allem dem Zweck, gewisse Vorstellungen sowie auch Deutungen von Vergangenen in der jeweiligen Gegenwart präsent zu halten (Krammer, 2008). Außerdem ist die Zuschreibung, ein Schlüsselbild für ein historisches Ereignis zu sein, wiederum als ein Anzeichen dahingehend zu interpretieren, dass bestimmten rückblickenden Deutungen selektiv Dominanz gegenüber anderen zugesprochen wird.

Aus geschichtsdidaktischer Sicht liegt daher der Fokus nicht primär auf dem Quellencharakter des Bildes und der Frage, «inwiefern Bilder [Ikonen] die Vergangenheit erhellen können, sondern [darauf,] was der Umgang damit über gegenwärtige Bedeutungen aussagt» (Körber, 2006, 170). Dementsprechend sind Bilder, die den aus der Gegenwart formulierten Status einer Ikone erhalten, als geschichtskulturelle Dokumente beziehungsweise als symbolhafte Verdichtungen für historische Darstellungen zu verstehen,¹ die auch als «narrative Abbrévatures» bezeichnet werden können (Rüsen, 1996, 512).

Des Weiteren ist mit Andreas Körber anzufügen, dass ikonische Fotos aufgrund ihres häufigen Verweises auf ein zeitliches davor und danach zwar Ansätze für Narrationen beziehungsweise Proto-Narrationen enthalten, selbst aber keine narrative Struktur haben, wie etwa Historien Gemälde (Körber, 2006). Daher ist davon auszugehen, dass auch Ikonen trotz ihrer vermeintlich eindeutigen und vor allem fotografischen Bildaussagen sowie ihres scheinbaren Schlüssel-

1 Hamann (2019) unterscheidet mit Verweis auf Ansätze der Visual History und speziell für Fotografien drei Funktionen: Bilder sind demnach Quellen der Forschung, Medien der Erinnerung wie auch Instrumente der Politik. D. h.

1. Bilder können Quellen für Konstruktion historischer Narrative sein (Vergangenheit)

2. dienen in sozialen Erinnerungsgemeinschaften als Symbole (Gegenwart)

3. werden als Bildakte/Waffe eingesetzt (Zukunft).

Für die in dieser Studie verwendeten Bildikonen sind v. a. die Ebenen 2 und 3 von Bedeutung, wobei natürlich auch immer der Entstehungskontext eine Rolle spielt.

charakters – «Ein Bild sagt mehr als tausend Worte» – dennoch eine gewisse Uneindeutigkeit besitzen. Im Hinblick auf die Rezeption der Bilder wird diese Vieldeutigkeit noch deutlicher. So wiesen bereits Roland Barthes sowie Stuart Hall auf die polysemischen Potentiale von Fotografien hin, ohne dies jedoch auf die mit Bildern verbundenen historischen Narrationen zu beziehen (Barthes, 1990, 33; Hall, 2004, 108). Das heißt, es bestehen für die Betrachter*innen grundsätzlich Spielräume, die eine andere als die vom Produzenten sowie der Produzentin des Bildes vorgesehene Rezeption ermöglichen (Schreiber, 2004, 48–49). Die optischen Zeichen und Signale müssen dementsprechend von den Betrachter*innen auch auf Basis der jeweiligen historischen Vorkenntnisse sowie Voreinstellungen erkannt, gelesen sowie interpretiert und in gewisser Weise dann auch dekonstruiert werden (Krammer, 2008, 167). Dies bedeutet wiederum, dass geschichtskulturell verschiedene historische Narrationen sowie Kenntnisse zu Fotoikonen bestehen, wobei empirisch zu erforschen ist, welche für die Proband*innen bekannt sind.

Wenn im Folgenden von Kenntnissen historischer Bildikonen die Rede ist, dann basiert dies auf der Annahme, dass sich der ikonische Status von Bildern nicht zuletzt aus der Bedeutungszuschreibung in der Geschichtskultur sowie aus der Einbettung in unterschiedliche historische Narrationen ergibt. Zugleich wird damit deutlich, wie sehr historische Kenntnisse auch immer geschichtskulturelle Kenntnisse sind, denn die Bedeutungszuschreibung zu einzelnen historischen Themen (z. B. «Canossa») oder auch Ikonen (z. B. Kniefall Brandts in Warschau) sind nur vor dem Hintergrund geschichtskultureller Debatten und Konventionen verständlich.

Außerdem stellt sich für dieses Projekt zudem die Frage, inwieweit sich die Kenntnisse von bekannten Bildern bei den Studierenden unterscheiden und welche weiterführenden Schlussfolgerungen daraus abzuleiten sind. Dabei werden nicht alle Bildgattungen und Ikonen aller Zeiten berücksichtigt, sondern diese Frage wird anhand von Fotos des 20. und 21. Jahrhunderts untersucht, die als Ikonen bezeichnet werden.

3 Methodische Vorgehensweise und Herausforderungen

3.1 Design des Erhebungstools

Die obigen Annahmen zu Schlüsselbildern, die «zu unserem festen Inventar gehören», sind eine z. T. auch normativ aufgeladene Bedeutungszuschreibung, die daraufhin zu hinterfragen ist, welches Kollektiv hier angesprochen wird und inwiefern sich diese Annahme empirisch bestätigen lässt. Eine Erhebung solcher historischer Kenntnisse lässt sich neben der Einstellungs-, Sozialisations- und Kompetenzforschung als eine der vier geschichtsdidaktisch-empirischen Forschungsperspektiven bezeichnen (Meyer-Hamme, 2016). Die vorliegende Untersuchung bietet eine Perspektive auf historische Kenntnisse, die in anderen Studien beispielsweise mit Multiple-Choice-Tests (Schroeder et al., 2012) oder Rangreihungsaufgaben (von Borries, 1995) erhoben wurden. Dabei ist nicht nur zentral, inwieweit die Bilder bekannt oder unbekannt sind, sondern mit welchen Begriffen die Ikonen bezeichnet werden und welche «narrativen Abbriviaturen» beziehungsweise Interpretationsansätze damit gegebenenfalls verbunden werden.

Für die Entwicklung des Erhebungstools wurden aus der Vielzahl an möglichen Bildern solche ausgewählt, die in den einschlägigen Sammlungen als Ikonen bezeichnet wurden (Paul, 2008; 2009). Darüber hinaus wurden nicht nur Fotografien von Männern, kriegerischen Szenen, Katastrophen und europäischen Ereignissen berücksichtigt. Schließlich wurden solche Bilder ausgewählt, die die zeitliche Spanne vom Beginn des 20. Jahrhunderts bis in die Lebenszeit der Befragten am beginnenden 21. Jahrhundert umfassten. Folgende 12 Bilder wurden ausgewählt:

1. Lenin während einer Ansprache vor Truppen in Moskau (1920). Das Bild erhielt eine besondere Aufmerksamkeit, weil es nicht nur eine der wenigen Fotografien von Lenin ist, sondern in der Originalfassung auch Trotzki und Kamenev zu sehen waren (Waschik, 2008, 252–259).
2. «Junge aus dem Warschauer Ghetto» (April/Mai 1943). Das Bild stammt aus dem «Stroop-Bericht», mit dem die SS die Niederschlagung des Widerstands sowie die gewaltsame Auflösung des Warschauer Ghettos dokumentierte (Knoch, 2009, 43).
3. Torhaus Auschwitz-Birkenau (Februar/März 1945). Dieses Bild steht symbolisch für die Verbrechen und massenhaften Tötungen sowie für die Gräueltaten der Shoah durch die Nationalsozialisten (Hamann, 2007, 92, 97).

4. Marilyn Monroe, die versucht, ihren durch den Wind hochwehenden Rock herunterzuhalten (1954). Besonders durch diese Aufnahme und deren Reproduktionen wurde Monroe zu einer «Glamour-Ikone des 20. Jahrhunderts» (Bronfen, 2011, 148–150).
5. Das Porträt von Ernesto «Che» Guevara (1960). Das vermeintlich am «häufigsten reproduzierte Foto der Geschichte» wurde erst nach «Ches» Tod zum Symbol von Studentenbewegungen sowie globalen Protestbewegungen. Inzwischen lässt sich das Gesicht Guevaras auf vielen Artikeln und Gegenständen sowie Werbeplakaten finden, ohne explizit ein Symbol des Protests sein zu müssen (Lahrem, 2008, 234–235).
6. «Sprung in den Westen» (1961). Das Foto zeigt den damaligen NVA-Soldaten Conrad Schumann, als dieser zwei Tage nach dem Bau der Berliner Mauer über den Stacheldraht in einen westlichen Sektor springt und somit aus der DDR flieht. Im Westen wird dieses Bild «als Sprung in die Freiheit» und somit als Symbol für die Flucht aus der DDR in den Westen vielfach rezipiert (Demke, 2009, 111–113).
7. Mondlandung (1969). Das am breitesten rezipierte Bild zeigt Edwin Aldrin (und nicht Neil Armstrong), wie er vor der amerikanischen Flagge salutiert und gewissermaßen den Mond für die USA in Besitz nimmt. Neben des Dokumentations- und Inszenierungscharakters der Bilder für die «wissenschaftlich-technische Sensation» gibt es langanhaltende Debatten um deren Authentizität (Heßler, 2008, 394–401).
8. Das Mädchen Kim Phuc (1972). Die Aufnahme zeigt das Mädchen Kim Phuc mit starken Verbrennungen, wie es nach einem Napalmangriff der südvietnamesischen Einheit ohne Kleidung flieht. Das Bild wurde zur Ikone der internationalen Protestbewegungen gegen den Vietnamkrieg (Paul, 2009, 91).
9. «Tankman» (Juni 1989). Das Bild zeigt, wie ein Mann sich auf dem Tiananmen-Platz («Platz des Himmlischen Friedens») mit Einkaufstüten vier Panzern entgegenstellt. Es ist eine der Ikonen für die blutige Niederschlagung der chinesischen Demokratiebewegung (Drechsel, 2008, 566–573).
10. «Mauerfall» (November 1989). Das Bild zeigt jubelnde Menschen, die auf der Mauerkrone vor dem Brandenburger Tor sitzen. Es wurde zur Ikone für die Grenzöffnung sowie z. T. sogar für die spätere Wiedervereinigung Deutschlands (Demke, 2009, 119).
11. Twin Towers 9/11 (11. September 2001). Das Bild zeigt einen brennenden Turm und ein Flugzeug kurz vor dem Einschlag in den zweiten Turm. Es ist

eine ikonografische Verdichtung der Ereignisse am 11. September (Münkler, 2009, 151–152).

12. Der Gefangene in Abu Ghraib (2003). Das häufig euphemistisch als «Kapuzenmann» bezeichnete Bild ist eines von vielen Folterbildern von Gefangenen der US-Armee während des Zweiten Irak-Krieges. Es zeigt, wie ein Gefangener dazu gezwungen wurde, mit einer «Kapuze» über dem Kopf auf einer Kiste zu stehen, während er an Stromkabel angeschlossen war, um ihm somit den Eindruck einer Exekution zu vermitteln (Paul, 2008, 702–709).

Man mag einwenden, dass diese Auswahl in gewisser Weise willkürlich und subjektiv ist und andere Bilder ebenso geeignet gewesen wären. Objektive Kriterien für eine solche Auswahl sind jedoch nicht zu definieren, vielmehr geht es darum, solche Bilder auszuwählen, denen einerseits eine große Bekanntheit zugesprochen wird und die andererseits mutmaßlich unter den befragten Studierenden nicht im gleichen Maße bekannt sind.

Als Erhebungstool wurde ein halboffener Fragebogen entworfen, der vor allem auf die Gewinnung quantitativer Daten abzielte.² In dem Survey konnten die Befragten zu jedem gezeigten Bild dann zwischen drei Kategorien wählen:

- «Dieses Bild habe ich noch nie gesehen.»
- «Dieses Bild habe ich mal gesehen, kann aber nichts weiter dazu sagen.»
- «Dieses Bild zeigt ... (bitte ergänzen).»

Bei der dritten Kategorie konnten die Befragten das Bild benennen und den Kontext beschreiben.³ Damit sollte auch gleichzeitig deutlich werden, welche Interpretationen die befragten Personen mit den Bildern verbinden, um somit auch qualitative Daten zu den möglichen Kenntnissen über die Ikonen zu sammeln. Um differenzierende Aussagen zwischen Gruppen herstellen zu können, wurden als Sozialdaten neben dem Geschlecht, auch die Erstsprache(n), die Semesteranzahl und die Zahl der Bücher zu Hause abgefragt. Die Frage nach den Büchern gilt als Marker für den Bildungshintergrund der Eltern. Befragt wurden Studierende

2 Bevor die Bilder einsehbar waren, wurden Proband*innen zudem nach zwei für sie relevanten Bildern des 20. Jahrhunderts befragt. Von den 151 hier erhaltenen und durchaus vielfältigen Antworten, sind mit 16 Prozent der Antworten die Twin Towers, mit 11 Prozent der «Mauerfall», 7 Prozent Kim Phuc, 4 Prozent die Mondlandung sowie der Kniefall Willy Brandts und mit 3 Prozent das Torhaus Auschwitz-Birkenau hervorzuheben ($N = 353$).

3 Zusätzlich enthielt jedes Bild auf dem Fragebogen eine weitere Zeile zur Angabe der Jahreszahlen.

des Lehramts Geschichte an mehreren Hochschulen in Deutschland ($N = 353$) in den Jahren 2017 bis 2018.

3.2 Die Auswertung mittels MaxQDA und die Herausforderung der «plausiblen Zuordnung» der Bildkenntnisse

Die Auswertung erfolgte codierend mittels der Analysesoftware MaxQDA. Diese Notwendigkeit ergab sich aufgrund der halboffenen formulierten Items zur Bildbeschreibung. Demzufolge galt es bei den Bildbeschreibungen zwischen «plausibel zuzuordnenden» und «nicht plausibel zuzuordnenden» Bildbeschreibungen zu unterscheiden, wobei nur die plausibel zuzuordnenden Beschreibungen als Bildbeschreibungen beziehungsweise Kenntnis interpretiert wurden. Die Definitionen für die Zuordnungen beziehungsweise Codierungen der Bildbeschreibungen wurden wiederum auf Basis der Ausführungen zu den Bildern in Pauls Ikonensammlung und den dort enthaltenen Kontextualisierungen erstellt. Da es sich um eine explorative Studie handelt, wurden Zweifelsfälle der Codierung in der Gruppe ausdiskutiert. Bei nicht plausiblen beziehungsweise zu vagen Bildbeschreibungen wurden induktiv in der Sprache der Befragten Codes entwickelt, die deren Deutungen verdeutlichen.

An dieser Stelle ist anzuführen, dass besonders Kenntnistests Aufgabenformate erfordern, die zwischen eindeutig «korrekten» sowie «falschen» Antworten unterscheiden lassen, um valide Aussagen über Kenntnisse treffen zu können; zum Teil hilft auch die Kategorisierung von teilweise korrekten Antworten (Meyer-Hamme, 2016, 91). Exemplarisch lässt sich an den Ikonen somit aufzeigen, wie schwierig sich eine Einordnung als eindeutig «richtig» oder eindeutig «falsch» für historische Kenntnistests gestaltet. Eine gewisse Schwierigkeit ergibt sich für die Codierungsentscheidungen aus den verschiedenen Schichten der Erinnerungsprozesse der Fotoikonen. Exemplarisch sei hier auf die Diskrepanz zwischen dem Entstehungs- und dem Erinnerungskontext des Bildes von Kim Phuc sowie die Einbettung in unterschiedliche Narrationen hingewiesen. Obwohl das Bild von der mit schlimmsten Verbrennungen gezeichneten Kim Phuc nach einem Napalmangriff von südvietnamesischen Einheiten gemacht wurde, erhielt es häufig die spätere Bildbeschreibung, während eines Angriffs von US-amerikanischen Truppen entstanden zu sein. Als solches wurde es auch zum Sinnbild für Proteste gegen den Vietnamkrieg sowie für die amerikanischen Kriegsverbrechen in diesem Krieg (Paul, 2009, 91). Dies deutet einerseits auf die zuvor herausgestellten Merkmale von Ikonen und der retrospektiven Zuschreibung die-

ses Status hin, führte in der Auswertung der Bildbeschreibungen jedoch dazu, zwischen «plausibel zuzuordnen» und «nicht plausibel zuzuordnen» zu differenzieren.

Die Auswertung der Bildbeschreibungen für dieses Beispiel wurde dann als plausibel kategorisiert, wenn das Bild in den Kontext des Vietnamkrieges gestellt wurde, was durch 60 Prozent der befragten Personen erfolgte (siehe Tabelle 2). In dieser Kategorie sind auch Bildbeschreibungen enthalten, die das Bild als einen Angriff von US-Truppen und/oder als «Massaker von My Lai» beschreiben. Bildbeschreibungen, die es in einen ganz anderen Kontext wie zum Beispiel Hiroshima/Nagasaki oder generell dem «Kalten Krieg» verorteten, wurden hingegen als «nicht plausible Bildbeschreibung» codiert sowie kategorisiert und somit nicht als Kenntnis gewertet, da sich bspw. an dem Bild die These eines «Kalten Krieges» für das Beispiel Vietnam widerlegen lässt. In dieser Hinsicht galt es für jede Ikone Kriterien für die Zuordnung in «plausibel zuzuordnen» und «nicht plausibel zuzuordnen» zu formulieren. Diese Auswertungen bieten daher auch einen Anlass für weitere qualitative Analysen.

Die in diesem Aufsatz vorgelegten quantitativen Auswertungen machen es nötig, die Stichprobe zu beschreiben. Erhoben wurden die Daten an drei Hochschulen,⁴ allerdings sind die Fallzahlen recht unterschiedlich, sodass eine standortvergleichende Analyse problematisch ist. Die Studierenden waren im Mittel 22.5 Jahre alt.⁵ Sie waren in der Mehrzahl weiblich und der Großteil hatte weniger als sechs Fachsemester Geschichte studiert.⁶ Etwas mehr als die Hälfte hat angegeben, dass es zu Hause weniger als 200 Bücher gibt.

4 Ein großer Dank geht an Bärbel Völkel und Felix Hinz für die Unterstützung.

5 54 % der Studierenden waren zum Zeitpunkt der Befragung zwischen 18 und 22 und 31 % zwischen 23 und 26 Jahre und eine dritte Gruppe von 13,5% um die 30 Jahre alt.

6 Hier wurden die Fachsemester insgesamt gezählt und nicht zwischen BA und MA unterschieden.

Tabelle 1: Beschreibung der Stichprobe

Gender	Weiblich	205
	Männlich	147
Anzahl der Bücher zu Hause	bis zu 200	187
	mehr als 200	162
Fachstudium	1–5 Semester	247
	Mehr als 5 Semester	95
Gesamt <i>N</i>		353

4 Ausgewählte Ergebnisse der Auswertung

4.1 (Un-)Bekanntheit der Bildikonen

Tabelle 2: Kenntnisse von Bildikonen (*N* = 353)⁷

	Bildbeschreibung «plausibel zuzuordnen» Prozent / Abs. Nennung	Bildbeschreibung «nicht plausibel zuzuordnen» Prozent / Abs. Nennung	Bild «mal gesehen, aber kann nichts weiter dazu sagen» Prozent / Abs. Nennung	«Bild noch nie gesehen» Prozent / Abs. Nennung	Gesamt Prozent / Abs. Nennung
9/11 «Twin Towers»	99.7 % / 352	0 % / 0	0.3 % / 1	0 % / 0	100 % / 353
«Mauerfall»	97.1 % / 343	0.5 % / 2	1.4 % / 5	0.8 % / 3	99.8 % / 353
Mondlandung	97.1 % / 343	0.8 % / 3	2 % / 7	0 % / 0	99.9 % / 353
Marilyn Monroe	91.7 % / 324	1.1 % / 4	6 % / 21	1.1 % / 4	99.9 % / 353
Torhaus Auschwitz-Birkenau	67.9 % / 239	9.3 % / 33	11.7 % / 41	11.1 % / 39	100 % / 352

⁷ Die Prozentangaben beziehen sich auf die Anteile der jeweiligen Kategorie von der Gesamtzahl der Stichprobe und sind in den Nachkommastellen gerundet.

	Bildbeschreibung «plausibel zuzuordnen» Prozent / Abs. Nennung	Bildbeschreibung «nicht plausibel zuzuordnen» Prozent / Abs. Nennung	Bild «mal gesehen, aber kann nichts weiter dazu sagen» Prozent / Abs. Nennung	«Bild noch nie gesehen» Prozent / Abs. Nennung	Gesamt Prozent / Abs. Nennung
«Che» Guevara	60.6 % / 213	10.8 % / 38	24.2 % / 85	4.3 % / 15	99.9 % / 351
Vietnam Kim Phuc	60.5 % / 213	13 % / 46	13.1 % / 46	13.4 % / 47	100 % / 352
«Sprung in den Westen»	37.9 % / 133	8.8 % / 31	22.2 % / 78	31.1 % / 109	100 % / 351
«Lenin»	28.9 % / 102	7.1 % / 25	25 % / 88	38.9 % / 137	99.9 % / 352
«Tankman» auf dem Platz des Himmlischen Friedens	18.3 % / 64	12.5 % / 44	42.2 % / 148	26.9 % / 94	99.9 % / 350
«Junge Warschauer Ghetto»	16.1 % / 57	26.6 % / 94	30 % / 106	27.2 % / 96	99.9 % / 353
Abu Ghraib «Kapuzenmann»	3.7 % / 13	11.6 % / 41	11.9 % / 42	72.8 % / 257	100 % / 353

Grundsätzlich lässt sich mit Blick auf die Verteilung der Häufigkeiten und angesichts der relativen Stabilität dieser über drei Standorte hinweg festhalten, dass für dieses Set an Bildern drei Arten von Ikonen unterschieden werden können:

1. Es gibt Fotografien, die nahezu jede*r Geschichtsstudierende kennt. Dazu zählt das Bild der brennenden Twin Towers, das Bild der Menschen auf der Berliner Mauer, das Bild der Mondlandung und auch das Bild Marilyn Monroes. Hier lässt sich zwar besonders mit Bezug auf das Bild der Mondlandung anführen, dass die Bilddarstellungen bereits einen Teil der Antworten enthalten mögen und sich somit unterschiedliche Bekanntheitsgrade erklären lassen beziehungsweise sich anhand der Bildikonen die historischen Kenntnisse der Proband*innen nicht valide genug erheben lassen mögen. Einzuwenden ist jedoch, dass 135 Personen und somit 38 Prozent der Befragten den Namen Neil Armstrong sowie 18 Prozent ebenfalls das exakte Jahr und nochmal 20 Prozent die Dekade der Mondlandung in ihrer Bildbeschreibung nennen konnten und somit genauere Kenntnisse zu diesem Ereignis aufwie-

- sen, die über eine reine Bildbeschreibung hinausgehen. Darüber hinaus gibt es vereinzelt Studierende, die nichts zu diesem Foto sagen können.
2. Es lässt sich eine Gruppe von Bildikonen feststellen, die vielen Studierenden sehr vertraut und bekannt sind, aber längst nicht jedem beziehungsweise jeder geläufig sind. Dazu zählt bezeichnenderweise das Torhaus des Vernichtungslagers Auschwitz-Birkenau, das Porträt von «Che» Guevara und das Bild von Kim Phuc mit ihren Verbrennungen. Bei diesen Bildern können etwa zwei Drittel der Befragten plausible Bildbeschreibungen liefern, die auf eine Kenntnis schließen lassen.
 3. Es gibt Bilder, deren Kontexte nur sehr vage beschrieben beziehungsweise nahezu unbekannt sind. Dazu zählen der «Sprung in den Westen» und das Bild von Lenin, die von einem Drittel der Befragten plausibel zugeordnet werden konnten, während der «Tankman» auf dem Platz des Himmlischen Friedens nicht einmal von einem Fünftel, der Gefolterte in Abu Ghraib nicht einmal von einem Zwanzigstel der Befragten plausibel beschrieben wurde. Wohlgemerkt: Alle diese Bilder gelten als anerkannte Ikonen, während die Ergebnisse dieser Befragung zeigen, dass nur wenige dieser Bilder den Studierenden des Lehramts Geschichte geläufig sind!

Es lohnt sich also ein etwas genauerer Blick auf die Antworten, wobei hierzu exemplarisch kontrastive Befunde zu den Bekanntheitsunterschieden zwischen den Ikonen herausgestellt werden. Zunächst mag mit Blick auf die nahezu allen bekannten Ikonen angefügt werden, dass es sich bei den Bildern der brennenden Twin Towers und dem «Mauerfall» um aktuelle Ereignisse aus der näheren Zeitgeschichte handelt und dies bereits ihre Bekanntheit erklären mag. Jedoch lässt sich die Bekanntheit nicht allein auf die Aktualität der Ereignisse zurückführen, denn im Fragebogen waren ebenfalls aktuellere oder zeitgleiche Ikonen wie der «Tankman» und der «Kapuzenmann», der im Jahre 2004 das Cover des Spiegels zierte, enthalten. Obwohl diese Bilder teilweise sogar direkt verbunden oder aus einem kontextuellen sowie zeitlichen Umfeld zum Beispiel des «Mauerfalls» und «9/11» stammen, sind sie kaum bis gar nicht bekannt. So ließe sich zwischen dem Terroranschlag vom 11. September als Auslöser für den Irakkrieg eine Verbindung zu dem Bild von der Folter des Gefangenen von Abu Ghraib herstellen. Hier ist außerdem anzumerken, dass 13 Prozent der Befragten um die 30 Jahre alt und somit zum Zeitalter der Veröffentlichung Jugendliche waren. Dennoch war Ihnen dieses größtenteils unbekannt. In dieser Hinsicht verwundert auch, dass der ebenfalls aus dem Jahre 1989 stammende «Tankman», der gewissermaßen das

Gegenbild der «friedlichen Revolution» in der DDR darstellt und dieser als «chinesische Lösung» vorausgeht, auch unter zukünftigen Geschichtslehrenden nur sehr wenigen geläufig ist. Insgesamt mag diese Diskrepanz ein Hinweis darauf sein, dass die je individuellen historischen Kenntnisse neben der Prägung durch zeitgeschichtliche Ereignisse ebenfalls eine nationale und mit Blick auf die Twin Towers stark westliche Prägung tragen.

In Bezug auf die Bekanntheit und die Entstehungszeitpunkte verwundert ebenfalls die Spanne zwischen der Bekanntheit des Torhauses von Auschwitz-Birkenau und dem «Jungen aus dem Warschauer Ghetto», wobei für diese Beispiele die Codierungsentscheidungen noch genauer zu diskutieren sind. Beispielsweise beschrieben 239 beziehungsweise 68 Prozent der befragten Personen das Torhaus Auschwitz-Birkenau als ein Bild des Vernichtungslagers Auschwitz. Hier ist jedoch anzumerken, dass Beschreibungen, die es lediglich als Bild eines Konzentrationslagers bezeichneten, als zu vage eingestuft wurden. Diese Gruppe umfasst 33 beziehungsweise zusätzlich 9,34 Prozent der Proband*innen. Demgegenüber konnten nur insgesamt 57 beziehungsweise 16 Prozent der Proband*innen den «Jungen aus dem Warschauer Ghetto» dessen Entstehungskontext zuordnen, wie er bspw. bei Hamann beschrieben ist (Hamann, 2007). Hier ist ebenfalls anzumerken, dass eine noch größere Gruppe eine generelle Einordnung in größere Kontexte vornahm. So wurden weitere 94 Antworten beziehungsweise 26 Prozent aufgrund von Bildbeschreibungen wie Deportationen, Verbrechen im Kontext des Zweiten Weltkrieges oder auch Judenverfolgung und Pogrome als zu vage gewertet und nicht als hinreichend plausibel codiert. Exemplarisch zeigt sich daran wiederum, dass eine scharfe Trennung ebenso schwierig wie notwendig bleibt.

Ein weiterer Aspekt bezüglich der Prägung der historischen Kenntnisse ergibt sich unter anderem im Hinblick auf die Bekanntheit der «Stil-Ikone» Monroe. Obwohl manche Bilder in vielen Schulbüchern und damit mutmaßlich auch im Geschichtsunterricht vorkommen, können diese längst nicht von allen Studierenden plausibel zugeordnet werden (z. B. Lenin). Demgegenüber lässt sich an der Bekanntheit Monroes oder auch der Mondlandung, die höchstwahrscheinlich nicht dem schulischen Kontext entsprungen ist, darlegen, dass sich historische Kenntnisse mutmaßlich sowohl aus der geschichtsunterrichtlichen als auch besonders aus der außerschulischen und weiteren geschichtskulturellen Sozialisation im Allgemeinen speisen. Dieser Befund lässt sich ebenfalls als die von von Borries vermutete Differenz zwischen offiziellen und inoffiziellen historischen Kenntnissen interpretieren (von Borries, 2012, 13).

Ein weiterer kontrastiver Befund zeigt sich an der Lücke zwischen den plausiblen Angaben zum Bild von Ernesto «Che» Guevara (60 %) und der Angabe, dass nur 4 Prozent der Befragten dieses Bild noch nie gesehen haben. Dies mag ein Anzeichen dafür sein, dass es sich hierbei um eine teilweise enthistorisierte Ikone handelt, die zwar schon nahezu jede*r gesehen hat, deren Entstehungs- und oder Verbreitungskontext aber weitgehend unbekannt ist. Generell mag dies ein Hinweis darauf sein, dass viele dieser Bilder vage bekannt sind, aber kaum historisch verortet werden können.

Ein ähnliches Bild ergibt sich bei einem genaueren Blick auf die Antworten zur Monroe-Ikone. Obwohl insgesamt über 90 Prozent der Proband*innen den Namen Monroes wiedererkannten, gaben nur 34 Prozent eine Einordnung in den größeren Kontext der 1950er-/60er-Jahre an und 58 Prozent nannten gar keine Angaben zum zeitlichen Kontext der Ikone, sodass für das Bild Monroes ebenfalls der Eindruck von einer enthistorisierten und nahezu zeitlosen Ikone entstehen kann.

An diesen Ergebnissen wird eine generelle Eigenschaft von Ikonen deutlich. Somit können Ikonen als symbolhafte Bilder für bestimmte Deutungen stehen, die in der Geschichtskultur tradiert werden, die sich über die Zeit immer mehr von den damit verbundenen historischen Narrationen sowie vom eigentlichen Entstehungskontext lösen.

4.2 Unterscheidung von Subgruppen

In einem weiteren Auswertungsschritt wurde der Datensatz auf mögliche Zusammenhänge zwischen plausiblen Bildbeschreibungen und den soziodemografischen Daten hin analysiert. Für diesen Auswertungsschritt dienten vor allem die weniger bis kaum bekannten Ikonen und die prozentuale Verteilung der hierfür getroffenen eindeutigen Bildbeschreibungen als interessante Anhaltspunkte. Für diesen Analyseschritt wurden in MaxQDA mit einer Kreuztabelle die Häufigkeiten der zuvor vergebenen Codes zur Einstufung der Bildkenntnisse noch einmal anhand der Auswahl verschiedener Dokumentvariablen wie Gender, Alter und Erstsprache(n) durchsucht, um weitere Differenzierungen hinsichtlich historischer Kenntnisse erkennen zu können. Dass sich anhand der Bildkenntnisse bemerkenswerte Differenzierungen erzielen lassen, wird vor allem mit Blick auf die oben aufgeführten Auswertungen ersichtlich, kann hier jedoch nur punktuell andiskutiert werden.

Tabelle 3: Plausible Nennungen in Abhängigkeit zu soziodemografischen Daten (prozentuale Werte & absolute Nennungen, N = 353)⁸

	Gender (N = 352)		Anzahl der Bücher (N = 349)		Anzahl Fachsemester (N = 342)		Erstsprache (N = 351)	
	Weiblich N = 205	Männlich N = 147	200 u. weniger N = 187	>200 N = 162	1–5 Semester N = 247	>5 Semester N = 95	Erst-sprache Deutsch N = 274	Mehrere Erst-sprachen N = 77
Twin Towers 99.7 % / 352	100 % / 204	100 % / 147	99 % / 186	100 % / 162	100 % / 247	95 % / 100	100 % / 274	99 % / 76
Mauerfall 97 % / 343	97 % / 198	98 % / 144	96 % / 179	99 % / 160	98 % / 242	96 % / 91	98 % / 242	97 % / 75
Mondlandung 97 % / 343	96 % / 196	99 % / 145	97 % / 181	98 % / 158	97 % / 239	98 % / 93	97 % / 266	95 % / 73
Monroe 92 % / 324	95 % / 195	87 % / 129	91 % / 172	91 % / 149	93 % / 232	90 % / 86	93 % / 255	88 % / 68
Auschwitz-Birkenau 68 % / 239	62 % / 128	74 % / 110	66 % / 124	68 % / 111	68 % / 169	66 % / 63	68 % / 188	63 % / 49
«Che» Guevara 61 % / 213	49 % / 102	76 % / 112	54 % / 101	66 % / 108	58 % / 144	66 % / 63	60 % / 166	59 % / 46
Kim Phuc 61 % / 213	50 % / 104	74 % / 109	63 % / 118	56 % / 92	61 % / 151	62 % / 59	62 % / 172	53 % / 41
«Sprung in den Westen» 38 % / 133	24 % / 50	55 % / 82	29 % / 56	45 % / 73	36 % / 89	42 % / 40	42 % / 117	19 % / 15
Lenin 29 % / 102	14 % / 29	48 % / 72	22 % / 43	33 % / 55	25 % / 64	36 % / 35	29 % / 82	24 % / 19
«Tankman» 18 % / 64	6 % / 13	34 % / 51	15 % / 29	20 % / 33	16 % / 40	23 % / 22	20 % / 56	10 % / 8

8 Die Prozentangaben sind gerundet und beziehen sich auf die Anteile der plausiblen Nennungen in Bezug auf die Gesamtzahlen der Subgruppen.

	Gender (N = 352)		Anzahl der Bücher (N = 349)		Anzahl Fachsemester (N = 342)		Erstsprache (N = 351)	
«Junge Warschauer Ghetto» 16 % / 57	9 % / 19	25 % / 38	17 % / 33	14 % / 24	14 % / 36	21 % / 20	16 % / 44	15 % / 12
«Kapuzen- mann» Abu Ghraib 4 % / 13	1 % / 3	6 % / 10	4 % / 8	3 % / 5	3 % / 8	3 % / 3	3 % / 10	3 % / 3

Als erstes Ergebnis ist der fast durchgängige und deutliche, teils riesige Unterschied entlang der Genderkategorie für die eindeutigen Bildbeschreibungen zu nennen (in der Tabelle 2 sind die Unterschiede >10 % fett gedruckt). Besonders deutlich sind die Unterschiede bei dem Bild von Lenin (34 % Unterschied) und dem «Sprung in den Westen» (31 %), aber auch bei dem «Tankman» (28 %), «Che» Guevara (27 %) und Kim Phuc (23 %) sind die Unterschiede bemerkenswert hoch. Ebenso bemerkenswert und erklärungsbedürftig sind die Unterschiede beim Torhaus Auschwitz-Birkenau (12 %) und – wenngleich auf einem viel niedrigeren Niveau – bei dem «Jungen aus dem Warschauer Ghetto» (16 %). Dieser generelle Trend der Bildkenntnisunterschiede entlang der Genderkategorie kehrt sich nur bei der Monroe-Ikone um (8 % Unterschied). Dass sich solche Differenzen im Umgang mit Geschichte entlang der Geschlechterkategorie zeigen, wurde auch in repräsentativen Studien erhoben (z. B. von Borries, 1995, 362).

Im Vergleich dazu erscheinen die Unterschiede für die weiteren erhobenen Sozialdaten eher gering beziehungsweise im erwartbaren Rahmen zu liegen: So sind die plausiblen Zuordnungen bei einer höheren Anzahl von Büchern (mehr als 200) bei dem «Sprung in den Westen» (16 % Unterschied), «Che» Guevara (12 %) und Lenin (13 %) deutlich vorhanden, kleinere Unterschiede in umgekehrter Richtung gibt es bei dem «Jungen im Warschauer Ghetto» (9 %) und Kim Phuc (7 %). Nennenswert sind zudem wenige Unterschiede entlang der Unterscheidung der Erstsprachen: So wird der «Sprung in den Westen» deutlich häufiger plausibel zugeordnet bei denjenigen, die nur Deutsch als Erstsprache angeben (23 %). Auch bei dem «Tankman» ist ein deutlicher Unterschied zu verzeichnen (10 %), aber bei allen anderen Bildern erklärt diese Differenzkategorie kaum Unterschiede. Dies gilt in noch größerem Maße für die Semesteranzahl. Zwar gibt

es mit der höheren Semesterzahl meist einen höheren Anteil an plausiblen Zuordnungen, aber nur ein Bild mit einem größeren Unterschied (Lenin: 11 %).

4.3 Ansätze für qualitative Auswertung: Bildbeschreibungen als narrative Abkürzungen?

Als dritten Aspekt der Auswertung lässt sich mit Blick auf die aufgrund des halb-offenen Formates ebenfalls generierten qualitativen Daten ein weiteres Analysefeld für die Untersuchung historischer Kenntnisse anhand von Bildikonen aufzeigen. Obwohl Fotos und in diesem Fall Ikonen von sich aus noch keine historische Narration liefern und auch der verwendete Prompt («Das Bild zeigt ...») zu keiner expliziten Bildinterpretation auffordert, wird aus der qualitativen Analyse der plausiblen sowie auch nicht plausiblen Bildbeschreibungen ersichtlich, dass bei fast allen bekannten Fotografien in den Bildbeschreibungen unterschiedliche Ausdrücke verwendet werden, die als «narrative Abkürzungen» interpretiert werden können. So ist zwar der Ausdruck «Mauerfall» die dominante Bildbeschreibung für die diesbezügliche Fotografie, sodass diese in 81 Prozent der gegebenen Beschreibungen verwendet wird. Dennoch werden andere Beschreibungen wie Grenzöffnung verwendet (8 %). Des Weiteren wird in nahezu jeder zehnten Bildbeschreibung die Gleichsetzung mit der später erfolgten Wiedervereinigung vollzogen (9 %), was bereits auf eine Verknüpfung von zwei differenten Ereignissen und somit einen narrativen Ansatz hinweisen mag. In dieser Hinsicht ist auch aufschlussreich, was nicht oder nur kaum genannt wird. So wird die Deutung «friedliche Revolution» in diesem Zusammenhang nur zweimal erwähnt.

Noch spannender und auch diverser sind diese Bruchstücke verkürzter Narrationen bei den weniger geläufigen Bilddokumenten. In dieser Hinsicht weisen auch vermeintlich eindeutige Bilder wie das Porträt Ernesto Guevaras ganz unterschiedliche Bildbeschreibungen auf. Obwohl in mehr als zwei Dritteln der plausiblen Nennungen nur der Name «Che» und/oder Guevara genannt wird (73 %), gibt es auch Antworten, die auf andere Narrationen verweisen. So wird er in 12 Prozent der plausiblen Beschreibungen als Revolutionär, in weiteren 12 Prozent als Freiheits-, Befreiungs- oder Widerstandskämpfer und in 3 Prozent der Antworten als Guerillero beziehungsweise Rebellenführer gedeutet, wobei sich dies z. T. überlagert mit der Zuschreibung, ein «Kommunist» zu sein. Dies unterstreicht die zuvor von Schreiber vertretene These zur Uneindeutigkeit von Bildern beziehungsweise zur Unbestimmtheit der Rezeption eines Bildes (Schrei-

ber, 2004, 48–49). Darüber hinaus kann dies ein Hinweis auf die Einbettung der Fotos in dahinterliegende geschichtskulturelle Interpretationen sein, die teils kontrovers zueinander sind. Genauer kann diese These aufgrund des verwendeten Fragebogens und Prompts aber nicht ausdifferenziert werden.

5 Diskussion: Zur Bedeutung deklarativer Kenntnisse von Bildikonen

Zur Beantwortung der Fragen, inwiefern es Bilder gibt, die jede*r kennt und inwieweit sich Bildikonen für die Erfassung historischer Kenntnisse eignen, lassen sich drei Thesen formulieren.

1. Es gibt einerseits bestimmte Fotos, die vielfach in der Geschichtskultur reproduziert werden und die nahezu jede*r (der befragten Geschichtslehramtsstudierenden) plausibel zuordnen kann, die also bekannt sind. Andererseits trifft dies nur auf wenige der ausgewählten Ikonen zu: längst nicht alle häufig reproduzierten Bilder sind den Befragten bekannt. Vielmehr zeigen sich deutliche Unterschiede in der Bekanntheit der Bilder, die erklärungsbedürftig sind und sich weder einfach auf die klassischen Bilder des Geschichtsunterrichts zurückführen lassen noch auf die Ikonen der eigenen Lebenszeit. Es zeigt sich, dass sich anhand von Bildern historische Kenntnisse erfassen lassen, die zugleich auf Unterschiede zwischen verschiedenen Gruppen hinweisen. Es ist also zu vermuten, dass sich in den Kenntnissen von Bildikonen unterschiedliche Sozialisationsprozesse und historische Identitäten widerspiegeln. Diese Anzeichen deuten darauf hin, dass sich historische Kenntnisse besonders aus der außerschulischen und weiteren geschichtskulturellen Sozialisation im Allgemeinen speisen. Damit einher geht auch die Frage danach, inwieweit sich ähnliche Strukturen für andere Ikonen zeigen, wobei die Anzahl der möglichen Bildmaterialien nahezu überwältigend scheint und auch zu beachten ist, dass die Zuschreibung, was eine Ikone ist, einem Wandel unterliegen kann (Hamann, 2019, 7).
2. Erklärungsbedürftig ist vor allem der Genderzusammenhang, der in dieser Befragung zu Tage tritt und größer ist als alle anderen Zusammenhänge. Zu diskutieren wäre, ob es sich um einen Methodeneffekt handelt, der sich aus der Bildauswahl ergibt. Zu klären wäre dann aber die Frage, welche alternativen Bilder eine bessere Auswahl darstellen. Ein Hinweis könnte hier die repräsentative Befragung zum «Geschichtsbewusstsein Jugendlicher» lie-

fern. Demnach interessieren sich die Mädchen massiv weniger für «Politik- und Personengeschichte», dafür etwas mehr für «Kunst- und Objektgeschichte» sowie «Alltags- und Sozialgeschichte» (von Borries, 1995, 362). Wenn dies zutrifft, wäre zu diskutieren, welche Bilder des 20. und beginnenden 21. Jahrhunderts für diese Geschichtssektoren als Ikonen anzusehen sind.

3. Im Hinblick auf die Frage danach, inwieweit den Befragten die historischen Zusammenhänge von Bildikonen bekannt sind beziehungsweise in welche Zusammenhänge sie die Bildikonen einordnen, ergibt sich eine weitaus komplexere Situation. Auch wenn für eine quantitative Auswertung mit Blick auf historische Kenntnisse eine Vielzahl an möglichen Deutungen als plausibel anerkannt wurde, so zeigt sich darin, dass die Bildikonen in ganz unterschiedlichen und z. T. widersprüchlichen Narrationen verortet werden. Damit stehen zumindest manche Ikonen nicht für *eine* Deutung, sondern vermutlich für mehrere, wie an der Komplexität der Entstehungs- und Erinnerungskontexte von Ikonen exemplarisch für das Bild von Kim Phuc gezeigt wurde. Hier zeigt sich erneut, dass historische Kenntnisse nicht einfach deklaratives Wissen darstellen, sondern immer auch in Narrationen eingebettet sind und daher andere Wissensbereiche mit tangieren. Es ist daher eine weiterführende qualitativ-empirische Frage, welche Narrationen mit Bildikonen verknüpft werden und inwiefern ihnen eine subjektive Bedeutung zugesprochen wird.

Es ist abschließend darauf hinzuweisen, dass diese Befunde über historische Kenntnisse nicht dahingehend missverstanden werden sollten, dass (wieder) die reine Einführung in deklaratives Wissen zu diesen und in weitere Ikonen im Geschichtsunterricht zu forcieren ist. Es sollen auch keineswegs einigen Gruppen, die anhand von soziodemografischen Daten gebildet wurden, historische Kenntnisse im Allgemeinen abgesprochen werden. Hier sei nur exemplarisch auf die Problematik einer normativen Setzung von historischen Kenntnissen verwiesen. So wurden und werden historische Kenntnisse immer wieder dazu missbraucht, eine Unterscheidung in gebildete sowie weniger gebildete Menschen vorzunehmen und somit gewisse gesellschaftliche Positionierungen zu rechtfertigen (Stearns, 1998). Darüber hinaus lässt sich anfügen, dass eine normative Setzung von Kenntnissen auch immer schon gewisse Deutungen sowie Orientierungen impliziert.

Ganz im Gegensatz dazu ging es uns mit dieser empirischen Untersuchung von Bildkenntnissen darum, deskriptiv zu erheben, wie eine Generation ange-

hender Geschichtslehrer*innen sich mit Geschichtskultur – untersucht am Beispiel der Bildikonen – auseinandersetzt und welche der Deutungen als relevant und/oder plausibel angesehen werden. Sollten diese Befunde nicht reine Zufallsfunde sein, so ergibt sich bei den Kenntnissen von Bildikonen ein vielversprechendes Forschungsfeld, das in seiner Bedeutung für das historische Lernen noch zu diskutieren ist. Es stellt sich allerdings auch die Frage, welche hochschuldidaktischen Konsequenzen daraus zu ziehen sind.

Literatur

- Ammerer, Heinrich (2019). Herausforderungen und Probleme im Umgang mit visuellen Repräsentationen im Schulbuch. In Christoph Kühberger/Roland Bernhard & Christoph Bramann (Hrsg.), *Das Geschichtsschulbuch: Lehren – lernen – forschen* (S. 125–146). Münster, New York: Waxmann.
- Barricelli, Michele & Lücke, Martin (Hrsg.) (2012). *Handbuch Praxis des Geschichtsunterrichts* (Bde. 1–2). Schwalbach/Ts.: Wochenschau.
- Barthes, Roland (1990). Rhetorik des Bildes. In Roland Barthes & Dieter Hornig (Übersetzer), *Der entgegenkommende und der stumpfe Sinn* (S. 28–46). Frankfurt a. M.: Suhrkamp.
- Borries, Bodo von (1995). Das Geschichtsbewußtsein Jugendlicher. *Erste repräsentative Untersuchung über Vergangenheitsdeutungen, Gegenwartswahrnehmungen und Zukunftserwartungen von Schülerinnen und Schülern in Ost- und Westdeutschland*. Weinheim: Juventa.
- Borries, Bodo von (2012). Historische Kenntnisse – historische Kategorien – historische Kompetenzen. In Christoph Kühberger (Hrsg.), *Historisches Wissen: Geschichtsdidaktische Erkundung zu Art, Tiefe und Umfang für das historische Lernen* (S. 9–32). Schwalbach/Ts.: Wochenschau.
- Borries, Bodo von (2015). «Subjektorientiertes» Geschichtslernen ist nur als «identitätsreflektierendes» wünschenswert! In Heinrich Ammerer/Thomas Hellmuth & Christoph Kühberger (Hrsg.), *Subjektorientierte Geschichtsdidaktik* (S. 93–130). Schwalbach/Ts.: Wochenschau.
- Boßmann, Dieter (1977). «Was ich über Adolf Hitler gehört habe ...» *Auszüge aus 3042 Aufsätzen von Schülern und Schülerinnen aller Schularten der Bundesrepublik Deutschland*. Frankfurt a. M.: Fischer.
- Borfen, Elisabeth (2011). Marilyn: Diva und Sexikone der 50 Jahre. In Gerhard Paul (Hrsg.), *Bilder, die Geschichte schrieben: 1900 bis heute* (S. 148–155). Göttingen: Vandenhoeck & Ruprecht.
- Bunnenberg, Christian (2020). Bewegte Bilder, bewegende Bilder, Bewegung in Bildern – Gedanken zur Einführung in den Themenschwerpunkt. *Zeitschrift für Geschichtsdidaktik*, 19, 4–12. <https://doi.org/10.13109/zfgd.2020.19.1.4>.
- Chapman, Arthur (2021). Introduction: Historical knowing and the «knowledge turn». In Arthur Chapman (Hrsg.), *Knowing history in schools: Powerful knowledge and the powers of knowledge* (S. 1–31). London: UCL Press. <https://doi.org/10.14324/111.9781787357303>.

- Demke, Elena (2009). Mauerbilder – Ikonen im Kalten Krieg. In Cécile Engel (Hrsg.), *Bilder im Kopf. Ikonen der Zeitgeschichte* (S. 108–119). Köln: Dumont.
- Drechsel, Benjamin (2008). Der Tank Man: Wie die Niederlage der chinesischen Protestbewegung von 1989 visuell in einen Sieg umgedeutet wurde. In Gerhard Paul (Hrsg.), *Das Jahrhundert der Bilder. Bd. 2: 1949 bis heute* (S. 574–581). Göttingen: Vandenhoeck & Ruprecht.
- Günther-Arndt, Hilke (2006). Conceptual Change Forschung: Eine Aufgabe für die Geschichtsdidaktik? In Hilke Günther-Arndt & Michael Sauer (Hrsg.), *Geschichtsdidaktik empirisch: Untersuchungen zum historischen Denken und Lernen* (S. 251–278). Berlin: Lit.
- Hall, Stuart (2004). Das Spektakel des Anderen. In Juha Koivisto & Andreas Merckens (Hrsg.), *Ideologie, Identität, Repräsentation. Ausgewählte Schriften 4* (S. 108–166). Hamburg: Argument.
- Hamann, Christoph (2007). *Visual History und Geschichtsdidaktik: Bildkompetenz in der historisch-politischen Bildung*. Herbolzheim: Centaurus.
- Hamann, Christoph (2019). *Fotografien im Geschichtsunterricht. Visual History als didaktisches Konzept*. Frankfurt a. M.: Wochenschau.
- Hasberg, Wolfgang (2012). Canossa – ein Erinnerungsort in Bedrängnis. Zur Dignität historischen Wissens in Deutungsgeschäft Geschichte. In Christoph Kühberger (Hrsg.), *Historisches Wissen: Geschichtsdidaktische Erkundung zu Art, Tiefe und Umfang für das historische Lernen* (S. 213–217). Schwalbach/Ts.: Wochenschau.
- Heßler, Martina (2008). Der Mond ist ein Ami: Bilder der Mondlandung und die Inszenierung der Wissenschaft. In Gerhard Paul (Hrsg.), *Das Jahrhundert der Bilder. Bd. 2: 1949 bis heute* (S. 394–401). Göttingen: Vandenhoeck & Ruprecht.
- Jeismann, Karl-Ernst (1977). Didaktik der Geschichte. Die Wissenschaft von Zustand, Funktion und Veränderung geschichtlicher Vorstellungen im Selbstverständnis der Gegenwart. In Erich Kosthorst & Karl-Ernst Jeismann (Hrsg.), *Geschichtswissenschaft: Didaktik – Forschung – Theorie* (S. 9–33). Göttingen: Vandenhoeck & Ruprecht.
- Knoch, Habbo (2009). Der Junge aus dem Warschauer Ghetto – Ikone des Entsetzens. In Cécile Engel (Hrsg.), *Bilder im Kopf. Ikonen der Zeitgeschichte* (S. 40–49). Köln: Dumont.
- Körber, Andreas (2006). Bilder als Quellen – Bilder als Darstellungen: Bilder zum Rekonstruieren von Geschichte, Geschichte in Bildern de-konstruieren. In Johannes Kirschenmann & Ernst Wagner (Hrsg.), *Bilder, die die Welt bedeuten: «Ikonen» des Bildgedächtnisses und ihre Vermittlung über Datenbanken* (S. 169–195). München: Kopaed.
- Krammer, Reinhard (2008). Über Fotos, die jeder kennt. Warum Bilder Karriere machen. In Jan Patrick Bauer/Johannes Meyer-Hamme & Andreas Körber (Hrsg.), *Geschichtslernen – Innovationen und Reflexionen: Geschichtsdidaktik im Spannungsfeld von theoretischen Zuspitzungen, empirischen Erkundungen, normativen Überlegungen und pragmatischen Wendungen. Festschrift für Bodo von Borries zum 65. Geburtstag* (S. 165–182). Kenzingen: Centaurus.
- Kuckartz, Udo (2016). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Weinheim: Juventa.
- Kühberger, Christoph (2012). Konzeptionelles Wissen als besondere Grundlage für das historische Lernen. In Christoph Kühberger (Hrsg.), *Historisches Wissen: Geschichtsdidaktik*

- tische Erkundung zu Art, Tiefe und Umfang für das historische Lernen (S. 33–74). Schwalbach/Ts.: Wochenschau.
- Lahrem, Stephan (2008). Eine globale Protestikone des 20. Jahrhunderts. In Gerhard Paul (Hrsg.), *Das Jahrhundert der Bilder. Bd. 2: 1949 bis heute* (S. 234–241). Göttingen: Vandenhoeck & Ruprecht.
- Mayer, Ulrich/Hans-Jürgen Pandel/Gerhard Schneider & Bernd Schönemann (Hrsg.) (2006). *Wörterbuch Geschichtsdidaktik*. Schwalbach/Ts.: Wochenschau.
- Meyer-Hamme, Johannes (2016). Im Spannungsfeld historischer Uneindeutigkeit, notwendiger Exaktheit und sozialer Erwünschtheit. Eine Re-Analyse von Fragebogen- und Testkonstruktionen in quantitativen Studien zum Geschichtsbewusstsein und historischem Lernen. In Holger Thünemann & Meik Zülsdorf-Kersting (Hrsg.), *Methoden geschichtsdidaktischer Unterrichtsforschung* (S. 89–113). Schwalbach/Ts.: Wochenschau.
- Meyer-Hamme, Johannes (2022). Schülerwissen in der geschichtsdidaktischen Forschung. In Georg Weißeno & Beatrice Ziegler (Hrsg.), *Handbuch Geschichts- und Politikdidaktik*. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-29673-5_16-1.
- Münkler, Herfried (2009). 9/11 – Das Bild als Waffe in einer globalisierten Welt. In Cécile Engel (Hrsg.), *Bilder im Kopf. Ikonen der Zeitgeschichte* (S. 150–161). Köln: Dumont.
- Nitsche, Martin (2021). Ansätze und Methoden der Professionsforschung in Geschichte. In Sebastian Barsch & Burkhardt Barte (Hrsg.), *Motivation – Kognition – Reflexion. Schlaglichter geschichtsdidaktischer Professionsforschung* (S. 18–35). Frankfurt a. M.: Wochenschau.
- Paul, Gerhard (2006). Von der historischen Bildkunde zur Visual History. Eine Einführung. In Gerhard Paul (Hrsg.), *Visual History: Ein Studienbuch* (S. 7–36). Göttingen: Vandenhoeck & Ruprecht.
- Paul, Gerhard (Hrsg.) (2008). *Das Jahrhundert der Bilder. Bd. 2: 1949 bis heute*. Göttingen: Vandenhoeck & Ruprecht.
- Paul, Gerhard (2008). Der «Kapuzenmann»: Eine globale Ikone des beginnenden 21. Jahrhunderts. In Gerhard Paul (Hrsg.), *Das Jahrhundert der Bilder. Bd. 2: 1949 bis heute* (S. 702–709). Göttingen: Vandenhoeck & Ruprecht.
- Paul, Gerhard (Hrsg.) (2009). *Das Jahrhundert der Bilder. Bd. 1: 1900 bis 1949*. Göttingen: Vandenhoeck & Ruprecht.
- Paul, Gerhard (2009). Die Macht der Bilder in der Mediengesellschaft. In Cécile Engel (Hrsg.), *Bilder im Kopf. Ikonen der Zeitgeschichte* (S. 91–21). Köln: Dumont.
- Reiche, Jürgen (2009). Bilder im Kopf. In Cécile Engel (Hrsg.), *Bilder im Kopf. Ikonen der Zeitgeschichte* (S. 10–17). Köln: Dumont.
- Resch, Mario (2021). Professionelle Kompetenz von Geschichtslehrkräften. In Beatrice Ziegler & Georg Weissenso (Hrsg.), *Handbuch Geschichts- und Politikdidaktik*. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-29673-5_18-1.
- Rüsen, Jörn (1996). Historische Sinnbildung durch Erzählen: Eine Argumentationsskizze zum narrativistischen Paradigma der Geschichtswissenschaft und der Geschichtsdidaktik im Blick auf nicht-narrative Faktoren. *Internationale Schulbuchforschung*, 18(4), 501–543.

- Schreiber, Waltraud (2004). *Bilder aus der Vergangenheit – Bilder der Vergangenheit?* Neuried: ars una.
- Schroeder, Klaus/Deutz-Schroeder, Monika/Quasten, Rita & Schulze Heuling, Dagmar (2012). *Später Sieg der Diktaturen? Zeitgeschichtliche Kenntnisse und Urteile von Jugendlichen*. Frankfurt a. M.: Peter Lang.
- Stearns, Peter N. (1998). *Why study history?* [https://www.historians.org/about-aha-and-membership/aha-history-and-archives/historical-archives/why-study-history-\(1998\)](https://www.historians.org/about-aha-and-membership/aha-history-and-archives/historical-archives/why-study-history-(1998)) (14.04.2022).
- Waschik, Klaus (2009). Wo ist Trotzki? Sowjetische Bildpolitik als Erinnerungskontrolle in den 1930er Jahren. In Gerhard Paul (Hrsg.), *Das Jahrhundert der Bilder. Bd. 1: 1900 bis 1949* (S. 252–259). Göttingen: Vandenhoeck & Ruprecht.
- Wineburg, Sam (2018). *Why learn history? When it's already on your phone*. Chicago: University of Chicago Press. <https://doi.org/10.7208/chicago/9780226357355.001.0001>.
- Ziegler, Beatrice (2012). Historische Konzepte und Kompetenzmodelle: Ein empirischer Zugang zur Bedeutung von Wissen für Kompetenzen. In Christoph Kühberger (Hrsg.), *Historisches Wissen: Geschichtsdidaktische Erkundung zu Art, Tiefe und Umfang für das historische Lernen* (S. 135–152). Schwalbach/Ts.: Wochenschau.
- Zülsdorf-Kersting, Meik (2007). *Sechzig Jahre danach: Jugendliche und Holocaust. Eine Studie zur geschichtskulturellen Sozialisation*. Berlin: Lit.

Historisches Lernen mit der App in die Geschichte

Fallbeispiele einer Vorstudie zur Untersuchung historischer Lernprozesse und Narrationen von Studierenden im digitalen Medium

Alexandra Krebs

1 Einleitung

Der digitale Wandel vollzieht sich rasant, Corona-bedingtes Homeschooling beziehungsweise Homeoffice scheinen diese Entwicklung zu verstärken. Für den Geschichtsunterricht gilt jedoch weiterhin, dass «von ›Online-Lernumgebungen‹ noch keine Rede sein kann» (Bernhard & Kühberger, 2018, 436) und er in einer «konventionellen Form» zu verharren scheint (Bernhardt & Neeb, 2021, 343). Problematisch sind zudem die aktuellen digitalen Lernangebote, da sie historisches Lernen meist auf ein Quizformat reduzieren, das auf Kennen- und Auswendiglernen von Fakten abzielt (Bernhardt & Neeb, 2020, 80). Des Weiteren mangelt es immer noch an «fundierte[n] empirischen Erkenntnissen zu den durch digitale (Unterrichts-)Medien angestoßenen Prozessen historischen Lernens [...]» (Schwabe, 2020, 351).

Die in diesem Beitrag vorgestellte Studie zur App in die Geschichte¹ zielt daher darauf ab, den Forschungsstand in diesem Bereich zu erweitern, indem sie das Nutzer*innenverhalten in der App in Bezug auf die historischen Lernprozesse als auch Narrationen untersucht, um so nicht nur Aussagen über dieses Lernangebot, sondern auch über die Gestaltung weiterer digitaler Angebote sowie grundsätzlich über Lernprozesse und Narrationen im Digitalen treffen zu kön-

1 Die neue Version der App in die Geschichte (2020/21; <https://www.app-in-die-geschichte.de/>) wurde in Zusammenarbeit mit der Qualitäts- und UnterstützungsAgentur – Landesinstitut für Schule des Landes Nordrhein-Westfalen (QUA-LiS NRW), dem Hauptarchiv der v. Bodelschwingschen Stiftungen Bethel in Bielefeld und der Universität Paderborn erstellt.

nen. Damit ist das Promotionsprojekt zur Entwicklung, aber vor allem auch zur Erforschung der App im Zeitalter der Digitalität von besonderer Relevanz.

Der Beitrag verfolgt das Ziel, das Design der Studie zur App in die Geschichte anhand zweier Fallbeispiele, also Gruppen, aus der Vorstudie mit Studierenden vorzustellen und hinsichtlich ihrer Potenziale und Grenzen zur Diskussion zu stellen. Dafür werden im ersten Teil die theoretischen Grundlagen und Forschungsstände skizziert. Der zweite Teil erläutert die Konzeption der App und zeigt damit zugleich die Möglichkeiten dieses Erhebungsinstruments. Darauf aufbauend wird das aktuelle Design der Studie vorgestellt und anhand von Beispielen konkretisiert, sodass dieses Studiendesign abschließend kritisch reflektiert und Alternativen angedacht werden können.

Grundsätzlich lassen die ersten, hier vorgestellten Einblicke in die Vorstudie, im Gegensatz zu anderen aktuellen digitalen Lernangeboten und Untersuchungen zu ihnen (Bernhardt & Neeb, 2020), zumindest vermuten, dass die App beziehungsweise das App-Konzept durchaus Potenziale für historisches Lernen im digitalen Medium birgt.

2 Theoretische Grundlagen und Forschungskontext

Der Forschungskontext, in welchen die empirische Studie zur App in die Geschichte eingebettet ist, setzt sich zunächst aus verschiedenen Studien zu Re-Konstruktionskompetenzen beziehungsweise zu Lernprozessen im Umgang mit historischen Quellen sowie zur Erstellung eigener historischer Narrationen im Geschichtsunterricht in analogen sowie digitalen Lernsettings zusammen. Aus deren Erkenntnissen wurden einerseits zunächst Rückschlüsse für die Konzeption der App gezogen (siehe Abschnitt 3.1) sowie andererseits zentrale Forschungsfragen der Studie zur App entwickelt.

Als theoretische Grundlage der App in die Geschichte und damit auch der empirischen Studie über sie dienen der Regelkreis des historischen Denkens nach Jörn Rüsen (1983) und die darin ausgedrückte Re-Konstruktion von Vergangenheit und ihre Deutung als zentrale Operationen historischen Denkens, wie sie im FUER-Kompetenzmodell modelliert sind (Körber, Schreiber & Schöner, 2007). Bereits in verschiedenen empirischen Studien wurde die Kompetenz von Schüler*innen sowie Studierenden, historische Quellen auf der Basis einer historischen Fragestellung zu analysieren und interpretieren, um darauf aufbauend eine eigene historische Narration zu verfassen, untersucht. Die Ergebnisse fallen

überwiegend ernüchternd aus: In einer Unterrichtsstunde arbeiten die Lernenden meist weniger als zehn Minuten mit den historischen Quellen (Hodel & Waldis, 2007, 136), diese sind viel zu oft «stromlinienförmig», kommen dem Gebot der Multiperspektivität nicht nach (von Borries, 2013) und führen letztlich zu einer von der Lehrperson vorkonstruierten Geschichtserzählung. Die elementarsten Kriterien der Quellenkritik werden oft außer Acht gelassen (Hodel & Waldis, 2007, 138). Das Ergebnis ist ein «weitestgehend affirmativ [...] bzw. positivistisch[er]» Umgang mit historischen Quellen (Spieß, 2014, 104; Wineburg, 1991) sowie das häufige Fehlen einer «analytischen Distanz» (Schönemann, Thünemann & Zülsdorf-Kersting, 2011, 89). Weitere Untersuchungen belegen, dass Aufgabenstellungen und historisches Thema die narrativen Performanzen von Lernenden wie auch Studierenden beeinflussen (Hodel, Waldis, Zülsdorf-Kersting & Thünemann, 2013; Monte-Sano & La Paz, 2012; Waldis, Marti & Nitsche, 2015) und fordern unter anderem, «unterstützende Lernbegleitungen» (Nitsche & Waldis, 2016, 31; Waldis, Nitsche & Gollin, 2020, 108) zu entwickeln.

Vergleicht man hierzu den Forschungsstand zu digitalen Lernsettings, fällt zunächst auf, dass dieser «weiterhin recht dünn» (Schwabe, 2018, 422) ist, nur vereinzelt Studien vorliegen und diese sich lediglich auf einzelne wenige digitale Lernangebote (vgl. u. a. Schreiber & Bertram, 2018; Gautschi & Lücke, 2018) beziehen. Allerdings konnte auch bereits eine Studie aus dem Jahr 2014, ähnlich wie in den zuvor beschriebenen analogen Settings, konstatieren, dass das quellenkritische Arbeiten dabei oft vernachlässigt werde: «So beschränkt sich die Auswertung der schriftlichen Quellen oder Oral-History-Interviews oftmals auf eine reine Inhaltswiedergabe. Was Aussagewert oder auch -grenzen einer Quelle ausmachen, ist in den meisten Projekten nicht Teil der Schülerprodukte» (Oswalt, Aspelmeier & Boguth, 2014, 84). Andererseits konnte auch schon gezeigt werden, dass zum Beispiel sogenannte *computer-based learning*-Umgebungen mit fachspezifischen Aufgabenstellungen durchaus geeignet sind, um gezielt Strategien im Umgang mit historischen Quellen zu schulen, wie etwa mit der App «Sourcer's Apprentice» (2002):

«After the Sourcer's Apprentice training, students mentioned more source features in their notes and answered more source questions correctly on a transfer test relative to students not given the Sourcer's Apprentice training.»
(Britt & Aglinskas, 2002, 511)

Neben Studien zu Lernprozessen bezüglich des Umgangs mit historischen Quellen sowie Narrationen sind zugleich Studien zum Umgang mit Darstellungen und zur Gestaltung von Lernprozessen im digitalen Medium für die Studie zur App in die Geschichte relevant, da die Lernenden in der App ebenso mit historischen Darstellungen arbeiten sowie ihren Lernprozess selbst gestalten (siehe Abschnitt 3.1). Diese sowie die aus ihnen abgeleitete weiteren Forschungsfragen beziehungsweise Analysekatégorien werden jedoch für die exemplarische Analyse der beiden Fallbeispiele in diesem Beitrag nicht betrachtet und daher an dieser Stelle nicht weiter ausgeführt.²

Die Studie zur App in die Geschichte schließt an die zuvor genannten Untersuchungen an beziehungsweise führt sie fort. Die zentralen Fragen des Forschungsvorhabens beschäftigen sich damit, wie die Lernenden die App in die Geschichte nutzen, das heißt: Wie recherchieren sie historische Darstellungen, wie wählen sie sie aus und wie gehen sie mit ihnen um? Wie recherchieren sie historische Quellen, wie wählen sie sie aus und wie gehen sie mit ihnen um? Wie erstellen sie eine eigene historische Narration? Wie gestalten sie den (gemeinsamen) Lern- und Forschungsprozess? In diesem Beitrag wird hieraus insbesondere die Teilfrage untersucht, wie die Lernenden in der App mit historischen Quellen umgehen, genauer: Nach welchen Kriterien wählen die Lernenden welche Quellen aus und inwieweit beziehungsweise wie analysieren sie diese quellenkritisch (siehe Abschnitt 4.5)?

3 Die App in die Geschichte³

Die App in die Geschichte ist nicht nur ein digitaler Lernraum für den Geschichtsunterricht, sie dient zugleich als Erhebungsinstrument für die hier vorgestellte empirische Studie. Daher ist es nötig, die Konzeption der App und die verschiedenen Tools näher zu beschreiben, um so auch die Potenziale des Erhebungsinstrumentes für die empirische Forschung deutlich zu machen.

2 Exemplarisch sei jedoch auf die Studien von Hodel (2013), Schwabe (2012) und Wineburg (2018) hingewiesen.

3 Eine detailliertere Beschreibung der App-Konzeption findet sich bei Krebs & Meyer-Hamme (2021) sowie der einzelnen Tools bei Krebs (2021).

3.1 Methodisch-didaktische Konzeption

Die Konzeption der App in die Geschichte versucht, auf die beschriebenen defizitären Studienergebnisse zu antworten und zugleich den damit verbundenen Forderungen gerecht zu werden. Aus diesem Grund führt die App unter anderem Prinzipien forschend-entdeckenden Lernens und Projektlernens sowie Archivarbeit in einer digitalen Lernumgebung zusammen und macht dadurch zudem Potenziale digitaler Medien für ein fachspezifisches historisches Lernen fruchtbar. Der Lernprozess innerhalb der App sowie die Aufgaben orientieren sich dabei am Regelkreis historischen Denkens nach Jörn Rüsen (1983) und zielen darauf ab, Kompetenzen historischen Denkens zu fördern (Körper, Schreiber & Schöner, 2007), indem die Lernenden ein historisches Thema wählen und eigene Fragestellungen entwickeln, diese anhand von selbst recherchierten historischen Quellen erforschen, um schließlich eine eigene historische Narration als Antwort auf ihre Fragen zu verfassen. In Bezug auf die bereits zitierten Studien soll den Lernenden in der App dadurch ermöglicht werden, intensiv mit historischen Quellen aus dem digitalen Archiv zu arbeiten, welches angelehnt an das analoge Herkunftsarchiv strukturiert ist. Die Quellen sind daher unter anderem multiperspektivisch sowie teils unvollständig und unleserlich. Lernende können so quellenkritisch und reflektiert historisch forschen, sodass plurale Narrationen entstehen, die eben keine vorkonstruierte Geschichtserzählungen der Lehrperson oder Schulbuchautor*innen sind und die es lohnt im Klassenraum zu verhandeln. Zudem wird dieser Lernprozess in der App nicht kleinschrittig vorgegeben, sondern kann von den Lernenden selbst gestaltet werden. Die App bietet im Sinne des Scaffolding unterstützende Hinweise und Hilfen (u. a. Recherchetipps, Methodenhinweise und Formulierungshilfen) an, auf welche die Nutzer*innen gegebenenfalls zugreifen können. Hiermit versucht die App der Forderung nach einer grundsätzlichen Offenheit historischen Lernens gerecht zu werden (von Borries, 1998, 78–96; Heuer, 2011, 43–58). Da es sich bei der App in die Geschichte außerdem um eine variable Lernplattform handelt, kann jede Gruppe die Gestaltung des eigenen Lernraums selbst vornehmen, indem sie aus verschiedenen Modulen (mit unterschiedlichen Forschungsaufträgen, Themen, Archiven), die für sie passenden auswählt. Diese sind für jeweils unterschiedliche außerschulische Lernorte (i. d. R. Archive) mit gegebenenfalls unterschiedlichen Arbeitsaufträgen oder völlig frei gestaltet.⁴

4 Es wurden bereits zwei Module für die App entwickelt, weitere sind geplant. Vgl. hierzu Krebs und Meyer-Hamme (2021) sowie Krebs (2020).

Im Rahmen der Studie nutzen jedoch alle Lerngruppen zur besseren Vergleichbarkeit den «Story-Modus Bethel». Darin wird zu Beginn ein Forschungsauftrag an die Lernenden adressiert, der am sogenannten Task-Based-History-Learning-Konzept ausgerichtet ist (Körber, Gärtner, Hartmann & Stork, 2021). Abgeleitet vom Task-Based-History-Learning-Konzept aus der englischen Fachdidaktik meint dieses «ein zugleich problemorientiertes, projektartig organisiertes geschichtsdidaktisches Lernkonzept», in welchem «echte Herausforderungen an authentischem Material aus der Lebenswelt der Lernenden» (ebd., 211) diese zu eigenen Lern- und Denkprozessen motivieren. Die ausgewählten Fälle beziehungsweise Problemfelder sollen daher möglichst realistisch beziehungsweise lebensweltlich sein und Gelegenheit bieten, «die Erfahrungen bei und mit ihrer probeweisen Anwendung in geschütztem als auch organisiertem Rahmen zu reflektieren» (ebd., 210). Der Forschungsauftrag des Story-Modus Bethel dreht sich deshalb um eine Debatte im Bielefelder Stadtrat über die Benennung einer Straße, die zwar fiktiv ist, jedoch auf einer tatsächlichen geschichtskulturellen beziehungsweise wissenschaftlichen Kontroverse basiert und zugleich anderen aktuellen Kontroversen um Straßennamen ähnelt: Die Parteien im Stadtrat können sich nicht einigen, nach welcher historischen Persönlichkeit die neue Straße im Bielefelder Stadtteil Bethel benannt werden soll. Daher soll untersucht werden, welche Person beziehungsweise ob eine der Personen überhaupt geeignet ist, Namensgeber der Straße zu werden.⁵ Die App-Nutzer*innen müssen nun in die Rolle von Historiker*innen schlüpfen und nach wissenschaftlichen Kriterien die Geschichte Bethels mithilfe digitalisierter Archivalien erforschen. Hierzu besuchen sie das Hauptarchiv der v. Bodelschwingschen Stiftungen Bethel in Bielefeld beziehungsweise das digitale Archiv der App in die Geschichte.

3.2 Tools in der App in die Geschichte

Der so gestaltete Forschungsauftrag wirft ein gegenwärtiges Problem auf, das es im Team zu lösen gilt. Die Arbeitsschritte in der App sehen vor, dass diese Forscher*innenteams im ersten Schritt im Aufgabentool gemeinsam Vorüberlegungen anstellen und ihre Forschung planen. D. h. sie führen zunächst Hintergrundrecherchen zum historischen Thema, zum Forschungsauftrag, zu Bethel

5 Zur gegenwärtigen Kontroverse der geschichtswissenschaftlichen Forschung bzw. geschichtskulturellen Debatte vgl. u.a. Hochmuth (1997); Benad, Schmuhl & Stockhecke (2016b); Degen (2014); Kalischek (2019) und Melter (2020).

und so weiter durch und formulieren eigene Leitfragen für ihre Recherche im digitalen Archiv. Ihre Überlegungen halten sie gemeinsam im Aufgabentool der App mithilfe verschiedener Etherpads fest. Die im digitalen Archiv recherchierten Quellen müssen sie analysieren, kontextualisieren und interpretieren, um abschließend die Ergebnisse in einem Gutachten für die Fraktionen des Stadtrates darzustellen und damit ihre Forschungsfrage(n) zu beantworten.

Das digitale Archiv der App ist angelehnt an das Hauptarchiv der v. Bodelschwinghschen Stiftungen in Bethel und nach verschiedenen Akten mit deren jeweiliger Signatur sortiert. Dadurch bleiben die Materialien beziehungsweise Quellen in der App im Gegensatz zu den meist geglätteten, also gekürzten, vorausgewählten, sinnvoll angeordneten und vereinfachten Quellen in Geschichtsschulbüchern multiperspektivische und widersprüchliche Zeugnisse der Vergangenheit (Lange & Lux, 2004, 47). Sie machen so in besonderer Weise den Konstruktcharakter von Geschichte deutlich und ermöglichen, den Lernenden ähnlich wie Historiker*innen zu arbeiten. Stoßen die Lernenden im digitalen Archiv auf eine für ihre Forschungsfragen relevante Quelle, fügen sie diese ihrer Merkleiste beziehungsweise der Zeitleiste hinzu. Die eigenständige Recherche und die notwendigen Auswahlentscheidungen sind zentraler Bestandteil des Lernprozesses, da sie zwangsläufig bereits mögliche historische Deutungen und Urteile am Ende des Forschungsprozesses beeinflussen. Die Kriterien der Quellenauswahl sollen daher auch im Journal festgehalten werden (siehe Abschnitt 4.7.1 sowie 4.8.1).

Eine erste Analyse und Auswertung der Quellen erfolgt in der App mithilfe einer Zeitleiste. Jedes Team erstellt eine eigene digitale Zeitleiste und fügt in diese nach und nach Quellen aus der Merkleiste ein. Mithilfe farbiger Markierungen und Notizen können die Nutzer*innen direkt auf der jeweiligen Quelle ihre Analyseergebnisse, Fragen oder Interpretationsansätze festhalten. Diese multimediale Zeitleiste kann zusätzlich mit Überschriften, Bildern, Texten oder Videos ergänzt werden. Sie verknüpft so die ausgewählten Quellen mit den ersten Analyse- und Interpretationsergebnissen. Zugleich dient die Zeitleiste der Vorstrukturierung für das Gutachten, welches die Nutzer*innen im Journal der App verfassen. Sie stellen darin ihre Ergebnisse in einem zusammenhängenden Text dar, ergänzen diesen mit Elementen aus der Zeitleiste, wie zum Beispiel zentralen Quellen oder Textstellen. Dieses Gutachten soll abschließend eine Bewertung beinhalten, in der die Lernenden eine Empfehlung für die Benennung der neuen Straße in Bethel abgeben.

4 Die empirische Studie zur App in die Geschichte

Die zentralen Fragen des Forschungsvorhabens beschäftigen sich damit, wie die Lernenden die App Geschichte nutzen, das heißt, wie sie in der App Geschichte erzählen und welche Rückschlüsse über historische Lernprozesse im digitalen Medium sich daraus ableiten lassen. Exemplarisch wird in diesem Beitrag ein Einblick in die Daten gegeben, und zwar im Hinblick auf die Teilfrage der Studie, wie die Nutzer*innen in der App mit historischen Quellen umgehen, insbesondere, wie sie historische Quellen recherchieren, auswählen und inwieweit beziehungsweise wie sie diese quellenkritisch analysieren.

Die App stellt hierfür ein innovatives Erhebungsinstrument dar, weil sich mit ihr sowohl die historischen Lernprozesse als auch die daraus resultierenden historischen Narrationen untersuchen lassen, indem quantitative (Logfiles, siehe Abschnitt 4.3 und 4.4) sowie qualitative Daten (Produkte aus der App wie u. a. digitale Zeitleisten und historische Narrationen), die während des Einsatzes der App im Unterricht entstehen, erhoben und ausgewertet werden.

4.1 Studiendesign

Ein *Mixed-Methods-Design*, also die Kombination beziehungsweise Integration qualitativer sowie quantitativer Methoden, scheint für die beschriebenen Forschungsfragen besonders geeignet, da sich durch die Kombination dieser verschiedenen Zugriffe komplexe Prozesse wie historisches Denken womöglich besser erfassen lassen (Prinz & Thünemann, 2016, 251). *Mixed-Methods-Ansätze* können dazu beitragen, ein genaueres Bild und ein tieferes Verständnis des Untersuchungsgegenstands zu gewinnen und zugleich mögliche «blinde Flecken» aufzudecken und Grenzen des einen methodischen Zugangs mit dem jeweils anderen aufzulösen (Köster & Thünemann, 2019, 29). Unter dem Begriff *Mixed Methods* finden sich aktuell in der Literatur bis zu 72 verschiedene Kombinationsmöglichkeiten, die sich jeweils in ihren verschiedenen Dimensionen (Implementation, Priorität, Integration, Rolle der theoretischen Perspektiven) unterscheiden (Prinz & Thünemann, 2016, 232–233). Entscheidend für die Auswahl sollte vor allem sein, dass das gewählte Design zur Beantwortung der Forschungsfragen angemessen ist. Die Vorstudie zur App in die Geschichte dient daher auch dazu, eine geeignete *Mixed-Methods-Variante* für die noch folgende Hauptstudie zu entwickeln.

Eine Variante, die im Folgenden an zwei Fallbeispielen gezeigt und danach zur Diskussion gestellt wird, ist ein sogenanntes sequenzielles *Mixed-Methods-*

Design, genauer auch als *Explanatory Design* bezeichnet (ebd., 233–234). Hierfür werden zunächst quantitative Logfiledaten der App-Nutzer*innen (u. a. hinsichtlich der Abfolge, der Aufenthaltsdauer und der Seitenansichten) analysiert und darauf aufbauend Hypothesen verschiedener Nutzertypen gebildet (siehe Abschnitt 4.3), was eine Strukturierung der qualitativen Daten ermöglicht. In einem zweiten Schritt werden die qualitativen Daten, also die Produkte aus der App (u. a. das Journal und die Zeitleisteneinträge), mithilfe der inhaltlich-strukturierenden Inhaltsanalyse (Kuckartz, 2018) mit MAXQDA (u. a. hinsichtlich des Umgangs mit historischen Quellen und Formen historischen Erzählens) untersucht. Dabei werden wiederum Ergebnisse der quantitativen Analyse für die Interpretation hinzugezogen, um ein tieferes Verständnis bezüglich der verschiedenen Auswertungskategorien zu gewinnen. Hierdurch soll die zu Beginn festgelegte quantitative Typisierung überprüft beziehungsweise revidiert werden. Im nächsten Abschnitt wird diese Designvariante anhand zweier Fallbeispiele aus der Vorstudie konkretisiert.

Für diese Vorstudie konnten 45 Studierende der Lehrämter Geschichte im Masterstudium gewonnen werden.⁶ Die Teilnehmenden nutzten die App gemeinsam in Kleingruppen mit jeweils drei bis vier Personen, sodass insgesamt Daten von 14 Teams erhoben werden konnten. Davon forschten sechs Gruppen zur Person Friedrich v. Bodelschwingh, die anderen acht zu Elisabeth Philipp. Zur besseren Vergleichbarkeit der Fälle werden in diesem Beitrag exemplarisch nur die Gruppen betrachtet, die zu Friedrich v. Bodelschwingh geforscht haben ($n = 6$).⁷

6 Personenbezogene Sozialdaten wie z. B. Alter und Geschlecht der Teilnehmenden wurden nicht erhoben, da dies im Rahmen der App-Nutzung nicht datenschutzkonform ist und diese Daten zudem in Bezug auf die Fragestellungen und Zielsetzungen der (Vor-)Studie nicht relevant sind.

7 Friedrich v. Bodelschwingh d. J. leitete ab 1910 die Betheler Anstalten bis zu seinem Tod im Jahre 1947. Prägte zunächst noch der «Mythos Bodelschwingh» das Bild in der öffentlichen Wahrnehmung, wird heute vor allem seine Rolle in der Zeit des Nationalsozialismus kontrovers diskutiert. So stehen u. a. seine prinzipielle Staatstreue und damit auch seine Treue zum NS-Regime sowie seine positive Haltung zur Eugenik bzw. Zwangssterilisation seinem Einsatz gegen die NS-«Euthanasie»-Aktionen, die teils als «unsichtbare Form des Widerstands» bewertet werden, gegenüber. Vgl. u. a. Benad (1997) sowie Benad, Schmuhl und Stockhecke (2016a).

4.2 Umfang der Daten

In der App wird das Verhalten der Nutzer*innen während der gesamten Zeit ihres Aufenthalts auf der Webseite in Form von Logfiles mithilfe der Software Matomo getrackt (siehe unten). Diese Daten sind sehr umfangreich, da die Nutzungszeit der Gruppen zwischen acht und 27 Stunden beträgt ($M = 17$ Stunden, $SD = 6$), während eines Zeitraumes von ca. vier bis sechs Wochen (siehe Abbildung 1).

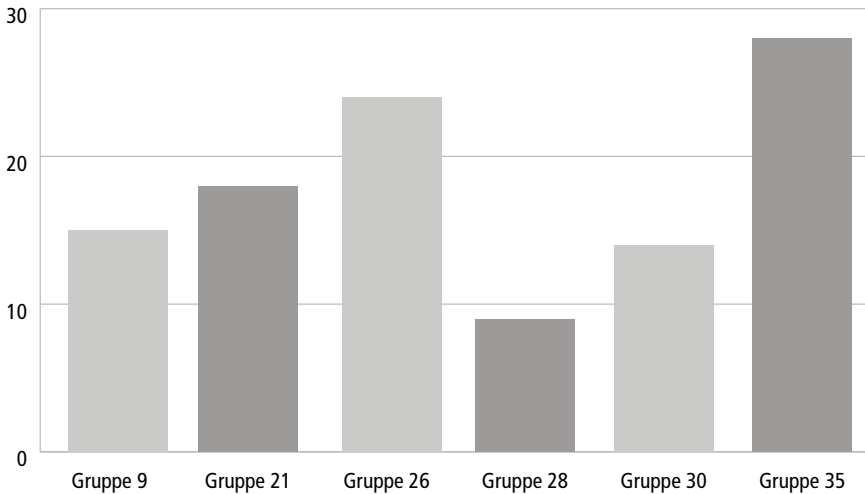


Abbildung 1: Aktive Nutzungszeit der Gruppen in der App in die Geschichte

Die erhobenen qualitativen Daten bestehen sowohl aus den von den Gruppen verfassten Texten (Vorüberlegungen, Planung und Journal) sowie der Merkliste mit den ausgewählten Quellen als auch den Einträgen in der Zeitleiste, die sich teilweise aus selbstverfassten Texten in Form von Annotationen und digitalisierten Archivalien zusammensetzen. Die Materialgrundlage der qualitativen Daten ist daher sehr umfangreich. Die ausgewählten Gruppen haben jeweils zwischen sieben und 27 Textseiten⁸ ($M = 13$, $SD = 7$) verfasst (siehe Abbildung 2), zwischen fünf und 83 Archivseiten der Merkliste ($M = 30$, $SD = 28$) hinzugefügt und zwischen acht und 22 Einträge in der Zeitleiste ($M = 11$, $SD = 4$) vorgenommen (siehe Tabelle 1).

⁸ Hierfür wurde die Wortanzahl aller geschriebenen Texte addiert. 250 Wörter in der App entsprechen etwa einer Textseite.

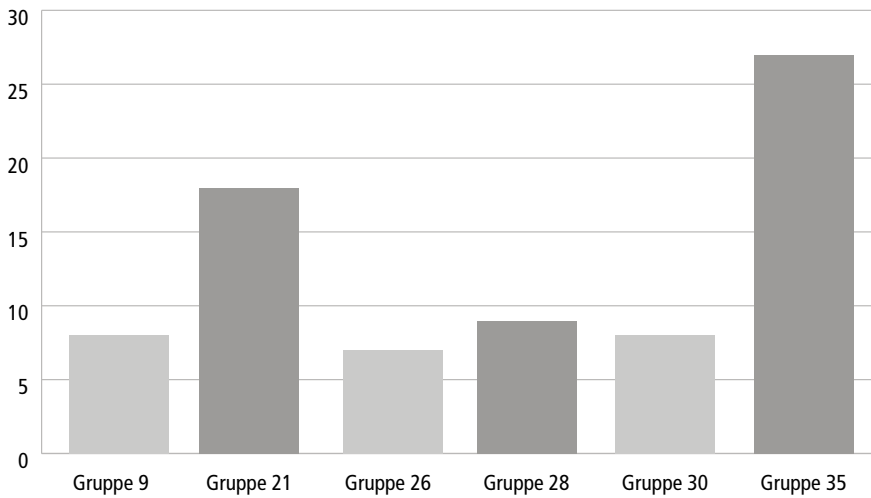


Abbildung 2: Die gesamte Anzahl der von den Gruppen verfassten Textseiten (Vorüberlegungen, Planung und Journal) (zur Umrechnung Seiten/Wörter s. Anm. 14)

Tabelle 1: Seitenanzahl der ausgewählten Quellen in der Merkliste je Gruppe und Anzahl der Einträge in der Zeitleiste je Gruppe. Eine Quelle kann eine oder mehrere Seiten umfassen und ein Zeitleisteneintrag kann keine, eine oder mehrere Quellen enthalten.

Gruppe	Seitenanzahl der ausgewählten Quellen in der Merkliste	Anzahl der Einträge in der Zeitleiste
Gruppe 9	22	11
Gruppe 21	5	22
Gruppe 26	53	9
Gruppe 28	9	8
Gruppe 30	8	8
Gruppe 35	83	12

4.3 Die Auswertung der Daten

Für die Hauptstudie ist geplant, die qualitativen Daten mithilfe der inhaltlich-strukturierenden Inhaltsanalyse nach Kuckartz mit MAXQDA zu analysieren und zu interpretieren. Die Vorstudie soll nicht nur dazu dienen, das Studiendesign zu erarbeiten, sondern auch eine erste Version des Kategoriensystems für die Ana-

lyse der qualitativen Daten der Hauptstudie zu erstellen. Diese wird derzeit noch anhand von deduktiven und induktiven Kategorien entwickelt. In den folgenden Fallbeispielen werden hieraus die Kategorien «Quellenrecherche», «Quellenauswahl» und «Umgang mit Quellen» exemplarisch in den Blick genommen (siehe Tabelle 2).

Tabelle 2: Beschreibung der ausgewählten Kategorien bezüglich der qualitativen (QUAL) und quantitativen Daten (QUAN)

Kategorie	QUAL (v. a. Texte)	QUAN (v. a. Logfiles)
«Quellenrecherche»	Aussagen und Reflexionen über die Recherche (u. a. eigene Leitfragen, Kriterien, Vorgehen)	Umfang der Recherche (d. h. Anzahl der Seitenansichten im digitalen Archiv)
«Quellenauswahl»	Aussagen und Reflexionen über die Quellenauswahl (u. a. Quellengattungen, [Multi-]Perspektivität, Entstehungskontext, Autor, Adressat)	Umfang der Quellenauswahl (d. h. Anzahl der ausgewählten Quellen in der Merkleiste)
«Umgang mit den Quellen»	(Quellenkritische ⁹) Aussagen und Reflexionen u. a. zu Widersprüchen zwischen den Quellen, historischen Kontexten, Quellengattungen, Perspektivität, Partialität, Selektivität, Passung der Quellen für die Fragestellung; sowie Informationsentnahme	Abfolge der Projektschritte hinsichtlich der Recherche, Auswertung und Darstellung der Ergebnisse (d. h. Aufenthalt/Aktivitäten im zeitlichen Verhältnis zueinander in den verschiedenen Tools: Aufgabentool, Archiv, Zeitleiste Journal)

Nach diesen Kategorien sollen ebenso die Logfiledaten analysiert werden (siehe Tabelle 2). Zunächst wird hierzu das Webanalyse-Tool Matomo (ehemals Piwik) verwendet, eine datenschutzkonforme Alternative zu Google Analytics. Diese Open-Source-Software wird klassischerweise für die Analyse des Nutzer*innenverhaltens in Onlineshops zu Marketingzwecken genutzt, um damit zum Beispiel eine Webseite optimal für verschiedene Verbraucher*innen und deren Interessen zu gestalten und so den Verkauf bestimmter Produkte zu steigern. Die Software kann jedoch, zumindest teilweise, auch für die Analyse des Nutzer*innenverhaltens in der App in die Geschichte eingesetzt werden. Mit ihr werden unter anderem die Logfiles im Hinblick auf die Anzahl und Dauer der Besuche

9 Quellenkritik bzw. quellenkritischen Reflexionen bedeutet nicht, dass die Quellen «lügen», sondern vielmehr die Standortgebundenheit einer Quelle zu untersuchen: «Davon ausgehend gilt es festzustellen, welchen Wert die Quelle für eine bestimmte Fragestellung hat bzw. welche Fragen sinnvoll an die Quelle gerichtet und mit ihrer Hilfe beantwortet werden können» (Spieß, 2014, 106).

oder Seitenaufrufe in der App sowie das Verhalten auf den verschiedenen Seiten oder Tools untersucht. Allerdings lassen sich mit Matomo nicht alle notwendigen Metriken für die Forschungsfragen beziehungsweise Kategorien analysieren, da diese für Marketingzwecke nicht erforderlich sind. Außerdem sind die Möglichkeiten der Darstellung der Daten beziehungsweise der Analyseergebnisse stark begrenzt. Für diese Zwecke werden in der Studie zur App daher von der Autorin selbst implementierte Python-Programme eingesetzt.

4.4 Fallbeispiele

Zwei Fallbeispiele geben im folgenden Teil einen exemplarischen Einblick in das Datenmaterial, sodass daran anschließend das gewählte Studiendesign kritisch reflektiert werden kann. Aus den verschiedenen Untersuchungsschwerpunkten werden hierfür die Quellenrecherche beziehungsweise -auswahl sowie der weitere Umgang mit historischen Quellen in den Blick genommen. Diese Aspekte stellen nämlich einen zentralen Bestandteil von Re-Konstruktionsprozessen dar und sind auch für den Lernprozess in der App wichtige Schwerpunkte. Für den exemplarischen Fallvergleich in diesem Beitrag werden nach dem Prinzip der maximalen Kontraste zwei Gruppen ausgewählt, die in der quantitativen Logfileanalyse die beiden äußersten Extreme darstellen. Grundlage bildet die Anzahl der Seitenansichten im Archiv, da sie zumindest oberflächlich darüber Auskunft gibt, wie intensiv die Gruppen historische Quellen recherchiert haben, das heißt, wie viele Seiten sie sich im Archiv angesehen haben.

4.5 Anzahl der Seitenansichten im Archiv

Anhand einer Logfileanalyse kann für die einzelnen Gruppen, die sich jeweils aus drei bis vier Personen zusammensetzen, die Anzahl der Seitenansichten im Archiv bestimmt werden. Bezüglich des zur Verfügung stehenden Materials ist anzufügen, dass sich im Archiv derzeit zwölf digitalisierte Akten mit einem Umfang von insgesamt 862 Blättern befinden. Jeder Klick auf ein neues Blatt in einer Akte wird dabei als neue Seitenansicht getrackt.

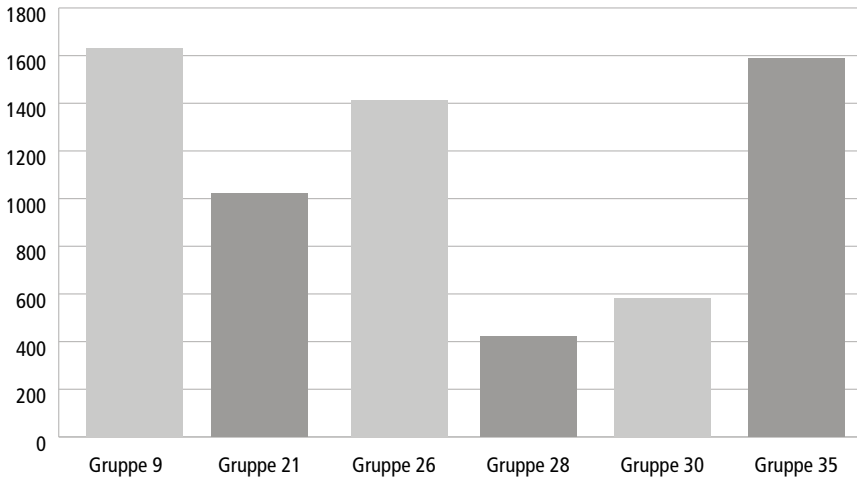


Abbildung 3: Darstellung der Anzahl der Seitenansichten im Archiv der verschiedenen Gruppen

Grundsätzlich ist festzustellen, dass die Anzahl der Seitenansichten bei allen ausgewählten Gruppen recht groß ist (>400). Es ist daher davon auszugehen, dass sich alle Gruppen in einem gewissen Umfang mit den Archivalien im digitalen Archiv auseinandergesetzt haben (siehe Abbildung 3). Vergleicht man allerdings die Anzahl der Seitenansicht zwischen den Gruppen, fallen beachtliche Unterschiede auf: Vier der Gruppen verbuchen mit ca. 1000 bis 1600 Ansichten eine doppelt bis vierfach so große Anzahl wie zwei andere Gruppen, die nur auf ca. 400 bis 600 Ansichten kommen. Da diese Unterschiede beträchtlich sind, scheint es sinnvoll, sie näher zu untersuchen. Für den kontrastiven Fallvergleich in diesem Beitrag sollen daher Gruppe 9, die mit 1634 Seitenansichten das Maximum darstellt (daher «Max»-Gruppe), und Gruppe 28, die mit 418 Seitenansichten das Minimum darstellt (daher «Min»-Gruppe), im Fokus der folgenden qualitativen Analyse stehen, bei der sowohl die Quellenrecherche und -auswahl als auch der Umgang mit den Quellen in den beiden historischen Narrationen in den Blick genommen werden.

4.6 Gruppe 9 – die «Max»-Gruppe

4.6.1 Recherche und Quellenauswahl

Gruppe «Max» wählt insgesamt neun Quellen, also digitalisierte Archivalien, aus dem Archiv aus und analysiert sowie interpretiert diese sowohl in der Zeitleiste als auch im Journal. Wie die Gruppe zu ihrer Auswahl gelangt ist, lässt sich anhand des Textes im Journal rekonstruieren. Die Gruppenmitglieder haben darin kollaborativ einen Text geschrieben sowie ergänzend hierzu ausgewählte Quellen zusammengestellt. Sie begründen die Auswahl ihrer Quellen neben dem Inhalt vor allem mit der Quellengattung:

«Ebenfalls ist festzuhalten, dass es sich bei den ausgewählten Quellen häufig um Briefwechsel mit befreundeten Kollegen wie beispielsweise Pastor Paul Braune handelt. Aufgrund der Quellengattung ist es am wahrscheinlichsten, dass Friedrich von Bodelschwingh in diesen Dokumenten seine Meinung weitgehend ungefiltert wiedergegeben hat. Zudem finden sich an einigen Stellen dieser privaten Korrespondenzen Aufforderungen, Aussagen vertraulich zu behandeln oder die entsprechenden Dokumente zu vernichten.» (Journal LG 9, S. 1: 1332)

Die Gruppe macht damit ihre quellenkritischen Reflexionen in Bezug auf die Quellengattung («private Korrespondenz») und die Adressaten der Briefe («befreundete Kollegen») für die Auswahl der Quellen fruchtbar. Sie erkennt, dass je nachdem, an wen v. Bodelschwingh seine Schreiben richtet, er unterschiedlich offen seine persönliche Haltung zum Ausdruck bringt. Diese Erkenntnis nutzt die Gruppe für die Auswahl der Quellen. Sie wählt sowohl private Briefe an Freunde (Pastor Paul Braune) als auch offizielle Anfragen und Briefe an Behörden sowie führende NS-Funktionäre (u. a. an Reichsmarschall Hermann Göring) aus (siehe Abbildung 4).

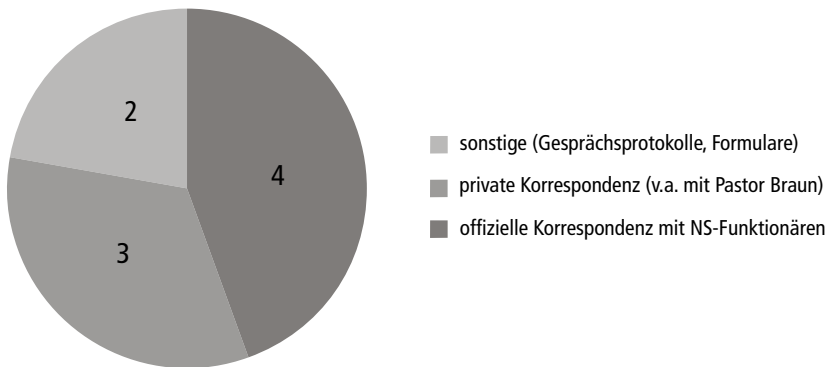


Abbildung 4: Verteilung der ausgewerteten Quellen der Gruppe 9 («Max») nach Quellengattungen

Es findet hier also ein Wechsel zwischen Quellenkritik und Quellenauswahl statt. Diese Deutung kann durch die wesentlich höhere Anzahl an Seitenansichten im digitalen Archiv gestützt werden. Die Gruppe ist bei der Quellenrecherche und -analyse nicht linear vorgegangen, sondern vielmehr rekursiv beziehungsweise iterativ und hat daher eine große Anzahl an Seitenansichten benötigt. Dies kann auch anhand der Abfolge der Arbeitsschritte anhand weiterer quantitativer Daten gezeigt werden (siehe *Abbildung 5*).

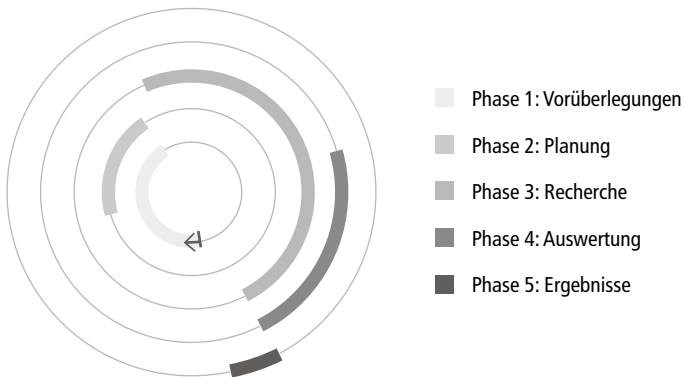


Abbildung 5: Abfolge der Projektschritte von Gruppe 9 («Max») in relativer Länge (absolute Gesamtzeit: 849 Minuten)

Die Grafik zeigt die Abfolge der fünf Projektschritte, die im Lernprozess in der App vorgegeben sind, sowie die relativen Längen zueinander. Der Vollkreis entspricht dabei der Gesamtlänge des jeweiligen Projekts. Die Daten lassen sich sowohl aus den Logfiles als auch aus den Videos der *Etherpads* ableiten, in denen der Schreibprozess der Texte protokolliert wird.¹⁰ Hieran zeigt sich, dass die «Max»-Gruppe zu einem großen Teil die Auswertung (Phase 4) während der Recherche (Phase 3) betrieben hat und dadurch die Schritte miteinander verzahnen und wechselseitig aufeinander beziehen konnte (siehe Abbildung 5).

4.6.2 Der Umgang mit historischen Quellen in der historischen Narration

Anhand des privaten Briefwechsels mit Pastor Paul Braune und weiterer vertraulicher Briefe beschreibt die Gruppe v. Bodelschwings Handlungen gegen die NS-«Euthanasie»-Programme und deutet an, dass diese als «Akt des Widerstands gegenüber des NS-Regimes» (Journal LG 9, S. 3: 402) beurteilt werden könnten. In einer Aussage Bodelschwings zu den nationalsozialistischen «Euthanasie»-Aktionen in einem Brief an einen nicht näher bekannten Ministerialrat erkennt sie jedoch auch einen möglichen Widerspruch zu ihrer Deutung:

«Offensichtlich plädierte von Bodelschwingh für eine Einstellung des Verfahrens, aber es erscheint dennoch so, dass er gegen ein modifiziertes Verfahren keinen

¹⁰ Die Darstellung der Abfolge der Projektschritte wurde mit einem von der Autorin selbst implementierten Python-Programm durchgeführt.

weiteren Widerstand geleistet hätte. Folglich ließe sich an dieser Stelle kritisch hinterfragen, ob es von Bodelschwingh tatsächlich vorrangig um die Rettung von Menschenleben ging oder um den Erhalt bzw. das Fortbestehen der eigenen, vom Vater gegründeten Anstalt ging. Anhand der Materialien im Archiv ließ sich hierauf keine Antwort finden.» (Journal LG 9, S. 4: 545).

Hierdurch zeigt sich, dass die Gruppe die Partialität historischer Quellen beziehungsweise der im digitalen Archiv vorhandenen Quellen erkennt. Nicht für alle ihre Fragen existieren geeignete Quellen, sodass sich auch nicht alle Fragen beantworten lassen.

4.7 Gruppe 28 – die «Min»-Gruppe

4.7.1 Recherche und Quellenauswahl

Die «Min»-Gruppe, die wesentlich weniger Seitenansichten im digitalen Archiv aufweist, beschränkt ihre Auswahl auf sieben Quellen, was nur einen kleinen Unterschied zur «Max»-Gruppe darstellt. Diese Auswahl begründet die Gruppe jedoch allein auf der inhaltlichen Ebene, nämlich mit dem möglichen Informationsgehalt der Quellen: «Wir haben im Archiv sieben Quellen gefunden, die die Absichten und Ziele v. Bodelschwinghs darstellen» (Journal LG 28, S. 1: 1587). Die Gruppe reflektiert daher zum Beispiel nicht, inwieweit von Bodelschwinghs Intentionen Einfluss auf den Aussagewert der Quellen haben und inwieweit daher ihre Quellenauswahl zu den Forschungsfragen passen.

Ebenso reflektiert die «Min»-Gruppe ihre Auswahl nicht im Hinblick auf die Quellengattungen und wählt sieben Quellen aus, die alle offizielle Briefe Bodelschwinghs an NS-Funktionäre sind, wie zum Beispiel an Reichsgesundheitsführer und Chef der Reichsärztekammer Leonardo Conti oder an Karl Brandt, Hitlers Begleitarzt und «Euthanasie»-Bevollmächtigten (siehe Abbildung 6).

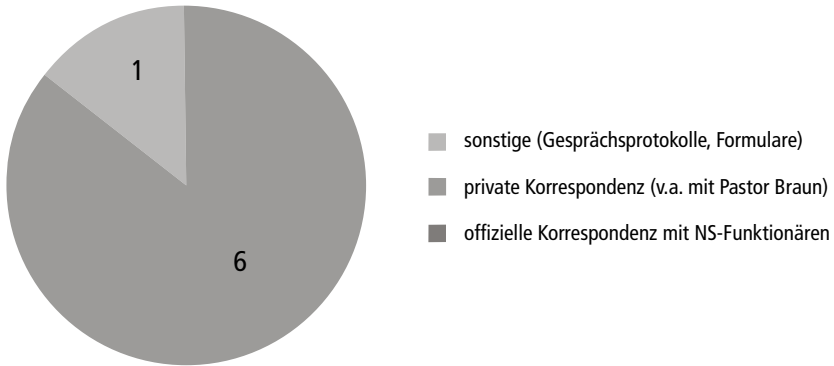


Abbildung 6: Verteilung der ausgewerteten Quellen der Gruppe 28 («Min») nach Quellengattungen

Diese Deutung lässt sich mit der geringeren Anzahl an Seitenansichten untermauern beziehungsweise erklärt die niedrige Zahl. Zudem zeigt die weitere Analyse der quantitativen Daten, dass diese Gruppe anders als die «Max»-Gruppe die Recherche der Quellen und deren Analyse getrennt durchgeführt und nicht wechselseitig aufeinander abgestimmt hat. Die Abfolge der Schritte «Recherche» und «Auswertung» (Phase 3 und 4) ist, wie in der Grafik dargestellt, linear (siehe Abbildung 7).

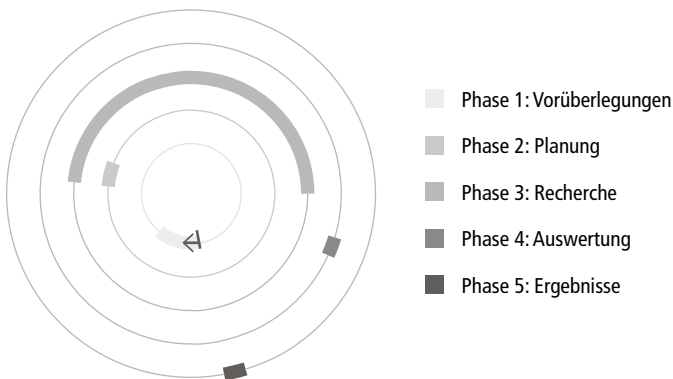


Abbildung 7: Abfolge der Projektschritte von Gruppe 28 («Min») in relativer Länge (absolute Gesamtzeit: 520 Minuten)

4.7.2 Der Umgang mit historischen Quellen in der historischen Narration

Die «Min»-Gruppe erkennt und reflektiert quellenkritisch in ihrer Narration im Journal das Problem, welches mit ihrer Quellenauswahl verbunden ist:

«Keine der ausgewählten und bearbeiteten Quellen gibt Aufschluss über die politische Positionierung von Friedrich von Bodelschwingh. [...] Genauso ist der Hitlergruß am Ende eines jeden Schriftstückes zu bewerten. Dies war zu jener Zeit Standard, wenn Briefe an Offizielle und Funktionäre geschrieben wurden. Es spielte die eigene politische Einstellung bei einem solchen Gruß keine Rolle. Die Einschätzung und Bewertung eines solchen teils schon standardisierten Schriftstückes ist zudem schwierig, da der Zweck hinter dem Inhalt nicht zweifelsfrei rekonstruiert werden kann.» (Journal LG 28, S. 3: 423)

Die Quellenauswahl beziehungsweise die ausgewählten Quellengattungen und die damit einhergehenden Einschränkungen führen in dieser Gruppe daher dazu, dass sie einen großen Teil der eigenen Fragen nicht oder nur teilweise beantworten kann. Beispielhaft wird dies in diesem Abschnitt deutlich:

«Die erste Leitfrage Gibt [sic] es Quellen, die eine rechtsextreme bzw. nationalsozialistische Positionierung Friedrichs unterstützen oder widerlegen? [sic] kann anhand der ausgewählten Quellen nicht abschließend geklärt werden.» (Journal LG 28, S. 3: 423)

Das Beispiel zeigt, neben mehreren weiteren, gut die zahlreichen quellenkritischen Reflexionen in der historischen Narration der «Min»-Gruppe, die bei der Quellenauswahl zuvor noch fehlten: So erkennt die Gruppe hier, dass sie ihre Fragestellungen anhand der Quellenauswahl nicht umfassend beantworten kann.

4.7.3 Vergleich der Fallbeispiele

Ausgehend von der großen Differenz der Seitenansichten im Archiv (Gruppe «Max»: 1634, Gruppe «Min»: 418) lassen sich Unterschiede der beiden Gruppen in Bezug auf Quellenrecherche, -auswertung und -auswahl feststellen (siehe Abbildung 8).

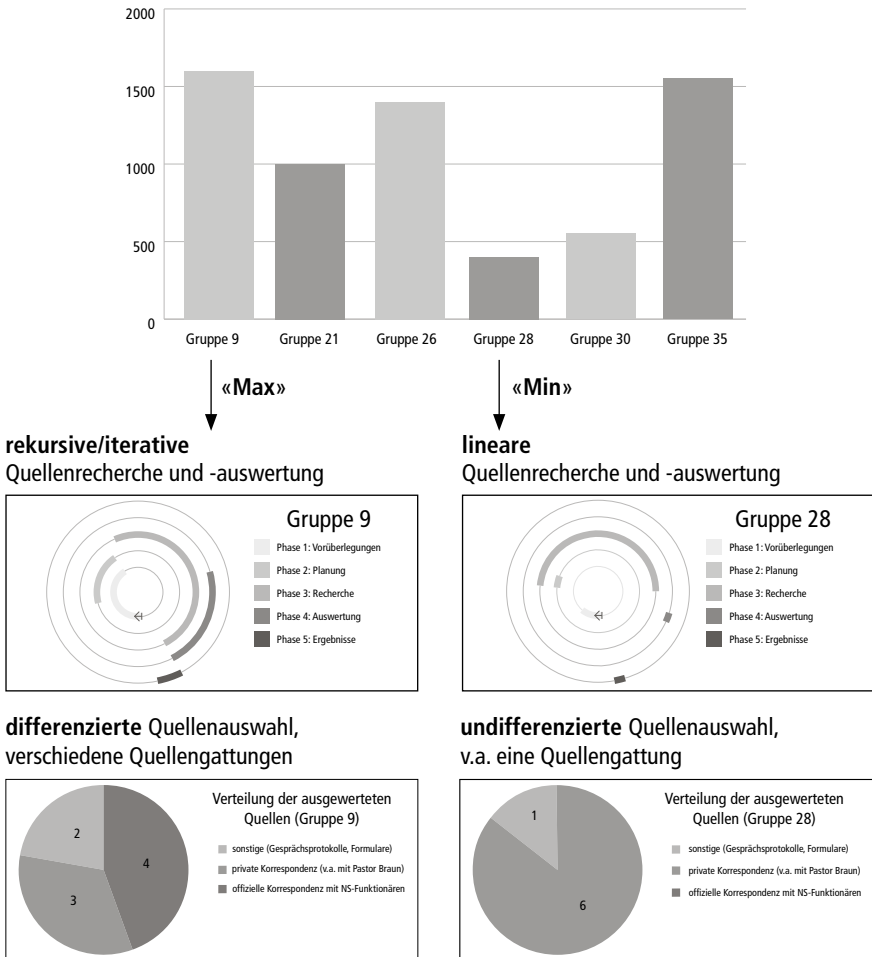


Abbildung 8: Zusammenstellung und Vergleich der Analyse- und Interpretationsergebnisse

Gruppe «Max» ist während der Recherche und Auswertung iterativ beziehungsweise rekursiv vorgegangen (Phase 3 und 4), hat diese Phasen also zu einem gewissen Anteil miteinander verknüpft beziehungsweise mehrmals durchgeführt.

Entsprechend liegen bei dieser Gruppe der Quellenauswahl bereits quellenkritische Überlegungen zugrunde. Gruppe «Min» hat dagegen linear zuerst die Quellenrecherche und danach die Auswertung vorgenommen. Der Quellenauswahl liegen keine quellenkritischen Überlegungen zugrunde. Demensprechend differiert ebenso die Auswahl der Quellen bezüglich der Quellengattung (siehe Abbildung 8). Gruppe «Max» hat bewusst verschiedene Quellengattungen (v. a. private und ebenso offizielle Briefe), Gruppe «Min» dagegen vor allem offizielle Briefe und keine private Korrespondenz ausgewählt und ausgewertet. Die Metrik der Seitenansichten kann daher in Kombination mit weiteren qualitativen und quantitativen Daten (siehe Abbildung 8), zumindest in diesen beiden Fällen, grundlegende Unterschiede der Quellenrecherche, -auswertung und -auswahl aufzeigen, die so gegebenenfalls auch für eine mögliche Typenbildung genutzt werden können. Inwieweit dies auch auf andere Gruppen zutrifft, gilt es allerdings noch zu untersuchen.

Keine beziehungsweise kaum Unterschiede zeigten sich hingegen beim Umgang mit den historischen Quellen in der abschließenden Narration, denn beide Gruppen gingen hier überwiegend quellenkritisch mit den ausgewählten Quellen um und reflektierten dementsprechend mehrmals zum Beispiel, inwieweit die jeweilige Quelle für die Beantwortung ihrer Frage(n), unter anderem im Hinblick auf die Quellengattung, genutzt werden darf.

4.8 Diskussion des Studiendesigns

An diesen Beispielen zeigt sich, dass vor allem die qualitative Interpretation der Daten mithilfe der quantitativen Analyse unterstützt werden kann, um so möglichst valide Ergebnisse zu erzielen. Kritisch bleibt jedoch zu hinterfragen, inwieweit die Strukturierung der Daten im ersten Schritt allein durch eine quantitative Analyse sinnvoll ist. Es konnten zwar Unterschiede zwischen den beiden Gruppen vor allem bezüglich der Quellenrecherche und -auswahl herausgestellt werden, inwieweit sich diese jedoch auch in den anderen Gruppen zeigen, muss noch untersucht werden. Problematisch könnte es nämlich sein, dass Gruppen, die laut quantitativer Metriken weit auseinander liegen, es bei der qualitativen Analyse nicht sind, sodass eher kleinere Differenzen ausgemacht werden und diesen gegebenenfalls ein zu großes Gewicht verliehen werden könnte. Viceversa könnten daher auch qualitative Unterschiede zwischen Gruppen, die laut quantitativer Metriken nah beieinander liegen, eher übersehen werden. Denkbar wäre daher stattdessen, in einem parallelen *Mixed-Methods-Design*, auch als *concur-*

rent triangulation bezeichnet (Köster & Thünemann, 2019, 26–27), wechselseitig die qualitativen und quantitativen Daten zu analysieren und für die Interpretation gegenseitig aufeinander zu beziehen, jedoch ohne bereits im Vorfeld eine Strukturierung der Daten vorzunehmen.

5 Fazit

Obwohl auf der Grundlage von lediglich zwei Fällen keinesfalls verallgemeinerbare Aussagen gemacht werden dürfen, lassen diese zumindest ein Potenzial der App erahnen. Bemerkenswert sind nämlich in beiden Gruppen besonders die quellenkritischen Reflexionen, sowohl bei der Quellenauswahl als auch in den historischen Narrationen. Bisherige Studien haben diesbezüglich eher gegenteilige Befunde herausgestellt. Von Borries etwa kritisiert in einem «Diskussionsbeitrag in Thesenform» (von Borries, 2007, 78) in Bezug auf die historischen Kompetenzen Lehramtsstudierender, dass «narrative Kompetenz» fehle und dies unter anderem auf eine «schematische Quellenkritik» zurückzuführen sei, bei der ««Informationsentnahme», aber nicht [...] sinnvolle «Einordnung», «Synthese» und «Präsentation» durchgeführt werden (ebd., 79). Ebenso konstatieren zum Beispiel Nitsche und Waldis, dass historisches Erzählen «auch für Geschichtslehrstudierende keine Tätigkeit [ist], die umstandslos vorausgesetzt werden kann» (Nitsche & Waldis, 2016, 30). Inwieweit zum Beispiel die eigenständige Quellenrecherche und -auswahl, wie sie die App in die Geschichte den Nutzer*innen ermöglicht, die narrativen Kompetenzen im Sinne der Re-Konstruktion fördern und dadurch die narrativen Performanzen beeinflussen, muss jedoch noch tiefergehend und umfangreicher untersucht werden. Bisher haben unter anderem Schreibinterventionsstudien gezeigt, dass «die narrative Kompetenz im Geschichtsunterricht von Gymnasiast*innen mittels gezielter materialbasierter Schreibintervention systematisch gefördert werden kann» (Waldis, Nitsche & Gollin, 2020, 108).

In der noch folgenden Hauptstudie gilt es vor allem die App-Nutzung von Schüler*innen näher zu untersuchen. Studien haben hierzu bisher nämlich überwiegend ernüchternde Ergebnisse gezeigt. So werden etwa in Abiturklausuren formale Quellenmerkmale zwar genannt, selten jedoch gewichtet und nie «mit der übrigen Interpretation systematisch verzahnt» (Schönemann, Thünemann & Zülsdorf-Kersting, 2011, 98). Und auch Unterrichtsbeobachtungen zeigten eher einen positivistischen Umgang mit historischen Quellen: «Dass sich Lehrkräfte

und Schülerinnen und Schüler mit Quellen ähnlich wie Historiker beschäftigt und dabei die erkenntnistheoretischen Prinzipien [...] den Umgang bestimmen, konnte im Rahmen der Studie nur in Ansätzen beobachtet werden.» (Spieß, 2014, 104). Inwieweit sich in der Studie zur App ebenso ein problematischer Umgang mit historischen Quellen bei Schüler*innen zeigt, wird daher ein wichtiger Untersuchungsaspekt sein.

Literatur

- Benad, Matthias (1997). Einleitung: Gesichtspunkte zur Orientierung. In Anneliese Hochmuth (Hrsg.), *Spurensuche: Eugenik, Sterilisation, Patientenmorde und die v. Bodelschwinghschen Anstalten Bethel 1929–1945* (S. XXI–XXXVI). Bielefeld: Bethel.
- Benad, Matthias/Schmuhl, Hans-Walter & Stockhecke, Kerstin (2016a). «... unter Einsatz aller unserer Kräfte Anwälte unserer Kranken sein»: Bethel und die nationalsozialistischen Krankenmorde – ein Überblick über den Stand der Forschung. In Matthias Benad/Hans-Walter Schmuhl & Kerstin Stockhecke (Hrsg.), *Bethels Mission (4): Beiträge von der Zeit des Nationalsozialismus bis zur Psychiatriereform* (S. 17–28). Bielefeld: Luther.
- Benad, Matthias/Schmuhl, Hans-Walter & Stockhecke, Kerstin (Hrsg.) (2016b). *Bethels Mission (4): Beiträge von der Zeit des Nationalsozialismus bis zur Psychiatriereform*. Bielefeld: Luther.
- Bernhard, Roland & Kühberger, Christoph (2018). «Digital history teaching»? Qualitativ empirische Ergebnisse aus 50 teilnehmenden Beobachtungen zur Verwendung von Medien im Geschichtsunterricht. In Thomas Sandkühler/Charlotte Bühl-Gramer/Anke John/Astrid Schwabe & Markus Bernhardt (Hrsg.), *Geschichtsunterricht im 21. Jahrhundert: Eine geschichtsdidaktische Standortbestimmung* (S. 425–440). Göttingen: V&R unipress.
- Bernhardt, Markus & Neeb, Sven (2020). Apps & Co – Grundlagen, Potenziale und Herausforderungen historischen Lernens in digitalen Lernumgebungen. *Zeitschrift für Didaktik der Gesellschaftswissenschaften*, 11(1), 65–82.
- Bernhardt, Markus & Neeb, Sven (2021). «Geschichte ganz smart – Geschichtslehrer*inne nausbildung mit digitalen Formaten innovativ gestalten». In Michael Beißwenger/Björn Bulizek/Inga Gryl & Florian Schacht (Hrsg.), *Digitale Innovationen und Kompetenzen in der Lehramtsausbildung* (S. 343–352). Duisburg: Universitätsverlag Rhein-Ruhr.
- Borries, Bodo von (1998). Geschichtslernen in offenen Lernformen und an außerschulischen Lernorten. In Günther Rohdenburg (Hrsg.), *Öffentlichkeit herstellen – Forschen erleichtern! Aufsätze und Literaturübersicht zur Archivpädagogik und historischen Bildungsarbeit* (S. 78–96). Hamburg: Körber-Stiftung.
- Borries, Bodo von (2007). «Geschichtsbewusstsein» und «Historische Kompetenz» von Studierenden der Lehramter Geschichte. *Zeitschrift für Geschichtsdidaktik*, 6, 60–83.

- Borries, Bodo von (2013). Zurück zu den Quellen? Plädoyer für die Narrationsprüfung. *Aus Politik und Zeitgeschichte*, 63(42/43). <http://www.bpb.de/apuz/170162/zurueck-zu-den-quellen-plaedoyer-fuer-die-narrationspruefung?p=all> (21.04.2022).
- Britt, Anne & Aglinskas, Cindy (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, 20(4), 485–522.
- Degen, Barbara (2014). *Bethel in der NS-Zeit: Die verschwiegene Geschichte*. Bad Homburg: VAS.
- Gautschi, Peter & Lücke, Martin (2018). Historisches Lernen im digitalen Klassenzimmer: Das Projekt «Shoa im schulischen Alltag». In Thomas Sandkühler/Charlotte Bühl-Gramer/Anke John/Astrid Schwabe & Markus Bernhardt (Hrsg.), *Geschichtsunterricht im 21. Jahrhundert: Eine geschichtsdidaktische Standortbestimmung* (S. 465–485). Göttingen: V&R unipress.
- Heuer, Christian (2011). «Gemeinsam erzählen»: Offener Unterricht, Aufgabenkultur und historisches Lernen. In Michele Barricelli/Axel Becker & Christian Heuer (Hrsg.), *Jede Gegenwart hat ihre Gründe: Geschichtsbewusstsein, historische Lebenswelten und Zukunftserwartung im frühen 21. Jahrhundert – Hans-Jürgen Pandel zum 70. Geburtstag* (S. 43–58). Schwabach/Ts.: Wochenschau.
- Hochmuth, Anneliese (Hrsg.) (1997). *Spurensuche: Eugenik, Sterilisation, Patientenmorde und die v. Bodelschwingschen Anstalten Bethel 1929–1945*. Bielefeld: Bethel.
- Hodel, Jan (2013). *Verkürzen und Verknüpfen: Geschichte als Netz narrativer Fragmente: Wie Jugendliche digitale Netzmedien für die Erstellung von Referaten im Geschichtsunterricht verwenden*. Bern: hep.
- Hodel, Jan & Waldis, Monika (2007). Sichtstrukturen im Geschichtsunterricht – die Ergebnisse der Videoanalyse. In Peter Gautschi (Hrsg.), *Geschichtsunterricht heute: Eine empirische Analyse ausgewählter Aspekte* (S. 91–142). Bern: hep.
- Hodel, Jan/Waldis, Monika/Zülsdorf-Kersting, Meik & Thünemann, Holger (2013). Schülernarrationen als Ausdruck narrativer Kompetenz. *Zeitschrift für Didaktik der Gesellschaftswissenschaften*, 4(2), 121–145.
- Kalischek, Ingo (2019, 5. Juni). Interview zu Euthanasie-Forschung: «Wenn es ernst wird, schaut Bethel weg». *Neue Westfälische*. https://www.nw.de/lokal/bielefeld/mitte/22473731_Interview-zu-Euthanasie-Forschung-Wenn-es-ernst-wird-schaut-Bethel-weg.html (27.02.2021).
- Körper, Andreas/Gärtner, Nico/Hartmann, Hanna & Stork, Annika (2021). Task-Based History Learning (TBHL) – ein Konzept für reflexive Lernaufgaben im Geschichtsunterricht? *Zeitschrift für Geschichtsdidaktik*, 20, 197–212.
- Körper, Andreas/Schreiber, Waltraud & Schöner, Alexander (Hrsg.) (2007). *Kompetenzen historischen Denkens: Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik. Kompetenzen: Grundlagen – Entwicklungen – Förderung* (Bd. 2). Neu-ried: ars una.
- Köster, Manuel & Thünemann, Holger (2019). The untapped potential of mixed-methods research approaches for German history education research. *History Education Research Journal*, 16(1), 25–34. <https://doi.org/10.18546/HERJ.16.1.03>.
- Krebs, Alexandra (2020). «App in die Geschichte». Schülerinnen und Schüler auf Spurensuche. *Geschichte lernen*, 33(194), 60–62.

- Krebs, Alexandra (2021). Die neue «App in die Geschichte». Ein digitaler Lernraum für Kooperationen zwischen Archiven und Lerngruppen. *Archivpflege in Westfalen-Lippe*, (93/94), 55–58.
- Krebs, Alexandra & Meyer-Hamme, Johannes (2021). Historisches Lernen digital. Die neue Version der App in die Geschichte. *Zeitschrift für Geschichtsdidaktik*, 20, 180–196.
- Kuckartz, Udo (2018). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (4. Aufl.). Weinheim: Beltz Juventa.
- Lange, Thomas & Lux, Thomas (2004). *Historisches Lernen im Archiv*. Schwalbach/Ts.: Wochenschau.
- Melter, Claus (Hrsg.) (2020). *Krankenmorde im Kinderkrankenhaus «Sonnenschein» in Bethel in der NS-Zeit? Forschungen zu Sozialer Arbeit, Medizin und «Euthanasie»*. Weinheim: Beltz Juventa.
- Monte-Sano, Chauncey & De La Paz, Susan (2012). Using writing tasks to elicit adolescents' historical reasoning. *Journal of Literacy Research*, 44(3), 273–299. <https://doi.org/10.1177/1086296X12450445>.
- Nitsche, Martin & Waldis, Monika (2016). Narrative Kompetenz von Studierenden erfassen – Zur Annäherung an formative und summative Vorgehensweisen im Fach Geschichte. *Zeitschrift für Didaktik der Gesellschaftswissenschaften*, 7(1), 17–35.
- Oswalt, Vadim/Aspelmeier, Jens & Boguth, Suzelle (2014). *Ich dachte, jetzt brennt gleich die Luft: Transnationale historische Projektarbeit zwischen interkultureller Begegnung und Web 2.0*. Schwalbach/Ts.: Wochenschau.
- Prinz, Doren & Thünemann, Holger (2016). Mixed-Methods-Ansätze in der empirischen Schul- und Unterrichtsforschung. Möglichkeiten und Grenzen für die Geschichtsdidaktik. In Holger Thünemann & Meik Zülsdorf-Kersting (Hrsg.), *Methoden geschichtsdidaktischer Unterrichtsforschung* (S. 229–253). Schwalbach/Ts.: Wochenschau.
- Rüsen, Jörn (1983). *Historische Vernunft. Grundzüge einer Historik I: Die Grundlage der Geschichtswissenschaft*. Göttingen: Vandenhoeck & Ruprecht.
- Schönemann, Bernd/Thünemann, Holger & Zülsdorf-Kersting, Meik (2011). *Was können Abiturienten? Zugleich ein Beitrag zur Debatte über Kompetenzen und Standards im Fach Geschichte* (2. Aufl.). Münster: Lit.
- Schreiber, Waltraud & Bertram, Christiane (2018). Ein multimediales Schulgeschichtsbuch in der Anwendung. Wie empirische Studien helfen können, Geschichtsunterricht besser zu verstehen. In Thomas Sandkühler/Charlotte Bühl-Gramer/Anke John/Astrid Schwabe & Markus Bernhardt (Hrsg.), *Geschichtsunterricht im 21. Jahrhundert: Eine geschichtsdidaktische Standortbestimmung* (S. 441–464). Göttingen: V&R unipress.
- Schwabe, Astrid (2012). *Historisches Lernen im World Wide Web: Suchen, flanieren oder forschen? Fachdidaktisch-mediale Konzeption, praktische Umsetzung und empirische Evaluation der regionalhistorischen Website Vimu.info*. Göttingen: V&R unipress.
- Schwabe, Astrid (2018). Womit? (Digitale) Medien des historischen Lernens. Einführung in die Sektion. In Thomas Sandkühler/Charlotte Bühl-Gramer/Anke John/Astrid Schwabe & Markus Bernhardt (Hrsg.), *Geschichtsunterricht im 21. Jahrhundert: Eine geschichtsdidaktische Standortbestimmung* (S. 413–424). Göttingen: V&R unipress.
- Schwabe, Astrid (2020). Historisches Lernen in Schulen der Vielfalt und Herausforderungen der Digitalisierung. In Sebastian Barsch/Bettina Degner/Christoph Kühberger &

- Martin Lücke (Hrsg.), *Handbuch Diversität im Geschichtsunterricht: Inklusive Geschichtsdidaktik* (S. 350–364).
- Spieß, Christian (2014). *Quellenarbeit im Geschichtsunterricht: Die empirische Rekonstruktion von Kompetenzerwerb im Umgang mit Quellen*. Göttingen: Vandenhoeck & Ruprecht.
- Waldis, Monika/Marti, Philipp & Nitsche, Martin (2015). Angehende Geschichtslehrpersonen schreiben Geschichte(n). *Zeitschrift für Geschichtsdidaktik*, 14, 63–86.
- Waldis, Monika/Nitsche, Martin & Gollin, Kristine (2020). «Schülerinnen und Schüler schreiben Geschichte» – Eine Interventionsstudie an Deutschschweizer Gymnasien. *Zeitschrift für Geschichtsdidaktik*, 19, 90–108.
- Wineburg, Sam (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83(1), 73–87. <https://doi.org/10.1037//0022-0663.83.1.73>.
- Wineburg, Sam (2018). *Why learn history (When it's already on your phone)*. Chicago: University of Chicago Press.

Kann man die Relevanz von Geschichte messen?

Erste Validierung der deutschsprachigen Version der Relevance of History Measurement Scale (RHMS-D)

Marcel Mierwald und Julia Waldeyer

1 Einleitung

Nicht selten wird aus geschichtstheoretischer Sicht (z. B. Rüsen, 2013; Stearns, 1998) und curricularer Perspektive (z. B. Bundesministerium für Bildung, Wissenschaft und Forschung, 2016; Deutschschweizer Erziehungsdirektorenkonferenz, 2018; Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2014; Verband der Geschichtslehrer Deutschlands, 2011) die lebenspraktische Relevanz von Geschichte darin gesehen, dass Menschen diese benötigen, um bspw. die gegenwärtige Welt zu verstehen, sich im zeitlichen Wandel zu orientieren, eine eigene Identität zu entwickeln, «Citizenship» auszubilden oder Handlungsfähigkeit zu erlangen, um etwa an geschichtskulturellen Diskursen teilzuhaben. Studien weisen jedoch darauf hin, dass gerade Mittelstufenschüler*innen Geschichte als eher wenig relevant ansehen (van Straaten, Wilschut & Oostdam, 2018) und häufig als etwas «Äußerliches» erleben, das sie selbst und ihr Leben wenig tangiert (Angvik & von Borries, 1997; Haeberli, 2005). Auch bei Lernenden am Ende der Mittelstufe und am Anfang der Oberstufe scheint das Relevanzempfinden bezüglich Geschichte eher mittelmäßig ausgeprägt zu sein (van Straaten et al., 2018; Waldis & Buff, 2007). Dabei könnte nicht zuletzt eine Rolle spielen, dass das Schulfach Geschichte durch Jugendliche im Vergleich zu anderen Fächern als weniger nützlich wahrgenommen wird (Wilschut, 2013) und die Bedeutsamkeit seiner Inhalte für Gegenwart und Zukunft selten erkannt werden (Haag & Götz, 2012). Neben diesen Befunden, die sich vor allem auf Deutschland, die Niederlande und die Schweiz beziehen, scheinen auch Schüler*innen aus England und Nordamerika unsicher darin zu sein, warum man sich mit Geschichte beschäftigen sollte (z. B. Foster, Ashby & Lee, 2008; Harris & Reynolds, 2014; Haydn & Harris, 2010; VanSledright, 1997). Darüber hinaus

deuten erste Befunde aus den Niederlanden und Spanien darauf hin, dass selbst Studienanfänger*innen des Lehramtes Geschichte und der Geschichtswissenschaft nur mittelmäßig bis moderat stark zustimmen, dass Geschichte für sie relevant ist (Miguel-Revilla, 2021; van Straaten et al., 2018).

Diese Befundlage ist zum einem insofern problematisch, als sich geschichts-didaktisch fragen ließe, inwiefern Kinder, Jugendliche und (junge) Erwachsene die eingangs beschriebenen Ziele historischer Bildung überhaupt erreichen und so etwa Zusammenhänge zwischen Vergangenheit, Gegenwart und Zukunft ziehen (Jeismann, 1980; Rösen, 2020). Pädagogisch-psychologisch sind diese Befunde zumindest hinderlich, da Wertüberzeugungen mit der Motivation und Leistung in einem Schulfach zusammenhängen und als Prädiktor für Entscheidungen bei der Kurs- und Studienwahl gelten (z. B. Gaspard, Nagengast & Trautwein, 2019). Zum anderen muss die Befundlage selbst problematisiert werden. So wurden die Daten auf der Grundlage unterschiedlicher theoretischer Konzeptionen und methodischer Herangehensweisen generiert und sind somit nur bedingt vergleichbar und aussagekräftig. Gerade für den deutschsprachigen Raum fehlt es an einem validen und reliablen Instrument, mit dem solche Relevanzüberzeugungen auch bei größeren Stichproben speziell für die Domäne Geschichte diagnostiziert und die Wirkung von Interventionen zur Relevanzförderung zukünftig untersucht werden können. Solch ein Fragebogen liegt mit der Entwicklung der *Relevance of History Measurement Scale* (RHMS) durch die Geschichtsdidaktiker van Straaten, Wilschut und Oostdam (2018) bisher nur für niederländische Schüler*innen und Studierende vor.

Vor diesem Hintergrund fragen wir in diesem Beitrag, inwiefern die deutschsprachige Variante der RHMS ein valides und reliables Instrument zur Erfassung von Relevanzüberzeugungen in der Domäne Geschichte darstellt. Hierfür prüften wir das übersetzte Instrument in einem ersten Versuch an einer Stichprobe von Studierenden der Geschichtswissenschaft, denen entsprechend vorliegenden Befunde mit niederländischen und spanischen Studierenden tendenziell hohe Zustimmungswerte attestiert werden sollten (Miguel-Revilla, 2021; van Straaten et al., 2018). Im Folgenden gehen wir auf den theoretischen Hintergrund sowie die Methode, Forschungsfrage und Ergebnisse dieser ersten Validierungsstudie ein und diskutieren abschließend deren Befunde.

2 Theoretischer Hintergrund

2.1 Relevanz im Fokus bisheriger geschichtsdidaktischer Theorie und Empirie

Die «Relevanz» von Geschichte wurde bisher vor allem zu einem Gegenstand der theoretischen Reflexion. Schönemann (2014, 166) geht unter dem gleichnamigen Lemma im Wörterbuch Geschichtsdidaktik darauf ein, dass die «Wichtigkeit, Erheblichkeit» von Geschichte und Geschichtswissenschaft Bestandteil der geschichtstheoretischen Debatte in den frühen 1970er-Jahren war. Dabei stritten Historiker (z. B. Kocka, 1974; Nipperdey, 1972) um die Bedeutsamkeit der Geschichte für die Gegenwart und Gesellschaft angesichts der aufstrebenden Sozialwissenschaften und ihrer schulischen Pendanten mit den Fächern Politik und Sozialkunde (Schönemann, 2014). Nach dem «Wert der Geschichte», so sei an dieser Stelle angemerkt, wird weiterhin in aktuelleren Büchern und Essays gefragt (z. B. Brechtken, 2020; Tanner, 2011). Auch die Geschichtsdidaktik der 1970er-Jahre, so Schönemann (2014, 168–169) weiter, befasste sich mit der Relevanz von Geschichte unter dem Blickwinkel der «Auswahl der Inhalte» und «Reform der Lehrpläne», der Einführung des Prinzips des Gegenwarts- und Zukunftsbezugs sowie der Abstimmung von Sach- und Lerner*innen-Ebene bei der Unterrichtsplanung. Viele dieser Punkte prägen die geschichtsdidaktische Diskussion und schulische Praxis bis heute (z. B. Mayer, Gautschi & Bernhardt, 2012).

Auch auf der Ebene der geschichtsdidaktischen Empirie fand die Relevanz von Geschichte Beachtung. So liegen vereinzelt Fragebogenskalen vor, die Aspekte der Relevanz von Geschichte und des Geschichtsunterrichts erfassen (z. B. Kindervater & von Borries, 1997; Waldis & Buff, 2007). Weiterhin existieren hierfür in der internationalen geschichtsdidaktischen Forschung verschiedene quantitative Umfragen mit geschlossenen und offenen Fragen sowie qualitative Interviews oder Studien, die qualitative und quantitative Verfahren kombinieren (z. B. Foster et al., 2008; Haerberli, 2005; Harris & Reynolds, 2014; Haydn & Harris, 2010; Van Sledright, 1997). Daneben ist für die deutschsprachige Geschichtsdidaktik die Studie von Daumüller und Seidenfuß (2017) hervorzuheben, denen es in narrativen Interviews um die Frage ging, welchen Nutzen Schulabgänger*innen aus ihrem Geschichtsunterricht für sich ziehen. Bei den hier grob skizzierten Ansätzen und der von Daumüller und Seidenfuß (2017, 39–51) unter dem «Relevanz-Label» synthetisierten Forschung bleibt jedoch vage, inwiefern es um die Relevanz von Geschichte allgemein oder die des Geschichts-

unterrichts speziell ging oder ob Interesse, Wissen und Konzepte bezüglich Geschichte untersucht wurden.

2.2 Eine neue geschichtsdidaktische Konzeptualisierung von Relevanz in Geschichte

Verglichen mit den bisherigen Ansätzen definieren van Straaten, Wilschut und Oostdam (2016, 4) Relevanz in Geschichte konzeptionell klarer als «allowing students to recognize and experience what history has to do with themselves, with today's society and their general understanding of human existence». Sie beziehen sich damit auf die generelle Bedeutsamkeit von Geschichte für das gegenwärtige und zukünftige Leben von Menschen und die Ansichten, die diese dazu vertreten können. Ihre Konzeptualisierung von Relevanz in der Domäne Geschichte basiert auf drei theoretischen Quellen (vgl. im Folgenden: van Straaten et al., 2016, 6–16; 2018, 103–104; Wilschut, 2019, 133–138). Aus bildungstheoretischen Überlegungen übernehmen sie Qualifikation, Sozialisation und Subjektwerdung als übergeordnete Bildungsziele. Anhand geschichtsdidaktischer Literatur legen die Forscher dar, dass historische Bildung einen Beitrag dazu leisten sollte, junge Menschen auf das zukünftige Leben vorzubereiten (z. B. politische Ideen und Prozesse verstehen oder Argumentationsfähigkeit erlangen). Weiterhin heben sie die Bedeutung von Geschichte für die Sozialisierung von Menschen in historisch gewachsenen Gesellschaften (z. B. Wissen um Institutionen, Traditionen und Denkweisen erwerben) und die Entwicklung einer eigenen Identität (z. B. tradierte Werten, Ideale und Überzeugungen reflektieren) hervor. Der konstruktivistischen Lerntheorie entnehmen sie als zweiter Quelle, dass Wissen aktiv konstruiert, bedeutungsvoll und transferfähig auf andere Kontexte sein sollte. Dies sei im Einklang mit Lernansätzen im Geschichtsunterricht, die bspw. das eigene historische Forschen der Schüler*innen anregen und ihnen ermöglichen, Verbindungen zwischen Vergangenheit, Gegenwart und Zukunft zu ziehen. Schließlich dienen van Straaten und Kollegen (2016) geschichtstheoretische Überlegungen zum Geschichtsbewusstsein und historischen Denken als dritte Quelle für ihre Konzeptualisierung von Relevanz in Geschichte. Dabei arbeiten sie heraus, dass Geschichte hilfreich und nötig ist, um mehr über das Menschsein in seiner zeitlichen Dimension zu erfahren und sich historisch orientieren zu können (z. B. Einsicht in menschliche Grunderfahrungen gewinnen, sich durch das Fremde mit dem Eigenen beschäftigen, Beeinflussung der Gegenwart durch die Vergangenheit erkennen).

Auf der Grundlage der skizzierten Quellen leiten sie drei Ziele historischen Lernens ab und operationalisieren das Relevanzkonstrukt mit den Teilaspekten «building a personal identity», «becoming a citizen» und «understanding the human condition» (van Straaten et al., 2016, 6–12; Wilschut, 2019, 136–137). «Building a personal identity» (IDE) ist näher dadurch beschrieben, inwiefern sich Menschen als Individuen mit einer eigenen Vergangenheit sehen und Werte, Einstellungen und Ideale durch Geschichte entwickeln. Im Teilaspekt «becoming a citizen» (CIT) geht es darum, inwiefern Menschen Geschichte nützlich erscheint, um gegenwärtige Ereignisse und Phänomene zu erklären und sich über aktuelle Verhältnisse ein begründetes Urteil zu bilden. Schließlich meint «understanding the human condition» (HUM), inwiefern sich Menschen über ihre eigene Historizität bewusst werden und ihre Erfahrungen mit der Vergangenheit auf gegenwärtige und zukünftige menschliche Probleme beziehen.

Die niederländischen Geschichtsdidaktiker entwerfen eine fachspezifische Theorie, die das Relevanzkonstrukt in Geschichte gegenüber domänenübergreifenden Ansätzen zur Definition von Wertüberzeugungen wie der Erwartungswert-Theorie (Wigfield & Eccles, 2000) aus der Pädagogischen Psychologie konturiert. Dabei zeigen sich einige Schnittpunkte mit der deutschsprachigen Geschichtsdidaktik, in der intensiv die Funktionen der Identitätsbildung und Orientierung (\approx IDE), die Befähigung zur geschichtskulturellen Teilhabe (\approx CIT) und die Anbahnung eines reflektierten Geschichtsbewusstseins (\approx HUM) als Ziele des historischen Lernens diskutiert werden (z. B. Meyer-Hamme, 2018; Rösen, 2013; Schreiber, 2002). Interessant ist, dass die theoriebasierte Operationalisierung von van Straaten und Kollegen (2016) eine gewisse Ähnlichkeit mit den «Relevanz-Konstruktionen» aufweist, die Daumüller und Seidenfuß (2017, 136) bei deutschsprachigen Schulabgänger*innen empirisch identifizierten. Zudem erscheint es zukünftig kompetenzdidaktisch gedacht lohnend, tiefergehend danach zu fragen, inwiefern Relevanzüberzeugungen im Fach Geschichte Bedingung oder Ausdruck historischer Orientierungskompetenz sind (z. B. Trautwein et al., 2017).

Neben der Theoriebildung gelang es van Straaten et al. (2018) mit der RHMS, Relevanzüberzeugungen von niederländischen Schüler*innen und Studierenden empirisch valide und reliabel zu erfassen. Ihre Ergebnisse deuten darauf hin, dass Geschichte mit zunehmendem Alter relevanter für Menschen wird und dieses Relevanzempfinden vor allem ab dem 14. bis hin zum 18. Lebensjahr stark ansteigt. Im Alter von 12 bis 13 Jahren gibt es vermutlich pubertär bedingt einen Abfall in der Beurteilung der Relevanz von Geschichte (siehe auch Angvik & von

Borries, 1997, B24–27; von Borries, 2001, 8). Weiterhin zeigt sich, dass Studienanfänger*innen (zumindest für das Lehramt Geschichte) eine stärkere Relevanzzuschreibung aufweisen als Oberstufenschüler*innen (van Straaten et al., 2018). Eine neuere Befragung mit einer stark gekürzten spanischen Fragebogenadaptation der RHMS bestätigt diesen Befund (Miguel-Revilla, 2021). Tendenziell weisen auch qualitative Studien auf höhere Relevanzzuschreibungen von älteren und in der Ausbildung fortgeschrittenen Personengruppen wie Schulabgänger*innen und Studierenden hin (Berg, 2019; Daumüller & Seidenfuß, 2014). Offen bleibt jedoch die Frage, inwiefern die Ergebnisse von van Straaten et al. (2018) sowohl messtheoretisch als auch inhaltlich auf deutsche Schüler*innen und Student*innen übertragbar sind. Die Grundlage zur Beantwortung dieser Frage bildet eine vorherige Validierung der Skala an deutschsprachigen Personen.

3 Forschungsfrage und Hypothesen

In diesem Beitrag gehen wir der übergeordneten Forschungsfrage nach, inwiefern es sich bei der deutschsprachigen Version der RHMS (im Folgenden RHMS-D) um ein valides Instrument zur Erfassung von Relevanzüberzeugungen in der Domäne Geschichte handelt.

Im Fokus der Konstruktvalidierung steht die Überprüfung der faktoriellen Validität der RHMS-D (Hartig, Frey & Jude, 2012). Diesbezüglich nehmen wir an, dass sich die faktorielle Struktur des niederländischen Originals der RHMS für seine deutschsprachige Übersetzung im Wesentlichen replizieren lässt. Genauer gehen wir davon aus, dass ein Drei-Faktoren-Modell die Daten angemessen beschreibt (Hypothese 1). Zur Prüfung der kriterialen Validität untersuchen wir Zusammenhänge der drei RHMS-Teilkonstrukte mit ausgewählten zeitgleich existierenden Außenkriterien. Wir wählten die Häufigkeit der Beschäftigung mit Geschichte und das individuelle Interesse an Geschichte als Außenkriterien, da frühere Studien (van Straaten et al., 2018; 2019) mittlere bis starke korrelative Zusammenhänge zu den Teilkonstrukten des Fragebogens ergaben. Darauf basierend vermuten wir, dass auch für den deutschsprachigen Fragebogen positive Korrelationen zwischen den Außenkriterien und den drei Teilkonstrukten existieren (Hypothese 2). Zuletzt untersuchen wir die Ausprägung der drei Relevanzüberzeugungen genauer. Hier gehen wir der Annahme nach, dass mit aufsteigendem Alter und längerer zeitlicher Dauer der Beschäftigung mit Geschichte auch die Zustimmung dazu, dass Geschichte relevant ist, zunimmt (Miguel-Revilla,

2021; van Straaten et al., 2018). Um diesen Aspekt der Validität, die sogenannte «known-group validity» (z. B. Hattie & Cooksey, 1984), zu untersuchen, teilen wir unsere Stichprobe in die Gruppen der B.A.- und der M.Ed.-Studierenden ein. Wir gehen generell von hohen Relevanzüberzeugungen bei beiden Gruppen aus und nehmen an, dass die M.Ed.-Studierenden höhere Zustimmungswerte zu den Teilkonstrukten der RHMS-D aufweisen als die B.A.-Studierenden (Hypothese 3).

4 Methode

4.1 Stichprobe und Vorgehen

In dieser ersten Pilotierung der RHMS-D testeten wir das Instrument an einer Stichprobe von 257 Studierenden (Alter: $M = 22.04$, $SD = 4.73$; $m = 135$, $w = 122$) der Geschichtswissenschaft. Die Studierenden wurden im Wintersemester 2019/2020 innerhalb von elf Lehrveranstaltungen an der Ruhr-Universität Bochum befragt. 213 Studierende (Alter: $M = 20.97$, $SD = 2.77$; $m = 113$, $w = 100$) befanden sich im Bachelor of Arts (Hochschulsemester: $M = 2.75$, $SD = 2.18$) und 44 Studierende (Alter: $M = 27.45$, $SD = 6.52$; $m = 22$, $w = 22$) im Master of Education (Hochschulsemester: $M = 10.86$, $SD = 3.45$).

Die Befragung fand in den ersten 20 Minuten der jeweiligen Lehrveranstaltungen statt, deren Dozent*innen Zeit für die Bearbeitung des Fragebogens einräumten. Der Erstautor dieses Beitrages stellte sich in den ersten 5 Minuten vor und erläuterte kurz das Anliegen der Befragung unter der Nutzung der immer gleichen Instruktion. In den anschließenden 15 Minuten erfolgte die freiwillige und anonyme Bearbeitung der ausgeteilten Kopien des Fragebogens durch die Studierenden. Der ausgeteilte Fragebogen umfasste neben der RHMS-D einige Fragen zur Erfassung soziodemografischer und studienbezogener Daten sowie zum Zwecke der Validierung weitere Skalen zur Erfassung der Außenkriterien.

4.2 Instrumente

4.2.1 Relevance of History Measurement Scale (RHMS-D)

Auf der Basis des niederländischen Fragebogens und dessen publizierter englischsprachiger Version (van Straaten et al., 2018) wurde, orientiert an Empfehlungen für Fragebogenübersetzungen im interkulturellen Rahmen (Harkness, 2003), eine deutschsprachige Version erstellt. Hierfür wurden gemeinsam mit

den Urhebern der RHMS, Arie Wilschut und Dick van Straaten, unterschiedliche Übersetzungen der Items aus dem Englischen und Niederländischen ins Deutsche vorgenommen sowie tabellarisch die Sprachvarianten gegenübergestellt, abgeglichen und diskutiert. Dabei erwies es sich als hilfreich, dass die niederländischen Forscher ihre Vorstellungen zu den ursprünglichen Items erläuterten und über gute Deutschkenntnisse verfügen.

Die deutschsprachige Version der RHMS besteht, wie das niederländische Original, aus 24 Items, die sich auf die drei Teilkonstrukte «building a personal identity» (IDE = 7 Items, z. B. «Geschichte beeinflusst, wie ich mich verhalte.»), «becoming a citizen» (CIT = 12 Items, z. B. «Durch Geschichte kann ich mir eine eigene Meinung über die Dinge bilden.») und «understand the human condition» (HUM = 5 Items; z. B. «Geschichte hilft uns nicht, die heutigen Probleme zu lösen.»; invers formuliert) beziehen. Die Proband*innen sind angehalten, auf einer sechsstufigen Likert-Skala ihre Übereinstimmung mit den positiv oder negativ vorformulierten Statements zu markieren (d. h. von 1= «stimme ich überhaupt nicht zu» bis 6 = «stimme ich voll und ganz zu»).

4.2.2 Außenkriterien zur Validierung

Bei der Auswahl der Außenkriterien orientierten wir uns an van Straaten et al. (2018), die die Häufigkeit der Beschäftigung mit Geschichte außerhalb der Schule beziehungsweise des Studiums mit einem Item («Wie oft sprechen Sie über Geschichte zu Hause oder mit anderen Leuten?») auf einer vierstufigen Likert-Skala von 1 = «niemals» bis 4 = «häufig» erfassten. Weiterhin erfragten wir das individuelle Interesse an Geschichte durch die Adaption einer Skala von Köller, Baumert und Schnabel (2000) für Mathematik. Diese fragt nach dem Interesse am «Gegenstand» Geschichte und beinhaltet fünf Items, die sich auf den emotionalen Aspekt, die Wichtigkeit und das Flow-Erleben beziehen (z. B. «Mich mit Geschichte zu beschäftigen, macht mir Spaß.»). Die Studierenden gaben hierzu ihre Einschätzungen auf einer vierstufigen Likert-Skala von 1 = «trifft überhaupt nicht zu» bis 4 = «trifft von und ganz zu» ab. Die Reliabilität der Skala liegt in dieser Studie bei einem Cronbachs Alpha von .70.

4.3 Statistische Analysen

Wie bei van Straaten et al. (2018) wurde eine Konfirmatorische Faktorenanalyse (CFA) durchgeführt, um zu prüfen, ob die ursprüngliche Faktorenstruktur in unseren Daten bestätigt werden kann. Die CFA ist ein hypothesenprüfendes

Verfahren zur Untersuchung der Frage, ob unsere Daten aus der Befragung der Studierenden mit dem theoretisch angenommenen Modell von drei unterschiedlichen Teilkonstrukten von Relevanz in Geschichte (d. h. Faktoren) und den dazugehörigen Items beziehungsweise Indikatoren übereinstimmt oder nicht. Dabei können auch verschiedene konkurrierende Modelle mit demselben Datensatz getestet werden, falls bspw. Indikatoren aufgrund niedriger Ladungen entfernt werden. Zur Überprüfung der Übereinstimmung, dem sogenannten Modellfit, werden verschiedene Gütekriterien beziehungsweise Fit-Maße (siehe weiter unten) verwendet (Moosbrugger & Schermelleh-Engel, 2012, 334–341). Alle Analysen zur Überprüfung der faktoriellen Struktur wurden mit R 4.0.2 (package lavaan; Rosseel, 2012) unter der Verwendung des Robust Maximum Likelihood Estimators (MLR) und des Full-Information Maximum Likelihood Algorithm (FIML; Enders, 2010) zur Berücksichtigung fehlender Werte durchgeführt. Aufgrund der Empfindlichkeit des Chi-Quadrat-Tests gegenüber der Stichprobengröße konzentrierten wir uns bezüglich der oben erwähnten Fit-Maße auf den Comparative Fit Index (CFI), den Tucker-Lewis Index (TLI) und den Root-Mean-Square Error of Approximation (RMSEA). Für TLI und CFI zeigen Werte von .90 oder höher einen akzeptablen Fit an, während Werte über .95 einen guten Fit anzeigen (Hu & Bentler, 1999). Für den RMSEA werden Werte von $< .05$ als guter Fit, Werte zwischen $.05$ und $< .08$ als angemessener Fit und Werte zwischen $.08$ und $.10$ als mittelmäßige Passung angesehen (Browne & Cudeck, 1993).

Die kriteriale Validität überprüften wir über Pearson-Korrelationen der drei Teilkonstrukte beziehungsweise Relevanzüberzeugungen mit den gewählten Außenkriterien. Gemäß Cohen (1988) sprechen signifikante Korrelationen von $r > .10$ für einen kleinen, $r > .30$ für einen mittleren und $r > .50$ für einen großen Zusammenhang. Für die Untersuchung der Ausprägung der Relevanzüberzeugungen vergleichen wir über t -Tests für abhängige Stichproben den Grad der Zustimmung zu den drei Skalen des Fragebogens. Über t -Tests für unabhängige Stichproben gingen wir der Frage nach, inwiefern statistisch signifikante Unterschiede zwischen B.A.- und M.Ed.-Studierenden hinsichtlich der drei Teilkonstrukte «building a personal identity», «becoming a citizen» und «understand the human condition» bestehen. Mehrfachvergleiche fanden zur Vermeidung einer α -Fehler-Kumulierung unter der Berücksichtigung der sogenannten Bonferroni-Korrektur statt. Die genannten Berechnungen wurden alle mit IBM SPSS Statistics 27 durchgeführt.

5 Ergebnisse

5.1 Überprüfung der faktoriellen Struktur der RHMS-D

Zunächst ergab die etablierte dreifaktorielle Struktur des RHMS-D-Fragebogens eine mittelmäßige Datenanpassung für die Gesamtstichprobe ($\chi^2 = 528.775$, $p < .001$, CFI = 0.78, TLI = 0.76, RMSEA = 0.07). Eine genauere Betrachtung der Ergebnisse zeigte, dass drei Indikatoren niedrige Faktorladungen ($< .40$) aufwiesen, sodass wir diese entfernten und ein alternatives Modell mit 21 Indikatoren (d. h. gleiche dreifaktorielle Struktur) testeten.

Danach zeigte das Drei-Faktoren-Modell eine akzeptable Passung auf die Daten ($\chi^2 = 272.692$, $p < 0.001$; CFI = 0.92; TLI = 0.90; RMSEA = 0.05; SRMR = 0.06). Die standardisierten Faktorladungen waren alle signifikant und reichten von .42 bis .76 (siehe Tabelle 1). Die Kovarianzen zeigten, dass die drei Faktoren bedeutungsvoll korreliert waren (siehe Tabelle 2). Insgesamt stimmten die Ergebnisse der CFA weitgehend mit denen der ursprünglichen Studie von van Straaten und Kollegen (2018) überein. Die drei Faktoren wiesen eine akzeptable bis gute interne Konsistenz auf: $\omega_{IDE} = .78$, $\omega_{CIT} = .80$, $\omega_{HUM} = .68$.

Tabelle 1: Ladungen des Drei-Faktoren-Modells für die Gesamtstichprobe

Faktor [Indikator]	Building a personal identity (IDE)	Becoming a citizen (CIT)	Understanding the human condition (HUM)
IDE			
[IDE1]	.51		
[IDE3] Bsp.: «Geschichte beeinflusst, wie ich mich verhalte.»	.69		
[IDE4]	.58		
[IDE5]	.66		
[IDE6]	.64		
[IDE7]	.57		
CIT			
[CIT2]		.57	
[CIT3] Bsp.: «Durch Geschichte kann ich mir eine eigene Meinung über die Dinge bilden.»		.53	
[CIT4]		.46	
[CIT5]		.42	
[CIT6]		.43	
[CIT7]		.63	
[CIT8]		.53	
[CIT9]		.60	
[CIT10]		.40	
[CIT11]		.56	
[CIT12]		.59	
HUM			
[HUM2] Bsp.: «Geschichte hilft uns nicht, die heutigen Probleme zu lösen.» (invers codiert)			.50
[HUM3]			.59
[HUM4]			.51
[HUM5]			.76

Tabelle 2: Faktorkorrelationen (Standardfehler in Klammern) für die Gesamtstichprobe

	Anzahl der Indikatoren	1	2
1 Building a personal identity (IDE)	6		
2 Becoming a citizen (CIT)	11	.66 (.07)	
3 Understanding the human condition (HUM)	4	.57 (.05)	.51 (.05)

5.2 Zusammenhänge zwischen Relevanzüberzeugungen und Außenkriterien

Es konnten erwartungskonforme korrelative Zusammenhänge zwischen den Relevanzüberzeugungen der Studierenden und der Häufigkeit ihrer Beschäftigung mit Geschichte sowie ihrem individuellen Interesse an Geschichte festgestellt werden (siehe Tabelle 3).

Tabelle 3: Korrelationen zwischen den Relevanzüberzeugungen und Außenkriterien für die Gesamtstichprobe ($N = 257$)

	Häufigkeit der Beschäftigung mit Geschichte	Individuelles Interesse an Geschichte
Building a personal identity (IDE)	.37**	.43**
Becoming a citizen (CIT)	.39**	.48**
Understanding the human condition (HUM)	.21**	.27**

Anmerkungen: Die Korrelationen basieren auf Pearsons r , $p < .01^{**}$, $p < .05^{*}$ (zweiseitig).

Spezifisch zeigten sich signifikante Korrelationen mit einem mittleren Effekt sowohl zwischen der Häufigkeit der Beschäftigung mit Geschichte und den Relevanzüberzeugungen der Studierenden bezüglich «building a personal identity» ($r = .37$, $p < .01$) und «becoming a citizen» ($r = .39$, $p < .01$) als auch ihrem individuellen Interesse an Geschichte und «building a personal identity» ($r = .43$, $p < .01$) und «becoming a citizen» ($r = .48$, $p < .01$). Die Relevanzüberzeugung der Studierenden zu «understand the human condition» korrelierten signifikant, aber mit einem niedrigen Effekt mit der Häufigkeit der Beschäftigung mit Ge-

schichte ($r = .21, p < .01$) und dem individuellen Interesse an Geschichte ($r = .27, p < .01$). Die Befunde sprechen dafür, dass, je häufiger die Studierenden sich mit Geschichte beschäftigten und je höher ihr individuelles Interesse an Geschichte ausgeprägt war, desto relevanter empfanden sie Geschichte auch für die Entwicklung einer eigenen Identität, das Werden eines Staatsbürgers und dem Verständnis für die menschliche Existenz (und vice versa).

5.3 Ausprägung der Relevanzüberzeugungen

Erwartungskonform wurden in beiden Gruppen relativ hohe Zustimmungswerte zu den drei Teilkonstrukten erzielt (siehe Tabelle 4). Auf die gesamte Stichprobe bezogen, vertraten die Studierenden stärker die Ansicht, dass Geschichte für «becoming a citizen» ($M = 4.69, SD = 0.60$) wichtiger sei als für «understanding the human condition» ($M = 4.05, SD = 0.79$), $t(256) = 12.74, p < .001, d = 0.90$, und «building a personal identity» ($M = 3.74, SD = 0.82$), $t(256) = -20.92, p < .001, d = 1.29$. In beiden Fällen handelt es sich um große Effekte. Zudem zeigte sich eine höhere Zustimmung zu «understanding the human condition» als zu «building a personal identity», $t(256) = 5.45, p < .001, d = 0.39$. Dabei handelt es sich um einen kleinen Effekt.

Tabelle 4: Mittelwerte und Standardabweichungen für die drei Teilkonstrukte (IDE, CIT und HUM) der RHMS-D

	Building a personal identity (IDE)		Becoming a citizen (CIT)		Understanding the human condition (HUM)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
B.A.-Studierende ($n = 213$)	3.68	0.85	4.66	0.63	4.05	0.80
M.Ed.-Studierende ($n = 44$)	4.03	0.60	4.86	0.40	4.07	0.73
Gesamte Stichprobe ($N = 257$)	3.74	0.82	4.69	0.60	4.05	0.79

Anmerkung: Die angegebenen Mittelwerte bewegen sich auf einer Skala von 1 («stimme ich überhaupt nicht zu») bis 6 («stimme ich voll und ganz zu»).

Mit Blick auf die Gruppenvergleiche zeigten die Studierenden im M.Ed. ($M = 4.03, SD = 0.60$) eine höhere Zustimmung bezüglich «building a personal identity» als

diejenigen im B.A. ($M = 3.68$, $SD = 0.85$), $t(83.10) = -3.23$, $p = .006$, $d = 0.43$. Dabei ist ein kleiner Effekt sichtbar. Der Levene-Test indizierte hier eine Ungleichheit der Varianzen ($F = 6.41$, $p = .012$), sodass ein t -Test mit «Welch-Korrektur» und adjustierten Freiheitsgraden genutzt wurde. Auch mit Blick auf «becoming a citizen» zeigten die Studierenden im M.Ed. ($M = 4.86$, $SD = 0.40$) eine höhere Zustimmung als die B.A.-Studierenden ($M = 4.66$, $SD = 0.63$), $t(93.80) = -2.83$, $p = .018$, $d = 0.34$. Es handelt sich abermals um einen kleinen Effekt. Hier wurden die Freiheitsgrade ebenfalls wegen fehlender Varianzhomogenität adjustiert ($F = 10.40$, $p = .001$). Mit Blick auf «understand the human condition» unterschieden sich M.Ed.- ($M = 4.07$, $SD = 0.73$) und B.A.-Studierende ($M = 4.05$, $SD = 0.80$) nicht statistisch signifikant, $t(255) = -0.17$, $p = 1.000$, $d = 0.03$. Wie angenommen zeigten die M.Ed.-Studierenden höhere Relevanzüberzeugungen als die B.A.-Studierenden, wobei die Unterschiede nur bei zwei von drei Teilkonstrukten statistisch signifikant waren.

6 Diskussion

Das Ziel dieses Beitrages war es, eine erste Validierung der deutschsprachigen Version der Relevance of History Measurement Scale (RHMS-D) vorzunehmen. Hierfür testeten wir das Instrument erstmals an einer Stichprobe von Studierenden der Geschichtswissenschaft. Die Ergebnisse der CFA ergaben, dass im Wesentlichen die dreifaktorielle Struktur des ursprünglich niederländischen Instruments auch für seine deutsche Version anhand der vorliegenden Stichprobe mit akzeptablen FIT-Werten repliziert werden konnte (van Straaten et al., 2018). Es liegt somit ein Instrument vor, mit dem konstruktvalide Relevanzüberzeugungen von Studierenden in der Domäne Geschichte erfasst werden können. Je Subskala mussten wir jedoch ein Item aufgrund niedriger Ladungen entfernen. Ein genauerer Blick auf die entfernten Items zeigt, dass diese sehr allgemein gehalten und wenig spezifisch formuliert sind (z. B. «Durch Geschichte lerne ich wenig über mich selbst.» [IDE2]). Möglicherweise führte dies dazu, dass die Items nicht wie intendiert von den Studierenden verstanden wurden. Hier müsste zukünftig an der Inhaltsvalidität der ausgeschlossenen Items, etwa durch sprachliche und inhaltliche Präzisierung, gearbeitet werden. Die Reliabilität der drei Faktoren ist gut, lediglich der Omega-Wert für die HUM-Skala spricht für eine noch akzeptable interne Konsistenz. Van Straaten und Kollegen (2018) weisen bereits darauf hin, dass eine Schwierigkeit bei der Instrumentenentwicklung

darin bestand, psychometrisch valide Items zu konstruieren, die sich mit der Bewältigung andauernder menschlicher Probleme befassen. Die niedrige Reliabilität könnte allerdings auch an der geringen Anzahl an Items in der Subskala selbst liegen. Zwischen den drei Subskalen fanden wir, wie im niederländischen Original (van Straaten et al., 2018), substanzielle Korrelationen, die für starke Zusammenhänge sprechen.

Des Weiteren ergaben sich niedrige bis zumeist mittlere Korrelationen zwischen den drei Relevanzüberzeugungen und der Häufigkeit der Beschäftigung mit Geschichte einerseits sowie dem individuellen Interesse an Geschichte andererseits. Dies zeigte sich ebenfalls in früheren Studien (van Straaten et al., 2018; 2019) und spricht recht plausibel dafür, dass eine häufige Beschäftigung mit Geschichte außerhalb der Schule oder Universität und ein großes Interesse an Geschichte allgemein auch mit hohen Wertüberzeugungen assoziiert sind, die Personen Geschichte zu schreiben. Schwächere Korrelationen ergaben sich nur bei dem Teilkonstrukt «understanding the human condition» und den Außenkriterien, was eventuell mit der vergleichsweise niedrigeren Reliabilität der Subskala zusammenhängt.

Es zeigten sich bei den deutschsprachigen Studierenden deskriptiv relativ hohe Ausprägungen der geschichtsbezogenen Relevanzüberzeugungen. Vergleichbar mit früheren Studien aus Spanien und den Niederlanden fallen diese moderat stark aus (Miguel-Revilla, 2021; van Straaten et al., 2018). Dies ist erwartungskonform, da es sich bei den Proband*innen um Personen handelt, die in der Regel das Studienfach Geschichte aus eigenem Antrieb beziehungsweise intrinsischer Motivation wählten und sich mit diesem eine gewisse Zeit lang beschäftigen (Kanert, 2014, 78). Wie in Befragungen von spanischen und niederländischen Schüler*innen und Studierenden zeigt sich auch bei den deutschsprachigen Studierenden die im Vergleich niedrigsten Zustimmungswerte bezüglich des Teilkonstrukts «building a personal identity» (Miguel-Revilla, 2021; van Straaten et al., 2018). Dies könnte darauf hindeuten, dass Zusammenhänge zwischen historischen Sachverhalten und der eigenen Identität eher weniger bewusst als zu den anderen Relevanzdimensionen gezogen werden oder, dass solche Zusammenhänge seltener in der Schule, den Medien oder universitären Lehrveranstaltungen (wirksam) thematisiert werden (Andrews, McGlynn & Mycock, 2009). Die berechneten Gruppenvergleiche ergaben erwartungskonform höhere Zustimmungswerte zu den drei Relevanzüberzeugungen bei den M.Ed.-Studierenden als bei den B.A.-Studierenden, wobei die Unterschiede nur bei zwei von drei Teilkonstrukten statistisch signifikant waren. Dies scheint der Tendenz zu entsprechen,

dass Menschen mit aufsteigendem Alter höhere Zustimmungswerte bezüglich ihrer Relevanzüberzeugungen in der Domäne Geschichte aufweisen (Haerberli, 2005; Haydn & Harris, 2010; van Straaten et al., 2018) und generell bei Studierenden, zumindest der Geschichtswissenschaft, hohe Ausprägungen vorhanden sind (Berg, 2019). Die kleinen Effektstärken bezüglich der gefundenen Unterschiede sprechen jedoch dafür, dass bei Studierenden zwischen dem B.A.- und dem M.Ed.-Studiengang der Zuwachs an Relevanzüberzeugungen nur noch gering zunimmt.

Natürlich sind auch die vorliegenden Befunde mit einigen Limitationen verbunden. Wie bereits van Straaten und Kollegen (2018) für das Originalinstrument einräumten, erfasst die RHMS-D lediglich drei mögliche Facetten von Relevanzzuschreibungen in der Domäne Geschichte. Von einer weiteren Theoriearbeit könnte zum einen die Skala «understanding the human condition» (z. B. Erweiterung der Skala durch konstruktkonforme Items) profitieren und zum anderen könnten weitere denkbare Facetten in das bestehende Instrument integriert werden. Weiterhin führten wir diese erste Validierung an Studierenden der Geschichtswissenschaft durch, weshalb eine Testung an anderen Stichproben aussteht. Daher überprüfen wir gegenwärtig die Validität der RHMS-D zusätzlich an einer Stichprobe von Schüler*innen, um die Funktionsfähigkeit auch an diesen zu testen. Für die Gruppenvergleiche innerhalb dieses Beitrages muss zudem auf die ungleiche Verteilung von B.A.- und M.Ed.-Studierenden hingewiesen werden. Hier stocken wir die Stichprobe momentan mit M.Ed.-Studierenden auf, um zukünftig belastbarere Befunde vorzulegen.

Die vorliegende Studie bietet der deutschsprachigen geschichtsdidaktischen Forschung nichtsdestotrotz ein valides und reliables Instrument, um Relevanzüberzeugungen in der Domäne Geschichte messen zu können. Dies ist wichtig, um zukünftig interkulturell vergleichende Studien durchzuführen und möglichen Zusammenhängen mit Kompetenzen historischen Denkens, wie vor allem der historischen Orientierungskompetenz, nachzugehen (Trautwein et al., 2017). Weiterhin bietet sich das Instrument an, um in Interventionsstudien zu untersuchen, inwiefern unter anderem mit verschiedenen historischen Themen, Lehr-Lern-Methoden oder geschichtsdidaktischen Prinzipien (z. B. dem Gegenwartsbezug) Relevanzüberzeugungen in Geschichte bei Lernenden beeinflusst werden können (Hulleman & Harackiewicz, 2020; van Straaten et al., 2018; 2019). Auf diese Weise könnte es möglich sein, Geschichte im Unterricht, innerhalb der Vermittlung in Museen oder in der geschichtswissenschaftlichen Lehre an Universitäten ein wenig relevanter zu machen.

Literatur

- Andrews, Rhys/McGlynn, Catherine & Mycock, Andrew (2009). Students' attitudes towards history: Does self-identity matter? *Educational Research*, 51(3), 365–377.
- Angvik, Magne & von Borries, Bodo (Hrsg.) (1997). *Youth and history: A comparative European survey on historical consciousness and political attitudes among adolescents* (Bd. B). Hamburg: Körber-Stiftung.
- Berg, Christopher W. (2019). Why study history? An examination of undergraduate students' notions and perceptions about history. *Historical Encounters*, 6(1), 54–71.
- Borries, Bodo von (2001). Geschichtsdidaktik am Ende des 20. Jahrhunderts. Eine Bestandsaufnahme zum Spannungsfeld zwischen Geschichtsunterricht und Geschichtspolitik. In Hans-Jürgen Pandel & Gerhard Schneider (Hrsg.), *Wie weiter? Zur Zukunft des Geschichtsunterrichts* (S. 7–32). Schwalbach/Ts.: Wochenschau.
- Brechtken, Magnus (2020). *Der Wert der Geschichte: Zehn Lektionen für die Gegenwart*. München: Siedler.
- Browne, Micheal W. & Cudeck, Robert (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, 21(2), 230–259.
- Bundesministerium für Bildung, Wissenschaft und Forschung (2016). *Allgemeinbildende Höhere Schule: Unterstufe Lehrplan. Geschichte und Sozialkunde/Politische Bildung*. <https://www.ris.bka.gv.at/eli/bgbl/II/2016/113> (21.04.2022).
- Cohen, Jacob (1988). *Statistical power analysis for behavioral sciences* (2. Aufl.). Hillsdale: Erlbaum.
- Daumüller, Markus & Seidenfuß, Manfred (2014). Geschichtsunterricht – was beliebt? Die Sicht der Schulabgängerinnen und Schulabgänger. In Michael Sauer/Charlotte Bühl-Gramer/Anke John/Marko Demantowsky & Alfons Kenkmann (Hrsg.), *Geschichtslernen in biographischer Perspektive: Nachhaltigkeit – Entwicklung – Generationendifferenz* (S. 115–137). Göttingen: Vandenhoeck & Ruprecht.
- Daumüller, Markus & Seidenfuß, Manfred (2017). *Endstation Geschichtsunterricht: Die Sicht von Schulabgängerinnen und Schulabgängern auf ihren Geschichtsunterricht*. Münster: Lit.
- Deutschschweizer Erziehungsdirektorenkonferenz (2018). *Lehrplan 21. Kanton Schaffhausen: Natur, Mensch, Gesellschaft*. https://sh.lehrplan.ch/lehrplan_printout.php?e=1&k=1&fb_id=6 (21.04.2022).
- Enders, Craig K. (2010). *Applied missing data analysis*. New York, London: Guilford.
- Foster, Stuart/Ashby, Ros & Lee, Peter (2008). *Usable historical pasts: A study of students' frameworks of the past: Full research report ESRC end of award report*. RES-000-22-1676. Swindon: ESRC.
- Gaspard, Hanna/Nagengast, Benjamin & Trautwein, Ulrich (2019). Erfassung von Wertüberzeugungen. In Hanna Gaspard/Ulrich Trautwein & Markus Hasselhorn (Hrsg.), *Diagnostik und Förderung von Motivation und Volition* (S. 87–100). Göttingen: Hogrefe.
- Haag, Ludwig & Götz, Thomas (2012). Mathe ist schwierig und Deutsch aktuell: Vergleichende Studie zur Charakterisierung von Schulfächern aus Schülersicht. *Psychologie in Erziehung und Unterricht*, 59, 32–46. <https://doi.org/10.2378/peu2012.arto3d>.

- Haerberli, Philippe (2005). Relating to history. An empirical typology. *International Journal of Historical Learning, Teaching and Research*, 5(1), 19–29.
- Harkness, Janet A. (2003). Questionnaire translation. In Janet A. Harknes/Fons van de Vijver & Peter Ph. Mohler (Hrsg.), *Cross-cultural survey methods* (S. 35–56). Hoboken: Wiley.
- Harris, Richard & Reynolds, Rosemary (2014). The history curriculum and its personal connection to students from minority ethnic backgrounds. *Journal of Curriculum Studies*, 46(4), 464–486. <https://doi.org/10.1080/00220272.2014.881925>.
- Hartig, Johannes/Frey, Andreas & Jude, Nina (2012). Validität. In Helfried Moosbrugger & Augustin Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 143–171). Berlin: Springer. https://doi.org/10.1007/978-3-642-20072-4_7.
- Hattie, John & Cooksey, Ray W. (1984). Procedures for assessing validities of tests using the «known-group» method. *Applied Psychological Measurement*, 8(3), 295–305.
- Haydn, Terry & Harris, Richard (2010). Pupil perspectives on the purposes and benefits of studying history in high school: A view from the UK. *Journal of Curriculum Studies*, 42(2), 241–261. <https://doi.org/10.1080/00220270903403189>.
- Hu, Li-tze & Bentler, Peter M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Hulleman, Chris & Harackiewicz, Judith M. (2021). The utility-value intervention. In Gregory W. Walton & Alia J. Crum (Hrsg.), *Handbook of wise interventions: How social-psychological insights can help solve problem* (S. 100–125). New York: Guilford.
- Jeismann, Karl-Ernst (1980). Geschichtsbewusstsein. Überlegungen zu einer zentralen Kategorie eines neuen Ansatzes der Geschichtsdidaktik. In Hans Süßmuth (Hrsg.), *Geschichtsdidaktische Positionen* (S. 179–222). Paderborn: UTB.
- Kanert, Georg (2014). Motivationale Beweggründe für die Wahl des (Geschichts)Lehrerberufs und des Studienfachs Geschichte. In Tobias Arand & Manfred Seidenfuß (Hrsg.), *Neue Wege – neue Themen – neue Methoden? Ein Querschnitt aus der geschichtsdidaktischen Forschung des wissenschaftlichen Nachwuchses* (S. 65–82). Göttingen: V&R unipress.
- Kindervater, Angelika & von Borries, Bodo (1997). Historical motivation and historical-political socialization. In Magne Angvik & Bodo von Borries (Hrsg.), *Youth and history: A comparative European survey on historical consciousness and political attitudes among adolescents* (Bd. A, S. 48–106). Hamburg: Körber-Stiftung.
- Kocka, Jürgen (1974). Geschichtswissenschaft heute – wozu noch Geschichte? In Karl Filser (Hrsg.), *Theorie und Praxis des Geschichtsunterrichts* (S. 24–35). Bad Heilbrunn: Klinkhardt.
- Köller, Olaf/Baumert, Jürgen & Schnabel, Kai Uwe (2000). Zum Zusammenspiel von schulischem Interesse und Lernen im Fach Mathematik: Längsschnittanalysen in den Sekundarstufen I und II. In Ulrich Schiefele & Klaus Peter Wild (Hrsg.), *Interesse und Lernmotivation – Untersuchungen zu Entwicklung, Förderung und Wirkung* (S. 163–183). Münster: Waxmann.
- Mayer, Ulrich/Gautschi, Peter & Bernhardt, Markus (2012). Themenbestimmung im Geschichtsunterricht der Sekundarstufen. In Michele Barricelli & Martin Lücke (Hrsg.),

- Handbuch Praxis des Geschichtsunterrichts* (Bd. 1, S. 378–404). Schwalbach/Ts.: Wochenschau.
- Meyer-Hamme, Johannes (2018). Was heißt «historisches Lernen»? Eine Begriffsbestimmung im Spannungsfeld gesellschaftlicher Anforderungen, subjektiver Bedeutungszuschreibungen und Kompetenzen historischen Denkens. In Thomas Sandkühler/Charlotte Bühl-Gramer/Anke John/Astrid Schwabe & Markus Bernhardt (Hrsg.), *Geschichtsunterricht im 21. Jahrhundert. Eine geschichtsdidaktische Standortbestimmung* (S. 75–92). Göttingen: Vandenhoeck & Ruprecht.
- Miguel-Revilla, Diego (2021). What is history education good for? A comparative analysis of students' conceptions about the relevance of history. *Journal of Curriculum Studies*, 54(1). <https://doi.org/10.1080/00220272.2021.1896784>.
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2014). *Kernlehrplan für die Sekundarstufe II Gymnasium/Gesamtschulen in Nordrhein-Westfalen. Geschichte*. Düsseldorf: Ritterbach. https://www.schulentwicklung.nrw.de/lehrplaene/upload/klp_SII/ge/KLP_GOST_Geschichte.pdf (21.04.2022).
- Moosbrugger, Helfried & Schermelleh-Engel, Karin (2012). Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In Helfried Moosbrugger & Augustin Kaleva (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 325–343). Berlin: Springer. https://doi.org/10.1007/978-3-642-20072-4_13.
- Nipperdey, Thomas (1972). Über Relevanz. *Geschichte in Wissenschaft und Unterricht*, 23, 577–596.
- Rosseel, Yves (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–3. <https://doi.org/10.18637/jss.v048.i02>.
- Rüsen, Jörn (2013). *Historik. Theorie der Geschichtswissenschaft*. Köln: Böhlau.
- Rüsen, Jörn (2020). *Geschichte denken: Erläuterungen zur Historik*. Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-658-29275-1>.
- Schönemann, Bernd (2014). Relevanz. In Ulrich Mayer/Hans-Jürgen Pandel/Gerhard Schneider & Bernd Schönemann (Hrsg.), *Wörterbuch Geschichtsdidaktik* (3. Aufl., S. 166–168). Schwalbach/Ts.: Wochenschau.
- Schreiber, Waltraud (2002). Reflektiertes und (selbst-)reflexives Geschichtsbewusstsein durch Geschichtsunterricht fördern – ein vielschichtiges Forschungsfeld der Geschichtsdidaktik. *Zeitschrift für Geschichtsdidaktik*, 1, 18–43.
- Stearns, Peter N. (1998). Why study history? [https://www.historians.org/about-aha-and-membership/aha-history-and-archives/historical-archives/why-study-history-\(1998\)](https://www.historians.org/about-aha-and-membership/aha-history-and-archives/historical-archives/why-study-history-(1998)) (21.04.2022).
- Straaten, Dick van/Wilschut, Arie & Oostdam, Ron (2016). Making history relevant by connecting past, present and future. *Journal of Curriculum Studies*, 48(4), 479–502. <https://doi.org/10.1080/00220272.2015.1089938>.
- Straaten, Dick van/Wilschut, Arie & Oostdam, Ron (2018). Measuring students' appraisals of the relevance of history. The construction and validation of the Relevance of History Measurement Scale (RHMS). *Studies in Educational Evaluation*, 56, 102–111. <https://doi.org/10.1016/j.stueduc.2017.12.002>.
- Straaten, Dick van/Wilschut, Arie/Oostdam, Ron & Fukkink, Ruben (2019). Fostering students' appraisals of the relevance of history by comparing analogous cases of an endu-

- ring human issue. A quasi-experimental study. *Cognition and Instruction*, 37(4), 512–533. <https://doi.org/10.1080/07370008.2019.1614590>.
- Tanner, Jakob (2011). Lehren der Geschichte – Lernen aus der Vergangenheit? In Hans-Ulrich Grunder (Hrsg.), *Aus der Geschichte lernen: Die historische Perspektive in der Ausbildung von Lehrerinnen und Lehrern* (S. 272–277). Baltmannsweiler: Schneider Hohengehren.
- Trautwein, Ulrich et al. (2017). *Kompetenzen historischen Denkens erfassen. Konzeption, Operationalisierung und Befunde des Projekts «Historical Thinking Competencies in History» (HiTCH)*. Münster: Waxmann.
- VanSledright, Bruce A. (1997). And Santayana lives on: Students' views on the purposes for studying American history. *Journal of Curriculum Studies*, 29(5), 529–557.
- Verband der Geschichtslehrer Deutschlands (2011). *Bildungsstandards Geschichte. Sekundarstufe I. Kompetenzmodell und synoptische Darstellung der Kompetenzen und verbindlichen Inhalte des Geschichtsunterrichts*. Entwurf. Vom Arbeitskreis Bildungsstandards dem VGD zur Diskussion vorgelegt auf dem Historikertag am 29. September 2010. https://www.geschichtslehrerforum.de/VGD_Bildungsstandards_Entwurf2010.pdf (21.04.2022).
- Waldis, Monika & Buff, Alex (2007). Die Sicht der Schülerinnen und Schüler – Unterrichtswahrnehmung und Interesse. In: Peter Gautschi/Daniel Moser/Kurt Reusser & Pit Wiher (Hrsg.), *Geschichtsunterricht heute. Eine empirische Analyse ausgewählter Aspekte* (S. 177–210). Bern: hep.
- Wigfield, Allan & Eccles, Jacquelynne, S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68–81. <https://doi.org/10.1006/ceps.1999.1015>.
- Wilschut, Arie (2019). The relevance of history to students: How to improve it? In Enrico Valseriati (Hrsg.), *Prospettive per la didattica della storia in Italia e in Europa* (S. 131–154). Palermo: New Digital Press.
- Wilschut, Arie (2013). *De taal van burgerschap* [Die Sprache der Staatsbürgerschaft]. Amsterdam: Hogeschool van Amsterdam.

GeDiKo-T

Zur Entwicklung und Implementation eines Kompetenztests für Geschichtslehrpersonen

Stefanie Zabold¹

1 Einleitung

Wandel, insbesondere, wenn er als Zeitenwandel erfahren wird, stellt an Geschichtslehrpersonen hohe Anforderungen: Sie müssen ihre Schüler*innen dazu befähigen, sich gerade auch durch die Auseinandersetzung mit Vergangenen in der jeweils eigenen Gegenwart zu orientieren (Rüsen, 2008; Körber, Schreiber & Schöner, 2007; Bräuer, 2021). Entsprechend darf Geschichtsunterricht nicht nur vergangenheitsbezogenes historisches Lernen initiieren, sondern muss über den historischen Inhalt hinausweisen und auch die unmittelbare Gegenwart und die (vermutlich) zu erwartende Zukunft im Blick behalten. Nur so wird «Orientierung durch Geschichte» möglich und Geschichtsunterricht kann die Sinnbildung für das eigene Leben unterstützen.

«Lehrerkompetenzen bilden die Voraussetzung für Schülerkompetenzen» (von Borries, 2007, 60; vgl. auch Baumert & Kunter, 2011, 163–192). Die Schlussfolgerung daraus ist, dass Geschichtslehrkräfte über die Fähigkeit, Fertigkeit und Bereitschaft verfügen müssen, das Historisch-denken-Lernen ihrer Schüler*innen zu ermöglichen. Bei diesen Fähigkeiten, Fertigkeiten und Bereitschaften handelt es sich um geschichtsdidaktische Kompetenzen (vgl. Hasberg, 2010, 175–179; Heuer, Körber, Schreiber & Waldis, 2019; Körber, Heuer & Schreiber, i. E.).

¹ Unter Mitarbeit von Wolfgang Wagner.

Wenngleich die Auseinandersetzung mit Kompetenzen von Geschichtslehrkräften noch eher jung ist (vgl. Hasberg, 2010, 164)², befasst sich die Geschichtsdidaktik seit einiger Zeit sowohl theoretisch mit Geschichtslehrerkompetenzen als auch mit methodischen Möglichkeiten, diese Kompetenzen zu erforschen beziehungsweise zu diagnostizieren und mit deren Entwicklung durch Lehrerbildung.³ Trotz dieses zunehmenden Interesses gilt, was Hover und Hicks in internationaler Perspektive feststellen, auch für den deutschsprachigen Raum, nämlich dass die Erforschung von Geschichtslehrerkompetenzen noch unsystematisch ist und erst wenige empirische Arbeiten vorliegen.⁴ Hier knüpft das vorliegende Projekt an, das die Entwicklung eines Tests zum Ziel hat, der Ausprägungen von geschichtsdidaktischen Kompetenzen messen will.

Der vorliegende Beitrag skizziert die Entwicklung des Kompetenztests GeDiKo-T, seine Pilotierung, Validierung und Standardisierung: Im ersten Schritt werden hierfür Möglichkeiten der Anknüpfung an den aktuellen Stand der Erforschung von Lehrerkompetenzen in der Geschichtsdidaktik aufgezeigt und die theoretische Grundlage des Tests geklärt. Darauf folgt die Besprechung der Test- und Aufgabenstruktur und die Erläuterung der Testentwicklung. Anhand von drei unterschiedlichen Aufgabenstellungen – zur geschichtsdidaktischen Methodenkompetenz, zur Strukturierungs- und Kategorisierungs- sowie zur Diagnosekompetenz – werden diese Einblicke vertieft. Die Darstellungen der Pilotierungsergebnisse und der darauf fußenden aktuellen Version des Tests runden das Kapitel zur Testentwicklung ab. Sodann wird beschrieben, wie der Test aktuell in Forschungsprojekten eingesetzt ist. Abschließend erfolgt die Diskussion seiner Einordnung in den geschichtsdidaktischen Forschungskontext.

2 Wolfgang Hasberg (2010) stellt in seiner den Stand der Empirie reflektierenden Habilitationsschrift noch 2010 fest, dass «in der Geschichtsdidaktik [...] der Lehrer selten zum Gegenstand des Interesses geworden [ist] – viel weniger noch zum Gegenstand der Forschung» (160). Michael Sauer (2019) sieht dies anlässlich der 19. Zweijahrestagung der Konferenz für Geschichtsdidaktik in Augsburg zum Thema «Zur Professionalisierung der Geschichtslehrerbildung – nationale und internationale Perspektiven» (19) weiterhin bestätigt.

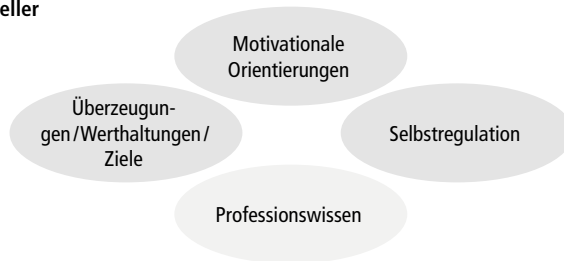
3 Vgl. Zeitschrift für Geschichtsdidaktik 2014; Beiträge in Sammelbänden zum Beispiel zu den Tagungen *geschichtsdidaktik empirisch 15* (Waldis & Ziegler, 2017) oder *geschichtsdidaktik empirisch 17* (Waldis & Ziegler, 2019), Einzelbeiträgen in einschlägigen Zeitschriften (wie *Didaktik der Gesellschaftswissenschaften*, *Unterrichtsforschung*) sowie erste Monografien (z. B. Kanert, 2014; Resch, 2018).

4 «Empirical work on history teacher education internationally is exceptionally limited» (van Hover & Hicks, 2018, 397).

2 Anknüpfungspunkte im Forschungsstand zur Messung von Kompetenzen bei Geschichtslehrpersonen

Die meisten empirischen Studien zu Kompetenzen bei Geschichtslehrpersonen orientieren sich an Kompetenzmodellierungen zur professionellen Kompetenz von Lehrkräften aus der Pädagogischen Psychologie/Allgemeindidaktik/Bildungsforschung (Baumert & Kunter, 2006; Blömeke, Kaiser & Lehmann, 2010 (TEDS-M-Studie); Kunter et al., 2011 (COACTIV-Studie)), die letztlich die Shulman'sche Dimensionierung von Professionswissen aufgreifen, also die Unterscheidung von Fachwissen, fachdidaktischem Wissen und pädagogischem Wissen (Shulman, 1986; 1987).

Aspekte professioneller Kompetenz



Kompetenzbereiche

Fachwissen
Fachdidaktorisches Wissen
Pädagogisch-psychologisches Wissen
Organisationswissen
Beratungswissen

Kompetenzfacetten

→	Tiefes Verständnis der Schulmathematik
↗	Erklärungswissen
→	Wissen über das mathematische Denken von Schülern
↘	Wissen über mathematische Aufgaben
↗	Wissen um Leistungsbeurteilung
→	Wissen über Lernprozesse
↘	Wissen über effektive Klassenführung

Abbildung 1: Aspekte professioneller Kompetenz aus dem COACTIV-Forschungsprogramm (Klumann, Kunter, Richter & Voss, o.J., 6).

Am Bezug der Geschichtsdidaktik auf das am Beispiel des Mathematik-Unterrichtens entwickelte und in der Lehrerforschung äußerst einflussreiche Modell professioneller Kompetenz von Lehrkräften wird zum Teil leise Kritik geübt. Sei-

denfuß verweist in seinem Einführungsbeitrag zum Themenheft «Forschungsfeld Geschichtslehrkräfte» der Zeitschrift für Geschichtsdidaktik zum Beispiel darauf, dass das Verständnis von «Fachwissen» als «tiefes Verständnis des Schulcurriculums» und als «Begriffswissen» weder kompatibel mit der Geschichtswissenschaft noch der Geschichtsdidaktik ist (Seidenfuß, 2014, 13).

Solche Grundsatzfragen werden in den einzelnen Studien der «COACTIV-Follower» eher wenig thematisiert. Sie fokussieren fachdidaktisches Wissen (vgl. Kanert & Resch, 2014; Resch, 2018; Resch, Heuer & Seidenfuß, 2019; Waldis, 2019) oder Fachwissen (Fenn et al., 2019; Wolf, Seiffert, Seifert, Rothland & Brauch, 2018). Oft wählen sie dabei Einzelaspekte aus, die sie explizit oder implizit als geschichtsdidaktische Wendung der Shulman'schen Dimensionierung ansehen.⁵ Zum Teil werden Wissensdimensionen und Geschichtslehrerhandeln gemeinsam betrachtet (vgl. Waldis, 2020), z. T. gesondert, wobei in Bezug auf geschichtsdidaktisches Handeln «Aufgabenstellungen» besonderes Interesse auf sich ziehen (vgl. Mägdefrau & Michler, 2018; Resch, 2018).

Methodisch bedienen sich die bislang vorliegenden Studien der Geschichtsdidaktik des breiten aus der empirischen Bildungsforschung bekannten Spektrums. Sie arbeiten quantitativ (z. B. Berufseinsteiger-Kompetenzen, vgl. Kanert, 2014), sind als Vignettenstudien angelegt (vgl. Kanert & Resch, 2014; Waldis, Nitsche, Marti, Hodel & Wyss, 2014), greifen mindestens als Grundlage für Testentwicklung qualitativ zu (vgl. Wolf et al., 2018) oder mixen Methoden (vgl. Waldis et al., 2020).

Die Fachspezifik als Ausgangspunkt zu wählen statt als einen Fall (neben Mathematik, Deutsch, Sport, Naturwissenschaften ...) ist der andere Weg, der in den letzten Jahren in der Geschichtsdidaktik verfolgt wird. Er wird insbesondere von der Gruppe um Christian Heuer (Graz), Andreas Körber (Hamburg) und Waltraud Schreiber (Eichstätt) vorangetrieben (Heuer et al., 2019; Körber et al., i. E.). Den Ausgangspunkt bildet dabei die theoretische Modellierung eines geschichtsdidaktischen Kompetenzstrukturmodells (GeDiKo). Bezüge zu den drei Dimensionen des Professionswissens werden ausgehend von den fachspezifischen Überlegungen hergestellt.

5 Der Heidelberger Forschungsverbund EKoL etwa «hat zum Ziel, Professionalisierungsprozesse in der Lehrerbildung empirisch zu untersuchen, wobei die Forschungsprozesse auf der Bestimmung von wesentlichen Determinanten fachdidaktischen Wissens und Könnens einerseits und der Untersuchung von Entwicklungsprozessen andererseits liegen. Dazu wird ein theoretisches Modell zu Professionalisierungsprozessen in der Lehrerbildung vorge schlagen» (Rutsch, Vogel, Seidenfuß, Dörfler & Rehm, 2017, 9).

3 GeDiKo als theoretische Grundlage für den Kompetenztest GeDiKo-T

Das Kompetenzstrukturmodell bezieht sich nicht nur auf Geschichtsunterricht, sondern auf geschichtsdidaktisches Handeln insgesamt. Den Bezugsrahmen bilden Aufgabengebiete und Handlungsfelder sowie die in der Schnittmenge darauf liegenden Tätigkeitsfelder.

Die Aufgabengebiete definieren geschichtsdidaktisches Handeln: Sie umfassen nicht nur die Ver-Mittlung, sondern auch Forschung in Theorie und Empirie oder Normsetzungen. Die unübliche Schreibweise des Aufgabenbereichs der «Ver-Mittlung» signalisiert, dass es nicht (nur) um ein Beibringen und Befähigen, um ein Vermitteln im landläufigen Sinne, geht. «Ver-Mittlung» zielt wesentlich auch auf eine reflektierte Auseinandersetzung mit unterschiedlichen Perspektiven, Deutungen, Wertungen, mit unterschiedlichen methodischen Zugriffsweisen und die Befähigung zum eigenständigen Umgang damit.

Handlungsfelder sind neben Schule unter anderem Museen, Gedenkstätten, dazu Print-, AV-, digitale Medien, Film, Theater, aber auch Universitäten, Akademien und andere Bildungseinrichtungen.⁶ Geschichtsunterricht ergibt sich als ein *Tätigkeitsfeld* geschichtsdidaktischen Handelns, verortet im Schnittbereich zwischen dem Handlungsfeld Schule und dem Aufgabengebiet der Ver-Mittlung.

6 Die geschichtsdidaktisch handelnden Akteur*innen adressieren spezifische Rezipient*innengruppen. Zu den Akteuren zählen etwa Lehrkräfte, Dozent*innen, Kurator*innen, Museumspädagog*innen, Filmemacher*innen, Youtuber*innen, Wissenschaftsjournalist*innen, aber auch Wissenschaftler*innen, die diese Handlungsfelder beforschen, darunter nicht zuletzt Vertreter der akademischen Geschichtsdidaktik oder der Bildungsforschung. Die adressierten Rezipient*innengruppen sind u. a. Schüler*innen, Studierende, Besucher*innen von Museen oder Gedenkstätten, wissenschaftliche Communitys sowie, weit weniger spezifisch, Leser*innen, Zuseher*innen, Zuhörer*innen, die z. B. im WWW auf Ergebnisse geschichtsdidaktischen Handelns «treffen».



Abbildung 2: Das Kompetenzstrukturmodell GeDiKo-T unter Bezugnahme auf Geschichtsunterricht

Im auf dieser theoretischen Basis erarbeiteten Kompetenzstrukturmodell liegt der Schwerpunkt auf der Modellierung der Kompetenzbereiche. Unterschieden werden prozedurale Kompetenzen und die Kompetenz, den Horizont des für geschichtsdidaktisches Handeln relevanten Wissens zu kategorisieren und zu strukturieren. Die Strukturähnlichkeit mit dem Kompetenzstrukturmodell historischen Denkens der FUER-Gruppe ergibt sich aus der disziplinären Nähe.

Zu den prozeduralen Kompetenzen zählen geschichtsdidaktische Fragekompetenz, geschichtsdidaktische Methodenkompetenz und geschichtsdidaktische Orientierungskompetenz. Die geschichtsdidaktische Kategorisierungs- und Strukturierungskompetenz umfasst den Umgang mit Wissensbeständen aus der Geschichtsdidaktik, der historischen Forschung, der Geschichtstheorie, aber auch der Erziehungswissenschaft, aus der Soziologie und anderen Gesellschaftswissenschaften oder aus Kulturwissenschaften. Die Tätigkeitsfelder (etwa der Geschichtsunterricht) haben Rückwirkungen auf die für das geschichtsdidaktische Handeln notwendigen Wissensbestände, deren Tiefe und Reichweite. An

dieser Stelle sind die drei Shulman'schen Dimensionen des «Professionswissen» verortet; für das geschichtsdidaktische Handeln von Geschichtslehrkräften ist aber ein weiterer Wissenshorizont notwendig. So kann historische Orientierungsfähigkeit einer zunehmend heterogenen Schülerschaft nicht ohne Kultur- und gesellschaftswissenschaftliches Wissen aufgebaut werden.

Bezogen auf dieses Kompetenzstrukturmodell wurde parallel der Eichstätter Geschichtslehrerkompetenztest GeDiKo-T entwickelt.

4 Zur Entwicklung des Eichstätter Lehrerkompetenztests GeDiKo-T⁷ – Stichprobe und Aufgabenstellungen

4.1 Stichprobe

Mit dem Test GeDiKo-T werden zunächst vornehmlich die geschichtsdidaktischen Kompetenzausprägungen von fertig ausgebildeten Geschichtslehrkräften beziehungsweise Referendar*innen in der zweiten Ausbildungsphase adressiert. Für diese Zielgruppe ist er in zweierlei Hinsicht nutzbar: Erstens ermöglicht er, den Ist-Stand der jeweils aktuellen Kompetenzausprägungen zu eruieren. Zweitens kann GeDiKo-T zum Beispiel im Rahmen von Interventionsprojekten genutzt werden, um die Veränderung von Kompetenzausprägungen, etwa durch Fortbildungsmaßnahmen, festzustellen. Die Stichprobe umfasste bewusst Geschichtslehrkräfte aller Schularten, da unabhängig vom Alter und von den Kompetenzausprägungen der Lernenden (Körber et al., 2007) immer das gleiche Ziel, die Förderung des Historisch-denken-Lernens (Meyer-Hamme, 2018) und damit historischer Kompetenzen, anvisiert wird. Die Annahme, dass alle Geschichtslehrkräfte über geschichtsdidaktische Kompetenzen verfügen müssen, um für ihre Schüler*innen einen gewinnbringenden kompetenzorientierten Geschichtsunterricht entwickeln und durchführen zu können, lag zugrunde.

7 Für die Testentwicklung haben wir uns u. a. an den Test-Entwicklungs-Strategien von Kingston, Scheuring und Kramer (2013) orientiert.

Die Verwendbarkeit des Tests in der Ausbildung, zum Beispiel während der ersten Phase an der Universität oder zur Diagnose geschichtsdidaktischer Kompetenzen am Übergang von der ersten in die zweite Ausbildungsphase⁸ wird derzeit erprobt.⁹

4.2 Aufgabenentwicklung

Der Test GeDiKo-T ist der Gruppe der Power- oder Niveautests zuzuordnen: Anhand der Aufgaben wird die Leistungsfähigkeit geschichtsdidaktischen Denkens, also geschichtsdidaktische Kompetenzausprägungen, ermittelt. Für die Bearbeitung der Aufgaben sind keine strikten zeitlichen Vorgaben gemacht (vgl. Schmidt-Atzert & Amelang, 2012); eine durchschnittliche Bearbeitungszeit von 30 bis 40 Minuten zeichnet sich ab.

Die Aufgaben und dazugehörigen Items des Eichstätter Lehrerkompetenztests beziehen sich auf die Kompetenzbereiche des zugrunde gelegten Theorie-modells GeDiKo. Zusätzlich berücksichtigt wird die in mehreren Kompetenzbereichen zu verortende Diagnosekompetenz der Lehrkräfte (vgl. auch Hasberg, 2010, 179; Pandel, 2007, 48). Der Test soll Kompetenzausprägungen anhand geschlossener (bzw. gebundener) Aufgabenformate erheben (vgl. Trautwein et al., 2017, 63; Köster, Bernhardt & Thünemann, 2016; Bühner, 2011, 110).

Bei der Entwicklung von Aufgaben zur Strukturierungs- und Kategorisierungskompetenz wurde unter anderem Bezug auf die Aufgaben des Heidelberger EKoL-Projekts (vgl. Kanert & Resch, 2014; Resch, 2018; Resch et al., 2019) genommen. Es wurde also auf bereits erprobte und der Forschungsdiskussion vorgelegte Aufgaben aufgebaut. Ausgewählt wurden Aufgaben, die das Verfügenkönnen über geschichtsdidaktisches und geschichtstheoretisches Wissen erfassen.¹⁰

Die den prozeduralen Kompetenzbereichen zugeordneten Aufgaben wurden dagegen neu entwickelt, ebenso die Aufgaben zur diagnostischen Kompetenz. Auf die Berücksichtigung der Qualitätsmerkmale geschichtsdidaktischer Reflektiertheit und (Selbst-/Fremd-)Reflexivität (Heuer et al., 2019, 106 ff.; Körber et al., i. E., Kapitel 2) wurde geachtet.

8 Zur Diskussion der Übergänge in der Geschichtslehrausbildung vgl. auch Erdmann (2013) oder Jung und Thünemann (2007).

9 Siehe Kapitel 6 zum derzeitigen Einsatz von GeDiKo-T in Forschungsprojekten.

10 Auch bezogen auf allgemeindidaktische Kompetenzmodelle wäre es unstrittig, dass damit Aspekte professioneller Lehrerkompetenz gemessen werden.

Bevor je eine Aufgabe aus den prozeduralen und kategorisierenden Kompetenzbereichen und eine Aufgabe zur Diagnosekompetenz vorgestellt und diskutiert werden, werden die Schritte der Testentwicklung (vgl. Kingston et al., 2013) beschrieben.

4.2.1 Entwicklungs- und Pilotierungsstufen des Tests GeDiKo-T

Die «Erstellung des Testentwurfs» (Bühner, 2011, 84), also der Pilotierungsfassung, nahm zehn aufeinander aufbauende Schritte in Anspruch, der elfte Schritt diente der «empirischen Überprüfung des Testentwurfs» (Bühner, 2011, 84) im Rahmen der Pilotierung, der zwölfte der Zusammenstellung des Tests GeDiKo-T (Bühner, 2011, 84). Folgend werden die Schritte aufgelistet und dann kurz charakterisiert.

1. Schritt: Entwicklung der Aufgabenideen auf der Grundlage des Theoriemo-
dells GeDiKo für die Zielgruppe Geschichtslehrkräfte in Sitzungen des Eich-
stätter Kompetenztestteams¹¹ unter der Leitung von Stefanie Zabold
2. Schritt: Operationalisierung der Aufgabenideen in den angezielten Aufga-
ben¹²
3. Schritt: Diskussion der konkreten Aufgaben in Eichstätter Teamsitzungen
(= inhaltliche Item-Validierung I)
4. Schritt: Erste Optimierung der Aufgaben
5. Schritt: Diskussion der optimierten Aufgaben in einem Item-Workshop mit
den «critical friends» Prof. Dr. Bodo von Borries und Prof. Dr. Béatrice Zieg-
ler (= inhaltliche Item-Validierung II)
6. Schritt: Zweite Optimierung der Aufgaben
7. Schritt: Cognitive Labs mit Lehrkräften (= inhaltliche Item-Validierung III)
8. Schritt: Dritte Optimierung der Aufgaben
9. Schritt: Experten-Rating mit den Mitgliedern des HiTCH-Konsortiums¹³ (=
inhaltliche Item-Validierung IV)
10. Schritt: Vierte Optimierung der Aufgaben

11 Unter der Federführung von Stefanie Zabold gehören zum Eichstätter Geschichtslehrerkom-
petenztest-Team Waltraud Schreiber, Matthias Hirsch, Tobias Langguth, David Naaß, Ina
Obermeyer, Susanne Sachenbacher und Michael Werner.

12 Schritt 1 und 2 entspricht der Item-Generierung (vgl. Kanert & Resch, 2014, 27–30).

13 Das HiTCH-Konsortium (HiTCH: Historical Thinking – Competencies in History) ging aus
einem BMBF-Projekt zur Entwicklung eines Kompetenztests für Schüler*innen hervor. Mitt-
lerweile arbeitet das Konsortium an einem Test für historische Kompetenzen der dritten
Generation: <http://www.hitch-projekt.de/>.

11. Schritt: Bundesweite, browserbasierte Pilotierung der Aufgaben mit dem Umfragetool Limesurvey (189 Items; $N = 186$; Geschichtslehrpersonen aus der dritten Phase)
12. Zusammenstellung des validierten und standardisierten Tests GeDiKo-T

Im Eichstätter Kompetenztestteam erfolgte nach ausführlicher Diskussion des Theoriemodells GeDiKo die Entwicklung von Aufgabenideen. Anschließend übernahmen die einzelnen Teammitglieder die Verantwortung für Aufgaben unterschiedlicher Kompetenzbereiche und legten erste Realisierungen vor (Schritt 1 und 2). Daran schloss sich ein Diskussionsprozess zu den Aufgabenentwürfen in unterschiedlichen Konstellationen an: Der Erstentwurf wurde im Eichstätter Team diskutiert (Schritt 3), der Zweitentwurf in den Item-Workshop mit den Critical Friends getragen (Schritt 5). Eine weitere Optimierungsrunde, teilweise noch einmal unter Einbindung der Critical Friends, schloss sich an (Schritt 6).

Schritt 7 bildeten Cognitive Labs (Werner & Schreiber, 2015), also die Erprobung der Aufgaben mit zur Zielgruppe gehörenden Geschichtslehrkräften. Dabei wurden die Aufgaben der Methode des lauten Denkens folgend bearbeitet und anschließend diskutiert. In zwei Runden, mit zeitlichem Abstand, wurden 17 Cognitive-Lab-Interviews in den Bundesländern Baden-Württemberg, Bayern, Berlin, Hessen und Nordrhein-Westfalen mit Lehrkräften unterschiedlicher Schularten (Gesamtschule, Gymnasium, Hauptschule, Mittelschule, Realschule, Sekundarschule) mit Proband*innen geführt, die zum Befragungszeitpunkt mehr als 10 Jahre Schulerfahrung hatten.

Tabelle 1: Sample für die Cognitive Labs zur Erprobung der Aufgaben mit zur Zielgruppe gehörenden Geschichtslehrkräften (Schritt 7)

Probandin/Proband	Geschlecht	Schulart	Bundesland
Proband 1	männlich	Hauptschule	Baden-Württemberg
Probandin 2	weiblich	Gymnasium	Baden-Württemberg
Probandin 3	weiblich	Realschule	Bayern
Proband 4	männlich	Realschule	Bayern
Proband 5	männlich	Realschule	Bayern
Proband 6	männlich	Mittelschule	Bayern
Probandin 7	weiblich	Mittelschule	Bayern
Proband 8	männlich	Gymnasium	Berlin

Probandin/Proband	Geschlecht	Schulart	Bundesland
Proband 9	weiblich	Gesamtschule	Hessen
Proband 10	männlich	Gesamtschule	Nordrhein-Westfalen
Probandin 11	weiblich	Gesamtschule	Nordrhein-Westfalen
Probandin 12	weiblich	Gesamtschule	Nordrhein-Westfalen
Probandin 13	weiblich	Sekundarschule	Nordrhein-Westfalen
Proband 14	männlich	Gymnasium	Nordrhein-Westfalen
Probandin 15	weiblich	Gesamtschule	Nordrhein-Westfalen
Probandin 16	weiblich	Hauptschule	Nordrhein-Westfalen
Proband 17	männlich	Gymnasium	Nordrhein-Westfalen

Den Cognitive Labs lag ein Interviewleitfaden zugrunde, der zum einen Fragen enthielt, die zwingend gestellt werden mussten: Wie verstehen Sie die Aufgabe? Was verstehen Sie unter ...? (Klärung des Verständnisses von in den Aufgaben verwendeten Fachbegriffen). Wie sind Sie bei der Lösung vorgegangen?

Ergänzend konnte die Interviewerin beziehungsweise der Interviewer situativ aus einem Pool weitere Fragen wählen, um das laute Denken der Proband*innen und die Diskussion zu intensivieren: Wieso kreuzen Sie genau so an? Wie kommen Sie darauf? Wie begründen Sie diese Lösung? Warum zögern Sie? Ist Ihnen etwas unklar? Zwischen welchen Alternativen schwanken Sie – warum? Wie sicher sind Sie sich hinsichtlich Ihrer Lösung? Ist die Aufgabe Ihrer Ansicht nach zu leicht, zu schwer oder passend – wie schätzen Sie die Aufgabenschwierigkeit ein? Haben Sie Anmerkungen zur Reihenfolge der Aufgaben?

Die Konsequenzen, die wir aus den transkribierten Interviews zogen (Schritt 8), waren vielfältig: Von der Änderung der Aufgabenreihenfolge, über den Austausch von Begrifflichkeiten bis hin zu mehr Klarheit in den Item-Formulierungen erfolgten zahlreiche, zum Teil umfassende Modifikationen, was die Zielgruppenspezifität der Aufgabenstellung klar erhöhte.

Bevor die Aufgaben in einer größeren Pilotierung eingesetzt wurden, erfolgte noch die inhaltliche Validierung in einem Experten-Rating (Schritt 9). 13 für die Geschichtsdidaktik ausgewiesene Experten (Professor*innen sowie wissenschaftliche Mitarbeiter*innen aus der Bundesrepublik Deutschland, Österreich und der Schweiz) haben die in den Schritten 1 bis 8 entwickelten Aufgaben geratet und kommentiert.

In den Test für die folgende Pilotierung aufgenommen wurden diejenigen Items, die mindestens von 75 Prozent der Expert*innen den Intentionen der Item-Konstrukteure folgend richtig beantwortet wurden.¹⁴ Zudem wurden die Aufgaben den Anmerkungen der Experten folgend ein weiteres Mal optimiert – in inhaltlicher wie sprachlicher Hinsicht. Vor Einsatz des Tests in der online-basierten Pilotierung erfolgte eine abschließende sprachliche Überprüfung, die sich ausschließlich auf Rechtschreibung und Grammatik bezog.¹⁵ An der Pilotierung waren 186 Lehrkräfte aus den Bundesländern Baden-Württemberg, Bayern, Berlin, Brandenburg, Hessen, Mecklenburg-Vorpommern, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland, Sachsen, Sachsen-Anhalt, Schleswig-Holstein, Thüringen und einige wenige aus Österreich beteiligt. Die Lehrer*innen unterrichten an verschiedenen Schularten, an der Grundschule, der Mittelschule, der Gesamtschule, der Realschule oder dem Gymnasium; z. T. sind sie auch in der Lehreraus- und -fortbildung aktiv. Gewonnen wurden sie durch Einladungen über die Bildungsserver der Länder.

4.2.2 Aufgabenstrukturen des Tests GeDiKo-T

Es wurde bereits erwähnt, dass im Test GeDiKo-T aus Gründen der Praktikabilität mit geschlossenen Aufgabenformaten gearbeitet wird. Dabei kommen verschiedene Ratingskalen zum Einsatz. Wenn von «sehr geeignet» bis «keinesfalls geeignet» oder «dem stimme ich unbedingt zu» bis «das finde ich überhaupt nicht» geurteilt werden soll, liegt der Fokus auf der Intensität (Rohrman, 1978). Um Richtig-falsch-Aufgaben (Bühner, 2011, 116–117) geht es bei den dichotomen Antwortformaten mit «ja»/«nein», «trifft zu»/«trifft nicht zu», «stimme zu»/«stimme nicht zu» beziehungsweise «geeignet»/«ungeeignet» oder «Perspektivenübernahme ist zustande gekommen»/«Perspektivenübernahme ist nicht zustande gekommen». Single-Choice-Aufgaben (Bühner, 2011, 116–117) kommen zum Einsatz, wenn die Testperson beispielsweise entscheiden muss, welcher Kompetenzbereich bei den Lernenden mit der anhand des Items getroffenen Aussage gefördert werden soll.

Einigen der Aufgaben sind weitere Materialien zugeordnet. Typische im Unterricht von Geschichtslehrkräften genutzte Materialien gehören dazu, wie Quellen oder zu den Themen verfasste geschichtliche Darstellungen. Durch die digi-

14 Vgl. Content-Review nach Wendler & Burrus, 2013, 284ff.

15 Vgl. Editing Review nach Wendler & Burrus, 2013, 287ff.

tale Aufbereitung des Tests ist auch der Einsatz von Vignetten möglich (vgl. Kanert & Resch, 2014; Waldis et al., 2014; allgemein: Seidel & Thiel, 2017). Dieses Format wurde gewählt, um Lehrkräften reale Unterrichtssituationen anzubieten, um damit deren Außensicht auf Unterricht zu erfassen (Seidel & Thiel, 2017): U. a. sollen sie beurteilen, ob das anvisierte Ziel der Unterrichtsstunde erreicht ist, nämlich das Hineinversetzen und die Perspektivenübernahme der Lernenden in Personen der Vergangenheit. Dafür beurteilen die Lehrkräfte, inwiefern schriftlich vorliegende Äußerungen von Schüler*innen das Erreichen des Stundenziels widerspiegeln.

4.2.3 Aufgabenbeispiel 1 zur geschichtsdidaktischen Strukturierungs- und Kategorisierungskompetenz

Wie bereits ausgeführt, wurde bei Aufgaben zur Strukturierungs- und Kategorisierungskompetenz Ansätze aus dem Heidelberger EKoL-Projekt («Effektive Kompetenzdiagnose in der Lehrerbildung», Seidenfuß et al.) aufgegriffen. Bezug genommen wurde auf den Forschungsschwerpunkt «Bestimmung von wesentlichen Determinanten fachdidaktischen Wissens und Könnens» (Rutsch, Vogel, Seidenfuß, Dörfler & Rehm, 2017, 9).

In einem Item-Workshop mit dem Heidelberger Kollegen Mario Resch wurden die Erfahrungen mit den EKoL-Items diskutiert. Die ausgewählten, und z. T. moderat überarbeiteten Testaufgaben beziehen sich auf zentrale für das geschichtsdidaktische Handeln relevante Konzepte wie: Historisches Erzählen, Quelle, Historische Darstellung, Verhältnis von Vergangenheit und Geschichte, Historische Sinnbildung, Re-Konstruktion, De-Konstruktion. Sie alle entsprechen geschichtsdidaktischen Basiskonzepten, die für eine reflektierte Unterrichtsplanung, -durchführung und -nachbereitung zentral sind. So müssen Geschichtslehrkräfte den Unterschied zwischen Vergangenheit und Geschichte beziehungsweise deren Beziehung kennen, wissen, dass Quellen wesentliche Informationsträger im Zusammenhang mit methodisch regulierter historischer Forschung sind oder sich der Orientierungsfunktion der Auseinandersetzung mit vergangenen Zeiten für das eigene Leben bewusst sein, um die entsprechenden Kompetenzen ihrer Schüler*innen durch geschichtsdidaktisches Lehrerhandeln fördern zu können. In der bearbeiteten Form durchliefen die Aufgaben die unterschiedlichen Entwicklungsschritte. Dabei erfuhren sie weitere Optimierungen. Unter anderem führten die Cognitive Labs zur Entscheidung, pro Begriffskonzept jeweils mit vier Items zu arbeiten.

Im Test werden die Lehrkräfte aufgefordert anzugeben, ob sie sich mit den jeweiligen Konzepten im Rahmen ihrer Aus- und/oder Fortbildungen befasst haben. Diese Zusatzfrage nahmen wir auf, weil zum Beispiel in Grund-, Mittel- und Gesamtschulen Lehrkräfte Geschichte oft fachfremd unterrichten. Das Antwortformat dieser Zusatzfrage ist dichotom: Die Proband*innen stimmen zu oder stimmen nicht zu.

4.2.4 Aufgabenbeispiel 2 zur geschichtsdidaktischen Methodenkompetenz

Das zweite Aufgabenbeispiel ist der geschichtsdidaktischen Methodenkompetenz zuzuordnen. Das Beispiel wird in der aktuellen Testversion nicht verwendet; es gibt aber Aufgaben und Items mit einer vergleichbaren Struktur.

Die Testpersonen werden hier in einem ersten Schritt dazu aufgefordert, einen Absatz aus einem Schweizer Schulbuchtext zu lesen. Dieser thematisiert als Kapiteleinführung die sich selbst zugeschriebene Neutralität der Schweiz im Ersten Weltkrieg. Die Schüler*innen werden im Text dazu aufgefordert, sich ein eigenes Urteil zu dieser Frage zu bilden:

«In diesem Kapitel geht es mir darum, dir unterschiedliche Meinungen zur Frage der Schweizer Neutralität während des Ersten Weltkriegs vorzulegen. Auf dieser Grundlage sollst du dir selbst ein Bild machen. Entscheide dann: In welchen Hinsichten war die Schweiz im Ersten Weltkrieg neutral, in welchen nicht?»

Auf diese Schulbuchpassage Bezug nehmend sollen die Lehrkräfte darüber nachdenken, was sie im weiteren Fortgang des so eingeleiteten Kapitels erwarten.

Die Aufgabe ist der geschichtsdidaktischen Methodenkompetenz zugeordnet. Indem die Lehrkräfte aufgefordert werden zu antizipieren, was in einem stringent aufgebauten Lehrwerk an Darstellungen und Materialien zu erwarten ist, wird überprüft, inwiefern sie Schulbuchtexte und Kapitelanlagen dekonstruierend analysieren können. Dabei steht nicht nur Inhaltliches im Zentrum, sondern insbesondere auch, welche Angebote gemacht werden müssen, damit Schüler*innen – wie gefordert – sich selbst ein Bild (hier von der Schweizer Neutralität) machen können.

Tabelle 2: Aufgabenbeispiel aus der Entwicklungsphase des GeDiKo-T zur geschichtsdidaktischen Methodenkompetenz.

Im an diesen Text anschließenden Kapitel wäre zu erwarten, dass ...					
		dem stimme ich unbedingt zu	dem stimme ich zu	das finde ich nicht	das finde ich überhaupt nicht
1.	... Materialien angeboten werden (z. B. Bildquellen), die illustrieren, dass die Schweiz während des Ersten Weltkriegs tatsächlich als Friedensinsel dargestellt wurde.	x			
2.	... gesagt wird, ob die Schweiz auch während des Ersten Weltkriegs weiterhin (wie vor dem Krieg) Rüstungsgüter verkaufte – auch an Kriegsmächte.		x		
3.	... gezeigt wird, ob die Schweiz weiterhin Nahrungsmittel oder Kohle von denselben Handlungspartnern bezog, wie vor dem Krieg – unabhängig davon, ob diese inzwischen am Krieg beteiligt waren.		x		
4.	... damals relevante völkerrechtliche Definition von Neutralität gegeben wird (Haager Landkriegsordnung von 1907).		x		
5.	... gezeigt wird, dass Neutralität im Kalten Krieg anders definiert wurde als während des Ersten Weltkriegs.				x
6.	... geschildert wird, inwiefern unter Schweizer Bürgern die Neutralität Zustimmung fand.		x		
7.	... die Ausbildung der Schweizer Armee dargestellt wird.				x
8.	... dass Belege vorgelegt werden, die zeigen, dass einzelne Schweizer Bürger für Frankreich und andere für Deutschland Partei ergriffen.			x	

Für die Auswertung werden die beiden zustimmenden und ablehnenden Aussagen zusammengefasst.

4.2.5 Aufgabenbeispiel 3 zur geschichtsdidaktischen Diagnosekompetenz

Eine weitere Aufgabe, die nach der Pilotierung Bestandteil des Tests geblieben ist, befasst sich mit der diagnostischen Kompetenz von Lehrkräften. Deman-

towsky und Waldis (2014) betonen die besondere Relevanz des Diagnostizierens unter anderem durch den Bezug auf die Hattie-Studie und hier insbesondere Hatties Wegweiser 3: «Wenn ich wissen will, wie Lernende im Unterricht denken, muss ich wissen, was sie denken. Dazu gehört sowohl die Diagnose von Lern- und Denkprozessen [...] als auch die Diagnose der Wissensstruktur» (Demantowsky & Waldis, 2014, 112). Theoretisch wie empirisch, so Waldis auch noch 2020 (Waldis, 2020, 438), sei das Diagnostizieren im Geschichtsunterricht wenig beforscht.

Das ist der Grund, warum auch hier eine Neuentwicklung der Aufgabe notwendig war. Getestet werden drei unterschiedliche Aspekte der Diagnosekompetenz von Lehrkräften:

- Positionierungen zur Bedeutsamkeit der Klärung der Lernausgangslage vor Beginn eines historischen Lernprozesses;
- zu berücksichtigende Aspekte bei der Diagnose der Lernausgangslage;
- Zusammenhang zwischen der Diagnose der Lernausgangslage und den jeweils gültigen Richtlinien.

Dabei werden die Lehrkräfte dazu aufgefordert, sich zu Statements, die die Diagnose der Lernausgangslage von Schüler*innen (Adamski, 2014; Barsch, 2020; Kühberger, 2016; Zabold, 2021) im Hinblick auf die drei benannten Schwerpunkte fokussieren, zu positionieren. Sie entscheiden sich, den getroffenen Aussagen zuzustimmen oder sie abzulehnen: Ich stimme zu/Ich stimme nicht zu.

Insgesamt bleibt die Ebene, auf der sich die Items bewegen, eine eher allgemeine. Sie betreffen Voraussetzungen und Rahmenbedingungen, nehmen aber noch keine konkreten Aussagen von Schüler*innen in den Blick oder erteilen den Auftrag, auf der Grundlage von Schüler*innenaussagen Entscheidungen für den weiteren Lernprozess zu treffen. Entsprechende Aufgaben sind in Bearbeitung und für einen Test GeDiKo-T 2.0 zu erwarten.

5 Der aktuelle Geschichtslehrerkompetenztest GeDiKo-T

Insgesamt wurden 136 Items des GeDiKo-T-Tests pilotiert. Anhand psychometrischer Methoden erfolgte eine Reduktion auf 71: Genutzt wurde dafür erstens eine unidimensionale Analyse basierend auf der Item-Response-Theorie (IRT) (Eingrenzung von signifikant positiv auf den latenten Faktor ladenden oder eine hohe Anzahl signifikanter Korrelationen mit den übrigen Testitems (über Median = 7 positive Interkorrelationen) aufweisenden Items). Zweitens wurden durch

exploratorische Faktoranalysen mit ein-, zwei- und dreifaktorieller Lösung weitere Items ausgeschlossen, die in keiner der drei Varianten eine signifikant positive Ladung auf einen der Faktoren aufwiesen. Auf Grundlage dieser Auswahl ergab die explorative Faktoranalyse eine plausible 2-Faktor-Lösung mit geringer Faktorkorrelation von $r(F_1, F_2) = .24^*$. (vgl. Bühner, 2011)

Zur inhaltlichen Interpretation der Faktoren wurden alle signifikant positiv auf einen der Faktoren ladenden Items ($F_1 k = 19, F_2 k = 28$) bezüglich formaler und inhaltlicher Merkmale untersucht. Eine formale Begründung der Faktorstruktur konnte dabei ausgeschlossen werden. Inhaltlich betrachtet ließen die Beobachtungen zu den Items je Faktor folgende Zusammenfassung zu:

- Faktor 1: Reflexion gegebener Unterrichtsplanung und -handlung
- Faktor 2: Geschichtsdidaktische Kategorisierungskompetenz als Bezugspunkt für Unterrichtsplanung, -handlung und -reflexion

Während Faktor 2 vor allem für die im GeDiKo-Modell ausgewiesene Strukturierungs- und Kategorisierungskompetenz steht, kommen in Faktor 1 in erster Linie die weiteren Kompetenzbereiche (= prozessbezogene Kompetenzbereiche) zum Ausdruck. Dies macht deutlich, dass die entwickelten Items dem gesetzten Ziel, geschichtsdidaktische Kompetenzen zu messen, gerecht werden.

6 Zum derzeitigen Einsatz von GeDiKo-T in Forschungsprojekten

Eine erste Implementation des Tests findet im Verbundprojekt KLUG (InKLUSiv Geschichte lehren; Prof. Dr. Waltraud Schreiber, Prof. Dr. Clemens Hillenbrand, Prof. Dr. Ulrich Trautwein; Förderung durch das BMBF)¹⁶ statt. Der Gediko-T-Test ist eines der Instrumente, die im Prä- und Postdesign eingesetzt wurden, um die Wirksamkeit des interdisziplinär erarbeiteten Fortbildungskonzepts zu überprüfen. Wegen der Coronapandemie musste die für das Schuljahr 2019/20 geplante Haupterhebung abgebrochen werden. Sie fand im Schuljahr 2020/2021 statt. Das KLUG-Projekt erlaubt es, den GeDiKo-T-Test mit qualitativen Daten zu Lehrerkompetenzen zu korrelieren. Die Ergebnisse werden 2022 vorliegen.

¹⁶ Vgl. hierzu die Projekt-Homepage: <https://klugprojekt.de/>.

Darüber hinaus wird GeDiKo-T für eine Studie zur ersten Phase der Lehrerbildung genutzt, ebenfalls im Prä-Post-Design: Erstsemester-Studierende des Lehramts Geschichte (Grund-, Mittelschule- und Realschule sowie Gymnasium) haben den Test vor und nach Besuch ihres ersten geschichtsdidaktischen Moduls bearbeitet. Die Erkenntnisse werden mit qualitativen Erhebungsmethoden im Sinne der Diagnose von Lernausgangslagen trianguliert. Die Triangulation aller Daten und der damit erzielten Ergebnisse lässt auf Erkenntnisse zur Nutzbarkeit des Tests in der ersten Phase der Lehrerbildung zu Beginn des Studiums hoffen.

7 Diskussion

Wie einleitend festgestellt, setzt die Empirie zu den Kompetenzen von Geschichtslehrkräften erst an. Auch wenn die Anzahl der Studien zunimmt und die methodische Qualität der Arbeiten, wie dies insgesamt für die Geschichtsdidaktik festgestellt werden kann, durchaus überzeugt (vgl. jüngst: Barsch & Barte, 2021), muss das Monitum ernst genommen werden, dass die Auseinandersetzung eher unsystematisch erfolgt.

Das Projekt GeDiKo-T leistet einen Beitrag zur Systematisierung. Diese ergibt sich daraus, dass mit dem Kompetenzstrukturmodell GeDiKo eine Theoriebasis zugrunde gelegt wird, die die Fachspezifik geschichtsdidaktischen Handelns betont. Der Gediko-T-Test enthält Aufgaben zu allen Kompetenzbereichen. Er bietet damit ein Set an, das versucht, geschichtsdidaktische Kompetenzausprägungen systematisch zu erfassen.

Dass die beiden in der Pilotierung herausgearbeiteten Faktoren die Grundstruktur des Modells stützen, nämlich die Unterscheidung prozeduraler und strukturierend-kategorisierender Kompetenzbereiche, kann als wertvoller erster Schritt der empirischen Validierung angesehen werden. Die aktuelle Nutzung der psychometrisch bereinigten Fassung in zwei weiteren Projekten ($N = \text{ca. } 200$, bzw. $\text{ca. } 50$) eröffnet die Chance, zeitnah zu prüfen, inwiefern sich die Faktoren bestätigen.

Auch für den Entwicklungsprozess des quantitativen Testinstruments wurde ein systematisches Vorgehen angestrebt. Dem State of the Art entspricht nicht nur der Entwicklungsprozess, sondern ebenso die Datenanalyse, die Wolfgang Wagner vom Hector-Institut für Empirische Bildungsforschung der Universität Tübingen durchgeführt hat. Trotz der – bezogen auf die hier im Fokus

stehende Zielpopulation – relativ umfangreichen Stichprobe sind den «Standard»-Analyseverfahren (mit Maximum-Likelihood-Schätzern) im IRT-Kontext in solchen Fällen typischerweise enge Grenzen gesetzt. Mithilfe eines Bayes-Schätzers konnten hier allerdings sogar komplexe mehrdimensionale IRT-Modelle analysiert werden.

Die Möglichkeiten der Triangulation des Gediko-T-Tests mit qualitativen Daten, die sich durch seinen Einsatz im KLUG-Projekt ergeben, eröffnet weitere Möglichkeiten der Systematisierung. Es handelt sich um Studien, die jeweils inhaltsanalytisch arbeiten. Das Datenmaterial der ersten Studie (Stefanie Hölzlwimmer) sind Aufgaben, die die Lehrkräfte im Anschluss an jede der 8 Fortbildungsphasen erledigen. Der Fokus der Analyse liegt auf der Kompetenzentwicklung der Lehrkräfte, wobei Reflektiertheit und Selbst-/Fremdreflexivität Qualitätskriterien sind. Das Datenmaterial der zweiten Studie (Susanne Sachbacher) ergibt sich aus der Beobachtung des Unterrichts von 4 Teilnehmer*innen der Fortbildung vor Beginn der Fortbildung, nach jeder der 8 Maßnahmen und 6 Wochen nach Beendigung der gesamten Reihe. Der jeweils zugrunde liegende Beobachtungsleitfaden zielt auf geschichtsdidaktische Kompetenzen (nach GeDiKo) und die Berücksichtigung der Basisdimension von Unterrichtsqualität (Klieme, Schümer & Knoll, 2001). Der Fokus der Analyse liegt hier auf der Entwicklung der Qualität des kompetenzorientierten Unterrichts von Geschichte.

Im Zuge des KLUG-Projekts kann zusätzlich geprüft werden, ob sich Zusammenhänge zwischen den Ergebnissen des GeDiKo-T-Tests und Schülerdaten ergeben. Im Prä-Post-Design werden – mithilfe von Aufgaben aus dem HITCH-Test – Kompetenzdaten erhoben und Daten zur Einschätzung der Qualität von Geschichtsunterricht anhand einer adaptierten Form des Fragebogens der Studie von Richard Göllner und Kollegen (2016). Beteiligt sind jeweils ca. 2000 Schüler*innen.

Außerdem wird anhand einer weiteren Studie (Lernstandsdiagnosen bei Studienanfänger*innen für ein Lehramt Geschichte) des kumulativ angelegten Habilitationsprojekts von Stefanie Zabold (Diagnose geschichtsdidaktischer Lehrerkompetenzen¹⁷) derzeit geprüft, ob der Gediko-T-Test grundsätzlich auch für Kompetenzmessungen in der Erstausbildung im Rahmen des Studiums einsetzbar ist.

17 Vgl. hierzu Zabold (2020).

Dem aktuellen Entwicklungsstand des Tests entsprechend können derzeit lediglich die ersten Facetten des Konstrukts geschichtsdidaktische Kompetenz erfasst werden. Weitere Aufgaben müssten entwickelt, validiert und standardisiert werden. Die Herausforderung wird sein, die Bearbeitungszeit des Tests bei ca. 30 Minuten zu halten, damit er im Kontext von Wirksamkeitsstudien (etwa von Fortbildungen) einsetzbar bleibt beziehungsweise bei Erhebungen von Lernfortschritten im Laufe des Lehramtsstudiums oder in der zweiten Phase.

Zu prüfen wird sein, ob es sinnvoll ist, neben einem gemeinsamen Anker an Aufgaben spezielle Aufgaben für die einzelnen Phasen, evt. auch für einzelne Schularten zu entwickeln.

Bei aller Offenheit, die dem Entwicklungsstand der Testentwicklung geschuldet ist, kann aber davon ausgegangen werden, dass der Gediko-T-Test einen Ansatzpunkt für weitere Forschungsarbeiten im quantitativen Bereich bietet, um das Konstrukt der geschichtsdidaktischen Kompetenz auf systematische Weise erfassen zu können.

Literatur

- Adamski, Peter (2014). *Historisches Lernen diagnostizieren. Lernvoraussetzungen – Lernprozesse – Lernleistungen*. Schwalbach/Ts.: Wochenschau.
- Barsch, Sebastian (2020). *Inklusiven Geschichtsunterricht planen*. Frankfurt a. M.: Wochenschau.
- Barsch, Sebastian & Barte, Burghard (Hrsg.) (2021). *Motivation – Kognition – Reflexion. Schlaglichter geschichtsdidaktischer Professionsforschung*. Frankfurt a. M.: Wochenschau.
- Baumert, Jürgen & Kunter, Mareike (2006). Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Baumert, Jürgen & Kunter, Mareike (2011). Das mathematikspezifische Wissen von Lehrkräften, kognitive Aktivierung im Unterricht und Lernfortschritte von Schülerinnen und Schülern. In Mareike Kunter/Jürgen Baumert/Werner Blum/Uta Klusmann/Stefan Krauss & Michael Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogrammes COACTIV* (S. 163–192). Münster: Waxmann.
- Blömeke, Sigrid/Kaiser, Gabriele & Lehmann, Rainer (2010). *TEDS-M 2008: Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.
- Borries, Bodo von (2007). «Geschichtsbewusstsein» und «Historische Kompetenz» von Studierenden der Lehramter «Geschichte». *Zeitschrift für Geschichtsdidaktik*, 6, 60–83.
- Brauer, Benjamin (2021). *Historische Orientierungsgelegenheiten: Eine Theorie zur Berücksichtigung von Diversität für die Anregung historischer Denkprozesse*. Manuskript der Dissertation. Katholische Universität Eichstätt-Ingolstadt, Eichstätt.

- Bühner, Markus (2011). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson.
- Demantowsky, Marko & Waldis, Monika (2014). John Hatties «Visible Learning» und die Geschichtsdidaktik. *Zeitschrift für Geschichtsdidaktik*, 13, 100–116.
- Erdmann, Elisabeth (2013). Typen der Geschichtslehrerbildung in Europa. In Susanne Popp/Michael Sauer/Bettina Alavi/Marko Demantowsky & Alfons Kenkmann (Hrsg.), *Zur Professionalisierung von Geschichtslehrerinnen und Geschichtslehrern. Nationale und internationale Perspektiven* (S. 371–386). Göttingen: V&R unipress.
- Fenn, Monika/Stebener, Isabelle & Urban, Stefanie (2019). Concept Maps zur Erfassung von professionellem Fachwissen bei (angehenden) Geschichtslehrkräften. In Waltraud Schreiber/Béatrice Ziegler & Christoph Kühberger (Hrsg.), *Geschichtsdidaktischer Zwischenhalt: Beiträge aus der Tagung «Kompetent machen für ein Leben in, mit und durch Geschichte» in Eichstätt vom November 2017* (S. 264–271). Münster, New York: Waxmann.
- Göllner, Richard/Wagner, Wolfgang/Klieme, Eckhard/Lüdtke, Oliver/Nagengast, Benjamin & Trautwein, Ulrich (2016). Erfassung der Unterrichtsqualität mithilfe von Schülerurteilen: Chancen, Grenzen und Forschungsperspektiven. In Bundesministerium für Bildung und Forschung. (Hrsg.), *Forschungsvorhaben in Anknüpfung an Large-Scale-Assessments* (S. 63–82). Berlin: Bundesministerium für Bildung und Forschung. https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/3/31222_Bildungsforschung_Band_44.pdf?__blob=publicationFile&%3Bv=3 (21.04.2022).
- Hasberg, Wolfgang (2010). Historiker oder Pädagoge? Geschichtslehrer im Kreuzfeuer der Kompetenzdebatte. *Zeitschrift für Geschichtsdidaktik*, 9, 159–179.
- Heuer, Christian/Körber, Andreas/Schreiber, Waltraud & Waldis, Monika (2019). GeDiKo – Professionstheoretische Überlegungen zur Modellierung geschichtsdidaktischer Kompetenzen. *Zeitschrift für Geschichtsdidaktik*, 18, 97–111.
- Hover, Stephanie van & Hicks, David (2018). History teacher preparation and professional development. In Scott Alan Metzger & Lauren McArthur Harris (Hrsg.), *The Wiley international handbook of history teaching and learning* (S. 391–418). New York: Wiley-Blackwell.
- Jung, Michael & Thünemann, Holger (2007). Welche Kompetenzen brauchen Geschichtslehrer? Für eine Debatte über fachspezifische Standards in der Geschichtslehrerausbildung. *Zeitschrift für Geschichtsdidaktik*, 6, 243–252.
- Kanert, Georg (2014). *Geschichtslehrerausbildung auf dem Prüfstand. Eine Längsschnittstudie zum Professionalisierungsprozess*. Göttingen: V&R unipress.
- Kanert, Georg & Resch, Mario (2014). Erfassung geschichtsdidaktischer Wissensstrukturen von Geschichtslehrkräften anhand eines vignettengestützten Testverfahrens. *Zeitschrift für Geschichtsdidaktik*, 13, 15–31.
- Kingston, Neal M./Scheuring, Sylvia T. & Kramer, Laura B. (2013). Test development strategies. In Kurt F. Geisinger/Bruce A. Bracken/Janet F. Carlson/Jo-Ida C. Hansen & Nathan R. Kuncel (Hrsg.), *APA handbook of testing and assessment in psychology* (Bd. 1, S. 165–184). Washington D. C.: American Psychological Association.
- Klieme, Eckhard/Schümer, Gundel & Knoll, Steffen (2001). Mathematikunterricht in der Sekundarstufe I. «Aufgabenkultur» und Unterrichtsgestaltung. In Eckhard Klieme & Jürgen Baumert (Hrsg.), *TIMSS Impulse für Schule und Unterricht: Forschungsbefunde*,

- Reforminitiativen, Praxisberichte und Video-Dokumente* (S. 43–57). Berlin: Bundesministerium für Bildung und Forschung.
- Klusmann, Uta/Kunter, Mareike/Richter, Dirk & Voss, Tamar (o. J.). *COACTIV-R Highlights: Ausgewählte Ergebnisse aus dem Forschungsprogramm für Studienteilnehmerinnen und -teilnehmer*. Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik. <https://ezw.uni-freiburg.de/wp-content/uploads/2020/07/coactiv-Broschüre.pdf> (21.04.2022).
- Körper, Andreas/Heuer, Christian & Schreiber, Waltraud. (i. E.). *Geschichtsdidaktische Kompetenzen – ein Strukturmodell*. Münster: Waxmann.
- Körper, Andreas/Schreiber, Waltraud & Schöner, Alexander (Hrsg.) (2007). *Kompetenzen historischen Denkens. Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik*. Neuried: ars una.
- Köster, Manuel/Bernhardt, Markus & Thünemann, Holger (2016). Aufgaben im Geschichtsunterricht. Typen, Gütekriterien und Konstruktionsprinzipien. *Geschichte lernen*, (174), 2–11.
- Kühberger, Christoph (2016). Fachdidaktische Diagnose als notwendige Voraussetzung im Umgang mit Heterogenität. Subjektorientierte Zugänge für das historische Lernen. In Christoph Kühberger/Silvia Kronberger & Manfred Oberlechner (Hrsg.), *Diversitätskategorien in der Lehramtsausbildung: Ein Handbuch* (S. 299–314). Innsbruck, Wien, Bozen: Studienverlag.
- Kunter, Mareike/Baumert, Jürgen/Blum, Werner/Klusmann, Uta/Krauss, Stefan & Neubrand, Michael (Hrsg.) (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Mägdefrau, Jutta & Michler, Andreas (2018). Arbeitsaufträge im Geschichtsunterricht – Diskrepanz zwischen Lehrerintention und didaktischem Potenzial? In Katharina Grannemann/Sven Oleschko & Christian Kuchler (Hrsg.), *Sprachbildung im Geschichtsunterricht: Zur Bedeutung der kognitiven Funktion von Sprache* (S. 137–155). Münster, New York: Waxmann.
- Meyer-Hamme, Johannes (2018). Was heißt «historisches Lernen»? Eine Begriffsbestimmung im Spannungsfeld gesellschaftlicher Anforderungen, subjektiver Bedeutungszuschreibungen und Kompetenzen historischen Denkens. In Thomas Sandkühler/Charlotte Bühl-Gramer/Anke John/Astrid Schwabe & Markus Bernhardt (Hrsg.), *Geschichtsunterricht im 21. Jahrhundert. Eine geschichtsdidaktische Standortbestimmung* (S. 75–92). Göttingen: V&R unipress.
- Pandel, Hans-Jürgen (2007). *Geschichtsunterricht nach PISA: Kompetenzen, Bildungsstandards und Kerncurricula*. Schwalbach/Ts.: Wochenschau.
- Resch, Mario (2018). *Aufgaben formulieren können. Entwicklung und Validierung eines Vignetentests zur Erfassung professioneller Kompetenz für historisches Lehren*. Frankfurt a. M.: Wochenschau.
- Resch, Mario/Heuer, Christian & Seidenfuß, Manfred (2019). Theoretische Modellierung und empirische Erfassung von geschichtsdidaktischem Wissen und Können bei angehenden Geschichtslehrpersonen. In Waltraud Schreiber/Béatrice Ziegler & Christoph Kühberger (Hrsg.), *Geschichtsdidaktischer Zwischenhalt: Beiträge aus der Tagung «Kom-*

- petent machen für ein Leben in, mit und durch Geschichte» in Eichstätt vom November 2017* (S. 242–257). Münster, New York: Waxmann.
- Rohrmann, Bernd (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie*, 9, 222–245.
- Rutsch, Juliane/Vogel, Markus/Seidenfuß, Manfred/Dörfler, Tobias & Rehm, Markus (2017). Professionalisierungsprozesse angehender Lehrkräfte untersuchen. In Juliane Rutsch/Markus Rehm/Markus Vogel/Manfred Seidenfuß & Tobias Dörfler (Hrsg.), *Effektive Kompetenzdiagnose in der Lehrerbildung: Professionalisierungsprozesse angehender Lehrkräfte untersuchen* (S. 9–22). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-20121-0_1.
- Rüsen, Jörn (2008). *Historische Orientierung: Über die Arbeit des Geschichtsbewusstseins, sich in der Zeit zurechtzufinden*. Schwalbach/Ts.: Wochenschau.
- Sauer, Michael (2013). Zur Professionalisierung von Geschichtslehrerinnen und Geschichtslehrern: Einführung in das Tagungsthema. In Susanne Popp/Michael Sauer/Bettina Alavi/Marko Demantowsky & Alfons Kenkmann (Hrsg.), *Zur Professionalisierung von Geschichtslehrerinnen und Geschichtslehrern: Nationale und internationale Perspektiven* (S. 19–38). Göttingen: V&R unipress.
- Schmidt-Atzert, Lothar & Amelang, Manfred (2012). *Psychologische Diagnostik*. Berlin: Springer.
- Seidel, Tina & Thiel, Felicitas (2017). Standards und Trends der videobasierten Lehr-Lernforschung. *Zeitschrift für Erziehungswissenschaft*, 20(supplement 1), 1–21. <https://doi.org/10.1007/s11618-017-0726-6>.
- Seidenfuß, Manfred (2014). Einführung in den Themenschwerpunkt. *Zeitschrift für Geschichtsdidaktik*, 13, 5–14.
- Shulman, Lee S. (1986). Those who understand: Knowledge growth in teaching. In *Educational Researcher*, 15(2), 4–31. <https://doi.org/10.3102/0013189X015002004>.
- Shulman, Lee S. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review*, 57(1), 1–2. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>.
- Trautwein, Ulrich et al. (2017). *Kompetenzen historischen Denkens erfassen. Konzeption, Operationalisierung und Befunde des Projekts «Historical Thinking Competencies in History (HiTCH)»*. Münster: Waxmann.
- Waldis, Monika (2019). Professionelle Unterrichtswahrnehmung als Annäherung an fachdidaktisches Wissen von Geschichtslehrpersonen. In Waltraud Schreiber/Béatrice Ziegler & Christoph Kühberger (Hrsg.), *Geschichtsdidaktischer Zwischenhalt: Beiträge aus der Tagung «Kompetent machen für ein Leben in, mit und durch Geschichte» in Eichstätt vom November 2017* (S. 223–241). Münster, New York: Waxmann.
- Waldis, Monika (2020). Geschichte in der Lehrerinnen- und Lehrerbildung. Professionalisierung im Umfeld von (Hoch-)Schule und Geschichtskultur. In Colin Cramer/Johannes König/Martin Rothland & Sigrid Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 436–443). Bad Heilbrunn: Klinkhardt.
- Waldis, Monika/Nitsche, Martin/Marti, Philipp/Hodel, Jan & Wyss, Corinne (2014). «Der Unterricht wird fachlich korrekt geleitet» – theoretische Grundlage, Entwicklung der Instrumente und empirische Erkundungen zur videobasierten Unterrichtsreflexion angehender Geschichtslehrpersonen. *Zeitschrift für Geschichtsdidaktik*, 13, 32–49.

- Waldis, Monika & Ziegler, Béatrice (Hrsg.) (2017). *Forschungswerkstatt Geschichtsdidaktik 15. Beiträge zur Tagung «geschichtsdidaktik empirisch 15»*. Bern: hep.
- Waldis, Monika & Ziegler, Béatrice (Hrsg.) (2019). *Forschungswerkstatt Geschichtsdidaktik 17. Beiträge zur Tagung «geschichtsdidaktik empirisch 17»*. Bern: hep.
- Wendler, Cathy & Burrus Jeremy (2013). The importance of editorial reviews in ensuring item quality. In Kurt F. Geisinger/Bruce A. Bracken/Janet F. Carlson/JoIda C. Hansen & Nathan R. Kuncel (Hrsg.), *APA handbook of testing and assessment in psychology* (Bd. 1, S. 283–291). Washington D. C.: American Psychological Association.
- Werner, Michael & Schreiber, Waltraud (2015). Testfragen befragen – Pretesting und Optimierung des Large-Scale-Kompetenztests «HiTCH» durch Cognitive Labs. In Monika Waldis & Béatrice Ziegler (Hrsg.), *Forschungswerkstatt Geschichtsdidaktik 13. Beiträge zur Tagung «geschichtsdidaktik empirisch 13»* (S. 153–164). Bern: hep.
- Wolf, Jörgen/Seiffert, Johanna/Seiffert, Andreas/Rothland, Martin & Brauch, Nicola (2018). Das geschichtsdidaktische Planungswissen von angehenden Geschichtslehrer/innen. Entwicklung und Pilotierung eines Testinstruments zur Messung des fachdidaktischen Planungswissens von Lehramtsstudierenden im Fach Geschichte. *Geschichte in Wissenschaft und Unterricht*, 69(7/8), 373–393.
- Zabold, Stefanie (2020). Diagnose geschichtsdidaktischer Lehrerkompetenzen in unterschiedlichen Phasen der Lehrerausbildung. In Petra Hiebl & Rainer Wenrich (Hrsg.), *Lehrer/-innenbildung und Bildungsforschung an der KU* (S. 30–31). Eichstätt: Universität Eichstätt-Ingolstadt. https://www.ku.de/fileadmin/1920/Publikation_L_B.pdf#page=30 (04.07.2022).
- Zabold, Stefanie (2021). Lernstandsdiagnose – Schlüssel zu historischem Denken und Grundlage erfolgreicher Lernprozesse. In Wolfgang Buchberger & Christoph Kühberger (Hrsg.), *Historisches Lernen in der Primarstufe: Standpunkte – Herausforderungen – Perspektiven* (S. 227–248). Innsbruck: Studienverlag.

Das «Aura-Erlebnis» in der Begegnung mit Zeitzeug*innen

Eine komplexe Lernerfahrung verstehen und standardisiert erfassen

Lisa Zachrich, Wolfgang Wagner, Christiane Bertram und Ulrich Trautwein

1 Einleitung

«Man spürt es vielmehr [...], man kann sich besser reinversetzen, wenn man den [Zeitzeugen] so sieht, und dann denkt, der war da wirklich, er hat es mitbekommen, er weiß das. Es ist ein anderes Erlebnis, als wenn man 'nen Text liest.»¹ So beschreibt eine Schülerin der 9. Klasse ihre Begegnung mit einem Zeitzeugen aus der ehemaligen DDR im Geschichtsunterricht. Tatsächlich stellt die Begegnung mit Zeitzeug*innen, so die Beobachtung von Obens und Geißler (2010), für Schüler*innen oft ein sehr bedeutsames Erlebnis dar, das sie kaum in Worte fassen können. Aber auch für die Geschichtsdidaktik ist gar nicht so einfach zu fassen, was die Begegnung mit Zeitzeug*innen so besonders macht: «In der Fachdebatte bedienen sich Experten deshalb häufig «Hilfskonstruktionen» wie «Aura» oder «Authentizität», ohne diese im Detail zu beschreiben» (Obens & Geißler, 2010, 59). Ein zentrales Charakteristikum besteht möglicherweise darin, dass Zeitzeug*innen das Potenzial haben, «nicht nur Köpfe, sondern auch Herzen der Jugendlichen zu erreichen» (Uhl, 2012, 59). Im Sinne Pestalozzis hieße das, dass die Begegnung mit Zeitzeug*innen neben Gedanken auch die Gefühle der Jugendlichen ansprechen kann (Klika, 2018, 79). Dass das Lernen über die Geschichte «etwas mit Emotionen zu tun» (Pandel, 2012, 148) hat, wird zunehmend in geschichtsdidaktischen Diskussionen thematisiert (z. B. Brauer & Lücke, 2013). Bislang ist jedoch umstritten, ob Emotionen eher «Katalysator» oder

1 Das Zitat stammt aus einem Interview mit einer Schülerin nach der Unterrichtsstunde mit einem Zeitzeugen im Rahmen einer Interventionsstudie zur Wirksamkeit von Zeitzeugenbefragungen (Bertram, Wolfgang & Trautwein, 2017).

«Sackgasse» für das Lernen der Schüler*innen im Geschichtsunterricht sind (Brauer, 2019, 278).

Die Lernerfahrung mit Zeitzeugenberichten stellt eine Unterrichtssituation dar, die diesen Zwiespalt – Katalysator oder Sackgasse – in besonderer Weise verdeutlicht. Das Ziel des Historisch-denken-Lernens beinhaltet unter anderem, Quellen kritisch hinterfragen zu können, um ein (selbst-)reflexives Geschichtsbewusstsein zu entwickeln (z. B. Schreiber, 2002, 27). Zeitzeugenberichte haben die Chance, Schüler*innen für Geschichte zu motivieren, indem sie die Vergangenheit «lebendig und authentisch erfahrbar machen» (Henke-Bockschatz, 2007, 354). Mit Blick auf das Ziel des Geschichtsunterrichts bergen sie allerdings auch das Risiko, zu überwältigen. Denn die Arbeit mit Zeitzeug*innen stelle nach Wierling (1997) «hohe Ansprüche an das Einfühlungsvermögen und die Quellen­distanz» (239). Diese Ambivalenz der Zeitzeugenbegegnungen konnten Bertram, Wagner und Trautwein (2017, 472–473) empirisch bestätigen. Schüler*innen berichteten nach einer Live-Begegnung mit einem*r Zeitzeug*in, dass sie mehr Spaß im Unterricht hatten, zeigten aber weniger Einsicht in den Konstruktcharakter von Geschichte als Schüler*innen, die mit videografierten oder schriftlichen Zeugnissen lernten .

Was macht also die Live-Begegnung mit einem*r Zeitzeug*in besonders? Bertram und Kollegen boten zur möglichen Erklärung ihrer Befunde das von Sabrow (2012) geprägte Konzept der «Aura der Authentizität» (27; Bertram et al., 2017, 472) an. Zachrich, Weller, Baron und Bertram (2020, 249) entwickelten zur Annäherung an diese Besonderheit ein konzeptuelles Modell, das die Merkmale der Lernerfahrung mit komplexen historischen Quellen (u. a. Zeitzeugenberichten) beschreibt und im Kern ein Zusammenspiel von kognitiver, emotionaler und physischer Involviertheit identifiziert. Die empirische Frage, was Kognition, Emotion und Physis in diesem Kontext ausmacht, ist bisher jedoch unbeantwortet. Dieser Frage stellen wir uns im vorliegenden Beitrag: Wir haben einen standardisierten Fragebogen zur Besonderheit der Begegnung mit einem*r Zeitzeug*in entwickelt und berichten hier Befunde zu dessen psychometrischer Qualität sowie zu Zusammenhängen mit Situations- und Personenmerkmalen.

2 Theoretischer Hintergrund

2.1 Zeitzeugenbegegnungen im Kontext historischen Lernens

Das Verständnis von Zeitzeug*innen unterscheidet sich in Geschichtswissenschaft, -didaktik und Erinnerungskultur (Bertram, 2020, 259–260). Dabei hilft es, die Zeitzeugenschaft in ihrer Komplexität wahrzunehmen und zu verstehen. Je nach Zielsetzung der Bildungsarbeit werden unterschiedliche Zeitzeugnisse genutzt: In der Aufarbeitung von und Erinnerung an Gewalt- und Unrechtsgeschichte sollen die Zeugnisse von Held*innen oder Opfern für die Geschichte sensibilisieren (von Plato, 2001, 136). In der sogenannten Oral History hingegen, die sich als Methode der Geschichtswissenschaft seit den 1950er-/1960er-Jahren etabliert hat, geht es vor allem darum, jenen Bevölkerungsgruppen eine Stimme zu geben, die in der traditionellen Geschichtsschreibung nicht gehört wurden und werden. Es sind also Alltagsgeschichten, die Einblicke in die vergangene Zeit geben sollen (z. B. Lanman & Wendling, 2006). Zeitzeug*innen unterscheiden sich folglich hinsichtlich verschiedener Dimensionen, zum Beispiel je nach Rolle in der Vergangenheit (z. B. Täter, Opfer ...), nach ihrer Expertise (Laie vs. Experte) und in dem zeitlichen Abstand, den sie zu ihren Erfahrungen in der Vergangenheit haben. Entsprechend unterscheiden sich die Berichte und auch die Begegnungen stark. Was den Berichten aber gemein ist: Sie ermöglichen Einblicke in die individuelle Wahrnehmung vergangener Erfahrungen einer Person. Diese natürlich subjektiven, aber in ihrer Erlebnisqualität «echten» Geschichten haben das Potenzial, Schüler*innen anders anzusprechen als Schulbuchtexte und das historische Denken (u. a. FUER-Kompetenzstrukturmodell, Schreiber et al., 2006) zu fördern (Henke-Bockschatz, 2007, 363).

Empirisch gibt es bisher allerdings nur wenige Erkenntnisse darüber, was und wie Schüler*innen mit und von den Begegnungen lernen. Dass Jugendliche den Zeitzeugnissen mehr Vertrauen schenken als den Historiker*innen, war ein Ergebnis der Studie *Youth & History* (Angvik & von Borries, 1997). Obens und Geißler (2010, 59) zogen aus ihren qualitativen Interviews mit Schüler*innen nach einer Begegnung mit Holocaust-Überlebenden das Fazit, dass viele der Jugendlichen nicht viel mehr aus den Gesprächen mitnahmen als eine emotionale Ergriffenheit. Auch Bertram et al. (2017) bestätigten mit ihrer Studie die Risiken und Potenziale der Zeitzeugenbegegnungen: In einer groß angelegten Interventionsstudie mit 962 Schüler*innen aus 35 Klassen an weiterführenden Schulen in Baden-Württemberg untersuchten die Autor*innen die Effekte einer Live-Be-

gegnung mit einem*r Zeitzeug*in der ehemaligen DDR im Vergleich zu Video- und Textversionen der Zeitzeugenberichte auf historische Sachkompetenzen und die Motivation der Schüler*innen. Lernende, die dem*der Zeitzeug*in live begegnet waren, berichteten zwar mehr Spaß an der Unterrichtseinheit, zeigten aber eine weniger ausgeprägte Einsicht in die Perspektivität von Zeitzeugenaussagen als Lernende, die mit Text- beziehungsweise Videoberichten arbeiteten (Bertram et al., 2017, 471–473). Brüning (2018, 372) schlussfolgerte in ihrer Studie, dass die Zeugnisse von Holocaust-Überlebenden, selbst wenn die Schüler*innen nur mit den Videos arbeiteten, sich für eine methodische Quellenkritik nicht eigneten. Es brauche eine angemessene Methodik im Umgang mit diesen besonderen Zeitzeugnissen.

2.2 Prozesse der Lernerfahrung

Eine notwendige Voraussetzung für die Entwicklung einer angemessenen Methodik ist, dass wir die besondere Lernerfahrung zunächst genauer verstehen. Diesem Forschungsdesiderat haben sich Zachrich et al. (2020) gewidmet und aus mehreren theoretischen Ansätzen ein Modell entwickelt, das beschreibt, was die Erfahrung mit komplexen historischen Quellen² ausmacht – was also hinter der «Aura der Authentizität» (Bertram, 2017, 473; Sabrow, 2012, 27) steckt (Zachrich et al., 2020, 244). Komplexe historische Quellen zeichnet aus, dass sie sich aus mehreren Medien und/oder Interaktionsebenen zusammensetzen und die Rezipient*innen auf kognitive, emotionale und physische Weise ansprechen (Zachrich et al., 2020). Zur Beschreibung der besonderen Erfahrung bedienten sich die Autor*innen mehrerer Theorien aus verschiedenen wissenschaftlichen Disziplinen: Die *Transportation Theory* (Green & Brock, 2000) und das *Model of Narrative Comprehension and Engagement* (Busselle & Bilandzic, 2009) kommen aus der Medienpsychologie und beschreiben das kognitive und emotionale Eintauchen in einen Medieninhalt, bei dem alle mentalen Kapazitäten auf die Inhalte des Narrativs gerichtet sind; die ästhetische Begegnung mit Kunstwerken beschreibt aus anthropologischer Perspektive ein Rahmenmodell zu den Motiven und Prozessen der Rezeption von Kunst (Csikszentmihalyi & Robinson, 1990);

2 Zeitzeug*innenberichte sind Quelle und Darstellung zugleich. Als Artefakt der Vergangenheit lässt sich der Bericht als Quelle dekonstruieren. Die Erzählung beinhaltet aber immer auch die eigene Interpretation der Erfahrung, dynamische Erinnerungen, und ist folglich auch Darstellung (z. B. Bertram, 2020).

Cameron und Gatewood (2003) sowie Latham (2013) identifizierten, ebenfalls aus anthropologischer Perspektive, mit dem Konzept *Numen-seeking* ein Motiv für den Besuch von Gedenkstätten, das eine nahezu spirituelle Begegnung beinhaltet; Historische Empathie ist ein Konzept aus der US-amerikanischen Geschichtsunterrichtsforschung und beschreibt die kognitiven und affektiven Komponenten von Empathie mit einem historischen Akteur; Presence, ein Konzept aus Psychologie und Informatik meint ein Gefühl, in einer *medienvermittelten* Welt zu sein, ohne das Medium wahrzunehmen (Lombard & Ditton, 1997). In Anlehnung an die Ausführungen dieser Konzepte entwickelten die Autor*innen das Modell und generierten die einzelnen Facetten. Dabei verknüpften sie die theoretischen Darstellungen mit empirischen Ergebnissen aus ihren Forschungsarbeiten.

Abbildung 1 zeigt das konzeptuelle Modell der Begegnung mit komplexen historischen Quellen. Im Kern der Begegnung stehen die kognitive, emotionale und physische Involviertheit der Lernenden, die durch die Begegnung hervorgerufen werden (bzw. werden können; Zachrich et al., 2020, 252–253). Das Modell bezieht sich beispielhaft auf Zeitzeugenbegegnungen und historische Orte. Die Lernerfahrungen können sich natürlich je nach Zeugnis und Quelle unterscheiden. Insbesondere in der physischen Involviertheit sind Unterschiede zwischen den Begegnungen denkbar. So sitzen die Schüler*innen im Gespräch mit Zeitzeug*innen in der Regel an einem festen Platz, während sie sich am historischen Ort bewegen. Das Modell zeigt jedoch die möglichen Verarbeitungsprozesse für diese besonderen Zeugnisse auf.

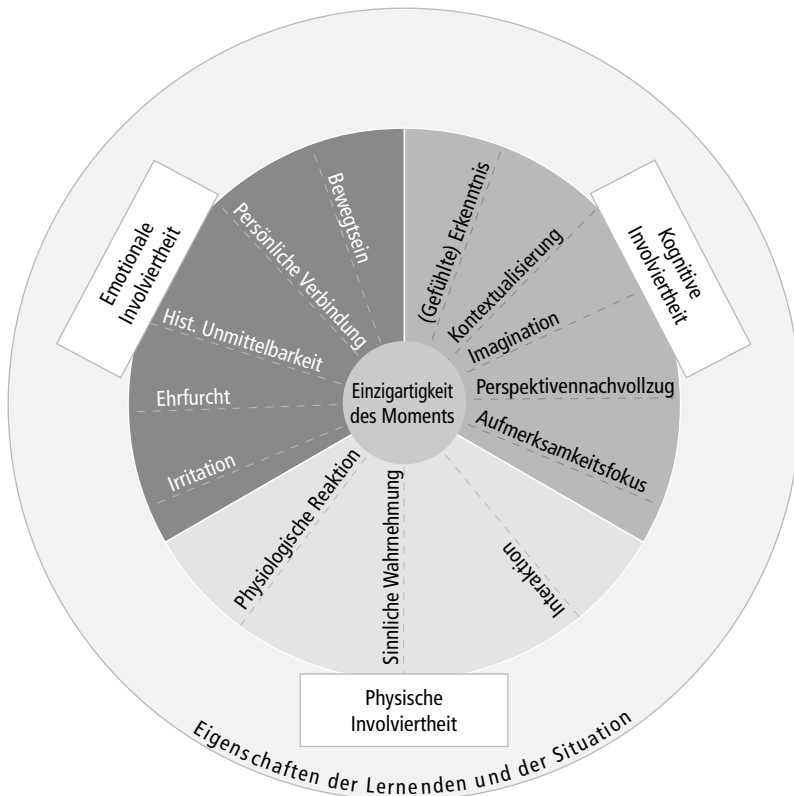


Abbildung 1: Modell zur Beschreibung der Begegnung mit komplexen historischen Quellen

Das Modell identifiziert fünf Elemente der kognitiven Involviertheit: (1) Aufmerksamkeitsfokus: die Lernenden richten all ihre Aufmerksamkeit auf die Begegnung (z. B. Csikszentmihalyi & Robinson, 1990, 28); (2) Imagination: die Vergangenheit ist für die Lernenden bildlich vorstellbar (z. B. Green & Brock, 2000, 703); (3) Perspektivennachvollzug: die Lernenden verstehen die Perspektive des historischen Akteurs (z. B. Endacott & Brooks, 2013, 43); (4) Kontextualisierung: die Lernenden verknüpfen die Informationen mit ihrem Vorwissen (z. B. Huijgen, van Boxtel, van de Grift & Holthuis, 2017, 112); (5) (gefühlte) Erkenntnis: die Lernenden glauben während der Begegnung, etwas über die Vergangenheit im Allgemeinen zu lernen (z. B. Busselle & Bilandzic, 2009, 330). Fünf Elemente machen laut des Modells die emotionale Involviertheit aus: (1) Bewegtes sein: die Ler-

nenden sind emotional ergriffen (z. B. Menninghaus et al., 2015, 12); (2) persönliche Verbindung: die Lernenden identifizieren sich mit dem Akteur (z. B. Brauer, 2016, 32); (3) historische Unmittelbarkeit: die Vergangenheit fühlt sich durch die Begegnung nah an (z. B. Lombard & Ditton, 1997, 7); (4) Ehrfurcht: die Lernenden empfinden eine tiefe Anerkennung (z. B. Cameron & Gatewood, 2003, 67); (5) Irritation: die Lernenden sind irritiert von den neuen Informationen (z. B. Brauer, 2019, 277). Drei Elemente setzen den Rahmen der physischen Involviertheit: (1) Sinnliche Wahrnehmung: Der*die Zeitzeug*in ist «zum Greifen» nah (z. B. Drozdzewski, De Nardi & Waterton, 2016, 447); (2) Physische Interaktion: die Lernenden können Antworten auf ihre eigenen Fragen erhalten; (3) Physiologische Reaktion: die Lernenden zeigen körperliche Reaktionen, haben Tränen in den Augen oder Gänsehaut (z. B. Latham, 2013, 15).

Zuletzt beschreibt die «Einzigartigkeit des Moments» ein Zusammenspiel kognitiver, emotionaler und physischer Involviertheit (Zachrich et al., 2020). Diese Elemente stellen mögliche Prozesse der Lernerfahrung dar, die nicht in jeder Begegnung und für jede Person von Bedeutung sind. Das «Aura-Erlebnis» ist demnach durch hohe Ausprägungen auf mehreren dieser Subskalen gekennzeichnet. Wir legen dabei zum jetzigen Zeitpunkt keine bestimmte Zusammensetzung des «Aura-Erlebens» über die Subskalen fest, weshalb wir im Folgenden in der Regel auf Basis der einzelnen Involviertheitsskalen argumentieren, nicht jedoch auf der Ebene des «Aura-Erlebens».

2.3 Bedingungen der besonderen Lernerfahrung

Welche Bedingungen erhöhen die Wahrscheinlichkeit, dass kognitive, emotionale und physische Involviertheit auftritt, und welche Konsequenzen hat sie? Wir haben hierzu das Modell von Zachrich et al. (2020) in ein Wirkungsmodell umstrukturiert und um Bedingungen und Konsequenzen erweitert (siehe Abbildung 2). Dabei unterscheiden wir bei den Bedingungen zwischen Merkmalen der Person und der Situation, denn theoretische wie empirische Arbeiten legen nahe, dass sowohl die Eigenschaften der Lernenden als auch die Merkmale der Situation einen Einfluss darauf haben, wie die Begegnung erlebt wird: In einer Studie von Green, Chatham und Sestir (2012, 48) berichteten Frauen, stärker in eine Geschichte einzutauchen als Männer. Gnams et al. (2014, 191) konnten zeigen, dass die Erfahrung von Transportation sowohl auf Personenmerkmale als auch auf situationale Merkmale zurückzuführen ist. Auch weil man davon ausgeht, dass die Unterrichtssituation einen Einfluss auf die Involviertheit in der Begeg-

nung hat, empfehlen Lehrbücher zur Geschichtsdidaktik, den Zeitzeugenbesuch intensiv mit den Schüler*innen vor- und nachzubereiten (z. B. Henke-Bockschatz, 2007, 354).

Die große Bedeutsamkeit der Begegnung schreiben Lehrkräfte und Schüler*innen unter anderem einer «Echtheit» und «Authentizität» des Zeitzeugenberichts zu. Empirische Untersuchungen in der Medienpsychologie bestätigten die Bedeutung der wahrgenommenen Echtheit (*perceived realism*) in Zusammenhang mit Erfahrungen wie Transportation (z. B. Green, 2004, 257), Immersion und Identifikation (z. B. Bilandzic & Busselle, 2011, 45) sowie emotionale Reaktionen der Rezipient*innen (z. B. Pouliot & Cowen, 2007, 250). Einen Unterschied der Erfahrung zwischen der Bezeichnung des Narratives als Fakt beziehungsweise Fiktion zeigte sich allerdings in einer Reihe von Experimenten nicht konsistent (Green & Brock, 2000, 718). Mierwald, Lehmann und Brauch (2018, 292) konnten in einer Interventionsstudie wiederum Hinweise finden, dass authentisches Lernmaterial zum Thema «Einfluss der NS-Herrschaft auf das Zusammenleben von Juden und Nichtjuden» einen positiven Einfluss auf epistemische Überzeugungen im Fach Geschichte hat. Allerdings wurden die vertonten Originalquellen von den Schüler*innen als authentischer wahrgenommen als die ursprünglich schriftlich vorliegenden Quellen. Die genannten Studien implizieren, dass eine differenzierte Betrachtung von Authentizität notwendig ist. Ob die wahrgenommene Authentizität des Zeitzeugenberichts eine Rolle für die Lernerfahrung spielt, bleibt mit Blick auf die inkonsistente Befundlage eine empirische Frage. Wir haben deshalb im Wirkungsmodell die wahrgenommene Authentizität als Mediator zwischen Bedingung und Lernerfahrung berücksichtigt.

In Anlehnung an Vorderer (1992, 84) nehmen wir zudem an, dass es neben der starken Involviertheit beim Zeitzeugenbericht auch einen kritisch-distanzierten Verarbeitungsmodus gibt. Diesbezüglich geben bisherige Forschungsarbeiten Hinweise darauf, dass Transportation positiv mit Persuasion zusammenhängt (Green & Brock, 2000, 719) und Rezipient*innen durch die Erfahrung des Eintauchens Inkonsistenzen in der Geschichte leichter akzeptieren, folglich weniger kritisch sind (Weber & Wirth, 2014, 136).

Die Studie von Bertram et al. (2017, 469) gab empirische Hinweise, dass die Begegnung mit Zeitzeug*innen einen Einfluss auf beispielsweise die Motivation oder das Lernen der Schüler*innen haben kann. Deshalb ergänzten wir das Modell schließlich um Konsequenzen der Lernerfahrung, die wir aber hier nicht untersuchten.

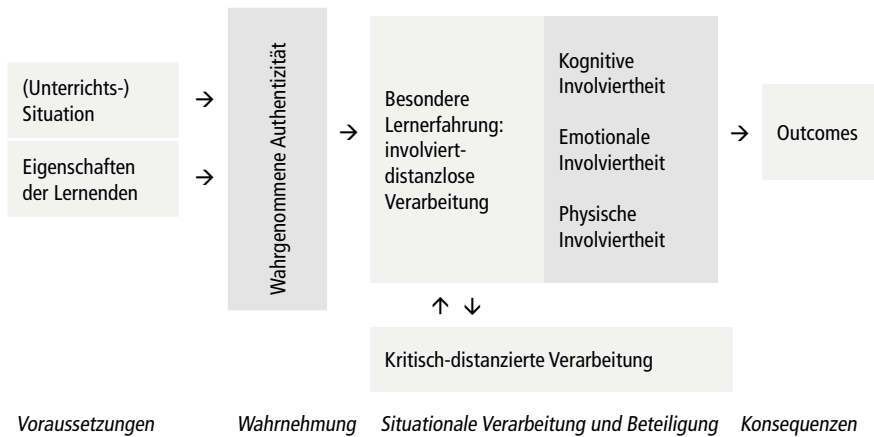


Abbildung 2: Wirkmodell zur besonderen Lernerfahrung mit Zeitzeug*innen

2.4 Fragestellung und Hypothesen

Obens und Geißler (2010) stellten fest, dass es an Untersuchungen in «nahezu allen Aspekten dieses Bildungsarrangements» (4) fehlt. Mit der vorliegenden Studie begegnen wir genau diesem Forschungsdesiderat und entwickelten als Grundlage für weitere empirische Studien ein standardisiertes Messinstrument zur Erfassung dieser besonderen Lernerfahrung. Dabei interessierte uns (i) was sich empirisch hinter der «Black Box» der Zeitzeugenbegegnung verbirgt und (ii) welche Prädiktoren auf Situations- und Personenebene die starke Involviertheit beim Zeitzeugenbericht vorhersagen können. Konkret orientierten wir uns dafür an den folgenden Fragen: (1) Lassen sich die verschiedenen theoretisch angenommenen Dimensionen der Erfahrung mit Zeitzeugenberichten empirisch trennen? (2) In welchem Zusammenhang stehen die erfassten Konstrukte mit der kritisch-distanzierten Verarbeitung? Wir erwarteten, dass eine starke Involviertheit negativ mit der kritisch-distanzierten Verarbeitung zusammenhängt. (3) Welche individuellen Faktoren der Schüler*innen können eine hohe Involviertheit beim Zeitzeugenbericht vorhersagen? Wir nahmen an, dass Mädchen eine höhere Involviertheit berichten als Jungen; zudem sollten Schüler*innen, die eine Frage gestellt haben, eine höhere Involviertheit berichten als Schüler*innen, die nur zuhörten. (4) In der vierten Fragestellung interessierte uns, welche situationalen Merkmale hohe Involviertheit vorhersagen können.

Mit Blick auf die Intensität der Vorbereitung hatten wir keine konkrete Erwartung, aber wir erwarteten, dass der zeitliche Abstand zwischen Begegnung und Studienteilnahme negativ mit der Involviertheit assoziiert ist. (5) Zuletzt interessierte uns, ob die Zusammenhänge zwischen den Bedingungen und der Begegnung durch die wahrgenommene Authentizität vermittelt werden.

3 Methode

3.1 Fragebogenentwicklung

Der Prozess der Fragebogenentwicklung vollzog sich in mehreren Schritten. Aufbauend auf den theoretischen Überlegungen entwickelten wir in Fokusgruppen mit Expert*innen aus der Bildungspraxis und -forschung, Geschichtsdidaktik und -wissenschaft standardisierte Items und passten Items aus validierten Fragebögen auf unseren Kontext an (z. B. Busselle & Bilandzic, 2009, 329–330). Der erste Entwurf umfasste 92 Items zu 17 Konstrukten der emotionalen, kognitiven und physischen Involviertheit, die auf einer sechsstufigen Rating-Skala beantwortet werden.

Die Pilotierung des Fragebogens fand mit $N = 296$ Studierenden aus sieben Veranstaltungen der Universität Tübingen statt. Als Impuls sahen die Studierenden einen fünfminütigen Videoclip von Kurt Gutmann, der im Zweiten Weltkrieg ein Kind war und von seinen Erfahrungen während des Krieges und in der Nachkriegszeit berichtet.³ Wir werteten die Daten mittels Reliabilitätsanalysen, konfirmatorischen und exploratorischen Faktorenanalysen aus. Wir kürzten den Fragebogen auf 60 Items für 14 Konstrukte. Der überarbeitete Fragebogen wurde in einem nächsten Schritt mit fünf Schüler*innen in *Cognitive Labs* (Hilton, 2017) überprüft. Aufgrund von Missverständnissen bei wenigen Items passten wir Formulierungen an. Die Validierung des Fragebogens mit Schüler*innen nach der Begegnung mit Live-Zeitzeug*innen steht als letzter Schritt der Fragebogenentwicklung im Fokus der vorliegenden Studie und wird deshalb im Folgenden im Detail beschrieben.

3 Das Zeitzeugenvideo wurde für eine Videoinstallation der Medienkünstler*innen Ina Rommel und Stefan Kraus gedreht und für diese Studie zu einem kurzen Clip zusammengeschnitten: <https://vimeo.com/233487804>.

3.2 Validierungsstudie

Das Kultusministerium des Landes Baden-Württemberg hat die Studie und die Erhebung der Daten genehmigt (Datum der Genehmigung: 16. April 2019, Aktenzeichen: 31-6499.20/1274). Die Ethikkommission für psychologische Forschung der Universität Tübingen bestätigte, dass die Studie mit den ethischen Standards der Forschung am Menschen übereinstimmt (Datum der Genehmigung: 21. März 2019, Aktenzeichen: Trautwein_2019_0215_149).

3.2.1 Forschungsdesign

Unsere Studie wurde von zwei Institutionen unterstützt, die uns den Zugang zu Schüler*innen nach einer Begegnung mit einem*r Live-Zeitzeug*innen ermöglichten. Deshalb enthält unsere Datengrundlage zwei unterschiedliche Settings von Zeitzeugenbegegnungen: (1) Das Koordinierende Zeitzeugenbüro ist ein Kooperationsprojekt der Bundesstiftung Aufarbeitung, der Stiftung Berliner Mauer, der Stiftung Gedenkstätte Berlin-Hohenschönhausen und der Beauftragten der Bundesregierung für Kultur und Medien. Das Büro vermittelt und finanziert DDR-Zeitzeug*innen bundesweit an Schulen. Den Zeitzeug*innen ist gemein, dass sie in der ehemaligen DDR aus verschiedenen Gründen inhaftiert waren. (2) Die Gedenkstätte Lindenstraße in Potsdam ist (unter anderem) ein ehemaliges Untersuchungsgefängnis des Ministeriums für Staatssicherheit. Die Gedenkstätte bietet individuelle Bildungsangebote für Schulklassen an, empfiehlt vor einem Gespräch mit einem*r Zeitzeug*in (ehemaliger Häftling) eine Führung durch das Gebäude. Zudem nutzten wir die Partnerschulen der Universitäten Konstanz und Tübingen, informierten über unser Projekt und riefen zur Teilnahme auf. Aufgrund terminlicher Schwierigkeiten im Schulalltag variierte der Erhebungszeitpunkt zwischen unmittelbar nach der Begegnung bis eine Woche später. Auch der Rahmen der Begegnung war nicht einheitlich: Einige Lehrkräfte luden den*die Zeitzeug*in in eine Klasse ein, während andere eine Großveranstaltung in der Aula organisierten, an der mehrere Klassen teilnahmen.

3.2.2 Stichprobe und Zeitzeug*innen

Insgesamt nahmen $N = 668$ Schüler*innen ($M = 16.11$ Jahre, $SD = 1.49$) aus 44 Lerngruppen der Klassenstufen 8 bis 13 an der Befragung teil. 52.1 Prozent der Befragten gaben an, weiblich zu sein, 45.8 Prozent männlich, 1.3 Prozent divers und

0.7 Prozent machten keine Angabe zu ihrem Geschlecht.⁴ Bei 29 Gruppen lagen zwischen der Begegnung mit den Zeitzeug*innen und der Erhebung weniger als 1 Tag, bei 8 Gruppen 1 bis 2 Tage, bei 3 Gruppen 3 bis 6 Tage und bei 4 Gruppen mehr als 6 Tage.

Die Daten wurden im Anschluss an Vorträge mit zehn Zeitzeug*innen gesammelt, die von unterschiedlich vielen Schüler*innen gesehen wurden. So haben 281 der 668 Schüler*innen die Geschichte desselben Zeitzeugens gehört, während die Geschichte eines anderen Zeitzeugen nur vier Schüler*innen gehört haben. Alle Zeitzeugenbesuche wurden als Vortrag mit anschließender Frageunde organisiert.

3.2.3 Instrument

Im Zentrum des Fragebogens stehen 50 Items, die 12 Konstrukte erfassen sollten. Die kognitive Involviertheit wurde über die Konstrukte Imagination, Aufmerksamkeit, Perspektivennachvollzug, (gefühlte) Erkenntnis und Kontextualisierung erhoben. Die emotionale Involviertheit wurde abgebildet durch die Konstrukte Bewegtsein, Identifikation, historische Unmittelbarkeit, Irritation und Einzigartigkeit des Moments. Die physische Involviertheit erfassten wir über die Konstrukte sinnliche Wahrnehmung und physiologische Reaktion.

Zur Erfassung der kritisch-distanzierten Verarbeitung adaptierten wir das von Bertram (2013) entwickelte Instrument zur Erfassung der Einsicht in die Perspektivität von Zeitzeugenaussagen. Exploratorische Faktorenanalysen ergaben, dass lediglich drei der vierzehn Items hohe Ladungen auf dem zentralen Faktor und keine substantiellen Nebenladungen auf weitere Faktoren aufwiesen. Die folgenden drei Items bildeten folglich in dieser Studie die kritisch-distanzierte Verarbeitung ab: «Zeitzeugen können uns wahre Auskünfte über die Vergangenheit geben, denn sie waren damals dabei» (invers codiert); «Viele Jahre später wissen Zeitzeugen noch genau, was früher alles passiert ist» (invers codiert); «Zeitzeugen vergessen möglicherweise, wie es genau gewesen ist.» Cronbachs Alpha der reduzierten Skala war akzeptabel ($\alpha = .68$).

Als Voraussetzungen der Lernerfahrung erhoben wir die Hintergrundvariablen Geschlecht und Interaktion zwischen Lernenden und Zeitzeug*in (Hast du eine Frage gestellt?). Zudem wurde in einem kurzen Lehrerfragebogen der Zeit-

4 Wenn wir im weiteren Verlauf der Studie von Mädchen und Jungen sprechen, beziehen wir uns auf jene Personen, die bei der Frage nach ihrem Geschlecht (weiblich, männlich, divers) das Kreuzchen bei Mädchen respektive Jungen gesetzt haben.

raum zwischen Besuch und Befragung und eine Einschätzung zur Vorbereitung des Besuchs erfragt (1 = wenig vorbereitet bis 4 = intensiv vorbereitet).

Die wahrgenommene Authentizität erfassten wir über drei Items, um drei Facetten abzubilden: Authentizität des Berichts: «Der Bericht des Zeitzeugen war für mich sehr glaubwürdig»; Authentizität der Person: «Der Zeitzeuge hat sich uns gegenüber ehrlich geöffnet»; Authentizität der Quelle: «Die persönlichen Geschichten von jemandem, der erlebt hat, wovon er berichtet, haben die Vergangenheit lebendig gemacht.» Die Items korrelierten im mittleren Bereich miteinander: Bericht und Person: $r = .25$; Bericht und Quelle: $r = .30$; Person und Quelle: $r = .23$.

3.2.4 Analyse

Für unsere erste Forschungsfrage begannen wir in einem ersten Schritt mit exploratorischen Faktorenanalysen mit Einfachladungsstruktur und testeten dazu jede mögliche Itemkombination für ein Set an Faktoren und die Veränderung der Modellgüte bei Ausschluss einzelner Items aus dem Messmodell. Aufgrund der hohen Anzahl möglicher Itemkombinationen mussten sich diese Analysen jeweils auf einen Teilbereich (also eine Itemauswahl) beschränken. In einigen Fällen überprüften wir, ob eine Erweiterung der Faktorenanzahl zu einer Verbesserung der Modellgüte führte. Ziel war hier, die Konstrukte so zu spezifizieren, dass die Items keine Mehrfachladung zu anderen Konstrukten aufwiesen. Diese Berechnungen dienten als Grundlage für eine weitere Kürzung der Itemanzahl pro Konstrukt. In einem zweiten Schritt überprüften wir die vorgeschlagenen Modelle der jeweils mithilfe eindimensionaler Items erfassten Konstrukte mit konfirmatorischen Faktorenanalysen und berechneten Cronbachs Alpha zur Beurteilung der Homogenität der Skalen. Wir untersuchten zudem mittels zwei Intraklassenkorrelationskoeffizienten ($ICC_{\text{unadjustiert}}$ und $ICC_{\text{adjustiert}}$), ob die unterschiedlichen Zeitzeug*innen im Sample für Unterschiede in der Involviertheit der Schüler*innen verantwortlich sind. Hierzu verglichen wir die relativen Varianteile der Skalen, die Unterschiede zwischen Klassen repräsentieren ($ICC_{\text{unadjustiert}}$), mit solchen aus einem Modell, in dem wir für Unterschiede zwischen Zeitzeug*innen (als Dummy-Variablen spezifiziert) kontrollierten ($ICC_{\text{adjustiert}}$). Eine relativ kleine $ICC_{\text{adjustiert}}$ würde dafür sprechen, dass die Unterschiede zwischen Klassen nicht (wesentlich) auf Unterschiede in der Schülerkomposition zurückgehen, sondern – bei deutlich höherer $ICC_{\text{unadjustiert}}$ – auf Unterschiede seitens der Zeitzeug*innen.

Zur Beantwortung der weiteren Forschungsfragen berechneten wir die latenten Korrelationen zwischen den Konstrukten und dem kritisch-distanzierten Faktor. Wir spezifizierten ein Regressionsmodell mit den persönlichen und situativen Merkmalen als Prädiktoren. Aus Gründen der Vergleichbarkeit der Modelle inkludierten wir alle hier im Fokus stehenden Variablen in jedem Modell als Korrelate, womit die Datengrundlage also jeweils identisch war. Zuletzt spezifizierten wir ein Mediationsmodell mit der wahrgenommenen Authentizität als Mediator. Wir verglichen, ob die standardisierten direkten Effekte aus dem Regressionsmodell nach Hinzunahme des Mediators kleiner wurden und prüften, ob die indirekten Effekte signifikant wurden. Die berichteten Regressionskoeffizienten wurden anhand der Varianz der abhängigen Variable sowie für kontinuierliche Prädiktoren zusätzlich an der Varianz der unabhängigen Variable standardisiert.

Wir bewerteten die Modellgüte nach dem *Root Mean Square Error of Approximation* (RMSEA), dem *Comparative Fit Index* (CFI) und dem *Standardized Root Mean Square Residual* (SRMR). Werte des RMSEA unter 0.05, CFI-Werte über 0.95 und SRMR-Werte unter 0.08 sprechen für eine akzeptable bis gute Modellpassung (Yu, 2002). Wir testeten die statistische Signifikanz jeweils auf dem Fünf-Prozent-Niveau. Die Analysen wurden mithilfe der Statistikprogramme Mplus 8.0 (Muthén & Muthén, 1998–2017) und SAS 9.4 (SAS Institute Inc., 2018) durchgeführt. Bei den konfirmatorischen Analysen in Mplus wurde die Clusterstruktur der Daten durch Korrektur der Standardfehler und Modelfit-Indizes über *Robust-Maximum-Likelihood*-Schätzung berücksichtigt.

4 Ergebnisse

4.1 Faktorenanalysen

Für die besondere Lernerfahrung standen 12 Konstrukte mit insgesamt 50 Items im Fokus dieses ersten Analyseschrittes. Nach der Prüfung, ob der Ausschluss von Items zu einer bedeutsamen Verbesserung der Modellpassung führte und Doppelladungen zu anderen Konstrukten vorhanden waren, schlossen wir schrittweise 10 Items (20 %) aus. Die Interkorrelationsstruktur der Konstrukte für sinnliche Wahrnehmung, Identifikation und Unmittelbarkeit wies einen Wert $r > 1$ auf, weshalb wir diese drei Konstrukte in einer exploratorischen Faktorenanalyse gemeinsam untersuchten. Wir extrahierten aus den Ergebnissen

einen Faktor mit vier Items, welcher das Eintauchen in die vergangene Zeit und das Spüren einer unmittelbaren Nähe zum*r Zeitzeug*in und zur Vergangenheit beschreibt und den wir als *historische Nähe* interpretieren. Die Items zum Konstrukt Irritation luden nicht eindeutig auf einen Faktor, weshalb diese ausgeschlossen wurden. Die Items des Konstrukts Imagination luden auf zwei Faktoren, weshalb wir die Items in die Faktoren *bildliche Imagination* und *abstrakte Vorstellungskraft* aufteilten. *Bildliche Imagination* beschreibt, dass Lernende lebendige Szenen zu dem Bericht im Kopf haben, während *abstrakte Imagination* erfasst, ob die Vergangenheit durch den Bericht für die Lernenden grundsätzlich vorstellbar war. Das Konstrukt der physiologischen Reaktion (mit zwei Items) schlossen wir aus, da die Items zu «stark» für die Begegnung mit DDR-Zeitzeug*innen formuliert waren (z. B. «Tränen in den Augen»). Denn diese Berichte waren häufig nicht so dramatisch, wie zum Beispiel jene von Holocaust-Überlebenden. Folglich arbeiteten wir in den weiteren Untersuchungen nur mit den Dimensionen der kognitiven und emotionalen Involviertheit.

Die Lernerfahrung mit den Zeitzeugenberichten zeichnet sich durch eine starke kognitive wie emotionale Involviertheit aus: Die kognitive Involviertheit lässt sich über die sechs Faktoren Aufmerksamkeitsfokus, bildliche Imagination, abstrakte Vorstellungskraft, Perspektivennachvollzug, gefühlte Erkenntnis und Kontextualisierung erfassen. Die emotionale Involviertheit beim Zeitzeugenbericht können wir über die drei Faktoren Bewegtsein, historische Nähe und Einzigartigkeit des Moments beschreiben. Das Modell mit diesen Konstrukten zeigt eine gute Modellpassung: RMSEA = 0.032, CFI = 0.968, SRMR = 0.039. Tabelle 1 gibt einen Überblick über die erfassten Konstrukte, ihre Kennwerte und Beispielimens. Ein Vergleich von $ICC_{\text{unadjustiert}}$ und $ICC_{\text{adjustiert}}$ zeigte, dass bei einer Kontrolle für Zeitzeug*innen immer noch etwa die Hälfte der Varianz auf Klassenebene verbleibt und somit die unterschiedlichen Zeitzeug*innen die Varianz der Involviertheit nur teilweise erklären können. Tabelle 2 zeigt die latenten Interkorrelationen zwischen den Faktoren. Mit Ausnahme von Kontextualisierung korrelierten alle Faktoren mittel bis stark miteinander. Die Interkorrelationen zeigten hierbei kein eindeutiges Muster mit Blick auf die Zuordnung zu emotionaler und kognitiver Involviertheit. Die Spezifikation von Second-Order-Faktoren führte zwar zu einer Verschlechterung der Modellpassung, wobei die Werte noch als akzeptabel betrachtet werden können: RMSEA = 0.039, CFI = 0.951, SRMR = 0.054.

Tabelle 1: Kennwerte der finalen Konstrukte zur Beschreibung der besonderen Lernerfahrung mit Zeitzeugenberichten

Be- reich	Konstrukt	Iteman- zahl	Beispielitem	N	M	SD	Cron- bachs Alpha	ICC unad- justiert	ICC adju- stiert	Range der std. Ladungen
Emotionale Involviertheit	Beweg- sein	3	Der Bericht des ZZ ging mir nahe.	667	4.01	1.04	.87	.20	.10	0.78–0.87
	Historische Nähe	4	Obwohl der ZZ von früher erzählt hat, habe ich eine Nähe zu dieser Zeit ge- spürt.	667	3.79	0.89	.78	.17	.07	0.62–0.72
	Einzig- artigkeit	4	Die Begegnung mit dem ZZ war für mich einzigartig.	668	4.03	0.98	.86	.15	.11	0.65–0.87
Kognitive Involviertheit	Aufmerk- samerkeits- fokus	3	Ich war bei dem ZZ- Bericht voll und ganz bei der Sache.	667	4.46	0.98	.86	.21	.12	0.71–0.89
	Bildliche Imagina- tion	3	Als ich dem ZZ zuge- hört habe, sind mir Szenen zu dem Be- richt in den Kopf ge- kommen.	668	4.39	1.03	.85	.09	.04	0.75–0.90
	Abstrakte Vorstel- lung	3	Für mich war die Zeit, von der der ZZ be- richtet hat, sehr gut vorstellbar.	668	4.68	0.73	.76	.10	.04	0.64–0.76
	Perspek- tivennach- vollzug	4	Ich kann die Empfin- dungen des ZZ ver- stehen.	668	4.86	0.72	.82	.13	.06	0.69–0.78
	Gefühlte Erkenntnis	5	Durch den ZZ-Bericht habe ich verstanden, wie es damals war.	666	4.66	0.71	.82	.07	.02	0.46–0.78
	Kontextua- lisierung	3	Ich habe die Begeg- nung mit meinem Wissen über die Zeit in Verbindung ge- bracht.	668	4.80	0.88	.88	.05	.04	0.71–0.80

Notiz. ICC_{unadjustiert} = relative Varianzanteile der Skalen, die Unterschiede zwischen Klassen repräsentieren; ICC_{adjustiert} = relative Varianzanteile der Skalen, wenn für Unterschiede zwischen Zeitzeug*innen kontrolliert wurde. Grundlage für den Range der standardisierten Ladungen ist das konfirmatorische Faktorenmodell mit eindimensionalen Faktoren.

Tabelle 2: Latente Interkorrelationen zwischen den Faktoren emotionaler und kognitiver Involviertheit

		1	2	3	4	5	6	7	8
1	Bewegtsein	1							
2	Historische Nähe	.75	1						
3	Einzigartigkeit	.76	.69	1					
4	Aufmerksamkeitsfokus	.55	.58	.61	1				
5	Bildliche Imagination	.54	.60	.50	.45	1			
6	Abstrakte Vorstellung	.51	.72	.54	.55	.70	1		
7	Perspektivennachvollzug	.58	.57	.53	.46	.55	.72	1	
8	Gefühlte Erkenntnis	.50	.57	.52	.38	.54	.78	.78	1
9	Kontextualisierung	.21	.21	.26	.27	.27	.31	.37	.30

Notiz. N = 668 Schüler*innen.

4.2 Zusammenhangsanalysen

4.2.1 Überwältigt vs. kritisch-distanziert

Die zweite Fragestellung suchte zu eruieren, wie die besondere Lernerfahrung mit der kritisch-distanzierten Verarbeitung zusammenhängt. Die Korrelationen bestätigten unsere Erwartungen: Die Involviertheit-Faktoren hingen statistisch signifikant negativ mit der kritisch-distanzierten Verarbeitung zusammen. Der Range lag im Bereich von $r = -.16$, $SE = 0.07$ mit Kontextualisierung bis $r = -.58$, $SE = 0.04$ mit gefühlter Erkenntnis. Das Modell zeigte gute Fit-Indizes: RMSEA = 0.032, CFI = 0.964, SRMR = 0.041.

4.2.2 Einfluss von Personen- und Situationsmerkmalen

Mit der dritten und vierten Fragestellung untersuchten wir, welche individuellen Faktoren der Schüler*innen und Situationsmerkmale eine starke Involviertheit vorhersagten. Hierfür spezifizierten wir ein Modell mit den Prädiktoren Geschlecht (dummy-codiert: 0 = Mädchen, 1 = Jungen) und Interaktion mit dem*der Zeitzeug*in, zeitlicher Abstand zwischen Begegnung und Erhebungstermin und der Intensität der Vorbereitung des Zeitzeugenbesuchs. Mädchen gaben konsistent an, stärker involviert zu sein als Jungen: Der stärkste Unterschied zeigte sich beim Faktor Bewegtsein: $b = -0.29$, $SE = 0.04$, $p < .001$. Der kleinste Unterschied

lag zwischen Geschlecht und bildlicher Imagination: $b = -0.11$, $SE = 0.04$, $p = .012$. Bei dem Faktor Kontextualisierung zeigte sich kein statistisch signifikanter Unterschied. Jungen hingegen verarbeiteten die Informationen aus dem Zeitzeugenbericht stärker kritisch-distanziert als Mädchen ($b = 0.17$, $SE = 0.06$, $p = .003$). Wenn Schüler*innen dem*r Zeitzeug*in eine Frage stellten, zeigten sie auf allen Konstrukten eine statistisch signifikant stärkere Involviertheit bei dem Bericht. Die Interaktion mit dem*r Zeitzeugen*in führte zu einer Reduktion der kritisch-distanzierten Verarbeitung ($b = -0.20$, $SE = 0.05$, $p < .001$). Für die Prädiktoren zeitlicher Abstand und Vorbereitung des Besuchs zeigten sich keine statistisch signifikanten Zusammenhänge. Der Determinationskoeffizient der latenten Variablen reichte von $.02$ (Kontextualisierung) $\leq R^2 \leq .15$ (Bewegtsein). Das Regressionsmodell wies gute Modellfit-Werte auf: RMSEA = 0.033 , CFI = 0.956 , SRMR = 0.038 . Die Ergebnisse bestätigten unsere Erwartungen mit Blick auf die Bedeutung persönlicher Merkmale für die Lernerfahrung. Für die Bedeutung der Intensität der Vorbereitung und den zeitlichen Abstand fanden wir kaum Hinweise. Das ist möglicherweise auf fehlende Varianz zurückzuführen (z. B. wurde kein Besuch «stark» vorbereitet).

4.2.3 Wahrgenommene Authentizität als Mediator

Mit der letzten Fragestellung wollten wir herausfinden, ob die wahrgenommene Authentizität die Effekte der Prädiktoren auf die Involviertheit mediiert. Hierfür ergänzten wir das Regressionsmodell um drei manifeste Items zur Erfassung der drei Facetten von Authentizität, verglichen die Unterschiede zwischen den totalen Effekten (c' , aus einem Regressionsmodell ohne Moderatoren) und den direkten Effekten und testeten die indirekten Effekte (z. B. Effekt von Geschlecht auf Bewegt sein mediiert über die Authentizität des Berichts) auf Signifikanz.⁵ Hierfür testeten wir, ob die Pfade von Prädiktor auf Mediator und von Mediator auf die abhängige Variable signifikant waren. Wir fanden das folgende Muster: Die direkten Effekte von Geschlecht auf Involviertheit wurden kleiner beziehungsweise waren nach Hinzunahme der Mediatoren nicht mehr statistisch signifikant. Die kleinste Differenz fanden wir beim Faktor Kontextualisierung ($c' - c = -0.08$), die größte Differenz zeigte sich beim Faktor Bewegtsein: $c' - c = -0.18$. Die indirekten Effekte von Geschlecht über die Authentizität des Berichts sowie über die Authentizität der Quelle auf alle Involviertheitskalen waren statistisch sig-

5 Wir verwenden hier die typischen Begriffe von Mediationsmodellen. Allerdings implizieren wir aufgrund des Studiendesigns keine kausalen Effekte.

nifikant – nicht jedoch über die Authentizität der Person (Ehrlichkeit). Auch der Effekt der Interaktion auf alle Involviertheitsfaktoren verringerte sich nach Hinzunahme der Mediatoren, jedoch nur marginal ($c' - c \leq 0.05$).

Die Effekte der Mediatoren (b) auf die Involviertheitsfaktoren waren insbesondere für die Facetten Authentizität des Berichts und der Quelle statistisch signifikant. Den stärksten Effekt hatte die Glaubwürdigkeit des Berichts auf den Perspektivennachvollzug ($b = 0.45$, $SE = 0.05$). Den kleinsten Effekt hatte Glaubwürdigkeit auf die Kontextualisierung ($b = 0.13$, $SE = 0.07$). Dieser war als einziger jedoch nicht statistisch signifikant. Die Authentizität der Quelle wiederum hatte den stärksten Effekt auf den Faktor historische Nähe ($b = 0.51$, $SE = 0.04$) und den schwächsten Effekt auf den Faktor Kontextualisierung ($b = 0.17$, $SE = 0.04$). Die Authentizität der Person zeigte lediglich statistisch signifikante Zusammenhänge mit der Kontextualisierung ($b = 0.12$, $SE = 0.04$), gefühlter Erkenntnis ($b = 0.12$, $SE = 0.04$) und kritisch-distanzierter Verarbeitung ($b = 0.14$, $SE = 0.06$).

Ein Vergleich der Determinationskoeffizienten in Regressions- und Mediationsmodell zeigte, dass durch Hinzunahme der Mediatoren deutlich mehr Varianz der Involviertheitsfaktoren erklärt werden konnte: $.09$ (Kontextualisierung) $\leq \Delta R^2 \leq .48$ (Perspektivennachvollzug). Das Mediationsmodell wies gute Modellfit-Werte auf: RMSEA = 0.032 , CFI = 0.958 , SRMR = 0.039 . Die Ergebnisse der Mediationsanalysen geben Hinweise, dass die wahrgenommene Authentizität, insbesondere gemessen über die Glaubwürdigkeit des Berichts und die allgemeine Zustimmung zur Bedeutsamkeit der Quelle, eine besondere Rolle für die Verarbeitung des Berichts spielen und die persönlichen Merkmale nicht unmittelbar auf die Verarbeitung wirken.

5 Diskussion, Limitationen und Ausblick

Ziele der vorliegenden Studie umfassten (1) die Entwicklung eines standardisierten Messinstruments, das es ermöglicht, die Lernerfahrung mit Zeitzeug*innen besser zu verstehen, (2) die Analyse, welche Zusammenhänge zwischen persönlichen und situativen Merkmalen und der Lernerfahrung sowie zwischen der wahrgenommenen Authentizität und der Lernerfahrung bestehen. Das von uns zugrunde gelegte theoretische Modell erwies sich als produktiv, weil tatsächlich viele Faktoren auch empirisch bestätigt werden konnten, auch wenn die finale Faktorstruktur teilweise abwich von den postulierten Faktoren. Insbesondere konnten wir die physische Involviertheit mit dem Fragebogen nicht reliabel er-

fassen. Zudem fanden wir zwar bestätigende Hinweise auf die Zuordnung der Dimensionen zu den Faktoren der kognitiven und emotionalen Involviertheit, allerdings waren diese weniger eindeutig als erwartet. Der standardisierte Fragebogen zur Erfassung der emotionalen und kognitiven Dimension des «Aura-Erlebnisses» in der Begegnung mit Zeitzeug*innen kann nun in weiteren Studien eingesetzt werden, auch wenn eine Weiterentwicklung und erneute Prüfung des Instruments sinnvoll erscheinen, um möglicherweise weitere Teilprozesse der Lernerfahrung identifizieren zu können.

Die vielleicht größte Einschränkung der vorliegenden Studie findet sich im Studiendesign: Die Thematik und das Setting der Zeitzeugenbegegnung variierten kaum, was letzten Endes zu einer eingeschränkten Varianz geführt haben könnte, die typischerweise dazu führt, dass die Anzahl und Ausprägtheit von Subdimensionen unterschätzt wird. Dies wiederum könnte dazu führen, dass die Ergebnisse der Studie nicht direkt auf andere Thematiken und das Setting übertragbar wären. Von daher wäre es wichtig zu untersuchen, ob beispielsweise die physische Involviertheit in einer Studie mit stärker variierendem Setting und mit anderen Messmethoden (z. B. Hautwiderstand) reliabler erfasst werden kann.

Für weitergehende Forschung bietet sich auch eine explizite Verknüpfung mit der modernen Forschung zur Unterrichtsqualität an (vgl. Kunter & Trautwein, 2013). Zeitzeugenberichte aktivieren die Lernenden kognitiv und emotional. Damit aber die Begegnung nicht mit einem Verlust an kritischer Distanz, sondern einem Zuwachs an historischem Denken einhergeht, braucht es eine Einbettung in ein kognitiv aktivierendes Unterrichtssetting auf einer Meta-Ebene sowie die konstruktive Unterstützung durch die Lehrkraft, damit Schüler*innen nicht mit ihren Emotionen alleingelassen werden. Trautwein, Schreiber und Hasenbein (2021) wiesen in ihrem Positionspapier zur Fachspezifität von Unterrichtsqualitätsmerkmalen explizit auf die besondere Bedeutung der konstruktiven Unterstützung im Geschichtsunterricht hin. Damit das Unterrichtssetting aber nicht nur zu einer «Kultivierung der Affekte» (von Borries, 1994, 90), sondern auch zu einem tiefgreifenden Verständnis über die Veränderbarkeit von Erinnerung und die Perspektivität von Zeitzeugenaussagen beiträgt, ist eine kognitiv aktivierende Rahmung der Begegnung notwendig. Nur wenn wir die Prozesse dieser besonderen Lernerfahrung in Gänze verstehen, können wir daraus eine angemessene Methodik und Gestaltung für den Unterricht ableiten, welche das «Aura-Erlebnis» mit Zeitzeug*innen zu einem «Katalysator» für das Lernen im Geschichtsunterricht machen.

Literatur

- Angvik, Martin & von Borries, Bodo (1997). *Youth and history. A comparative European survey on historical consciousness and political attitudes among adolescents*. Hamburg: Körber-Stiftung.
- Bertram, Christiane (2013). Chancen und Risiken von Zeitzeugenbefragung – Entwicklung eines Messinstruments für eine Interventionsstudie. In Jan Hodel/Monika Waldis & Béatrice Ziegler (Hrsg.), *Forschungswerkstatt Geschichtsdiaktik 12. Beiträge zur Tagung «geschichtsdiaktik empirisch 12»* (S. 108–119). Bern: hep.
- Bertram, Christiane (2020). Live oder medial vermittelt: Was lernen Schülerinnen und Schüler mit Zeitzeuginnen und Zeitzeugen? In Anne-Berinke Rothstein & Stefanie Pilzwegger-Steiner (Hrsg.), *Entgrenzte Erinnerung* (S. 247–266). Oldenburg: De Gruyter.
- Bertram, Christiane/Wagner, Wolfgang & Trautwein, Ulrich (2017). Learning historical thinking with oral history interviews: A cluster randomized controlled intervention study of oral history interviews in history lessons. *American Educational Research Journal*, 54(3), 444–484. <https://doi.org/10.3102/0002831217694833>.
- Bilandzic, Helena & Busselle, Rick W. (2011). Enjoyment of films as a function of narrative experience, perceived realism and transportability. *Communications*, 36(1), 29–50. <https://doi.org/10.1515/COMM.2011.002>.
- Borries, Bodo von (1994). Von gesinnungsbildenden Erlebnissen zur Kultivierung der Affekte? Über die Ziele und Wirkungen von Geschichtslernen. *Studien zur internationalen Schulbuchforschung*, 76, 67–92.
- Brauer, Juliane (2016). «Heiße Geschichte»? Emotionen und historisches Lernen in Museen und Gedenkstätten. In Sarah Willner/Georg Koch & Stefanie Samida (Hrsg.), *Doing History – Performative Praktiken in der Geschichtskultur* (S. 29–44). Münster: Waxmann.
- Brauer, Juliane (2019). Gefühlte Geschichte? Emotionen, Geschichte und historisches Lernen. *Vierteljahresschrift für wissenschaftliche Pädagogik*, 95, 272–283. <https://doi.org/10.30965/25890581-09501022>.
- Brauer, Juliane & Lücke, Martin (Hrsg.) (2013). *Emotionen, Geschichte und historisches Lernen – Geschichtsdiaktische und geschichtskulturelle Perspektiven*. Göttingen: V&R impress.
- Brüning, Christina Isabel (2018). *Holocaust Education in der heterogenen Gesellschaft*. Frankfurt a. M.: Wochenschau.
- Busselle, Rick & Bilandzic, Helena (2009). Measuring narrative engagement. *Media Psychology*, 12(4), 321–347. <https://doi.org/10.1080/15213260903287259>.
- Cameron, Catherine M. & Gatewood, John B. (2003). Seeking numinous experiences in the unremembered past. *Ethnology*, 42(1), 55–71.
- Csikszentmihalyi, Mihaly & Robinson, Rick E. (1990). *The art of seeing: An interpretation of the aesthetic encounter*. Los Angeles: Getty Center for Education in the Arts.
- Drozdowski, Danielle/De Nardi, Sarah & Waterton, Emma (2016). Geographies of memory, place and identity: Intersections in remembering war and conflict. *Geography Compass*, 10(11), 447–456. <https://doi.org/10.1111/gec3.12296>.
- Endacott, Jason & Brooks, Sarah (2013). An updated theoretical and practical model for promoting historical empathy. *Social Studies Research and Practice*, 8(1), 41–58.

- Gnambs, Timo/Appel, Markus/Schreiner, Constanze/Richter, Tobias & Isberner, Maj Britt (2014). Experiencing narrative worlds: A latent state-trait analysis. *Personality and Individual Differences*, 69, 187–192. <https://doi.org/10.1016/j.paid.2014.05.034>.
- Green, Melanie C. (2004). Transportation into narrative worlds: The role of prior knowledge and perceived realism. *Discourse Processes*, 38(2), 247–266. <https://doi.org/10.1207/s15326950dp3802>.
- Green, Melanie C. & Brock, Timothy C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5), 701–721. <https://doi.org/10.1037/0022-3514.79.5.701>.
- Green, Melanie C./Chatham, Christopher & Sestir, Marc A. (2012). Emotion and transportation into fact and fiction. *Scientific Study of Literature*, 2(1), 37–59. <https://doi.org/10.1075/ssl.2.1.03gre>.
- Henke-Bockschatz, Gerhard (2007). Zeitzeugenbefragung. In Ulrich Mayer/Hans-Jürgen Pandel & Gerhard Schneider (Hrsg.), *Handbuch Methoden im Geschichtsunterricht* (S. 354–369). Schwalbach/Ts.: Wochenschau.
- Hilton, Charlotte E. (2017). The importance of pretesting questionnaires: a field research example of cognitive pretesting the Exercise referral Quality of Life Scale (ER-QLS). *International Journal of Social Research Methodology*, 20(1), 21–34.
- Huijgen, Tim/van Boxtel, Carla/van de Grift, Wim & Holthuis, Paul (2017). Toward historical perspective taking: Students' reasoning when contextualizing the actions of people in the past. *Theory and Research in Social Education*, 45(1), 110–144. <https://doi.org/10.1080/00933104.2016.1208597>.
- Klika, Dorle (2018). Bildung und Emotion – Historisch-systematische Zugänge. In Martin Huber & Sabine Krause (Hrsg.), *Bildung und Emotion* (S. 75–90). Wiesbaden: Springer.
- Kunter, Mareike & Trautwein, Ulrich (2013). *Psychologie des Unterrichts*. Stuttgart: UTB.
- Lanman, Barry A. & Wendling, Laura M. (2006). *Preparing the next generation of oral historians*. Lanham: Altamira Press.
- Latham, Kiersten F. (2013). Numinous experiences with museum objects. *Visitor Studies*, 16(1), 3–20. <https://doi.org/10.1080/10645578.2013.767728>.
- Lombard, Matthew & Ditton, Theresa (1997). At the heart of it all: the concept of presence. *Journal of Computer-Mediated Communication*, 3(2), 1–36.
- Menninghaus, Winfried/Wagner, Valentin/Hanich, Julian/Wassiliwizky, Eugen/Kuehnast, Milena & Jacobsen, Thomas (2015). Towards a psychological construct of being moved. *PLoS ONE*, 10(6), 1–33. <https://doi.org/10.1371/journal.pone.0128451>.
- Mierwald, Marcel/Lehmann, Thomas & Brauch, Nicola (2018). Changing epistemological beliefs in student labs: Authentic learning materials as a chance to foster the development of academically adequate beliefs in the domain of history? *Unterrichtswissenschaft*, 46(3), 279–297. <https://doi.org/10.1007/s42010-018-0019-7>.
- Obens, Katharina & Geißler-Jagodzinski, Christian (2010). «Dann sind wir ja auch die letzte Generation, die davon profitieren kann». Erste Ergebnisse einer empirischen Mikrostudie zur Rezeption von Zeitzeugengesprächen bei Jugendlichen/jungen Erwachsenen. https://www.vielfalt-mediathek.de/wp-content/uploads/2020/12/forschungsbericht_zeitzeugengesprache.pdf (16.06.2021).

- Pandel, Hans-Jürgen (2012). *Geschichtsdidaktik. Eine Theorie für die Praxis*. Schwalbach/Ts.: Wochenschau.
- Plato, Alexander von (2001). Chancen und Gefahren des Einsatzes von Zeitzeugen im Unterricht. *BIOS Zeitschrift für Biographieforschung, Oral History und Lebensverlaufsanalysen*, 14(2), 134–139.
- Pouliot, Louise & Cowen, Paul S. (2007). Does perceived realism really matter in media effects? *Media Psychology*, 9(2), 241–259. <https://doi.org/10.1080/15213260701285819>.
- Sabrow, Martin (2012). Der Zeitzeuge als Wanderer zwischen den Welten [The contemporary witness as wayfarer between two worlds]. In Martin Sabrow & Norbert Frei (Hrsg.), *Die Geburt des Zeitzeugen nach 1945* (S. 13–32). Göttingen: Wallstein.
- Schreiber, Waltraud (2002). Reflektiertes und (selbst-)reflexives Geschichtsbewusstsein durch Geschichtsunterricht fördern – ein vielschichtiges Forschungsfeld der Geschichtsdidaktik. *Zeitschrift für Geschichtsdidaktik*, 1, 18–34..
- Schreiber, Waltraud et al. (2006). *Historisches Denken. Ein Kompetenz-Strukturmodell*. Neuried: ars una.
- Trautwein, Ulrich/Schreiber, Waltraud & Hasenbein, Lisa (2021). *Erkennt und nutzt die geschichtsdidaktische Forschung das Potenzial der Basisdimensionen der Unterrichtsqualität? Unveröffentlichtes Manuskript*.
- Uhl, Heidemarie (2012). Orte und Lebenszeugnisse – «Authentizität» als Schlüsselkonzept in der Vermittlung der NS-Verfolgungs- und Vernichtungspolitik. In Martin Rössner & Heidemarie Uhl (Hrsg.), *Renaissance der Authentizität? Über die neue Sehnsucht nach dem Ursprünglichen* (S. 258–284). Bielefeld: transcript.
- Vorderer, Peter (1992). *Fernsehen als Handlung. Fernsehfilmrezeption aus motivationspsychologischer Perspektive*. Berlin: Edition Sigma.
- Weber, Patrick & Wirth, Werner (2014). When and how narratives persuade: The role of suspension of disbelief in didactic versus hedonic processing of a candidate film. *Journal of Communication*, 64(1), 125–144. <https://doi.org/10.1111/jcom.12068>.
- Wierling, Dorothee (1997). Oral History. In Klaus Bergmann/Klaus Fröhlich/Anette Kuhn/Jörn Rösen & Gerhard Schneider (Hrsg.), *Handbuch der Geschichtsdidaktik* (S. 236–239). Seelze: Kallmeyer.
- Yu, Ching-Yun (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (30. Aufl.). Los Angeles: University of California.
- Zachrich, Lisa/Weller, Allison/Baron, Christine & Bertram, Christiane (2020). Historical experiences: A framework for encountering complex historical sources. *History Education Research Journal*, 17(2), 243–275. <https://doi.org/10.14324/HERJ.17.2.08>.

Abstracts

«Pictures Everyone Knows?»: An Empirical Study of Historical Knowledge Research Using Image Icons With Trainee History Teachers

Daniel Fastlabend und Johannes Meyer-Hamme

This empirical study explored how far historical knowledge can be gathered through image icons, using photographic icons from the twentieth and twenty-first centuries. This question was investigated by conducting a quantitative survey among trainee history teachers. This paper introduces the basic aspects of historical knowledge research as well as the characteristics of photographic icons. After discussing questionnaire design and the corresponding evaluation methodology, it considers exemplary contrastive findings on the (un)familiarity of iconic images. These are also differentiated between subgroups in order to discuss the characteristics of iconic images and to draw conclusions about historical knowledge.

Concept Mapping and the Constructivist Grounded Theory: Capturing Historical Knowledge with Semantic Networks

Burghard Barte

Contrary to the lively discussion and the theoretical modelling of historical knowledge, to date only few empirical results on how higher education learners structure their historical knowledge are available. This article exemplifies concept mapping as a data collection method suited to capturing historical concepts and the use of historical terms. Taking into account text linguistics, the developed analytical approach allows abductive as well as deductive-inductive analyses of concept maps. Implementation follows the three main coding phases of Kathy Charmaz's grounded theory methodology. To illustrate the analysis, coded segments from concept maps forming part of the author's research are presented. As a result, the article discusses the potential that concept maps offer for capturing historical terms and their use by learners to structure historical knowledge.

The «Aura Experience» in Encounters with Contemporary Witnesses: Understanding and Measuring a Complex Learning Experience*Lisa Zachrich, Wolfgang Wagner, Christiane Bertram und Ulrich Trautwein*

Pupils attach special importance to meeting contemporary witnesses. This paper describes the development of a standardised questionnaire to assess the pedagogical-psychological processes forming part of the «aura experience» in encounters with eyewitnesses as well as their connections with situational and personal characteristics. Data for this validation study came from 668 pupils attending grades 8 to 13 in Baden-Württemberg and Brandenburg who met contemporary witnesses from the former GDR. As theoretically assumed, we were able to empirically capture pupils' learning experience through their cognitive and emotional involvement. However, we also found deviations from the theoretical model: We were unable to empirically map physical involvement. We found indications of the importance of personal characteristics, but not of situational characteristics. The standardised questionnaire can now be used in further studies.

Conducting Dialogic Classroom Discussions: Case Analysis for Developing Professional Competency in History Teachers*Matthias Zimmermann, Miriam Moser, Anke Wischgoll, Kurt Reusser und Christine Pauli*

Classroom discussions are the predominant social form in Swiss history teaching. However, largely teacher-centred and quasi-monological discussion patterns provide only limited subject-specific learning opportunities. This paper describes one history teacher's trajectory in acquiring dialogic teaching skills during one year in-service professional development program (TPD). In a mixed-methods research design, video analyses from different phases of the TPD were used to investigate the development of a) the teacher's discourse actions with regard to quality features of dialogic teaching and b) argumentative qualities of learners' contributions. The results show that TPD had positive effects, including the significantly increased speaking time of learners and the use of dialogic teaching strategies by the teacher, which helped learners better substantiate their in-class contributions.

Promoting Didactic Knowledge and Skills for the Construction of History Assignments: an Intervention Study

Mario Resch, Christian Heuer und Hendrik Lohse-Bossenz

This paper examines how well a university course enabled prospective history teachers to develop (1) the didactic skills needed to construct assignments for the history classroom and (2) subject knowledge (historical knowledge and didactic knowledge). As part of an intervention study, students at the University of Heidelberg and Heidelberg University of Teacher Education were surveyed in a pre-post comparison group design (experimental group with a focus on task construction, $n = 22$ with follow-up survey $n = 5$; control group with a general focus on history didactics, $n = 15$) to determine their history teaching skills via a vignette test and their subject knowledge via a knowledge test. Both groups exhibited statistically significant intervention effects in historical knowledge and subject-specific didactic knowledge. In contrast, only the experimental group showed statistically significant developments with regard to subject-specific didactic skills for task construction. The findings of the follow-up survey indicate differential learning effects and a long-term impact of the intervention.

GeDiKo-T: Developing and Implementing a Competency Test for History Teachers

Stefanie Zabold

History teachers must be able to promote their students' historical thinking skills and, therefore, need subject-specific didactic skills, among other competencies. The GeDiKo-T enables collecting data on corresponding competencies. Teacher training and further education can draw practical benefits from this approach. This paper shows how the GeDiKo-T was developed based on a theoretical model (GeDiKo, History Didactic Competencies) in 12 steps. It also discusses how the model has been piloted, validated and standardised. Three sample tasks on subject-specific didactic structuring and competency categorisation, methodological competency and diagnostic competency provide insights into the task structure. In the context of GeDiKo-T, we are conducting various qualitative studies to further systematise research into history teacher competencies by means of triangulation. The paper concludes with a discussion of this aspect, among several others.

Historical Learning Using the «App in die Geschichte» [App into History]: Case Studies From a Preliminary Study of Historical Learning Processes and Narratives in a Digital Environment*Alexandra Krebs*

The more digitalized society becomes, the more critical digital aspects of learning and teaching become. In particular, the Covid-19 pandemic has highlighted the key role of digital learning tools, instruments, and teaching methods. Unfortunately, analog teaching forms still seem to persist in historical learning in schools. The few attempts at digital historical learning are often limited to simple quizzes or knowledge-based tasks. Furthermore, digital historical learning processes remain underresearched. Only a handful of studies exist to date. The project «App in die Geschichte» [App into History] aims to close this gap by developing and analyzing an innovative learning concept. This article presents a mixed-methods study of user behavior in the application. It analyses historical learning processes and narrations and presents two cases from a preliminary study involving university students to discuss the study design. Initial results indicate the considerable potential of the application to promote historical learning.

Historical Learning with Originals? On the Realisation of a Source Orientation in Museums*Andrea Brait*

The article presents some of the results of the «Historical Learning between School and Museum» research project. It examines the significance which history teachers attribute to museum exhibits when visiting museums in the context of school history education and the ways in which the pupils work with the exhibits during these visits. The analysis is based on history lessons in Austrian high schools and the permanent exhibitions of Austrian provincial museums. On the one hand, it can be seen that the Austrian history teachers who were interviewed frequently referred to the central value of exhibits but did not expect an analysis of these exhibits to be a starting point for a re-construction. On the other hand, it turned out that during the museum visits, exhibits were analysed especially when the students themselves conducted a guided tour or when they completed a worksheet provided by their teachers.

HiTCH II: What Further Developing the HiTCH Test Teaches Us about Historical Competencies

Wolfgang Wagner, Christiane Bertram and the HiTCH-Konsortium

The HiTCH project aims to devise and to constantly further develop a test for capturing and surveying historical competencies. The psychometric quality of the first version of HiTCH («HiTCH I») was tested in 2014 as part of a validation study. This also involved using further competency tasks in three additional test booklets (extension test booklets, EH). This paper first examines to what extent the newly developed tasks also capture the same construct and thus can be used to expand the HiTCH I task pool (HiTCH II). This held true for a total of 136 EH items. Second, it investigates to what extent these tasks capture EH-specific aspects and, third, whether these can be explained on the basis of item characteristics. We found clear indications of an additionally recorded dimension in EH 1 and EH 3. As for the item characteristics, significant impact was found solely for concreteness but not for real-life closeness (text versus image).

Is It Possible to Measure the Relevance of History? Initial Validation of the German-Language Version of the Relevance of History Measurement Scale (RHMS-D)

Marcel Mierwald und Julia Waldeyer

Although the practical relevance of history for people's lives may be said to be high from a theoretical and curricular perspective, this does not necessarily coincide with empirical findings on individual perception. However, we still lack an instrument, especially for German-speaking history didactics, for reliably and validly measuring whether people are convinced that history is relevant to their lives. We therefore first tested the psychometric qualities of the Relevance of History Measurement Scale (RHMS-D), translated from Dutch, with 257 students (approx. 22 years old; B.A. = 213/M.Ed. = 44). Data analysis showed that the survey questionnaire is able to validly and reliably measure three subspects of the relevance of history. We also found significant correlations with the surveyed external criteria and differences between B.A. and M.Ed. students in the perception of relevance, thus confirming expectations. We present and discuss the findings and implications for research and practice.

Promoting Reading in the History Class? An Intervention Study on the Effectiveness of Adaptive Language Support in Developing Subject-Specific Non-Fiction Texts*Helene Bergmann*

As a source of information and knowledge, specialist texts serve as a basis for historical teaching and learning processes also in the digital age. However, the linguistic features of factual texts and historical sources may impede comprehension, making the reading process susceptible to disruption, not only for learners with language difficulties. With learning groups becoming increasingly heterogeneous, history teaching is, therefore, challenged to develop differentiating language support that adequately supports both learners with reading difficulties and those with good reading skills in understanding subject-specific texts. The PhD project presented in this paper attempts to find empirical evidence for which forms of differentiating, subject-integrated language support effectively help pupils understand subject-specific non-fiction texts, and thus also the historical thinking and understanding. The paper focuses on the subject-specific use of reading strategies and investigates the effect of an adaptive historical learning setting based on strategy work. It presents the conceptualisation and initial results of the (quasi-)experimental study, which investigates the effects of subject-specific scaffolding measures. These are systematically varied according to learning and language prerequisites, among others, by employing predominantly quantitative methods.

Developing Multiperspectivity through Writing: Learning to Narrate Using Textual Procedures*Georg Marschnig*

Multiperspectivity is considered one of the core principles of history didactics and is also required normatively by school curricula. The habilitation project «Multiperspektivität als sprachliche Herausforderung» [Multiperspectivity as a Linguistic Challenge] focuses on implementing multiperspectivity in a writing setting that has been designed to help learners narrate multiperspectively with the help of textual procedures. This paper presents the theoretical foundations, the empirical framework and the initial findings of the project.

Promoting Narrative Competence and Epistemological Beliefs in History Classes? On the Effectiveness of a Writing Intervention in German-Swiss Grammar Schools

Martin Nitsche, Monika Waldis, Kristine Gollin, Philipp Marti und Manuel S. Hubacher

Experimental studies testing historical writing instruction that promotes students' narrative competence are still missing. We, therefore, conducted an intervention study in upper secondary schools in German-speaking Switzerland to foster students' narrative competence. Using a quasi-experimental research design, participants' narrative competence was assessed using material-based writing tasks. Because prior research indicated that students' epistemological beliefs are involved when they were asked to write historically, we also investigated students' epistemological beliefs by serving a closed-ended questionnaire. Regression analyses indicate that our intervention significantly impacted students' narrative competence. However, we found neither significant intervention effects on participants' epistemological beliefs nor correlations between them and students' narrative competence.

What Works? A Research Programme on the Effectiveness of Contemporary Witnesses in History Lessons

Christiane Bertram

Witnesses involve the opportunity and the risk of touching learners in a special way and thus of influencing learning outcomes: These are the main findings of the first randomised controlled intervention study on the effectiveness of contemporary witnesses in history teaching. Based on conceptualising the involvement of contemporary witnesses and the requirements for randomised controlled field trials (RCFT), this paper presents the first randomised controlled intervention study involving contemporary witnesses and three resulting research projects: (1) development of the «Aura Instrument»: Capturing special learning experiences with contemporary witnesses; (2) the interview project «Generation 1975»; and (3) the new randomised DFG contemporary witness study (in progress). The paper concludes with a discussion of the limitations and potentials of randomised and controlled intervention studies.

Which Scaffolds Could Help? Insights into an Intervention Study Aimed at Promoting Historical Explanation*Christoph Bramann, Carina Wöhlke und Nicola Brauch*

Explaining history can be understood as a complex, domain-specific linguistic activity that presents students with diverse challenges. However, little research exists on which aids («scaffolds») might support students in turning their explanations into language. The article presents an exploratory student laboratory study that compared using three different scaffolding tools (domain-specific — linguistic — integrating / integratively domain-specific and linguistic) for revising scripts for explanatory videos produced for lower upper secondary students ($N = 80$). Overall, the scaffolds were accepted by students, with integrative scaffolding being used most frequently. Further, many students found the scaffolds dealing with causality and temporality particularly accessible.

From Barbed Wire to Nivea Tins: How Carrying Out an Archaeological Excavation Can Trigger Conceptual Changes in Students' Historical Thinking*Johanna Sachse*

In everyday life, children and young people very often come (unconsciously) into contact with archaeology, with excavations and the «detective-like search» for clues of earlier life being followed by excitement. My dissertation and this article explore this intrinsic motivation and seek to render it fruitful for history didactics. The principal research question is to what extent conceptual changes in relation to archaeology as a science and to historical thinking can be initiated in pupils as a result of carrying out an archaeological excavation. This paper first explains the conceptual change approach on which the study is based, and then examines the first partial data of the dissertation within this theoretical framework. 12th grade students carried out an archaeological excavation at a former concentration camp in Bremen. Before and after the excavation, individual interviews were conducted with the participating students, who photographically recorded their impressions during the intervention. These data will be used to discover on which levels conceptual changes or extensions occurred.

Biografische Angaben

Burghard Barte ist wissenschaftlicher Mitarbeiter der Professur für Didaktik der Geschichte an der Universität Passau. Nach dem Studium des Lehramts an Gymnasien in den Fächern Geschichte und Politik-Wirtschaft an der Leibniz-Universität Hannover war er als wissenschaftlicher Mitarbeiter der Abteilung Didaktik der Geschichte an der Christian-Albrechts-Universität zu Kiel tätig. Seine Arbeitsschwerpunkte sind die Professionalisierung angehender Geschichtslehrpersonen, Theorien historischen Lernens und Wissens sowie digitales historisches Lehren und Lernen.

Helene Bergmann, M. A., ist wissenschaftliche Mitarbeiterin am Lehrstuhl für Geschichtsdidaktik der Universität Göttingen. Von 2015 bis 2021 war sie am Institut für Politik- und Geschichtswissenschaft der Pädagogischen Hochschule Freiburg i. Br. in der Lehrkräfteausbildung tätig. Seit 2020 betreibt sie im Rahmen des Forschungsverbunds HeLPS ein Dissertationsprojekt zu «Leseförderung im Geschichtsunterricht der Sekundarstufe I».

Christiane Bertram ist Juniorprofessorin für Fachdidaktik in den Sozialwissenschaften an der Binational School of Education, Universität Konstanz. Ihre Forschungsschwerpunkte sind Interventionsstudien, Kompetenzerfassung und Zeitzeug*innen und Oral History.

Andrea Brait, Priv.-Doz. MMag. Dr., ist assoziierte Professorin für Zeitgeschichte und Geschichtsdidaktik an der Universität Innsbruck sowie Lehrbeauftragte an der Universität Wien und verschiedenen Pädagogischen Hochschulen. In ihrem Habilitationsprojekt hat sie sich mit der Einbettung von Museumsbesuchen in den Geschichtsunterricht in Österreich beschäftigt.

Christoph Bramann, Mag., ist Studienreferendar für die Fächer Geschichte und Deutsch in Frankfurt. Zuvor war er unter anderem als wissenschaftlicher Mitarbeiter und Lektor im Bereich Geschichtsdidaktik an der Universität Salzburg und der Universität Bochum tätig. Seine Forschungsschwerpunkte liegen im Bereich der Schulbuchforschung, der Erforschung fachlicher Aufgabenkultur und Sprachbildung sowie der empirischen Geschichtsdidaktik.

Nicola Brauch, Prof. Dr., ist seit 2014 Professorin für Geschichtsdidaktik an der Ruhr-Universität Bochum. Seit 2012 ist sie im Kontext des HITCH Projekts an der Entwicklung des Kompetenztests für historisches Denken beteiligt. Ihre Interessen liegen außerdem im Bereich der Geschichtskultur und -theorie, der fachlichen Sprachbildung und der Thematisierung sensibler Themen im Geschichtsunterricht. Einen empirischen Arbeitsschwerpunkt hat die Bochumer Geschichtsdidaktik in der Unterrichtsforschung im Alfried Krupp-Schülerlabor.

Daniel Fastlabend ist wissenschaftlicher Mitarbeiter am Bereich für Theorie und Didaktik der Geschichte an der Universität Paderborn. Seine Forschungsschwerpunkte liegen im Bereich empirischer Geschichtsbewusstseins- sowie Geschichtskulturforschung, insbesondere vor dem Hintergrund der Migrationsgesellschaft und der Betrachtung internationaler Perspektiven.

Kristine Gollin ist wissenschaftliche Mitarbeiterin am IVP NMS in Bern.

Christian Heuer, Univ.-Prof. Dr. phil., ist Professor für Geschichtsdidaktik an der Karl-Franzens-Universität Graz. Seine Forschungsschwerpunkte liegen in den Bereichen der Theorie und Geschichte der Geschichtsdidaktik, der Pragmatik des historischen Lehrens und Lernens, der Orte und Medien der Geschichtskultur, der empirischen Professions- und Geschichtsunterrichtsforschung und der Lehrer*innenbildungsforschung.

Manuel S. Hubacher, MA, ist wissenschaftlicher Mitarbeiter am Zentrum Politische Bildung und Geschichtsdidaktik der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz in Aarau.

Alexandra Krebs arbeitet neben ihrer Tätigkeit als Lehrerin für Latein und Geschichte an einem Gymnasium als abgeordnete Lehrkraft an der Universität Paderborn am Arbeitsbereich Theorie und Didaktik der Geschichte. Zudem ist sie Mitglied einer Fortbildungskommission für historisch-politische Bildung der QUA-LiS NRW. Seit Oktober 2022 ist sie Fellow for Digital History (der Gerda Henkel Stiftung) am German Historical Institute (Washington D.C.) und dem Roy Rosenzweig Center for History and New Media der George Mason University (Fairfax, VA). Ihre Forschungsschwerpunkte liegen im Bereich digitaler Lernangebote für den Geschichtsunterricht sowie außerschulischer Lernorte.

Hendrik Lohse-Bossenz, JProf. Dr. phil, ist Juniorprofessor für Psychologie mit dem Schwerpunkt Lehr-Lern-Forschung an der Pädagogischen Hochschule Heidelberg. Seine Forschungsinteressen liegen im Bereich der kompetenzorientierten Professionalisierungsforschung von pädagogischen Fachkräften in unterschiedlichen Bildungsinstitutionen.

Georg Marschnig, Mag. Dr. phil., ist Senior Scientist für Geschichtsdidaktik am Institut für Geschichte der Karl-Franzens-Universität Graz. Seine Forschungsschwerpunkte sind Geschichtskultur und Public History, Geschichtsbewusstsein und Identität, Geschichtsdidaktik und Sprache, Erinnerung, Gedächtnis, Gedenken, Quellen und Kompetenzorientierung sowie lokale und regionale Zeitgeschichte.

Philipp Marti, Dr., ist wissenschaftlicher Mitarbeiter am Zentrum Politische Bildung und Geschichtsdidaktik der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz in Aarau.

Johannes Meyer-Hamme, Prof. Dr., ist Professor für Theorie und Didaktik der Geschichte an der Universität Paderborn. Seine Forschungsschwerpunkte umfassen Fragen zur Geschichtstheorie und zu Theorien historischen Lernens, insbesondere unter Berücksichtigung gesellschaftlicher Heterogenität, ebenso wie Fragen empirischer Forschung zum Geschichtsbewusstsein in der Gesellschaft und Kompetenzen historischen Denkens mit qualitativen und quantitativen Methoden.

Marcel Mierwald, Dr., ist wissenschaftlicher Mitarbeiter und Lehrbeauftragter am Arbeitsbereich Didaktik der Geschichte der Ruhr-Universität Bochum. Gegenwärtig ist er im Oral-History-Projekt «Menschen im Bergbau» zur Förderung des kritisch-reflektierten Umgangs mit Zeitzeug*innen-Interviews durch die digitale Lernplattform MiBLabor und im HiTCH-Projekt zur Entwicklung eines Kompetenztests für historisches Denken beteiligt. Seine Forschungsinteressen umfassen unter anderem das Lernen mit (digitalen) Medien, die empirische Unterrichtsforschung und die Sprachbildung im Fach Geschichte.

Miriam Moser ist Studienbereichsleiterin und Dozentin im Bereich Erziehungswissenschaften Sek I an der Pädagogischen Hochschule St. Gallen. Forschungsinteressen: Dialogisches Lernen, Unterrichtskommunikation und student agency.

Martin Nitsche, Dr., ist wissenschaftlicher Mitarbeiter am Zentrum Politische Bildung und Geschichtsdidaktik der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz in Aarau.

Christine Pauli, Prof. Dr., ist Professorin für Allgemeine Didaktik am Zentrum für Lehrerinnen- und Lehrerbildung der Universität Freiburg/Fribourg. Hauptverantwortliche Projektleiterin im vom SNF finanzierten Projekt «Socrates 2.0 – Lernwirksame Klassengespräche führen». Forschungsinteressen: Unterrichtsqualität, Unterrichtskommunikation und Lehrpersonenbildung.

Mario Resch, Dr. phil., ist Akademischer Rat im Fach Geschichte an der Pädagogischen Hochschule Heidelberg. Seine Forschungsinteressen liegen in den Bereichen der empirischen Professionalisierungsforschung, der Aufgabenkultur im Geschichtsunterricht und der Pragmatik historischen Lehrens und Lernens.

Kurt Reusser, Prof. em. Dr., ist emeritierter Professor für Pädagogische Psychologie und Didaktik an der Universität Zürich und seit 2013 Gastprofessor an der Leuphana Universität Lüneburg. Arbeitsgebiete: (Videobasierte) Lern- und Unterrichtsforschung, Didaktik, Lehrerinnen- und Lehrerbildung.

Johanna Sachse, Universität Bremen. Lehramtsstudium der Frankoromanistik und Geschichtswissenschaft in Bremen und Lille (Frankreich). Mehrjährige studentische Tätigkeit in der Landesarchäologie Bremen. Abschluss des 2. Staatsexamens und seitdem Lehrerin an einem Bremer Gymnasium. Seit 2017 Promotion an der Universität Bremen zur Einbindung von Archäologie in den Geschichtsunterricht.

Ulrich Trautwein, Prof. Dr., ist Professor für Empirische Bildungsforschung, Direktor des Hector-Instituts für Empirische Bildungsforschung sowie Co-Direktor des LEAD Graduate School & Research Network an der Universität Tübingen. Zu seinen Forschungsinteressen gehören unter anderem Fragen der Educational Effectiveness, der Unterrichtsqualität und der Motivationsentwicklung, die er unter anderem im Rahmen von großen Längsschnittstudien sowie interdisziplinär angelegten Projekten untersucht.

Wolfgang Wagner, Dr., ist wissenschaftlicher Mitarbeiter am Hector-Institut für Empirische Bildungsforschung an der Universität Tübingen. Im Zentrum seiner Forschungsinteressen stehen Fragen, die sich den Bereichen Educational Effectiveness, Unterrichtsqualität und Kompetenzmodellierung zuordnen lassen.

Julia Waldeyer, Dr., ist wissenschaftliche Mitarbeiterin in der Arbeitsgruppe Bildungspsychologie der Ruhr-Universität Bochum. Sie ist gegenwärtig unter anderem an der Evaluation des BMBF-Projekts DiAL:OGe «Digitalisierung in der Ausbildung von Lehramtsstudierenden: Orientierung und Gestaltung ermöglichen» sowie an diversen Projekten zum Thema Lern- beziehungsweise Studien-erfolg beteiligt. Ihre Forschungsinteressen liegen primär im Bereich des selbst-regulierten Lernens.

Monika Waldis, Prof. Dr., ist Leiterin des Zentrums Politische Bildung und Geschichtsdidaktik der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz in Aarau.

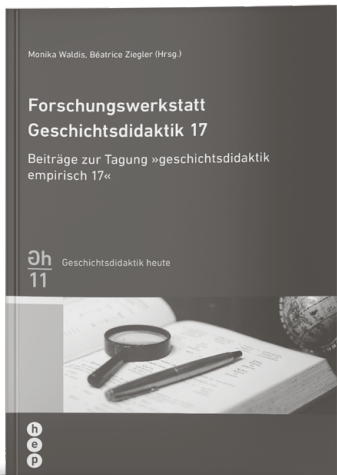
Anke Wischgoll, Dr., ist wissenschaftliche Mitarbeiterin am Institut für Entwicklung und Erforschung des Mathematikunterrichts an der TU Dortmund. Forschungsinteressen: Unterrichtskommunikation im Mathematikunterricht der Sekundarstufe.

Carina Wöhlke, Dr., ist wissenschaftliche Mitarbeiterin in der Didaktik der Physik an der Ruhr-Universität Bochum. Nach dem Studium der Mathematik und Physik (Lehramt Gymnasium) an der Universität Hamburg war sie 2015–2019 als wissenschaftliche Mitarbeiterin mit dem Ziel der Promotion in der Physikdidaktik der Universität Hamburg tätig. Ihre Forschungs- und Arbeitsschwerpunkte liegen im Bereich der Professionalisierung von Lehrkräften (professionelle Unterrichts-wahrnehmung, Mathematisierung), Bildung für nachhaltige Entwicklung und fachliche Sprachbildung im Physikunterricht sowie der Hochschullehre.

Stefanie Zabold, Dr., ist wissenschaftliche Assistentin am Lehrstuhl für Theorie und Didaktik der Geschichte der Katholischen Universität Eichstätt-Ingolstadt und Grundschullehrerin. Ihre Schwerpunkte sind schulische und außerschulische Kompetenzforschung. Im empirischen Bereich hat sie bisher qualitativ gearbeitet. In erster Linie mit dem Test GeDiKo-T beziehungsweise der Mitarbeit an HiTCH 3.0 erweitert sie ihren Schwerpunkt um quantitative Forschungsperspektiven. In theoretischer Hinsicht hat sie auf der Grundlage ihrer empirischen Forschungen zur Ausdifferenzierung des FUER-Modells beigetragen und befasst sich aktuell vor allem mit der Modellierung von geschichtsdidaktischen Lehrkompetenzen. Ihre pragmatischen Schwerpunkte liegen insbesondere in den Feldern Schule und Ausstellung/Museum.

Lisa Zachrich ist Doktorandin am Hector-Institut für Empirische Bildungsforschung und der LEAD Graduate School & Research Network in Tübingen. Das Master-Studium für Empirische Bildungsforschung und Pädagogische Psychologie in Tübingen und die Tätigkeit als studentische Hilfskraft in den Geschichtsprojekten von Christiane Bertram weckten ihr Interesse für Forschungsarbeiten an der Schnittstelle von Empirischer Bildungsforschung und historisch-politischer Bildung. In ihrem Dissertationsprojekt beschäftigt sich Lisa Zachrich mit verschiedenen Aspekten der besonderen Lernerfahrung von Schüler*innen mit Zeitzeugenberichten.

Matthias Zimmerman, Dr. phil, ist Lehrbeauftragter an der Pädagogischen Hochschule St. Gallen und Doktorassistent am Zentrum für Lehrerinnen- und Lehrerbildung der Universität Freiburg/Fribourg. Forschungsinteressen: Unterrichtsgespräche, Geschichtsdidaktik und Lehrpersonenfortbildungen.

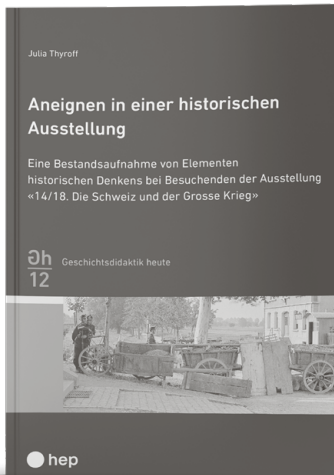


Monika Waldis, Béatrice Ziegler (Hrsg.)

Forschungswerkstatt Geschichtsdidaktik 17

**Beiträge zur Tagung
«geschichtsdidaktik empirisch 17»**

Der Band «Forschungswerkstatt Geschichtsdidaktik 17» versammelt verschiedene Beiträge der sechsten Ausgabe der Tagung «geschichtsdidaktik empirisch». Sie decken einerseits die drei Themenfelder ab, mit denen sich die empirische Geschichtsdidaktik intensiv befasst. Im Blick sind die historischen Kompetenzen der Schülerinnen und Schüler und deren Messung, Fragen der Lehrerprofessionalität sowie das Verstehen von Lehr-Lern-Vorgängen im geschichtskulturellen Kontext. Darüber hinaus führen die Beiträge die Diskussionen über einen virulenten Forschungsaspekt weiter, der an der Tagung im Rahmen der Keynotes aufgegriffen wurde: die «Übersetzung» zwischen Theorie und Empirie.



Julia Thyroff

Aneignen in einer historischen Ausstellung

Eine Bestandsaufnahme von Elementen historischen Denkens bei Besuchenden der Ausstellung «14/18. Die Schweiz und der Grosse Krieg»

Was denken Museumsbesuchende beim Rundgang durch eine historische Ausstellung? Worauf achten sie, welche Assoziationen äussern sie? Und welche Bezüge stellen sie zwischen der Ausstellung, der darin erzählten Geschichte und ihrer eigenen Person und Gegenwart her?

Mithilfe der Methode des Lauten Denkens untersucht die Autorin Aneignungsweisen bei Besuchenden der Ausstellung «14/18. Die Schweiz und der Grosse Krieg». Dabei zeigt sie Varianten des Umgangs mit Historischem auf und bietet einen Einblick in das Zusammenspiel von Ausstellung und individuellen Perspektiven der Besuchenden.

