

Martin Nitsche, Monika Waldis (Hrsg.)

Historisches Denken erforschen

gh
15

Geschichtsdidaktik heute





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 Zentrum für Demokratie Aarau

Prof. em. Dr. Markus Furrer
Em. Professor für Geschichte und Geschichtsdidaktik an der
Pädagogischen Hochschule Luzern

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

Prof. Dr. Nadine Ritzer
Professorin am Zentrum für Educational Governance und
Demokratiebildung der Pädagogischen Hochschule Zürich

Martin Nitsche und Monika Waldis (Hrsg.)

Historisches Denken erforschen

Reihe «Geschichtsdidaktik heute», Band 15





*Publiziert mit Unterstützung des Schweizerischen Nationalfonds
zur Förderung der wissenschaftlichen Forschung.*

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

ISBN Print: 978-3-0355-2658-5

DOI: <https://doi.org/10.36933/9783035526585>



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

Inhaltsverzeichnis

Einführung	7
Historisches Denken erforschen	9
<i>Martin Nitsche und Monika Waldis</i>	
Assessment historischen Denkens	29
Historische Kompetenzen im Umgang mit Medienberichten über den 6. Januar 2021	31
<i>Paul G. Fitchett, Carolin Hestler und Franziska Rein</i>	
Das Repertory-Grid-Interview als qualitativ empirische Forschungsmethode in der Geschichtsdidaktik	51
<i>Franziska Rein</i>	
Den Umgang mit Quellen und Darstellungen digital beforschen	72
<i>Jan Scheller, Jonas Schobinger, Martin Nitsche</i>	
Digitalisierung und historisches Denken	97
Künstliche Intelligenz im Fach Geschichte	99
<i>Christiane Bertram, Zarah Weiß, Lisa Zachrich, Ramon Ziai</i>	
«In ihrer Zeitung [...] versuchten sie an alle heranzukommen und sie an ihre tolle Kultur und an sie selbst zu erinnern»	124
<i>Kristopher Muckel</i>	

Sprachsensible Erforschung historischen Denkens	147
Historisches Erklären lernen?	149
<i>Tobias Flink</i>	
Zur sprachbasierten Diagnose historischen Denkens in Texten von Schüler*innen	171
<i>Viola Schrader</i>	
Abstracts	193
Autor*innen	197
Gutachter*innen der Beiträge	200

Einführung

Historisches Denken erforschen

Zur Ausgangslage

Martin Nitsche und Monika Waldis

1 Einleitung

Historisches Denken gilt oft als anthropologische Tatsache. So wird angenommen, es finde zu jeder Zeit und an jedem Ort statt, da Menschen vergangene Geschehnisse vergegenwärtigen müssen, um daraus Schlüsse in der Gegenwart für die Zukunft zu ziehen. Dadurch werde der gegenwärtige Moment mit Vergangenheit und Zukunftserwartungen verbunden, wodurch menschliches Handeln eine Richtung erhalte (z. B. Bracke et al., 2018; Körber, Schreiber & Schöner, 2007). Rüsen (1982) bezeichnet diesen Vorgang als «Sinnbildung über Zeiterfahrung» und betont zugleich seine lebenspraktische und akademische Bedeutung. Darüber hinaus fasst er historisches Denken als Fähigkeit des Geschichtsbewusstseins auf, das sich im Modus historischen Erzählens vollziehe und in einem kommunikativen Kontext eingebettet sei (ebd., S. 72). Diese Herleitung wird inzwischen auch im englischsprachigen Diskurs aufgegriffen, wenngleich auch andere Konzepte (s. u.) herangezogen werden (z. B. Lévesque & Clark, 2018).

Somit lässt sich das Erlernen historischen Denkens als Ziel historischer Bildung schlechthin auffassen. Allerdings liegen nur wenige systematische Einsichten zu üblichen Formen, Progressions- und Graduierungslogiken vor (Ammerer, 2022; Carretero & Perez-Manjarrez, 2022), was unter anderem daran liegt, dass sich historisches Denken als mentaler Vorgang von Subjekten individuell ausdrückt und nicht direkt beobachten lässt. Als «Black Box» (Luhmann, 1984, 156) ist es nur mittelbar zugänglich und erweist sich in der Folge sowohl für die pädagogische Praxis als auch die wissenschaftliche Analyse als schwer fassbar. Zugleich sind diagnostische Einblicke unabdingbar für die Bestimmung von Lernzielen, die Ausgestaltung von Fördersettings und Lernaufgaben sowie die Planung und Evaluation des Geschichtsunterrichts. Des Weiteren interessieren geschichtskulturell überformte Denkformen und Denkmuster, da sie in enger Beziehung zur individuellen und gesellschaftlichen Identität stehen und

somit die Grundlage für die Gestaltung gesellschaftlicher Lebensbereiche bilden. Nicht zuletzt deshalb stellt die geschichtsdidaktische Analyse historischen Denkens, die im englischsprachigen Kontext auch als Assessment bezeichnet wird (s.u.), ein zentrales Thema geschichtsdidaktischen Arbeitens dar.

1.1 Theoretische Konzeptionen und Modellierungen

Trotz der beschriebenen Bedeutung historischen Denkens wird es je nach Kontext unterschiedlich konzeptualisiert. Dabei sind wissenschaftliche Paradigmen etwa der Wissenspsychologie, Pädagogischen Psychologie, Geschichtstheorie und (Sozial-)Anthropologie grundlegend, die in verschiedenen wissenschaftlichen, politischen und geschichtskulturellen Diskursen wurzeln (Bracke et al., 2018).

Im deutschsprachigen Raum stellt seit den frühen 1970er-Jahren das Konzept des Geschichtsbewusstseins eine zentrale Referenz dar. Ursprünglich von Schörken im Diskurs prononciert, entwickelte es eine zentrale Integrationskraft als theoretisches Konzept und Grundlage für empirische Studien (Borries, 1995; Jeismann, 1988; Schörken, 1972). Nach Jeismann (1988) bildet Geschichtsbewusstsein die Instanz zur Reflexion der eigenen historischen Bedingtheit. Es ist konzipiert als ein zyklischer Prozess der Bildung, Infragestellung und Umbildung von Vorstellungen hinsichtlich vergangener, gegenwärtiger und zukünftiger Ereignisse und Zustände und impliziert Formen subjektiver Wahrnehmung, Wertung und Urteilsbildung. Rösen (1982) wiederum unterfütterte das Konzept mit geschichtstheoretischen Überlegungen zum Erzählbegriff. Folgt man Rösens (1983) Matrix des historischen Denkens, dann stehen zu Beginn eines historischen Denkaktes Bedürfnisse der Orientierung der menschlichen Lebenspraxis in der Zeit, hervorgerufen durch Kontingenzerfahrungen wie etwa Kriege, Umweltkatastrophen, Wirtschaftskrisen oder weitere Störungen alltäglicher Lebensvollzüge. In der Folge richten Menschen ihren Blick auf bestimmte Vergangenheiten, um sich in der Auseinandersetzung mit Quellen und Darstellungen Rechenschaft über diese Vergangenheiten abzulegen und Orientierungsoptionen in der Gegenwart für die Zukunft auszuloten.

Beide theoretischen Zugänge dienten im Zuge der PISA-Debatte und der Diskussion um Bildungsstandards (Klieme et al., 2003; Weinert, 2001) als Grundlage für die Ausarbeitung historischer Kompetenzmodelle (Barricelli, Gautschi & Körber, 2012; Gautschi, 2009; Körber et al., 2007; Pandel, 2005) und sind bis heute für das Assessment historischen Denkens leitend. Zugleich machten sich Vertreter*innen daran, historisches Wissen in Anknüpfung an wissens-

psychologische Konzepte sowie mit Bezug auf Second-Order-Konzepte (Carretero & Perez-Manjarrez, 2022) ausdifferenzieren (z. B. Bracke et al., 2018; Kühberger, 2012). Die vorliegenden Kompetenzmodelle bilden zudem den Ausgangspunkt für die Implementierung des Kompetenzbegriffs in Curricula, Lernstandserhebungen und Prüfungsvorgaben (Bertram et al., 2022; Kultusministerkonferenz, 2005), wobei in der praktischen Umsetzung teilweise eine Einführung auf kategoriales und prozedurales Wissen vorgenommen wurde (z. B. D-EDK, 2016).

Im Kontext der angloamerikanischen *history education* bildete lange Zeit die Konturierung relevanter *second-order-concepts* wie *evidence*, *significance* oder *explanation* den Ausgangspunkt und die theoretische Grundlage historischen Denkens. Sie wurden in mehreren Forschungsprojekten mit Kindern und Jugendlichen seit Mitte der 1970er-Jahre fundiert (Ashby, Lee & Dickinson, 1997; Carretero & Perez-Manjarrez, 2022; Seixas, 2017; Seixas & Morton, 2013; Wineburg, 1998a). Die Grundlage bildete dabei die Annahme, Schulunterricht solle Einsichten in die Struktur der hinter den Schulfächern stehenden Disziplinen und deren Wissensformen ermöglichen (Lévesque & Clark, 2018) oder sicherstellen, dass die Schüler*innen bis zum Ende der regulären Schulzeit jene Konzepte erlernen, die beim Umgang mit Geschichte involviert sind (Lee & Chapman, 2015). Entsprechend spielten im englischsprachigen Kontext mentale Prozesse von Historiker*innen eine wichtige Rolle während der Konzeption historischen Denkens (Lévesque & Clark, 2018). Die Basis legte Wineburg (1991a, 1991b, 1998b, 2001) in einer Serie von Vergleichsstudien mit Historiker*innen und Schüler*innen in den 1990er-Jahren, indem er die drei Heuristiken *sourcing*, *corroboration* und *contextualization* als wesentliche Aspekte historischen Denkens herausarbeitete. Im Kontext von nachfolgenden Ausarbeitungen etablierten sich umfassendere Konzepte historischen Denkens, in denen Teilaspekte ausdifferenziert wurden wie etwa die «historical thinking concepts» (Seixas, 2017; Seixas & Morton, 2013) oder das *historical reasoning*, das in einem aktuellen Ansatz zusätzlich den Blick auf historisches Wissen (*first – and second – order knowledge*) und subjektive Personenmerkmale wie epistemologische Beliefs und Geschichtsinteresse lenkt (z. B. van Boxtel & van Drie, 2018). Beide Konzeptionen bilden bis heute den Ausgangspunkt zahlreicher Studien (z. B. Nokes & De La Paz, 2023; Stoel, van Drie & van Boxtel, 2017), Curricula und schulischer Förderprogramme (Lévesque & Clark, 2018). Seit den 2000er-Jahren setzte auch im englischsprachigen Diskurs die Reflexion der außerschulischen Umgangsformen mit und der lebenspraktischen Bedeutung von Geschichte ein. In der Folge

wurde das Konzept des Geschichtsbewusstseins aufgegriffen (Grever & Adriansen, 2019; Nordgren, 2019; Popa, 2022; Seixas, 2004), wenngleich es häufig synonym zu historischem Denken verwendet wird (Gómez-Carrasco, Rodríguez-Medina, López-Facal & Monteagudo-Fernández, 2022).

Die Konzeptionen historischen Denkens sind nach wie vor Gegenstand theoretischer Auseinandersetzung. Offen ist grundsätzlich, worin die Domänenspezifik besteht, also ob die Historizität des Denkens formal vor dem Hintergrund einer narrativen Geschichtstheorie (z. B. Körber et al., 2007), funktional aufgrund der mentalen Ausrichtung auf Vergangenes und Geschichte(n) (van Boxtel & van Drie, 2018) oder inhaltlich anhand des Aussagegehalts der Ausprägungen historischen Denkens zu bestimmen ist (Geiss, 2018), zumal zugehörige Vorgänge nicht allein fachimmanent, sondern ebenfalls übergreifend (z. B. Lesekompetenz) fundiert (Trautwein et al., 2017) und zudem gesellschaftlich etwa für die politische Bildung oder Demokratieerziehung relevant sind (Barton, 2023; Lange, 2004). Weiterhin ist strukturell fraglich, inwiefern historisches Denken holistisch etwa mittels «narrativer Kompetenz» (z. B. Barricelli, 2005), Teiloperationen historischen Denkens (z. B. Körber et al., 2007) oder geschichtsspezifischen Sprachhandlungen operationalisiert werden sollte (Hartung, 2018). Zugleich wird der Einbezug historischen Wissens und subjektiver Konstrukte wie Interesse oder Beliefs in den Modellen unterschiedlich gehandhabt (Barricelli et al., 2012; van Boxtel & van Drie, 2018; Bracke et al., 2018; Lévesque & Clark, 2018). Schliesslich herrscht lerntheoretisch kein Konsens darüber, wie das Erlernen historischen Denkens so konzipiert werden kann, dass es sich in diversen Kontexten beschreiben lässt (z. B. Fenn & Zülsdorf-Kersting, 2023; Monte-Sano & Reisman, 2016).

Darüber hinaus werden in den letzten Jahren ausgehend von der Orientierungsfunktion von Geschichte in Individualisierungs- und Gruppenbildungsprozessen und diversitätstheoretischen Bezügen (Bräuer, 2022; Thorp & Persson, 2020) Brückenschläge zur Wahrnehmung und Deutung sozialer und gesellschaftlicher Strukturen vorgenommen, etwa unter Rückgriff auf erinnerungskulturelle oder inklusionspädagogische Konzepte oder mit Blick auf Geschichte(n) in der Migrationsgesellschaft (Georgi, Lücke, Meyer-Hamme & Spielhaus, 2022; Nordgren & Johansson, 2015; Völkel, 2017). Im historischen Rückblick zeigt sich, dass das Bildungsziel «historisches Denken lernen» Konjunkturen unterworfen ist, wobei der Grad der (Trans-)Disziplinarität und die gesellschaftliche Funktion immer wieder Gegenstand von Diskussionen sind. Im Zuge internationaler Bestrebungen, demokratische Werte als Fundament für

Bildungsprozesse stärker zu verankern, stehen derzeit politische und emanzipatorische Konzepte wie Teilhabe, Partizipation, Kritik, demokratisches Zusammenleben oder nachhaltiges Handeln im Vordergrund (z. B. Rapanta et al., 2021). Überfachliche und transversale Kompetenzen, wie sie in den 1970er-Jahren in Konzepten der Lernzielorientierung bereits entwickelt wurden und aktuell in Vorschlägen zu *future skills* (OECD, 2020) oder zu Demokratiekompetenzen (Barrett, 2016; Richardson & Milovidov, 2019) gefordert werden, bilden einen weiteren bildungspolitischen Rahmen des vorliegenden Zusammenhangs (Nordgren, 2017). Zukünftig ist somit offen, inwiefern veränderte gesellschaftliche Bedürfnisse und Herausforderungen neue Konzepte historischen Denkens prägen werden.

1.2 Empirische Zugänge zum historischen Denken

Die Untersuchung historischer Denkprozesse dient unterschiedlichen Zielen und Funktionen. In formalen Bildungskontexten werden individualdiagnostische Zugänge basierend auf beiläufigen Beobachtungen von Lernaktivitäten, mündlichen Beiträgen in Klassengesprächen oder schriftlichen Aufgabenlösungen häufig praktiziert. Sie informieren Lehrpersonen über den Fortgang von Lernprozessen. Die summative Beurteilung von Lerner*innen mit Blick auf die Lernzielerreichung mittels Prüfungen wird ebenso von Lehrpersonen vollzogen; manchmal auch im Vergleich mehrerer Klassen. Wissenschaftliche Analysen setzen oftmals qualitativ an und führen Fallanalysen oder -vergleiche durch. Schulübergreifende Lernzielüberprüfungen sowie (internationale) Large-Scale-Testungen des historischen Denkens sind nicht sehr zahlreich. Diese dienen dazu, verallgemeinerbare Aussagen zur Wirksamkeit von Unterrichtsmassnahmen oder der Effektivität eines Bildungssystems zu machen. Folglich werden zur Analyse historischen Denkens Methoden verwendet, die (a) sich in einem Kontinuum zwischen unstandardisiert und standardisiert bewegen und (b) den individuellen oder kollektiven Umgang mit Geschichte fokussieren (Shavelson et al., 2008; Shemilt, 2018).

Standardisierte Erhebungen historischen Denkens, die psychometrischen Anforderungen hinsichtlich Designs und Güte genügen und somit Vergleiche historischen Denkens von Individuen und Gruppen zulassen, sind eher selten. Unter den ersten Projekten sticht etwa die internationale Vergleichsstudie *Youth and History* hervor (Angvik & Borries, 1997; Borries, 1995). In den USA wird im Rahmen des *National Assessment of Educational Progress* (NEAP) seit den

2000er-Jahren nicht nur Faktenwissen, sondern auch *historical analysis* and *interpretation* als Operationen historischen Denkens geprüft (Bertram & Wagner, 2017). Als eines von wenigen Ländern kennt Schweden zentralisierte Geschichtstests, die laufend weiterentwickelt werden (Eliason, Alvé, Yngvéus & Rosenlund, 2015). Indes stiessen geschlossene Geschichtstests etwa auf die Kritik, dass sie ein objektivistisches Geschichtsverständnis befördern (Rosenlund, 2023) oder eher Faktenwissen statt Operationen historischen Denkens erfassen (VanSledright, 2014). Ebenso wird vermutet, dass die geschlossenen Testaufgaben mittels Lesekompetenz gelöst werden können (Bertram & Wagner, 2017). In Abgrenzung dazu beanspruchte die deutschsprachige HiTCH-Studie, zentrale Aspekte historischen Denkens im Sinn Körbers u. a. (2007) bei Schüler*innen der 9. Klassenstufe aus Österreich, Deutschland und der Schweiz zu erfassen. Zwar gelang dies in statistisch plausibler Weise, jedoch konnte nur ein Gesamtkonstrukt «historische Kompetenz» statistisch modelliert werden, statt, wie ebenfalls angestrebt, Teilbereiche zu identifizieren (Trautwein et al., 2017). Obwohl der Test inzwischen auszugsweise in anwendungsorientierten Studien zur Feststellung der Kompetenzentwicklung bei Schüler*innen Verwendung findet und damit seine Praxisnähe unterstrichen werden konnte (Bertram, 2017; Schreiber & Bertram, 2018), bleibt die statistische Differenzierung in Teilbereiche historischen Denkens herausfordernd.

Fortschritte konnten jüngst insbesondere in Untersuchungen von Teildimensionen historischer Kompetenz verzeichnet werden, wobei die Studienautor*innen teilweise einen Wechsel zu qualitativen Zugängen vollzogen. So sind mehrere Studien zu historischem Fragen entstanden (Behrendt & Brauch, 2020; Logtenberg, 2012; Schobinger, Scheller & Nitsche, 2023). Hinsichtlich der historischen Orientierung und Urteilsbildung hat Ammerer (2022) eine erste empirische Arbeit zur Lernprogression vorgelegt. Parallel dazu haben Autor*innen im angloamerikanischen Kontext Konzepte zu *ethical judgements* analysiert (Milligan, Gibson & Peck, 2018; Yoon, 2022). Zudem wurden vorliegende kognitive Modelle konzeptuell etwa hinsichtlich emotional-affektiver Prozesse ausdifferenziert (Borries, 2014; Zachrich, Weller, Baron & Bertram, 2020). Eine methodologische Erweiterung stellt die Erfassung historischen Denkens mittels offener, materialbasierter Schreibaufgaben dar (Hodel, Waldis, Zülsdorf-Kersting & Thünemann, 2013; Waldis, Marti & Nitsche, 2015; Waldis, Nitsche & Gollin, 2020), deren Bearbeitungsergebnisse mittels Kombination aus deduktivem Codieren und Textrating hinsichtlich historischer Performanzen untersucht werden können.

In weiteren Beiträgen wurden zudem anhand des *assessment triangle* (Pellegrino, Chudowsky & Glaser, 2001) Grundlagen der Erfassung historischen Denkens diskutiert (Ercikan & Seixas, 2015; Köster, Bernhardt & Thünemann, 2016) oder spezifische methodologische Probleme (z. B. Testung prozeduraler Aspekte) in den Blick gerückt (Reich, 2009). Ausserdem dienen verschiedene Ansätze der Erforschung historischen Denkens aus einer spezifischen Perspektive (z. B. Sprachsensibilität) oder hinsichtlich eines bestimmten Kontextes (z. B. Digitalisierung) (z. B. Barsch, Lutter & Meyer-Heidemann, 2019; Handro & Schönemann, 2022).

Gemeinsam ist den dargestellten Herangehensweisen, dass sie auf Vorgehensweisen basieren, mit denen historisches Denken von Subjekten identifiziert wird, um dessen Ausprägungsgrade und kontextuelle Bedingtheiten zu erfassen bzw. entsprechende Fördermassnahmen zu identifizieren (Shemilt, 2018; VanSledright, 2015). Dabei gilt es inzwischen als Konsens, dass zumindest das Zielkonstrukt (z. B. historisches Denken), das Erhebungsformat und die damit verbundenen Aufgabenstellungen (z. B. materialbasierte Schreibaufgaben) sowie das Beurteilungsinventar (z. B. Kategoriensystem) definiert und miteinander im Forschungsprozess so verknüpft werden müssen, dass historisches Denken nachvollziehbar zur Performanz gebracht und untersucht werden kann (Ercikan & Seixas, 2015; Köster et al., 2016; Nitsche, Gollin & Waldis, 2017; VanSledright, 2015).

Zukünftig könnten in den Studien stärker die Generalisierbarkeit der Ergebnisse und ihre Funktion für Deskription, Analyse und Graduierung historischer Denkprozesse reflektiert werden. So ermöglichen die vorliegenden Arbeiten häufig qualitative Einblicke in Ausprägungen historischen Denkens, lassen aber aufgrund des Studiendesigns Verallgemeinerungen kaum zu. Die bisherigen Überprüfungen von Fördermassnahmen historischen Denkens werden in Interventionsstudien zudem anhand von kleinen bis mittelgrossen Stichproben ($n < 20$ Gruppen) durchgeführt, wobei Erhebungsdesign, Thema und Materialauswahl oftmals variieren. Unterschiedliche Untersuchungskontexte sowie heterogene methodische Standards tragen zu einer weiteren Erhöhung der Vielfalt von Studienergebnissen bei. Umgekehrt lassen sich aus quantifizierenden Verfahren bisher noch zu selten pragmatische Schlussfolgerungen für die Geschichtsvermittlung oder vertiefte Einblicke in diagnostischer Hinsicht gewinnen (z. B. Bertram et al., 2022), sodass es zukünftig wünschenswert wäre, triangulative Designs oder Mixed Methods zu nutzen, um Chancen und Grenzen unstandardisierter und standardisierter sowie individual- und gruppenbezogener Analyseverfahren historischen Denkens zu kombinieren.

2 Die Beiträge des Bandes

Die Artikel dieses Bandes gehen die Untersuchung historischen Denkens in unterschiedlicher Art und Weise an und sind den Abschnitten «Assessment historischen Denkens», «Digitalisierung und historisches Denken» sowie «Sprach-sensible Erforschung historischen Denkens» basierend auf dem Schwerpunkt der Beiträge zugeordnet, wenngleich auch Querbezüge zwischen den Themenbereichen festzustellen sind. Im Abschnitt «Assessment historischen Denkens» stellen *Paul G. Fitchett*, *Carolin Hestler* und *Franziska Rein* ihre Studie zur Performanz historischer Kompetenzen von (angehenden) Geschichtslehrpersonen im Umgang mit Medienberichten über den 6. Januar 2021 vor. Historische und professionelle Kompetenz erfassen die Autor*innen entlang gängiger deutschsprachiger Modelle (z. B. Barricelli et al., 2012; Körber, 2022; Nitsche, 2021; Waldis, 2020) und koppeln sie mit englischsprachigen Konzepten der *media literacy* sowie der *intercultural communication competence*. In der Folge vergleichen sie deutsche und US-amerikanische Proband*innen, die sie bei der Auseinandersetzung mit dem Erhebungsthema beobachteten. Anschliessend verbindet *Franziska Rein* in ihrem Artikel zum Repertory-Grid-Interview die qualitativ empirische Forschungsmethode und Forschungsansätze zu Assessment und Diversitätssensibilität (z. B. Barsch, Degner, Kühberger & Lücke, 2020; VanSledright, 2014). Sie erläutert die Zusammenstellung des Repertory-Grid-Interviews für die Erhebung und Auswertung historischen Denkens aufgrund von Vorarbeiten und illustriert an Fallbeispielen dessen geschichtsdidaktische Verwendungsmöglichkeiten zur Diagnose von historischen Denkprozessen. Dabei profiliert die Autorin das Verfahren als partizipative Methode, die sich dazu eignet, die individuellen historischen Sinnkonstruktionen von Lernenden zu erschliessen. Dadurch rückt der Beitrag eine mögliche Vorgehensweise in den Blick, mit der die Ansprüche einer inklusiven geschichtsdidaktischen Forschung zukünftig eingelöst werden könnten. *Jan Scheller*, *Jonas Schobinger* und *Martin Nitsche* tragen zum Bereich Assessment Teilergebnisse ihrer Studie zur Analyse historischer Denkprozesse vor. Sie beobachteten Personen mit diverser Expertise beim Umgang mit Quellen und Darstellungen und arbeiteten qualitative Unterschiede beim historischen Denken durch die Proband*innen heraus. Die Studie verbindet Zugänge des Assessments historischen Denkens mit einer lerntheoretischen Perspektive (z. B. Fenn & Zülsdorf-Kersting, 2023; Shemilt, 2018), da sie anhand des vorliegenden Forschungsstandes und der Datenanalysen darauf abzielt, Kategorien zur qualitativen Differenzierung historischer Denkprozesse

und zugehöriger Operationen zu konturieren. Sie stellt zudem ein Beispiel dar, wie historisches Denken im digitalen Kontext erhoben werden kann (z. B. Barsch et al., 2019), auch wenn dies nicht den Schwerpunkt des Beitrags bildet.

Zwei Artikel adressieren direkt den Themenbereich «Digitalisierung und historisches Denken». *Christiane Bertram* stellt anhand von zwei Studien eines grösseren Forschungsvorhabens ihre Erfahrungen beim Einsatz von künstlicher Intelligenz (KI) zur Auswertung historischer Schüler*innentexte vor. Dazu entfaltet die Autorin zwei computerlinguistische Methoden: das *automatic content assessment* und das *linguistic complexity assessment*, welche dazu eingesetzt wurden, die inhaltliche Richtigkeit und die linguistische Komplexität der Texte einzuschätzen. Abschliessend diskutiert sie die Chancen und Grenzen der Verfahren für das zukünftige Assessment historischen Denkens. *Kristopher Muckel* präsentiert Teile einer Interventionsstudie, in der er die Wirkung von digitalisierten historischen Zeitungen gegenüber solchen im Faksimiledruck und didaktisierten Materialien auf das historische Denken von deutschen Schüler*innen der Klassenstufe 10 untersucht. Die Studie nutzt materialbasierte Schreibaufgaben zur Testung historischen Denkens und wertet die entstandenen Schüler*innentexte mittels qualitativer Inhaltsanalyse sowie quantitativ hinsichtlich der Wirkung der Intervention aus.

Den Abschluss des Bandes bilden zwei Untersuchungen im Abschnitt «Sprachsensible Erforschung historischen Denkens» (z. B. Handro & Schönmann, 2022). *Tobias Flink* stellt in seinem Beitrag eine Studie vor, in deren Zentrum die Vermittlung historischer Erklärungen steht. Dazu situiert der Autor die Textform in der Literatur und stellt angelehnt an die Conceptual-Change-Forschung seinen Zugang zur Erfassung des Konzepts «historische Erklärung» anhand von Daten von Schüler*innen vor, die an einer Intervention zum historischen Erklären teilgenommen haben. Dabei zeigt er auf, wie sich Konzeptänderungen bei Schüler*innen untersuchen lassen und welche Veränderungen sichtbar werden. *Viola Schrader* beschreibt die Entstehung eines Instruments zur Diagnose fachlicher und überfachlicher Sprachhandlungen beim historischen Argumentieren und illustriert diese anhand von Beispielen aus ihrer Dissertation. Für diese Analyse verknüpft die Autorin Konzepte historischen Denkens mit sprachsensiblen Ansätzen, stellt die Erhebungs- und Auswertungsmethoden vor und berichtet anhand historischer Argumentationen von deutschen Schüler*innen ausführlich von der Typenbildung, welche die Grundlage für das dargestellte Diagnoseinventar bildet. Zum Schluss werden dessen Anwendungsmöglichkeiten geschichtsdidaktisch kontextualisiert und diskutiert.

Insgesamt illustrieren die Beiträge, welchen Weg die (deutschsprachige) Geschichtsdidaktik bei der Analyse historischen Denkens inzwischen zurückgelegt hat. Das Konstrukt wird inzwischen sowohl theoretisch als auch empirisch in vielfältiger Hinsicht adressiert. Theoretische Konzepte wie historische Kompetenzen und sprachsensibler Geschichtsunterricht sowie professionstheoretische Zugänge bilden dabei produktive Ausgangspunkte. Empirisch dominieren die Erfassung historischen Denkens mittels schriftlicher Schüler*innenprodukte sowie der Zugang der qualitativen Inhaltsanalyse. Die Analyse mündlicher Beiträge sowie standardisierte Tests bilden noch immer die Ausnahme. So basieren die vorliegenden Studien oftmals auf materialbasierten Aufgaben mit vorab mehr oder weniger klar festgelegtem Lösungshorizont. Die Produkte der Studienbeteiligten werden von den Forschenden zumeist kategorial ausgewertet und teilweise quantifiziert, was dazu führt, dass in der Regel kleine Stichproben oder Fallanalysen vorliegen. Daraus resultieren vertiefte Einsichten zu Ausprägungen historischen Denkens und dessen kontextueller Bedingtheit wie dem Aufgabenkontext. Repräsentative Aussagen zum Stand historischen Denkens in ausgewählten (Schüler*innen-)Populationen fehlen nach wie vor. Die Erwartung, dass computerunterstützte Verfahren zukünftig die aufwendigen Analysen grösserer qualitativer Datensätze ermöglichen werden, schürt der Beitrag von Bertram. Allerdings dürfte die aktuell noch immer geringe Datengrundlage kaum hinreichend sein, um automatisierte Inhaltsanalysen zu programmieren, wie sie derzeit mit KI-unterstützten Systemen in Angriff genommen werden (z. B. Bertram, Weiss, Zachrich & Ziai, 2021). In dieser Hinsicht sind weitere Projekte notwendig. Jedoch gelingt es den vorliegenden Studien, Vorgehensweisen auszuloten, die sich für die geschichtsunterrichtliche Diagnostik und die Beurteilung von Performanzen historischen Denkens in schulischen Settings eignen. Wünschbar wären in dieser Hinsicht nicht nur Forschungsprojekte, die den Umgang mit wesentlichen Bereichen historischen Denkens erfassen, sondern die die Entwicklung operationaler Aspekte wie historisches Fragen und weiterer zugehöriger Konstrukte wie metakognitive Strategien oder Beliefs aufklären. Darüber hinaus dürfte die Entwicklung historischen Denkens im regulären Geschichtsunterricht dadurch gefördert werden, dass lern- und professionstheoretische Erkenntnisse mit partizipativen Forschungsdesigns verbunden werden, bei denen Lehrkräfte als Beteiligte die Chance haben, ihre Erfahrungen mit Unterrichtsentwicklung zu verknüpfen.

Einige Felder der Erforschung historischen Denkens sollten zukünftig stärker profiliert werden. Mit Blick auf digitales Geschichtslernen ist etwa trotz ers-

ter Studien offen (z. B. Goulding, 2021; McGrew, 2022; Mierwald, 2023), inwiefern digitale Kontexte historisches Denken grundlegend oder graduell verändern. Darüber hinaus scheinen vor dem Hintergrund des Klimawandels, der sich diversifizierenden Gesellschaften, von Kriegen usw. Untersuchungen nötig, die diese Herausforderungen adressieren und zugleich sowohl dem individuellen historischen Denken von Personen als auch den kontextuellen Anforderungen in Wissenschaft, Gesellschaft und Umwelt gerecht werden.

Wir hoffen, dass wir mit dem vorliegenden Band die Erforschung historischen Denkens befördern und anregen können. Wir danken den Autor*innen für die Beiträge und den Gutachter*innen für die hilfreichen Feedbacks. Bei Manuel S. Hubacher bedanken wir uns für seine redaktionelle Unterstützung während der Zusammenstellung des Bandes und bei den Mitarbeitenden des hep Verlages für ihr Engagement während des Publikationsprozesses. Allen Lesenden wünschen wir eine anregende Lektüre und interessante Auseinandersetzung mit den vorliegenden Studien zum historischen Denken.

Literatur

- Ammerer, Heinrich (2022). *Konzepte historischen Denkens und ihre Entwicklungslogik. Eine Studie zur Genese historischer Verständnishorizonte*. Frankfurt / M.: Wochenschau.
- Angvik, Magne & Borries, Bodo von (Hrsg.) (1997). *Youth and history: a comparative European survey on historical consciousness and political attitudes among adolescents*. Hamburg: Körber-Stiftung.
- Ashby, Rosalyn / Lee, Peter & Dickinson, Alaric (1997). How children explain the «why» of history. The Chata research project on teaching history. *Social Education*, 61(1), 17–21.
- Barrett, Martyn (2016). *Competencies for democratic culture: living together as equals in culturally diverse democratic societies*. Strasbourg: Council of Europe.
- Barricelli, Michele (2005). *Schüler erzählen Geschichte: Narrative Kompetenz im Geschichtsunterricht*. Schwalbach / Ts.: Wochenschau.
- Barricelli, Michele / Gautschi, Peter & Körber, Andreas (2012). Historische Kompetenzen und Kompetenzmodelle. In Michele Barricelli & Martin Lücke (Hrsg.), *Handbuch Praxis des Geschichtsunterrichts* (Bd. 1, S. 207–235). Schwalbach / Ts.: Wochenschau.
- Barsch, Sebastian / Degner, Bettina / Kühberger, Christoph & Lücke, Martin (Hrsg.) (2020). *Handbuch Diversität im Geschichtsunterricht. Inklusive Geschichtsdidaktik*. Frankfurt a. M.: Wochenschau.

- Barsch, Sebastian / Lutter, Andreas & Meyer-Heidemann, Christian (Hrsg.). (2019). *Fake und Filter. Historisches und politisches Lernen in Zeiten der Digitalität*. Frankfurt a. M.: Wochenschau.
- Barton, Keith C (2023). Learning in history and social studies. In Robert J. Tierney, Fazal Rizvi & Kadriye Ercikan (Hrsg.), *International Encyclopedia of Education* (4. Aufl., S. 315–322). Amsterdam: Elsevier. <https://doi.org/10.1016/B978-0-12-818630-5.14044-8>
- Behrendt, Lena & Brauch, Nicola (2020). 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. Theoretischen Grundlagen, empirische Forschungsmethoden und aktuelle Anwendungsgebiete* (S. 90–97). Münster & New York: Waxmann.
- Bertram, Christiane (2017). *Zeitzeugen im Geschichtsunterricht. Chance oder Risiko für historisches Lernen? Eine randomisierte Interventionsstudie*. Schwalbach / Ts.: Wochenschau.
- Bertram, Christiane & Wagner, Wolfgang (2017). Der Weg von NAEP zu HiTCH. Erfassung historischer Kompetenzen in standardisierten Formaten am Beispiel von Aufgaben der US-Geschichte. In Monika Waldis & Béatrice Ziegler (Hrsg.), *Forschungswerkstatt Geschichtsdidaktik 15. Beiträge zur Tagung «geschichtsdidaktik empirisch 15»* (S. 206–220).
- Bertram, Christiane / Wagner, Wolfgang / Kühberger, Christoph / Hülsen, Karin / Stello, Benjamin & Trautwein, Ulrich (2022). *Wie gut können Schülerinnen und Schüler historisch denken? Vorläufige Ergebnisse eines Projekts des IQSH mit der HiTCH-Gruppe*. Kiel: IQSH.
- Bertram, Christiane / Weiss, Zarah / Zachrich, Lisa & Ziai, Ramon. (2021). Artificial intelligence in history education. Linguistic content and complexity analyses of student writings in the CAHisT project (Computational assessment of historical thinking). *Computers and Education: Artificial Intelligence*, 100 038. <https://doi.org/10.1016/j.caeai.2021.100038>
- Borries, Bodo von (1995). *Das Geschichtsbewusstsein Jugendlicher. Erste repräsentative Untersuchung über Vergangenheitsdeutungen, Gegenwartswahrnehmungen und Zukunftserwartungen von Schülerinnen und Schülern in Ost- und Westdeutschland*. Weinheim und München: Juventa.
- Borries, Bodo von (2014). *Zwischen «Genuss» und «Ekel»: Ästhetik und Emotionalität als konstitutive Momente historischen Lernens (mit Beiträgen von Johannes Meyer-Hamme)*. Schwalbach / Ts.: Wochenschau.
- 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. <https://doi.org/10.1002/9781119100812.ch6>

- Bracke, Sebastian / Flaving, Colin / Jansen, Johannes / Köster, Manuel / Lahmer-Gebauer, Jennifer / Lankes, Simone / ... Zülsdorf-Kersting, Meik. (2018). *Theorie des Geschichtsunterrichts*. Schwalbach/Ts.: Wochenschau.
- Bräuer, Benjamin (2022). Reflektiertes und orientierungswirksames historisches Denken als geschichtsdidaktisches Ziel. In Georg Weißeno & Béatrice Ziegler (Hrsg.), *Handbuch Geschichts- und Politikdidaktik* (S. 93–107). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-29668-1_7
- Carretero, Mario & Perez-Manjarrez, Everado (2022). Learning history. In R. Keith Sawyer (Hrsg.), *The Cambridge Handbook of the Learning Science* (3. Aufl., S. 523–542). New York: Cambridge University Press.
- D-EDK (2016). *Lehrplan 21. Gesamtausgabe*. Luzern: Deutschschweizer Erziehungsdirektoren-Konferenz. Abgerufen von https://v-ef.lehrplan.ch/container/V_EF_DE_Gesamtausgabe.pdf
- Eliason, Per / Alvé, Frederik / Yngvéus, Cecilia & Rosenlund, David (2015). Historical consciousness and historical thinking reflected in large-scale assessment in Sweden. In Kadriye Ercikan & Peter Seixas (Hrsg.), *New directions in assessing historical thinking* (S. 171–182). New York: Routledge.
- Ercikan, Kadriye & Seixas, Peter (2015). Issues in designing assessments of historical thinking. *Theory Into Practice*, 54(3), 255–262. <https://doi.org/10.1080/00405841.2015.1044375>
- Fenn, Monika & Zülsdorf-Kersting, Meik (2023). Historisches Lernen – Ansätze einer Lerntheorie für den Geschichtsunterricht. In Monika Fenn & Meik Zülsdorf-Kersting (Hrsg.), *Geschichts-Didaktik. Praxishandbuch für den Geschichtsunterricht* (S. 54–97). Berlin: Cornelsen.
- Gautschi, Peter (2009). *Guter Geschichtsunterricht. Grundlagen, Erkenntnisse, Hinweise*. Schwalbach/Ts.: Wochenschau.
- Geiss, Peter (2018). Objektivität als Zumutung. Überlegungen zu einer postnarrativistischen Geschichtsdidaktik. *Zeitschrift für Geschichtsdidaktik*, 17, 27–41. <https://doi.org/10.13109/zfgd.2018.17.1.27>
- Georgi, Viola B. / Lücke, Martin / Meyer-Hamme, Johannes & Spielhaus, Riem (Hrsg.) (2022). *Geschichten im Wandel: neue Perspektiven für die Erinnerungskultur in der Migrationsgesellschaft*. Bielefeld: transcript.
- Goulding, James (2021). Historical thinking online: An analysis of expert and non-expert readings of historical websites. *Journal of the Learning Sciences*, 30(2), 204–239. <https://doi.org/10.1080/10508406.2020.1834396>

- Gómez-Carrasco, Cosme J. / Rodríguez-Medina, Jairo / López-Facal, Ramón & Monteagudo-Fernández, José (2022). A review of literature on history education: An analysis of the conceptual, intellectual and social structure of a knowledge domain (2000–2019). *European Journal of Education*, 57(3), 497–511. <https://doi.org/10.1111/ejed.12508>
- Grever, Maria & Adriaansen, Robbert-Jan (2019). Historical consciousness: the enigma of different paradigms. *Journal of Curriculum Studies*, 51(6), 814–830. <https://doi.org/10.1080/00220272.2019.1652937>
- Handro, Saskia & Schönemann, Bernd (Hrsg.) (2022). *Sprachsensibler Geschichtsunterricht. Geschichtsdidaktische Forschungsperspektiven und -befunde*. Münster: LIT.
- Hartung, Olaf (2018). Sprachhandeln und kognitive Prozesse von Schülerinnen und Schülern beim Schreiben über Geschichte. In Katharina Grannemann, Christian Kuchler & Sven Oleschko (Hrsg.), *Sprachbildung im Geschichtsunterricht. Zur Bedeutung der kognitiven Funktion von Sprache* (S. 67–89). Münster New York: Waxmann.
- Hodel, Jan / Waldis, Monika / Zülsdorf-Kersting, Meik & Thünemann, Holger (2013). Schülernarrationen als Ausdruck historischer Kompetenz. *Zeitschrift für Didaktik der Gesellschaftswissenschaften*, 4(2), 121–145.
- Jeismann, Karl-Ernst (1988). Geschichtsbewußtsein als zentrale Kategorie der Geschichtsdidaktik. In Gerhard Schneider (Hrsg.), *Geschichtsbewußtsein und historisch-politisches Lernen* (Bd. 1, S. 1–24). Pfaffenweiler: Centaurus.
- Klieme, Eckhard / Avenarius, Hermann / Blum, Werner / Döbrich, Peter / Gruber, Hans / Prenzel, Manfred / ... Vollmer (2003). *Zur Entwicklung nationaler Bildungsstandards. Expertise*. (Bundesministerium für Bildung und Forschung (BMBF), Hrsg.). Berlin: Bundesministerium für Bildung und Forschung. Abgerufen von <http://www.intranet.bbsii-kl.de/media/Unterricht/Lehr-%20und%20Lernforschung/Klieme%20u.%20a.%20Expertise%20Bildungsstandards.pdf>
- Körper, Andreas (2022). Kompetenzmodelle in der Geschichtsdidaktik. In Georg Weißeno & Béatrice Ziegler (Hrsg.), *Handbuch Geschichts- und Politikdidaktik* (S. 3–16). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-29668-1_1
- 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*, 29(174), 2–11.
- Kühberger, Christoph (Hrsg.) (2012). *Historisches Wissen. Geschichtsdidaktische Erkundungen zu Art, Tiefe und Umfang für das historische Lernen*. Schwalbach / Ts.: Wochenschau.

- Kultusministerkonferenz (2005). Einheitliche Prüfungsanforderungen in der Abiturprüfung Geschichte. Beschluss der Kultusministerkonferenz vom 01. 12. 1989 i. d. F. vom 10. 02. 2005. Abgerufen von https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1989/1989_12_01-EPA-Geschichte.pdf
- Lange, Dirk (2004). *Historisch-politische Didaktik. Zur Begründung historisch-politischen Lernens*. Schwalbach / Ts.: Wochenschau.
- Lee, Peter & Chapman, Arthur (2015). Learning history as seeing the world differently. Some thoughts from a small island. In Heinrich Ammerer, Thomas Hellmuth & Christoph Kühberger (Hrsg.), *Subjektorientierte Geschichtsdidaktik* (S. 340–355). Schwalbach / Ts.: Wochenschau.
- Lévesque, Stéphane & Clark, Penney (2018). Historical thinking: Definitions and educational applications. In Scott Alan Metzger & Lauren McArthur Harris (Hrsg.), *The Wiley International Handbook of History Teaching and Learning* (S. 117–148). London: Wiley. <https://doi.org/10.1002/9781119100812.ch5>
- Logtenberg, Albert (2012). *Questioning the past: student questioning and historical reasoning*. Ipskamp: University of Amsterdam.
- Luhmann, Niklas (1984). *Soziale Systeme: Grundriss einer allgemeinen Theorie*. Frankfurt a. M.: Suhrkamp.
- McGrew, Sarah (2022). Internet or archive? Expertise in searching for digital sources on a contentious historical question. *Cognition and Instruction*, 40(4), 488–516. <https://doi.org/10.1080/07370008.2021.1908288>
- Milligan, Andrea / Gibson, Lindsay & Peck, Carla L. (2018). Enriching ethical judgments in history education. *Theory & Research in Social Education*, 46(3), 449–479. <https://doi.org/10.1080/00933104.2017.1389665>
- Mierwald, Marcel (2023). Historisches Lehren und Lernen mit digitalen Medien – Geschichtsdidaktische Impulse für die Digital Humanities. In Christina Antenhofer, Christoph Kühberger & Arno Strohmeyer (Hrsg.), *Digital Humanities in den Geschichtswissenschaften* (S. 461–475). Wien & Köln: Böhlau.
- Monte-Sano, Chauncey & Reisman, Abby (2016). Studying Historical Understanding. In Lyn Corno & Eric M. Anderman (Hrsg.), *Handbook of Educational Psychology* (3. Aufl., S. 281–294). New York: Routledge.
- Nitsche, Martin (2021). Ansätze und Methoden der Professionsforschung in Geschichte. In Sebastian Barsch & Burghard Barte (Hrsg.), *Motivation – Kognition – Reflexion. Schlaglichter auf geschichtsdidaktischer Professionsforschung* (Bd. 12, S. 18–35). Frankfurt a. M.: Wochenschau.

- 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. <https://doi.org/10.13109/zfgd.2017.16.1.235>
- Nokes, Jeffery D. & De La Paz, Susan (2023). Historical argumentation: Watching historians and teaching youth. *Written Communication*, 074108832211486. <https://doi.org/10.1177/07410883221148679>
- Nordgren, Kenneth (2017). Powerful knowledge, intercultural learning and history education. *Journal of Curriculum Studies*, 49(5), 663–682. <https://doi.org/10.1080/00220272.2017.1320430>
- Nordgren, Kenneth (2019). Boundaries of historical consciousness: a Western cultural achievement or an anthropological universal? *Journal of Curriculum Studies*, 51(6), 779–797. <https://doi.org/10.1080/00220272.2019.1652938>
- Nordgren, Kenneth & Johansson, Maria (2015). Intercultural historical learning: a conceptual framework. *Journal of Curriculum Studies*, 47(1), 1–25. <https://doi.org/10.1080/00220272.2014.956795>
- OECD (2020). *OECD Lernkompass 2030. OECD-Projekt Future of Education and Skills 2030. Rahmenkonzept des Lernens*. Abgerufen von https://www.oecd.org/education/2030-project/contact/OECD_Lernkompass_2030.pdf
- Pandel, Hans-Jürgen (2005). *Geschichtsunterricht nach PISA. Kompetenzen, Bildungsstandards und Kerncurricula*. Schwalbach / Ts.: Wochenschau.
- Pellegrino, James W. / Chudowsky, Naomi & Glaser, Robert. (2001). *Knowing what Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.
- Popa, Nathalie (2022). Operationalizing historical consciousness: A review and synthesis of the literature on meaning making in historical learning. *Review of Educational Research*, 92(2), 171–208. <https://doi.org/10.3102/00346543211052333>
- Rapanta, Chrysi / Vrikki, Maria & Evagorou, Maria (2021). Preparing culturally literate citizens through dialogue and argumentation: Rethinking citizenship education. *The Curriculum Journal*, 32(3), 475–494. <https://doi.org/10.1002/curj.95>
- Reich, Gabriel A. (2009). Testing historical knowledge: Standards, multiple-choice questions and student reasoning. *Theory & Research in Social Education*, 37(3), 325–360. <https://doi.org/10.1080/00933104.2009.10473401>
- Richardson, Janice & Milovidov, Elizabeth (2019). *Digital citizenship education handbook: being online, well-being online, rights online*. Strasbourg: Council of Europe.

- Rosenlund, David (2023). The nature of historical knowledge in large-scale assessments – a study of the relationship between item formats and offerings of epistemic cognition in the Swedish national test in history. *Frontiers in Education*, 8. <https://doi.org/10.3389/educ.2023.1253926>
- Rüsen, Jörn (1982). Die vier Typen des historischen Erzählens. In Reinhart Koselleck, Heinrich Lutz & Jörn Rüsen (Hrsg.), *Formen der Geschichtsschreibung* (S. 514–605). München: dtv.
- Rüsen, Jörn (1983). *Historische Vernunft: Grundzüge einer Historik*. Göttingen: Vandenhoeck & Ruprecht.
- Schobinger, Jonas / Scheller, Jan & Nitsche, Martin (2023). Historisches Fragen aufgabenbasiert und digital beforschen. In Heike Krösche, Andrea Brait & Klaus Oberhauser (Hrsg.), *Neue Aufgabekultur im Geschichtsunterricht? Theoretische Zugänge und empirische Befunde* (S. 251–268). Frankfurt a. M.: Wochenschau.
- Schorcken, Rolf (1972). Geschichtsdidaktik und Geschichtsbewußtsein. In Hans Süssmuth (Hrsg.), *Geschichtsunterricht ohne Zukunft?* (S. 87–101). Stuttgart: Klett.
- 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* (S. 441–464). Göttingen: V & R unipress. <https://doi.org/10.14220/9783737008914.441>
- Seixas, Peter (Hrsg.) (2004). *Theorizing historical consciousness*. Toronto: University of Toronto Press.
- Seixas, Peter (2017). A model of historical thinking. *Educational Philosophy and Theory*, 49(6), 593–605. <https://doi.org/10.1080/00131857.2015.1101363>
- Seixas, Peter & Morton, Tom (2013). *The Big Six Historical Thinking Concepts*. Toronto: Nelson Education.
- Shavelson, Richard J. / Young, Donald B. / Ayala, Carlos C. / Brandon, Paul R. / Furtak, Erin Marie / Ruiz-Primo, Maria Araceli / ... Yin, Yue (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education*, 21(4), 295–314. <https://doi.org/10.1080/08957340802347647>
- Shemilt, Denis (2018). Assessment of learning in history education: Past, present, and possible futures. In Scott Alan Metzger & Lauren McArthur Harris (Hrsg.), *The Wiley International Handbook of History Teaching and Learning* (S. 449–471). New York: Wiley & Blackwell.

- Stoel, Gerhard / Drie, Jannet van & Boxtel, Carla van (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>
- Thorp, Robert & Persson, Anders (2020). On historical thinking and the history educational challenge. *Educational Philosophy and Theory*, 52(8), 891–901. <https://doi.org/10.1080/00131857.2020.1712550>
- Thünemann, Holger & Jansen, Johannes (2018). Historisches Denken lernen. In Sebastian Bracke, Colin Flaving, Johannes Jansen, Manuel Köster, Jennifer Lahmer-Gebauer, Simone Lankes, ... Meik Zülsdorf-Kersting (Hrsg.), *Theorie des Geschichtsunterrichts* (S. 71–106). Frankfurt am Main: Wochenschau.
- Trautwein, Ulrich / Bertram, Christiane / Borries, Bodo von / Brauch, Nicola / Hirsch, Matthias / Schröter, Kathrin / ... Zuckowski, Andreas (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 and Understanding: Innovative Designs for New Standards*. New York: Routledge.
- VanSledright, Bruce A. (2015). Assessing for learning in the history classroom. In Kadriye Ercikan & Peter Seixas (Hrsg.), *New Direction in Assessing Historical Thinking* (S. 75–89). New York: Routledge.
- Völkel, Bärbel (2017). *Inklusive Geschichtsdidaktik. Vom inneren Zeitbewusstsein zur dialogischen Geschichte*. Schwalbach/Ts.: Wochenschau.
- Waldis, Monika (2020). Geschichte in der Lehrerinnen- und Lehrerbildung. Geschichtskultur. In Colin Cramer / Johannes König / Martin Rothland & Sigrid Blömeke, *Handbuch Lehrerinnen- und Lehrerbildung* (S. 436–442). Wien u. a.: UTB. <https://doi.org/10.35468/hblb2020-051>
- Waldis, Monika / Marti, Philipp & Nitsche, Martin (2015). Angehende Geschichtslehrpersonen schreiben Geschichte(n). Zur Kontextabhängigkeit historischer Narrationen. *Zeitschrift für Geschichtsdidaktik*, 14, 63–86. <https://doi.org/10.13109/zfgd.2015.14e.1.63>
- 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>
- Weinert, Franz E. (Hrsg.) (2001). *Leistungsmessungen in Schulen*. Weinheim: Beltz.
- Wineburg, Samuel (1991a). 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>

- Wineburg, Samuel (1991b). On the reading of historical texts: Notes on the breach between school and academy. *American Educational Research Journal*, 28(3), 26. <https://doi.org/10.3102/00028312028003495>
- Wineburg, Samuel (1998a). Die psychologische Untersuchung des Geschichtsbewusstseins. In Jürgen Straub (Hrsg.), *Erzählung, Identität und historisches Bewußtsein. Die psychologische Konstruktion von Zeit und Geschichte. Erinnerung, Geschichte, Identität 1* (S. 298–337). Frankfurt a. M.: Suhrkamp.
- Wineburg, Samuel (1998b). 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
- Wineburg, Samuel. (2001). *Historical thinking and other unnatural acts. Charting the future of teaching the past*. Philadelphia: Temple University Press.
- Yoon, Jong-Pil (2022). Moral judgment in history education and historical positionality as a moral evaluator. *Theory & Research in Social Education*, 1–23. <https://doi.org/10.1080/00933104.2022.2117672>
- Zachrich, Lisa / Weller, Allison / Baron, Christine & Bertram, Christiane (2020). Historical experiences: A framework for encountering complex historical sources. *History Education Research Journal*, 17(2). <https://doi.org/10.14324/HERJ.17.2.08>

Assessment historischen Denkens

Historische Kompetenzen im Umgang mit Medienberichten über den 6. Januar 2021

Eine US-amerikanisch-deutsche Studie mit Lehramtsstudierenden und Lehrkräften

Paul G. Fitchett, Carolin Hestler und Franziska Rein

1 Einleitung

Der Angriff auf das US-Kapitol am 6. Januar 2021 markiert einen Wendepunkt in der amerikanischen Geschichte. Die USA positionierten sich weltweit jahrelang als Bollwerk westlicher Demokratie; dieses demokratische Ansehen wurde mit der Gewalt am Capitol Hill in Frage gestellt. Unmittelbar nach den Ereignissen des 6. Januar 2021 erschienen zahlreiche Berichte in den unterschiedlichsten Medien (z. B. Brückner, 2021; Fandos, Nicholas & Cochrane, 2021; McKay 2021; «Sturm aufs Parlament», 2021). Diese Medien bieten alternative und kontroverse Narrative bzw. Rechtfertigungen für die Ereignisse an. Mehr als ein Jahr später sind der Anschlag und die damit verbundene mediale Aufarbeitung des Ereignisses Teil der nationalen Geschichte der Vereinigten Staaten mit internationaler Resonanz, so auch in der deutschen Medienlandschaft.

Die verschiedenen Medienberichte lassen sich unter dem Stichwort Kontroversität (Salewski, 2014) in einem kompetenzorientierten Geschichtsunterricht behandeln, indem zeitgenössische Ereignisse und deren mediale Verarbeitung kritisch reflektiert werden. Die unterrichtlich zu schulenden Kompetenzen wie beispielsweise Methodenkompetenz (Schreiber et al., 2006) beschränken sich nämlich nicht auf historische Quellen, sondern schliessen zeitgenössische Darstellungen wie Zeitungsberichte mit ein. Damit Lehrkräfte dies fördern können, benötigen auch sie entsprechende Kompetenzen. In der vorliegenden Studie werden daher die Kompetenzen von (angehenden) Geschichts- bzw. Sozialkundelehrkräften in Deutschland und in den Vereinigten Staaten in ihrem Umgang mit kontroversen Medienberichten über den 6. Januar erforscht. Im Vordergrund stehen ihre Media Literacy, die historische Methodenkompetenz sowie die interkulturelle Kommunikationskompetenz (ICC). Die leitende For-

schaftsfrage lautet: Inwiefern bewerten deutsche und US-amerikanische Geschichts- und Sozialkundefachlehrer*innen die medialen Darstellungen des Anschlags auf das US-Kapitol am 6. Januar kritisch? Über welche Strategien verfügen sie, um diese Bewertung vorzunehmen und zu begründen? Um dieser Fragestellung nachzugehen, sind zunächst ausgewählte Aspekte der disziplinspezifischen Kompetenzdebatte heranzuziehen. Die Forschungsergebnisse werden anschliessend vor dem Hintergrund von Media Literacy, historischer Methodenkompetenz und Intercultural Communication Competence analysiert.

1.1 Historische Kompetenzen in historischen Lern- und Lehrprozessen

Im geschichtsdidaktischen Diskurs werden zahlreiche Kompetenzmodelle diskutiert. So geht Hans-Jürgen Pandels (Pandel, 2005) Kompetenzmodell über schulisches Geschichtslernen hinaus. Er unterscheidet verschiedene Kompetenzen, wird aber aufgrund einer fehlenden Theoriefundierung kritisiert (Trautwein et al., 2017, 39). Das FUEP-Kompetenzmodell unterscheidet zwischen historischer Methodenkompetenz, historischer Orientierungskompetenz, historischer Sachkompetenz und historischer Fragekompetenz (Schreiber et al., 2006). Das Modell von Michael Sauer (2006) fokussiert u. a. den Umgang mit Medien und Methoden von Schüler*innen beim historischen Denken. Dieses eng gefasste Modell betrachtet schulisch institutionalisiertes Geschichtslernen. Peter Gautschis Kompetenzmodell konzentriert sich auf die synthetischen Prozesse historischen Denkens und schulisch institutionalisiertes Lernen. Analytische Prozesse werden nicht näher beleuchtet (Gautschi, 2009). Das vorzustellende Forschungsprojekt bezieht sich auf das FUEP-Kompetenzmodell, da dieses nicht nur schulische Lern- und Lehrprozesse umfasst und sowohl in Bildungsplänen als auch in der Lehrkräfteausbildung anwendbar und weit verbreitet ist. Das FUEP-Modell wurde disziplinspezifisch breit rezipiert.¹ Dieses Modell ist detailliert abgeleitet, ausgearbeitet und begründet. Die zum FUEP-Modell gehörigen Kompetenzen (Körper, Schreiber & Schöner, 2007) der historischen Fragekompetenz, historischen Orientierungskompetenz und historischen Sachkompetenz stehen aufgrund der Fragestellung nicht im Fokus der Studie. Die Studie fragt nach der Rekonstruktions- und Dekonstruktionskompetenz der Teilnehmer*innen.

¹ Beispielsweise wurde es der HiTCH-Studie zugrunde gelegt (siehe Trautwein et al., 2017).

Nur wenn Lehrkräfte über entsprechende Kompetenzen verfügen, können sie professionell agieren (Rutsch, Rehm, Seidenfuß, Vogel & Dörfler, 2018, 10) und Schüler*innen bei ihrem Kompetenzerwerb unterstützen.² Lehrkräfte aller Ausbildungsphasen müssen also in der Kompetenzdebatte berücksichtigt werden: «Nicht vergessen werden sollte, dass es neben den Schüler- auch Lehrerkompetenzen gibt» (Pandel, 2014, 118). Heuer et al. mahnen an, dass die Fachlichkeit bei Lehrkräften oftmals vernachlässigt wird, obwohl diese besonders die professionelle Lehrer*innentätigkeit prägt (Heuer, Resch & Seidenfuß, 2019, 58). Darauf reagiert die vorliegende Studie; sie fokussiert das explizierbare geschichtsdidaktische Professionswissen der (angehenden) Geschichtslehrkräfte. Didaktisches Wissen wird in den Vereinigten Staaten weitgehend mit Shulmans (1986) Theorie des pädagogischen Inhaltswissens (PCK) in Verbindung gebracht, die davon ausgeht, dass sich die Art des Wissens, um ein Fach zu unterrichten, vom eigentlichen Fachinhalt unterscheidet. In der Geschichtsdidaktik wurde die PCK-Forschung, -Wissenschaft und -Lehrentwicklung weitgehend von der Stanford History Education Group geleitet (Seixas, 2017). Wineburg (2001) und seine ehemaligen Studierenden (Monte-Sano & Budano, 2013; Reisman, 2012) betonen die Relevanz der Beschaffung (Bestimmung des Wertes, der Gültigkeit und der Schlussfolgerung von Quellenmaterial), die Kontextualisierung (Analyse der Quelle innerhalb des historischen Zeitraums, in dem sie erstellt wurde) und die Bestätigung (Bewertung der Quelle neben konkurrierenden und übereinstimmenden Aussagen) als Ressourcen der Lehrkräfte. Aus dem deutschsprachigen Raum liegen erste Untersuchungen zu Kompetenzen von (angehenden) Lehrkräften vor (Fenn & Urban, 2021; Heuer, Resch & Seidenfuß, 2019; Resch & Heuer, 2019; Rutsch, Rehm, Seidenfuß, Vogel & Dörfler, 2018). Bisher sind noch keine Studien bekannt, welche die Kompetenzen von US-amerikanischen und deutschen (angehenden) Geschichtslehrkräften bei der Analyse von kontemporären Medienberichten vergleichend untersuchen.

1.2 Das Heidelberger Geschichtslehrerkompetenzmodell

Das Heidelberger Geschichtslehrerkompetenzmodell basiert auf der Konzeption von Jürgen Baumert und Eveline Wuttke (2016). Es ist in der geschichtsdidaktischen Debatte um Kompetenzen von Geschichtslehrkräften einschlägig und

² Daneben spielen ihre Überzeugungen, Motivation, professionelle Selbstkontrolle und ihr professionelles Wissen eine Rolle (z. B. Heuer, Resch & Seidenfuß, 2017, 28).

gegenüber dem Potsdamer Modell (Fenn & Urban, 2021) elaborierter. Während sich das Potsdamer Modell auf das Fachwissen konzentriert, untersucht das Heidelberger Modell die fachdidaktischen Fähigkeiten, Fertigkeiten sowie Bereitschaften von Lehrkräften theoriegeleitet, fundiert und systematisch. Daher fassen die theoretischen Überlegungen des Beitrags auf diesem Konzept und unterscheiden zwischen Geschichtswissen und geschichtsdidaktischem Wissen (Heuer, Resch & Seidenfuß, 2017).

Für (angehende) Lehrkräfte erscheint sowohl historisches als auch geschichtsdidaktisches Wissen relevant (Heuer et al., 2017, 28–30). Geschichtswissen umfasst das Wissen über den fachlichen Inhalt des Historischen, wobei geschichtsdidaktisches Wissen notwendig ist, «um historisches Denken und die es auszeichnenden Methoden und Kategorien zu lehren, die historischen Inhalte für Lernende verständlich aufzubereiten [...], um sie zur eigenen historischen Sinnbildung über die Erfahrung von Zeit zu befähigen (sinnbildend und sinnverstehend)» (Heuer, Resch & Seidenfuß, 2019, 61). Zum historischen Wissen gehören konsensuale Erkenntnisse der Disziplin der Geschichtswissenschaft, Geschichtstheorie und Geschichtsdidaktik. Es ist davon auszugehen, dass dieses Wissen stark von kulturellen und sozialen Aspekten geprägt wird. Geschichtsdidaktisches Wissen wiederum verbindet Geschichtswissen mit interdisziplinärem pädagogisch-psychologischem Wissen. Es ist notwendig, um die Erkenntnisse der Geschichtswissenschaft verstehbar werden zu lassen und einen Zugang für die Lernenden zu erschliessen (Heuer, Resch & Seidenfuß, 2017, 32). Dazu gehören Wissen um disziplinspezifische Kategorien, Konzepte, Begriffe und Prinzipien, wie beispielsweise (Multi-)Perspektivität (Salewski, 2014), Gegenwartsbezug (Heuer, 2014) oder Quellenorientierung (Beilner, 2014), sowie ein Methodenwissen, um historische Lern- und Lehranlässe zu gestalten. Geschichtslehrkräfte müssen u. a. kompetent darin sein, Quellen und Darstellungen zu analysieren und sie auf ihr didaktisches Potenzial hin zu prüfen (Heuer et al., 2017, 33). Dieser Schritt hängt mit Perspektivität zusammen, denn «perspective-taking is hardly an operation separate from reading sources» (Seixas, 2017, 601). Diese Annahmen sind grundlegend für das Forschungsdesign der vorliegenden Studie und schlagen einen Bogen zur historischen Methodenkompetenz nach dem FUER- Kompetenzmodell.

2 Das Forschungsdesign

Die leitende Forschungsfrage wird von folgenden Unterfragen begleitet:

- Welche Methoden benutzen die Teilnehmer*innen, um die Medienberichte einem Land und einer politischen Richtung zuzuordnen? Spielt die Ausprägtheit der Methodenkompetenz und der ICC hierfür eine Rolle?
- Welche Instrumente nutzen die Teilnehmer*innen, um die Medienberichte kritisch zu analysieren und zu reflektieren?

Dafür wurden die Daten in leitfragengestützten Interviews erhoben. Zur Vorbereitung auf die Interviews lasen die Teilnehmer*innen vier Medienberichte über die Ereignisse des 6. Januar 2021 in Washington und wurden aufgefordert, deren Herkunft (deutsch oder US-amerikanisch) zu identifizieren und sie politisch einzuordnen, siehe unten.³ Um die Vielschichtigkeit der Darstellung politischer Ereignisse präsentieren zu können, wurde je ein US-amerikanisches (Fandos & Cochrane, 2021; McKay, 2021) und ein deutsches Beispiel (Brückner, 2021; «Sturm aufs Parlament», 2021) für das eher linke bzw. rechte politische Spektrum gewählt. Die Interviews fanden mit allen Teilnehmer*innen in englischer Sprache statt⁴ und wurden folgendermassen strukturiert: Zunächst sollte die Zuordnung der Medienberichte in das Raster erläutert und die Unterschiede der Berichte erklärt werden. Anschliessend wurde nach der Glaubwürdigkeit der dargestellten Informationen gefragt. Danach sollten die Teilnehmenden ihre Sicht auf die mediale Berichterstattung und die Geschehnisse am 6. Januar 2021 darstellen. Es folgte eine Frage dazu, woher sie ihre Informationen bezogen haben. Abschliessend wurden die Probanden nach ihrer Einschätzung des Einflusses des Ereignisses auf die Demokratie und die Rolle der Medien darin gefragt. Bei den Forschungsteilnehmer*innen handelte es sich um vier US-amerikanische Lehrkräfte, zwei davon im Vorbereitungsdienst für Sozialkundeunterricht und zwei im Schuldienst, sowie fünf deutsche Geschichtslehramts-Masterstudierende. Hier werden Geschichtslehrkräfte in drei unterschiedlichen (Ausbildungs-)Phasen – aktiver Schuldienst, Studium und Vorbereitungsdienst – betrachtet. Das limitiert die Forschungsergebnisse, da offen ist, inwiefern sich die Kompetenzen aufseiten der deutschen Geschichts-

3 Die Namen der Medien oder Webseiten wurden zuvor unkenntlich gemacht. Die deutschen Berichte wurden von Erstsprachler*innen in das US-Amerikanische übersetzt.

4 Alle deutschsprachigen Studierenden studieren das Fach Englisch.

lehramtsstudierenden im Laufe des Studiums oder Vorbereitungsdienstes weiterentwickeln. Es wird davon ausgegangen, dass bereits im Studium Kompetenzen angelegt werden, über welche ausgebildete Geschichtslehrkräfte in der Lehrtätigkeit ebenfalls verfügen. Daher erscheint es legitim, Lehrkräfte unterschiedlicher Ausbildungsphasen zu befragen, um deren Kompetenzen zu untersuchen. Die Interviews wurden nach der Qualitativen Inhaltsanalyse (Mayring, 2015) getrennt voneinander ausgewertet und dann zueinander in Bezug gesetzt. Diese Methode wurde gewählt, da sie sich besonders für die theoriegeleitete, auf Konzepten basierende Analyse von Interviews eignet. Grundsätzlich wären auch andere Erhebungs- oder Auswertungsmethoden wie beispielsweise die Grounded Theory (Strauss & Juliet, 2010) denkbar. Da diese sich aber eher eignet, um bisher unbekannte Phänomene zu untersuchen und Theorien zu generieren (Strauss & Juliet, 2010), erscheint die Qualitative Inhaltsanalyse eher zielführend. Die Kategorien und Codes, mit denen die Interviews ausgewertet werden, wurden deduktiv aus der Theorie mit Rücksicht auf das Erkenntnisinteresse entwickelt und ergeben sich nicht aus den beobachteten Inhalten.⁵ Alle folgenden Ergebnisse der Studie sind rein qualitativ und nicht repräsentativ zu verstehen. Um die Vermutungen der Untersuchung zu überprüfen, sind weitere Forschungsstudien nötig.

3 Kompetenzen in der Lehrkräfteausbildung

3.1 Media Literacy

3.1.1 Theorie

Es gibt aktuell noch keine konsensuale Definition von Media Literacy. Nach dem Modell von Baacke (Baacke, 1997) gehören zur Media Literacy die vier Dimensionen Medienkunde, Medienkritik, Mediennutzung und Mediengestaltung. Dieses Modell muss mittlerweile um digitale Medien erweitert werden, was beispielsweise Treumann et al. unternommen haben (Treumann et al., 2007). Für die vorliegende Untersuchung erscheint besonders das Modell von Scheibe und Rogow (Scheibe & Rogow, 2011) vielversprechend, da es an die Überlegungen von Baacke anschliesst, wobei sich die sieben Dimensionen besonders anschlussfähig an geschichtsdidaktische Überlegungen zu historischen Kompetenzen

5 Die einzelnen Kategorien finden sich im Fliesstext in den jeweiligen Ausführungen zur empirischen Auswertung wieder.

zeigen. Zur Media Literacy gehört 1) *understanding* als die Fähigkeit und Fertigkeit, die grundlegenden, expliziten Botschaften von Medien zu verstehen als Vorstufe zur Fähigkeit, analytische Fragen dazu zu stellen, 2) *awareness* als Aufmerksamkeit, um die Rolle von Medienbotschaften im eigenen Leben zu erkennen, 3) *analysis* als die Fähigkeit und Fertigkeit, Botschaften von Medien durch kritisches und freies Denken zu entschlüsseln, 4) *evaluation* als Fähigkeit und Fertigkeit, informierte und begründete Urteile über den Gehalt und die Nützlichkeit der Medienbotschaft zu fällen, 5) *creation* als Fähigkeit und Fertigkeit, Medienbotschaften zu erzeugen und verschiedene Medienformate zu nutzen, 6) *reflection* als die Fähigkeit und Fertigkeit, zu erkennen, wie persönliche Erfahrungen und Werte den Umgang mit Medienbotschaften beeinflussen und wie Medien sich selbst und andere beeinflussen, und 7) *participation* als die Fähigkeit und Fertigkeit, an gemeinschaftlichen Aktivitäten rund um die virtuelle Medienwelt zu partizipieren (Scheibe & Rogow, 2011, 20). Critical Media Literacy ist ein Teil der Media Literacy und würde, so die Einschätzung der Autor*innen, jede Dimension des Modells nach Scheibe und Rogow um ein kritisches und (selbst-)reflexives Moment erweitern. Critical Media Literacy meint «active inquiry and critical thinking about the messages we receive and create» (Alliance for Media Literate America, 2007, 3). Diese Fähigkeit kann durchaus dazu beitragen, dass sich Personen aus einer machtkritischen Perspektive mit den Darstellungen von Medien und deren Botschaften beschäftigen und diese kritisch hinterfragen.

Ulf Kerber erschließt historische Medienbildung bereits 2016 geschichtsdidaktisch und leitet ein Kompetenz-Strukturmodell davon ab (Kerber, 2016). Die oben genannten Überlegungen zur Media Literacy und die disziplinspezifischen Überlegungen von Ulf Kerber lassen sich auf die historische Methodenkompetenz des FUER-Kompetenzmodells übertragen bzw. daran anschließen. So lässt sich begründet annehmen, dass bei den Dimensionen 1) *understanding*, 2) *awareness*, 3) *analysis*, 4) *evaluation*, 5) *creation* und 6) *reflection* grundlegende Operationen vollzogen werden und Kompetenzen notwendig sind, die in der historischen Methodenkompetenz als Kompetenz zur Rekonstruktion und Dekonstruktion zu finden sind. Wird Dimension 7) *participation* erweitert und umfasst sie nicht nur virtuelle Aspekte, sondern auch die Fähigkeit und Fertigkeit, sich im eigenen Leben zu orientieren und beispielsweise an gemeinschaftlichen geschichtskulturellen Einrichtungen teilzuhaben, lässt sich auch diese Dimension der Media Literacy geschichtsdidaktisch anbinden. Media Literacy könnte also als ein Bestandteil von historischer Methodenkompetenz erschlossen werden.

Wie der Vergleich der obigen Ergebnisse zeigt, ordneten die amerikanischen Teilnehmer*innen die Artikel häufiger korrekt in die vorgegebenen Quadranten ein. Kein*e deutsche*r Teilnehmer*in schätzte die Herkunft aller Artikel richtig ein. Bedeutet das auch, dass sie unterschiedliche Kompetenzen aufwiesen?

Alle Teilnehmer*innen argumentierten bei der Frage der Zuordnung mit sprachlichen Hinweisen, sodass davon ausgegangen werden kann, dass die amerikanischen Studierenden einen gewissen Vorteil bei der Einteilung hatten. Den amerikanischen Teilnehmer*innen fiel beispielsweise auf, dass Medienberichte, die sie als *conservative* beschrieben, dazu tendierten, die Verantwortung von Einzelpersonen abzulenken und gleichzeitig die Demonstrant*innen des 6. Januar eher positiv zu schildern. So gibt beispielsweise David zum Beitrag von Fox News an: «*They refer to people as <Trump loyalists>: again, instead of calling someone an insurrectionist.*»⁸ Hier wird deutlich, wie subtil manche sprachlichen Hinweise in Medien sind. Zweitsprachenlernenden fallen solche Details wohl erst nach jahrelanger Übung auf. Für die deutschen Teilnehmenden stand die Ausgewogenheit der Sprache im Mittelpunkt. Sie identifizierten die Äußerungen als *inflammatory* oder *emotionally driven*⁹ und ordneten sie eher dem rechten politischen Spektrum zu. Für diese Studierenden war die Äusserung von Emotionen in den Artikeln in erster Linie ein Zeichen für den amerikanischen Ursprung des Textes. Ihnen schienen deutsche Medien «*less emotional in general about news*»¹⁰ zu berichten. Was David dem konservativen Artikel attestierte, nämlich dass er eine unterschwellige Bewertung des Geschehens vornehme, wurde von den amerikanischen Befragten als ein typisches Element linksgerichteter Medien angebracht. So äusserte Michelle die Ansicht, der Artikel aus der New York Times behaupte, dass «*the siege was a climax of a weeklong campaign by Mr. Trump*».¹¹

Nach Maßgabe des vorgestellten Modells der Media Literacy lässt sich sagen, dass alle Aussagen über das basale *understanding* hinausgingen. Es zeigte sich aber, dass die deutschen Studierenden aufgrund der sprachlichen Hürde Probleme mit der Kategorie *analysis* hatten. Die Äusserungen von Maria über die Art der verwendeten Sprache hingegen macht deutlich, dass sie ein begrün-

8 Interview mit David (US) vom 01.06.2021, durchgeführt von Paul Fitchett.

9 Interview mit Maria (D) vom 09.04.2021, durchgeführt von Carolin Hestler und Franziska Rein.

10 Interview mit Maria (D) vom 09.04.2021, durchgeführt von Carolin Hestler und Franziska Rein.

11 Interview mit Michelle (US) vom 06.05.2021, durchgeführt von Paul Fitchett.

detes Urteil über die Nützlichkeit der Medien (*evaluation*) aus ihrer persönlichen Sicht (*reflection*) trifft. Davids Aussage zeigt, dass er sich der Rolle von Medien im eigenen Leben bewusst ist (*awareness*), indem er zu einer eher positiven Darstellung der Ereignisse des 6. Januar kommt und das Kollektiv stärker in den Vordergrund rückt als einzelne Akteur*innen.

3.2 Historische Methodenkompetenz

3.2.1 Theorie

Waltraud Schreiber versteht Re- und Dekonstruktionskompetenz als die «Basisoperationen des historischen Denkens» (Schreiber, 2008, 198). Auch für die vorliegende Studie sind Rekonstruktions- und Dekonstruktionskompetenz die zentralen Operationen zur Analyse zeitgenössischer Medienberichte. Dem FUER-Modell liegt die konsensuale Annahme zugrunde, dass es sich bei Geschichte immer um eine dekonstruierbare Konstruktion handelt. Bei der Rekonstruktionskompetenz steht, so Schreiber, der re-konstruierende Umgang mit Quellen sowie die Entwicklung einer historischen Narration im Vordergrund (Schreiber, 2006, 200). Dazu gehört die Fähigkeit und Fertigkeit, eine innere und äußere Quellenkritik vornehmen zu können (Schreiber, 2006, 202). Für die vorliegende Analyse der Kompetenzen der (angehenden) Lehrkräfte ist besonders die innere Quellenkritik interessant. Mit einer inneren Quellenkritik lässt sich eine Aussage darüber treffen, was der*die Autor*in sagt und meint. Im Fokus steht hier, die Quelle in den (historischen) Kontext einzuordnen und vor diesem Hintergrund zu verstehen (Schreiber, 2006, 203–204).

Eine De-Konstruktion einer historischen Narration folgt einer Fragestellung; sie erfasst zum einen ihre Oberflächenstruktur, also die Elemente der Narration. «Die inhaltliche Struktur zu erheben und jenseits der Formalie gattungs-, medien- und adressatenspezifische Aspekte herauszuarbeiten (sic!) sind die Hauptziele des Teilschritts der De-Konstruktion» (Schreiber, 2006, 226). Mit der Analyse der Tiefenstruktur der Quelle nähert man sich dem an, welche Intentionen beispielsweise hinter der Narration stehen; es geht um implizite oder latente Motivationen. Subtile Orientierungsangebote werden herausgearbeitet. Darüber hinaus lässt sich die Argumentationsstruktur nachvollziehen und interpretieren. Indem zusätzliches Material oder Wissen herangezogen wird, lässt sich die Triftigkeit der historischen Narration überprüfen; hier könnten ausgewählte Passagen überprüft und eingeordnet werden, um die Urteile über

die historische Narration zu untermauern. Hier gilt es, selbst zu einem Ergebnis zu kommen und Vergleiche anzustellen (Schreiber, 2006, 226–229).

«Die Antworten auf historische Fragen haben notwendig Konstruktcharakter» (Schreiber, 2008, 201). Den Unterschied zwischen der Vergangenheit und einer gegenwärtig konstruierten Geschichte müssen Schüler*innen sowie Lehrkräfte wahrnehmen, wenn sie kompetent mit Narrationen umgehen. Geschichten als plausible, triftige und valide Darstellungen des Vergangenen aus unterschiedlichen Sichtweisen (Schreiber, 2008, 201) können dabei durchaus kontrovers ausfallen. Waltraud Schreiber mahnt an, dass «mündige Bürger [...] in der Lage sein [sollten], sich zu den unterschiedlichen Deutungen und Sinnbildungen bewusst zu verhalten, sie zu beurteilen und ggf. zu kritisieren, die in unterschiedlichen Gruppen üblichen Sichtweisen abzugleichen oder zu ersetzen» (Schreiber, 2008, 199). De- und Rekonstruktionskompetenz sind die entscheidenden Fähigkeiten und Fertigkeiten für die komplexen Operationen.

Bei der historischen Methodenkompetenz geht es zum einen darum, eine historische Narration zu entwickeln, zum anderen darum, Strukturen von Narrationen wahrzunehmen und zu erfassen. Es handelt sich um ein Zusammenspiel von Rekonstruktion und Dekonstruktion, wobei bei der Entwicklung historischer Narrationen die synthetische Rekonstruktion, beim Erfassen der Strukturen hingegen die analytische Dekonstruktion im Vordergrund steht (Schreiber, 2008, 204). Der Fokus der vorliegenden Studie liegt aufgrund der Fragestellung auf der Dekonstruktion von Medienberichten.

3.2.2 Die Ergebnisse der Studie

In den Interviews finden sich Belege für verschiedene Methoden innerhalb der Dekonstruktionskompetenz, die die Befragten anwendeten bzw. anzuwenden versuchten. Der amerikanische Lehrer David gab an, dass er die Frage der *Triftigkeit* einer Darstellung an der Anzahl der verwendeten Zitate und deren Ursprung ausmache. Für die Glaubwürdigkeit sprach seiner Meinung nach, wenn Zitate aller politischen Positionen genannt wurden.¹² Während die amerikanischen Teilnehmer*innen deutlich öfter eigene Gefühle und eigene Beobachtungen, z. B. selbst Gesehenes im TV, als Begründung für die Glaubwürdigkeit von Darstellungen heranzogen, war das Vorgehen bei den deutschen Studierenden eher intersubjektiv und kriteriengeleitet. Aber auch hier gibt es Beispiele, die zeigen,

¹² Vgl. Interview mit David (US) vom 01.06.2021, durchgeführt von Paul Fitchett.

dass die Methodenkompetenz theoretisch verstanden, aber nicht angewandt wurde. So ist das folgende Zitat von Meltem sehr vage bezüglich der genauen Kriterien für Triftigkeit: *«I definitely see the ones that I put in, on the German liberal side, A and D article. I think they're the best, because [...] they're like lots of information in there. Like even if you haven't heard about anything they just say everything. Like what happened, why did it, why did this happen and everything. That's what I feel like article A and D (too). For article B [...], this sounds hard, but it's like trash kind of [...].»*¹³ In einer früheren Passage sagte sie, dass Quellen jeglicher Art nicht die Wahrheit darstellen, sondern nur einen Standpunkt abbilden. Es wird deutlich, dass sie wusste, dass ein multiperspektivisches Vorgehen und eine Ansammlung von unterschiedlichen Informationen ein Kriterium darstellen, um die Glaubwürdigkeit von Medienberichten zu bewerten. Es gelang ihr aber noch nicht vollständig, dies anzuwenden. Das Zitat impliziert die Annahme, dass man nur die Artikel mit vielen Informationen lesen müsse, um herauszufinden, was passiert sei. Auf der theoretischen Ebene wusste sie, dass das nicht zutrifft, aber sie stolperte bei der Anwendung auf den Zeitungsbericht über die sogenannte Theorie-Praxis-Lücke (Wineburg, 1991; Heuer, Resch & Seidenfuß, 2019). Ähnlich wie den US-amerikanischen scheint den deutschen Studierenden die Ausgewogenheit der Positionen entscheidend für die Glaubwürdigkeit eines Textes zu sein. So sagte Karl: *«So, I think article A and C they (are) the most reliable ones. Because I think they show [...] both (perspectives) and they have a lot of sources.»*¹⁴ Einige Befragte gingen noch einen Schritt weiter und bezogen die Quellenangaben der Zitate in ihre Überlegungen mit ein: *«I mean you can take that quotes and build up your own opinion. So that's the politicians are the original source, or their quotes are the original source and you can try to build up your own opinion. And Reuters I mean, I googled Reuters and I think they're, they're quite a reliable source, because I think they're for financial and political news.»*¹⁵ Patricia, von der diese Aussage stammt, zeigte einen hohen Grad an *Methodenkompetenz*. Sie legte bei der Beurteilung der Glaubwürdigkeit Wert auf eine multiperspektivische Sichtweise. Darüber hinaus sollen die Quellen im Original verwendet werden. Und um die Glaubwürdigkeit, nicht nur des Artikels an sich,

13 Interview mit Meltem (D) vom 09.04.2021, durchgeführt von Carolin Hestler und Franziska Rein.

14 Interview mit Karl (D) vom 09.04.2021, durchgeführt von Carolin Hestler und Franziska Rein.

15 Interview mit Patricia (D) vom 09.04.2021, durchgeführt von Carolin Hestler und Franziska Rein.

sondern auch der zitierten Quellen tiefergehend festzustellen, betrieb sie Hintergrundrecherche. Mit diesem kriteriengeleiteten Ansatz unterschied sich ihre Aussage von der Einschätzung Meltems deutlich. Darüber hinaus achteten die deutschen Studierenden auf Struktur, Intention und Inhalt des untersuchten Artikels. Sie überprüften die Medienberichte auf Lücken in den Argumentationslinien und fanden teilweise triftige Gegenargumente. Dorothee konnte neben der Methoden- auch Elemente der Orientierungskompetenz in der Befragung anwenden. Sie nahm eine kritische, kriteriengeleitete, distanzierte Sicht ein: «[...] it [the article] completely neglects or claims that it was not a coup, it was only a few who resorted to violence and it called the election [...] rigged. It didn't say that there is no proof for the election being stolen. [...The other article is] reliable because I can see that they were paying attention to being transparent. I can see that they were trying to state where they got their information from.»¹⁶ Dorothee zeigte mit Wendungen wie «they were paying attention to being transparent» oder «were trying to state»¹⁷ ein hohes Maß an Distanziertheit und Abstraktionsvermögen; sie verstand die Artikel als Konstruktionen, die sich der Wahrheit allenfalls annähern. Sie lobte das Bemühen um Neutralität, war sich aber bewusst, dass diese nie vollständig erreicht werden kann.

3.3 Intercultural Communication Competence

3.3.1 Theorie

Das Forschungsprojekt ist in einen internationalen und interkulturellen Kontext eingebettet. Alle deutschsprachigen Studierenden erwerben bzw. vertiefen durch ihr Englischstudium, so die Annahme, auch eine Intercultural Communicative Competence, kurz ICC.¹⁸ «Intercultural Communicative Competence (ICC) needs to be incorporated in the language curriculum if educators hope to help students develop an appreciation for the language and culture studied, an awareness of their own culture, and the development of skills that will allow them to be competent, adaptable, communicators» (López-Rocha, 2016, 105). Sandra López-Rocha betont die Relevanz des Kontextes, in dem eine fremde Sprache produziert oder rezipiert wird. Der kulturelle Kontext hat entscheiden-

16 Interview mit Dorothee (D) vom 09. 04. 2021, durchgeführt von Carolin Hestler und Franziska Rein.

17 Interview mit Dorothee (D) vom 09. 04. 2021, durchgeführt von Carolin Hestler und Franziska Rein.

18 Alle deutschsprachigen Studierenden studieren Geschichte und Englisch auf Lehramt.

den Einfluss auf die Bedeutung des Sinngehalts, was oft nachrangig behandelt wird. ICC ist von zentraler Bedeutung, wenn der Kontext in die Bedeutung des Sinngehalts einfließen soll (López-Rocha, 2016, 106). Neben diesen Überlegungen lässt sich ICC mit Blick auf Michael Byrams Kompetenzmodell folgendermaßen zergliedern: 1) *knowledge*, 2) *attitudes*, 3) *(political) education and critical cultural awareness*, 4) *skills of interpreting and relating* und 5) *skills of discovering and interacting* (Byram, 2008). Da es das vornehmliche Ziel der vorliegenden Studie ist, zu erfahren, wie Studierende deutsche und US-amerikanische Medienberichte analysieren und sie vor dem Hintergrund ihrer unterschiedlichen Bedeutungsgehalte interpretieren, erscheint es notwendig, dass die Studierenden in der Lage sind, auch die jeweiligen kulturellen Kontexte in ihre Analysen und Interpretationen einfließen zu lassen und diese zu reflektieren.

3.3.2 Die Ergebnisse der Studie

Die vorliegende Studie zeigt, dass für das Verständnis einer Sprache mehr benötigt wird als lexikalische Kenntnisse. Alle deutschen Studierenden verstanden die Inhalte der vorgelegten Berichte. Um die Medien aber dem räumlichen und politischen Ursprung zuordnen zu können, benötigten sie mitunter ein Bewusstsein für Feinheiten und kulturelle Unterschiede. So erkannte zum Beispiel Maria den Spiegel-Artikel als ein deutsches Medium, da dort der deutsche Außenminister zitiert wird. In einem amerikanischen Bericht über ein amerikanisches Ereignis würde niemals ein deutscher Außenminister als Quelle angegeben werden, so Maria. Ihr fiel auch auf, dass im selben Artikel Uhrzeiten anders angegeben werden als in amerikanischen Artikeln: «[...] *they also use local time [...] often they'll say Eastern standard or Western standard or I think Central time instead of like local time just because the US is set up in two I think four or five different time zones, so as an American (...) I don't think that would be used as much.*»¹⁹ Maria, die Familie in den USA hat und dort teilweise aufgewachsen ist, hat ein tieferes Verständnis kultureller Unterschiede (*knowledge, attitudes, education and critical cultural awareness*). Den anderen deutschen Befragten fielen diese Feinheiten nicht auf.

Ein weiteres Beispiel zeigt, dass vermeintliche Kompetenzen aus dem Bereich des ICC zu sog. *false friends* führen konnten. Alle deutschen Befragten interessierten sich nach eigenen Angaben für die US-amerikanische Kultur und

19 Interview mit Maria (D) vom 09.04.2021, durchgeführt von Carolin Hestler und Franziska Rein.

Sprache. Dass dieses vermeintliche Wissen (*knowledge*) über Kultur bei der Einordnung in die Quadranten hinderlich sein konnte, zeigte die Zuordnung von Artikel B: Drei der deutschen Studierenden begründeten ihre (falsche) Zuordnung zur amerikanischen Seite ähnlich: «*I thought the Reichstag thing as well was so cringy in a way, because I was like that's they always do this, that they say*» und «*Oh, but we're not that bad. Look at what Germany did in World War II. Like this is, this is NOTHING against THAT compared to that.*»²⁰ Meltem, Maria und Patricia bezogen sich auf den hier genannten historischen Bezug zum Reichstagsbrand 1933. Sie verwendeten dies als Hinweis, dass der Artikel amerikanisch sein müsse, denn ihrer Meinung nach käme ein Vergleich der Ereignisse des 6. Januar mit dem Reichstagsbrand einer Relativierung und dadurch einer Verharmlosung nationalsozialistischer Taten gleich. Sie glaubten, dass eine solche Deutung im heutigen Deutschland nicht toleriert werden würde. Tatsächlich ist der Artikel aber deutschen Ursprungs, publiziert von der sog. *Junge Freiheit*, einem dem rechten politischen Spektrum zuzuordnenden Online-Medium. Die befragten Studierenden lasen vermutlich keine Berichte dieser Art. Ein weiterer Grund könnte sein, dass alle fünf Befragten Geschichte studieren und daher für die Verwendung von unangemessenen Vergleichen und nationalsozialistischen Ausdrücken sensibilisiert sind. Sie zeigen zwar *skills of interpreting and relating, political education and critical cultural awareness* sind für sie aber wohl hinderlich für die Meinungsfindung.

Aus der Einordnung in das amerikanische Spektrum schlossen Meltem und Maria auf amerikanische Medien im Allgemeinen. Solche Vergleiche seien keine Einzelfälle. So sagte Maria über amerikanische Medien aus dem konservativen Spektrum: «*They ignore kind of fascist signs in front of them.*»²¹ Erfahrungen, welche die Studierenden in der Vergangenheit exemplarisch machten, führten zu einer Übergeneralisierung und damit zu einem falschen Schluss bezüglich der Einordnung des Artikels. So entstehende Vorurteile behindern *discovering and interacting* und sind damit hinderlich für die Erlangung von ICC.

20 Interview mit Meltem (D) vom 09.04.2021, durchgeführt von Carolin Hestler und Franziska Rein.

21 Interview mit Maria (D) vom 09.04.2021, durchgeführt von Carolin Hestler und Franziska Rein.

4 Fazit

Die Erkenntnisse der Studie sind zunächst kritisch zu diskutieren und ihre Gültigkeit ist zu limitieren. Zum einen ist dies auf das kleine Sample zurückzuführen. Daher besitzen die Aussagen keinen Anspruch auf Repräsentativität, Verallgemeinerbarkeit oder Vergleichbarkeit. Die Forschungsteilnehmer*innen befinden sich in unterschiedlichen (Ausbildungs-)Phasen, was die Erkenntnisse zusätzlich schwächt. Dennoch erscheint der Vergleich legitim, da die Kompetenzen von Lehrkräften im Studium grundgelegt werden. Die Daten wurden ausserdem nicht im Intercoderverfahren ausgewertet, sodass keine Intercoderreliabilität im engeren Sinne gegeben ist. Allerdings wurden die Daten von einer Person ausgewertet und von zwei weiteren Personen geprüft und anschliessend weiterverarbeitet, sodass eine intersubjektive Übereinstimmung über die Zuweisung zu den jeweiligen Codes vorliegt.

Der Umgang mit den Medienberichten fiel auf deutscher bzw. US-amerikanischer Seite sehr unterschiedlich aus. Den deutschen Studierenden fiel die Einordnung der Medienberichte schwerer als den US-amerikanischen (angehenden) Lehrkräften. Grund dafür ist vermutlich weniger ein Mangel an Methodenkompetenz, wie ursprünglich gedacht. Die Analyse ergab, dass es meist inhaltliche und sprachliche Feinheiten wie Zeitangaben und die Konnotationen einzelner Ausdrücke waren, die für die fehlerhafte Zuordnung verantwortlich waren. Im weitesten Sinne lässt sich diese unterschiedliche Einordnung auf Abweichungen im Bereich der sprachlichen Kompetenz bzw. ICC zurückführen. Darüber hinaus könnten die politischen Einstellungen der Befragten dazu führen, dass extreme Positionen nicht als solche wahrgenommen werden, sondern in den Quadranten zentraler positioniert werden. Eine solche Korrelation müsste in einer Folgeuntersuchung analysiert werden.

Unterschiede wurden auch bezüglich der verwendeten Instrumente zur kritischen Analyse der Artikel deutlich. So argumentierten die US-amerikanischen Studierenden emotionaler und erfahrungsbasiert. Die deutschen Studierenden gingen stärker theorie- und kriteriengeleitet vor. Sie wendeten Methoden der Quellenanalyse an, indem sie nach dem Ursprung, der Originalität und Kriterien der Auswahl von Zitaten fragten. Allen war bewusst, dass es sich bei den Medienberichten um Konstruktionen handelte. Es wurde deutlich, dass die deutschen Studierenden Methodenkompetenzen anwendeten, die sie im Laufe ihres Studiums eingeübt hatten. Allerdings wurde auch hier eine Graduierung innerhalb der Befragten deutlich.

Von diesen ersten Ergebnissen ausgehend sind weitere Forschungsarbeiten durchzuführen, um die Hochschullehre daran auszurichten und die Theorie-Praxis-Lücke weiter zu erforschen. Darüber hinaus erscheint es zielführend, Überlegungen unterschiedlicher Kompetenzdebatten aufeinander zu beziehen und füreinander anschlussfähig werden zu lassen. So könnten beispielsweise die geschichtsdidaktisch erschlossene historische Methodenkompetenz und Media Literacy weiter theoretisch aufeinander bezogen werden, um die domänenspezifischen Theoriemodelle zusätzlich zu fundieren. Die Pilotstudie hat neben ersten Erkenntnissen viele Fragen aufgeworfen, die in kommenden Studien vertieft werden sollen. Unter anderem soll die Bedeutung der Sprache in einer Folgestudie der Forschungsgruppe reduziert werden, indem das Vorgehen bei der Analyse von Bildquellen aus dem Nationalsozialismus von Studierenden verschiedener Länder untersucht wird. Darüber hinaus erscheint es mit Blick auf die Ergebnisse der Studie zielführend, wenn angehende Geschichts- und Sozialkundefachkräfte darin geschult werden, eine Sensibilität für sprachliche Bedeutungen zu entwickeln und diese kritisch zu hinterfragen. Zur Lehrkräfteausbildung sollte neben dem historischen Kontextwissen auch das Wissen um die Historisierbarkeit von Begriffen gehören.

Literatur

- Alliance for Media Literate America (2007). *The core principles of media literacy education*. Abgerufen von https://mediaeducationlab.com/sites/default/files/AMLA-Core-Princ-MLE_o.pdf
- Baacke, Dieter (1997). *Medienpädagogik*. Tübingen: Niemeyer.
- Beilner, Helmut (2014). Quellen, schriftliche. In Ulrich Mayer, Hans-Jürgen Pandel, Gerhard Schneider, & Bernd Schönemann (Hrsg.), *Wörterbuch Geschichtsdidaktik* (3. Aufl., S. 162–163). Schwalbach/Ts.: Wochenschau.
- Brückner, Thorsten (2021, 8. Januar). *Das war ein Ausdruck der Ohnmacht und kein «Umsturzversuch»*. Abgerufen von <https://jungfreiheit.de/debatte/kommentar/2021/ausdruck-ohnmacht-umsturzversuch/>
- Byram, Michael (2008). *From foreign language education to education for intercultural citizenship: Essays and reflections*. Clevedon: Multilingual Matters.
- Fandos, Nicholas & Cochrane, Emily (2021, 6. Januar). After pro-trump mob storms capitol, congress confirms Biden's win. *The New York Times*. Abgerufen von <https://www.nytimes.com/2021/01/06/us/politics/congress-gop-subvert-election.html>

- Fenn, Monika & Urban, Stefanie (2021). Das Potsdamer Modell des erweiterten Fachwissens für den schulischen Kontext Geschichte: Explorative Prüfung in einer Delphi-Studie. In Sebastian Barsch & Burghard Barte (Hrsg.), *Motivation – Kognition – Reflexion: Schlaglichter geschichtsdidaktischer Professionsforschung*. Frankfurt a. M.: Wochenschau.
- Gautschi, Peter (2009). *Guter Geschichtsunterricht: Grundlagen, Erkenntnisse, Hinweise*. Schwalbach/Ts.: Wochenschau.
- Heuer, Christian (2014). Gegenwartsbezug. In Ulrich Mayer, Hans-Jürgen Pandel, Gerhard Schneider, & Bernd Schönemann (Hrsg.), *Wörterbuch Geschichtsdidaktik* (3. Aufl., S. 76–77). Schwalbach/Ts.: Wochenschau.
- Heuer, Christian / Resch, Mario & Seidenfuß, Manfred (2017). «What do I have to know to teach history well?»: Knowledge and expertise in History teaching – a proposal. *Yesterday and Today*, 18, 27–41.
- Geschichte unterrichten können?! Forschungen zur Professionalisierung von Geschichtslehrpersonen. *heiEDUCATION Journal. Transdisziplinäre Studien zur Lehrerbildung*, (4), 57–68. <https://doi.org/10.17885/HEIUP.HEIED.2019.4.24020>
- Kerber, Ulf (2016). *Historische Medienbildung als theoretisches Kompetenz-Strukturmodell für eine Integration der Medienbildung in die Fachdidaktik des Faches Geschichte in Baden-Württemberg – ein transdisziplinärer Ansatz* (Dissertation). Pädagogische Hochschule Karlsruhe, Karlsruhe. Abgerufen von <https://nbn-resolving.org/urn:nbn:de:bsz:751-opus4-651>
- Körper, Andreas, Schreiber, Waltraud & Schöner, Alexander (Hrsg.) (2007). *Kompetenzen historischen Denkens: Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik. Kompetenzen* (Bd. 2). Neuried: Ars Una.
- López-Rocha, Sanda (2016). Intercultural communicative competence: creating awareness and promoting skills in the language classroom. In Cecilia Gorla, Oranna Speicher, & Sascha Stollhans (Hrsg.), *Innovative language teaching and learning at university: enhancing participation and collaboration* (S. 105–111). Dublin: Research-publishing.net.
- Mayring, Philipp (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12. Aufl.). Weinheim: Beltz.
- McKay, Hollie (2021, 8. Januar). How Wednesday’s Capitol riot came to fruition and who made it happen. *Fox News*. Abgerufen von <https://www.foxnews.com/politics/how-wednesdays-capitol-riot-come-to-fruition>
- Monte-Sano, Chauncey & Budano, Christopher (2013). Developing and enacting pedagogical content knowledge for teaching history: An exploration of two novice teachers’ growth over three years. *Journal of the Learning Sciences*, 22(2), 171–211.

- Pandel, Hans-Jürgen (2005). *Geschichtsunterricht nach PISA: Kompetenzen, Bildungsstandards und Kerncurricula*. Schwalbach/Ts.: Wochenschau.
- Pandel, Hans-Jürgen (2014). Kompetenz. In Ulrich Mayer, Hans-Jürgen Pandel, Gerhard Schneider, & Bernd Schönemann (Hrsg.), *Wörterbuch Geschichtsdidaktik* (3. Aufl., S. 117–118). Schwalbach/Ts.: Wochenschau.
- Reisman, Abby (2012). Reading like a historian: A document-based history curriculum intervention in urban high schools. *Cognition and Instruction*, 30(1), 86–112.
- Resch, Mario & Heuer, Christian (2019). Fachwissen und geschichtsdidaktisches Wissen und Können bei Lehramtsanwärterinnen und Lehramtsanwärttern zu Beginn der zweiten Ausbildungsphase. In Monika Waldis & Béatrice Ziegler (Hrsg.), *Forschungswerkstatt Geschichtsdidaktik 17: Beiträge zur Tagung «Geschichtsdidaktik empirisch 17»* (S. 107–118). Bern: hep.
- Rutsch, Juliane / Rehm, Markus / Seidenfuß, Manfred / Vogel, Markus & Dörfler, Tobias (2018). Professionalisierungsprozesse angehender Lehrkräfte untersuchen. In Juliane Rutsch, Markus Rehm, Markus Vogel, Manfred Seidenfuß, Tobias Dörfler & Stefan Krauss (Hrsg.), *Effektive Kompetenzdiagnose in der Lehrerbildung: Professionalisierungsprozesse angehender Lehrkräfte untersuchen* (S. 9–25). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-20121-0_1
- Salewski, Melanie (2014). Multiperspektivität. In Ulrich Mayer, Hans-Jürgen Pandel, Gerhard Schneider & Bernd Schönemann (Hrsg.), *Wörterbuch Geschichtsdidaktik* (3. Aufl., S. 143–144). Schwalbach/Ts.: Wochenschau.
- Sauer, Michael (2006). Kompetenzen für den Geschichtsunterricht: Ein pragmatisches Modell als Basis für die Bildungsstandards des Verbandes der Geschichtslehrer. *Informationen für den Geschichts- und Gemeinschaftskundelehrer*, (72), 7–20.
- Scheibe, Cynthia L. & Rogow, Faith. (2011). *The Teacher's Guide to Media Literacy: Critical Thinking in a Multimedia World*. Thousand Oaks: SAGE.
- Schreiber, Waltraud. (2006). Kompetenzbereich historische Methodenkompetenz. In Waltraud Schreiber, Andreas Körber, Bodo v. Borries, Reinhard Krammer, Sibylla Leutner-Ramme, Sylvia Mebus, Alexander Schöner & Béatrice Ziegler (Hrsg.), *Historisches Denken: Ein Kompetenz-Strukturmodell* (Bd. 1, 2. Aufl., S. 194–235). Neuried: Ars Una.
- Schreiber, Waltraud. (2008). Ein Kompetenz-Strukturmodell historischen Denkens. *Zeitschrift für Pädagogik*, 54(2), 198–212.
- Schreiber, Waltraud, Körber, Andreas, Borries, Bodo v., Krammer, Reinhard, Leutner-Ramme, Sibylla, Mebus, Sylvia, Schöner, Alexander & Ziegler, Béatrice (Hrsg.). (2006). *Historisches Denken: Ein Kompetenz-Strukturmodell* (Bd. 1, 2. Aufl.). Neuried: Ars Una.

- Seifried, Jürgen & Wuttke, Eveline (2016). Professionelle Kompetenzen von Lehrkräften: Das Beispiel kognitive Aktivierung. *bwp@*, (Profil 4). Abgerufen von <https://www.bwpat.de/profil-4/beitraege/seifried-wuttke>
- Seixas, Peter (2017). A Model of Historical Thinkin. *Educational Philosophy and Theory*, 49(6), 593–605.
- Shulmann, Lees S. (1986). Those who understand: Knowledge growth in teaching. *Educational Reasearcher*, 15(2), 4–14.
- Sturm aufs Parlament: Polizei und Nationalgarde sichern Kapitol (2021, 7. Januar). *Der Spiegel*. Abgerufen von <https://www.spiegel.de/ausland/washington-sturm-auf-us-kapitol-donald-trump-schickt-nationalgarde-a-7cbfcfd0-7384-40d3-837e-fcoe94e-de522>
- Trautwein, Ulrich / Bertram, Christiane / Borries, Bodo von / Brauch, Nicola / Hirsch, Matthias / Klausmeier, Kathrin (...) Zuckowski, Andreas (2017). *Kompetenzen historischen Denkens erfassen: Konzeption, Operationalisierung und Befunde des Projekts «Historical Thinking – Competencies in History» (HiTCH)*. Münster: Waxmann.
- Treumann, Klaus P. / Meister, Dorothee / Sander, Uwe / Burkatzki, Eckhard / Hagedorn, Jörg / Kämmerer, Manuela (...) Wegener, Claudia (2007). *Medienhandeln Jugendlicher: Mediennutzung und Medienkompetenz; Bielefelder Medienkompetenzmodell*. Wiesbaden: VS Verl. für Sozialwissenschaften.
- Wineburg, Sam (1991). On the reading of historical texts: Notes on the breach between school and academy. *American Educational Research Journal*, 28, 495–519.
- Wineburg, Sam (2001). *Historical thinking and other unnatural acts*. Philadelphia: Temple University Press.

Das Repertory-Grid-Interview als qualitativ empirische Forschungsmethode in der Geschichtsdidaktik

Franziska Rein

1 Einleitung

In der geschichtsdidaktischen Forschung klafft nach wie vor in mancherlei Hinsicht eine Lücke, was empirische Erkenntnisse angeht (Barsch, Degner, Kühberger & Lücke, 2020, 12)¹, obwohl im vergangenen Jahrzehnt eine verstärkte Forschungstätigkeit zu beobachten ist. Die bestehenden Leerstellen haben zur Folge, dass theoretische und pragmatisch orientierte Überlegungen noch nicht konsequent mit Beobachtungen beispielsweise aus der Forschung zu historischer Sinnbildung von Lernenden überprüft werden können. Bei dem überwiegenden Teil der empirischen Forschungsarbeiten zeigt sich zudem eine Orientierung auf den gymnasialen Bildungsbereich (Barsch et al., 2020, 10), was die Verallgemeinerbarkeit ihrer Ergebnisse fraglich erscheinen lässt. Insbesondere Schüler*innen mit dem Anspruch auf sonderpädagogische Förderung fanden in der Vergangenheit noch selten Eingang in geschichtsdidaktische Untersuchungen (Rein, 2021, 21). Seit Kurzem ist ein Umdenken und eine Veränderung in der Forschungslandschaft zu beobachten, sodass erste empirische Zugänge sowie Erkenntnisse für historische Lern- und Lehrprozesse dieser Schüler*innenschaft vorliegen (so beispielsweise Barsch, 2013; 2016b; Degner, 2016; Krause, 2021). Möchte man mehr über diese Lernenden erfahren, so müssen weitere empirische Studien durchgeführt werden. Dafür ist auch das Methodenrepertoire diversitätssensibel zu reflektieren. Der vorliegende Beitrag erschließt Repertory-Grid-Interviews (Fromm, 1995) fachspezifisch aus einer diversitätssensiblen Perspektive. Damit reagiert er auf die Warnung von Wolfgang Hasberg, der befürchtet, dass die Disziplin Gefahr laufe, methodisch dilettantisch (Has-

¹ Barsch et al. (2020) verweisen auf einen Mangel an empirischen Studien über Kinder und Jugendliche mit intellektuellen Beeinträchtigungen.

berg, 2014b) zu operieren, wenn sie Forschungsmethoden ihrer Bezugsdisziplinen ohne weitere Reflexionsprozesse übernehme.

Dazu werden zunächst theoretische Hintergründe und Merkmale von Repertory-Grid-Interviews dargestellt sowie deren Potenziale für partizipative Forschung erschlossen. Anschließend wird die Interviewmethodik aus geschichtsdidaktischer Perspektive erörtert. Wie sich diese Interviews, insbesondere mit Blick auf die kognitiv-sprachliche Diversität der Lernenden, durchführen und auswerten lassen, zeigt ein Fallbeispiel von Rein (2021). Die abschliessenden Überlegungen diskutieren Ressourcen und Barrieren der Methodik aus allgemeiner sowie diversitätssensibler Perspektive und geben einen Ausblick.

2 Das Repertory-Grid-Interview

2.1 Theoretischer Hintergrund und geschichtsdidaktische Bezüge

Grundlegende Idee der Personal Construct Psychology, auf der die Repertory-Grid-Interviewmethodik theoretisch basiert, ist, dass Menschen aktiv ihre Umwelt über das In-Beziehung-Setzen von Aspekten konstruieren und diese in Spannungsfeldern verorten. So erzeugen sie persönliche Sinnkonstruktionen, die es ihnen erlauben, sich zu orientieren, zu handeln und ihre Zukunft besser vorherzusehen (Kelly, 1986, 133). Die Umweltaspekte werden in Bezug auf Ähnlichkeiten und Unterschiede reflektiert, wodurch dichotome Sinnzuweisungen entstehen (Kelly, 1986, 75). Menschen konstruieren beispielsweise andere Menschen mit der Deutung freundlich – unfreundlich oder sympathisch – unsympathisch. Diese Gegensatzpaare wenden sie an, wenn sie Menschen begegnen.

Einige Grundlagen der Personal Construct Psychology sind in ihren Kernideen in geschichtsdidaktischen Diskursen auszumachen (Rein, 2021, 336–338). George Kelly betrachtet jeden Menschen als Wissenschaftler*in bzw. Forscher*in, der*die seine*ihre persönliche Zukunft aufgrund der persönlichen Sinnzuweisungen, der persönlichen Konstrukte, vorhersagt und entsprechend handelt. Die persönlichen Konstrukte sind Ergebnisse von Deutungsprozessen, die sich dezidiert auf die Vergangenheit in Form von Erfahrungen sowie als Erwartungen auf die persönliche Zukunft beziehen, und sie dienen der Orientierung (Kelly, 1955, 12–14). Die Verknüpfung der Zeitebenen und Orientierung spielen in geschichtsdidaktischen Annahmen eine zentrale Rolle – hier lassen

sich Bogen in die geschichtsdidaktische Forschung schlagen (Jeismann, 1997, 42; Hasberg, 2014a, 34).

Die Personal Construct Psychology lässt sich einer konstruktivistischen Erkenntnistheorie zuschreiben (Kelly, 1986, 19). Durch die aktive Konstruktion der Umweltaspekte gelingt es, die eigene Zukunft besser vorherzusagen. Der Mensch passt seine Orientierungs- und Handlungsmöglichkeit ständig an (Kelly, 1986, 19). Die Sichtweisen auf die eigene Wirklichkeit sind also relativ, subjektiv und veränderbar. Bärbel Völkel erschliesst Grundideen des Konstruktivismus systematisch für die Geschichtsdidaktik (Völkel, 2014) und diskutiert, welchen Einfluss er für das Verständnis des Fachs und historische Lernprozesse hat. Saskia Handro bezieht konstruktivistische Überlegungen auf die eigene Lebensgeschichte:

«Mit Blick auf den Geschichtsunterricht erfordert dies, dass das Wissen vom erkennenden Subjekt aus aufgebaut wird, dass die Lebenswirklichkeit und der Erfahrungsbezug der Schüler den Ausgangspunkt historischen Fragens und historischer Reflexion bilden, dass historisches Lernen der Erweiterung des individuellen Erfahrungsspielraums des Lerners um die historische Dimension und damit der Konstruktion der eigenen Lebensgeschichte dient [...]» (Handro, 2014, 120).

Repertory-Grid-Interviews finden genau das heraus: Sie erforschen, wie und welches Wissen aufgebaut wird und wie Inhalte reflektiert werden. Auf beiden Ebenen, der biografisch-historischen und elaboriert-historischen Ebene, tragen die Sinnzuweisungen zur Orientierung bei (Rein, 2021, 138–157). Die Personal Construct Psychology umfasst sowohl implizite, nichtbewusste als auch explizite, bewusste Bedeutungszuweisungen. Mithilfe der darauf basierenden Repertory-Grid-Methodik lässt sich beispielsweise erforschen, welche Effekte nichtbewusste oder / und bewusste Reflexionsprozesse auf die Vergangenheitsdeutung und Sinnzuweisung von Schüler*innen besitzen. Repertory-Grid-Interviews folgen dabei einer bestimmten Vorgehensweise, die im Folgenden dargestellt wird.

2.2 Grundzüge und Merkmale

Bei der Repertory-Grid-Methodik handelt es sich um einen Sammelbegriff für Interviewverfahren, um persönliche Sinnkonstruktionen verschiedenster Umweltaspekte, sog. Elemente, als Spannungsfelder zu erheben und auszuwerten. Ziel der Erhebung und Auswertung ist es, zu erfahren, über welche persönlichen

Sinnkonstruktionen der*die Forschungsteilnehmer*in verfügt und welche Reichweite sie haben (Kelly, 1986, 72). Alle Elemente werden während der Erhebung bei einem Ranking innerhalb von Spannungsfeldern verortet (Riemann, 1987, 94). In der anschliessenden Auswertung werden die Beziehungen zwischen den Umweltaspekten und persönlichen Sinnkonstruktionen interpretiert, analysiert sowie Hypothesen für die Weltsicht bzw. deren Ursachen oder Konsequenzen erzeugt. Während der Erhebung sind die Spannungsfelder und das Ranking zu protokollieren, um in der Auswertung zum einen diese Protokolle zu interpretieren; zum anderen lassen sich die Daten händisch oder mit dem eigens entwickelten Computerprogramm Grid Suite² weiteranalysieren. Nach der Auswertung sind die Ergebnisse gemeinsam mit den Forschungsteilnehmer*innen kommunikativ zu validieren (Scheer, 1993, 25).

Um sicherzustellen, dass tatsächlich die Sinnkonstruktion der Forschungsteilnehmer*innen erhoben wird, sollten alle Spannungsfelder beschrieben und erklärt werden. Das gewährleistet bei der anschliessenden Auswertung, dass die Erkenntnisse möglichst nah bei den subjektiven Theorien der Forschungsteilnehmer*innen liegen und nicht die Perspektiven des oder der Forschenden abbilden. Bei Repertory-Grid-Interviews greifen die Erhebung und Auswertung der Daten ineinander. Darin zeigt die Methodik gegenüber anderen Verfahren einen deutlichen Vorteil: Die subjektiven Theorien werden nicht aufwendig durch Verdichtungen, Interpretationen und Bewertungen aus dem Interviewmaterial herausgefiltert und ausgewertet, wie beispielsweise bei der Qualitativen Inhaltsanalyse (Mayring, 2010). Vielmehr liegen diese Daten als persönliche Sinnkonstruktionen unmittelbar und expliziert durch die Erhebung vor.

Es gibt zahlreiche Möglichkeiten, Repertory-Grid-Interviews zu adaptieren und an heterogene Lernvoraussetzungen von Schüler*innen anzupassen. So ist es beispielsweise denkbar, die subjektiven Sichtweisen mittels vorgegebener oder gemeinsam mit von dem*der Forschungsteilnehmer*in festgelegten Elementen zu erheben. Diese Elemente sind auf unterschiedlichen Aneignungsebenen basal-perzeptiv, konkret-gegenständlich, anschaulich oder abstrakt-begrifflich (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2009, 13) denkbar. Man kann die Skala des Rankings anpassen oder auf sie verzichten und die Elemente ausschliesslich dichotom einordnen lassen. Nach eigenen Erkenntnissen ist es in einem Repertory-Grid-Interview nicht möglich, auf Ver-

2 <http://www.gridsuite.de/45614/home.html>; alle Abbildungen des vorliegenden Beitrags wurden mit dem Computerprogramm GridSuite erzeugt.

balsprache zu verzichten. Es ist aber nicht ausgeschlossen, dass es ohne Verbal-sprache auskommen könnte. Darüber sollte beispielsweise mit Blick auf Schüler*innen mit sog. komplexer Behinderung als Teilnehmer*innen an geschichtsdidaktischen Forschungsprojekten weiter nachgedacht werden.³

2.3 Das Repertory-Grid-Interview in der partizipativen Forschung

Repertory-Grid-Interviews lassen sich in partizipativen Forschungssettings nutzen. Partizipative Forschung meint, dass *mit* den Menschen anstatt *über* sie oder *für* sie geforscht wird (Bergold & Thomas, 2010, 333). Hinter partizipativen Forschungszugängen steckt keine eigenständige Methodik; es handelt sich vielmehr um einen Forschungsstil, der den Anspruch besitzt, praxisrelevant zu sein und mit den Forschungsteilnehmer*innen als Mitforschende gemeinsam zu forschen (Bergold & Thomas, 2010, 333). Diese Herangehensweise erlaubt die aktive Mitarbeit der Forschungsteilnehmenden als Mitforschende, was in der geschichtsdidaktischen Empirie bislang noch wenig Beachtung findet. Waltraud Schreiber et al. erwähnen diesen Zugang zwar (Schreiber, Ziegler & Kühberger, 2019, 336), er ist aber noch nicht breit rezipiert und aufgegriffen worden. Bisher ist noch kein beendetes partizipatives Forschungsvorhaben in der Disziplin bekannt. Es erscheint gewinnbringend, fachspezifische Forschungsmethoden auf ebendiese partizipativen Ressourcen hin zu befragen, um tatsächlich praxisrelevanten Fragen nachzugehen.

Bei allen Teilschritten eines Repertory-Grid-Interviews lässt sich partizipativ vorgehen. Das Festlegen der Fragestellung, die Wahl der Elemente, die Erhebung der Sinnkonstruktionen wie auch die Auswertung kann gemeinsam mit den Mitforschenden durchgeführt werden. Das erscheint für Erkenntnisse über persönliche Sinnzuweisungen besonders aussichtsreich, möchte man doch möglichst nah an die Sichtweisen der jeweiligen Person gelangen und etwas über ihre Perspektive auf die Wirklichkeit erfahren. So liesse sich beispielsweise bei der Erforschung der Sinnkonstruktion von Schüler*innen oder auch Lehrkräften die genaue Fragestellung gemeinsam festlegen. Darüber hinaus kann gemeinsam entschieden werden, welche Elemente für die Erhebung der persönlichen Sinnkonstruktionen gewählt werden. Die Auswahl der Elemente ist wich-

3 Eine detaillierte Erörterung sowie Anwendung der Methodik aus sonderpädagogischer Perspektive ist bei Rein (2021) nachzuvollziehen.

tig und bestimmt den Inhalt des Interviews. Sie lässt sich neben der eigentlichen Auswertung der persönlichen Konstrukte zusätzlich interpretieren. Mit einer Frage zur Zusammenstellung der Elemente begründet der*die Forschungsteilnehmer*in die Auswahl, woran sich das Erheben der persönlichen Konstrukte als Spannungsfelder anschliesst. Diese Äusserungen und persönlichen Konstrukte lassen sich in Bezug setzen und interpretieren sowie durch eine kommunikative Validierung belegen.

3 Das Repertory-Grid-Interview aus geschichtsdidaktischer Perspektive

3.1 Qualitativ empirische geschichtsdidaktische Forschung

Bisher, so Saskia Handro und Bernd Schönemann, zeichnet sich fachspezifische Forschung weniger durch ihre spezifischen Methoden, sondern eher durch die Zugänge, Kategorien und Fragestellungen aus. Ein spezifisches, repräsentatives Methodensystem existiere nicht (Handro & Schönemann, 2002, 4–7). Typischerweise werden in der Disziplin Interviews geführt, um etwas über die Perspektiven der Akteur*innen eines bestimmten Forschungsfelds zu erfahren (Barsch, 2013; Meyer-Hamme, 2016, 2016; Thünemann, 2016; Yildirim, 2017). Dafür gibt es verschiedene Interviewtechniken.⁴ Um Repertory-Grid-Interviews von den bisher gängigen Interviewtechniken abzugrenzen, werden diese im Folgenden kurz dargestellt.

Bei narrativen Interviews strukturieren die Forschungsteilnehmer*innen das Interview und dessen Inhalt; halbstandardisierte Interviews werden sowohl durch die Forschungsteilnehmenden wie auch die Forschenden selbst strukturiert (Köster, 2016, 33). Narrative Interviews verzichten gegenüber leitfadengestützten Interviews auf die Struktur des Leitfadens und agieren offen (Köster, 2016, 39). Sie beziehen sich methodisch wie auch inhaltlich auf Narrativität (Barricelli, 2014) als einen zentralen Aspekt der Disziplin. Diese narrativen Interviews erfordern von der*dem Interviewteilnehmenden komplexe kognitiv-sprachliche Fähigkeiten. Denn das Interview wird massgeblich von dem*der Interviewteilnehmenden strukturiert und ist somit sehr von seinen*ihrer Kom-

4 Der vorliegende Beitrag kann aufgrund der theoretischen Rahmung nur einen groben Überblick über die qualitativ-empirische disziplinspezifische Forschung mit Interviews geben. Eine ausführlichere Darstellung findet sich beispielsweise in Barricelli & Sauer (2015).

petenzen abhängig. Hier liegt, so die Einschätzung der Autorin, für Personen, deren kognitiv-sprachliche Fähigkeiten eingeschränkt sind, beispielsweise Schüler*innen mit Förderanspruch, eine große Hürde.

Halbstandardisierte und leitfadenstrukturierte Interviews erscheinen für geschichtsdidaktische Forschungsvorhaben vielseitig anschlussfähig (Köster, 2016, 35). Sie lassen sich in zahlreiche Typen untergliedern, von denen exemplarisch einige zu erwähnen sind, um anschliessend die Rekonstruktion subjektiver Theorien (Scheele & Groeben, 1988) darzustellen. Um zu ergründen, wie beispielsweise geschichtskulturelle Medien wirken, lassen sich fokussierte Interviews führen (Köster, 2016, 35–36). Mithilfe des problemzentrierten Interviews sind «subjektive Sichtweisen auf gesellschaftlich relevante Probleme» (Köster, 2016, 38) zu erheben. Aus geschichtsdidaktischer Perspektive könnten das etwa kontroverse zeitgeschichtliche oder historische Themen sein. Im episodischen Interview lassen sich z. B. subjektive Theorien (episodisches Wissen) und erlerntes (semantisches) Wissen gegenüberstellen und interpretieren (Köster, 2016, 39). Mithilfe von Expert*inneninterviews kann bestimmtes Erfahrungswissen von Personen, die als Expert*innen für einen bestimmten Realitätsausschnitt gelten, untersucht werden, so etwa Geschichtslehrkräfte in Bezug auf ihr didaktisch-methodisches Fachwissen. Es geht dabei weniger um subjektive Theorien, sondern vielmehr um fachspezifisches Wissen (Köster, 2016, 39). Diese halbstandardisierten Interviews, die mit einem Leitfaden agieren, gelten als gängige Erhebungsmethode zur Erforschung subjektiver Theorien, so auch in der Geschichtsdidaktik (Thünemann, 2016, 44). Bei der Rekonstruktion subjektiver Theorien handelt es sich um ein halbstandardisiertes Interviewverfahren, welches die persönlichen Sinnkonstruktionen durch Fragen und eine anschliessende Strukturlegetechnik erhebt. Durch diese Methode werden die subjektiven Theorien während des Interviews visualisiert und validiert (Scheele & Groeben, 1988, 18–67). Sie wird selten gewählt (Köster, 2016, 37). Holger Thünemann verzichtet bewusst auf diesen Teilschritt bei der Datenerhebung. Er befürchtet, dass der*die Forschende Einfluss nehmen könnte, sodass die Ergebnisse nicht mehr den ursprünglichen Sinnkonstruktionen der Forschungsteilnehmer*innen entsprechen. Diese Gefahr könnte, so die Einschätzung der Autorin, durch ein entsprechendes Vorgehen beim Strukturlegerverfahren minimiert oder gar ausgeschlossen werden. Diese Methodik sei zum einen «eine für geschichtsdidaktische Zugänge enorm fruchtbare Zielperspektive», zum anderen für die Theorie produktiv (Köster, 2016, 37). Sie basiert auf der gleichen theoretischen Grundlage wie die Repertory-Grid-Methodik. Repertory-Grid-Interviews gehen teilstrukturiert vor (Fromm, 2010, 524). Der zentrale

Unterschied zwischen den gängigen Interviewverfahren und teilstrukturierten Repertory-Grid-Interviews besteht in der Erhebung der Daten als persönliche Konstrukte bzw. Spannungsfelder über den Vergleich von Elementen, die anschliessend gerankt werden. Darin ähnelt die Methodik der Strukturlegetechnik.

Da es für die Auswertung (von Interviews) zahlreiche verschiedene Verfahren gibt, beschränkt sich die anschliessende Darstellung exemplarisch auf die sehr gängige Qualitative Inhaltsanalyse von Interviews nach Mayring (2010), die u. a. Manuel Köster (Köster, 2016), Sebastian Barsch (Barsch, 2016a) und Holger Thünemann (Thünemann, 2016) geschichtsdidaktisch erschliessen. Die Qualitative Inhaltsanalyse bezieht sich nicht nur auf die Auswertung von Daten, die mittels Interviews erhoben werden; sie ist auch für die Auswertung anders gewonnener Daten einsetzbar (Köster, 2016, 51). Bei der Qualitativen Inhaltsanalyse geht es vornehmlich darum, die Daten zu strukturieren, um ihnen durch einen kategoriengeleiteten Zugang Erkenntnisse abgewinnen zu können (Mayring, 2016, 114–121). Die Strukturierung und Auswertung der Daten geschieht standardisiert. Diese starre Standardisierung stand in der Vergangenheit immer wieder in der Kritik (Köster, 2016, 51), wobei Sebastian Barsch betont, dass es zahlreiche Variationen und Abwandlungen gibt (Barsch, 2016a, 207) – von einer Standardisierung sei nicht unbedingt zu sprechen. Mithilfe der Qualitativen Inhaltsanalyse lassen sich «subjektive Denkmuster, Argumentationen und Herleitungen» (Barsch, 2016a, 217) untersuchen, was disziplinspezifisch relevant erscheint. Die Repertory-Grid-Methodik unterscheidet sich in der Auswertung von der Qualitativen Inhaltsanalyse, da die subjektiven Theorien nicht erst aus Daten extrahiert werden müssen, sondern als persönliche Konstrukte und durch das Ranking der Elemente unmittelbar vorliegen.

3.2 Das Repertory-Grid-Interview in der geschichtsdidaktischen Forschung

Bodo von Borries (2002, 13) betont die Möglichkeit, sich an der Lehr- und Lernforschung ausserhalb des deutschsprachigen Diskurses zu orientieren. In seiner Darstellung über internationale Forschungsarbeiten findet die Repertory-Grid-Methodik keinen Eingang. Sie ist ein Verfahren, das überwiegend im englischsprachigen Raum und in der psychologischen Forschung Anwendung findet. Ihre Einführung in das disziplinspezifische Forschungsinstrumentarium lässt sich als Antwort auf Bodo von Borries' Vorschlag verstehen, den Radius zu erweitern, um sich methodisch zu orientieren. Gleichzeitig verschafft man sich in der vorwiegend englischsprachigen Psychologie als Nachbardisziplin Anregungen.

Manuel Köster (2016, 37–38) verweist darauf, dass bei halbstandardisierten Interviews oftmals Schwierigkeiten, beispielsweise beim Gebrauch historischer Begriffe, auftreten können. Solche Herausforderungen minimiert die Repertory-Grid-Methodik nach Einschätzung der Autorin, da sie mit Begriffen arbeitet, die von den Interviewteilnehmer*innen selbst erzeugt bzw. benannt werden. Gegenüber der Befürchtung von Holger Thünemann in Bezug auf die Strukturlegetechnik sieht die Autorin in der strukturierten Erhebung der persönlichen Konstrukte und dem anschließenden Ranking der Elemente durch die Repertory-Grid-Methodik gerade den Vorteil, dass der*die Forschende wenig in diesen Prozess eingreifen muss und kann. Denn die Repertory-Grid-Methodik stellt die Interviewteilnehmer*innen als aktive Subjekte in den Fokus aller Betrachtungen, so auch bei der Erhebung der Daten (Fromm & Paschelke, 2010, 11–12). Die Erhebungs- und Auswertungsmethodik ist weniger als andere Verfahren von der Interpretation und Perspektive der Forschenden abhängig. Dieser Zugang erlaubt eine *echte Subjektorientierung*, so die These. Christoph Kühberger (2015, 36) verortet diesen didaktischen Grundsatz zwischen Theorie, Empirie und Pragmatik. Subjektorientierte empirische Forschung stellt, so Christoph Kühberger weiter, ein leitendes Prinzip dar. «Forschungsfragen und -designs werden daher zentral auf die Subjekte des historischen Denkens ausgerichtet, um über wissenschaftliche Instrumentarien Einsichten zumindest in Teilaspekte, welche eine Relevanz für historische Lernprozesse besitzen, zu erlangen.» (Kühberger, 2015, 36) Zudem erlauben Repertory-Grid-Interviews einen mehrdimensionalen Zugang zu den persönlichen Sinnkonstruktionen der Forschungsteilnehmenden. Die Elemente der Interviews, die stellvertretend für die zu untersuchenden Umweltaspekte stehen, können durch unterschiedliche Sinneseindrücke verschiedener ästhetischer Dimensionen angeboten werden, was aktuell diskutiert wird (Barsch & Kühberger, 2020; Degner & Franz, 2020). Die Repertory-Grid-Methodik lässt sich nicht nur verbal und visuell, sondern zusätzlich oder alternativ auch haptisch, olfaktorisch, akustisch-auditiv durchführen und zeigt sich anschlussfähig an jüngste geschichtsdidaktischen Forschungsüberlegungen. Gerade im Hinblick auf Lernende mit heterogenen Fähigkeiten bietet die Methodik zahlreiche Möglichkeiten, ein Forschungsdesign zu entwickeln, das sich gezielt an die Fähigkeiten der Lernenden anpassen lässt. Das Verfahren eignet sich über die dargestellten Möglichkeiten in der schulpraktischen Forschung hinaus auch für andere Gruppen. So lassen sich mithilfe der Repertory-Grid-Methode empiriegestützte Erkenntnisse über Lehramtsstudierende, Geschichtslehrkräfte oder auch Lehramtsanwärter*innen bzw. Refe-

rendar*innen gewinnen, mit deren Hilfe sich die verschiedenen Phasen der Lehrkräfteaus- bzw. -weiterbildung, Studium, Fortbildungen und Ausbildung im Vorbereitungsdienst bzw. Referendariat verbessern lassen.

Es ist ein Anliegen empirischer geschichtsdidaktischer Forschung, Vergangenheitsdeutungen, Gegenwartswahrnehmungen und Zukunftserwartungen z. B. von Schüler*innen zu untersuchen. Damit verfolgt sie das Ziel, Aussagen über zentrale Aspekte des Geschichtsbewusstseins zu gewinnen. Denn Vergangenheitsdeutungen werden spätestens seit den Überlegungen von Karl-Ernst Jaismann (1997) als ein relevantes Konzept des Geschichtsbewusstseins verstanden, die vorbiografische Vergangenheit deutend zu reflektieren. Empirische Forschung darüber blickt mittlerweile auf eine rund dreissigjährige Geschichte zurück und beginnt u. a. mit Studien von Bodo von Borries (1995). Forschungen in diesem Feld fokussieren überwiegend darauf, die subjektive Deutung der Vergangenheit zu untersuchen, um Rückschlüsse auf die Gegenwartswahrnehmung sowie Zukunftserwartungen zu ziehen. Die Repertory-Grid-Methodik wurde bisher erst einmal eingesetzt, um Deutungen als spezifisch historische Sinnbildungsleistung mit Bezug auf die vorbiografische Vergangenheit zu erforschen (Völkel 2007, 2013). Neben dieser Studie ist eine weitere geschichtsdidaktische Untersuchung mit der Repertory-Grid-Methodik als Forschungsdesign bekannt. Hier ist eine erste systematische Einführung der Repertory-Grid-Methodik in die geschichtsdidaktische Forschungslandschaft zu finden (Rein, 2021). Die Ergebnisse dieser Studie mit Schüler*innen mit dem Anspruch auf sonderpädagogische Förderung im Bereich der geistigen Entwicklung verweisen darauf, dass das Verfahren bei Schüler*innen mit besonderen Lernbedürfnissen einsetzbar ist. Allen Forschungsteilnehmer*innen gelingt es im Interview, ihre eigene Biografie zu deuten. Ihre persönlichen Sinnkonstruktionen sind teilweise stabil, teilweise werden sie durch neue Erlebnisse verändert, was die Zukunft der Schüler*innen und ihre Erwartungshaltung beeinflusst. Diese Erkenntnisse sowie die von Bärbel Völkel vorgelegten Veröffentlichungen unterstreichen die Vermutung, dass das Verfahren grundsätzlich für die Erforschung historischer Sinnbildungsprozesse geeignet ist.

3.3 Ein Anwendungsbeispiel

Das Anwendungsbeispiel entstammt der bereits genannten Studie, die Erkenntnisse über die persönliche Sinnkonstruktion gewinnen möchte (Rein, 2021). Sie geht der Fragestellung nach, welche Sichtweisen Schüler*innen, die

im Förderschwerpunkt geistige Entwicklung beschult werden und sich im Übergang von der Schule in das Arbeitsleben befinden, in Bezug auf Arbeitssituationen in der Schule und ihrem Berufsorientierungspraktikum haben. Die Studie zeigt mithilfe der Repertory-Grid-Methodik, dass bei allen untersuchten Schüler*innen differenzierte historische Sinnbildungsprozesse zu beobachten sind (Rein, 2021, 302–331). Für den vorliegenden Beitrag wird exemplarisch und knapp das Vorgehen bei der Erhebung und Auswertung der schulischen Arbeitssituation der Schülerin Lena dargestellt. So lässt sich zeigen, inwiefern sich mithilfe der Forschungsmethodik ihre persönlichen Sinnkonstruktionen zu einer vergangenen biografischen Erfahrung erheben lassen.⁵

3.3.1 Die Auswahl und Präsentation der Elemente

Die Elemente wurden entsprechend der Fragestellung ausgewählt. Für das Forschungsvorhaben wurden Lena zehn Fotografien⁶ von vergangenen Situationen als Quellen bzw. Elemente verschiedener Arbeitssituationen vorgelegt. Ausgehend von der Forschungsfrage sollen die Fotografien die persönlichen Arbeitserfahrungen von Lena abbilden und daher ihrer biografischen Vergangenheit entstammen. Die Elemente sprechen die Ebene der visuellen Wahrnehmung an und verweisen auf konkrete Situationen, die Lena erlebt hat: «Mittagessen Schule», «Pausenhof», «Arbeiten mit Chefin Café», «Arbeitsraum Café», «Pause Schule», «Holz schleifen», «Geschirr spülen», «Kassieren», «Holz ölen» und «Tisch wischen». Auf allen Fotografien (außer Pausenhof und Arbeitsraum Café) ist Lena abgebildet, wie sie einer Tätigkeit nachgeht.⁷ Da sich die Erhebung auf die persönliche Sinnkonstruktion der eigenen biografischen Vergangenheit bezieht, wurden keine historischen Quellen im Sinne von Zeugnissen, welche die eigene Lebenszeit überschreiten, ausgewählt.

3.3.2 Die Erhebung der persönlichen Konstrukte

Im Zuge der Durchsicht der Fotografien fasste Lena ihre Arbeitserfahrungen in Spannungsfeldern zusammen, die sie mit den Situationen verbindet. Folgende Sinnkonstruktionen wurden erhoben: «ausruhen – lernen», «nicht gut – gut», «kein Geschirr – Geschirr», «arbeiten – nicht arbeiten», «nichts essen – essen»,

5 Siehe Rein (2021, 220–236) für eine detaillierte Darstellung und Interpretation der Daten.

6 Aus Datenschutzgründen muss auf die Abbildung der Fotografien verzichtet werden.

7 Da bei den beiden Elementen Pausenhof und Arbeitsraum Café der Raum und nicht die Tätigkeit im Vordergrund stehen soll, wird hier darauf verzichtet, Lena auf der Fotografie abzubilden.

«sich gut fühlen – sich schlecht fühlen», «konzentriert – nicht konzentriert», «ausruhen – laut», «entspannt – nicht entspannt».

3.3.3 Das Ranking der Elemente an den persönlichen Konstrukten

Das Ranking der Elemente ist genuiner Bestandteil von Repertory-Grid-Interviews (u. a. Fromm, 1995). Lena lag eine fünfstufige Skala⁸ vor, mit deren Hilfe sie die Elemente in den Rahmen der persönlichen Sinnkonstruktionen einordnete. Diese Rankingskala soll abbilden, wie Lena die einzelnen Quellen im Spannungsfeld der jeweiligen Sinnkonstruktion verortet. Damit werden Informationen über die einzelnen Quellen gesammelt, die sich anschliessend zueinander in Beziehung setzen lassen. Durch die fünf Stufen der Skala hat Lena die Möglichkeit, eine klare Zuweisung vorzunehmen, Abstufungen zu nutzen oder die Quellen neutral einzuordnen. Die einzelnen Pole der Spannungsfelder werden jeweils oben und unten an der Skala angelegt und jedes einzelne Element auf der Skala eingeordnet. Die Zuweisung der Elemente zur entsprechenden Rankingstufe wurde während der Erhebung protokolliert. Dem Prinzip des Rankings entsprechend ordnete Lena jedes Element anhand der Spannungsfelder ein und wies so jeder Situation einen spezifischen Sinn zu. Die folgende Skala macht dies anschaulich.

sehr	ein bisschen	mittel	eigentlich nicht	gar nicht

Abb. 1: Mögliche Rankingskala (eigene Darstellung)

3.3.4 Die Auswertung der Ergebnisse mit GridSuite

Für die Auswertung der Ergebnisse wurde auf das Computerprogramm GridSuite zurückgegriffen. Ziel der Auswertung war es, die persönlichen Sinnkonstruktionen nachzuvollziehen und Hypothesen für ihr Zustandekommen oder ihre Konsequenzen für die Orientierung von Lena in zukünftigen Arbeitssituationen zu entwickeln. Mithilfe von GridSuite lassen sich verschiedene Darstellungen erzeugen, die es erlauben, Sinnzusammenhänge herzustellen und zu interpretieren, welche möglicherweise ohne die Aufbereitung der Daten und deren Visualisierung unerkannt bleiben würden. Exemplarisch werden im Folgenden zwei Darstellungen abgebildet, die Teil der Auswertung waren.

8 Die Rankingskala war zusätzlich mit Piktogrammen visualisiert, die aus Lizenzgründen nicht veröffentlicht werden können.

1	Mittagessen Schule	Pausenhof	Pause Schule	Kassieren	Arbeiten mit Chefin	Schleifen	Geschirr spülen	Tisch wischen	Arbeitsraum Caffé	Holz ölen	5
ausruhen											lernen
nicht gut											gut
kein Geschirr											Geschirr
arbeiten											nicht arbeiten
nichts essen											Essen
sich gut fühlen											sich schlecht fühlen
konzentriert											nicht konzentriert
ausruhen											laut
entspannt											nicht entspannt

Abb. 2: Elemente oben, persönliche Konstrukte links und rechts als Spannungsfelder

PCA (Varimax) Arbeit Schule:: Lena
 Achsen: -1.01 to +1.01

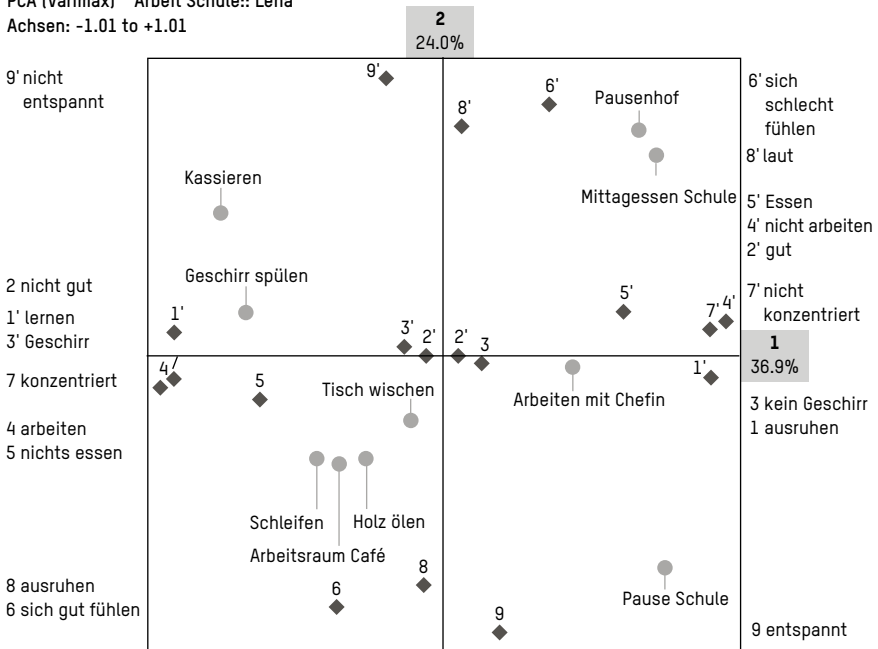


Abb. 3: Elemente sowie persönliche Konstrukte als Konstruktpole verteilt in den Quadranten

Mithilfe dieser Darstellungen liessen sich die Elemente und Sinnkonstruktionen in Bezug zueinander setzen und Hypothesen für die Bezüge bzw. deren Einfluss auf zukünftige Bedeutungszuweisungen anhand der Sinnzuweisungen ableiten. Die Ergebnisse der Auswertung wurden in einer anschliessenden kommunikativen Validierung mit Lena besprochen und daraufhin geprüft, ob sie für Lena zutreffen.⁹

4 Ressourcen und Barrieren von Repertory-Grid-Interviews für empirische geschichtsdidaktische Forschung

Anhand des Fallbeispiels konnte gezeigt werden, dass sich Repertory-Grid-Interviews eignen, um Sinnbildungsergebnisse über die eigene biografische Vergangenheit zu erheben und auszuwerten. Die Vergangenheitsdeutungen beziehen sich auf die eigene Lebenszeit der Schülerin Lena und gehen damit über den geschichtsdidaktischen theoretischen wie auch empirischen Diskurs über lebenszeitüberschreitende Vergangenheitsdeutungen hinaus. Damit wird das geschichtsdidaktische Konzept der Vergangenheitsdeutung um die Deutung biografischer Erfahrungen erweitert, die mit der Repertory-Grid-Methodik untersucht werden kann.

Ein besonderer Vorteil der Repertory-Grid-Methodik liegt darin, dass die Datenauswertung bereits dort beginnt, wo bei anderen Verfahren noch die Datenerhebung steht. Die Erhebung und deren Auswertung greifen also ineinander. Durch die spezifische Methodik der Erhebung der persönlichen Sinnkonstruktionen ist ein Interpretations- bzw. Analyseschritt, wie er beispielsweise in der Qualitativen Inhaltsanalyse vorgesehen ist, überflüssig. Mit der Methodik gelingt es, unmittelbar und ohne Umwege an nichtbewusste oder bewusste Sinnkonstruktionen zu gelangen. Die Erhebung ist somit, wie bereits dargestellt, weniger von den jeweiligen Sichtweisen und Interpretationen des*der Forscher*in abhängig als andere Verfahren. Hier verbirgt sich ein enormes Potenzial der Methodik, das insbesondere in Bezug auf qualitative Forschungsvorhaben nicht hoch genug einzuschätzen ist. Mit der Repertory-Grid-Methodik gelangt man direkt an die subjektiven Theorien der Forschungsteilnehmer*innen und nicht an die des oder der Forschenden. Ausserdem sind die verschiedenen Vorgehensweisen bei der Auswertung mit-

9 Für eine detaillierte Auswertung der Daten siehe Rein (2021, 220–236).

einander kombinierbar. Man kann die Protokolle aus dem Interview auswerten, die Daten händisch strukturieren oder mit GridSuite arbeiten. Die Erkenntnisse verschiedener Auswertungsmethoden ergänzen sich gegenseitig bei der Interpretation und lassen sich somit validieren. Die Repertory-Grid-Methodik zeigt sich dadurch gegenüber anderen Erhebungsmethoden recht ökonomisch und erzeugt belastbare Daten. Das ist ein Vorteil, wenn es darum geht, schonend und bewusst mit Ressourcen in der Forschung umzugehen.

Die kommunikative Validierung (Rein, 2021, 301–302) nach der Auswertung der Daten sichert die konsequente Orientierung an den Forschungsteilnehmer*innen und an ihren persönlichen Sichtweisen auf die Welt. Zudem profitiert die Forschungsmethodik davon, sich vielfältig in Abhängigkeit von der Fragestellung und den Lernvoraussetzungen der Interviewteilnehmer*innen adaptieren zu lassen. So können die Forschenden auf unterschiedliche Aneignungsniveaus eingehen. Gerade durch Visualisierungen (der Elemente und der Rankingskala) wird die Methodik anschaulich und unterstützt die Forschungsteilnehmer*innen bei der Erhebung. Diese Gedanken sind insbesondere aus einer diversitätssensiblen Perspektive sehr bedeutsam. Im geschichtsdidaktischen Bereich erscheint die Datenerhebung anhand von Visualisierungen besonders anschlussfähig, da elaborierte historische Sinnbildung grundsätzlich über die Beschäftigung mit historischen Quellen, Darstellungen oder anderen Medien geschieht (Pandel, 2014, 177). Ebendiese Quellen oder Darstellungen können Gegenstand und Ausgangspunkt eines Repertory-Grid-Interviews sein. Die Möglichkeit, auch andere Sinneskanäle in der Erhebung anzusprechen und zu Gegenständen der Erhebung zu machen, erscheint aus diversitätssensibler Sicht ebenfalls relevant. Darüber hinaus lässt sich die Rankingskala in ihrer Komplexität anpassen. Man kann zwischen einer dichotomen Einordnung der Elemente oder einer sehr fein abgestuften Rankingskala wählen. So kann das Ranking an die jeweilige Forschungsfrage bzw. das Erkenntnisinteresse und die Fähigkeiten der Forschungsteilnehmer*innen angepasst werden. Weiter kann die Repertory-Grid-Methodik, wie gezeigt, partizipativ eingesetzt werden und praxisrelevanten Forschungsfragen nachgehen. Darüber hinaus erscheint die Verschränkung der Theorie und Empirie durch die Repertory-Grid-Methodik von besonderem Interesse, da sich zentrale Ideen der Personal Construct Psychology und geschichtsdidaktische Grundannahmen dezidiert aufeinander beziehen lassen. Es stellt ein Qualitätsmerkmal dar, mit einer empirischen Forschungsmethodik zu arbeiten, die sich theoretisch in der Disziplin fundieren lässt, so die These.

Bei all den Ressourcen der Methodik zeigen sich aber auch Barrieren. So lassen sich mittels Repertory-Grid-Interviews kaum Erkenntnisse über subjektive Sichtweisen auf Umweltaspekte gewinnen, die nicht als Elemente in der Erhebung abzubilden sind. Hier besitzt die Forschungsmethodik beispielsweise gegenüber narrativen Interviews den Nachteil, dass sie nur einen begrenzten Ausschnitt der Deutungsweisen der Forschungsteilnehmer*innen erforscht. Andere Deutungen, die möglicherweise ebenfalls relevant und aussagekräftig für die Forschungsfrage bzw. das Erkenntnisinteresse wären, bleiben so unter Umständen im Verborgenen. Eine weitere Grenze der Methodik liegt in der formal festgelegten Struktur der Erhebung. Die Methodik muss von den Forschungsteilnehmer*innen nach dem vorgestellten Schema angewendet werden. Ansonsten besitzen die Daten keine oder nur eingeschränkte Aussagekraft. Es ist vorstellbar, diese starre Struktur aufzuweichen und die persönlichen Sinnkonstruktionen eher in einem Gespräch zu erheben und anschliessend abzuleiten. Dann verliert die Methodik aber ihren Vorteil gegenüber anderen Verfahren, unmittelbar und ökonomisch an die persönlichen Sinnzuweisungen zu gelangen. Auch ist bislang noch nicht erörtert, ob und inwiefern die Methodik ohne analoge Lautsprache einsetzbar ist. Vor dem Hintergrund diversitätssensibler Forschung erscheint es notwendig, dass auch Schüler*innen, die sich nicht analog sprechend äußern (können), an geschichtsdidaktischen Forschungsvorhaben teilhaben können. Für die Repertory-Grid-Methodik ist es eine Herausforderung, Daten ohne Lautsprache zu erheben. Dies ist zwar beispielsweise durch ein nonverbales Gruppieren der Elemente zur Darstellung der persönlichen Konstrukte oder nonverbales Einordnen der Elemente anhand einer Rankingskala vorstellbar, allerdings müsste die Aussagekraft einer solchen Datenerhebung näher erörtert werden. Der Einsatz der Methode bei Schüler*innen mit komplexer Behinderung ist also weiter zu diskutieren.

5 Fazit und Ausblick

Ausgehend von der vorgestellten empirischen Untersuchung lässt sich resümieren, dass sich Repertory-Grid-Interviews für qualitative empirische geschichtsdidaktische Forschungsansätze eignen, da sie unmittelbar und umstandslos die persönlichen Sinnkonstruktionen der Forschungsteilnehmer*innen erheben und damit weitgehend auf Interpretationen der Forschenden verzichten. Sie liefern ökonomisch belastbare Daten und können somit das disziplinspezifische Methodenrepertoire erweitern. Dieses Urteil bezieht sich auf die Erforschung

eigener biografischer Vergangenheitsdeutungen einer Schülerin. Das schränkt diese Aussage zunächst ein, da zu wenige Untersuchungen mit einem ähnlichen Forschungsdesign vorliegen, welche vorbiografische Vergangenheitsdeutungen betrachten. Mit Blick auf die beiden Veröffentlichungen von Bärbel Völkel lässt sich zum einen vermuten, dass sich die Methode dafür ebenfalls eignet (Völkel 2007, 2013). Vorstellbar ist beispielsweise, dass historische Quellen, welche aus der vorbiografischen Lebenszeit entstammen, als Elemente eingesetzt werden. Zum anderen wäre zu fragen, ob sich die Methodik zur Erforschung von Geschichtskultur eignen würde, wenn Produkte der Geschichtskultur als Elemente fungieren. Um Aussagen darüber treffen zu können, sind weitere Studien und Methodenreflexionen nötig.

Das Repertory-Grid-Verfahren zur Datenerhebung und -auswertung kann in Abhängigkeit von der Fragestellung und den Lernvoraussetzungen der Forschungsteilnehmer*innen adaptiert werden und erscheint aus diversitätssensibler Perspektive für disziplinspezifische Forschungsvorhaben vielversprechend. Mithilfe der Repertory-Grid-Methodik können Erkenntnisse über die Sinnkonstruktion von Schüler*innen gewonnen werden, die bislang noch selten im Fokus geschichtsdidaktischer Forschungsprojekte standen. Theoretische und pragmatische Überlegungen lassen sich mithilfe der Forschungsmethodik empirisch überprüfen, was Rückschlüsse auf die Gültigkeit der Theoriekonzepte und die Durchführung von Lern- und Lernprozessen in heterogenen Schüler*innengruppen zulässt. Bisher wurde noch nicht erörtert, ob die Forschungsmethodik für die Datenerhebung mit Schüler*innen, die sich nicht analog lautsprachlich äußern (können), geeignet ist. Sollte die Forschungsmethodik nicht dafür einsetzbar sein, würde das ihre Reichweite schmälern.

Die Methodik lässt durch die bestimmte Interviewstruktur nur die Erhebung der Sinnzuweisungen von zuvor festgelegten Inhalten zu; die Auswahl der Elemente legt die inhaltliche Rahmung der Interviews fest und begrenzt diese. Hier eröffnen andere Interviewverfahren mehr Freiräume und einen breiteren Erkenntnisgewinn. Repertory-Grid-Interviews lassen sich partizipativ verwenden, was die Teilhabe der Forschungsteilnehmer*innen als Mitforschende sowie praxisrelevante Erkenntnisse sichert. Dadurch können Fragestellungen gefunden und Einsichten gewonnen werden, was ohne die Mitforschenden nicht gelingen würde, so die Vermutung. Das hält nach Ansicht der Autorin für empirische Forschungsarbeiten der Disziplin neue Perspektiven und Chancen bereit.

Literatur

- Barricelli, Michele (2014). Narrativität. In Ulrich Mayer, Hans-Jürgen Pandel, Gerhard Schneider & Bernd Schönemann (Hrsg.), *Wörterbuch Geschichtsdidaktik* (3. Aufl., S. 149–150). Schwalbach/Ts.: Wochenschau.
- Barricelli, Michele & Sauer, Michael (2015). Empirische Lehr-Lern-Forschung im Fach Geschichte. In Georg Weißeno (Hrsg.), *Empirische Forschung in gesellschaftswissenschaftlichen Fachdidaktik: Ergebnisse und Perspektiven* (S. 185–200). Wiesbaden: Springer Fachmedien.
- Barsch, Sebastian (2013). Historische Imagination von Schülerinnen und Schülern mit «Lernbehinderung». In Hodel, Jan, Waldis, Monika, Ziegler, Béatrice (Hrsg.), *Forschungswerkstatt Geschichtsdidaktik 12: Beiträge zur Tagung «geschichtsdidaktik empirisch 12»* (S. 95–107). Bern: hep.
- Barsch, Sebastian (2016a). Die Qualitative Inhaltsanalyse als Methode der geschichtsdidaktischen Forschung. In Holger Thünemann & Meik Zülsdorf-Kersting (Hrsg.), *Methoden geschichtsdidaktischer Unterrichtsforschung* (1. Aufl., S. 206–228). Schwalbach/Ts.: Wochenschau.
- Barsch, Sebastian. (2016b). Schülerinnen und Schüler mit dem Förderschwerpunkt Lernen denken historisch: Einblicke in empirische Forschung. In Bettina Alavi & Martin Lücke (Hrsg.), *Geschichtsunterricht ohne Verlierer!?: Inklusion als Herausforderung für die Geschichtsdidaktik* (S. 71–84). Schwalbach/Ts.: Wochenschau.
- Barsch, Sebastian/Degner, Bettina/Kühberger, Christoph & Lücke, Martin. (2020). Einleitung – Diversität im Geschichtsunterricht – Zugänge zu einer inklusiven Geschichtsdidaktik. In Sebastian Barsch, Bettina Degner & Christoph Kühberger (Hrsg.), *Handbuch Diversität im Geschichtsunterricht: Inklusive Geschichtsdidaktik* (S. 9–24). Schwalbach/Ts.: Wochenschau.
- Barsch, Sebastian & Kühberger, Christoph (2020). Mit allen Sinnen Lernen? Zur Vielseitigkeit des historischen Lernens. In Sebastian Barsch, Bettina Degner & Christoph Kühberger (Hrsg.), *Handbuch Diversität im Geschichtsunterricht: Inklusive Geschichtsdidaktik* (S. 385–404). Schwalbach/Ts.: Wochenschau.
- Bergold, Jarg & Thomas, Stefan (2010). Partizipative Forschung. In Günter Mey & Katja Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 333–344). Wiesbaden: Springer Fachmedien.
- Borries, Bodo von (1995). *Das Geschichtsbewusstsein Jugendlicher: Erste repräsentative Untersuchung über Vergangenheitsdeutungen, Gegenwartswahrnehmungen und Zukunftserwartungen von Schülerinnen und Schülern in Ost- und Westdeutschland*. Weinheim u. a.: Juventa.

- Borries, Bodo von (2002). Lehr-/Lernforschung in europäischen Nachbarländern – ein Stimulus für die deutschsprachige Geschichtsdidaktik?! In Saskia Handro & Bernd Schönemann (Hrsg.), *Methoden geschichtsdidaktischer Forschung* (S. 13–49). Münster: LIT.
- Degner, Bettina (2016). Narrative Kompetenz im inklusiven Geschichtsunterricht?!. Ein Unterrichtsversuch. In Bettina Alavi & Martin Lücke (Hrsg.), *Geschichtsunterricht ohne Verlierer!?: Inklusion als Herausforderung für die Geschichtsdidaktik* (S. 85–100). Schwalbach/Ts.: Wochenschau.
- Degner, Bettina & Franz, Eva-Kristina (2020). Ästhetik und historisches Lernen. In Sebastian Barsch, Bettina Degner & Christoph Kühberger (Hrsg.), *Handbuch Diversität im Geschichtsunterricht: Inklusive Geschichtsdidaktik* (S. 82–92). Schwalbach/Ts.: Wochenschau.
- Fromm, Martin (1995). *Repertory Grid Methodik: Ein Lehrbuch*. Weinheim: Dt. Studien-Verl.
- Fromm, Martin (2010). Grid-Methodik. In Günter Mey & Katja Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 524–537). Wiesbaden: Springer Fachmedien.
- Fromm, Martin & Paschelke, Sarah (2010). *GridPractice: Anleitung zur Durchführung und Auswertung von Grid-Interviews*. Norderstedt: Books on Demand.
- Handro, Saskia (2014). Konstruktivismus. In Ulrich Mayer, Hans-Jürgen Pandel, Gerhard Schneider, & Bernd Schönemann (Hrsg.), *Wörterbuch Geschichtsdidaktik* (3. Aufl., S. 119–120). Schwalbach/Ts.: Wochenschau.
- Handro, Saskia & Schönemann, Bernd (2002). Zur Einleitung. In Saskia Handro & Bernd Schönemann (Hrsg.), *Methoden geschichtsdidaktischer Forschung* (S. 3–10). Münster: LIT.
- Hasberg, Wolfgang (2014a). Historisches Lernen für alle. In Sebastian Barsch & Wolfgang Hasberg (Hrsg.), *Inklusiv – exklusiv: Historisches Lernen für alle* (S. 11–39). Schwalbach/Ts.: Wochenschau.
- Hasberg, Wolfgang (2014b). Unde venis?: Betrachtungen zur Zukunft der Geschichtsdidaktik. In Tobias Arand & Manfred Seidenfuß (Hrsg.), *Neue Wege – neue Themen – neue Methoden?: Ein Querschnitt aus der geschichtsdidaktischen Forschung des wissenschaftlichen Nachwuchses* (1. Aufl., S. 15–62). Göttingen: V & R Unipress.
- Jeismann, K.-E. (1997). Geschichtsbewusstsein – Theorie. In Klaus Bergmann, Klaus Fröhlich, Annette Kuhn, Jörn Rüsen & Gerhard Schneider (Hrsg.), *Handbuch der Geschichtsdidaktik* (5. Aufl., S. 42–44). Seelze-Velber: Kallmeyer'sche.
- Kelly, George (1955). *The Psychology of Personal Constructs*. New York: Norton.
- Kelly, George (1986). *Die Psychologie der persönlichen Konstrukte*. Paderborn: Junfermann.

- Köster, Manuel (2016). Methoden empirischer Sozialforschung aus geschichtsdidaktischer Perspektive. Einleitung und Systematisierung. In Holger Thünemann & Meik Zülsdorf-Kersting (Hrsg.), *Methoden geschichtsdidaktischer Unterrichtsforschung* (1. Aufl., S. 9–62). Schwalbach/Ts.: Wochenschau.
- Krause, Susan (2021). Blinde Zugänge zur Vergangenheit?!: Untersuchung zur historischen Imagination blinder Schüler*innen als Beitrag inklusiver Geschichtsdidaktik. In Oliver Musenberg, Raphael Koßmann, Marc Ruhlandt, Kristina Schmidt & Seda Uslu (Hrsg.), *Historische Bildung inklusiv: Zur Rekonstruktion, Vermittlung und Aneignung vielfältiger Vergangenheiten* (S. 237–254). Bielefeld: transcript.
- Kühberger, Christoph (2015). Subjektorientierte Geschichtsdidaktik: Eine Annäherung zwischen Theorie, Empirie und Pragmatik. In Heinrich Ammerer, Thomas Hellmuth & Christoph Kühberger (Hrsg.), *Subjektorientierte Geschichtsdidaktik* (S. 13–47). Schwalbach: Wochenschau.
- Mayring, Philipp (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (11. Aufl.). Weinheim: Beltz. Abgerufen von http://www.content-select.com/index.php?id=bib_view&ean=9783407291424.
- Mayring, Philipp (2016). *Einführung in die qualitative Sozialforschung: Eine Anleitung zu qualitativem Denken* (6. Aufl.). Weinheim u. a.: Beltz. Abgerufen von http://content-select.com/index.php?id=bib_view&ean=9783407294524.
- Meyer-Hamme, Johannes (2016). «Ja also, das war 'ne gute Stunde»: Qualitätsmerkmale von Geschichtsunterricht aus Schülerperspektive. In Johannes Meyer-Hamme, Holger Thünemann & Meik Zülsdorf-Kersting (Hrsg.), *Was heißt guter Geschichtsunterricht? Perspektiven im Vergleich* (2. Aufl., S. 24–40). Schwalbach/Ts.: Wochenschau.
- Ministerium für Kultus, Jugend und Sport Baden-Württemberg (2009). *Bildungsplan Schule für Geistigbehinderte*. Abgerufen von <https://km-bw.de/Lde/Startseite>.
- Pandel, Hans-Jürgen (2014). Sinnbildung. In Ulrich Mayer, Hans-Jürgen Pandel, Gerhard Schneider & Bernd Schönemann (Hrsg.), *Wörterbuch Geschichtsdidaktik* (3. Aufl., S. 177). Schwalbach/Ts.: Wochenschau.
- Rein, Franziska (2021). *Historisches Lernen im Förderschwerpunkt geistige Entwicklung: Eine Studie zur Sinnbildung durch die eigene Lebensgeschichte* (Dissertation). Göttingen: V & R unipress.
- Riemann, Rainer (1987). *Struktur und Organisation persönlicher Konstrukte*. Regensburg: Roderer.
- Scheele, Brigitte & Groeben, Norbert (1988). *Dialog-Konsens-Methoden zur Rekonstruktion Subjektiver Theorien: Die Heidelberger Struktur-Lege-Technik (SLT), konsensuale Ziel-Mittel-Argumentation und kommunikative Flussdiagramm-Beschreibung von Handlungen*. Tübingen: Francke.

- Scheer, Jörn W. (1993). Planung und Durchführung von Repertory Grid-Untersuchungen. In Jörn W. Scheer & Ana Catina (Hrsg.), *Einführung in die Repertory Grid Technik: Band 1. Grundlagen und Methoden* (S. 24–40).
- Schreiber, Waltraud / Ziegler, Béatrice & Kühberger, Christoph (2019). Der Versuch einer Bilanz und einige Gedanken für die Zukunft. 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 15.–17. November 2017* (S. 333–336). Münster: Waxmann.
- Thünemann, Holger (2016). «Ganz, ganz historisch gedacht»: Merkmale guten Geschichtsunterrichts aus Lehrerperspektive. In Johannes Meyer-Hamme, Holger Thünemann & Meik Zülsdorf-Kersting (Hrsg.), *Was heißt guter Geschichtsunterricht? Perspektiven im Vergleich* (2. Aufl., S. 41–56). Schwalbach/Ts.: Wochenschau.
- Völkel, Bärbel (2007). «So einen Mann wünsche ich mir!» – Zur subjektiven Verarbeitung eines historischen Inhalts. In Judith Martin & Christoph Hamann (Hrsg.), *Geschichte, Friedensgeschichte, Lebensgeschichte: Festschrift für Peter Schulz-Hageleit* (S. 31–44). Herbolzheim: Centaurus.
- Völkel, Bärbel (2013). Krisenmanagement über den «Umweg» Geschichte – Wie Jungen anschlussfähige Traditionen in die Geschichte hinein konstruieren. In Martin Lücke (Hrsg.), *Helden in der Krise: Didaktische Blicke auf die Geschichte der Männlichkeiten* (S. 31–65). Berlin: LIT.
- Völkel, Bärbel (2014). *Wie kann man Geschichte lehren? Die Bedeutung des Konstruktivismus für die Geschichtsdidaktik* (3. Aufl.). Schwalbach/Ts.: Wochenschau.
- Yildirim, Lale (2017). *Der Diasporakomplex* (Dissertation). Bielefeld: transcript.

Den Umgang mit Quellen und Darstellungen digital beforschen

Eine Pilotstudie

Jan Scheller, Jonas Schobinger, Martin Nitsche

1 Einleitung

In der Geschichtsdidaktik besteht weitgehend ein Konsens darüber, dass im Geschichtsunterricht historisches Denken gelernt werden soll. Auch wenn die Konzeptionen für «historisches Denken» (Bracke et al., 2018; Körber, Schreiber & Schöner, 2007), «narrative Kompetenz» (Barricelli, 2005; Nitsche & Waldis, 2016; van Norden, 2011), *historical understanding* (VanSledright, 2014) oder *historical reasoning* (van Drie & van Boxtel, 2008) Unterschiede aufweisen, wird in allen der Umgang mit Quellen und Darstellungen als wesentlicher Bestandteil historischen Denkens hervorgehoben, um z. B. nach Antworten auf historische Fragen zu suchen. Die theoretisch sowie sozioempirisch vorgenommenen Operationalisierungen dieser Denkoperation differieren allerdings beträchtlich. Zusätzlich wird auf der theoretischen Ebene betont, dass die Operationen oft nur heuristisch zu trennen seien (Thünemann & Jansen, 2018). Auch sozioempirisch deutet sich bisher an, dass sich historische Kompetenzen nicht trennscharf darstellen lassen (Trautwein et al., 2017). Daher lässt sich schlussfolgern, dass das trennscharfe Identifizieren von historischen Operationen herausfordernd ist. Zudem liegen bisher kaum deutschsprachige Studien vor, in denen historisches Denken und dessen Teilaspekte im Allgemeinen oder der Umgang mit historischen Medien¹ im Spezifischen qualitativ differenziert werden konnten.

Das vom Schweizer Nationalfonds (SNF) geförderte Projekt «Research of Learning Processes in History (RiCH)» (2020–2024) nimmt sich dieser beiden Herausforderungen an und fragt ausgehend vom Konzept der narrativen Kompetenz danach, wie dessen Teilaspekte operationalisiert und qualitativ unter-

¹ Historische Medien werden im Folgenden als Oberbegriff für Quellen und Darstellungen verstanden (Günther-Arndt, 2015).

schieden werden können.² Dazu werden Personengruppen unterschiedlicher historischer Expertise mit historischen Medien konfrontiert sowie aufgefordert, eine historische Frage zu entwickeln und diese medienbezogen zu beantworten.

Im Beitrag werden die Ergebnisse der Pilotstudie hinsichtlich des Umgangs mit Medien vorgestellt. Konkrete Forschungsfragen sind, 1.) welche Operationen des historischen Denkens beim Umgang mit historischen Medien anhand von Protokollen Lauten Denkens identifiziert werden können und 2.) inwiefern sich Unterschiede in den operationalen Ausprägungen erkennen lassen. Dafür wurden Historiker*innen, geschichtsinteressierte Lai*innen, Geschichtsstudierende und Schüler*innen auf Klassenstufe neun mit Medien zur Schweizer Neutralität im 1. Weltkrieg konfrontiert sowie dazu aufgefordert, eine historische Frage zu entwickeln und die Materialien diesbezüglich zu untersuchen. Die Datenauswertung erfolgte gemäss der strukturierenden qualitativen Inhaltsanalyse nach Philipp Mayring (2010).

2 Theoretischer Rahmen

Als Heuristik der Analysen dient ein bereits vorliegendes Modell narrativer Kompetenz, in dem das Konstrukt «als zentrales menschliches Vermögen historischen Denkens» aufgefasst wird (Nitsche & Gollin, 2020, 320). Darin gilt historisches Denken als Interaktionsprozess zwischen historischen Operationen (s. u.) unter Einbezug historischen Wissens erster (z. B. Ereignisse) und zweiter Ordnung (z. B. Dauer und Wandel) sowie subjektiver Konstrukte wie geschichtstheoretischen Beliefs oder (situationalem) Interesse. Ziel dieses Vorgangs sei es demnach unter anderem, historische Narrationen diverser Gattungen (z. B. Erzählungen, Argumentationen) zu entwickeln oder mit diesen umzugehen (ebd.).

Für die Fundierung des Konzepts erfolgte eine Adaption des Prozessmodells historischen Denkens der FUER-Gruppe, indem dessen Teilbereiche mithilfe von Indikatoren aus internationalen empirischen Studien operationalisiert wurden. Dazu wurde auf deutsch- und englischsprachige Forschungsliteratur zurückgegriffen, sodass das Konzept für die Interpretation unserer Analysen besonders geeignet erscheint, um *eine* Theorie geschichtsspezifischer Kognition, die sowohl geschichtstheoretisch als auch empirisch eingebettet ist, empirisch zu prüfen und gegebenenfalls zu erweitern.

2 Fördernummer: 100019_192414.

Laut den Autor*innen dient narrative Kompetenz der Bewältigung von gegenwärtigen Irritationen (z. B. hinsichtlich Identität[en], Geschichte[n]) oder gesellschaftlichen Anforderungen (z. B. schulische Aufgaben), die sich mittels historischer Bezüge bearbeiten lassen. Die Operationen des Konstruktes lassen sich, so die Autor*innen weiter, in vier Bereiche unterscheiden: 1) Subjekte überführen Irritationen oder Anforderungen in *Interessen*, indem sie Fragen oder Hypothesen formulieren oder solche von anderen Subjekten erkennen. 2) Während der *De-Konstruktion* identifizieren Personen Aussagen über Vergangenes in Quellen und Darstellungen und vergleichen sie formal. 3) Auf Operationen der *Re-Konstruktion* wird zurückgegriffen, sobald Subjekte mittels Medienkritik die Plausibilität von Aussagen etwa hinsichtlich vergangener Situation oder zeitgenössischer Funktion der Medien einschätzen, wozu das Kontextualisieren mittels historischen Wissens sowie Medienvergleiche nötig sind. Zudem nutzen sie die Konstruktion von Zeitlichkeit mittels Chronologie und Deutungsmustern (z. B. Dauer und Wandel), um historische Aussagen zu strukturieren. 4) Für die historische *Orientierung* greifen Subjekte schliesslich auf Sinnbildungsmuster (z. B. kritische, genetische) zurück, wenn sie zukünftiges Handeln begründen. Zudem beurteilen sie die Relevanz von Vergangenem für heute (Nitsche & Gollin, 2020, 321).

Der Umgang mit historischen Medien sollte sich idealtypisch über alle Operationen narrativer Kompetenz erstrecken, da durch die Konfrontation mit ihnen historische Fragen ausgelöst und auf diese mittels Analyse von Medien Antworten gefunden werden können. Dieser Vorgang lässt sich aufgrund der in der Einleitung ausgeführten theoretischen und sozioempirischen Erfahrungen voraussichtlich nicht einem Aspekt narrativer Kompetenz trennscharf zuordnen. Im Fokus dieses Beitrags stehen daher sowohl Operationen der De-Konstruktion als auch der Re-Konstruktion, also das Identifizieren historischer Aussagen, der Medienvergleich, das Kontextualisieren sowie die Medienkritik.

3 Forschungsstand

Der Forschungsstand zum historischen Denken während des Umgangs mit Medien wächst in jüngster Zeit stetig. Die vorgenommenen Operationalisierungen unterscheiden sich dabei je nach Forschungsansatz, Erhebungsmethoden und gewähltem theoretischem Rahmen. In der deutschsprachigen Literatur wurden häufig theoretisch fundierte Operationalisierungen vorgenommen, die sich an der historischen Methode orientieren. Daraus lassen sich Operationen wie äus-

sere und innere Quellenkritik sowie Quelleninterpretation bzw. Beschreibung, Deutung und Interpretation ableiten (Bernhardt, 2011; Sauer, 2012; Scheller, 2019b). Im Rahmen des FUER-Modells wurden Re-Konstruktionsprozesse anhand von Quellen und De-Konstruktionsprozesse mit Blick auf Darstellungen konturiert und in verschiedenen sozioempirischen Ansätzen operationalisiert. Die Teilbereiche der Re-Konstruktionskompetenz orientieren sich an jenen der historischen Methode (Schreiber, 2007), die De-Konstruktionskompetenz wurde mittels Analyse der Oberflächen- und Tiefenstruktur operationalisiert (Kühberger, 2013; Kühberger, Neureiter & Wagner, 2018; Scheller, 2019a; Schöner, 2013a, 2013b).

In der englischsprachigen Literatur lassen sich grob vier Forschungsansätze ausmachen, die zum Teil auch miteinander kombiniert werden. Zunächst nutzten verschiedene Studien fachunspezifische Operationen wie *reading*, *rereading*, *summarizing* oder *inferring* (z. B. van Drie, van Driel & van Weijen, 2021; Holdinga, Janssen & Rijlaarsdam, 2021; Nokes & Kesler-Lund, 2019). Beim zweiten Ansatz wurden oft die fachspezifischen Heuristiken *sourcing*, *corroboration* und *contextualization* aus den Arbeiten Samuel Wineburgs (z. B. 1991) verwendet, um den Umgang mit historischen Medien zu erfassen (z. B. Nokes, Dole & Hacker, 2007; Reisman, Brimsek & Hollywood, 2019; Wiley, Griffin, Steffens & Britt, 2020). In Arbeiten des dritten Forschungsstranges erfolgte die Adaption eines allgemeinen Leseverständnismodells für Geschichte. Dabei wird davon ausgegangen, dass während des Lesens verschiedene mentale Modelle gebildet werden, etwa hinsichtlich des Mediums (*document model* bzw. *representation of the subtext*), des Medieninhaltes (*representation of the text*) und des historischen Ereignisses (*representation of the event* bzw. *situation model*) (Britt, Perfetti, Van Dyke & Gabrys, 2000; Wineburg, 1994). Für Studien der vierten Richtung wurden weitere fachspezifische Konzepte wie *evidence* (Monte-Sano, 2010; Samuelsson & Wendell, 2016; Tally & Goldenberg, 2005), *relevance* (van Drie et al., 2021), *significance* (Rantala & Khawaja, 2018) und *trustworthiness* bzw. *reliability* (McGrew, 2022; Rouet, Britt, Mason & Perfetti, 1996; Rouet, Favart, Britt & Perfetti, 1997) aufgegriffen, um diese Aspekte während der Geschichtskonstruktion anhand von Denk- oder Schreibprodukten sichtbar zu machen. Auch das Erfassen der Perspektive des*der Urheber*in eines Mediums (*historical perspective taking*) (Seixas, 2017) oder der in den Medien aufgeführten historischen Akteur*innen (*historical empathy*) kamen dabei in den Blick (z. B. Lee, 2013; Marcus, Paxton & Meyerson, 2006).

Die qualitative Differenzierung historischen Denkens erfolgte bislang in der deutschsprachigen Diskussion vor allem theoretisch (Klose, 2004; Pandel, 2013). Auf dieser Grundlage und unter Rückgriff auf Lawrence Kohlbergs Stufen der Moralentwicklung entfaltete etwa Andreas Körber im Rahmen des FUER-Modells Kriterien der Graduierung von Kompetenzausprägungen. Dabei führt er insbesondere das Kriterium *Komplexität* auf und unterscheidet zwischen einfachen, monodimensionalen, multi- und polydimensionalen sowie komplex vernetzten Ausprägungen (Körber, 2007).

In sozioempirischen Studien wurden qualitative Unterschiede für Teilaspekte des historischen Denkens beim Umgang mit Medien in den Blick genommen, woraus sich fünf weitere Kriterien anhand der dort vorgenommenen Stufungen ableiten lassen. Nach dem ersten Kriterium der *Vollständigkeit* gingen insbesondere Autor*innen vor, welche das Leseverständnis bei historischen Texten untersuchten. Sie nutzten u. a. die Anzahl der benannten historischen Akteur*innen, Ereignisse und diesbezügliche Zusammenhänge als Qualitätsmerkmale (z. B. Britt & Sommer, 2004; Le Bigot & Rouet, 2007). In Untersuchungen der zweiten Richtung wurde implizit das Kriterium der *Wissenschaftsförmigkeit* herangezogen. Zum Beispiel unterschied Markus Bernhardt (2007) zwischen spontanen und abgesicherten Deutungen anhand von Laut-Denk-Protokollen während einer Bildanalyse. Auch Matthias Martens (2010) Graduierungen der Wahrnehmung von Perspektivität in Darstellungen lassen sich dem Kriterium der Wissenschaftsförmigkeit zuordnen, da der Autor unter anderem analysierte, inwiefern Lernenden der Konstruktcharakter von Geschichte bewusst ist. Andere Forschende nutzten als drittes Kriterium die *Begründbarkeit*. Zum Beispiel stuften Marcel Mierwald (2020) bei historischen Argumentationen sowie Martin Nitsche und Monika Waldis (2016) bei historischen Erzählungen, inwiefern evidente Verweise auf Medien vorgenommen wurden. Als viertes Kriterium lässt sich die *Funktionalität* ableiten. So untersuchte Jeffrey Nokes (2017) anhand von Essays, inwiefern die Heuristik *sourcing* genutzt wurde. Dabei unterschied er, ob die Erkenntnisse der Operation für das Beantworten der Aufgabe genutzt oder lediglich genannt wurden. Schliesslich differenzierten zahlreiche Arbeiten danach, inwiefern eine bestimmte Operation überhaupt angewandt wurde, woraus sich das fünfte Kriterium *Vorhandensein* ableiten lässt. Als Beispiel kann die Studie von Abbi Reisman und Kolleg*innen (2019) aufgeführt werden, welche die Heuristik *corroboration* untersuchten.

Zusammenfassend liegen für die Operationalisierung des Umgangs mit Medien vier Ansätze vor, die für die vorliegende empirische Analyse herangezogen

werden: fachunspezifische und -spezifische Heuristiken, ein historisch gewendetes Leseverstehensmodell sowie historische Konzeptbezüge. Zudem scheint es, als ob für die Analyse qualitativer Unterschiede von Operationen historischen Denkens Kriterien wie Vollständigkeit, Wissenschaftsförmigkeit, Begründbarkeit, Komplexität, Funktionalität und Vorhandensein geeignet sind.

4 Methoden

4.1 Sample

Von den Annahmen ausgehend, dass historisches Denken sozial konstruiert ist und wissenschaftsorientiert ausgebildet werden sollte (Körber et al., 2007), wurde der klassische Noviz*innen-Expert*innen-Vergleich erweitert, welcher international bereits mehrfach mit Blick auf Geschichte Verwendung fand (z. B. Wineburg, 1991). Um gesellschaftlich übliches historisches Denken möglichst kontrastreich zu untersuchen, wurde die institutionelle Verankerung als Auswahlkriterium herangezogen. So konnten für die Pilotstudie je ein Deutschschweizer und ein deutscher Historiker als Experten gewonnen werden, wobei die erste Person zum Thema der Untersuchung arbeitet, was analytisch günstig scheint, da bekannt ist, dass Expert*innen, die zum Untersuchungsthema arbeiten, anders vorgehen als themenfremde Expert*innen (Wineburg, 1998). Als Semi-Expert*innen nahmen zwei Deutschschweizer Geschichtsstudierende teil. Um Formen historischen Denkens zu berücksichtigen, die vermutlich elaborierter sind als bei Schüler*innen, aber nicht so professionalisiert wie bei Historiker*innen, wurden ausserdem zwei geschichtsinteressierte Lai*innen rekrutiert, welche sich für Geschichte interessieren und ehrenamtlich in Deutschschweizer Museen tätig sind. Als Noviz*innen wurden zwei Deutschschweizer Schüler*innen der 9. Klasse befragt. Eine ausführlichere Darstellung der Personaldaten kann Tabelle 1 entnommen werden. Für die Hauptstudie werden neben einer grösseren Stichprobe weitere Personengruppen einbezogen (Schüler*innen auf Sekundarstufe I und II verschiedener Altersstufen sowie ausseruniversitäre Historiker*innen), um eine theoretische Sättigung zu erreichen (z. B. Strübing, 2014).

Gruppe	Expertise (Themen und formaler Bildungsstand bzw. -gang)	Alter	Geschlecht (m, w, d)
Historiker*in (PE 01)	Geschichtskultur zur Schweiz im 1. Weltkrieg u. a., Habilitation	42,8	m
Historiker*in (PE 02)	Mittelalter, Habilitation	38,9	m
Student*in (TP 08)	Bachelor Deutsch + Geschichte, 14. Semester	28,2	w
Student*in (TP 09)	Matura, 7. Semester	25,7	m
Lai*in (PE 03)	Lehrabschluss, Museumsguide	50,9	w
Lai*in (PE 04)	Lehrabschluss und Weiterqualifikation, Museumsguide	46,8	m
Schüler*in (PE 05)	Klassenstufe 9, Realschule (Grundniveau)	15,3	m
Schüler*in (PE 06)	Klassenstufe 9, Sekundarschule (mittleres Niveau)	15,6	w

Tab. 1: Personaldaten

4.2 Datenerhebung

Aufgrund der Covid-19-Pandemie wurde ein digitales Erhebungssetting gewählt, dessen technische Funktionsweise mit Schweizer Geschichtsstudierenden erprobt wurde. Die Proband*innen waren via Webex-Videokonferenz mit den Erhebenden verbunden. Die Datenerhebung erfolgte an zwei aufeinanderfolgenden Terminen. Während des ersten Termins wurden die Personenmerkmale der Teilnehmenden (z. B. Alter) mittels digitalen Fragebogens erfasst. Anschliessend erprobten sie Lautes Denken anhand einer Übungsaufgabe, bevor sie mit der Hauptaufgabe konfrontiert und aufgefordert wurden, alles zu äussern, was ihnen in den Sinn kommt. Das Verfahren wurde ausgewählt, da es für die Erfassung kognitiver Prozesse als besonders geeignet gilt. Dennoch sind Herausforderungen wie der Einfluss der Methode auf Denkprozesse bei der Dateninterpretation zu beachten (Konrad, 2010; Thyroff, 2022). Das Vorgehen und Laute Denken der Beteiligten wurden audiovisuell aufgezeichnet. Anschliessend erfolgte ein Interview über die Vorgehensweise der Teilnehmenden. Zudem erhielten sie an einem 2. Termin die Möglichkeit, ihre historische Frage schriftlich zu beantworten. Abschliessend wurden sie erneut zu ihrer Vorgehensweise interviewt. Für diesen Beitrag erfolgte eine Auswertung der Daten von Termin 1.

Die Proband*innen erhielten an Termin 1 eine medienbasierte Aufgabe bestehend aus zwei Aufträgen. Der erste Auftrag forderte sie auf, die Medien hinsichtlich interessanter Stellen zu untersuchen, um eine historische Frage zu entwickeln. Der zweite zielte darauf, die Medien hinsichtlich der historischen Frage zu analysieren. Als Thema wurde die Schweizer Neutralität im Ersten Weltkrieg ausgewählt, da Neutralität in der Schweiz latent diskutiert wird, mit Blick auf den Ersten Weltkrieg jedoch selten. Dadurch sollte ein geschichtskulturell relevantes Thema angeboten werden, das jedoch nicht von Kontroversen oder sozial erwünschten Urteilen überlagert ist, die historische Denkprozesse bestimmen können, wie am Beispiel des Nationalsozialismus in Studien mit deutschsprachigen Schüler*innen mehrfach herausgearbeitet wurde (z. B. Köster, 2013). Die Medien wurden digital auf SWITCHdrive zur Verfügung gestellt. Es handelt sich um die folgenden sieben:

- eine historiografische Darstellung zur Schweizer Neutralität,
- eine Textquelle zu Neutralitätsmassnahmen von 1914,
- je eine Textquelle zur militärischen Zusammenarbeit mit Deutschland bzw. Frankreich,
- je eine Postkarte von 1915 und 1916 mit Verbildlichungen der Neutralität und
- eine journalistische Darstellung zur heutigen Sicht auf die Schweizer Neutralitätsgeschichte.

Die Darstellungen wurden so ausgewählt, dass dem Prinzip der Kontroversität Rechnung getragen wird. Sie enthalten sowohl kontroverse Vergangenheitsdeutungen als auch Gegenwartsorientierungen. Die Quellen wurden multiperspektivisch ausgesucht und stammen aus offiziellen und nicht offiziellen Kontexten. Dadurch sollten vielfältige Auseinandersetzungen mit dem historischen Thema ermöglicht werden.

4.3 Datenaufbereitung und -analyse

Für den Beitrag wurden die audiovisuellen Daten des ersten Termins mittels der Software F4x automatisch transkribiert und anschliessend von Hilfskräften geprüft. Die Transkripte wurden in MaxQDA eingepflegt und qualitativ mittels strukturierender Inhaltsanalyse ausgewertet, da dieser Zugriff die Analyse spezifischer Aspekte ermöglicht (Mayring, 2010). Für die Erhebung von Operationen narrativer Kompetenz wurde deduktiv-induktiv vorgegangen. Zunächst bestand die Herausforderung, dass theoretisch hergeleitete Operationen oft-

mals nur heuristisch zu trennen sind (Thünemann & Jansen, 2018). Um dies zu berücksichtigen und ausserdem theoretische Offenheit sicherzustellen, wurden anfangs relativ grosse Kategorien gebildet, welche bei der anschliessenden Datenanalyse durch den Erst- und Zweitautor verfeinert wurden und zukünftig weiter ausdifferenziert werden sollen. Dabei wurde auf eine grundsätzliche Trennschärfe zwischen den Kategorien geachtet, im Material waren jedoch überschneidende Codierungen möglich. Die Codiereinheiten schwanken entsprechend der Ausführlichkeit der geäusserten Gedanken von wenigen Worten bis zu mehreren Sätzen. Um die Codier-Übereinstimmung zu prüfen, wurden 50 Prozent der Pilotdaten (n = 4) von den Analysierenden getrennt codiert. Dabei wurden zufriedenstellende Reliabilitätswerte erzielt. In Tabelle 2 werden die Kategorien sowie ihre Reliabilitätswerte aufgeführt.

Kategorie	Definition	K
Kontextualisieren	Eigenes Vorwissen wird einbezogen bzw. auf fehlendes Vorwissen verwiesen.	0.74
Medieninhalt erschliessen	Auseinandersetzung mit dem Medium mit dem Ziel, den Inhalt zu verstehen	0.58
Medienvergleich	Verschiedene Medien werden miteinander verglichen.	0.88
Medienkritisches Vorgehen	Formale Merkmale (z. B. Autor*in, Medienart) werden genannt bzw. genutzt, um Aussagen zur Perspektive, Relevanz oder Glaubwürdigkeit des Mediums zu treffen.	0.63
Aussagen hinsichtlich Frage identifizieren	Identifizierung inhaltlicher Aussagen, die für die Beantwortung der historischen Frage als relevant oder irrelevant eingeschätzt werden	0.87

Tab. 2: Operationen des historisch denkenden Umgangs mit Medien

Für die Identifizierung von Qualitätsunterschieden wurden die aus dem Forschungsstand herausgearbeiteten und in Kapitel 3 aufgeführten Kriterien an den Gesamtprozess sowie an einzelne Operationen angelegt und geprüft, ob sich Beispiele für Qualitätsunterschiede finden lassen. Dabei wurde explorativ vorgegangen und eine konsensuelle Übereinstimmung zwischen den Autor*innen hergestellt.

5 Ergebnisse

Entsprechend den Forschungsfragen besteht der Ergebnisbericht aus zwei Teilen. Zunächst wird dargestellt, welche Operationen des Umgangs mit Medien identifiziert werden konnten. Anschließend wird darauf eingegangen, inwiefern sich in den Daten Unterschiede in den operationalen Ausprägungen erkennen lassen.

5.1 Der Umgang mit historischen Medien

Die Operationen wurden mit ihren jeweiligen Definitionen und Reliabilitäten bereits in Tabelle 2 aufgeführt. Im Folgenden werden die einzelnen Operationen mit Ankerbeispielen und ersten Ansätzen der Subtypisierung vorgestellt.

Die Kategorie *Kontextualisieren* konnte aus der Theorie narrativer Kompetenz (vgl. Kapitel 2) übernommen und mit zufriedenstellender Reliabilität identifiziert werden. Sie wurde codiert, wenn Vorwissen zur Auseinandersetzung mit Medien einbezogen wurde. Beispielsweise nutzt eine Person ihr Vorwissen, um sich den Inhalt einer Postkarte zu erschliessen:

Kommen wir zu Material F. Das ist nochmal eine Postkarte aus der Zeit des Ersten Weltkrieg, die stammt von 1916. Die hier stammt von 1915. Die ist ein bisschen neuer. Interessantes Buch, muss ich mir vielleicht einmal besorgen. Flaggen auf der linken Seite von vorn nach hinten: Frankreich, Russland, das dürfte Italien sein, Belgien, Serbien. Vereinigtes Königreich, Grossbritannien war doch da schon im Krieg eingetreten? Die haben doch, glaube ich die, die Kriegserklärung gleichzeitig abgegeben. (PEO4, Pos. 74)

Die Operation *Medieninhalt erschliessen* wurde codiert, wenn eine Auseinandersetzung mit einem Medium stattfand, um dessen Inhalt zu erfassen. Die Reliabilität befindet sich lediglich im unteren zufriedenstellenden Bereich, was mehrere Ursachen hat. Bereits im Kapitel Forschungsstand wurde auf die Schwierigkeit verwiesen, fachunspezifische und fachspezifische Operationen des «Lesens» von Medien empirisch zu differenzieren. Ausserdem scheint es herausfordernd, analytische und synthetisierende Operationen voneinander zu trennen. Auch das Differenzieren zwischen *representation of the text* und *repre-*

sentation of the event nach dem Modell von Samuel Wineburg (1994) scheint in den vorliegenden Daten herausfordernd, da beides in den Laut-Denk-Protokollen ineinander übergeht. Die folgenden Beispiele sollen dies illustrieren:

Ja, also das sind einfach nur Hypothesen, viel mehr Material wird zumindest hier in der Quelle nicht ersichtlich, ob's das überhaupt gibt oder nicht. Wir gehen mal weiter. Also hie-, dass ist eigentlich so ne hypothetisch angezeigte Verletzung des in der Quelle oder Material A angezeigten Naturrechts, äh, Neutralitätsrechts, nicht Naturrecht. (TP09, Pos. 133)

[liest] «Er darf den kriegführenden Staaten keine Truppen und Operationsbasen zur Verfügung stellen, ihnen keinen Durchmarsch gestatten, aus staatseigenen Beständen kein Kriegsmaterial liefern.» [liest nicht mehr] Das finde ich noch interessant, aus staatseigenen Beständen. [schweigt, liest leise] Aus staatseigenen Beständen, okay, die Wirtschaft darf weiterhin liefern, so wie ich das verstehe. (PE04, Pos. 31–33)

Im ersten Beispiel wird eine Zusammenfassung des Quelleninhalts vorgenommen («hypothetisch angezeigte Verletzung des in der Quelle [...] angezeigten [...] Neutralitätsrechts»), welche sich gemäss Wineburg der Textebene zuordnen lässt. Diese enthält jedoch gleichzeitig eine Deutung des historischen Sachverhalts (Verletzung des Neutralitätsrechtes), sodass ebenfalls eine mentale Repräsentation des historischen Ereignisses vorliegen dürfte. Das zweite Beispiel liesse sich schon eher als schlussfolgernde Aussage über das historische Ereignis (*representation of the event*) verstehen («okay, die Wirtschaft darf weiterhin liefern»). Allerdings waren solche Beispiele zu selten, um eine reliable Zuordnung zu ermöglichen. Ähnlich verhält es sich bei der Unterscheidung zwischen Zusammenfassen (*summarize*) und Schlussfolgern (*inferring*), welche Jeffery Nokes und Alisa Kesler-Lund (2019) vornahmen.

Als vorläufiges Fazit lässt sich festhalten, dass die Analyse des Inhaltsererschliessens von Medien herausfordernd bleibt. Dagegen konnten weitere Denktionen des Umgangs mit Quellen und Darstellungen mit guter Reliabilität identifiziert werden. So nutzten Proband*innen den *Medienvergleich*, der dem Bereich der Re-Konstruktion des Modells narrativer Kompetenz entnommen wurde und Ähnlichkeiten mit Wineburgs (1991) *corroboration* aufweist. Die Daten zeigen, dass der Medienvergleich drei Funktionen diene. Im nachfolgenden Beispiel wird ein Medium mit einem anderen verglichen, um den Inhalt zu erfassen:

Dann hier das zweite ist wieder auch eine bildliche Darstellung. [tippt] Und hier ist interessant, scheint mir interessant, im Anfang diese Selbstverkleinerung, diese Darstellung des Krieges, inklusive globaler Aspekt. [...] während oben eben eher so karikaturhaft, ist es hier quasi, ist es hier eigentlich sehr ernst dargestellt und aber irgendwie abhängig, überhaupt nicht abhängig vom Agieren [der] Schweiz [...] (PE01, Pos. 31–32)

Des Weiteren wurde der Vergleich ebenfalls eingesetzt, um eine historische Frage anhand von fünf Medien abzuleiten, indem Bezüge zwischen diesen hergestellt und Differenzen wahrgenommen wurden:

Da ist A und B sind halt dann vorwiegend, vorwiegend äh, wie soll ich sagen, ein bisschen diese Grundlagentexte, die Ausgangslage beschreiben. Ähm. Während gerade die anderen beiden Textquellen und auch eine der Auszug des Artikels, gut der ist jetzt hier weniger, ähm thematisiert, also vor allem, vor allem [...] in diesem Zusammenhang wäre es natürlich, wäre es natürlich interessant, äh zu Material C und Material D hat auch, so diese Fragen ein bisschen in den Vordergrund zu stellen und E also die Postkarten insbesondere, die erste der beiden Postkarten stellen hier natürlich, also dieses Problem, von dem die Schweiz äh, oder dass sich der Schweiz während des Ersten Weltkriegs eröffnet natürlich schön dar. [...] Und, schon aufgrund, schon aufgrund dieser, dieser Textstellen E und D ist dann halt auch die Frage, wie wahrheitsgetreu dann diese Postkarte in Material F ist. (PE04, Pos. 128–129)

Nachdem diese Person mittels Medienvergleich eine historische Frage entwickelt hat, nutzt sie diese Operation ebenso, um Antworten auf die historische Frage zu finden:

[...] Und diese Neutralität, diese strikte Neutralität nicht immer vollständig eingehalten worden ist, ich denke, das ist so ein bisschen der Schluss, den ich aus den Quellen ähm, C und D ziehe. Dass diese Neutralität nicht vollständig eingehalten worden ist. Die Schweiz aber halt auch Zugeständnisse machen musste, um schlussendlich nicht vom Seil zu fallen. Wie es auf dieser Postkarte so schön dargestellt ist. (PE04, Pos. 153)

Ebenfalls mit zufriedenstellender Reliabilität liess sich die Denkoperation *medienkritisches Vorgehen* identifizieren. Sie ist im Modell narrativer Kompetenz dem Bereich der Re-Konstruktion entnommen. Als vergleichbare Konstrukte finden sich in der deutschsprachigen Literatur die *Quellenkritik* (z. B. Pandel, 2006) sowie in der englischsprachigen das *sourcing* (Wineburg, 1991). Das *medienkritische Vorgehen* wurde codiert, wenn formale Merkmale wie Urheber*in, Gattung, Entstehungszeit bzw. -ort oder Adressat*innen genannt oder genutzt wurden, um Aussagen über die Perspektive, Relevanz oder Glaubwürdigkeit eines Mediums zu machen. Beispielsweise nutzt die folgende Person die Autor*innenangabe, um Vermutungen über die Perspektive anzustellen:

Ah, Andre Langie. Ah, vielleicht war das ein Franzose, der oder er kam aus dem französischen Teil [...] und war auch deshalb gegen die Deutschen. (TP 02, Pos. 42–43)

Nachdem die Proband*innen eine historische Frage entwickelt hatten, wandten sie sich oft erneut den Medien zu, um Aussagen zu identifizieren, die sie für die Beantwortung der historischen Frage nutzen wollten. Diese Operation (*Aussagen hinsichtlich Frage identifizieren*) aus dem Modell narrativer Kompetenz (Nitsche & Gollin, 2020) konnte mit sehr guter Reliabilität in den Daten identifiziert werden. Beispielsweise formulierte PEO1 die historische Frage, inwiefern die Schweiz im Ersten Weltkrieg neutral war. Anschliessend geht er die einzelnen Medien nacheinander durch und identifiziert für seine historische Frage relevante Aussagen:

Dann Material C, ist jetzt hinsichtlich meiner Frage. Hier wird deutlich, dass der Geheimdienst, der [...] Geheimdienst hier offenbar, eng mit den Deutschen kooperiert, wie ja bekannt, historisch bekannt [...] Material D. Hier sieht man quasi, dass, dass aber auch die Neutralität eben auch ermöglicht, ermöglicht auch, quasi der, der französischen Offiziersdelegation, Ansprüche zu erheben. [...] Dann, in Material E wird deutlich [...] (PEO1, Pos. 83–86)

Zusammenfassend konnten während des Umgangs mit historischen Medien die Operationen *Kontextualisieren*, *Medienvergleich*, *medienkritisches Vorgehen* und *Aussagen hinsichtlich Frage identifizieren* mit guter Reliabilität kategorisiert und zum Teil schon in Suboperationen ausdifferenziert werden. Herausfordernd bleibt es, das *inhaltliche Erschliessen von Medien* fachspezifisch zu konturieren.

5.2 Identifizierung qualitativer Unterschiede des historisch denkenden Umgangs mit Medien

Im Folgenden werden erste Ergebnisse hinsichtlich der Frage vorgestellt, inwiefern die operationalen Ausprägungen qualitativ unterschieden werden können. Die Konsultation des Forschungsstandes ergab, dass die Kriterien *Vollständigkeit, Wissenschaftsförmigkeit, Begründbarkeit, Komplexität, Funktionalität und Vorhandensein* zur Bestimmung von Unterschieden des Gesamtprozesses oder der Teiloperationen historischen Denkens geeignet scheinen, weshalb sie für die Datenanalyse herangezogen wurden.

Mit Blick auf den Gesamtprozess lässt sich mithilfe von Dokumentporträts aus MaxQDA die unterschiedliche Nutzung von Denkoperationen darstellen. Dabei werden die Reihenfolge und die Länge der Codierungen grafisch dargestellt. Weisse Stellen verdeutlichen, dass in diesen Abschnitten keine der benannten Kategorien zugeordnet wurden. Aus dieser Grafik geht allerdings nicht hervor, wann sich Codierungen überlappen. In Abbildung 1 sind die visualisierten Vorgehensweisen eines Experten (PEo1) und Novizen (PEo6), welche sich deutlich voneinander unterscheiden, nebeneinander angeordnet. Der Vergleich zeigt, dass beide auf eine unterschiedliche Anzahl von Operationen zurückgreifen. So besteht die Herangehensweise des Schülers aus drei Denkoperationen (gepunktet = *Medieninhalt erschliessen*, weiss = *medienkritisches Vorgehen*, kariert = *Medienvergleich*), während der Historiker zusätzlich *kontextualisiert* (schraffiert) sowie *Aussagen für Fragen identifiziert* (schwarz). Diese Unterscheidung orientiert sich am Kriterium *Vorhandensein*.

Dokumentporträts von PEo1 (Historiker) und PEo6 (Schüler) aus MaxQDA

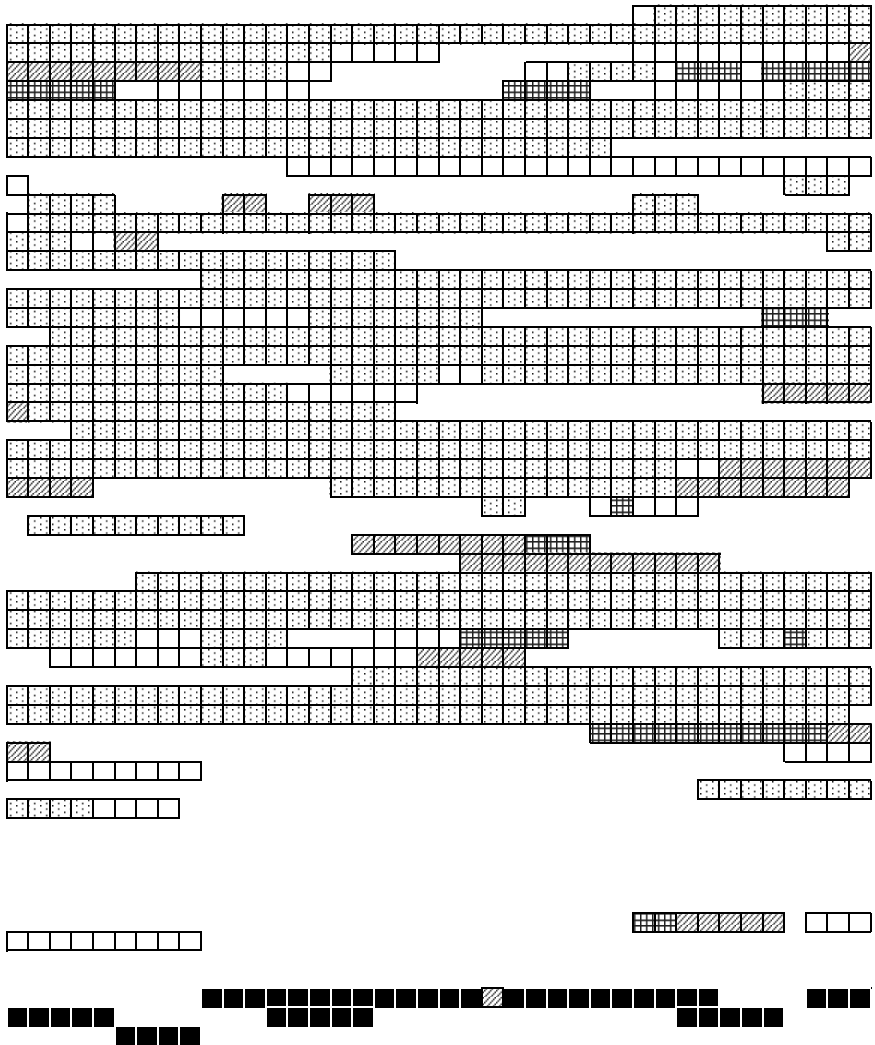


Abb. 1: Gesamtprozess PE01

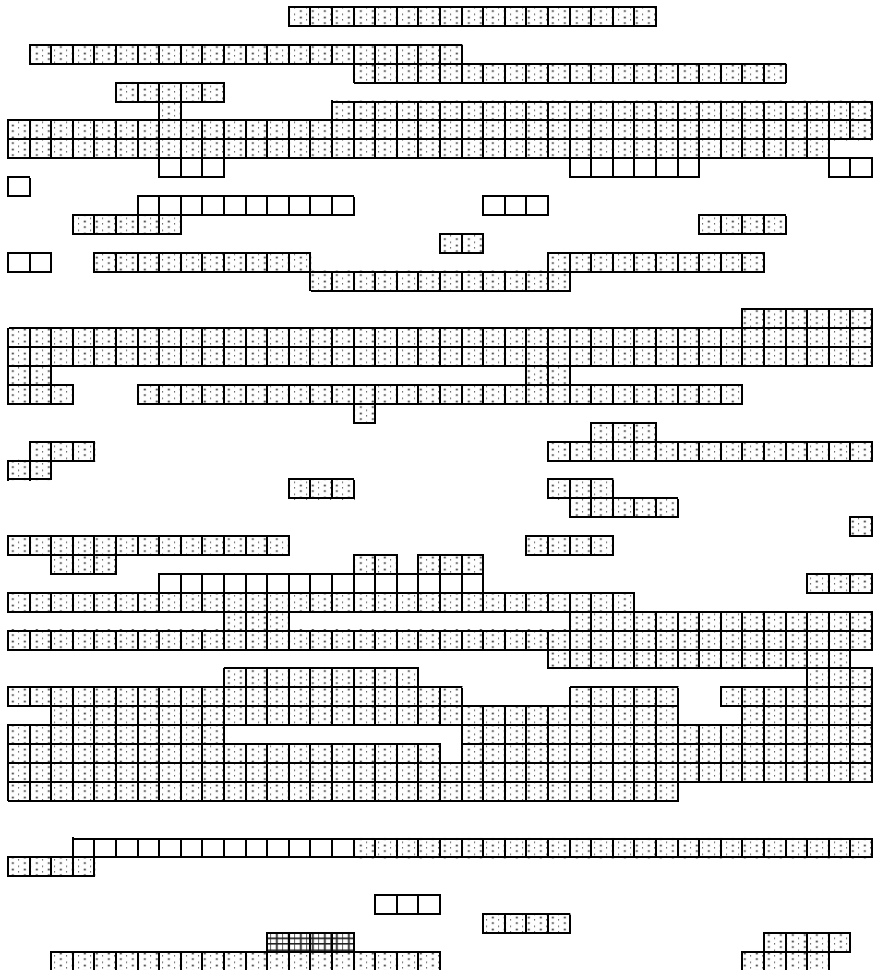


Abb. 2: Gesamtprozess PE06

Neben der Anzahl der genutzten Denkopoperationen zeigte sich bei der Analyse mit Blick auf den Gesamtprozess ebenfalls, dass manche Proband*innen in kurzer Zeit auf verschiedene Denkopoperationen zurückgriffen, während sie sich mit einem Medium auseinandersetzten. Andere Personen nutzten dagegen lediglich eine oder zwei Denkopoperationen. Insofern könnte von einem komplexeren oder weniger komplexen Umgang mit der Quelle bzw. der Darstellung gesprochen werden (Kriterium *Komplexität*). Beispielsweise nutzte die Laiin PEO3 die Operationen *Medieninhalt erschliessen*, *Medienvergleich* und *Kontextualisieren* bei der Auseinandersetzung mit der Quelle *Material D*:

[liest] «Material D: Die Führung in der Schweiz beschäftigte sich mit der Frage, was man tun sollte, wenn die Schweiz angegriffen würde. Hierzu fanden auch Absprachen statt.» [liest nicht mehr] Also dass [...] angegriffen würde, das finde ich genau jetzt, weil das, äh, Material C genau das Gegenteil, ähm, gesagt hat. Dass die Deutschen ja die Schweiz brauchten, um wichtige Dokumente zu übersetzen, dann nehme ich nicht an, dass sie die Schweiz angreifen würden. Die Schweiz war ohnehin ein Klacks auf ihrem Plan und wahrscheinlich lag auch viel Geld bei uns auf den Banken. Zu viel, zu viel Vermögen hier in der Schweiz, dass sie angegriffen hätte werden sollen. (PEO3, Pos. 39)

Zusammenfassend können bei der Analyse des Gesamtprozesses bisher die Kriterien *Vorhandensein* und *Komplexität* herangezogen werden, um Unterschiede zu erfassen. Dabei scheint es, dass Personen mit höherer Expertise mehr Operationen verwenden und diese stärker miteinander verschränken, wenn sie sich mit Medien auseinandersetzen.

Bezogen auf einzelne Operationen lassen sich weitere Kriterien anwenden, um Unterschiede zu identifizieren. Hinsichtlich der Operation *medienkritisches Vorgehen* lässt sich zwischen der blossen Nennung formaler Merkmale und deren Nutzung zur Erschließung der Perspektivität des Mediums unterscheiden. Diese Differenzierung orientiert sich am Kriterium der *Funktionalität* und fragt danach, inwiefern die formalen Merkmale für die Auseinandersetzung mit Medien genutzt werden. Während im ersten Beispiel ein Historiker aus den formalen Merkmalen eine konkrete Perspektive herausarbeitet, ist sie im zweiten Beispiel bei einer Studentin eher allgemein gehalten:

[...] ich recherchiere, wer ist eigentlich der Autor [...] Also das ist quasi politikwissenschaftlich-staatsrechtlicher Zugang zur Neutralität, der hier im HLS [Historisches Lexikon der Schweiz, d. V.] formuliert ist. (PE01 Pos. 15)

Jetzt sehe ich, es geht wieder um einen Text und, ähm, ich lese mal schnell den Titel: [liest] «Auszug aus dem Artikel «Ein Grundprinzip im Wandel: Neutral – was heisst das schon?», der 2018 auf der Homepage des SRF erschienen ist.» [liest nicht mehr] Also 2018, das heisst schon mal, es geht um einen, ähm, wenn man das so sagen kann, modernen Text. (TPO8 Pos. 138)³

In Anlehnung an die Studie von Jeffrey Nokes (2017) lässt sich *medienkritisches Vorgehen* ebenfalls nach dem *Grad der Begründung* unterscheiden. Im Folgenden finden sich zwei Beispiele, in denen Aussagen über die Plausibilität des Mediums getroffen wurden. Während im ersten Beispiel von einer Laiin widersprüchliche Aussagen zur Glaubwürdigkeit gemacht werden, nutzt eine Studentin im zweiten Beispiel ihr Vorwissen, um die Glaubwürdigkeit begründet einzuschätzen:

Das sind meine zwei Fragen, die ich jetzt beantwortet habe, wobei Wikipedia halt eben, äh, mit Vorsicht zu geniessen ist. Aber ich nehme nicht an, dass da irgend ein Unsinn steht. (PE03, Pos. 123)

Also, ähm, der Titel seit mir, dass es sich um ene Lexikonartikel handelt und ich kenn da Historische Lexikon der Schweiz au und [...] kenn das als e allgemein anerkannti, verlässlechi, [...] Ressource. (TPO8, Pos. 9)

Zusammenfassend eignen sich die anhand des Forschungsstandes herausgearbeiteten Kriterien zur Identifikation von Unterschieden. Je nach konkreter Operation können zum Teil mehrere Kriterien herangezogen werden. Beispielsweise lassen sich bisher für die Operation *medienkritisches Vorgehen* Qualitätsunterschiede anhand der Kriterien *Funktionalität* und *Begründbarkeit* festmachen. Der Gesamtprozess der Teilnehmenden der Pilotstudie kann hinsichtlich der Kriterien *Vorhandensein* sowie *Komplexität* näher beleuchtet werden.

3 Übersetzung aus Schweizerdeutscher Mundart.

6 Diskussion

Im Beitrag wurde bei verschiedenen Personengruppen untersucht, inwiefern sich Operationen des Umgangs mit historischen Medien sowie Unterschiede in den operationalen Ausprägungen narrativer Kompetenz identifizieren lassen. Dafür sollten Schüler*innen, Lai*innen und Historiker*innen anhand von Quellen und Darstellungen zur Schweizer Neutralität im 1. Weltkrieg eine historische Frage entwickeln sowie diese mittels der Medien beantworten. Die Protokolle Lauten Denkens wurden transkribiert und mittels qualitativer Inhaltsanalyse analysiert.

Dabei konnten bisher die Bereiche *Kontextualisieren*, *Medienvergleich*, *medienkritisches Vorgehen* sowie *Aussagen hinsichtlich historischer Frage identifizieren* aus dem Modell narrativer Kompetenz als Operationen identifiziert werden. Die von Wineburg (1994) und Britt et al. (2000) vorgenommene Unterscheidung zwischen medien- und ereignisbezogenen Aussagen bzw. von Nokes und Kessler-Lund (2019) zwischen Zusammenfassen und Schlussfolgern fiel anhand der Aussagen während des Lauten Denkens jedoch schwer, da sie sich häufig beiden Aspekten zuordnen liessen. Dadurch erscheint es, wie bereits mehrfach in der Literatur vermutet (z. B. Bracke et al., 2018), herausfordernd, zwischen analytischen und synthetisierenden Elementen bei der *Medienerschliessung* zu unterscheiden. Hinsichtlich der Identifikation operationaler Unterschiede erwiesen sich die Kriterien *Komplexität* und *Vorhandensein* beim Gesamtprozess sowie die Kriterien *Funktionalität* und *Begründbarkeit* bei der Operation *medienkritisches Vorgehen* als hilfreich.

In der laufenden Hauptstudie kommt es folglich darauf an, geschichtsspezifische und allgemeine Aspekte der Medienerschliessung sowie weitere Suboperationen der untersuchten Konstrukte theoretisch zu konkretisieren und empirisch rückzubinden. Zudem sollen qualitative Unterschiede von Aspekten historischen Denkens weiter ausdifferenziert werden, um davon ausgehend empirisch fundierte Entwicklungskonzepte zu konstruieren und normativ zur Diskussion zu stellen. Dazu wird das Sample unter Einbezug von Schüler*innen unterschiedlichen Alters auf der Sekundarstufe I und II sowie von ausseruniversitären Historiker*innen erweitert, um eine theoretische Sättigung zu erreichen. Insgesamt werden in jeder Personengruppe 6 bis 12 Personen in die Untersuchung einbezogen.

Mit Blick auf die Limitationen sei betont, dass ein qualitatives Vorgehen generell keinen Anspruch auf Repräsentativität erheben kann, was sich bereits daraus ergibt, dass lediglich eine Aufgabe gestellt wurde und bekannt ist, dass historisches Denken an den jeweiligen Aufgabenkontext gebunden ist (z. B. Nitsche, Gollin & Waldis, 2017). Der Fokus dieser Pilotstudie war folglich auf narrative Performanz gerichtet und zielte nicht auf die Frage, über welche historischen Denkoperationen die Teilnehmenden generell in allen möglichen Situationen verfügen. Das überschaubare Sample ermöglicht darüber hinaus noch keine empirisch gesättigten Aussagen hinsichtlich der zur Lösung des Aufgabensettings üblichen Operationen historischen Denkens über die Expertisegruppen hinweg oder bezüglich der qualitativen Unterschiede zwischen diesen Gruppen. Nicht zuletzt birgt Lautes Denken als Erhebungsmethode das grundsätzliche Problem, dass das Denken schneller abläuft als das Sprechen (Thyroff, 2022). Hinsichtlich der Datenanalyse wurde deutlich, dass die Kategorien, die in anderen Studien anhand von Texten oder Interviews entwickelt wurden, bei den vorliegenden Laut-Denk-Protokollen kategorial nicht immer trennscharf und intersubjektiv reliabel identifiziert werden konnten. Darüber hinaus ergibt sich die Herausforderung, zum Teil unvollständig ausgesprochene Gedanken historischen Operationen zuzuordnen zu müssen, was zu nicht übereinstimmenden Codierungen führen kann. Trotz der erwähnten Einschränkungen kann argumentiert werden, dass das Erhebungssetting geeignet ist, um Operationen historischen Denkens während des Umgangs mit Medien sowie operationale Unterschiede zu beforschen, sodass im Rahmen der Hauptstudie weitere Einblicke in den Prozess historischen Denkens sowie diesbezügliche Qualitätsunterschiede zu erwarten sind.

Literatur

- Barricelli, Michele (2005). *Schüler erzählen Geschichte: Narrative Kompetenz im Geschichtsunterricht*. Schwalbach/Ts.: Wochenschau.
- Bernhardt, Markus (2007). Vom ersten auf den zweiten Blick. Eine empirische Untersuchung zur Bildwahrnehmung von Lernenden. *Geschichte in Wissenschaft und Unterricht*, 7(8), 417–432.
- Bernhardt, Markus (2011). «Ich sehe was, was du nicht siehst!». Überlegungen zur Kompetenzentwicklung im Geschichtsunterricht am Beispiel der Bildwahrnehmung. In Saskia Handro & Bernd Schönemann (Hrsg.), *Visualität und Geschichte* (S. 37–53). Berlin: LIT.

- Bracke, Sebastian / Flaving, Colin / Jansen, Johannes / Köster, Manuel / Lahmer-Gebauer, Jennifer / Lankes, Simone / ... Zülsdorf-Kersting, Meik (2018). *Theorie des Geschichtsunterrichts*. Schwalbach/Ts.: Wochenschau.
- Britt, M. Anne / Perfetti, Charles A. / Van Dyke, Julie & Gabrys, Gareth (2000). The Sourcerer'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.
- Britt, M. Anne & Sommer, Jodie (2004). Facilitating textual integration with macro-structure focusing tasks. *Reading Psychology*, 25(4), 313–339.
- Drie, J. van / Driel, J. van & Weijen, D. van (2021). Developing students' writing in History: Effects of a teacher-designed domain-specific writing instruction. *Journal of Writing Research*, 13(2), 201–229.
- 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.
- Günther-Arndt, Hilke (2015). Ein neuer geschichtsdidaktischer Medienbegriff angesichts des digitalen Wandels? In Christoph Pallaske (Hrsg.), *Medien machen Geschichte. Neue Anforderungen an den geschichtsdidaktischen Medienbegriff im digitalen Wandel* (S. 17–36). Berlin: Logos.
- Holdinga, Lieke / Janssen, Tanja & Rijlaarsdam, Gert (2021). The relationship between students' writing process, text quality, and thought process quality in 11th-grade history and philosophy assignments. *Written Communication*, 38(4), 544–586.
- Klose, Dagmar (2004). *Klios Kinder und Geschichtslernen heute. Eine entwicklungspsychologisch orientierte konstruktivistische Didaktik der Geschichte*. Hamburg: Dr. Kovac.
- Konrad, Klaus (2010). Lautes Denken. In Günter Mey & Katja Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 476–490). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Körper, Andreas (2007). Graduierung: Die Unterscheidung von Niveaus der Kompetenzen historischen Denkens. In Andreas Körper, Waltraud Schreiber & Alexander Schöner (Hrsg.), *Kompetenzen historischen Denkens. Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik* (S. 415–472). Neuried: ars una.
- 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 (2013). *Historisches Textverstehen. Rezeption und Identifikation in der multiethnischen Gesellschaft*. Berlin: LIT.

- Kühberger, Christoph (Hrsg.) (2013). *Geschichte denken. Zum Umgang mit Geschichte und Vergangenheit von Schüler/innen der Sekundarstufe I am Beispiel «Spielfilm»*. Empirische Befunde – Diagnostische Tools – Methodische Hinweise. Innsbruck u. a.: Studien-Verl.
- Kühberger, Christoph / Neureiter, Herbert & Wagner, Wolfgang (2018). Umgang mit Darstellungen der Vergangenheit. Historische De-Konstruktionskompetenz empirisch messen. *Geschichte in Wissenschaft und Unterricht*, (7/8), 418–434.
- Le Bigot, Ludovic & Rouet, Jean-François (2007). The Impact of Presentation Format, Task Assignment, and Prior Knowledge on Students' Comprehension of Multiple Online Documents. *Journal of Literacy Research*, 39(4), 445–470.
- Lee, Mimi (2013). Promoting historical thinking using the explicit reasoning text. *The Journal of Social Studies Research*, 37, 33–45.
- Marcus, Alan S. / Paxton, Richard J. & Meyerson, Peter (2006). «The Reality of it All»: History Students Read the Movies. *Theory & Research in Social Education*, 34(4), 516–552.
- Martens, Matthias. (2010). *Implizites Wissen und kompetentes Handeln. Die empirische Rekonstruktion von Kompetenzen historischen Verstehens im Umgang mit Darstellungen von Geschichte*. Göttingen: V & R unipress.
- Mayring, Philipp. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (11. Aufl.). Weinheim: Beltz.
- McGrew, Sarah. (2022). Internet or archive? Expertise in searching for digital sources on a contentious historical question. *Cognition and Instruction*, 34(4), 488–516.
- Mierwald, Marcel. (2020). *Historisches Argumentieren und epistemologische Überzeugungen: Eine Interventionsstudie zur Wirkung von Lernmaterialien im Schülerlabor*. Wiesbaden: Springer VS.
- Monte-Sano, Chauncey (2010). Disciplinary literacy in history: an exploration of the historical nature of adolescents' writing. *Journal of the Learning Sciences*, 19(4), 539–568.
- 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. (2017). Exploring patterns of historical thinking through eighth-grade students' argumentative writing. *Journal of Writing Research*, 8(3), 437–467.

- 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.
- Nokes, Jeffery D. & Kesler-Lund, Alisa. (2019). Historians' social literacies: How historians collaborate and write. *The History Teacher*, 52(3), 369–410.
- Norden, Jörg van (2011). *Was machst Du für Geschichten? Didaktik eines narrativen Konstruktivismus*. Freiburg: Centaurus.
- Pandel, Hans-Jürgen (2006). *Quelleninterpretation. Die schriftliche Quelle im Geschichtsunterricht*. (3. Aufl.). Schwalbach/Ts.: Wochenschau.
- Pandel, Hans-Jürgen. (2013). *Geschichtsdidaktik. Eine Theorie für die Praxis*. Schwalbach/Ts.: Wochenschau.
- Rantala, Jukka & Khawaja, Amna (2018). Assessing historical literacy among 12-year-old Finns. *The Curriculum Journal*, 29(3), 354–369.
- Reisman, Abby/Brimssek, Emily & Hollywood, Claire (2019). Assessment of historical analysis and argumentation (AHAA): A new measure of document-based historical thinking. *Cognition and Instruction*, 37(4), 534–561.
- Rouet, Jean-François/Britt, M. Anne/Mason, Robert A. & Perfetti, Charles A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology*, 88(3), 478–493. <http://dx.doi.org/10.1037/0022-0663.88.3.478>
- Rouet, Jean-François/Favart, Monik/Britt, M. Anne & Perfetti, Charles A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and instruction*, 15(1), 85–106.
- Samuelsson, Johan & Wendell, Joakim (2016). Historical thinking about sources in the context of a standards-based curriculum: a Swedish case. *The Curriculum Journal*, 27(4), 479–499.
- Sauer, Michael (2012). *Geschichte unterrichten. Eine Einführung in die Didaktik und Methodik* (10. Aufl.). Seelze-Velber: Klett.
- Scheller, Jan (2019a). Ein Diagnoseraster für die De-Konstruktionskompetenz nach dem FUER-Modell. In Monika Waldis & Béatrice Ziegler (Hrsg.), *Forschungswerkstatt Geschichtsdidaktik 17. Beiträge zur Tagung «geschichtsdidaktik empirisch 17»*. (S. 192–204). Bern: hep.
- Scheller, Jan (2019b). Rekonstruktion historischer Denkopoperationen. Schüler/-innen und Studierende analysieren die zugrundeliegenden Intentionen und Orientierungsabsichten eines Plakats. In Christine Pflüger (Hrsg.), *Die Komplexität des kompetenzorientierten Geschichtsunterrichts – Aktuelle geschichtsdidaktische Forschungen*. (S. 17–34). Göttingen: V & R.

- Schöner, Alexander (2013a). Die Eichstätter Schulbuchanalysen. Zur Methode kategorialer Inhalts- und Strukturanalysen. In Waltraud Schreiber / Alexander Schöner & Florian Sochatzy, *Analyse von Schulbüchern als Grundlage empirischer Geschichtsdidaktik* (S. 66–90). Stuttgart: Kohlhammer.
- Schöner, Alexander (2013b). Inhaltsbezogene Kategorien als «Grammatik historischen Denkens» und «inhaltliches Strickmuster» historischer Narrationen. In Waltraud Schreiber / Alexander Schöner & Florian Sochatzy, *Analyse von Schulbüchern als Grundlage empirischer Geschichtsdidaktik* (S. 91–101). Stuttgart: Kohlhammer.
- Schreiber, Waltraud (2007). Kompetenzbereich historische Methodenkompetenzen. In Andreas Körber, Waltraud Schreiber & Alexander Schöner (Hrsg.), *Kompetenzen historischen Denkens. Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik* (Bd. 2, S. 194–235). Neuried: Ars Una.
- Seixas, Peter (2017). A model of historical thinking. *Educational Philosophy and Theory*, 49(6), 593–605.
- Strübing, Jörg (2014). *Grounded Theory: Zur sozialtheoretischen und epistemologischen Fundierung eines pragmatistischen Forschungsstils* (3. Aufl.). Wiesbaden: Springer VS.
- Tally, Bill & Goldenberg, Lauren B. (2005). Fostering historical thinking with digitized primary sources. *Journal of Research on Technology in Education*, 38(1), 1–21.
- Thünemann, Holger & Jansen, Johannes (2018). Historisches Denken lernen. In Sebastian Bracke, Colin Flaving, Johannes Jansen, Manuel Köster, Jennifer Lahmer-Gebauer, Simone Lankes, ... Meik Zülsdorf-Kersting, *Theorie des Geschichtsunterrichts* (S. 71–106). Frankfurt am Main: Wochenschau.
- Thyroff, Julia (2022). Lautes Denken. In Georg Weißeno & Béatrice Ziegler (Hrsg.), *Handbuch Geschichts- und Politikdidaktik* (443–457). Wiesbaden: Springer Fachmedien.
- Trautwein, Ulrich / Bertram, Christiane / Borries, Bodo von / Brauch, Nicola / Hirsch, Matthias / Schröter, Kathrin / ... Zuckowski, Andreas (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 and Understanding: Innovative Designs for New Standards*. New York: Routledge.
- 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 (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.

Wineburg, Samuel (1994). The cognitive representation of historical texts. In Gaea Leinhardt, Isabel Beck & Catherine Stainton (Hrsg.), *Teaching and Learning in History* (S. 85–135). Hillsdale, NJ: Erlbaum.

Wineburg, Samuel (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science*, 22(3), 319–346.

Digitalisierung und historisches Denken

Künstliche Intelligenz im Fach Geschichte

Kann ein Computer Texte von Schüler*innen sprachlich und inhaltlich beurteilen?

Christiane Bertram¹, Zarah Weiß, Lisa Zachrich, Ramon Ziai

1 Theoretischer Rahmen – historisches Denken erfassen

Im Geschichtsunterricht sollen keine «Fakten» oder vorgegebenen Deutungen vermittelt und abgeprüft werden, sondern es geht um die Förderung historischer Kompetenzen, die für ein kritisches, reflektiertes und reflexives «Geschichtsbewusstsein» notwendig sind – so steht es in den Bildungsplänen aller Bundesländer, der Schweiz und Österreichs. Schüler*innen sollen mit problemorientierten Fragestellungen historische Materialien methodisch kontrolliert erschließen und zu einem begründeten historischen Urteil kommen, das einen Bezug zur eigenen Lebenspraxis hat. Im deutschsprachigen Diskurs wurde dieser Ansatz seit den 1970er-Jahren in dem Konzept des «Geschichtsbewusstseins» beschrieben, das nach der klassischen Definition von Jeismann (1997) einen Zusammenhang zwischen «Vergangenheitsdeutung, Gegenwartsverständnis und Zukunftsperspektive» (42) herstellen soll. Zugrunde liegt diesem Konzept die Narrativitätstheorie (vgl. Danto, 1965/1980; Ricœur, 1983/1988), die die «Narrativität» als wesentliches Strukturmerkmal von «Geschichte» herausgearbeitet hat. Als nach dem «PISA-Schock» im Jahr 2000 für das Fach Geschichte «Kompetenzen» formuliert werden sollten (Klieme et al., 2003), diente die Narrativitätstheorie und das Konzept des Geschichtsbewusstseins als Ausgangspunkt für die Formulierung verschiedener Kompetenzmodelle. In einigen Kompetenzmodellen werden das «historische Erzählen» und die «narrative Kompetenz»

¹ Die Ergebnisse des CAHisT-Projekts wurden 2021 von Christiane Bertram, Zarah Weiß, Lisa Zachrich und Ramon Ziai in dem internationalen Journal *Computers & Education* unter der Überschrift «Artificial Intelligence in History Education. Linguistic Content and Complexity Analyses of Student Writings in the CAHisT Project (Computational Assessment of Historical Thinking)» veröffentlicht. Der hier eingereichte Beitrag stellt eine stark gekürzte und auf die deutsche Fachdidaktik zugeschnittene Version des Beitrags dar.

betont (u. a. Barricelli, 2005; Gautschi, 2009; Pandel, 2015), in anderen das «historische Denken», das im FUER-Strukturmodell verstanden wird als ein kritischer und reflektierter Umgang mit vorliegenden Narrativen («De-Konstruktion») und als quellenbasierte Erstellung einer eigenen Narration («Re-Konstruktion»). Einig sind sich die Kompetenzmodelle darin, dass nicht die Reproduktion von Fakten, sondern die Verknüpfung von Ereignissen zu einer sinnvollen Erzählung im Zentrum stehen. Sprachliche Fähigkeiten sind im Fach Geschichte unerlässlich, um historische Argumente zu formulieren (u. a. Mierwald & Brauch, 2015; Nitsche & Waldis, 2016), aber auch um Quellen und Darstellungen, die häufig in sprachlicher Form vorliegen, zu erschliessen (u. a. Bertram & Kolpatzik, 2019; Handro, 2013, 2015).

Üblicherweise werden in Schulleistungstudien wie PISA oder NEPS und in grossangelegten Forschungsstudien standardisierte Aufgabenformate wie z. B. Multiple-Choice-Tests eingesetzt. Ob solche Formate für die Erfassung der komplexen kognitiven Prozesse des historischen Denkens geeignet sind, wird allerdings vielerorts in Frage gestellt (Körper et al., 2008; Martinez, 1999; Smith, 2017; Smith et al. 2019; VanSledright, 2013) und es wird dafür plädiert, Ergebnisse des historischen Denkens (auch) mit offenen Aufgabenformaten zu erfassen, um eine *construct underrepresentation* (Messick, 1995) zu vermeiden. Die von der Stanford History Education Group entwickelten *History-Assessment-of-Thinking* (HAT-)Aufgaben beziehen sich beispielsweise auf eine oder zwei historische Quellen, die mit kurzen und prägnanten Fragen erschlossen werden (Breakstone et al., 2015; Smith, 2019). Im Unterricht, aber auch in der empirischen Forschung zur Kompetenzerfassung, werden die Quellen häufig mit gezielten Fragen präsentiert, um die Lernenden bei der Argumentation zu unterstützen (Mierwald & Brauch, 2015; Nitsche & Waldis, 2016). Allerdings stellt die nachvollziehbare und überprüfbare Bewertung der Schüler*innentexte eine Herausforderung dar (vgl. Schönemann et al., 2010). Um die psychometrischen Gütekriterien der Objektivität, Reliabilität und Validität in qualitativen Studien sicherzustellen, wird häufig der Nachweis einer Interrater-Reliabilität verwendet. Dies ist aufwendig und zeitintensiv, sodass sich der Einsatz offener Aufgabenformate in der Forschung meist auf eher kleinere Stichproben beschränkt und ein Einsatz in grossangelegten Schulleistungs- oder Interventionsstudien kaum möglich erscheint (Nokes & De La Paz, 2018; Radinsky, Goldman, & Pellegrino, 2015).

Daher untersuchen wir im CAHisT-Projekt (CAHisT = Computational Assessment of *Historical Thinking*) das Potenzial der maschinellen Auswertung von Schüler*innentexten. Der entscheidende Vorteil der künstlichen Intelligenz ist, dass der Computer – vorausgesetzt, er hat die richtigen Muster vorher gelernt – eine grosse Menge an qualitativen Daten auswerten kann. Wir nutzen hierfür die computerlinguistischen Methoden des *natural language processing*, die sich an der «menschlichen Sprachverarbeitung» orientieren, und untersuchen die Dimensionen der inhaltlichen Validität (*content validity*) und der sprachlichen Komplexität (*language performance*) von Schüler*innentexten im Fach Geschichte. In der Studie 1 haben wir mit den Methoden des *content assessment* gearbeitet, um Vorhersagen über die Richtigkeit der Schüler*innenantworten zu treffen.

2 Grundsätzliches zum Datenmaterial und methodischen Vorgehen

2.1 Aufgabenstellung und Datenmaterial

Im CAHisT-Projekt wurden Daten aus einer Interventionsstudie zur Wirksamkeit des Lernens mit Zeitzeug*innen verwendet (Bertram, Wagner & Trautwein, 2017). Die Intervention bestand aus einer sechsstündigen Unterrichtseinheit über die Friedliche Revolution in der Deutschen Demokratischen Republik (DDR), die darauf abzielte, die Fähigkeit des historischen Denkens der Schüler*innen zu fördern (insbesondere ihre Sach- und Methodenkompetenzen). Im Posttest bearbeiteten die Schüler*innen ($N = 962$) neben den standardisierten Testaufgaben sieben offene Aufgaben (Bertram et al., 2015). Für den de- und re-konstruierenden Zugriff auf drei Textdokumente zur Demonstration am 9. Oktober 1989 in Leipzig wurden die Aufgaben auf Basis der Sechs-Felder-Matrix des FUER-Modells (Körber et al., 2007, 863) formuliert.

Die ersten drei Aufgaben bezogen sich auf einen Schulbuchtext zum 9. Oktober 1989, den die Lernenden im Hinblick auf die drei Fokussierungen («Vergangenheit», «Geschichte» und «Gegenwart/Zukunft») de-konstruierten. Sie sollten die in dem Text genannten «Fakten» notieren (A1), interpretierende Textpassagen identifizieren (A2) und die Botschaft des Textes herausarbeiten (A3). Im zweiten Teil des Aufgabenbereichs sollten die Lernenden in vier Aufgaben die Ereignisse am 9. Oktober 1989 re-konstruieren. In zwei weiteren Quel-

len, dem «Aufruf an alle Leipziger», der von Kurt Masur u. a. direkt vor der Demonstration verlesen wurde, und dem Bericht des Zeitzeugen Christian Führer, damals Pfarrer an der Nikolaikirche, sollten sie weitere Informationen zur historischen Situation identifizieren (A4 und A5). In der folgenden Aufgabe (A6) sollten sie auf der Basis ihres gesamten Wissens zu der Friedlichen Revolution das friedliche Ende der Demonstration erklären und abschliessend in einer kreativen Schreibaufgabe die Bedeutung des historischen Ereignisses reflektieren (A7). Um den Umfang des Datenmaterials darzustellen, gibt Tabelle 1 einen Überblick über die Anzahl der Antworten, die für jede Aufgabe gesammelt wurden.

Aufgabe	Anzahl der Antworten	Wörter pro Antwort	
		<i>M</i>	<i>SD</i>
1	895	25.46	± 16.03
2	895	15.68	± 15.63
3	893	15.01	± 11.22
4	888	11.75	± 8.80
5	890	13.72	± 10.32
6	890	25.97	± 17.09
7	856	25.81	± 19.42

Tab. 1: Deskriptive Übersicht des Datenmaterials

Anmerkung: Da Schüler*innen krank waren, keine elterliche Erlaubnis hatten oder die Aufgabe nicht bearbeiten wollten (die Teilnahme an den Testungen war freiwillig), war die Anzahl der Antworten pro Aufgabe geringer als die Anzahl der Schüler*innen, die an der Studie teilgenommen haben.

2.2 Methodisches Vorgehen in der Computerlinguistik

Da der hier vorgestellte Ansatz einige Methoden der Computerlinguistik nutzt, die in der Geschichtsdidaktik wenig bekannt sind, soll zunächst die Grundidee des maschinellen Lernens vorgestellt werden. Wie «lernt» ein Computer Muster aus Textdaten, wie kommt er zu einer Bewertung und wie wird die maschinelle Bewertung überprüft? In Abbildung 1 wird der grundsätzliche Ablauf computerlinguistischer Methoden dargestellt.

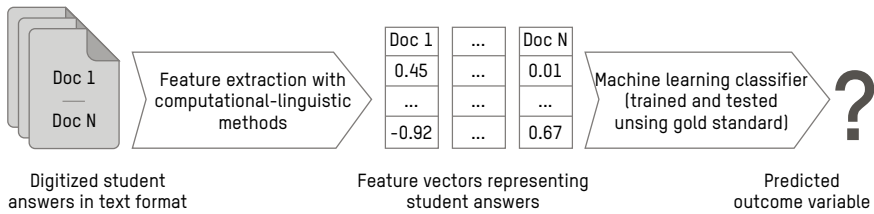


Abb. 1: Workflow für maschinelles Lernen

Der Prozess umfasst zwei Hauptschritte: (1) das Trainieren des Computers und (2) die Testung der Vorhersagen des Computers.

Im ersten Hauptschritt wird der «Trainingsdatensatz» mit den relevanten Dokumenten für die jeweilige Fragestellung und den jeweiligen Analysen «gefüttert». Wichtig ist, dass eine ausreichend grosse Anzahl von Schüler*innenantworten hinsichtlich der Ergebnisvariablen (der «Features») manuell gekennzeichnet wird. Die manuell kodierte Labels dienen als «Goldstandard»; anhand dieser Beispiele «lernt» der Computer, welche Merkmalswerte welchen Typen von Kennzeichnungen entsprechen. So kann er beispielsweise lernen, dass richtige Antworten viele Wörter enthalten, die sich mit der Zielantwort überschneiden (vgl. Studie 1), oder dass Schüler*innen bei sehr komplexen Aufgaben systematisch mehr Nebensätze pro Satz verwenden (vgl. Studie 2). Die extrahierten Features stellen die jeweiligen Antworten der Schüler*innen in Zahlen und Werten dar. Auf der Grundlage eines Merkmalsvektors für jede Schüler*innenantwort wird ein *Classifier* für maschinelles Lernen trainiert.

Im zweiten Hauptschritt beurteilt der Computer mithilfe dieses Classifiers ihm unbekannte Daten im Testdatensatz. Dieser muss ebenfalls die manuellen «Goldstandard»-Labels für die Ergebnisvariable enthalten, für deren Vorhersage der Classifier trainiert wurde. Die Leistung (Vorhersagekraft) des Classifiers wird eingeschätzt, indem die Vorhersagen, die der Classifier auf der Grundlage der Merkmalsvektoren der Schüler*innenantworten macht, mit den Goldstandard-Labels verglichen werden. Je höher der Prozentsatz der richtigen Vorhersagen (die so genannte Genauigkeit), desto besser ist die Leistung des Classifiers. Dabei wird erwartet, dass eine Klassifikation, die auf gelernten Textmerkmalen basiert, mehr korrekte Vorhersagen erzielt als zufälliges Raten (sog. *random baseline*) oder die jeweils häufigste Bezeichnung (sog. *majority baseline*).

Wichtig ist, dass der Trainingsdatensatz und der Testdatensatz nicht identisch sein und keine Datenpunkte gemeinsam haben dürfen. Gemeinsam ist al-

len Merkmalen, die aus den Texten extrahiert werden, die sich in unseren beiden Studien auf die sprachliche Komplexität oder die inhaltliche Richtigkeit der Antwort beziehen, dass die Merkmale der Antwort mit numerischen Werten ausgedrückt werden und zu einer numerischen Repräsentation der Schüler*in-entexte, aus denen sie extrahiert wurden, den sogenannten Merkmalsvektoren, zusammengefasst werden. Dabei basieren die computerlinguistischen Methoden, die die Richtigkeit des Inhalts vorhersagen, auf anderen Textmerkmalen als die der sprachlichen Komplexität. So kann beispielsweise der prozentuale Anteil an Wörtern, den eine Schüler*innenantwort mit einer vordefinierten Zielantwort gemeinsam hat, als Textmerkmal für die Bewertung des Inhalts der Antwort verwendet werden (vgl. Studie 1), während die durchschnittliche Anzahl der Nebensätze pro Satz ein Kriterium für die Bewertung der sprachlichen Komplexität darstellt (vgl. Studie 2). In den folgenden beiden Abschnitten werden der Hintergrund, das Design, die Methoden und die Ergebnisse der beiden Studien vorgestellt.

3 Studie 1 – automatische Bewertung der inhaltlichen Richtigkeit

3.1 Grundlage: Kodiersystem und Interrater-Reliabilität

Die manuelle Inhaltsbewertung wurde an einer Stichprobe von $N = 141$ Schüler*innenantworten (von den $N = 962$ Schüler*innen im Gesamtdatensatz) durchgeführt. Zur Verfügung standen: (a) manuell entworfene Zielantworten für jede Aufgabe als Referenz, (b) ein Kodiersystem auf der Grundlage der Zielantworten und (c) Bewertungen für die Schüler*innenantworten, die auf der Grundlage des Kodiersystems manuell erstellt worden waren. Die Zielantworten sowie das Kodiersystem für alle Aufgaben wurden abgeleitet aus: (1) dem Kodiersystem der manuellen Auswertung im Rahmen der Vorstudie (Bertram et al., 2015), (2) den Modellantworten einer Geschichtslehrkraft und (3) den Antworten einer Nichtexpertin. Das finale Kodiersystem war entsprechend der Aufgabenstellung zunehmend komplex. In den Aufgaben 1 bis 5 wurde pro genannten historischen Fakt bzw. pro Beobachtung jeweils ein Punkt gegeben, wenn jeweils mindestens zwei semantische Rollen (z.B. Akteur*in, Handlung oder Ort) angegeben wurden. Aus der Gesamtsumme der Punkte ergab sich die Antwortpunktzahl für die Aufgaben 1 bis 5. Bei der Bewertung der Aufgabe 6 wurde zusätzlich die Anzahl der genutzten Quellen (Lehrbuch, Appell, Augenzeugenbericht oder andere) be-

rücksichtigt (von einem Punkt, wenn nur einer, bis zu vier Punkten, wenn alle gegebenen Texte genutzt wurden). Zudem wurde das Argumentationsniveau von niedrig bis hoch (L1, L2, L3) eingeschätzt. Die Bewertung des Argumentationsniveaus erfolgte anhand der Anzahl der verwendeten Diskurskonnectoren wie z. B. Konjunktionen, die Nebensätze oder Hauptsätze oder andere Satzglieder einleiten und damit Verknüpfungen zwischen verschiedenen Aspekten und Argumenten herstellen (Göpferich & Neumann, 2016).

Zwei Rater*innen kodierten auf der Grundlage dieses Kodiersystems 141 Schüler*innenantworten. Um die interne Validität des manuellen Ratings zu testen, wurden Pearson-Korrelationen (ein numerisches Assoziationsmass) und ein (quadratisch) gewichtetes Kappa (ein Mass für geordnete Kategorien, das typischerweise bei Analysen im Computational Assessment verwendet wird) berechnet (Artstein & Poesio, 2008). Tabelle 2 zeigt die Ergebnisse der Interrater-Übereinstimmung für die manuelle Bewertung der Schüler*innenantworten in Übereinstimmung mit unserem Kodierungssystem. Die Übereinstimmung wurde für jede Aufgabe separat berechnet. Für die Aufgaben 6 und 7 wurde die Übereinstimmung separat für die verschiedenen Dimensionen berechnet.

Aufgaben	Pearson correlation	Weighted Kappa
Aufgabe 1	.90	.89
Aufgabe 2	.96	.96
Aufgabe 3	.86	.82
Aufgabe 4	.85	.84
Aufgabe 5	.90	.90
Aufgabe 6		
– Einzelargumente	.75	.75
– Anzahl genutzter Quellen	.83	.83
– Argumentationsniveau	.79	.76
Aufgabe 7		
– Einzelargumente	.88	.88
– Anzahl genutzter Quellen	.89	.88
– Argumentationsniveau	.88	.88

Tab. 2: Interrater-Übereinstimmung zwischen zwei Rater*innen bezogen auf die sieben Aufgaben

Anmerkung: Pearson-Korrelation und weighted Kappa zwischen den manuellen Ratings für jede Aufgabe einschliesslich der Subdimensionen in den Aufgaben 6 und 7. $N = 141$.

Die Ratings zeigten insgesamt eine starke positive Pearson-Korrelation, die von $r = .75$ (Aufgabe 6 Argumentationsniveau) bis $r = .96$ (Aufgabe 2) reichte. Anhand der Kappa-Werte lässt sich erkennen, dass Aufgabe 3 offenbar schwerer konsistent zu bewerten war, während die Rater*innen bei den Aufgaben 2 und 5 sogar Kappa-Werte über $.9$ erreichten, was als nahezu perfekte Übereinstimmung gilt. Zu beachten ist hierbei, dass das von uns verwendete gewichtete Kappa Unstimmigkeiten zwischen benachbarten Kategorien (z. B. 4 vs. 5 Punkte) geringfügiger berücksichtigt und nicht benachbarte Kategorien (z. B. 4 vs. 7 Punkte) exponentiell stärker gewichtet.

3.2 Methodisches Vorgehen: Das CoMiC-Verfahren

Eine besondere Herausforderung stellt in der Computerlinguistik die «Bedeutung» einer Antwort dar, denn es gibt immer eine Vielzahl von Möglichkeiten, um einen Sachverhalt zu bezeichnen. Grundsätzlich orientieren sich die Methoden des *Natural Language Processing* an den Methoden der menschlichen Sprachverarbeitung. Bei einem *Tutoring-System* werden die Antworten der Schüler*innen direkt mit vorgegebenen Referenzantworten verglichen (Dzikovska et al., 2013). Bei anderen Systemen wie dem frei verfügbaren *CoMiC-System* (Comparing Meaning in Context, Meurers et al., 2011) lernt die Maschine durch die Interpretation von Antworten, welche Antwort richtig ist. Das CoMiC-System arbeitet in drei Schritten (Ziai, 2018):

1. manuelles Annotieren von sprachlichen Einheiten (Wörter, Chunks und grammatikalische Relationen) in der Schüler*innen- und Zielantwort auf verschiedenen Abstraktionsebenen.
2. Finden von Übereinstimmungen in sprachlichen Einheiten zwischen Schüler*innen- und Zielantwort auf der Grundlage von Schritt 1.
3. Klassifizierung der Schüler*innenantwort auf der Grundlage der Anzahl und der Art der gefundenen Übereinstimmungen (siehe Tabelle 3 für eine Liste von Merkmalen) unter Verwendung eines «überwachten maschinellen Lernverfahrens» (Ziai, 2018, 167).

Den ersten Schritt, die manuelle Kodierung, haben wir oben vorgestellt. Wie findet nun der Computer die Übereinstimmungen zwischen den Antworten der Schüler*innen und der jeweiligen Zielantwort, wenn doch die gleichen Sachverhalte unterschiedlich ausgedrückt werden können? Dies soll am Beispiel des Features *Keyword Overlap* gezeigt werden. Hier wird berechnet, wie viele mög-

liche grammatikalische Relationen zwischen den Schüler*innenantworten und den Zielantworten auf der Grundlage eines sogenannten «Dependency Parse», das grammatikalische Beziehungen zwischen Wörtern identifiziert, hergestellt werden können. Wir nutzen für die Veranschaulichung ein Beispiel aus unseren Daten. Im ersten Arbeitsauftrag sollten die Schüler*innen die im Schulbuch genannten «Fakten» notieren. Das CoMiC-System vergleicht nun die Antwort des*der Schüler*in mit der korrekten Zielantwort, um zu entscheiden, ob die Schüler*innenantwort (SA) die Frage tatsächlich beantwortet oder nicht.

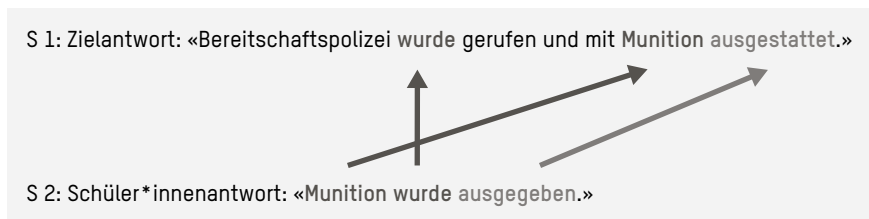


Abb. 2: Veranschaulichung des CoMiC-Systems

Anmerkung: Die Abbildung zeigt die vereinfachte Version eines Alignments einer Zielantwort und einer Schüler*innenantwort bezogen auf die erste Aufgabe.

In Abbildung 2 wird verdeutlicht, dass die Schüler*innenantwort einen Teil des Inhalts der Zielantwort ausdrückt, dass es aber Unterschiede in der Formulierung der Antwort gibt. Das Wort «Munition» ist als semantische Rolle in beiden Antworten enthalten, das Wort «ausgestattet» jedoch nicht. Ihm am nächsten kommt das Wort «ausgegeben», das in diesem Kontext als ähnlich genug eingestuft wurde. Es handelt sich hier um ein aligniertes Schlüsselwort, das zum Merkmal *Keyword Overlap* (Merkmal 1 in Tabelle 3) beiträgt. Da die Schüler*innenantwort dem durch die Zielantwort dargestellten Sachverhalt nahe genug kam und zwei semantische Rollen identifiziert werden konnten, wurde ein Punkt für diese Antwort vergeben. Tabelle 3 zeigt die numerischen Merkmale, die CoMiC aus den Alignments als Ergebnis der Übereinstimmung sprachlicher Einheiten (Schritt 2) abgeleitet hat. Im Wesentlichen handelt es sich hier um eine Reihe von Zahlen, die für jede Schüler*innenantwort extrahiert wurden und die Qualität der Antwort bezogen auf bestimmte inhaltliche Aspekte der «Zielantwort(en)» messen.

Feature	Beschreibung
1. keyword overlap	Prozentangabe der mit der Zielantwort überlappenden Keywords
2./3. token overlap	Prozentangabe der mit der Zielantwort überlappenden token (= kleinste sprachliche Einheiten)
4./5. Chunk Overlap	Prozentangabe der mit der Zielantwort übereinstimmenden syntaktischen Sinneinheiten (basierend auf einem OpenNLP)
6./7. triple overlap	Prozentangabe der mit der Zielantwort überlappenden grammatikalischen Relationen
8. «token match»	Prozentangabe der mit der Zielantwort identisch übereinstimmenden Token
9. similarity match	Prozentangabe der mit der Zielantwort übereinstimmenden Token als Ergebnis einer auf dem Textkorpus basierenden Assoziation (PMI-IR, Turney, 2001)
10. type match	Prozentangabe von mit der Zielantwort übereinstimmenden Wörtern («Typen» im Gegensatz zu «token») nach der «GermaNet hierarchy» (Hamp & Feldweg, 1997)
11. lemma match	Prozentangabe von mit der Zielantwort übereinstimmenden Token, die dieselbe Grundform haben (z. B. gehen, gingen)
12. synonym match	Prozentangabe der mit der Zielantwort übereinstimmenden Token, die nach der «GermaNet synset» dasselbe Konzept ausdrücken
13. variety of matches	Anzahl unterschiedlicher Arten von Übereinstimmungen in den Features 8 bis 12

Tab. 3: CoMiC-Machine-Learning-Features

Für jede Aufgabe extrahierten wir zunächst, wie oben erläutert, die in Tabelle 3 aufgeführten CoMiC-Merkmale und teilten dann die kodierte Stichprobe von 141 Schüler*innenantworten in 100 Antworten für das Training und 41 für den Test auf. Dies entsprach ungefähr einer 70/30-Aufteilung in Trainings- und Testdaten. Anschliessend trainierten wir einen Support-Vector-Regression(SVR)-Ansatz (unter Verwendung der Pakete kernlab und caret in R) und optimierten die Hyperparameter der SVR mit Hilfe einer 10-fachen Kreuzvalidierung auf dem Trainingssatz. Schliesslich bewerteten wir die Leistung sowohl auf der Trainingsmenge, die für die Erstellung des Modells verwendet wurde, als auch auf der ungesehenen Testmenge, um festzustellen, wie gut das Modell auf neue Daten reagiert.

3.3 Ergebnisse

Da wir die manuellen Ratings noch nicht zu einer gemeinsamen Referenz konsolidiert hatten, beschlossen wir, CoMiC für jeden Bewerter einzeln zu trainieren und zu bewerten. Insgesamt zeigt Tabelle 4, dass bei allen Aufgaben eine erhebliche Übereinstimmung besteht.

Task	Rater 1 vs. CoMiC		Rater 2 vs. CoMiC		Rater 1 vs. Rater 2
	Train	Test	Train	Test	
1	.80	.76	.89	.84	.80
2	.81	.77	.96	.76	.81
3	.69	.76	.82	.64	.69
4	.68	.59	.84	.72	.68
5	.91	.86	.90	.86	.91
6	.86	.63	.83	.68	.86
7	.77	.79	.88	.66	.88

Tab. 4: Übereinstimmung der Rater*innen 1 und 2 mit den Vorhersagen durch das CoMiC-System

Anmerkung: *Weighted Kappa* zwischen CoMiC und jedem Rating und für jede Aufgabe.

Bei den Aufgaben 1 und 5, in denen «Fakten» oft wörtlich aus dem Text übernommen wurden, erreichte unser System eine teilweise nahezu perfekte Übereinstimmung ($.76 < \text{Übereinstimmungskoeffizient} < .91$). Bei den Aufgaben 3 und 7 hingegen, in denen es um die eigenständige Bewertung ging, war der automatische Ansatz weniger erfolgreich. Dennoch war das System in der Lage, die guten Ergebnisse aus dem Trainingsset mit einem akzeptablen (und erwarteten) Leistungsabfall bei den völlig unbekanntem Testdaten zu reproduzieren; das automatische System passte sich erfolgreich an die menschlichen Bewertungen an. Die automatische Bewertung von frei geschriebenen Aufgaben ist also prinzipiell möglich, wenn auf der Basis des manuellen Ratings eine ausreichende Menge an Trainingsdaten zur Verfügung gestellt wird.

4 Studie 2 – linguistische Komplexitätsanalysen

In offenen Aufgaben wird von Schüler*innen nicht nur verlangt, inhaltlich richtige Antworten zu formulieren, sondern auch logisch und kohärent zu argumentieren. Daher untersuchten wir in der zweiten Studie den Zusammenhang zwischen der Aufgabenkomplexität und der sprachlichen Komplexität der Schüler*innentexte. Dabei gingen wir davon aus, dass insbesondere komplexere historische Denkprozesse durch die Verwendung komplexerer und vielfältigerer sprachlicher Formen ausgedrückt werden (vgl. die Untersuchung zur Aufgabenkomplexität in Geschichtslehrbüchern von Kühberger et al., 2019). Im folgenden Abschnitt stellen wir die theoretischen Grundlagen, das methodische Vorgehen und die Ergebnisse der linguistischen Komplexitätsanalysen vor.

4.1 Theoretischer Rahmen – wie wird die linguistische Komplexität erfasst?

Die sprachliche Komplexität lässt sich in die Teilbereiche der syntaktischen, lexikalischen, semantischen, morphologischen und diskursiven Komplexität unterteilen. Die *morphologische Komplexität* zeigt sich beispielsweise an dem Grad der Nominalisierung, u. a. an dem Verhältnis der abgeleiteten Substantive pro Substantiv. Ein anderes Beispiel – diesmal für das Mass der *lexikalischen Komplexität* – ist das Typ-Token-Verhältnis, in dem die einzigartigen Wortformen in einem Text (die Typen) in ein Verhältnis zu der Gesamtzahl der Wörter (den Token) gesetzt wird. Als ein Mass für die *syntaktische Komplexität* hingegen gilt das Verhältnis von Nebensätzen pro Satz. Tabelle 5 zeigt jeweils ein Beispielmass («Feature») für die verschiedenen Arten von sprachlicher Komplexität. Insgesamt wurde in unserer Studie mit 295 Features aus sieben linguistischen Bereichen gearbeitet. Wir beschränken uns auf jeweils ein Beispiel pro Bereich und verweisen bei Interesse auf weiterführende Literatur.

Linguistische Bereiche	Beispiel-Features	Literatur
Satzbaukomplexität	Nebensätze pro Satz	Kyle (2016); Wolfe-Quintero et al. (1998)
Phrasale Komplexität	Verbmodifikatoren pro Verb	Paquot (2019); Kyle (2016)
Lexikalische Komplexität	Type-Token-Verhältnis	Kyle (2019); McCarthy & Jarvis (2010)
Morphologische Komplexität	Abgeleitete Substantive pro Substantiv	Brezina & Pallotti (2019); De Clercq & Housen (2019)
Diskursive Komplexität	Konnektive pro Satz	Crossley et al. (2019); Barzilay & Lapata (2008)
Sprachgebrauch	Durchschnittliche Häufigkeit von Lemmata	Kyle (2019); Chen & Meurers (2018)
Kognitive Verarbeitung	Mittlere kognitive Integrationskosten	Weiss & Meurers (2019); Shain et al. (2016)

Tab. 5: Bereiche und Beispielfeatures der sprachlichen Komplexität

Hinsichtlich des Zusammenhangs zwischen der sprachlichen Komplexität und der Aufgabenkomplexität gehen wir davon aus, dass eine angemessene sprachliche Komplexität umso wichtiger wird, je komplexer die Aufgabe ist. Dies soll am folgenden Beispiel aus dem Datenmaterial deutlich werden, das sich auf die Aufgabe 6 bezieht, in der eine Antwort darauf gegeben wird, warum die Demonstration am 9. Oktober 1989 in Leipzig friedlich endete.

«Indem die DDR-Bürger nicht gewalttätig wurden, taten sie genau das Gegenteil von dem, was von ihnen erwartet wurde. Das hatte die DDR-Regierung nicht erwartet. Besonders die Sache mit den Kerzen, die Hoffnung für alle zeigen sollten, war sehr überraschend und brachte viele zum Nachdenken.»

Im ersten Satz wird der Gewaltverzicht des Volkes als Ursache für die Nichteinmischung der Regierung durch den kausalen Konnektor «indem» benannt. Die Argumentation über die Wirkung der mit den Kerzen ausgedrückten Symbolik wird durch einen Relativsatz realisiert, der die Substantivphrase «die Kerzen» modifiziert. Wir sehen also, dass die komplexe Aufgabe, in der nach einem Sinnzusammenhang gefragt wurde, anspruchsvolle sprachliche Formen wie Relativsätze, Substantivkonstruktionen oder Konnektoren erforderlich macht, die einen Zusammenhang zwischen den Ereignissen herstellen. Daher gehen wir in

der zweiten Studie der Frage nach, ob die sprachliche Komplexität der Schüler*innenantworten die Komplexität der geforderten zugrundeliegenden Prozesse des historischen Denkens widerspiegelt.

4.2 Methodisches Vorgehen

Zunächst wurden die sieben Aufgaben hinsichtlich der Komplexität des historischen Denkens in drei Kategorien eingeordnet: Die geringste Komplexität wurde der Identifikation von Sachinformationen (C₁; Aufgaben 1, 4 und 5) zugeschrieben, die weniger historische Kompetenz erfordert als das Erschliessen der implizit vermittelten Botschaft einer Quelle (C₂; Aufgaben 2 und 3). Die Aufgaben, bei denen die Schüler*innen über den friedlichen Ausgang der Demonstration in Leipzig (Aufgabe 6) und die heutige Relevanz der Erinnerung an die Demonstration (Aufgabe 7) nachdenken sollten, erhielten den höchsten Rang (C₃; Aufgaben 6 und 7). Diesem manuell vorgenommenen Rating stand das maschinelle Rating gegenüber. Der Algorithmus schätzte aus allen Schüler*innenantworten ($N = 5.390$) mithilfe des Systems von Weiss und Meurers (2018) die sprachliche Komplexität auf der Basis von insgesamt 295 aus dem Material extrahierten linguistischen Massen aus den Bereichen der Syntax, Lexik, Morphologie, des Diskurses, Sprachgebrauchs und der kognitiven Verarbeitung.

Um die Aufgabenkomplexität aus den Antworten der Schüler*innen vorherzusagen, wurde in der Software R ein Machine-Learning-Klassifikationsalgorithmus trainiert. Dazu wurde aus allen Schüler*innenantworten eine ausgewogene Teilmenge hinsichtlich der Aufgabenkomplexität erstellt, indem 1.500 Antworten der Komplexität C₁ zufällig ausgewählt wurden. Dies war notwendig, da die vollständigen Daten 2.398 Antworten auf Aufgaben mit der Komplexität C₁, aber nur 1.467 Antworten auf Aufgaben mit der Komplexität C₂ und 1.525 Antworten auf Aufgaben mit der Komplexität C₃ enthielten, was jedes auf diesen Daten trainierte Modell verfälschen würde. Nach dem Entfernen von Ausreißern ergab dies einen Arbeitsdatensatz mit $N = 4.403$ Antworten. Anschliessend wurden alle Komplexitätsmerkmale, die keine ausreichende Variabilität über den ausgewogenen Satz von Schüler*innenantworten aufwiesen, entfernt, indem verlangt wurde, dass der häufigste Merkmalswert in höchstens 80 % der Daten vorkommt. Dadurch verringerte sich die Anzahl der Merkmale auf 187, für die wir ihre jeweiligen Z-Scores berechneten.

Schliesslich wurden die Daten im Verhältnis 70/30 in einen Trainingsatz ($N = 3.084$) und einen Testsatz ($N = 1.319$) aufgeteilt. Wir trainierten eine Sup-

port Vector Machine (SVM) mit einem Polynom-Kernel (unter Verwendung der R-Pakete `caret`, `randomForest`, `kernlab` und `e1017`) auf der Trainingsmenge unter Verwendung einer 10-fachen Kreuzvalidierung zur Abstimmung der Hyperparameter. Die endgültigen Hyperparameterwerte waren Polynomgrad = 2, Skala = 0,01 und Kosten = 0,25. Um die Klassifizierungsleistung dieses Classifiers zu bewerten, berechneten wir die Genauigkeit unseres Modells auf den Testdaten und verglichen sie mit der Majority-Baseline, d. h. dem Prozentsatz richtiger Vorhersagen, den wir erhalten würden, wenn wir immer die häufigste Aufgabenkomplexität im Testsatz vorhersagen würden. Um weitere Erkenntnisse darüber zu gewinnen, welche linguistischen Merkmale sich über die verschiedenen Aufgabenkomplexitäten hinweg unterscheiden, haben wir die 187 Komplexitätsmerkmale nach ihrem durchschnittlichen Informationsgewinn eingestuft, wiederum unter Verwendung einer 10-fachen Kreuzvalidierung, um die informativsten Masse für die Identifizierung der Aufgabenkomplexität über die verschiedenen linguistischen Komplexitätsdomänen hinweg zu ermitteln (vgl. Abbildung 3).

4.3 Ergebnisse

Die Majority-Baseline, die die am häufigsten vorkommende Aufgabenkomplexität in den Testdaten (C_3) vorhersagt, ergab eine durchschnittliche Genauigkeit von 34,12 %, während unser Klassifikationsmodell eine durchschnittliche Genauigkeit von 85,37 % ergab und damit die Majority-Baseline weit übertraf. Die Konfusionsmatrix der Klassifikation in Tabelle 6 zeigt die Übereinstimmung der menschlichen Einschätzung der Aufgabenkomplexität (C_1 , C_2 , C_3) mit der Einschätzung des Algorithmus auf Basis der linguistischen Features. Die Diagonale zeigt die korrekt klassifizierten Textstellen über die Komplexitätsstufen der Aufgabe hinweg.

		Manuelle Einschätzung			
		C1	C2	C3	Sum
SVM- Classifier	C1	383	29	25	437
	C2	25	358	40	423
	C3	33	41	385	459
Sum		441	428	450	1.319

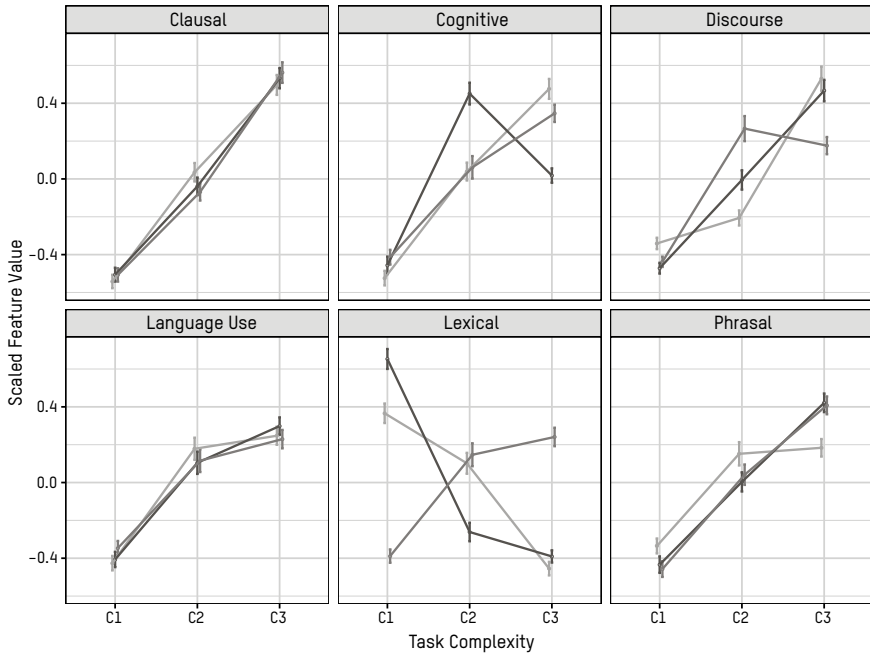
Tab. 6: Komplexität der Aufgaben und der Antworten (Ergebnisse als Konfusionsmatrix)

Anmerkung: Konfusionsmatrix für die Support Vector Machine Classifier (SVM-Classifler) auf den Testdaten ($N = 1.319$). Die übereinstimmende Vorhersage des SMV-Classifiers und der manuellen Einschätzung für die einzelnen Aufgabenkomplexitätsstufen wird durch Fettdruck hervorgehoben.

Die meisten Schüler*innenantworten wurden von dem Modell richtig klassifiziert. Fehlklassifikationen traten meist auf benachbarten Niveaus auf, mit Ausnahme der Antworten auf Aufgaben der Komplexität C1. Eine genauere Betrachtung ergab, dass die beobachteten Fehlklassifikationen zwischen den nicht benachbarten Aufgabenkomplexen C1 und C3 durch die Ähnlichkeit der Antworten auf die Aufgaben A5 (C1) und A6 (C3), die sich aufgrund der zugrunde liegenden Inhalte konzeptionell ähnlich waren, erklärt werden. In Aufgabe 5 notierten die Schüler*innen neue Informationen aus dem Bericht des Zeitzeugen Christian Führer, Pfarrer der Nikolaikirche in Leipzig, der die unerwarteten «Kerzen statt Steine» als Charakteristikum der Demonstration am 9. Oktober 1989 eindrücklich beschreibt. Als die Schüler*innen in der sechsten Aufgabe gefragt wurden, warum die Demonstration friedlich endete, erklärten dies viele mit den Kerzen in den Händen der Menschen, die ihre friedlichen Absichten zeigten. Alles in allem spiegelte also die sprachliche Komplexität der Antworten die Komplexität der Aufgaben wider.

Im nächsten Schritt untersuchten wir, welche sprachlichen Merkmale den höchsten Informationsgewinn hatten, d. h. welche Merkmale am informativsten waren, um die Komplexitätsstufen der Aufgaben zu unterscheiden. Wir entschieden uns dafür, die drei informativsten Merkmale aus jedem der linguistischen Komplexitätsbereiche genauer zu betrachten. Die morphologische Komplexität schlossen wir aus, da keines ihrer Merkmale informativ für die Unterscheidung zwischen den Ebenen der Aufgabenkomplexität war. Damit blieben insgesamt 18 Merkmale der sprachlichen Komplexität übrig, die wir genauer untersuchten. Wie gut diese Merkmale die Komplexität der Aufgabe er-

klären können, wird in Abbildung 3 dargestellt. Auf der x-Achse ist die Aufgabenkomplexität abgetragen, auf der y-Achse die Erklärungskraft des jeweiligen linguistischen Merkmals, sodass sich eine Verlaufslinie über die drei Komplexitätsstufen ziehen lässt.



Feature

- Clausal: Clauses per sentence
- Clausal: Verbs per sentence
- Clausal: Words per sentence
- Cognitive: Longest dependency per sentence
- Cognitive: PID
- Cognitive: Total IC at finite verb
- Discourse: Causal connectives per sentence
- Discourse: One-word connectives per sentence
- Discourse: Pronouns per noun
- Language Use: Subtlex-DE: average type frequency
- Language Use: dlexDB: average lemma frequency
- Language Use: dlexDB: percentage of frequent words
- Lexical: Lexical TTR
- Lexical: Nouns per word
- Lexical: Verbs per noun
- Phrasal: Adj. and adv. modifiers per verb modifier
- Phrasal: Coverage of verb modifier types
- Phrasal: Mean number of verb dependents (excl. modals)

Abb. 3: Verlauf der aussagekräftigsten Masse aller informativen Merkmalsgruppen bezogen auf die Aufgabenkomplexität

Anmerkung: Wir haben die Z-Scores für alle Komplexitätsmerkmale berechnet, um vergleichbare Merkmalswerte zu erhalten. Die Diagramme zeigen die Mittelwerte der Merkmale über die Aufgabenkomplexität hinweg. Die Balken geben 95 %-Konfidenzintervalle an.

Drei Beobachtungen möchten wir herausstellen: (1) Zum einen sehen wir, dass die komplexeren Aufgaben mit einer komplexeren Sprache insbesondere auf der syntaktischen und der diskursiven Ebene einhergingen. So zeigte sich bezogen auf die Aussagekraft der Satzkomplexität (*Clausal*), dass die drei untersuchten Merkmale (Verhältnis zwischen Nebensätzen, Verben und Worten jeweils pro Satz) positiv mit der Aufgabenkomplexität korrelierten. Dies ging einher mit der vermehrten Verwendung von Konnektiven, mit steigenden Anforderungen hinsichtlich der Abhängigkeiten und mit einer zunehmenden Elaboration und Variation bei der phrasalen Komplexität (z. B. Gebrauch von Adjektiven oder Modifikation von Verben). (2) Zudem zeigte sich, dass bei Antworten auf komplexere Aufgaben wesentlich häufiger gebräuchlichere Worte verwendet wurden als bei weniger komplexen Aufgaben (Sprachgebrauch) und dass die lexikalische Vielfalt in Form des Type-Token-Verhältnisses bei höherer Aufgabenkomplexität (lexikalische Komplexität) abnahm. Dies könnte damit zusammenhängen, dass die Schüler*innen bei Aufgaben, in denen sie freier schrieben, eher eigene Formulierungen nutzten, wohingegen Aufgaben der Klasse C1 häufig durch Zitate aus dem Quellenmaterial beantwortet wurden. (3) Darüber hinaus stellten wir fest, dass das Schreiben der Schüler*innen mit zunehmender Aufgabenkomplexität ko-referenzieller wurde. Dies zeigte sich u. a. an dem Verhältnis von Pronomen zu Substantiven, an der propositionalen Ideendichte, am Verhältnis von Verben zu Substantiven und an der Zunahme der Pronomen im Vergleich zu den Substantiven.

5 Diskussion

In den beiden Studien wurde untersucht, ob computerlinguistische Methoden bei der inhaltlichen und sprachlichen Bewertung von aufgabenbezogenen Schüler*innentexten unterstützen können. Die sieben Aufgaben bezogen sich auf drei historische Materialien, deren Aussagegehalt in den ersten fünf Aufgaben gesichert wurde. In den beiden abschliessenden Aufgaben sollten die Lernenden ihr eigenes Sach- und Werturteil zu Verlauf und Bedeutung der Demonstration am 9. Oktober 1989 in Leipzig im Kontext der Friedlichen Revolution begründen. Während die ersten fünf Aufgaben textnah anhand der gegebenen Materialien beantwortet wurden, zielten die letzten beiden mit dem Sach- und Werturteil zu einem Phänomen aus dem «Universum des Historischen» (Gautschi, 2009) auf den Kern des historischen Denkens.

In der ersten Studie zur inhaltlichen Richtigkeit wurde die Güte der Antworten, die wir als historische «Narrationen» verstanden haben, von zwei Rater*innen auf der Basis eines detaillierten Codesystems beurteilt. Mit diesem Kodiersystem mit Musterantworten und manuell konnotierten Antworten wurde das CoMiC-System trainiert. In den Testdaten zeigte sich, dass die maschinelle und menschliche Bewertung der Aufgaben, die durch Angaben im Text beantwortet werden konnten (Aufgaben 1 bis 5), in höherem Masse übereinstimmten als die beurteilenden und bewertenden Aufgaben (Aufgabe 6 und 7), doch ist der Grad an Übereinstimmung zwischen dem manuellen und maschinellen Rating doch bemerkenswert (Aufgabe 7: Rater 1 und CoMiC: weighted Kappa von .79; Rater 1 und Rater 2: weighted Kappa von .88). Prinzipiell scheint es also möglich zu sein, offene Aufgaben hinsichtlich ihrer inhaltlichen Validität automatisch zu bewerten – vorausgesetzt, ein ausgefeiltes Kodiersystem liegt vor und es wurden ausreichend viele Antworten manuell annotiert. In der zweiten Studie wurde die sprachliche Komplexität der Schüler*innenantworten untersucht. Hier zeigte sich, dass es grundsätzlich möglich ist, die Aufgabenkomplexität anhand der sprachlichen Komplexität einer Schüler*innenantwort vorherzusagen, und dass die Antworten mit zunehmender Aufgabenkomplexität kohärenter und syntaktisch komplexer werden. Allerdings lassen diese Auswertungen, die sich auf den Gesamtdatensatz (N = 962 Schüler*innen) beziehen, (noch) keine Rückschlüsse auf den Zusammenhang zwischen der sprachlichen Komplexität der Antwort und der Passung zum jeweiligen Inhalt zu. Das Computermodell «versteht» die Antworten nicht und kann die «Angemessenheit» bezogen auf die Aufgabenstellung nicht eigenständig beurteilen. Hier zeigt sich eine grundsätzliche Herausforderung beim Umgang mit den Methoden des maschinellen Lernens. In Anbetracht der zukünftigen Bedeutung der künstlichen Intelligenz für die Bildungsforschung (Luan et al., 2020) – möglicherweise auch für die Geschichtsdidaktik – sollen abschliessend die Ergebnisse in den beiden Studien hinsichtlich der Chancen und Risiken, die mit der computerbasierten Einschätzung von offenen Schüler*innenantworten verbunden sind, reflektiert werden.

Ein Vorteil des maschinellen Lernens besteht darin, dass nach einer anfänglichen Investition in die Entwicklung des Kodiersystems und in die Kodierung durch menschliche Rater*innen mit dem hier vorgestellten Verfahren offene Aufgaben auf eine grosse Menge von Antworten hochskaliert werden können. Damit könnten in grossangelegten Studien im Fach Geschichte, in denen derzeit aus ökonomischen Gründen ausschliesslich standardisierte Testformate eingesetzt werden (z. B. Bertram et al., 2022), in Zukunft auch offene Aufgaben

verwendet werden, die Sach- und Werturteile erfassen können. Allerdings ist die Kritik, die an standardisierten Instrumenten geübt wird, zum Teil auch auf die computerbasierte Auswertung übertragbar. Denn eine prinzipielle Einschränkung der Methode des «maschinellen Lernens» besteht darin, dass ein Computer nichts selbstständig «weiss» oder «verstehet», sondern lediglich mit vorhandenen Daten trainiert wird. In der ersten Studie wurde der Datensatz auf das CoMiC-System trainiert; im Testlauf schätzte das Modell nicht die inhaltliche Richtigkeit der Antwort selbstständig ein, sondern suchte (nur) nach Ähnlichkeiten mit dem Trainingsdatensatz. Das hierfür notwendige Kodierungssystem, der «Goldstandard», lässt allerdings wenig Spielraum für Interpretationen, denn die Rater*innen müssen sich einig sein, wie eine Schüler*innenantwort zu bewerten ist, damit der Computer die Daten präzise bewerten kann. Daher sind auch offene Aufgaben, die computerbasiert ausgewertet werden, einem gewissen Richtigkeitsstandard (vgl. Eid et al., 2010) verpflichtet. Zudem hat das CoMiC-Modell in den Trainingsdaten bezogen auf die offenen Aufgaben zwar eine Vielzahl potenziell zutreffender Antworten kennengelernt, doch unerwartete Antworten, in denen beispielsweise mit einem ungewöhnlichen historischen Vergleich argumentiert wird, der in dem Trainingsdatensatz nicht genannt wurde, oder auch die Antworten von Overachievern, die weit über den Rahmen der erwarteten Antwort hinausgehen, können von dem Modell eventuell nicht zutreffend eingeschätzt werden. Somit bergen automatisierte Bewertungsverfahren das Risiko, dass auf gültige, vielleicht zum Nachdenken anregende Antworten, die ausserhalb des Spektrums der erwarteten Antworten liegen, eventuell nicht angemessen eingegangen werden kann.

Das (vorläufige) Fazit für den Umgang mit den Methoden des maschinellen Lernens lautet daher, dass regelmässige Überprüfungen des maschinellen Ratings durch menschliche Kontrollen notwendig sind, beispielsweise indem Antworten, die zu einem erheblichen Teil von der erwarteten Antwort abweichen, für eine menschliche Überprüfung gekennzeichnet werden.² Generell sollte die Bedeutung der menschlichen Kontrolle beim Einsatz der Methoden des maschinellen Lernens anerkannt werden, um zu gewährleisten, dass z. B. unbeabsichtigte Verzerrungen im System erkannt und unklare Fälle oder Fälle mit hohem Risiko überprüft werden. Unter Berücksichtigung dieser Einschränkungen zei-

2 Für die Zwecke der vorliegenden Studie war dies kein praktisches Problem, da keine der manuell kodierten Schüler*innenantworten ($N = 331$) über die erwarteten Antworten hinausging, doch könnte dies in den übrigen Daten ($N = 631$) der Fall sein.

gen die präsentierten Ergebnisse, dass qualitative Methoden der Kompetenzerfassung und quantitative computerlinguistische Analysen kombiniert werden können, sodass damit die Möglichkeit eröffnet wird, zukünftig offene Aufgabenformate, die auch historische Sach- und Werturteile umfassen, in grossangelegten Studien einzusetzen und computerbasiert auszuwerten.

Literatur

- Artstein, Ron & Poesio, Massimo (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4), 555–596. <https://doi.org/10.1162/coli.07-034-R2>
- Barricelli, Michele (2005). *Schüler erzählen Geschichte. Narrative Kompetenz im Geschichtsunterricht. Forum Historisches Lernen*. Schwabach/Ts.: Wochenschau.
- 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* (S. 165–180). Bern: hep.
- 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>
- Bertram, Christiane & Kolpatzik, Andrea (Hrsg.) (2019). *Sprachsensibler Geschichtsunterricht. Von der Theorie über die Empirie zur Pragmatik*. Frankfurt a. M.: Wochenschau.
- Bertram, Christiane / Wagner, Wolfgang / Kühberger, Christoph / Hülsen, Karin / Einhaus, Nadja / Stello, Benjamin & Trautwein, Ulrich (2022). *Wie gut können Schüler:innen historisch denken? Auf dem Weg zu empirisch verankerten Bildungsstandards im Fach Geschichte. Ein Projekt des IQSH mit der HiTCH-Gruppe*. Kiel: Institut für Schulqualität in Schleswig-Holstein (IQSH).
- Breakstone, Joe / Wineburg, Sam & Smith, Mark (2015). Formative assessment using library of congress documents. *Social Education*, 79(4), 178–182.
- Brezina, Vaclav & Pallotti, Gabriele (2019). Morphological complexity in written L2 texts. *Second Language Research*, 35(1), 99–119. <https://doi.org/10.1177/0267658316643125>
- Chen, Xiaobin & Meurers, Detmar (2018). Worthäufigkeit und Lesbarkeit: Predicting the text-level Readability with a lexical-level attribute. *Journal of Research in Reading*, 41(3), 486–510. <https://doi.org/10.1111/1467-9817.12121>
- Clercq, Bastien de & Housen, Alex (2019). The development of morphological complexity: A cross linguistic study of L2 French and English. *Second Language Research*, 35(1), 71–97. <https://doi.org/10.1177/0267658316674506>

- Crossley, Scott A. / Kyle, Kristofer & Dascalu, Mihai (2019). The tool for the automatic analysis of cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods*, 51, 14–27. <https://doi.org/10.3758/s13428-018-1142-4>
- Danto, Arthur C. (1980). *Analytische Philosophie der Geschichte*. Frankfurt a. M.: Suhrkamp (Original erschienen 1965).
- Dzikovska, Myroslava O. / Nielsen, Rodney D. / Brew, Chris / Leacock, Claudia / Giampiccolo, Danilo / Bentivogli, Luisa / ... Dang, Hoa T. (2013). SemEval-2013 task 7: The joint student response analysis and 8th recognizing textual entailment challenge. In Suresh Manandhar, & Deniz Yuret (Hrsg.), *Second joint conference on lexical and computational semantics (*SEM), volume 2: Proceedings of the seventh international workshop on semantic evaluation (SemEval 2013)* (S. 263–274). Atlanta: Association for Computational Linguistics. Abgerufen von <https://www.aclweb.org/anthology/volumes/S13-2/>
- Eid, Michael, Gollwitzer, Mario & Schmitt, Manfred (2010). *Statistik und Forschungsmethoden. Lehrbuch*. Weinheim: Beltz.
- Gautschi, Peter (2009). *Guter Geschichtsunterricht: Grundlagen, Erkenntnisse, Hinweise*. Schwalbach/Ts.: Wochenschau.
- Göpferich, Susanne & Neumann, Imke (2016). Writing competence profiles as an assessment grid? – Students' L1 and L2 writing competences and their development after one semester of instruction. In Susanne Göpferich & Imke Neumann (Hrsg.), *Developing and assessing academic and professional writing skills* (S. 103–140). Peter Lang. <https://doi.org/10.3726/978-3-653-06614-2>
- Handro, Saskia (2013). Sprache und historisches Lernen. Dimensionen eines Schlüsselproblems des Geschichtsunterrichts. In M. Becker-Mrotzek, K. Schramm, E. Thürmann, & H. J. Vollmer (Hrsg.), *Sprache im Fach. Sprachlichkeit und fachliches Lernen* (S. 317–334). Waxmann: Münster.
- Handro, Saskia (2015). Sprache(n) und historisches Lernen. Zur Einführung. *Zeitschrift für Geschichtsdidaktik*, 14, 5–24.
- Housen, Alex / Kuiken, Folkert & Vedder, Ineke (2012a). Complexity, accuracy and fluency: Definitions, measurement and research. In Alex Housen, Folkert Kuiken, & Ineke Vedder (Hrsg.), *Dimensions of L2 performance and proficiency: Complexity, accuracy in SLA* (S. 1–20). John Benjamins Publishing Company. <https://doi.org/10.1075/llt.32.01hou>
- Housen, Alex / Kuiken, Folkert & Vedder, Ineke (2012b). *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins. <https://doi.org/10.1075/llt.32>

- Jeismann, Karl-Ernst (1977). Didaktik der Geschichte. Die Wissenschaft von Zustand, Funktion und Veränderung geschichtlicher Vorstellungen im Selbstverständnis der Gegenwart. In Erich Kosthorst (Hrsg.), *Geschichtswissenschaft. Didaktik – Forschung – Theorie* (S. 9–33). Göttingen: Vandenhoeck & Ruprecht.
- Körber, Andreas / Schreiber, Waltraud & Schöner, Alexander (2007). *Kompetenzen historischen Denkens: Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik*. Neuried: ars una. Abgerufen von <http://edoc.ku-eichstaett.de/1715/>
- Körber, Andreas / Borries, Bodo von / Pflüger, Christine / Schreiber, Waltraud & Ziegler, Béatrice (2008). Sind Kompetenzen historischen Denkens messbar? In Volker Frederking (Hrsg.), *Schwer messbare Kompetenzen. Herausforderungen für die empirische Fachdidaktik* (S. 65–84). Baltmannsweiler: Schneider.
- Kühberger, Christoph / Bramann, Christoph / Weiss, Zarah & Meurers, Detmar (2019). Task complexity in history textbooks: A multidisciplinary case study on triangulation in history education research. *History Education Research Journal*, 16(1), 139–57. <https://doi.org/10.18546/HERJ.16.1.12>
- Kyle, Kristofer (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* (Dissertation). Georgia State University, Atlanta. Abgerufen von https://scholarworks.gsu.edu/alesl_diss/35/
- Kyle, Kristofer (2019). Measuring lexical richness. In S. Webb (Hrsg.), *The Routledge handbook of vocabulary studies*, (S. 454–476). Routledge. <https://doi.org/10.4324/9780429291586>
- Luan, Hui, / Lai, Hollis / Gobert, Janice / Yang, Stephen / Ogata, Hiroaki / Baltés, Jackey / ... Tsai, Chin-Chung (2020). Challenges and future directions of big data and artificial intelligence in education. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.580820>
- Martinez, Michael E. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34, 207–218. https://doi.org/10.1207/s15326985ep3404_2
- McCarthy, Philip M. & Jarvis, Scott (2010). MTL-D, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods* 42(2), 381–392. <https://doi.org/10.3758/BRM.42.2.381>
- Meurers, Detmar / Ziai, Ramon / Ott, Niels & Kopp, Janina (2011). Evaluating answers to reading comprehension questions in context: Results for German and the role of information structure. In Sebastian Padó & Stefan Thater (Hrsg.), *Proceedings of the TextInfer 2011 workshop on textual entailment* (S. 1–9). Association for Computational Linguistics. Abgerufen von <https://www.aclweb.org/anthology/W11-2401>

- Messick, Samuel (1995). Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- Mierwald, Marcel & Brauch, Nicola (2015). Historisches Argumentieren als Ausdruck historischen Denkens. Theoretische Fundierung und empirische Annäherungen. *Zeitschrift für Geschichtsdidaktik*, 14, 104–120. doi: 10.13109/zfgd.2015.14e.1.104
- 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), S. 15–31.
- Nokes, Jeffery D. & De La Paz, Susan (2018). Writing and argumentation in history education. In S. A. Metzger & L. M. Harris (Hrsg.), *The Wiley international handbook of history teaching and learning* (S. 551–578). John Wiley & Sons. <https://doi.org/10.1002/9781119100812.ch21>
- Pandel, Hans-Jürgen (2005). *Geschichtsunterricht nach PISA. Kompetenzen, Bildungsstandards und Kerncurricula*. Schwalbach/Ts.: Wochenschau Verlag.
- Paquot, Magali (2019). Die phraseologische Dimension in der zwischensprachlichen Komplexitätsforschung. *Second Language Research*, 35(1), 121–145. <https://doi.org/10.1177/0267658317694221>
- Radinsky, Josh / Goldman, Susan R. & Pellegrino, James W. (2015). Commentary historical thinking: In search of conceptual and practical guidance for the design and use of assessments of student competence. In K. Ercikan & P. Seixas (Hrsg.), *New directions in assessing historical thinking* (S. 89–101). Routledge. <https://doi.org/10.4324/9781315779539>
- Ricœur, Paul (1988). *Zeit und Erzählung*. (Bd. 1). München: Fink. (Original erschienen 1983).
- Schönemann, Bernd, Thünemann, Holger & Zülsdorf-Kersting, Meik (2010). *Was können Abiturienten? Zugleich ein Beitrag zur Debatte über Kompetenzen und Standards im Fach Geschichte*. Berlin: LIT.
- Shain, Cory / van Schijndel, Marten / Futrell, Richard / Gibson, Edward & Schuler, William (2016). Speicherzugriff während inkrementeller Satzverarbeitung verursacht Lesezeitlatenz. In Dominique Brunato, Felice Dell'Orletta, Giorgio Venturi, Thomas François, & P. Blache (Hrsg.), *Proceedings of the workshop on computational linguistics for linguistic complexity (CL4LC)*, (S. 49–58). The COLING 2016 Organizing Committee. <https://www.aclweb.org/anthology/W16-4106/>
- Smith, Mark D. (2017). Cognitive validity: Can multiple-choice items tap historical thinking processes?. *American Educational Research Journal*, 54(6), 1256–1287. <https://doi.org/10.3102/0002831217717949>

- Smith, Mark / Breakstone, Joe & Wineburg, Sam (2019). History assessments of thinking: A validity study. *Cognition and Instruction*, 37(1), 118–144. <https://doi.org/10.1080/07370008.2018.1499646>
- VanSledright, Bruce A. (2013). *Assessing historical thinking and understanding. Innovative designs for new standards*. New York: Routledge. <https://doi.org/10.4324/9780203464632>
- Weiss, Zarah & Meurers, Detmar (2018). Modeling the readability of German targeting adults and children: An empirically broad analysis and its cross-corpus validation. In E. M. Bender, L. Derczynski, & P. Isabelle (Hrsg.), *Proceedings of the 27th international conference on computational linguistics*, (S. 303–317). Association for Computational Linguistics. Abgerufen von <https://www.aclweb.org/anthology/C18-1026/>
- Weiss, Zarah & Meurers, Detmar (2019). Broad linguistic modeling is beneficial for German L2 proficiency assessment. In A. Abel, A. Glaznieks, V. Lyding, & L. Nicolas (Hrsg.) *Widening the scope of learner corpus research: Selected papers from the fourth learner corpus research conference*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Wolfe-Quintero, Kate / Inagaki, Shunji & Kim, Hae-young (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu: University of Hawai'i Press.
- Ziai, Ramon (2018). *Short answer assessment in context: The role of information structure* (Dissertation). Universität Tübingen, Tübingen.

«In ihrer Zeitung [...] versuchten sie an alle heranzukommen und sie an ihre tolle Kultur und an sie selbst zu erinnern»

(Digitalisierte) historische Zeitungen als Quellen im Geschichtsunterricht¹

Kristopher Muckel

1 Einleitung

Waren Zeitungen in ihrer Anfangszeit im 17. Jahrhundert vorrangig eine Zuverdienstmöglichkeit für Drucker in stark frequentierten Ortschaften, entwickelte sich insbesondere im Umfeld der Krisen und Konflikte des 18. Jahrhunderts ein reges Pressewesen (Stöber, 2005, 61–77), aus dem im 19. Jahrhundert eine Massenpresse entstand, deren Produkte mit Fug und Recht die Stellung von Leitmedien für sich beanspruchten. Konnte die gedruckte Zeitung diese Rolle in der Presselandschaft, die sich nach dem Zweiten Weltkrieg in den westdeutschen Besatzungszonen bzw. der Bundesrepublik Deutschland entwickelte, seit den 1950er-Jahren zum Teil wieder einnehmen (Bösch, 2019, 6–9), geriet die traditionelle Zeitung spätestens in den 2000er-Jahren mehr und mehr unter den Druck digitaler Angebote (Fleischhacker, 2014, 135–145).

Darf der (gedruckten) Zeitung damit in der Gegenwart – wenn überhaupt – nur noch begrenzt eine prägende Rolle für den gesellschaftlichen Diskurs zugeschrieben werden, ist ihre Bedeutung für die politische und gesellschaftliche Entwicklung mindestens im 19. und 20. Jahrhundert wohl unumstritten. Inwiefern daraus auch ein Potenzial für historisches Lernen in gegenwärtigem Geschichtsunterricht erwachsen kann, ist die zentrale Frage des Projektes *Historische Zeitungen und digitale Zeitungsarchive im Geschichtsunterricht*. In dessen Rahmen werden im Folgenden bisherige geschichtsdidaktische Auseinandersetzungen mit historischen Zeitun-

¹ Das Zitat im Titel ist eine Lernendenäußerung (Dokument SaErUm04 1) im Rahmen der unter 3 beschriebenen Studie.

gen diskutiert (2). Auf Grundlage dieser Vorarbeit wird die im Projekt durchgeführte empirische Studie vorgestellt (3) und ein Teilbereich der Auswertung im Detail präsentiert (4). Abschliessend werden deren Ergebnisse diskutiert, in den Kontext des Gesamtprojektes eingeordnet und weiterführende Fragen aufgeworfen (5).

2 Die historische Zeitung in der geschichtsdidaktischen Diskussion

Die geschichtsdidaktische Auseinandersetzung mit historischen Zeitungen konzentrierte sich bislang vor allem auf aus den definitorischen Eigenschaften von Zeitungen abgeleiteten Annahmen zu ihrer Relevanz für historisches Lernen sowie auf entsprechende Praxishinweise und Unterrichtsentwürfe (Kuchler, 2016, 34–39). Zur Beschreibung der Wesensmerkmale der Zeitung haben sich die von Groth vorgeschlagenen Merkmale der Periodizität, Universalität, Aktualität und Publizität allgemein durchgesetzt (Groth, 1960, 101–102). Sie bieten gerade aus gegenwärtiger Perspektive den Vorteil, Zeitungen nicht über ihre Materialität zu definieren, sodass ihr Anwendungsbereich auch auf digital publizierte Materialien erweitert werden kann, die den Anspruch verfolgen, regelmässig über alle von ihren Leser*innen als relevant betrachtete Themen aktuell zu berichten und (zumindest potenziell) frei zugänglich zu sein.

Insgesamt deutet die bisherige geschichtsdidaktische Auseinandersetzung mit historischen Zeitungen, die insbesondere von Michael Sauer geprägt worden ist (z. B. Sauer, 2006), darauf hin, dass diese im Geschichtsunterricht insbesondere im Hinblick auf die Vermittlung von alltags- und lokalgeschichtlichen Themen einerseits sowie andererseits in Bereichen mit recht deutlich divergierenden Perspektiven, die sich in verschiedenen Zeitungen niederschlagen können, nutzbringend für historisches Lernen eingesetzt werden können. Allerdings scheint dies bislang nicht zu passieren (Kuchler, 2016, 39–43), was umso mehr erstaunt, da historische Zeitungen bereits seit geraumer Zeit in enormer Zahl frei im Netz verfügbar sind.

In der Verknüpfung von Zeitungsdigitalisaten mit digitalen Suchumgebungen liegen, so die im Weiteren verfolgte These, Potenziale für das historische Lernen, die über den beschriebenen Nutzen der historischen Zeitung im Geschichtsunterricht hinausgehen. Es wird angenommen, dass sich diese aus den heuristischen²

2 Der Begriff der Heuristik wird bereits grundlegend definiert bei Droysen (1868, 15) sowie darauf aufbauend auch von Rüsen (2013, 171).

Anforderungen ergeben, die sich stellen, wenn Lernende eigenständig Zeitungen zur Beantwortung konkreter Fragestellungen recherchieren und auswählen. So verlangt eine Recherche, die im Sinne Bernsens (2017, 314) über die bloße Suche nach Sachinformationen hinausgeht, zunächst eine grundlegende Einsicht in die Suchlogik des jeweiligen Digitalarchivs, mithin ein Verständnis der Funktionsweise der Suchfunktionen und Navigationsmöglichkeiten im entsprechenden Hypertext, als welcher das Digitalarchiv anzusehen ist (Kramersitsch, 2007, 130–147). Diese fachunspezifisch anmutende Kompetenz wird dadurch Teil eines historischen Denkprozesses, dass sie auf den historischen Gegenstand bezogen wird (Hodel, 2016, 155), also auf zentrale Aspekte der historischen Quellenkritik³ zurückgreifen muss, um die Brauchbarkeit der aufgefundenen Suchergebnisse für die Fragestellung zu überprüfen.

3 Vorstellung Studiendesign und Auswertungsinstrument

Die Überprüfung dieser theoretischen Annahmen zum Wert historischer Zeitungen und digitaler Zeitungsarchive für das historische Lernen ist das wesentliche Erkenntnisinteresse des Projektes. Es ordnet sich damit in den Kontext der Medienvergleichsstudien mit geschichtsdidaktischem Erkenntnisinteresse ein, wobei zweischrittig vorgegangen wird und in der ersten Forschungsfrage zunächst der grundsätzliche Nutzen historischer Zeitungen für das historische Lernen im Mittelpunkt steht:

Können im Vergleich zwischen Lernenden, die mit didaktisierten Quellen gearbeitet haben (Gruppe Q), und Lernenden, die mit verschiedenen Formen historischer Zeitungen gearbeitet haben (Gruppe Z), Unterschiede in Schreibprodukten zu identischen Lernaufgaben identifiziert werden?

Die Offenheit im Hinblick auf den Zeitungseinsatz verfolgt das Ziel einer zunächst grundlegenden Identifikation möglicher Potenziale des Mediums für das historische Lernen. Auf die Unterscheidung der verschiedenen Einsatzformen zielt die zweite Frage ab:

Können im Vergleich zwischen Lernenden, die mit Faksimiledrucken historischer Zeitungsausgaben (Gruppe ZA), Lernenden, die mit digitalisierten Faksimiles der gleichen Zeitungen auf Tablets (Gruppe ZD), und Lernenden, die

3 Überlegungen zu einer Quellenkritik im Digitalen, deren Notwendigkeit z.B. Wineburg (2019, 117–118) betont, befinden sich, so der Befund bei Föhr (2017, 8), noch im Anfangsstadium.

selbstständig in Digitalarchiven nach historischen Zeitungen recherchiert haben (Gruppe DA), Unterschiede in Schreibprodukten zu identischen Lernaufgaben identifiziert werden?

Um Prozesse des historischen Lernens sicht- und messbar zu machen, wurden diese anhand ihrer Manifestationen in Lernendennarrationen zu Schreibaufgaben untersucht. Dabei wurde von der Prämisse ausgegangen, dass jede Interpretation der Vergangenheit grundsätzlich dadurch eine Narration darstellt, dass sie sich von der Gegenwart aus deutend Bereichen der Vergangenheit zuwendet (Ankersmit, 1994, 33, 40) und dazu erklärende und argumentierende Elemente integriert (Rüsen, 1986, 37–42). Die Beschreibung der Lernendentexte als Narrationen umfasst somit deren gesamte Spannweite an Sprachhandlungen, die durch die deutende Ausrichtung auf die Vergangenheit historisch werden.

Die Erhebung der Lernendentexte erfolgte im Rahmen einer quasi-experimentellen⁴ Interventionsstudie⁵ im Prä-Post-Design⁶ im ersten Halbjahr des Schuljahres 2020/2021. 63 Schüler*innen in vier Lerngruppen der Jahrgangsstufe 10 eines nordrhein-westfälischen Gymnasiums haben daran teilgenommen. In zwei Pilotierungen wurden im zweiten Halbjahr des vorausgegangenen Schuljahres Aufgabenstellungen für die Schreibaufgaben, die Auswertungsinstrumente, verschiedene Möglichkeiten der Darreichung der Zeitungen an die Lernenden (Druckformate, Dateiformate, Endgeräte) mit anderen Schüler*innen als jenen in der Haupterhebung erprobt. Eine ursprünglich geplante Randomisierung der Treatmentgruppen auf Grundlage einer zu Schuljahresbeginn durchzuführenden Fragebogenerhebung und der fachlichen Leistungen der beteiligten Lernenden im vorausgegangenen Schuljahr konnte aufgrund der schulbezogenen Einschränkungen der Coronapandemie nicht umgesetzt werden. Zudem musste der Erhebungszeitraum erheblich verkürzt und die methodische Ausgestaltung der Unterrichtseinheiten drastisch auf individuelle Lernformate, die keine Platzwechsel der Lernenden im Raum erforderlich machten, eingeschränkt werden. Auch wenn die Vergleichbarkeit der Erhebungsgruppen damit nicht auf dem vorgesehenen Niveau sichergestellt werden konnte, wird

4 Zur Charakterisierung quasiexperimenteller Settings vgl. z.B. Döring und Bortz (2016, 193–194).

5 Eine allgemeine Definition der Intervention als gezielten Eingriff in ein System bietet Leutner (2012, 18), zur Spannweite von Interventionsmassnahmen im pädagogischen Feld vgl. Souvignier und van Dignath Ewijk (2010, 14).

6 Den Einsatz von Prä-Post-Designs für den Nachweis kurzfristiger treatmentbedingter Veränderungen diskutiert Hagenauer (2010, 245).

die Aussagekraft der Hauptstudie durch den Fokus auf die individuellen relativen Veränderungen in der Bearbeitung der Schreibaufgaben zwischen den Erhebungszeitpunkten sowie die Kontrolle der Störeinflüsse durch flankierende Begleitstudien abgesichert.

In deren Rahmen wurden die Schüler*innen mittels eines halboffenen Fragebogens sowohl grundsätzlich im Hinblick auf ihre Einstellung zum Fach Geschichte als auch bezüglich ihrer allgemeinen und schulischen Erfahrung mit Zeitungen befragt. Zur Sicherung der Vergleichbarkeit des Unterrichts, der von den schulischen Lehrkräften auf Grundlage von durch den Verfasser erstellten Materialien durchgeführt worden ist, wurde der Unterricht systematisch beobachtet und protokolliert. Für die Haupterhebung ergab sich schlussendlich das in Abb. 1 dargestellte Design:

Hauptstudie (geplant Mitte August bis Dezember 2020, durchgeführt Oktober bis Dezember 2020)

Vier Lerngruppen der JGS EF (10) eines Gymnasiums in NRW

<p>Vorlauf</p> <ul style="list-style-type: none"> • Thematische Einführung im Unterricht • Schulung der Lehrkräfte • Schulung der Beobachter*innen, Beginn der Unterrichtsbeobachtung • Klärung der Rahmenbedingungen 	<p>Prätest</p> <ul style="list-style-type: none"> • Erhebung der Trifftigkeit von Lernendentexten → materialgestützte Schreibaufgabe (zwei didaktisierte Quellen) • Erhebung von Störeinflüssen auf das Schreibergebnis → Halboffener Fragebogen SuS 	<p>Treatment</p> <ul style="list-style-type: none"> • Inhaltsgleiche und methodenvergleichbare, aber materialdifferenzierte Unterrichtserteilung durch die schulischen Lehrkräfte in den vier Erhebungsgruppen <ul style="list-style-type: none"> • digitale Zeitungsarchive • (ausgewählte) digitalisierte Zeitungsausgaben • (ausgewählte) Faksimiledrucke • didaktisierte Quellen • Systematische Unterrichtsbeobachtung 	<p>Posttest</p> <ul style="list-style-type: none"> • Erhebung der Trifftigkeit von Lernendentexten → materialgestützte Schreibaufgabe (auf Grundlage der Materialien des Treatments) • Erhebung von Störeinflüssen auf das Schreibergebnis → Halboffener Fragebogen SuS
<p>6 UE (45 Min.)</p>	<p>2,5 UE</p>	<p>10 UE</p>	<p>2,5 UE</p>

Abb. 1: Überblick Design Hauptstudie

Das Treatment selbst erfolgte in drei Schritten. So erhielten die Gruppen erstens eine Einführung in die Arbeit mit ihren jeweiligen Medien sowie eine grundlegende Schulung für die Verwendung multipler Dokumente für eine Schreibauf-

gabe (z. B. Goldman, 2004, 343–344).⁷ Die auf diesem Wege eingeführten Strategien zum Umgang mit historischen Zeitungen (DA/ZD/ZA), deren Digitalisaten (DA/ZD) sowie zur Recherche in Digitalarchiven (DA) wurden in einem zweiten Schritt unter Anleitung der Lehrkräfte ebenso trainiert wie jene zum Umgang mit den verschiedenen Formen didaktisierter Quellen, die in der entsprechenden Gruppe (Q) zum Einsatz gekommen sind. In diesem Kontext wurde drittens auch die inhaltliche Grundlage für die Schreibaufgabe des Posttests gelegt.

Im Folgenden wird ein Teilbereich der Auswertung der Lernendentexte im Fokus stehen. Eingebettet war die Erhebung in den Themenbereich «Fremdsein, Vielfalt und Integration – Migration am Beispiel des Ruhrgebiets im 19. und 20. Jahrhundert» des Inhaltsfeldes I des Kernlehrplans des Landes Nordrhein-Westfalen für die gymnasiale Oberstufe (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2014, 23). Da aufgrund des verkürzten Erhebungszeitraums die Einführung von für die Lernenden neuen Aufgabentypen nicht mehr realisierbar gewesen und von schulischer Seite eine Einbindung der Klausurvorbereitung in die Erhebungsphase eingefordert worden war, wurde für die Erhebungsaufgaben trotz der Kritik, die unter anderem Schönemann, Thünemann und Zülsdorf-Kersting (2011, 17–18) vorgebracht haben, auf Aufgabentypen im Sinne der Einheitlichen Prüfungsanforderungen für das Abitur zurückgegriffen (Kultusministerkonferenz, 2005, 10). Aufbauend auf einer Analyseaufgabe zu den nachfolgend genannten Quellen wurden für die beiden Testzeitpunkte die in Tabelle 1 zusammengefassten Aufgaben formuliert. Die so weit als möglich gleich gehaltenen Aufgabenstellungen sollten die entstandenen Schreibprodukte sowohl hinsichtlich ihres Umfangs als auch des Umgangs mit der Aufgabenstellung vergleichbar machen. Ein Gewöhnungseffekt wurde dafür vor dem Hintergrund in Kauf genommen, dass die erweiterte Materialgrundlage der zweiten Erhebung einen solchen abschwächt und eine bloße Reproduktion aus der Erinnerung an den ersten Erhebungszeitpunkt verhinderte.

7 Der Forschungsstand zum Einsatz multipler Dokumente als Grundlage für Schreibaufgaben kann hier nicht umfassend diskutiert werden, wird aber in der Gesamtstudie berücksichtigt.

Erster Erhebungszeitpunkt			
Material- grundlage	<ul style="list-style-type: none"> - Die Polen im rheinisch-westfälischen Steinkohlen-Bezirke, Auszug aus Gau «Ruhr und Lippe» des Alldeutschen Verbandes (1901, 55) - Übersetzungen aus dem Wiarus Polski, Nr. 1, Bochum, den 1. Januar 1901 (Institut für Zeitungsforschung, Dortmund Hz301) 		
Aufgaben- stellungen	<ol style="list-style-type: none"> 1. Erläutere zunächst kurz, was aus Sicht der Migrationsforschung unter Integration zu verstehen ist, und 2. beurteile auf dieser Grundlage, inwiefern die Bedingungen für eine solche im Ruhrgebiet gegeben waren (beziehe dich dabei explizit auf die beiden vorliegenden Quellen). 3. Bewerte abschliessend, inwiefern (aus heutiger Sicht) die Haltungen der sich selbst als deutsch und polnisch bezeichnenden Gruppierungen nachvollziehbar sind. Beziehe ggf. auch aktuelle Beispiele mit ein. 		
Zweiter Erhebungszeitpunkt			
Material- grundlage	Digitalarchive	Faksimiles historischer Zeitungen (analog bzw. als Digitalisat)	didaktisierte Quellen (vor- ausgewählte, gekürzte, kommentierte und trans- skribierte Quellen)
	<p>Recherche für den Ereigniszeitraum um 1905 in:</p> <p>Zeitungen aus dem Ruhrgebiet, z. B.</p> <ul style="list-style-type: none"> - Velberter Zeitung, - Dortmunder Zeitung; <p>Regionalzeitungen aus der Umgebung des Ruhrgebiets, z. B.</p> <ul style="list-style-type: none"> - Münsterischer Anzeiger (alle in https://zeitpunkt.nrw/); <p>reichsweit erscheinenden Zeitungen, z. B.</p> <ul style="list-style-type: none"> - Vossische Zeitung (in http://zefys.staatsbibliothek-berlin.de/) <p>Gewerkschaftszeitungen, z. B.</p> <ul style="list-style-type: none"> - Bergarbeiterzeitung (in: http://library.fes.de/gewerkschaftszeitschrift/) 	<p>Gesamtausgaben der:</p> <p>Zeitungen aus dem Ruhrgebiet:</p> <ul style="list-style-type: none"> - Velberter Zeitung (11. 01. 1905), - Dortmunder Zeitung (11. 01. 1905); <p>Regionalzeitung aus der Umgebung des Ruhrgebiets</p> <ul style="list-style-type: none"> - Münsterischer Anzeiger (11. 01. 1905) <p>reichsweit erscheinenden Zeitung</p> <ul style="list-style-type: none"> - Vossische Zeitung (12. 01. 1905 Morgenausgabe) <p>Gewerkschaftszeitung</p> <ul style="list-style-type: none"> - Bergarbeiterzeitung (14. 01. 1905) 	<ul style="list-style-type: none"> - gekürzter Quellenauszug aus Henning und Sieg (1997, 26–27) - drei gekürzte Quellenauszüge aus Tenfelde und Urban (2010, 483–485)

	Zweiter Erhebungszeitpunkt
Aufgabenstellungen	Arbeitsaufträge: <ol style="list-style-type: none">1. Erläutere zunächst noch einmal kurz, was aus Sicht der Migrationsforschung unter Integration zu verstehen ist, und2. beurteile auf dieser Grundlage, inwiefern eine solche mit Blick auf die polnischsprachigen Migranten im Ruhrgebiet stattgefunden hat (beziehe dich dabei explizit auf die bearbeiteten Quellen bzw. deine Notizen dazu).3. Bewerte abschliessend, inwiefern (aus heutiger Sicht) die Haltungen der sich selbst als deutsch und polnisch bezeichnenden Gruppierungen nachvollziehbar sind. Beziehe ggf. auch aktuelle Beispiele mit ein.

Tab. 2: Materialgrundlage und Aufgabenstellungen

Die Auswertung der auf Grundlage dieser Aufgabenstellungen entstandenen Texte erfolgte im Rahmen einer evaluativen qualitativen Inhaltsanalyse nach Kuckartz (2018, 123–140). Im Sinne einer Operationalisierung als Messbarmachung eines theoretischen Konzeptes (Döring & Bortz, 2016, 228) sind dazu zunächst Indikatoren zu benennen, anhand derer historische Denkleistungen in Texten identifiziert und ausgewertet werden können. Ausgehend von der Beschreibung der «mentalen Operationen der Sinnbildung» (Rüsen, 2013, 35) als Wahrnehmung, Deutung, Orientierung und Motivation⁸ werden auf struktureller Ebene die fünf Bereiche der Quellenbezüge, Bezüge zu externen Wissensbeständen, vergangenheitsbezogenen Deutungsleistungen sowie der gegenwarts- bzw. zukunftsbezogenen Deutungsleistungen und der Identifikation historischer Aktanten benannt, mit denen der Anspruch einhergeht, historische Denkleistungen auf der Satzebene als Kodiereinheit (Kuckartz, 2018, 41–44) erschöpfend zu erfassen, wobei Doppelcodierungen möglich und üblich sind. Sätze mit Quellenbezügen sind dabei jene, in denen Schreibende explizit oder implizit auf die zur Verfügung gestellten Quellen rekurren. In Abgrenzung davon werden unter Bezügen zu externen Wissensbeständen sämtliche Verweise auf andere unterrichtliche oder ausserunterrichtliche Informationsquellen verstanden. Deutungen bezeichnen hier Sätze, die dadurch einen narrativierenden Charakter haben, dass sie Ereignisse in der Vergangenheit sinnbildend miteinander verbinden (vergangenheitsbezogene Deutungsleistungen) bzw. vergangene Ereignisse mit der Gegenwart, z. B. mit gegenwärtigen Wertmassstäben und Positionen, verknüpfen (gegenwarts- bzw. zukunftsbezogene Deutungsleis-

8 Der Aspekt der Motivation wird hier als aussernarratives Handeln als Konsequenz des historischen Denkprozesses (Rüsen, 2013, 36) nicht weiter verfolgt.

tungen).⁹ Bei der Identifikation historischer Aktanten werden abschliessend auf Wortebene als Kodiereinheit die Träger und Ziele historischer Prozesse markiert, die von den Schreibenden benannt werden.

Die Analyse der Qualität der identifizierten Manifestationen historischer Denkleistungen erfolgt auf Grundlage eines auf den Triftigkeiten nach Rüsen basierenden Analyserasters (Rüsen, 1990, 87–91).¹⁰ Diese Kriterien, anhand derer Erzählungen im oben dargelegten Sinne hinsichtlich ihrer Zustimmungsfähigkeit und damit auch ihres narrativen Wahrheitsanspruchs überprüft werden können (Rüsen, 2013, 57–58), sind in der Geschichtsdidaktik bereits ausgiebig diskutiert worden.¹¹ Relativer Konsens herrscht hierbei bei der Definition der empirischen Triftigkeit, die sich nach Rüsen in einem «überprüfbaren Tatsachenbezug» (ebd., 60) niederschlägt, der durch die Offenlegung einerseits der der Narration zugrundeliegenden Quellen sowie ihrer Provenienz, andererseits der Auswahlentscheidungen der Quellenkritik, die zur Verwendung genau dieser Quellen geführt haben, gewährleistet wird. Neben die Auseinandersetzung mit der Perspektivgebundenheit der Quellen tritt die Reflexion sowohl des Standpunktes, von dem aus erzählt wird, als auch der leitenden «Hinsichten» (Rüsen, 1986, 104) im Sinne mehr oder weniger objektiver Beurteilungsmassstäbe, die Rüsen unter dem Aspekt der normativen Triftigkeit zusammenfasst (Rüsen, 2013, 61–62). Erst durch deren Verknüpfung mit der Gegenwart kann der Erzählung ein Sinn im Sinne einer Orientierungsfunktion für gegenwärtige Menschen zukommen (Bergmann, 2008, 144). Ebendiese bildet den Kern der bei Rüsen eher vage bleibenden Kategorie der narrativen Triftigkeit, in deren Rahmen die Gültigkeitsbedingungen von Narrationen sowie ihre Eignung für Perspektiv-erweiterungen zu untersuchen sind (Rüsen, 1983, 114).

Ausgehend von den vorgestellten Definitionen von Triftigkeit werden die so identifizierten strukturellen Merkmale historischer Denkleistungen hinsichtlich ihrer Ausprägungen evaluiert. Mit Blick auf die Treatmentgruppen besteht

9 Zum Begriff der Deutung vgl. mit direktem Bezug zu historischem Lernen Rüsen (1989, 99). Vom Begriff des Urteils wird hier Abstand genommen, da unter Deutungen auch Aussagen verstanden werden, die die Kriterien des Urteils z. B. nach Becker (2017, 322–323) (noch) nicht erfüllen.

10 Rüsen (2013, 57–62) spricht nicht mehr von Triftigkeiten, sondern von Plausibilitäten. Hier wird dennoch weiterhin der synonyme und etablierte Begriff der Triftigkeit verwendet, die synonymen Begriffsbedeutungen beschreibt z. B. Körber (2018, 22).

11 Exemplarisch für Operationalisierungen der Triftigkeitskriterien sei hier nur verwiesen auf die Arbeiten von Waldis, Marti und Nitsche (2015), Sochatzy, Schöner und Schreiber (2013) sowie Hasberg (2007).

die Grundannahme der Auswertung darin, dass die verwendeten Medien «[m]it ihren medienspezifischen Eigenschaften [...] diese Sinnbildungsprozesse» (Hodel, 2016, 155) prägten. Die Lernendertexte sind dementsprechend als Resultat der Kombination individueller historischer Denkleistungen und Medieneigenschaften zu betrachten (ebd., 157). Dem Rechercheprozess der Lernenden in den Digitalarchiven kommt dabei eine Sonderrolle zu, da es sich dabei ebenfalls um eine Kombination wie die gerade beschriebene handelt, bei der die Lernenden sowohl im Rahmen der Entscheidung für adäquate Suchbegriffe und der Auswahl zielführender Suchparameter als auch bei der Selektion der Suchresultate zumindest implizit auf Triftigkeitskriterien im beschriebenen Sinne zurückgreifen müssen (ebd., 157–158). Damit führt die Recherche die Teilbereiche des historischen Denkens und der (historischen) Medienkompetenz zusammen,¹² wobei davon ausgegangen wird, dass sich dieser ebenfalls (auch) triftigkeitsbasierte Prozess in der konkreten Ausgestaltung der Texte, die auf seiner Grundlage entstehen, so niederschlägt, dass er sich wiederum triftigkeitsanalytisch nachweisen lässt. Erwartet wird, dass u. a. der Einblick in die Vielzahl potenzieller Quellen zu einer Sensibilisierung hinsichtlich divergierender leitender Hin-sichten im Bereich der normativen Triftigkeit sowie einer vertieften Reflexion der Gültigkeit von Aussagen zu historischen Sachverhalten im Sinne der narrativen Triftigkeit beiträgt.

Die Triftigkeiten sind in Subcodes operationalisiert worden, die auf einer abgestuften Ratingskala von eins bis drei vergeben werden. Dies sei hier exemplarisch anhand des Merkmals «vergangenheitsbezogene Deutungsleistungen» entfaltet.

12 Dementsprechend beschreibt Kerber (2017, 73) in seiner Modellierung einer historischen Medienkompetenz die Recherche als Querschnittsaufgabe, die sich aus deren geschichtsdidaktischen Inhalts- sowie «Handlungsfelder[n] geschichtsunterrichtlicher Medienpädagogik» (Kerber, 2017, 53) speist.

Denkleistung	Vergangenheitsbezogene Deutungsleistung (VD)		
Triftigkeit	Prüfung der empirischen Triftigkeit	Prüfung der normativen Triftigkeit	Prüfung der narrativen Triftigkeit
Subcode (Kürzel)	Qualität vergangenheitsbezogener Deutungsleistungen (VDRelGM)	Qualität leitender Hinsichten (VLHRelGM)	Reichweite vergangenheitsbezogener Deutungsleistungen (VRRelGM)
Leitfrage	Wie intersubjektiv nachvollziehbar wird die vergangenheitsbezogene Deutungsleistung ausgeführt?	Auf welcher Grundlage werden Deutungen historischer Ereignisse vorgenommen?	Mit welchem Gültigkeitsanspruch werden vergangenheitsbezogene Deutungsleistungen formuliert?
Ratings	intransparent (1) – transparent (2) – transparent und belegt (3)	intransparent (1) – transparent (2) – wertend (3)	intransparent (1) – transparent (2) – wertend (3)
Ankerbeispiel (Dokumentchiffre Erhebungszeitpunkt) (Rating)	Theoretisch waren die Bedingungen für eine Integration gegeben, das Problem waren nur die preussischen Einwohner und die Regierung, da diese die Polen eindeutschen wollten, so dass diese ihre Kultur aufgeben. (AnWorg24 1) (2)	Eine solche Integration wurde im Zusammenhang mit den polnischen Migranten nicht angestrebt. (YvMaen20 2) (2)	In ihre Zeitung «Wiarus Polski», welche auf polnisch verfasst ist, versuchen sie heranzukommen und sie an ihre tolle Kultur und an sie selbst zu erinnern. (SaErum04 1) (1)

Tab. 3: Subcodes und Ratings zum Code «vergangenheitsbezogene Deutungsleistungen»

4 Auswertung am Beispiel des Teilbereichs «vergangenheitsbezogene Deutungsleistungen»

Die Codiererergebnisse der qualitativen Inhaltsanalyse wurden einer Inter- und Intracoderüberprüfung unterzogen. Für erstere haben zwei an der Studie beteiligte Hilfskräfte, die durch den Verfasser geschult worden sind, zehn Prozent der Gesamtstichprobe eigenständig codiert. Für die Identifikation der Manifestationen historischer Denkleistungen konnte ein κ nach Brennan und Prediger (1981, 692–694) von .90 bis .96 ($M = .92$) erreicht werden; über das gesamte Codesystem ergab sich mit $M\kappa = .79$ ein zufriedenstellende Wert (Kuckartz, 2018, 210). Für die

Intracoderüberprüfung wurden nach Abschluss des Codierprozesses acht Dokumentenpaare und damit ca. 14,5 % der Gesamtstichprobe zufällig ausgewählt und erneut codiert. Hier ergaben sich mit Werten zwischen $\kappa = .95$ und $\kappa = .98$ ($M = .97$) für die Identifikation der Manifestationen historischer Denkleistungen sowie von $M\kappa = .88$ für das gesamte Codesystem ebenfalls befriedigende Ergebnisse.

Für die weitere Auswertung sind die Resultate der qualitativen Inhaltsanalyse quantifiziert worden, um sie einer inferenzstatistischen Untersuchung zugänglich zu machen und auf diesem Wege relevante Unterschiede zwischen den Erhebungszeitpunkten und Treatmentgruppen zu identifizieren. Die absolute Anzahl der Manifestationen historischer Denkleistungen pro Text¹³ N_x wurde zur Berechnung relativer Codehäufigkeiten in Bezug zu den Kennwerten Satzzahl bzw. Wörterzahl gesetzt:

$$X_{Rel} = \frac{N_x}{N_{Satz/Wörter}}; 0 \leq X_{Rel} \leq 1$$

Zur Bestimmung der Qualität der identifizierten Elemente historischer Denkleistungen sind die relativen Häufigkeiten der einzelnen Ratings zu einem gewichteten Mittel¹⁴ zusammengefasst worden, bei dem die Ratingstufen (1–3) als Gewichte fungierten.

$$X_{RelGM} = \frac{(X_{RelR1} * 1) + (X_{RelR2} * 2) + (X_{RelR3} * 3)}{6}$$

Mit Blick auf die erste Forschungsfrage ist die Leistungsveränderung im Hinblick auf die relative Häufigkeit und die Umsetzung vergangenheitsbezogener Deutungsleistungen innerhalb der zu untersuchenden Gruppen, also denjenigen Lernenden, die mit didaktisierten Quellen gearbeitet haben ($Q, n = 14$), sowie der Gesamtheit all jener, die mit den verschiedenen Formen historischer Zeitungen gearbeitet haben ($Z, n = 41$), zu prüfen. Unter anderem Dimitrov und Rumrill

13 Im Hinblick auf die Auswertung der Absolutwerte ist die Frage zu diskutieren, inwiefern die bloße, möglicherweise redundante Anzahl der Manifestationen historischer Denkleistungen in längeren Texten einen relevanten Qualitätsunterschied im Vergleich zu kürzeren Texten mit damit auch zwangsläufig weniger entsprechenden Manifestationen darstellt. Vor diesem Hintergrund wird hier von einer Darstellung der Absolutwerte Abstand genommen, im Kontext der Gesamtstudie werden diese jedoch ausführlich berücksichtigt.

14 Die mathematischen Eigenschaften des gewichteten Mittels stellt z. B. Bullen (2003, 63) dar.

(2003, 160) schlagen insgesamt vier Vorgehensweisen für den Vergleich von Gruppen in Prätest-Posttest-Designs vor, wobei sie insbesondere die Analyse von Differenzscores sowie die Durchführung von Kovarianzanalysen (ANCOVA) als relevante Methoden herausstellen (ebenso auch bei van Breukelen, 2006, 920). Auch wenn die grossen Vorbehalte, die gerade in älterer Literatur gegen die Verwendung von Differenzscores im Hinblick auf ihre mangelhafte Reliabilität angeführt werden (Cronbach & Furby, 1970, 78), in aktuelleren Diskussionen so nicht mehr vertreten werden (z. B. Kim & Steiner, 2021, 1361–1365), wird hier insbesondere mit Blick auf die zugrundeliegenden Vorannahmen¹⁵ der ANCOVA als Messinstrument der Vorzug gerade für nicht randomisierte Studien gegeben (z. B. Dimitrov & Rumrill, 2003, 163). Daher wird für alle Vergleiche, bei denen die Voraussetzungen für eine ANCOVA (Bortz & Schuster, 2010, 311–314) erfüllt sind, dieses Verfahren ausgewählt, andernfalls wird ein Gruppenvergleichstest über die Differenzscores $t_2 - t_1$ der Variablen berechnet. Im Folgenden werden die Ergebnisse der Berechnung von ANCOVAs über die Posttestscores der Codes zu vergangenheitsbezogenen Deutungsleistungen (Innersubjektfaktor: Gruppenzugehörigkeit; Kovariate: Prätestscore) berichtet. Die unadjustierten deskriptiven Werte, die um das Prätestergebnis bereinigten Werte sowie die Resultate der ANCOVAs sind in Tabelle 3 zusammengefasst:

15 Die Vorannahmen der ANCOVA, insbesondere die Regression zur Mitte, sowie die Untersuchung von Differenzwerten, die Annahme gleicher Veränderungen bei gleichem Treatment, beschreibt z. B. van Breukelen (2006, 924–925).

Variable vergangenheitsbezogene Deutungsleistungen (Rel.) VDRel $F(1, 52) = 5.30, p = .025, \text{partielles } \eta^2 = .09$					
	unadjustiert			adjustiert	
Gruppe	<i>n</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SE</i>
Q	14	.39	.15	.41	.04
Z	41	.53	.17	.53	.03

Variable Qualität vergangenheitsbezogener Deutungsleistungen (Rel.) VDRelgM gewichtetes Mittel $F(1, 52) = 5.17, p = .027, \text{partielles } \eta^2 = .09$					
	unadjustiert			adjustiert	
Gruppe	<i>n</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SE</i>
Q	14	.13	.05	.13	.02
Z	41	.18	.06	.17	.01

Variable Qualität leitender Hintersichten (Rel.) VLHRelgM gewichtetes Mittel $F(1, 52) = 5.16, p = .025, \text{partielles } \eta^2 = .09$					
	unadjustiert			adjustiert	
Gruppe	<i>n</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SE</i>
Q	14	.12	.05	.13	.02
Z	41	.17	.07	.17	.01

Variable Reichweite vergangenheitsbezogener Deutungsleistungen (Rel.) VRRelgM gewichtetes Mittel $F(1, 52) = 4.38, p = .041, \text{partielles } \eta^2 = .08$					
	unadjustiert			adjustiert	
Gruppe	<i>n</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SE</i>
Q	14	.08	.03	.09	.01
Z	41	.12	.06	.12	.01

Tab. 4: Unadjustierte und adjustierte deskriptive Statistiken sowie Ergebnisse der ANCOVAs zum gruppenweisen (Q, Z) Vergleich der Posttestscores (Forschungsfrage 1)

Die Ergebnisse der ANCOVAs zeigen durchweg einen signifikanten Unterschied¹⁶ mit mittlerem Effekt nach Cohen (2013, 79–81) zwischen den Posttestergebnissen der Z- und jenen der Q-Gruppe, wobei die Ergebnisse der Zeitungsgruppen sowohl im Hinblick auf die relative Häufigkeit vergangenheitsbezogener

¹⁶ Als Signifikanzniveau bei den durchgeführten zweiseitigen Tests wird hier und im Folgenden $p < 0,05$ verwendet. Es sei betont, dass aufgrund der geringen Stichprobengröße und des Designs Signifikanz als Hinweis auf eine statistische Tendenz verstanden wird, die unter Vorbehalt zu interpretieren ist.

Deutungsleistungen als auch auf die definierten Triftigkeitsmerkmale derselben im Mittel durchweg höher ausgefallen sind. Dementsprechend scheint es berechtigt, einen Zusammenhang zwischen diesen Resultaten und dem Einsatz der verschiedenen Formen historischer Zeitungen anzunehmen.

Zur konkreten Interpretation des Befunds ist es mit Blick auf die zweite Forschungsfrage notwendig, den Leistungszuwachs der einzelnen Gruppen, die mit historischen Zeitungen gearbeitet haben (Recherche in Digitalarchiven DA, $n = 17$; Faksimiles historischer Zeitungen in digitalisierter Form ZD, $n = 12$; gedruckte Faksimiles historischer Zeitungen $n = 12$, ZA) zu vergleichen. Auch dazu wurden ANCOVAs über die Posttestscores der beschriebenen Codes (Innersubjektfaktor: Zugehörigkeit Zeitungsgruppe; Kovariate: Prätestscore) berechnet. Die Ergebnisse sind in Tabelle 4 aufgeführt. Da die Scores der Variable zur Reichweite vergangenheitsbezogener Deutungsleistungen die Voraussetzungen der ANCOVA nicht erfüllten,¹⁷ wurde für diese Variable ein Kruskal-Wallis-H-Test (Kruskal & Wallis, 1952, dazu auch Bortz & Lienert, 2008, 222–228) über die Differenzscores $t_2 - t_1$ berechnet,¹⁸ der in Tabelle 5 berichtet wird.

17 Die Regressionssteigungen der Stichprobenscores sind für diese Variable nicht gegeben, darüber hinaus sind auch die Residuen nicht normalverteilt, was zusammen bei unterschiedlich grossen Stichproben nach Bortz und Schuster (2010, 311) dazu führt, dass die ANCOVA kontraindiziert ist.

18 Von der Verwendung einer ANOVA wurde aufgrund einiger Ausreisser in zwei der drei Gruppen Abstand genommen.

Variable vergangenheitsbezogene Deutungsleistungen (Rel.) VDRel $F(2, 37) = 5.47, p = .008, \text{partielles } \eta^2 = .23$					
Gruppe	unadjustiert			adjustiert	
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SE</i>
DA	17	.45	.15	.45	.03
ZD	12	.63	.14	.64	.03
ZA	12	.54	.18	.54	.04

Variable Qualität vergangenheitsbezogener Deutungsleistungen (Rel.) VDRelgM gewichtetes Mittel $F(2, 37) = 57.35, p = .002, \text{partielles } \eta^2 = .28$					
Gruppe	unadjustiert			adjustiert	
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SE</i>
DA	17	.16	.05	.14	.01
ZD	12	.21	.07	.23	.02
ZA	12	.17	.17	.17	.02

Variable Qualität leitender Hinsichten (Rel.) VLHRelgM gewichtetes Mittel $F(2, 37) = 3.86, p = .030, \text{partielles } \eta^2 = .17$					
Gruppe	unadjustiert			adjustiert	
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SE</i>
DA	17	.15	.05	.14	.01
ZD	12	.19	.06	.20	.01
ZA	12	.17	.08	.19	.01

Tab. 5: Unadjustierte und adjustierte deskriptive Statistiken sowie Ergebnisse der ANCOVAS zum gruppenweisen (DA, ZD, ZA) Vergleich der Posttestscores (Forschungsfrage 2)

Variable Reichweite vergangenheitsbezogener Deutungsleistungen (Rel.) VRRelgM.diff Differenzscore ($t_2 - t_1$) der gewichteten Mittel $\chi^2(2) = 8.88, p = .01219$			
Gruppe	<i>n</i>	<i>M</i>	<i>SD</i>
DA	17	.00	.05
ZD	12	.06	.05
ZA	12	-.01	.07

Tab. 6: Deskriptive Statistiken sowie Ergebnisse des Kruskal-Wallis-H-Tests zum gruppenweisen (DA, ZD, ZA) Vergleich der Differenzscores (Forschungsfrage 2)

¹⁹ Die Verwendung der χ^2 -Verteilung an dieser Stelle sowie den Verzicht auf die Angabe einer Effektstärke für den Omnibustest erläutern Rasch, Friese, Hofmann und Naumann (2021, 120–121).

Die Übersicht zeigt, dass sich die Posttestscores der Gruppen, die mit historischen Zeitungen gearbeitet haben, untereinander signifikant mit starkem Effekt voneinander unterscheiden. Die für die Ergebnisse der ANCOVAS Bonferroni-korrigierte Post-hoc-Analyse ergab einen signifikanten Unterschied zwischen der DA- sowie der ZD-Gruppe für alle drei getesteten Variablen (VDRel: $p = .006$, $M_{\text{Diff}} = -.19$; VDRelGM: $p = .002$, $M_{\text{Diff}} = -.08$; VLHGGM: $p = .043$, $M_{\text{Diff}} = -.06$). Für die Variable zur Qualität vergangenheitsbezogener Deutungsleistungen wurde zudem der Unterschied zwischen den Gruppen DA und ZA signifikant ($p = .045$, $M_{\text{Diff}} = .06$).

Für die Post-hoc-Analyse des Kruskal-Wallis-H-Tests über die Variable zur Reichweite vergangenheitsbezogener Deutungsleistungen wurden Dunn-Bonferroni-Tests (Dunn, 1964) berechnet. Diese ergaben signifikante Unterschiede zwischen den Gruppen DA und ZD ($z = -2.72$, $p = .020$, $r = .51$) sowie zwischen ZA und ZD ($z = 2.50$, $p = .038$, $r = .51$)

5 Diskussion und Fazit

Zusammengefasst ergibt sich damit für diesen Ausschnitt aus der Gesamtheit der untersuchten Variablen ein recht deutliches Bild. Dabei ist jedoch zu beachten, dass an dieser Stelle die Wechselwirkungen zwischen den verschiedenen Manifestationen historischer Denkleistungen sowie zwischen ihren qualitativen Ausprägungen nicht mitberücksichtigt werden können, obwohl ihnen für eine umfassende Interpretation der vorgestellten Ergebnisse zum Teilbereich der vergangenheitsbezogenen Deutungsleistung eine nicht zu unterschätzende Bedeutung zukäme. Darüber hinaus sind im Rahmen der Fragebogen- und Beobachtungsstudie erhobene potenzielle Störeinflüsse auf die gewonnenen Ergebnisse, wie beispielsweise der Einfluss der verschiedenen Lehrkräfte oder das grundsätzliche Interesse der Lernenden an Geschichte, noch nicht in die konkrete Auswertung aufgenommen worden. Als zentrale Ergebnisse sind vor diesem Hintergrund die beiden folgenden Aspekte festzuhalten.

Erstens hat der Gruppenvergleich zwischen den Zeitungsgruppen und der Q-Gruppe (Forschungsfrage 1) gezeigt, dass erstere relativ zur Textlänge sowohl im Mittel signifikant mehr entsprechende Deutungsleistungen formuliert als auch diese besser belegt bzw. sicherer argumentativ in den Gesamttext eingebunden hat (Variable VDRelGM, Merkmal der empirischen Triftigkeit). Darüber hinaus konnten die Schreibenden der Z-Gruppen die leitenden Hinsichten, auf deren Grundlage sie ihre Deutungsleistungen formulierten, deutlicher offen-

legen (Variable VLHRelgM, Merkmal der normativen Triftigkeit) sowie deren Gültigkeitsansprüche in Grundzügen in ihren Texten reflektieren (Variable VRRRelgM, Merkmal der narrativen Triftigkeit). Der damit angedeutete Zusammenhang zwischen dem Einsatz historischer Zeitungen und dem Umgang mit vergangenheitsbezogenen Deutungsleistungen wurde in den Untersuchungen zu Forschungsfrage 2 konkretisiert.

Die entsprechenden Vergleiche für die Digitalarchive-, Digitalisate- und Faksimiledrucke-Gruppen haben nachgewiesen, dass es zu einem signifikanten Unterschied mit durchweg starkem Effekt auch zwischen den Zeitungsgruppen gibt. Zum anderen konnte durch die Post-hoc-Analysen gezeigt werden, dass die signifikanten Unterschiede für alle untersuchten Teilbereiche zwischen derjenigen Gruppe, die mit digitalisierten Zeitungsfaksimiles gearbeitet, und derjenigen, die eigenständig in Digitalarchiven nach Zeitungen recherchiert hat, bestehen. Für die Untersuchungen der Merkmale empirischer und narrativer Triftigkeit ergaben sich derartige Unterschiede zudem auch zwischen ersterer und der Gruppe, in der Faksimiledrucke historischer Zeitungen zum Einsatz gekommen sind, wobei mit Blick auf die deskriptiven Werte diejenigen Lernenden, die Zeitungsdigitalisate verwendet haben, durchweg die nach den Massstäben der Auswertung besseren Ergebnisse produziert haben.

Im Rahmen der untersuchten Stichprobe zeichnet sich damit als Tendenz ab, dass gerade der Einsatz von digitalisierten Zeitungsfaksimiles in einem positiven Zusammenhang mit der Art und Weise steht, auf die Lernende vergangenheitsbezogene Deutungsleistungen gestalten. Im Gegensatz dazu scheinen die insbesondere auf den Bereich der narrativen Triftigkeit ausgerichteten Ziele, die auf theoretischer Ebene für die Recherche in digitalen Zeitungsportalen angenommen worden sind, in keiner Weise erreicht worden zu sein. Im Gegenteil liegen die deskriptiven Werte hier durchweg unter denjenigen der anderen Zeitungsgruppen bzw. für den Bereich der Reichweite der vergangenheitsbezogenen Deutungsleistungen auf nahezu dem gleichen Niveau wie die Faksimiledrucke-Gruppe. Diese Konstellation ist abschliessend insbesondere mit Blick auf die Verbindung des Mediums Zeitung mit der Medialität des Digitalen zu diskutieren.

Die Gruppenvergleiche im Rahmen der ersten Forschungsfrage haben als Tendenz aufgezeigt, dass der Einsatz historischer Zeitungen in den verschiedenen Darreichungsformen die bislang theoretisch formulierten Ansprüche zumindest im Rahmen dieser Untersuchung in Bezug auf die Formulierung vergangenheitsbezogener Deutungsleistungen erfüllen zu können scheint. Auch im

Einzelvergleich der deskriptiven Kennwerte übertreffen die Ergebnisse der Zeitungsgruppen in diesem Bereich diejenigen der Q-Gruppe. Die inferenzstatistische Untersuchung, inwiefern diese Unterschiede der Posttestscores signifikant sind, wird an anderer Stelle zu berichten sein. Zu fragen bleibt aber, warum die relativ simple Form des Digitalisats der historischen Zeitung eine relativ deutliche positive Tendenz aufweist, während die Recherchetätigkeit im digitalen Raum tendenziell zu im Durchschnitt der Zeitungsgruppen eher schlechteren Resultaten führt. Als zentraler Faktor soll hier das Verhältnis von Komplexität und Nützlichkeit der jeweiligen medialen Form der historischen Zeitung zur Erklärung herangezogen werden. So standen einerseits die Lernenden, die mit gedruckten Faksimiles zu arbeiten hatten, vor der Herausforderung, in grossen Material-, d. h. hier Papiermengen, einen Überblick zu gewinnen und zu behalten. Die Möglichkeiten der digitalen Endgeräte (z. B. Lesezeichen setzen, Screenshots erstellen usw.) standen ihnen nicht zur Verfügung. Andererseits muss in Betracht gezogen werden, dass die Recherchetätigkeit in den Digitalarchiven einen so grossen Teil der Aufmerksamkeit der Lernenden in Anspruch genommen hat, dass es ihnen nicht mehr möglich war, mit dem gefundenen Material ähnlich der ZD-Gruppe zu arbeiten oder gar den Prozess der Recherche und den Einblick in den Fundus potenziell zu verwendender Quellen in das Schreibergebnis einfließen zu lassen. An dieser Stelle ist auch das verkürzte Treatment zu berücksichtigen, das möglicherweise nicht ausgereicht hat, um die Lernenden in einer Weise mit den entsprechenden digitalen Heuristiken vertraut zu machen, um die theoretisch erwarteten Ergebnisse zu erreichen.

Es sei damit hier als abschliessendes Fazit und zugleich als These formuliert, dass die identifizierten Unterschiede einerseits, wie beschrieben, mit den geschichtsdidaktischen Potenzialen der historischen Zeitung in Verbindung stehen. Andererseits scheint die Vereinfachung des Arbeitsprozesses, die mit der Verwendung der Zeitungsdigitalisate einhergeht, die Verfügbarkeit dieser Potenziale für die Lernenden zu erhöhen. Im Gegensatz dazu liegt die Annahme nahe, dass die Komplexitätssteigerung, die aus den Rechercheaufgaben resultierte, im durchgeführten Setting zu einer partiellen Überforderung beigetragen hat, die der Wirksamkeit des Mediums Zeitung für das historische Lernen tendenziell abträglich gewesen ist. In der weiteren Auswertung der erhobenen Daten ist nun zu überprüfen, inwiefern sich derartige Beobachtungen auch im Hinblick auf die übrigen Manifestationen historischer Denkleistungen machen lassen. Darüber hinaus dürfte von höchstem Interesse sein, im Zuge der Auswertung der die Haupterhebung flankierenden Fragebogenstudie zu untersu-

chen, ob die Einschätzung der Verwendung des Mediums Zeitung seitens der Lernenden einerseits und der medialen Darreichung andererseits die aufgestellte These der Komplexitätsreduktion und Nützlichkeit durch den Einsatz der Zeitungsdigitalisate stützt. Festgehalten werden darf aber bis hierhin und unter Berücksichtigung aller dargelegten Einschränkungen der Aussagekraft der Studie, dass die vorgestellten Ergebnisse als Hinweis darauf gelesen werden können, dass der historischen Zeitung ein tatsächliches Potenzial für historisches Lernen zukommen kann.

Literatur

- Ankersmit, F. (1994). *History and tropology: The Rise and fall of metaphor* ([ACLS Humanities E-Book edition]). Berkeley: University of California Press. Abgerufen von <http://ark.cdlib.org/ark:/13030/kt9k4016d3/>.
- Becker, Axel (2017). Historische Urteilsbildung. In Michele Barricelli & Martin Lücke (Hrsg.), *Handbuch Praxis des Geschichtsunterrichts* (2. Aufl., Bd. 1, S. 316–325). Schwalbach/Ts.: Wochenschau.
- Bergmann, Klaus (2008). *Geschichtsdidaktik: Beiträge zu einer Theorie historischen Lernens* (3. Aufl.). Schwalbach/Ts.: Wochenschau.
- Bernsen, Daniel (2017). Online recherchieren. In Daniel Bernsen & Ulf Kerber (Hrsg.), *Praxishandbuch Historisches Lernen und Medienbildung im digitalen Zeitalter* (S. 313–319). Bonn: BPB.
- Bortz, Jürgen / Lienert, Gustav A. (2008). *Verteilungsfreie Methoden in der Biostatistik* (3. Aufl.). Heidelberg: Springer.
- Bortz, Jürgen / Schuster, Christof (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Berlin: Springer.
- Bösch, Frank (2019). Zeitungs-Geschichte: Zum Wandel eines Mediums. *Praxis Geschichte*. (2), 4–9.
- Brennan, Robert L. & Prediger, Dale J. (1981). Coefficient Kappa: Some uses, missuses, and alternatives. *Educational and Psychological Measurement*, 41(3), 687–699.
- Broekelen, Gerard J. P. van (2006). ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *Journal of Clinical Epidemiology*, 59(9), 920–925.
- Bullen, P. S. (2003). *Handbook of means and their inequalities*. Boston: Kluwer Academic.
- Cohen, Jacob (2013). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hoboken: Taylor and Francis.

- Cronbach, Lee J. & Furby, Lita (1970). How we should measure «change» – or should we? *Psychological Bulletin*, 74(1), 68–80.
- Dimitrov, Dimitar M. & Rumrill, Phillip D. (2003). Pretest-posttest designs and measurement of change. *Work*, 20(2), 159–165.
- Döring, Nicola & Bortz, Jürgen (2016), *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin: Springer.
- Droysen, Johann G. (1868). *Grundzüge der Historik*. Leipzig: Velt.
- Dunn, Olive J. (1964). Multiple comparisons using rank sums. *Technometrics*, 6(3), 241. <https://doi.org/10.2307/1266041>
- Fleischhacker, Michael (2014). *Die Zeitung: Ein Nachruf*. Wien: Brandstätter.
- Föhr, Pascal (2017). *Historische Quellenkritik im Digitalen Zeitalter*. Abgerufen von https://edoc.unibas.ch/64111/1/F%C3%B6hr_Pascal-Historische_Quellenkritik_im_Digitalen_Zeitalter-2018.pdf.
- Gau «Ruhr und Lippe» des Alldeutschen Verbandes (Hrsg.) (1901). *Die Polen im rheinisch-westfälischen Steinkohlen-Bezirk: mit einem statistischen Anhang, einer Sammlung polnischer Lieder und zwei Karten*. München: Lehmann.
- Goldman, Susan R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In Nora Shuart-Faris & David Bloome (Hrsg.), *Uses of intertextuality in classroom and educational research* (S. 313–347). Greenwich: Information Age Pub.
- Groth, Otto (1960). *Das Wesen des Werkes*. Berlin: de Gruyter.
- Hagenauer, Gerade (2010). Kurzzeitinterventionen versus Langzeitinterventionen. In Tina Hascher & Bernhard Schmitz (Hrsg.), *Pädagogische Interventionsforschung: Theoretische Grundlagen und empirisches Handlungswissen* (S. 243–251). Weinheim, München: Juventa.
- Hasberg, Wolfgang (2007). Die Entzauberung der Hrosvith von Gandersheim – oder: Dekonstruktion als Akt entdeckenden historischen Lernens. In Wolfgang Hasberg & Wolfgang E. J. Weber (Hrsg.), *Geschichte entdecken: Karl Filser zum 70. Geburtstag* (S. 211–242). Berlin: LIT.
- Henning, Hansjoachim, & Sieg, Uwe (1997). Quellensammlung zur Geschichte der deutschen Sozialpolitik 1867 bis 1914: IV. Abteilung (1905–1914) (Bd. 3). Darmstadt: Wissenschaftliche Buchgesellschaft. Abgerufen von <http://www.urn.fi/urn:nbn:de:he-bis:34-2018040555125>.
- Hodel, Jan (2016). Geschichte – Medien – Suche: Kompetenter Umgang mit komplexen Herausforderungen. In Saskia Handro & Bernd Schönemann (Hrsg.), *Aus der Geschichte lernen?: Weisse Flecken der Kompetenzdebatte* (S. 153–168). Berlin: LIT.
- Kerber, Ulf (2017). Historische Medienbildung – ein transdisziplinäres Modell für den Geschichtsunterricht. In Daniel Bernsen & Ulf Kerber (Hrsg.), *Praxishandbuch Historisches Lernen und Medienbildung im digitalen Zeitalter* (S. 45–82). Bonn: BPB.

- Kim, Yongnam & Steiner, Peter M. (2021). Gain scores revisited: A graphical models perspective. *Sociological Methods & Research*, 50(3), 1353–1375. <https://doi.org/10.1177/0049124119826155>
- Körper, Andreas (2018). Geschichte – Spielen – Denken: Kontingenzverschiebungen und zweiseitige Triftigkeiten. *MedienPädagogik Zeitschrift für Theorie und Praxis der Medienbildung*, 19–35.
- Krameritsch, Jakob (2007). *Geschichte(n) im Netzwerk: Hypertext und dessen Potenziale für die Produktion, Repräsentation und Rezeption der historischen Erzählung*. Münster: Waxmann.
- Kruskal, William H. & Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583. <https://doi.org/10.2307/2280779>
- Kuchler, Christian (2016). Zwischen geschichtsdidaktischer Forschung und aktuellem Unterrichtseinsatz: Historisches Lernen und Zeitung. In Christian Kuchler & Benjamin Städter (Hrsg.), *Zeitungen von gestern für das Lernen von morgen? Historische Tagespresse im Geschichtsunterricht* (S. 31–49). Göttingen.
- Kuckartz, Udo (2018). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung* (4. Aufl.). Weinheim: Beltz Juventa.
- Kultusministerkonferenz (2005). Einheitliche Prüfungsanforderungen in der Abiturprüfung: Geschichte. Abgerufen von https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1989/1989_12_01-EPA-Geschichte.pdf am 08. 11. 2022.
- Leutner, Detlev (2012). Perspektiven pädagogischer Interventionsforschung. In Eckart Severing & Reinhold Weiss (Hrsg.), *Qualitätsentwicklung in der Berufsbildungsforschung* (1. Aufl., S. 17–28). Bielefeld: Bertelsmann.
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (2014). Kernlehrplan für die Sekundarstufe II Gymnasium / Gesamtschule in Nordrhein-Westfalen: Geschichte. Abgerufen von https://www.schulentwicklung.nrw.de/lehrplaene/upload/klp_SII/ge/KLP_GOSt_Geschichte.pdf am 08.11.2022.
- Rasch, Björn / Friese, Malte / Hofmann, Wilhelm & Naumann, Ewald (2021). *Quantitative Methoden 2: Einführung in die Statistik für Psychologie, Sozial- & Erziehungswissenschaften* (5. Aufl. 2021). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Rüsen, Jörn (1983). *Historische Vernunft: Die Grundlagen der Geschichtswissenschaft*. Göttingen: V & R.
- Rüsen, Jörn (1986). *Rekonstruktion der Vergangenheit: Die Prinzipien der historischen Forschung*. Göttingen: V & R.
- Rüsen, Jörn (1989). *Lebendige Geschichte: Foren und Funktionen des historischen Wissens*. Göttingen: V & R.

- Rüsen, Jörn (1990). *Zeit und Sinn: Strategien historischen Denkens*. Frankfurt am Main: Fischer.
- Rüsen, Jörn (2013). *Historik: Theorie der Geschichtswissenschaft*. Köln: Böhlau.
- Sauer, Michael (2006). «Was sich begeben und zugetragen hat»: Die Zeitung als Quelle im Geschichtsunterricht. In Markus Bernhardt, Gerhard Henke-Bockschatz, & Michael Sauer (Hrsg.), *Bilder – Wahrnehmungen – Konstruktionen: Reflexionen über Geschichte und historisches Lernen. Festschrift für Ulrich Mayer zum 65. Geburtstag* (S. 242–255). Schwalbach/Ts.: Wochenschau.
- 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-Verl.
- Schwabe, Astrid (2016). Multimedia statt Papier und Druckerschwärze – das «alte» Medium Zeitung in der digitalen Welt und der Geschichtsunterricht. In Christian Kuchler & Benjamin Städter (Hrsg.), *Zeitungen von gestern für das Lernen von morgen?: Historische Tagespresse im Geschichtsunterricht* (S. 161–177). Göttingen.
- Sochatzky, Florian, Schöner, Alexander, & Schreiber, Waltraud (Hrsg.) (2013). *Analyse von Schulbüchern als Grundlage empirischer Geschichtsdidaktik*. Stuttgart: Kohlhammer.
- Souvignier, Elmar & Dignath Ewijk, Charlotte van (2010). Pädagogische Interventionsforschung – ein historischer Rückblick. In Tina Hascher & Bernhard Schmitz (Hrsg.), *Pädagogische Interventionsforschung: Theoretische Grundlagen und empirisches Handlungswissen* (S. 12–30). Weinheim, München: Juventa.
- Stöber, Rudolf (2005). *Deutsche Pressegeschichte: Von den Anfängen bis zur Gegenwart* (2., überarb. Aufl.). Konstanz: UVK.
- Tenfelde, Klaus, & Urban, Thomas (Hrsg.). (2010). *Das Ruhrgebiet – Ein historisches Lesebuch: Zusammengestellt und bearbeitet von den Mitarbeiterinnen und Mitarbeitern des Instituts für soziale Bewegungen der Ruhr-Universität Bochum und der Stiftung Bibliothek des Ruhrgebiets*. Essen: Klartext.
- Waldis, Monika / Marti, Philipp & Nitsche, Martin (2015). Angehende Geschichtslehrpersonen schreiben Geschichte(n). *Zeitschrift für Geschichtsdidaktik*, 14(1), 63–86. <https://doi.org/10.13109/zfgd.2015.14e.1.63>
- Wineburg, Sam (2019). Warum historische Kompetenzen für die Auswertung von digitalen Quellen nicht ausreichend sind. In Sebastian Barsch, Andreas Lutter, & Christian Meyer-Heidemann (Hrsg.), *Fake und Filter: Historisches und politisches Lernen in Zeiten der Digitalität* (S. 105–120).

Sprachsensible
Erforschung
historischen Denkens

Historisches Erklären lernen?

Exploration von Konzepten von Schüler*innen und ihrer Veränderbarkeit

Tobias Flink

1 Problemaufriss

Empirische Studien zeigen, dass viele Schüler*innen Probleme damit haben, adäquate fachspezifische Sprachhandlungen im Geschichtsunterricht zu produzieren (Schönemann, Thünemann & Zülsdorf-Kersting, 2011, 124; Hartung, 2013, 281; Hodel, Waldis, Zülsdorf-Kersting & Thünemann, 2013, 138–139; Husemann 2022) und/oder betonen die Notwendigkeit von sprachlichen oder strukturierenden Hilfsmassnahmen für einen funktionalen Gebrauch sprachlicher Diskurs- und Darstellungsmuster (Mierwald & Brauch, 2015, 110). Ein möglicher Grund für das teilweise schwache Abschneiden kann in den Konzepten der Lernenden zu der jeweiligen Sprachhandlung vermutet werden. So fragt die Narratio-Gruppe, «ob Schülerinnen und Schüler im bisherigen Unterricht überhaupt darin geschult wurden, Narrationen zu verfassen, die fachdidaktischen Kriterien entsprechen?» (Hodel et al., 2013, 143). Auf die Notwendigkeit, das Sprachhandlungswissen Lernender als Faktor für deren Performanz zu berücksichtigen, hat jüngst auch Charlotte Husemann (2022) hingewiesen. Und Mareike-Cathrine Wickner (2019, 137) hat herausgestellt, dass ihre Proband*innen «eher über sprachliche als über fachsprachliche Schreibfähigkeiten verfügen». Sie betont allerdings – entgegen den eher kritischen Einschätzungen der oben angeführten Studien –, dass auf den allgemeinen Sprachfähigkeiten aufgebaut werden könne und die fachspezifischen Schreibhandlungen adäquat lösbar seien, «sofern mit den Lernenden kriteriengeleitet erarbeitet wird, welche fachsprachlichen Anforderungen mit derlei Aufgaben einhergehen» (2019, 142). Schülerseitige Konzepte scheinen demnach einen zentralen Faktor für die Qualität und die Fachspezifik der Performanz der jeweiligen Sprachhandlung darzustellen.

Blickt man auf den normativen Rahmen für Geschichtsunterricht in Deutschland, dann stellen die Operatoren, zu denen *erklären* gehört, einen ersten Ansatzpunkt zur Bildung von Konzepten zu spezifischen Sprachhandlungen dar. Diese werden in den Einheitlichen Prüfungsanforderungen in der Abiturprüfung Geschichte (2005, 7) definiert als «handlungsinitiierende Verben, die signalisieren, welche Tätigkeiten beim Lösen von Prüfungsaufgaben erwartet werden». Der Definition nach vor allem als normative Instrumente für transparente(re) Aufgabenstellungen zur Überprüfung fachspezifischer Kompetenzen / Performanzen von Lernenden gedacht, werden Operatoren seitens der (geschichts-)didaktischen Forschung massgeblich aufgrund der mit ihnen verbundenen Intransparenz kritisiert; diese Kritik bezieht sich u. a. auf

- die «entdidaktisiert[e] Veröffentlichung» (Peters, 2015, 301), also ihre Nennung ohne zuordenbare Autor*innen und didaktisches Fundament,
- die mangelnde Trennschärfe zwischen den einzelnen Verben (Dammer, 2007, 54),
- die teilweise widersprüchliche Zuordnung zu den jeweiligen Anforderungsbereichen (vgl. zu «darstellen» Hartung, 2013, 36),
- die unklare «fachliche [...] Relevanz» (Handro, 2013, 320) sowie
- ihre Vielzahl, die einen pragmatischen Gebrauch im Geschichtsunterricht erschwere (Schönemann et al., 2011, 17–18).

Eine zielführende Nutzung im (Geschichts-)Unterricht werde des Weiteren dadurch erschwert, dass die Definitionen der Operatoren zwar sprachliche und strukturelle Formen der zu produzierenden Texte einforderten, diese jedoch in der Schule nicht gelehrt würden (Feilke, 2012, 4).

In der geschichtsdidaktischen Forschung ist spätestens seit dem PISA-Schock ein zunehmendes Bewusstsein für die sprachlichen Anforderungen des Faches erkennbar: Beispielsweise charakterisiert Saskia Handro (2013, 317) das Verhältnis von historischem Lernen und Sprache als «Schlüsselproblem [...] des Geschichtsunterrichts». Auch die Operatoren rücken in den Fokus der Forschung. So fordern Sven Oleschko und Anastasia Moraitis (2012, 34), diese «aufzuschlüsseln», um die Anforderung, die an den Schüler gestellt wird, klar zu definieren»; auch Tülay Altun und Katrin Günther (2018, 15) betonen die Notwendigkeit, Operatoren «transparent» zu machen, und Olaf Hartung (2016, 189) plädiert dafür, in Aufgabenstellungen zusätzlich zu den Operatoren «Angaben über Zieltextformate» zu formulieren. Die gestiegene Bedeutsamkeit der Analyse sprachlicher und fachlicher Zusammenhänge zeigt sich zudem in einer

stetig wachsenden Zahl empirischer Studien, die die theoretische Modellierung einzelner Operatoren (z. B. zu *begründen* Altun / Günther, 2018) bzw. die Diagnose (z. B. zu *beschreiben* Oleschko, 2015) oder die instruktionsgestützte Erhebung der erwarteten Sprachhandlung u. a. zur Evaluation von unterstützenden Massnahmen (z. B. zu *argumentieren* Mierwald & Brauch, 2015) zum Ziel haben.

Vor diesem Hintergrund und angesichts des Potenzials, das in schülerseitigen Konzepten zu historischem Erklären liegt, leiten zwei Forschungsfragen den vorliegenden Werkstattbericht:

1. Wie sind schüler*innenseitige Konzepte zu historischem Erklären beschaffen?
2. Können Schüler*innen Spezifika historischen Erklärens lernen und zustimmungsfähig deklarieren, und wenn ja, in welchem Masse?

Diese Fragen werden explorativ diskutiert. In einem ersten Schritt wird die Sprachhandlung des *Erklärens* fachspezifisch für den Kontext des Geschichtsunterrichts operationalisiert. Zudem werden empirische Ergebnisse sowie Massnahmen zur Förderung schüler*innenseitigen historischen Erklärens diskutiert (Abschnitt 2). Anschliessend wird, auf lerntheoretische Überlegungen des *Conceptual Change*-Ansatzes zurückgreifend (Abschnitt 3), ein Design zur Erhebung und gezielten Veränderung schüler*innenseitiger Konzepte zu historischem Erklären vorgestellt (Abschnitt 4). Anhand zweier Schüler*innen der neunten Klasse wird exemplarisch untersucht, wie die Konzepte beschaffen sind und inwiefern sie sich über die Studie verändern (Abschnitt 5). Abschliessend werden die explorativen Analysen als Hinweise zur Beantwortung der beiden Forschungsfragen herangezogen und weitere Forschungsdesiderate aufgezeigt (Abschnitt 6).

2 Historisches Erklären

Die Frage, wie historische Erklärungen beschaffen sind, hat Geschichtsphilosophie und -theorie spätestens seit Johann Gustav Droysens *Grundriss der Historik* (1868) und seiner dort vertretenen These vom Dualismus zwischen Erklären und Verstehen vor grosse Herausforderungen gestellt (Hausmann, 1991, 149).

Eine zentrale definitorische Herausforderung, die auch für den schulischen Kontext relevant ist, besteht in grundlegenden Gemeinsamkeiten von geschichtswissenschaftlichen und allgemeinen, überwiegend alltäglichen Arten des Erklärens: Erklären ist «ein Konzept, das in einer diffusen Weise zwischen alltäglicher und theoretischer Sprache angesiedelt ist» (Ehlich, 2009, 11). Im All-

tag erklären wir tagtäglich, sind also mit grundlegenden Merkmalen aller Arten von Erklärungen vertraut: Eine Erklärung kann modelliert werden als eine mündliche oder schriftliche Kommunikationssituation zwischen wenigstens zwei Personen, in der Informationen übermittelt werden. Dabei werden Erklärungen unterteilt in das Thema, das sogenannte Explanandum («was zu erklären ist») und in den erklärenden Teil, das Explanans («das, was erklärt»). Der oder die Erklärende verfolgt das Ziel, dass der oder die Adressierte(n) einen bestimmten Sachverhalt verstehen bzw. dass die Zusammenhänge – im etymologischen Sinne des Wortes – *klar* («*claritas*» = «Klarheit») werden (Klein, 2009, 25–28).

Die Debatte um historisches Erklären brachte mehrere Modelle hervor und behandelte immer auch die Eigenständigkeit der Geschichtswissenschaft; erst im Gegensatz von Verstehen und Erklären, dann in dessen versuchter Überwindung durch die einheitliche Erklärform des H-O-Modells, zuletzt in der Abgrenzung von den Sozialwissenschaften (vgl. Haussmann, 1991, und Lorenz, 1997). Den bis heute zustimmungsfähigsten Vorschlag unterbreitete Arthur Danto (1965/1974, 321–322), der in der Erzählung die genuin historische Art des Erklärens sieht. Jörn Rüsen (2013, 162) kommt zu einer ähnlichen Einschätzung: «Mit der Einführung des Narrativitätsarguments in die Geschichtstheorie ist dann die Debatte über das, was historisches Erklären ist, an ihr Ende gelangt.» Dieser Auffassung ist jedoch nur bedingt zuzustimmen. Zwar erfreut sich die narrative Erklärung grosser Zustimmung in Geschichtsdidaktik, -theorie und -philosophie (vgl. z. B. Pandel, 2010, 88; Barricelli, 2012, 258; Gerber, 2012). Aber die (deutschsprachige) Geschichtsdidaktik steht noch am Anfang, wenn es darum geht, narratives Erklären geschichtstheoretisch fundiert im Kontext kompetenzorientierten Geschichtsunterrichts sowie in Abgrenzung zu alltäglichem Erklären zu modellieren.

Unter Berücksichtigung geschichtstheoretischer, philosophischer und linguistischer Positionen wird narratives Erklären wie folgt operationalisiert: Eine narrative Erklärung liegt in Form einer triadischen Geschichte mit Einleitung, gegliedertem Hauptteil und Schluss vor (Danto, 1965/1974). Sie erklärt die Veränderung eines Referenzsubjekts / eines Gegenstandes in der Vergangenheit (als genuin historisches Explanandum, Danto, 1974; Rüsen, 1986) durch die zustimmungsfähige Heranziehung mehrerer Propositionen (Gründe, Faktoren etc.; Jahr, 2000), also durch ein multikausales Geflecht von Ursachen- und Wirkungsbeziehungen, im Explanans (Pandel, 2017). Diese Beziehungen werden durch funktionale sprachliche Verknüpfungen zwischen den einzelnen Propositionen sowie auf das Explanandum ausgedrückt (Jahr, 2000). Die Propositionen

werden aus historischen Quellen gewonnen und müssen zur Überprüfung ihrer empirischen Plausibilität belegt werden (Rüsen, 2013). Zudem gewinnt das jeweils raumzeitlich singuläre Explanandum durch Einordnung in historische Theorien und Forschungsergebnisse an Zustimmungsfähigkeit (Hempel, 1942; Welskopp, 2002; Rüsen, 2013). Herangezogene Quellen und Darstellungen müssen hinsichtlich ihrer Perspektivität kritisch untersucht und die ihnen entnommenen Propositionen sprachlich distanziert wiedergegeben werden (Rüsen, 2013; Pandel, 2017). Des Weiteren müssen Erklärende aufgrund ihrer Standortgebundenheit ihre eigene Perspektive reflektieren und für die Lesenden der Erklärung aufdecken (Rüsen, 2013). Diese Merkmale bilden den Gegenstand des Treatments (4.) und werden in der Auswertung (5.) in Kode-Form vorgestellt.

Neben der fundierten Operationalisierung stellt die Entwicklung von Massnahmen, um Schüler*innen in ihren Konzepten und Performanzen zielgerichtet zu fördern, ein weiteres Desiderat dar. Für die hier im Fokus stehenden Konzepte historischen Erklärens kann an empirische Ergebnisse angeknüpft werden. So stellten Studien aus den 1990er-Jahren heraus, dass insbesondere jüngere Schüler*innen v. a. intentionale Faktoren und Gründe im Explanans bevorzugten, ältere Lernende strukturelle politische und wirtschaftliche Faktoren anführten (Carretero, Jacott & Limón, 1994, 369; Lee, Ashby & Dickinson, 1998, 239–240). Ola Halldén (1998, 132) zeigte, dass weniger reflektierte Schüler*innen dazu neigten, Strukturen zu personalisieren, und damit unterkomplex erklärten sowie die Interdependenzen zwischen Akteur*innen und Strukturen nicht berücksichtigten. Christian Mathis (2013, 192–211) stellte ebenfalls simplifizierende Erklärungsmuster der Lernenden fest, die vor allem von präsentistischen Sichtweisen wie teleologischem Fortschrittsdenken sowie von Verkennung epistemologischer Spezifika des Faches wie Retrospektivität, Konstruktivität und Historizität geprägt seien. Einen ersten Hinweis auf Konzepte zu den Operatoren *beschreiben* und *erklären* liefern Husemanns Interviews mit Gesamtschüler*innen der Mittelstufe. Die Ergebnisse sind überwiegend ernüchternd: Die meisten Schüler*innen können Bedeutung und Zweck von Operatoren nicht anführen (2022, 147); die Autorin konstatiert «erhebliche [...] Leerstellen in der Konzeptualisierung einer Erklärung» (2022, 198).

Während diese Studien wichtige Erkenntnisse u. a. zu möglichen nicht-zustimmungsfähigen Annahmen von Schüler*innen liefern, bieten sie mit Ausnahme Husemanns keine Hinweise zur möglichen Struktur und Veränderbarkeit von Konzepten zu historischen Erklärungen. Hier ist die Studie von Gerhard Stoel, Jannet van Drie und Carla van Boxtel hervorzuheben; sie beobachteten

durch die Förderung epistemologischer Beliefs und *second order-concepts* bei den an der Intervention teilnehmenden Elftklässler*innen kleine signifikante Effekte in der Performanz historischen Erklärens (2017, 332). Allerdings verbesserten sich die Kontrollgruppen in gleichem Masse, obwohl bei ihnen die Prinzipien nicht expliziert wurden (Stoel et al. 2017, 329). Beide Gruppen profitierten demnach v. a. von kollaborativen Settings und Analogiebildungen mit anschließendem abstrahierendem Transfer auf Geschichte. Die Massnahmen gehen auf Arthur Chapman (2017) zurück. Er stellt seine Überlegungen zu multikausalem Erklären und der Notwendigkeit, Faktoren zu gewichten, anhand einer im englischsprachigen Raum bekannten Anekdote zu dem Tod des Kamels Alfons vor. Mit Blick auf die Zusammenhänge sprachlichen und fachlichen Lernens (Handro, 2013; Wickner, 2019) erscheint zudem die Förderung von Sprachhandlungswissen einleuchtend. Erste explorative Ansätze finden sich bei Nicola Brauch, Lena Heine und Christoph Bramann mittels Scaffolds (2020, 161), allerdings ist deren Einsatz zur Unterstützung historischen Erklärens nicht unumstritten (vgl. Evans & Pate, 2007). Zudem zeigt Husemann (2022, 325) bei ihren Proband*innen kleine Effekte durch die Förderung von Sprachhandlungswissen, wenngleich sich einige überdurchschnittlich gute Teilnehmende ihres Samples durch das Treatment sogar verschlechterten.

Hinsichtlich der Förderung schüler*innenseitiger Konzepte zu historischem Erklären besteht also weiterhin Forschungsbedarf. Da Lernenden aufgrund der intransparenten normativen Vorgaben ein direkter Zugang zu den fachspezifischen Merkmalen mindestens erschwert und schlimmstenfalls verwehrt ist, kommt den schüler*innenseitigen Konzepten zu den Sprachhandlungen grosse Bedeutung zu. Daher bieten sich zur Förderung Strategien zur Reflexion des eigenen Konzepts historischen Erklärens bzw. kausalen Denkens an. Für einen grösseren Erfolg, so argumentiert auch die *Conceptual Change*-Forschung (s. u.), müssen alltägliche Konzepte allerdings zunächst aufgedeckt, in ihren Differenzen zu den favorisierten spezifischeren Modellierungen deutlich und ihre Grenzen für Schüler*innen plausibel gemacht werden.

3 Lernen als Conceptual Change

Insbesondere angesichts der Tatsache, dass die mit den Operatoren angestrebten fachspezifischen Sprachhandlungen meist in alltäglichen (Vor-)Formen bekannt sind, erscheint es vielversprechend, beim Vorwissen und den Präkonzepten von Lernenden anzusetzen; also zu fragen, was sie unter der jeweiligen

Sprachhandlung verstehen, und die Antworten daraufhin zu untersuchen, über welche nicht-zustimmungsfähigen, alltäglichen, aber auch fachspezifischen Wissens Elemente sie verfügen. Daran anknüpfend stellt sich die Frage, ob bzw. wie man diese vorhandenen Wissens Elemente gezielt hin zu einem wissenschaftsspezifischen Konzept historischen Erklärens verändern kann. Beide Fragen adressieren das Forschungsparadigma des *Conceptual Change*, ins Deutsche als «Veränderung von Wissensstrukturen» übersetzt (Stark, 2003, 133).

Das *Conceptual Change*-Paradigma folgt moderat-konstruktivistischen Annahmen und versteht Lernen als bereichsspezifische Auf-, Aus- und Umbauprozesse von Wissen (Gruber & Stamouli, 2015, 33). Dabei steht das Individuum als Konstrukteur des eigenen Wissens im Mittelpunkt, an dessen Vorwissen ebene Prozesse ansetzen (Gropengießer & Marohn, 2018, 57). Während ältere *Conceptual Change*-Ansätze von einer radikalen, vollständigen Ersetzung ausgingen, weisen jüngere Forschungsergebnisse darauf hin, dass sich ein solcher radikaler Wechsel nur selten vollzieht; vielmehr würden Konzepte durch neues Wissen angereichert, sodass wissenschaftliche und alltägliche Konzept(anteile) nebeneinander bestünden (vgl. z. B. Ammerer, 2022, 11).

Die klassischen, überwiegend naturwissenschaftlichen Modellierungen und Untersuchungen von *Conceptual Change*-Prozessen blicken überwiegend auf inhaltsbezogene Konzepte und Begriffe, die übertragen auf die Geschichtsdidaktik, im Bereich der *first-order concepts* liegen (Günther-Arndt, 2006). Aber auch prozedurales Wissen scheint veränderbar (Fenn & Zülsdorf-Kersting, 2023, 90). Wissen, so wird angenommen, ist im Langzeitgedächtnis in Form übergeordneter Konzepte, z. B. zu «erklären», gespeichert und kann u. a. für Problemlösungen aktiviert werden (Schraw, 2006, 255). Das individuelle Konzept stellt also die Grundlage für die Performanz der entsprechenden Operation, Sprachhandlung o. Ä. dar. Zwar ist das Verhältnis von Konzepten und Kompetenzen bzw. Wissen und Können noch nicht final geklärt (Fenn & Zülsdorf-Kersting, 2023, 11). Fokussiert man allerdings die kognitiven Aspekte der Weinert'schen Kompetenzdefinition, dann kann die Performanz als Ausdruck des zugrundeliegenden Konzepts verstanden werden (Gautschi, 2010, 74).

Das skizzierte Verständnis von Lernen erscheint für die vorliegende Untersuchung als äusserst vielversprechend: Wie empirische Befunde zeigen, kommen Schüler*innen bereits mit einem normalerweise alltäglich geprägten und oft nicht den fachspezifischen Kriterien entsprechenden Konzept von *Erklären* in den Geschichtsunterricht. Die hier untersuchten Neuntklässler*innen, die in der Regel noch keine schriftlichen historischen Erklärungen verfassen muss-

ten, werden mit ihrem alltäglichen Konzept im mündlichen Geschichtsunterricht vermutlich trotzdem erfolgreich gewesen sein (Gropengießer & Marohn, 2018, 52), weswegen sie ihre epistemologischen Überzeugungen zum Fach Geschichte sowie ihre Konzepte historischen Erklärens höchstwahrscheinlich noch nicht hinterfragen mussten. Es ist also anzunehmen, dass Lernende im wahrsten Sinne des Wortes mit Alltagsvorstellungen zum Erklären in den Unterricht kommen. Diese sind aufgrund der oben skizzierten Gemeinsamkeiten (geschichts-)wissenschaftlichen und alltäglichen Erklärens nicht per se zurückzuweisen; Setting, Zielsetzung sowie die Bestandteile Explanandum und Explanans bestehen in beiden Kontexten. Allerdings reichen sie spätestens in schriftlichen Settings nicht aus und müssen durch fachspezifische Wissens-elemente ausgebaut und ergänzt werden. Somit ist ein radikaler *Conceptual Change* als vollständige Konzeptveränderung oder -ersetzung weder erwartbar noch wünschenswert. Vielmehr ist anzunehmen, dass die Untersuchungsteilnehmenden einzelne Wissensbausteine (also hier: Merkmale der Operationalisierung) in ihre Konzepte integrieren bzw. einzelne Wissensmerkmale ergänzen («*knowledge*» bzw. «*conceptual enrichment*», Limón, 2002, 260) oder alltägliche Merkmale durch wissenschaftsspezifische Merkmale ersetzen.

Um diese Lernprozesse anzuregen und zu fördern, greift das vorliegende Vorhaben (s. Abschnitt 4) auf bewährte Massnahmen zur Herbeiführung eines *Conceptual Change* zurück. Zentral erscheint die Kontrastierung des eigenen, (überwiegend) alltäglichen Konzepts mit dem zustimmungsfähigeren wissenschaftlichen Konzept – durch einen sogenannten kognitiven Konflikt (Larsson 2013, 19). Solche Konflikte werden in der Regel von den Lehrkräften als Angebot initiiert. Da sich die Konzepte im Alltag jedoch oft bewährt haben, sind die Konzepte gegen solche Veränderungsversuche meist sehr resistent. Daher ist die Förderung metakognitiven Wissens notwendig: Lernende benötigen Strategien und Wissen, um die fehlende Adäquatheit ihres Konzepts zu erkennen und das angebotene anzunehmen (Murphy & Mason, 2006, 314). Den Studien Chapmans und Stoels et al. folgend sowie einhergehend mit konstruktivistischen Lerntheorien (Gropengießer & Marohn, 2018, 60) sind kollaborative Lernsettings und die Bildung von Analogien zu den Schüler*innen bekannten Kontexten lernförderlich. In einer Durchsicht empirischer Studien wurden jüngst die explizite Modellierung der entsprechenden Operation, die Analyse von Beispieltexten, die Arbeit mit Merkblättern sowie mit multiperspektivischen Materialarrangements als weitere erfolgreiche Massnahmen identifiziert (Nitsche, Waldis, Gollin, Marti & Hubacher, 2023, 106).

4 Methodik

4.1 Design und Sample

Um die Konzepte der Lernenden zu historischem Erklären zu erheben (Forschungsfrage 1) sowie daran anschliessend eine Wissensveränderung herbeiführen zu können (Forschungsfrage 2), wurde eine quasi-experimentelle Studie nach dem Modell des *Nonequivalent Control Group Designs* von Campbell und Stanley konzipiert (Reichardt, 2019, 115). Gerade in pädagogischen Kontexten ist dieses Design aufgrund der ökonomischen Anwendbarkeit und der Wahrung externer Validität eines der am häufigsten verwendeten (Reichardt, 2019, 159), insbesondere zum Vergleich von Schulklassen (Huber, 2019, 194). Unter anderem wurde die Studie in zwei neunten Gymnasialklassen, einer Experimental- (E9) und einer Kontrollgruppe (K9) durchgeführt.

	Kontrolle Störvariablen ¹	Prätest	Treatment	Posttest	Follow-up- Test
E9 (n = 29)	0	0	X	0	0
K9 (n = 27)	0	0		0	0

Tab. 1: Design

Das Vorhaben folgt dem klassischen Ablauf eines Experiments mit Prä-, Post- und Follow-up-Test mit fünf Unterrichtsstunden als Treatment in den Experimentalgruppen zwischen Prä- und Posttest. Im Treatment wurden die Experimentalgruppen dem Prinzip forschend-entdeckenden Lernens (Wolter, 2018) folgend mit den operationalisierten Merkmalen historischen Erklärens vertraut gemacht. Sie wandten die historische Methode an authentischen Quellen sowie konzipierten Materialien an und erklärten sukzessive ein genuin historisches Explanandum, den Ausbruch und Ablauf des Frankfurter Aufstandes 1848. Sie erwarben also Wissen über narratives Erklären *in actu* (vgl. hierzu auch Giest,

1 Da mit dem quasi-experimentellen Design Einschränkungen hinsichtlich der Vergleichbarkeit der Ergebnisse in den Tests zu erwarten sind, wurde dem Prätest ein dreiteiliger Test vorgeschaltet, um potenzielle Störvariablen erheben und kontrollieren zu können: ein Fragebogen zu soziodemografischen Merkmalen, ein C-Test zur Erhebung der Sprachfähigkeiten (Grotjahn, 2002) sowie der HiTCH-Teilttest zur Erhebung fachspezifischer Sach- und Methodenkompetenzen (Trautwein et al., 2017).

2011, 70–73). Das Treatment folgte in seiner Konzeption einer Zusammenstellung empirisch validierter Methoden (s. Abschnitte 2 und 3) wie der Herbeiführung eines kognitiven Konflikts, der Explizierung der Merkmale, Gruppenarbeiten und Plenumsdiskussionen, der Bildung von Analogien sowie dem Transfer der kennengelernten Merkmale und Techniken auf Phänomene aus dem Alltag der Lernenden und insbesondere der Förderung metakognitiven und prozeduralen Wissens zu historischem Erklären. Tabelle 2 gibt einen Überblick über den Ablauf des Treatments, die intendierten Erkenntnisziele sowie die jeweils verwendeten Konstruktionsprinzipien.

Stunde	Ziele: Die SuS lernen ...	Konstruktionsprinzipien
1) Was ist historisches Erklären?	... dass historische Erklärungen in Form einer Geschichte vorliegen. ... dass historische Explananda Veränderungen eines Gegenstandes / Referenzsubjekts in der Vergangenheit darstellen.	Kognitiver Konflikt (Ebene der Konzepte) Modelltext Analogiebildung Explizierung: Merksätze
2) Wie arbeiten Historiker*innen? I: Perspektivität von Quellen	... dass Quellen perspektivisch sind und Historiker*innen für eine Erklärung mehrere Materialien heranziehen, ... dass und wie man sich sprachlich von Quellenaussagen abgrenzen muss.	Forschend-entdeckendes Lernen Kognitiver Konflikt (Ebene der Quellen: Biperspektivisches Arrangement) Analogiebildung Explizierung: Merksätze
3) Wie arbeiten Historiker*innen? II: Bezug auf Theorien	... dass Historiker*innen zur Einordnung ihres singulären Explanandums allgemeinere Theorien und Darstellungen benötigen.	Modelltext Metadiskussion Explizierung: Merksätze
4) Wie strukturiere ich eine historische Erklärung?	... wie Erklärungen auf der Makroebene strukturiert sind, ... welche Bedeutung sprachlichen Verknüpfungen auf der Mikroebene zukommt.	Modelltext Metadiskussion Explizierung: Merksätze
5) Wie verorte ich mich?	... dass auch sie eine eigene Perspektive auf den zu erklärenden Gegenstand haben und dass sie diese reflektieren und aufdecken sollten.	Reflexion: Gemeinsamkeiten und Unterschiede der Erklärungen? Explizierung: Merksätze

Tab. 2: Ablauf des Treatments

4.2 Instrumente

Die Untersuchungsteilnehmenden wurden in den drei Tests dazu aufgefordert, aufzuschreiben, was sie unter *Erklären* im Fach Geschichte verstehen. Dabei wird der Annahme gefolgt, dass die aufgeschriebenen Antworten Rückschlüsse auf die zugrundeliegenden Konzepte erlauben. Die Antworten werden angelehnt an Mayrings Verfahren der skalierenden Strukturierung (2015, 106) deduktiv auf ihre Zustimmungsfähigkeit im Sinne der Operationalisierung eingeschätzt. So ist anzunehmen, dass die Antworten der Lernenden durch Anführung von Merkmalen auf einem Kontinuum zwischen alltäglichen und fachspezifischen Erklärungen liegen. Als Indikatoren dienen die Merkmale der Operationalisierung, die in Kodes überführt und in einem Kodierleitfaden mittels Ankerbeispielen definiert wurden. Ergänzend wurde die Spezifik von Explanandum bzw. Explanans kodiert. Zwar gehen beide Kategorien als grundlegende Bestandteile von Erklärungen in der Operationalisierung auf, jedoch war anzunehmen, dass im Prätest sowie in den Kontrollgruppen der teilweise hohe abstrakte Anspruch der beiden Kategorien nicht erreicht, dafür aber spezifische Angaben gemacht würden. Im Interrating zweier Kodierender an 30 zufällig ausgewählten Beispielen (15 % des Materials, was i. d. R. als ein ausreichender Anteil gilt, Schreier, 2012, 198) wurden Werte für Cohens Kappa zwischen 0.73 und 1.00 erzielt, was nach Landis und Koch als *substantial* (0.61–0.8) bzw. *almost perfect* (0.81–1.00) interpretiert werden kann (1977, 165) und dem Kodierleitfaden eine (sehr) gute Anwendbarkeit bescheinigt. Im Auswertungsprozess werden aus Gründen der Übersichtlichkeit nur vorhandene Kodes eingetragen.

Kode	Beschreibung
1.1 Form	Eine historische Erklärung liegt in Form einer Geschichte mit triadischer Struktur von Einleitung, Hauptteil und Schluss vor.
1.2 Explanandum als temporal-kausale Veränderung in der Vergangenheit	Das Explanandum wird als kausal bedingte Veränderung (eines Gegenstandes / Referenzsubjekts) in der Vergangenheit mit möglichen Bezügen zur Gegenwart charakterisiert.
1.3 Theoriebezug	Das Explanandum wird als ein konkreter Fall einer allgemeinen Entwicklung spezifiziert.
1.4 Multipropositionales Explanans	Historische Explananda werden multikausal erklärt, also durch mehrere Propositionen im Hauptteil.
1.5 Mikroebene	Zwischen den Propositionen sind funktionale zustimmungsfähige sprachliche Verknüpfungen notwendig.
1.6 Materialbezug / empirische Triftigkeit	Die Propositionen stammen aus Quellen (und Darstellungen), auf die verwiesen werden muss.
1.7 Perspektivenreflexion/ normative Triftigkeit	Da Quellen und Darstellungen perspektivisch sind, müssen mehrere Materialien für die Erklärung herangezogen und die gewonnenen Propositionen sprachlich distanzisiert (z. B. durch den Konjunktiv) wiedergegeben werden.
1.8 Reflexion eigener Standortgebundenheit	Der retrospektive Sehepunkt des* der Erklärenden muss offengelegt und reflektiert werden.
2.1 spezifisches Explanandum	Die Antwort beinhaltet ein als spezifisch für den Geschichtsunterricht / die Geschichtswissenschaft einzuschätzendes Explanandum, z. B. «Vergangenheit», «vergangene Ereignisse».
2.2 unspezifisches Explanandum	Die Antwort beinhaltet ein für die Geschichtswissenschaft / den Geschichtsunterricht unspezifisches bzw. ein alltägliches Explanandum, z. B. «Begriffe», «Sachverhalt».
3.1 spezifisches Explanans	Die Antwort beinhaltet eine oder mehrere als spezifisch für den Geschichtsunterricht / die Geschichtswissenschaft einzuschätzende Proposition(en) des Explanans, z. B. «kausaler Zusammenhang».
3.2 unspezifisches Explanans	Die Antwort beinhaltet eine oder mehrere als unspezifisch für den Geschichtsunterricht / die Geschichtswissenschaft einzuschätzende Proposition(en) des Explanans, z. B. «Informationen».

Tab. 3: Auszug aus dem Kodierleitfaden

5 Explorative Analyse

Die folgende explorative Analyse illustriert die Anwendbarkeit des Kodierleitfadens anhand der Antworten zweier Lernender der neunten Experimentalklasse, E9.5 und E9.15, über die drei Testzeitpunkte. Es wird analysiert, inwiefern sich die dahinterliegenden Konzepte verändert haben. Die beiden Proband*innen wurden exemplarisch ausgewählt, um erste Ergebnisse hinsichtlich sowohl unterschiedlich spezifischer Antworten im Prätest als auch durch das Treatment angestossene Lernprozesse (E9.5) bzw. deren überwiegendes Ausbleiben (E9.15) aufzuzeigen. Die Antworten wurden unter Beibehaltung der Originalschreibweise, also inklusive möglicher sprachlicher Fehler, transkribiert.

5.1 Beispiel 1 – Schüler E9.15

Antwort	Kodes
Durch «erklären» bringt man anderen was über Geschichte ^{2.1/3.2} bei. (E9.15_1.1, Prätest)	^{2.1} Spezifisches Explanandum ^{3.2} Unspezifisches Explanans
Unter «erklären» im Fach Geschichte verstehe ich, das übermitteln von Informationen die mit Quellen bewiesen werden können ^{1.6} . Mit diesen Information ^{3.2} kann man sich z. B. den Ursprung oder sonstiges von dem Gegenstand ^{2.2} herleiten. Mit diesen Ergebnissen über den Gegenstand kommt man eine Antwort oder Lösung auf die Frage oder das Problem. (E9.15_2.1, Posttest)	^{1.6} Materialbezug / empirische Triftigkeit ^{2.2} Unspezifisches Explanandum ^{3.2} Unspezifisches Explanans
Unter «erklären» im Fach Geschichte verstehe ich das Erzählen/ übermitteln von historischen Ereignissen ^{2.1/3.2} . Diese Ereignisse sollte man anhand von Beweisen etc. wiederlegen ^{1.6} können. (E9.15_3.1, Follow-up-Test)	^{1.6} Materialbezug / empirische Triftigkeit ^{2.1} Spezifisches Explanandum ^{3.2} Unspezifisches Explanans

Tab. 4: Antworten von Schüler E9.15

Blickt man auf die drei Antworten von E9.15, so fällt zunächst die unterschiedliche Länge auf: Die Antwort im Posttest (50 Wörter) ist ausführlicher als im Follow-up- (23 Wörter) und Prätest (9 Wörter). Interessanterweise verhält sich

hier die Qualität (Fachspezifik historischen Erklärens) wie die Quantität: E9.15_2.1 ist spezifischer als E9.15_3.1 und vor allem als E9.15_1.1.

Im Prätest findet sich, wohlwollend kodiert, eine spezifische Angabe zum Explanandum («was über Geschichte», Kode 2.1 spezifisches Explanandum), die Angabe zum Explanans wird tautologisch mit «Durch Erklären» gefüllt (3.2 unspezifisches Explanans). Merkmale der Operationalisierung bzw. epistemologische Spezifika fehlen. Die Antwort zeugt von einem recht unspezifischen Verständnis historischen Erklärens, lediglich allgemeine Merkmale wie das Setting und das Ziel, anderen etwas beizubringen, sind induktiv identifizierbar.

Dies ändert sich nach dem Treatment im Posttest: Das Explanandum ist spezifisch(er), der «Ursprung» eines «Gegenstands» soll hergeleitet werden und damit implizit die temporalen und kausalen Bezüge innerhalb der Vergangenheit. Den abstrakten Begriff «Gegenstand» haben die Lernenden im Zuge des Treatments kennengelernt und E9.15 scheint ihn verinnerlicht zu haben. Da das Abstraktum jedoch nicht zustimmungsfähig konkretisiert wird, wurde der Kode 3.2 unspezifisches Explanandum vergeben. Einen erfolgreichen Einfluss hatte das Treatment aber offenbar in der Vermittlung der Bedeutsamkeit empirischer Triftigkeit (Kode 1.6): E9.15 betont die Notwendigkeit, die Propositionen – unspezifisch als «Informationen» (Kode 3.2) bezeichnet – mit Quellen zu beweisen.

Die Antwort im Follow-up-Test, drei Monate nach dem Posttest, stellt gewissermassen eine Mischung der Antworten aus Prä- und Posttest dar: Der Versuch, das Abstraktum des Treatments (Veränderung eines Gegenstandes in der Vergangenheit) zu deklarieren, wird hier zugunsten eines spezifischen Explanandums, «historische Ereignisse», aufgegeben (Kode 2.1). Das Explanans ist erneut nicht spezifiziert (Kode 3.2). Das Merkmal der empirischen Triftigkeit scheint hingegen behalten worden zu sein: Sprachlich wohlwollend kodiert, sollten die zu erklärenden Ereignisse «anhand von Beweisen etc.» belegt («wiederlegen») werden können.

5.2 Beispiel 2 – Schülerin E9.5

Antwort	Kodes
<p>Unter «erklären» im Fach Geschichte verstehe ich, dass man genau erklärt, was passierte. (Also sehr umfangreich) Man sollte auch die Hintergründe und Folgen^{1,4} erwähnen, da diese sehr relevant sind, damit verstanden wird, warum es zu einer bestimmten Situation kommt^{1,2}. Man sollte sehr ausführlich alles begründen & erzählen. (E9.5_1.1, Prätest)</p>	<p>^{1,2} Explanandum als temporal-kausale Veränderung in der Vergangenheit ^{1,4} Multipropositionales Explanans</p>
<p>Ich verstehe unter «erklären» im Fach Geschichte, dass man alles beachten muss. Man sollte sich an die Fakten halten und mit Beweisen oder Quellen arbeiten.^{1,6} Ebenfalls sollte man gucken, was die Gründe dazu waren^{1,4}, dass es zu der Situation o. Ä. kam^{1,2}. Ebenfalls sollte der Verlauf und die Folgen erläutert werden. Man sollte mit unterschiedlichen Quellen gearbeitet werden, damit man andere Sichten hat^{1,7}, wie es für manche Leute eben die Situation war. (E9.5_2.1, Posttest)</p>	<p>^{1,2} Explanandum als temporal-kausale Veränderung in der Vergangenheit ^{1,4} Multipropositionales Explanans ^{1,6} Materialbezug / empirische Triftigkeit ^{1,7} Perspektivenreflexion / normative Triftigkeit</p>
<p>Unter «erklären» im Fach Geschichte verstehe ich, dass man sich an verschiedenen Punkten orientiert. Man sollte die Situation erklären, wieso es dazu kam, wer daran beteiligt war. und welche Folgen es mit sich nahm.^{1,2 / 3.1} Man muss sich an die Fakten halten und Beweise haben^{1,6}. Man verwendet zum Beispiel Ausschnitte aus der Zeitung der jeweiligen Zeit^{1,6}, doch dann sollte man sich auch verschiedene anschauen, damit man vielleicht sieht, wo es Meinungsverschiedenheiten gibt^{1,7}. (E9.5_3.1, Follow-up-Test)</p>	<p>^{1,2} Explanandum als temporal-kausale Veränderung in der Vergangenheit ^{1,6} Materialbezug / empirische Triftigkeit ^{1,7} Perspektivenreflexion / normative Triftigkeit ^{3.1} Spezifisches Explanans</p>

Tab. 5: Antworten von Schülerin E9.5

Die Schülerin E9.5 unterscheidet sich in drei Hinsichten massgeblich von E9.15. Auffallend ist *erstens* die größere Ausführlichkeit der Antworten (47 Worte im Prätest, 71 Worte im Post- und im Follow-up-Test). *Zweitens* ist bereits die Antwort im Prä-Test deutlich zustimmungsfähiger: Als temporal-kausale Ver-

änderung in der Vergangenheit wurde «warum es zu einer bestimmten Situation kommt» mit Kode 1.2 zustimmungsfähig kodiert. Auch das Explanans ist bereits multipropositional («Hintergründe und Folgen») angelegt (Kode 1.4). E9.5 deklariert also vor dem Treatment zustimmungsfähig im Sinne der Operationalisierung die beiden grundlegenden Bestandteile historischen Erklärens. Beides wird, in auffallend ähnlichen Formulierungen («dass es zu der Situation kam», «Hintergründe und Folgen»), auch im Posttest angeführt (Kodes 1.2 und 1.4). Hinzu kommen, *drittens*, zwei weitere Merkmale der Operationalisierung im Posttest, was den Schluss auf ein erfolgreiches Treatment erlaubt: So wird die Bedeutsamkeit empirischer Triftigkeit (Kode 1.6) hervorgehoben («Man sollte sich an die Fakten halten und mit Beweisen oder Quellen arbeiten»). Zudem führt die Schülerin Aspekte der normativen Triftigkeit (Kode 1.7) an, wenn sie die Notwendigkeit multiperspektivischen Arbeitens betont: «Man sollte mit unterschiedlichen Quellen gearbeitet werden, damit man andere Sichten hat, wie es für manche Leute eben die Situation war.»

Bemerkenswert ist, dass im Follow-up-Test beide erstmals im Posttest deklarierten und auf das Treatment zurückzuführenden Merkmale, empirische (Kode 1.6) und normative Triftigkeit (Kode 1.7), erneut angeführt werden. Die Formulierungen sind wieder ähnlich jenen im Posttest, die Angabe zur Quellenarbeit wird sogar exemplarisch «zum Beispiel Ausschnitte aus der Zeitung der jeweiligen Zeit» mit Bezug zum Treatment ausdifferenziert. Wie bei E9.15, der die Bedeutsamkeit empirischer Triftigkeit verinnerlicht zu haben scheint, deutet dies bei E9.5 auf eine Aufnahme der beiden Merkmale in das Konzept historischen Erklärens hin. Stabil bleibt auch im Follow-up-Test die Formulierung des Explanandums als temporal-kausale Veränderung in der Vergangenheit (Kode 1.2 «die Situation erklären, wieso es dazu kam»). Lediglich die Angabe zum Explanans wird weniger zustimmungsfähig, der multipropositionale Aspekt geht verloren: «wieso es dazu kam, wer daran beteiligt war» (Kode 3.1 spezifisches Explanans).

6 Fazit und Ausblick

Die explorative Analyse zweier Lernender hat die Anwendbarkeit des Kodierleitfadens zur Analyse schüler*innenseitiger Konzepte historischen Erklärens gezeigt. Zwar sind die Antworten im Prätest hinsichtlich der Zustimmungsfähigkeit von Explanandum und Explanans heterogen. Aber selbst die elaborierte Antwort E9.5_1.1 beinhaltet keine epistemologischen Spezifika des Faches. Er-

wartungskonform sind beide Antworten im Prätest relativ unspezifisch und lassen sich so in bereits bestehende Erkenntnisse zu schüler*innenseitigem historischem Erklären einordnen (Forschungsfrage 1). Mit Blick auf Forschungsfrage 2 zeigen beide Schüler*innen, dass epistemologische Besonderheiten der Disziplin als Grundlage für historisches Erklären im Treatment vermittelt werden konnten. Im Post- und auch drei Monate später im Follow-up-Test deklariert E9.15 zustimmungsfähig ein, E9.5 zwei Merkmale der Operationalisierung. Das in den Antworten deklarierte zustimmungsfähige Wissen über historisches Erklären nimmt von Prä- zu Posttest bei beiden Proband*innen zu und lässt auf einen Wissensausbau (*«knowledge enrichment»*) durch Aufnahme einzelner operationalisierter Merkmale schliessen. Die vergleichbaren Formulierungen über alle Testzeitpunkte von E9.5 lassen zudem den Schluss zu, dass bestehende Konzeptanteile dauerhaft angelegt und resistent gegen Veränderungen zu sein scheinen, jedoch auch ergänzt werden können. Für die Pragmatik des (sprachsensiblen) Geschichtsunterrichts bedeutet dies, dass Konzepte fachspezifischer Sprachhandlungen durch das hier entwickelte Treatment, u. a. durch gezielte Förderung metakognitiven Wissens über Prinzipien und Methodik der dem Unterricht zugrundeliegenden Disziplin, gefördert werden können; die hier vorgestellten Analyseergebnisse zeigen, dass Lernende neu erworbenes Wissen (in unterschiedlicher Tiefe und Breite) deklarieren können. Mit dieser Erkenntnis sagt der vorliegende Beitrag selbstredend noch nichts über den Einfluss des Konzepts auf die Performanz historischen Erklärens aus. Allerdings kann vor dem Hintergrund der vorgestellten empirischen Befunde und lerntheoretischen Annahmen die Hypothese formuliert werden, dass die Förderung des Konzepts auch zu besserer Performanz führt. Dies wird u. a. im zugrundeliegenden Vorhaben untersucht.

Die Befunde verweisen auf weiteren Forschungsbedarf: Obwohl die Antworten beider Schüler*innen zustimmungsfähiger werden, fällt auf, dass nur Teile der Operationalisierung angeführt werden. Das Treatment war also nur in Teilen erfolgreich; neben einer Überprüfung der Entwicklungen im Gesamtsample müssten in Zukunft also insbesondere auch individuelle Lernmerkmale einzelner Schüler*innen berücksichtigt und statistisch analysiert werden, beispielsweise um herauszufinden, warum das Treatment bei E9.5 erfolgreicher verlief als bei E9.15.

Literatur

- Altun, Tülay & Günther, Katrin (2018). *Begründen* als Arbeitsauftrag im Geschichtsunterricht. In Katharina Grannemann, Sven Oleschko & Christian Kuchler (Hrsg.), *Sprachbildung im Geschichtsunterricht: Zur Bedeutung der kognitiven Funktion von Sprache* (S. 157–176). Münster: Waxmann.
- Ammerer, Heinrich (2022). *Konzepte historischen Denkens und ihre Entwicklungslogik. Eine Studie zur Genese historischer Verständnishorizonte*. Frankfurt a. M.: Wochenschau.
- Barricelli, Michele (2012). Narrativität. In Michele Barricelli & Martin Lücke (Hrsg.), *Handbuch Praxis des Geschichtsunterrichts* (Bd. 1, S. 255–280). Schwalbach/Ts.: Wochenschau.
- 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: Sprachliche Vielfalt und historisches Lernen* (S. 137–163). Göttingen: Vandenhoeck & Ruprecht.
- Carretero, Mario / Jacott, Liliana & Limón, Margarita (1994). Historical knowledge: Cognitive and instructional implications. In Mario Carretero & James Voss (Hrsg.), *Cognitive and instructional processes in history and the social sciences* (S. 357–376). Hillsdale NJ: Erlbaum.
- Chapman, Arthur (2017). Causal explanation. In Ian Davies (Hrsg.), *Debates in history teaching* (2. Aufl., S. 130–143). London: Routledge.
- Dammer, Karl-Heinz (2007). Bildung nach Maß: Über den tieferen Sinn der Forderung nach Bildungsstandards im Abitur. *Pädagogische Korrespondenz*, (37), 44–62.
- Danto, Arthur Coleman (1974). *Analytische Philosophie der Geschichte*. (J. Behrens, Übers.) Frankfurt a. M.: Suhrkamp. (Original erschienen 1965)
- Ehlich, Konrad (2009). Erklären verstehen – Erklären und Verstehen. In Rüdiger Vogt (Hrsg.), *Erklären: Gesprächsanalytische und fachdidaktische Perspektiven* (S. 11–24). Tübingen: Stauffenburg.
- Evans, Jennifer / Pate, Gemma. (2007). Does scaffolding make them fall? Reflecting on strategies for developing causal argument in Years 8 and 11. *Teaching History*, (128), 18–29.
- Feilke, Helmuth (2012). Bildungssprachliche Kompetenzen – fördern und entwickeln. *Praxis Deutsch*, (233), 4–13.
- Fenn, Monika & Zülsdorf-Kersting, Meik (2023). Historisches Lernen: Ansätze einer Lerntheorie für den Geschichtsunterricht. In Monika Fenn & Meik Zülsdorf-Kersting (Hrsg.), *Geschichtsdidaktik. Praxishandbuch für den Geschichtsunterricht* (S. 54–97). Berlin: Cornelsen.

- Gautschi, Peter (2010). Wissen – Voraussetzung und Ergebnis von historischem Lernen. *Zeitschrift für Didaktik der Gesellschaftswissenschaften*, 1, 67–90.
- Gerber, Doris (2012). *Analytische Metaphysik der Geschichte: Handlungen, Geschichten und ihre Erklärung*. Berlin: Suhrkamp.
- Gerstenmaier, Jochen / Mandl, Heinz (1995). Wissenserwerb unter konstruktivistischer Perspektive. *Zeitschrift für Pädagogik* 41(6), 867–888.
- Giest, Hartmut (2011). Wissensaneignung, Conceptual Change und die Lehrstrategie des Aufsteigens vom Abstrakten zum Konkreten. *E-Journal for Activity Theoretical Research in Germany*, 4, 65–100. Abgerufen von <https://docplayer.org/130516117-Wissensaneignung-conceptual-change-und-die-lehrstrategie-des-aufsteigens-vom-abstrakten-zum-konkreten.html>.
- Gropengießer, Harald & Marohn, Annette (2018). Schülervorstellungen und *Conceptual Change*. In Dirk Krüger, Ilka Parchmann & Horst Schecker (Hrsg.), *Theorien in der naturwissenschaftlichen Forschung* (S. 49–68). Berlin: Springer.
- Grotjahn, Rüdiger (2002). Konstruktion und Einsatz von C-Tests: Ein Leitfaden für die Praxis. In Rüdiger Grotjahn (Hrsg.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen* (Bd. 4, S. 211–225). Bochum: AKS.
- Gruber, Hans & Stamouli, Eleni (2015). Intelligenz und Vorwissen. In Elke Wild & Jens Möller (Hrsg.), *Pädagogische Psychologie* (S. 25–44). 2. Aufl. Berlin / Heidelberg: Springer.
- Günther-Arndt, Hilke (2006) Conceptual Change-Forschung: Eine Aufgabe für die Geschichtsdidaktik? In Dies. & Michael Sauer (Hrsg.), *Geschichtsdidaktik empirisch: Untersuchungen zum historischen Denken und Lernen* (S. 251–278). Berlin: LIT.
- Halldén, Ola (1998). Personalization in historical descriptions and explanations. *Learning and Instruction*, 8(2), 131–139.
- Handro, Saskia (2013). Sprache und historisches Lernen: Dimensionen eines Schlüsselproblems des Geschichtsunterrichts. In Michael Becker-Mrotzek, Karen Schramm, Eike Thürmann & Helmut Johannes Vollmer (Hrsg.), *Sprache im Fach: Sprachlichkeit und fachliches Lernen* (S. 317–334). Münster: Waxmann.
- Hartung, Olaf (2013). *Geschichte schreiben lernen: Empirische Erkundungen zum konzeptionellen Schreibhandeln im Geschichtsunterricht*. Berlin / Münster: LIT.
- Hartung, Olaf (2016). Generische Lernaufgaben im Geschichtsunterricht – oder: die «zwei Seiten» einer Gattungskompetenz. In Saskia Handro & Bernd Schönemann (Hrsg.), *Aus der Geschichte lernen? Weiße Flecken der Kompetenzdebatte* (S. 187–198). Berlin: LIT.
- Haussmann, Thomas. (1991). *Erklären und Verstehen: Zur Theorie und Pragmatik der Geschichtswissenschaft*. Frankfurt a. M.: Suhrkamp.

- Hempel, Carl G. (1942). The function of general laws in history. *The Journal of Philosophy*, 39(2), 35–48. Abgerufen von <https://doi.org/10.2307/2017634>.
- Hodel, Jan / Waldis, Monika / Zülsdorf-Kersting, Meik & Thünemann, Holger (2013). Schülernarrationen aus Ausdruck historischer Kompetenz. *Zeitschrift für Didaktik der Gesellschaftswissenschaften*, 4(2), 121–145.
- Huber, Oswald (2019). *Das psychologische Experiment. Eine Einführung* (7. Aufl.). Bern: Hogrefe.
- Husemann, Charlotte (2022). *Geschichte beschreiben, Geschichte erklären. Eine Untersuchung fachsprachlicher Konzepte und fachlicher Sprachhandlungsfähigkeit von Gesamtschüler*innen der Sekundarstufe I*. Göttingen: Vandenhoeck & Ruprecht.
- Jahr, Silke (2000). Vertextungsmuster Explikation. In Klaus Brinker, Gerd Antos, Wolfgang Heinemann & Sven F. Sager (Hrsg.), *Text- und Gesprächslinguistik: Ein internationales Handbuch zeitgenössischer Forschung*. (1. Bd., S. 385–397). Berlin / New York: De Gruyter.
- Klein, Josef (2009). Erklären-Was, Erklären-Wie, Erklären-Warum: Typologie und Komplexität zentraler Akte der Welterschließung. In Rüdiger Vogt (Hrsg.), *Erklären: Gesprächsanalytische und fachdidaktische Perspektiven* (S. 25–36). Tübingen: Stauffenburg.
- Kultusministerkonferenz (2005). *Einheitliche Prüfungsanforderungen in der Abiturprüfung Geschichte*. Abgerufen von https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1989/1989_12_01-EPA-Geschichte.pdf.
- Landis, Richard J. & Koch, Gary G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33(1), 159–174.
- Larsson, Åsa (2013). *A structural view on conceptual change. integration, differentiation, and contextualization as fundamental aspects of individual meaning making*. Stockholm: Stockholm University Press.
- Lee, Peter / Dickinson, Alaric & Ashby, Rosalyn (1998). Researching children's ideas about history. In James Voss & Mario Carretero (Hrsg.), *International review of history education. Learning and reasoning in history* (Bd. 2, S. 227–251). London: Woburn.
- Limón, Margarita (2002). Conceptual change in history. In Margarita Limón & Lucia Mason (Hrsg.), *Reconsidering conceptual change: Issues in theory and practice* (S. 259–289). Dordrecht: Springer.
- Lorenz, Chris (1997). *Konstruktion der Vergangenheit: Eine Einführung in die Geschichtstheorie*. Köln: Böhlau.
- Mathis, Christian (2015). *«Irgendwie ist doch da mal jemand geköpft worden»: Didaktische Rekonstruktion der Französischen Revolution und der historischen Kategorie Wandel*. Baltmannsweiler: Schneider Hohengehren.

- Mayring, Philipp (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12. Aufl.). Weinheim: Beltz.
- Mierwald, Marcel & Brauch, Nicola (2015). Historisches Argumentieren als Ausdruck historischen Denkens. Theoretische Fundierung und empirische Annäherungen. *Zeitschrift für Geschichtsdidaktik*, 14(1), 104–120.
- Murphy, Kerry & Mason, Lucia (2006). Changing knowledge and beliefs. In Patricia A. Alexander & Philip Winne (Hrsg.), *Handbook of educational psychology* (2. Aufl., S. 305–324). Mahwah, NJ: Erlbaum.
- Nitsche, Martin/Waldis, Monika/Gollin, Kristine/Marti, Philipp & Hubacher, Manuel (2023). Narrative Kompetenz und epistemologische Beliefs im Geschichtsunterricht fördern? Zur Wirksamkeit einer Schreibintervention an Deutschschweizer Gymnasien. In Monika Waldis & Martin Nitsche (Hrsg.), *Geschichtsdidaktisch intervenieren* (S. 97–127). Bern: hep.
- Oleschko, Sven & Moraitis, Anastasia (2012). Die Sprache im Schulbuch: Erste Überlegungen zur Entwicklung von Geschichts- und Politikbüchern unter Berücksichtigung sprachlicher Besonderheiten. *Bildungsforschung*, 9(1), 11–46.
- Oleschko, Sven (2015). Herausforderungen einer domänenspezifischen Sprachdiagnostik im Kontext historischen Lernens. In *Zeitschrift für Geschichtsdidaktik*, 14(1), 87–103.
- Pandel, Hans-Jürgen (2010). *Historisches Erzählen: Narrativität im Geschichtsunterricht*. Schwalbach/Ts.: Wochenschau.
- Pandel, Hans-Jürgen (2017). *Geschichtstheorie: Eine Historik für Schülerinnen und Schüler – aber auch für ihre Lehrer*. Schwalbach/Ts.: Wochenschau.
- Peters, Jelko (2015). Schriftliches Argumentieren in den Fächern Deutsch und Geschichte: Ein Vergleich der Operatoren für die Abiturprüfung. In Alexandra Budke / Miriam Kuckuck / Michael Meyer / Frank Schäbitz / Kirsten Schlüter & Günther Weiss (Hrsg.), *Fachlich argumentieren lernen: Didaktische Forschungen zur Argumentation in den Unterrichtsfächern* (S. 298–310). Münster: Waxmann.
- Reichardt, Charles S. (2019). *Quasi-experimentation. A guide to design and analysis*. New York: Guilford.
- Rüsen, Jörn (1986). *Rekonstruktion der Vergangenheit: Grundzüge einer Historik II: Die Prinzipien der historischen Forschung*. Göttingen: Vandenhoeck & Ruprecht.
- Rüsen, Jörn (2013). *Historik: Theorie der Geschichtswissenschaft*. Köln: Böhlau.
- Schönemann, Bernd / Thünemann, Holger & Zülsdorf-Kersting, Meik (2011). *Was können Abiturienten? Zugleich ein Beitrag zur Debatte um Kompetenzen und Standards im Fach Geschichte* (2. Aufl.). Berlin: LIT.

- Schraw, Gregory (2006). Knowledge: Structures and Processes. In Patricia A. Alexander & Philip H. Winne (Hrsg.), *Handbook of Educational Psychology* (2. Aufl., S. 245–263). Mahwah, NJ: Erlbaum.
- Schreier, Margrit (2012). *Qualitative content analysis in practice*. Los Angeles: Sage.
- Stark, Robin (2003). Conceptual Change: kognitiv oder situiert? *Zeitschrift für Pädagogische Psychologie*, 17(2), 133–144.
- Stoel, Gerhard, L. / van Drie, Jannet & van Boxtel, Carla (2017). The effects of explicit teaching of strategies, second-order concepts, and epistemological under-pinnings on students' ability to reason causally in history. *Journal of Educational Psychology*, 109(3), 321–337.
- Trautwein, Ulrich / Bertram, Christiane / Borries, Bodo von / Brauch, Nicola / Hirsch, Matthias / Klausmeier, Kathrin / ... Zuckowski, Andreas (2017). *Kompetenzen historischen Denkens erfassen: Konzeption, Operationalisierung und Befunde des Projekts «Historical Thinking – Competencies in History» (HiTCH)*. Münster / New York: Waxmann.
- Welskopp, Thomas (2002). Erklären. In *Lexikon der Geschichtswissenschaft: Hundert Grundbegriffe* (S. 81–84). Stuttgart: Reclam.
- Wickner, Mareike-Cathrine (2019). Über die Vorzüge einer textsortenbasierten Schreibförderung im Geschichtsunterricht. Forschungsergebnisse aus dem SchriFT-Projekt. In Heike Roll / Markus Bernhardt / Christine Enzenbach / Hans Fischer / Erkan Gürsoy / Heiko Krabbe / ... Işıl Uluçam-Wegmann (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.
- Wolter, Heike (2018). *Forschend-entdeckendes Lernen im Geschichtsunterricht*. Frankfurt a. M.: Wochenschau.

Zur sprachbasierten Diagnose historischen Denkens in Texten von Schüler*innen

Eine empirisch gestützte Typologie

Viola Schrader

1 Problematisierung

Das Verhältnis von Sprache und historischem Denken ist konstitutiv (z. B. Günther-Arndt, 2010; Bernhardt, 2020, 32–36). Angesichts dieses theoretisch bedeutsamen Verhältnisses wurde der Sprache im Geschichtsunterricht in der Geschichtsdidaktik, insbesondere seit PISA, bereits einige Aufmerksamkeit geschenkt (z. B. Bernhardt & Conrad, 2018; Sandkühler & Bernhardt, 2020). Die deutschsprachige Geschichtsbewusstseinsforschung hat in den vergangenen 20 Jahren etwa das historische Begriffsverständnis (z. B. Langer-Plän, 2003; Schönemann, Thünemann & Zülsdorf-Kersting, 2011), das historische Textverstehen (z. B. von Borries, Fischer, Leutner-Ramme & Meyer-Hamme, 2005; Köster, 2013; Handro & Kilimann, 2019) sowie narrative Kompetenzen (z. B. BarriCELLI, 2005; Hodel, Thünemann, Waldis & Zülsdorf-Kersting, 2013; Waldis, Marti & Nitsche, 2015) der Lernenden als sprachgebundene Herausforderungen fachlichen Lernens konturiert. Wie die geschichtsdidaktischen Studien gezeigt haben, sind dabei Sprachhandlungen Lernender für die fachspezifische Diagnose und Beurteilung historischer Denk- und Lernleistungen zentral (z. B. Schönemann et al., 2011, 124; Hartung & Memminger, 2017, 6–9; Waldis, Nitsche & Gollin, 2020, 90–93, 106–108). In der Geschichtsdidaktik wurde und wird deswegen das Anliegen formuliert, ein sprachbasiertes Instrument zur Diagnose historischen Denkens zu entwickeln (z. B. Günther-Arndt, 2010, 45–46; Bernhardt & Wickner, 2015, 281–288; Handro, 2018, 21–22, 25; Marti, Gollin & Waldis, 2019, 166–167; Nitsche & van Norden, 2020). In welchem Zusammenhang sprachliche und fachliche Kompetenzen dabei stehen, wo und wie sprachliche Strukturen in Äußerungen von Lernenden auf historische Denkstrukturen verweisen und wie sprachliches Handeln Lernender letztendlich auch als ein Indikator für die graduierende Beurteilung historischen Denkens in Empirie und Pragmatik genutzt

werden kann, gilt es in der Geschichtsdidaktik weiterhin zu erforschen bzw. müssen vorhandene Ergebnisse im Hinblick auf die fachspezifische Diagnose systematisiert werden.

Vor diesem Hintergrund werden im Beitrag zentrale Ergebnisse der Studie «Historisches Denken und sprachliches Handeln» (Schrader, 2021) als Diskussionsbeitrag zur sprachbasierten Diagnose historischen Denkens vorgestellt. In der Studie wurden Texte Lernender der 10. Gymnasialklasse zum Vergleich und zur Beurteilung kontroverser Historikertexte hinsichtlich folgender Fragen untersucht: Welche konkreten sprachlichen Strukturen zeichnen die Texte der Lernenden aus? Wie repräsentiert sich im Sprachgebrauch auf Text-, Satz- und Wortebene historisches Denken? Wie lässt sich ein fachspezifisches Sprachhandeln von einem fachunspezifischen abgrenzen?

Als didaktische Konsequenzen der Ergebnisse stellten sich ferner folgende Fragen: Was bedeuten die herausgestellten Zusammenhänge für die Diagnose historischen Denkens in sprachlichen Handlungen Lernender im Geschichtsunterricht? Wo bestehen eventuelle sprachlich verankerte Graduierungsmöglichkeiten? Um diese Fragen beantworten zu können, wurden die Ergebnisse der Textanalysen in einer Typologie verdichtet, die vier Texttypen als heterogene Repräsentationsformen historischen Denkens und sprachlichen Handelns beschreibt. Im Beitrag werden diese Typen aufgegriffen, um an ihnen empirisch basierte Perspektiven für die Entwicklung eines fachspezifischen Diagnoseinstruments zu gewinnen. Um die Typologie in dieser Hinsicht zu kontextualisieren, erfolgt zunächst die Skizzierung bisheriger relevanter Forschungszusammenhänge sowie die Darlegung zentraler theoretischer Grundlagen und methodischer Entscheidungen der Studie.

2 Forschungszugänge

Für die Fragestellungen der Studie waren Forschungen besonders relevant, die sich an der Schnittstelle von Linguistik und Fachdidaktik bewegen, da es um die Generierung sowohl geschichtsdidaktischer als auch sprachlicher Kategorien und den Anschluss an entsprechende Untersuchungsansätze ging. Dies ist v. a. in der internationalen geschichtsdidaktischen Forschung der Fall, wo Sprachhandlungen von Schüler*innen in fachlichen Denkprozessen im Fokus stehen. Auf kognitionspsychologischem Fundament werden Referenzkonzepte historischen Denkens wie *historical literacy* (z. B. Shanahan & Shanahan, 2008; Montesano, 2010) oder *historical reasoning* (z. B. van Drie & van Boxtel, 2008; 2018)

ausgehend von den sprachlichen Aktivitäten der Schüler*innen als domänen-spezifische Problemlösungsstrategien betrachtet. Lese- und Schreibprozesse im Fach stehen hierbei als kognitiv-epistemische Prozesse im Zentrum des Erkenntnisinteresses (z. B. Wineburg, 1991; Young & Leinhardt, 1998). Empirische Forschungen, so auch zunehmend im deutschsprachigen Raum, fokussieren die Förderung domänenspezifischer Lese- und Schreibfähigkeiten (z. B. Voss & Wiley, 1997; Hartung 2013; Handro & Kilimann, 2019; Mierwald, 2020; Waldis et al., 2020). In diesem Forschungszugang wurde für die Studie «Historisches Denken und sprachliches Handeln» ein besonderes heuristisches, aber auch kategoriengebendes Potenzial gesehen: Erstens werden lernenden- und unterrichtsorientiert konkrete Sprachhandlungen historischen Denkens fokussiert, zweitens wird das Verhältnis von Sprache, Kognition und Epistemologie (vgl. auch Kühberger, 2016) aufgegriffen und so drittens die Schnittstelle sprachlicher und fachlicher Leistungen beschrieben bzw. beschreibbar. In der internationalen Geschichtsdidaktik liegen sowohl theoretische als auch empirische Ansätze vor, Sprachhandlungen in dieser Hinsicht mit historischem Denken zu verknüpfen (z. B. Handro, 2013; 2018; Hartung, 2013; Bernhardt & Conrad, 2018). Im viel rezipierten «Framework for Analyzing Students' Reasoning About the Past» vernetzen van Drie und van Boxtel (2008, 90) die drei Sprachhandlungen *describe*, *compare*, *explain* mit unterschiedlichen Operationen und Kategorien historischen Denkens. So stellen sich die Sprachhandlungen dann als fachspezifisch dar, wenn sie in einem argumentativ-begründenden Modus realisiert werden, der gegenstandsbezogen auf die historische Frage rekurriert, Quellen und Darstellungen berücksichtigt, historisch kontextualisiert sowie historisch-gegenstandsbezogene Begriffe adäquat verwendet. Während diese Komponenten konkret gegenstandsbezogen sind, bedarf es zudem der Anwendung von Kategorien historischen Denkens, die sich z. B. im Herausstellen von Triftigkeit, der Differenzierung von Zeitdimensionen oder der Markierung von Perspektivität sprachlich äussern und eng mit epistemologischen Überzeugungen verbunden sind. Die fachlich-epistemische Komponente der Sprachhandlungen wird darüber hinaus im «Prozessmodell epistemischen Sprachhandelns im Geschichtsunterricht» von Handro (2013, 325; 2018, 30) durch die Historische Methode näher konturiert und in sprachliche Teilhandlungen untergliedert. Die Modellierungen basieren auf einem unterschiedlichen theoretischen Fundament, teilen aber die Gemeinsamkeit, historisches Denken bzw. Erzählen in verschiedene Teiloperationen zu differenzieren, die auf sprachlicher Repräsentationsseite als kognitiv-epistemische Sprachhandlungen historischen Denkens

zum Ausdruck kommen. So wird historisches Denken über den Narrationsprozess in fachspezifische Sprachhandlungen differenzierbar und damit auch für Unterricht und empirische Forschung operationalisierbar. In den vergangenen Jahren wurden bereits verschiedene Sprachhandlungen für historisches Denken theoretisch und empirisch weiter spezifiziert und kategorial differenziert (z. B. Günther, 2020; Mierwald, 2020; Schrader, 2021; 2022; Husemann, 2022).

Vor dem Hintergrund dieser für die Studie zentralen Forschungszusammenhänge werden im Folgenden der theoretische Rahmen sowie die forschungsmethodischen Entscheidungen konkretisiert, die für die Analyse der Texte der Schüler*innen und die daran entwickelte Typologie leitend waren.

3 Theoretischer Rahmen: Sprachhandlungen historischen Denkens

Ausgehend von den internationalen geschichtsdidaktischen sowie interdisziplinären Forschungszugängen wird der Zusammenhang historischen Denkens und sprachlichen Handelns in der Studie über das Konstrukt der Sprachhandlung theoretisch modelliert. Dieser Fokus basiert auf einer geschichtsdidaktischen und linguistischen Theoriebildung, die Sprachhandlungen, wie sie im Unterricht als Operatoren (beschreiben, erklären, vergleichen, beurteilen etc.) gebraucht werden, als kognitiv-sprachliche Schnittstelle versteht (z. B. Vollmer, 2011; Thürmann, 2012; Dalton-Puffer, 2013; Handro, 2013; Bernhardt & Wickner, 2015; Bernhardt & Conrad, 2018; Oleschko & Grannemann, 2018). Dieser Annahme nach können sprachliche Strukturen folglich als Repräsentationsformen spezifischer Denkleistungen verstanden werden. Ausgehend von dieser theoretischen Grundannahme können Sprachhandlungen dann ein Kernelement fachlich-sprachlicher Beurteilung sein, wenn sie über fachliche Kriterien als spezifische sprachliche Repräsentationsformen historischen Denkens präzisiert werden. Es stellt sich für die Forschung (und den Unterricht) die Frage, mit welchen konkreten sprachlichen und fachlichen Kategorien Texte von Schüler*innen an dieser Schnittstelle untersucht werden können.

Mit der linguistischen Perspektive wird die sprachliche Dimension, also die Repräsentationsebene kognitiver Leistungen, theoretisch fundiert und empirisch untersuchbar (z. B. Mainzer-Murrenhoff et al., 2021). In seinem Textprozedurenmodell hat Feilke (2015, 63) Texthandlungstypen auf der Ausdrucksseite in Handlungsmuster und sprachliche Ausdrucksmittel differenziert und damit sprachbasierte Zugänge zu kognitiven Sprachhandlungsprozessen theoretisch

ermöglicht (auch Bachmann & Feilke, 2014). So wird Ausdrucksmitteln wie konjugierten Verben, Adjektiven, dem Tempus- und Modusgebrauch und Konnektoren eine konkrete Handlungsfunktion zugeschrieben. Der Linguistik ging und geht es dabei aber weniger um spezifische Zusammenhänge historischen Denkens und sprachlichen Handelns, sondern um die Benennung sprachlicher Herausforderungen der Fächer, auch im Vergleich unterschiedlicher Sprachregister im schulischen Kontext (z. B. Schleppegrell, 2004; Mainzer-Murrenhoff, Drumm & Heine, 2021). Von daher sensibilisieren die Studien dafür, dass sich fachliches Denken auf allen sprachlichen Ebenen, also der Text-, Satz- und Wortebene, repräsentiert. Ausgehend von dieser Grundannahme bieten sie (fachunspezifische) sprachliche Kategorien der Analyse und Diagnose an. Folgt man der funktionalen pragmatischen Linguistik, spielt ferner der Kontext, in dem eine Sprachhandlung realisiert wird, eine entscheidende Rolle, indem er die spezifische Ausführung einer Sprachhandlung mitbestimmt. Sprachhandlungsmuster und sprachliche Ausdrucksmittel auf Text-, Satz- und Wortebene stellen damit keine grammatikalisch-isolierten Einheiten dar, sondern sie sind in dieser Hinsicht immer Zeugnisse optionaler, individueller Sinnbildungsprozesse im spezifischen Gebrauchskontext (z. B. Dalton-Puffer, 2013). Im Unterricht wird dieser Kontext besonders durch die Aufgabenstellung und zugehörige Materialauswahl bestimmt. Indem Schüler*innen im Unterricht sprachlich handeln, z. B. bei der Produktion schriftlicher Texte, erbringen sie also aufgaben- und materialgebunden spezifische Denkleistungen und versuchen diese sprachlich zu repräsentieren. In dieser Hinsicht ist die fachdidaktische Perspektive entscheidend, die durch bestimmte historische Fragestellungen und Zielsetzungen, Bezugsmaterial und Aufgaben erstens den fachbezogenen Kontext definiert und zweitens fachliche Kategorien als spezifische historische Denkstrukturen bestimmt. Diese können, wie oben dargestellt, angelehnt an van Drie und van Boxtel (2008) etwa auf Triftigkeit, Perspektivität und Zeit bezogen werden oder mit Handro (2018) über die Historische Methode an Teilschritte des historischen Erkenntnisprozesses gekoppelt sein.

4 Forschungsmethodische Entscheidungen

Ausgehend von den theoretischen Rahmungen war es für die Studie entscheidend, einen fachlichen Material- und Aufgabenkontext zu schaffen sowie passende Sprachhandlungen zu benennen, um die Proband*innen zum konkreten Handeln aufzufordern. Das zentrale Untersuchungsinstrument der Studie bil-

det folgende materialbasierte Schreibaufgabe: «2013 haben der australische Historiker Christopher Clark und der deutsche Historiker Gerd Krumeich² unterschiedliche Positionen zur Kriegsschuldfrage vertreten. Lies bitte beide Textauszüge und bearbeite dann schriftlich folgende Aufgabe: Vergleiche die Positionen der beiden Historiker miteinander und beurteile ihre Argumente in Bezug auf die Kriegsschuldfrage.» Die Aufgabe war Teil eines Arbeitsblattes, das neben der Benennung der historischen Frage (100 Jahre danach: Wer war schuld am Ersten Weltkrieg?) auch einen kurzen Textauszug zur Kontextualisierung der Kontroverse (vgl. Bergmann, 2016, 56–59) bot und dem die beiden Historikertexte auf je einer Seite folgten. Zu dieser Schreibaufgabe wurden im Herbst 2015 in insgesamt fünf Geschichtskursen der 10. Klasse an drei Gymnasien im Umkreis des Ruhrgebiets 115 Texte Lernender erhoben. Um kontextualisierende Daten zu erhalten, wurde die Erhebung der Schreibprodukte durch einen kurzen Selbsteinschätzungsfragebogen (Auskünfte u. a. zum Interesse an Geschichte, thematisches Vorwissen, Vorerfahrungen zum Schreiben in Geschichte) vor dem Schreibprozess sowie einen individuellen, schriftlichen Aufgabenkommentar nach der Bearbeitung begleitet.

Die Auswahl des Untersuchungssamples wurde von drei Entscheidungen besonders geleitet: Erstens sollten die Proband*innen sowohl in Bezug auf das historische Denken als auch hinsichtlich der eigenständigen Textproduktion über Vorwissen und Vorerfahrungen verfügen. Zweitens sollte sich dem explorativen Zugriff der Studie gemäss im Sample ein möglichst heterogenes Spektrum dieser Vorwissensbestände und Vorerfahrungen widerspiegeln. Und drittens sollte die Erhebung in einer Jahrgangsstufe stattfinden, in der Schriftlichkeit im Geschichtsunterricht zunehmend relevant wird. Dies ist in NRW mit Beginn der Oberstufe durch die Einführung von Klausuren als Prüfungsformat der Fall.

Um die Theorie der Sprachhandlungen diagnostisch (für den Geschichtsunterricht) anschlussfähig zu machen, wurden zwei Sprachhandlungen exemplarisch ins Zentrum der Untersuchung gestellt: Vergleichen und Beurteilen. Ihre fachspezifisch kontextuelle Rahmung erfolgte über das geschichtsdidaktische Prinzip der Kontroversität (vgl. Bergmann, 2016, 40–44), sodass Vergleichen und Beurteilen epistemisch an die Erkenntnis des Konstruktcharakters historischer Deutung und Urteilsbildung gekoppelt sind. Vergleichen als stärker

2 Es handelt sich um Auszüge aus Clark (2013, 715–718) sowie um eine gekürzte Version des Interviews mit Gerd Krumeich in der gekürzten Version in Kellerhoff (2013).

analytische Referenz auf Fremdtexte und Beurteilen als stärker subjektgebundene Handlung legen dabei zwei kognitiv wie sprachlich unterschiedliche Handlungsschwerpunkte. In Anlehnung an die Komponenten des Modells von van Drie und van Boxtel (2008) wurden die beiden Sprachhandlungen ferner mit konkreten historischen Denkoperationen und -kategorien wie historischem Fragebezug, Perspektivgebundenheit, Triftigkeit, Zeitdifferenz und den Dimensionen der Sach- und Werturteilsbildung fachspezifisch konkretisiert (vgl. Schrader, 2021; 2022). Kontroversität als multiperspektivische Ausdrucksform leitete nicht nur exemplarisch das fachspezifische, epistemische wie kognitive Verständnis der Sprachhandlungen an, sondern gab schliesslich auch die Auswahl und Strukturierung des Erhebungsimpulses vor, indem etwa die historische Frage deutlich werden sollte, aber auch Perspektiven in den Historikertexten, Quellen- und Darstellungsbezüge sowie explizite Urteilsbildungen erkennbar sein sollten.

Die 115 erhobenen Texte umfassten im Schnitt 273 Wörter, bei einem Spektrum von maximal 568 und minimal 33 Wörtern (vgl. Schrader, 2021, 86–89, 100–102). Die Textanalyse erfolgte qualitativ hermeneutisch nach Verfahren der Qualitativen Inhaltsanalyse nach Mayring (2015) mit dem Ziel der Strukturierung des Materials. In Anlehnung an die geschichtsdidaktische und linguistische Theoriebildung wurden die sprachlichen Strukturen in den Texten der Lernenden dabei auf drei Ebenen analysiert, die schrittweise eine deduktiv-induktive Kategorienbildung im Zusammenhang geschichtsdidaktischer und linguistischer Kategorien schärften: 1. Beschreibung/Systematisierung unterschiedlicher sprachlicher Merkmale auf Wort-, Satz- und Textebene beim Vergleichen und Beurteilen der Historikertexte; 2. Bezug auf die Gewichtung der sprachlichen Merkmale hinsichtlich Repräsentationsformen von Kategorien historischen Denkens; 3. sprachbezogene Rekonstruktion epistemologischer Überzeugungen.

Induktiv wurden aus den Texten erstens insbesondere linguistische Merkmale herausgearbeitet, in denen sich Unterschiede und Gemeinsamkeiten beim Vergleichen und Beurteilen repräsentierten und so über die sprachliche Oberfläche beschreiben liessen. In diesem Zusammenhang wurden Textmuster herausgestellt und systematisiert, die z. B. den Aufbau der Texte der Lernenden, die Referenz auf die Historikertexte und die Repräsentation eigener Urteilsbildungen beschreiben und in denen sich jeweils sprachliche Mittel (wie Verbtempus, Modus, Adjektive) syntaktisch und semantisch unterschiedlich abbilden. Insbesondere deduktiv war zweitens das Verfahren, die herausgestellten sprachli-

chen Repräsentationen im Hinblick auf die Strukturen und Kategorien historischen Denkens wie den historischen Fragebezug, Perspektivität, Triftigkeit, Zeitdifferenz (vgl. van Drie & van Boxtel, 2008) und historische Sach- und Werturteilsbildung zu systematisieren und in verschiedenen Ausprägungen zu beschreiben. Durch die ersten beiden Analysedimensionen wurden drittens epistemologische Überzeugungen (zum Geschichtsverständnis, zum Verständnis der Historikertexte als historische Darstellungen sowie zum eigenen Urteilsverständnis) rekonstruiert und wiederum systematisch mit der von den Schüler*innen gebrauchten Sprache in Beziehung gesetzt (vgl. Schrader, 2021, 103–106). In der gleichrangigen Berücksichtigung geschichtsdidaktischer und linguistischer Kategorien liegt eine Besonderheit in der Anwendung der Qualitativen Inhaltsanalyse: Die Sprache wurde nicht im Zuge der fachlich-inhaltlichen Kategorienbildung zunehmend reduziert und verdichtet, sondern explizit in ihren Strukturen auf Text-, Satz- und Wortebene in Bezug zu den Kategorien historischen Denkens gesetzt.³ Die linguistischen Kategorien liegen insgesamt quer zu den Kategorien historischen Denkens (vgl. Abb. 1), weil sie jeweils – in spezifischer Semantik – für die Ausprägung historischen Denkens relevant werden. Kategorienbildung und Kodierung am Material wurden dabei wiederholt mit unterschiedlichen Gruppen von Expert*innen überprüft (Mayring, 2015; 123–129).

Kategorien historischen Denkens	Linguistische Kategorien
<i>im Textteil vergleichen</i>	Konnektoren – Prädikate (v. a. reedeinleitende Verben) – Tempusgebrauch – Modusgebrauch – Gebrauch von Adjektiven und Adverbien – Begriffsgebrauch
Wahrnehmung einer offenen historischen Frage	
Kennzeichnung von Perspektivität	
Herausstellung von Triftigkeit	
Markierung von Zeitdifferenz	Merkmale sprachlicher Register – Gebrauch von Phrasen und Idiomen
<i>im Textteil Beurteilen</i>	
Sachorientierte Urteilsbildung und Argumentation	
Wertorientierte Urteilsbildung und Argumentation	

Tab. 1: Kategoriensystem

3 Mayring (2015, 99–102) nennt hier den Begriff der «formalen Strukturierung».

Die drei analytischen Teilschritte liessen im Vergleich der Texte der Schüler*innen insgesamt vier syntaktisch und semantisch unterschiedliche sprachliche Muster erkennen, die in den entsprechenden Texten jeweils auf bestimmte Ausprägungen der untersuchten Kategorien historischen Denkens hinwiesen und sich unterschiedlich in Bezug zu epistemologischen Überzeugungen setzen liessen. Mit einer didaktischen Systematisierungsabsicht wurden diese Muster als Synthese der Teilbefunde in Form von Typen verdichtet. Die Typen sind in ihrer Zuspitzung somit zwar idealtypisch, haben aber je reale (Teil-)Entsprechungen in den Texten der Lernenden bzw. bilden sich in einzelnen Texten sogar nahezu vollständig ab. Sie werden im Folgenden ausführlich beschrieben.

5 Historisches Denken und sprachliches Handeln: eine Typologie

Die Namen der Typen nehmen auf die Art des entstandenen Textprodukts Bezug (von der Textwiedergabe über die Textanalyse hin zur Meinungsäußerung und Darstellungskritik) und charakterisieren das dabei eingenommene Verhältnis zur Kontroverse (von referierend über argumentationsorientiert bis zu faktenorientiert und historisch-diskursorientiert). Die folgende Einzeldarstellung der Typen geht zunächst immer von einer kognitiv-sprachlichen Charakterisierung aus und beschreibt dann die sprachliche Repräsentationsfläche. Nach einem Ankerbeispiel, das Charakteristika des Typs spiegelt, folgt die Interpretation im Hinblick auf Aspekte historischen Denkens in den Typen (vgl. Schrader, 2021, 246–251). Sprachliche Besonderheiten werden in den Ankerbeispielen zudem hervorgehoben.

5.1 Typ I – referierende Textwiedergabe

Kognitiv-sprachlich ist dieser Typ durch die wortnahe Historikertextwiedergabe bei einer lediglich basalen Vergleichsleistung und nur kurzen Urteilsbildungen mit diffusen Wertungen oder resümierenden Aussagen charakterisiert. Auf der sprachlichen Repräsentationsebene geht dieser Typ mit wortnahen, z. T. wörtlichen Wiedergaben einzelner Historikeraussagen in neutral-berichtender Referenz einher. Es herrschen reedeinleitende Verben des Sagens und Berichtens vor und zudem wird fast vollständig auf einordnende oder beurteilende Adjektive verzichtet. Historikeraussagen und vergangenheitsbezogene Darlegungen werden durch Präsens- und Präteritumgebrauch formal getrennt, dar-

über hinaus lassen die Texte keine zeitmarkierenden Sprachmittel erkennen. Die Urteilsaussagen haben die sprachliche Gestalt formelhafter Abschlussätze. So beginnt ein Schüler seinen Text folgendermassen:

*«**Im ersten Auszug berichtet** Christopher Clark, dass Deutschland und seine Bündnispartner nur wegen dem Versailler Vertrag im Mittelpunkt für die Kriegsschuld seien. Er **berichtet außerdem**, dass nicht ein einzelner Staat die gesamte Verantwortung für den ersten Weltkrieg haben kann, sondern dass alle Protagonisten von 1914 beteiligt sind. [...] **Im zweiten Auszug beschreibt** Gerd Krumeich in einem Interview einer Zeitung die Gegenposition: Gerd Krumeich ist der Meinung, dass Deutschland und seine Verbündeten die Hauptschuld am Ausbruch des ersten Weltkriegs haben. [...].» (Schrader, 2021, 315)*

Historisches Denken repräsentiert sich in diesen reproduzierenden, wenig analytischen Sprachhandlungen kaum. Die historische Frage, Begriffe und Metaphern sind lediglich Teil der wiederholten Historikertextpassagen. Perspektivität wird marginalisiert, indem die Historikertexte v. a. als Berichterstattungen historischer Tatsachen gekennzeichnet werden. In diesem Berichtcharakter ist auch die Triftigkeit offenbar hinreichend gesichert. Zeitdimensionen werden basal differenziert. Allenfalls als Aspekt sprachlicher Differenzierung werden die Tätigkeiten der Historiker im Präsens, der historische Sachverhalt im Präteritum referiert oder aber es verbleibt bei der Nennung der Jahreszahl 1914. Zusammengenommen stellen sich die Historikertexte damit primär als fachunspezifische Sachtexte dar, deren Urteilsbildung, ebenso wie die eigene, lediglich als wiederholende Tatsachenbestätigung erscheint.

5.2 Typ II – argumentationsorientierte Textanalyse

Kognitiv-sprachlich ist dieser Typ durch ein elaboriertes, aber fachunspezifisches Argumentationskonzept charakterisiert. Der deutliche Textschwerpunkt liegt dabei auf dem Vergleichen und seinen fremdtextbezogenen analytischen Anforderungen. Die Urteilsbildungen beziehen sich übergeordnet und auf inhaltlicher Metaebene auf die Schlüssigkeit der Gesamtargumentation. Auf der sprachlichen Repräsentationsebene fällt in diesem Typ die Häufung von Begriffen wie These, Argument oder Gegenposition auf. Ebenso werden die Aussagen der Historiker direkt gegenübergestellt, was durch konzessive oder adversative Konnektoren und reedeinleitende Verben wie erwidern und argumentieren rea-

lisiert wird. Andererseits werden Passivstrukturen gebraucht, um die Inhalte der Historikeraussagen ohne deren direkte Nennung wiedergeben zu können. Präsens und Präteritum kennzeichnen den Bezug zur Deutung oder zum historischen Sachverhalt, allerdings bedingt der Fokus auf die argumentative Metaebene auch einen häufigen Präsensgebrauch. Die Urteilsbildungen werden eingeleitet mit Adverbien wie abschliessend oder insgesamt und die Argumentationen durch Adjektive wie schlüssig, sachlich auf der inhaltlichen Metaebene beurteilt bzw. im Komparativ (viel schlüssiger als) verglichen. So schreibt ein Schüler:

*«Christopher Clark ist der Meinung, dass alle die im Krieg beteiligt waren, auch Schuld am Krieg tragen. Als **erstes Argument entkräftet** er die Fischer-Kontroverse, indem er sie **erklärt** und anschliessend die **Schuldfrage infrage stellt**. [...] Gerd Krumeich vertritt die Meinung, dass allein Deutschland die Schuld am Krieg trägt. [...] Insgesamt lässt sich sagen, dass **Clarks Argumentation viel schlüssiger** ist als Krumeichs **Argumentation**, da Clark **sachlich** bleibt im Gegensatz zu Krumeich. Außerdem erwähnt Clark auch **die Gegenposition**. Krumeich **jedoch nicht**.» (Schrader, 2021, 315–316)*

In diesem Typ zeigt sich Argumentationskompetenz, aber Ansätze historischer Denkstrukturen prägen sich höchstens als zufälliges Nebenprodukt aus. Der Fokus liegt insgesamt auf den Inhalten, weniger auf Elementen von Perspektivität oder Deutung. Werden z. B. Metaphern erläutert, geschieht dies im Hinblick auf ihren allgemeinen, inhaltlichen Aussagegehalt und nicht in ihrem historischen Deutungsgehalt. Die historische Frage wird explizit, aber ausschliesslich als Element der Aufgabenstellung wahrgenommen, weshalb dieser Typ auch mit einer Wiederholung der Aufgabenstellung in den Text einleitet. Perspektivität wird als externe Wahlmöglichkeit gekennzeichnet, wobei sich die Historiker feststehenden Sichtweisen anschliessen und diese argumentativ vertreten. Nicht immer treten die Historiker dabei als handelnde und deutende Individuen auf, z. T. wird auch nur der historische Sachverhalt benannt. Damit einher gehen Geltungsansprüche, die im historischen Wahrheitsbezug oder in bereits bekannten Quellenbezügen liegen. Mit dem Fokus auf die Wiedergabe der inhaltlichen Aussagen in den Historikertexten verbleiben die Lernenden weitgehend im Präsens, sodass die Zeitdifferenz der Vergangenheitsdeutung sprachlich nivelliert wird.

5.3 Typ III – faktenorientierte Meinungsäußerung

Kognitiv-sprachlich ist dieser Typ durch die faktenorientierte und durchweg positivistische Äußerung der eigenen Meinung zur Kriegsschuldfrage charakterisiert, wobei die Historikertexte und ihr Vergleich nur als eklektische und bereits auch urteilsdurchzogene Ausgangsbasis gelten. Auf der sprachlichen Repräsentationsebene fällt in diesem Typ besonders die starke Präsenz des urteilenden Ichs auf. Ergänzt werden Subjektpronomen der 1. Pers. Sing. durch einen inflationären Gebrauch des Verbs meinen sowie Verben der Imagination und Vorstellungskraft, die die Urteile begründen. Die Texte zeichnen sich dabei durch konzeptionelle Mündlichkeit aus. Begründungen werden oft alltags-sprachlich formuliert oder inhaltliche Füllwörter (irgendwie, viele meinen) halten die Argumentation aufrecht. Positionierungen werden als richtig und falsch beurteilt und Argumentationen als parteiisch gekennzeichnet. Verben zur Kennzeichnung von Motiven und Absichten (beschuldigen, sich verhalten, versuchen) kennzeichnen den Typ. Der Wechsel von Präteritum und Präsens erfolgt unsystematisch und sprunghaft, sodass Ereignis- und Deutungsebene ineinander übergehen. So schreibt eine Schülerin:

*«Die beiden Historiker Christopher Clark und Gerd Krumeich beziehen Stellung zu der Kriegsschuldfrage des ersten Weltkriegs und vertreten beide eine unterschiedliche Position zu diesem heiklen und defensiven Thema. **Quantität und Qualität dieser Aussagen sind von den Historikern abhängig.** [...] Beide **sind auf jeden Fall der Meinung, dass Deutschland mitgemacht hat** (am Krieg beteiligt war) also auch Schuld hat. Krumeich **allerdings meint**, die Deutschen waren hellwach und wussten **was sie da taten.** [...] **Während Clark meint**, dass die Deutschen als Schlafwandler – wachsam, aber blind mit Alpträumen **durch die Gegend laufen und Krieg führten. Meiner Meinung nach ist das Deutsche Reich aber nicht der Auslöser für den ersten Weltkrieg, da der Krieg durch ein Attentat ausgelöst wurde.** [...] **Auch wenn ich persönlich finde, dass es nicht gleich einen riesigen ersten Weltkrieg als Folge hätte geben sollen.** [...] **Es sollte keinen Krieg geben.**» (Schrader, 2021, 316)*

Bezogen auf das historische Denken bildet sich hier grundlegend ein positivistisches Geschichtsverständnis ab. Dies hat verschiedene Konsequenzen für den Umgang mit den Historikertexten und die Ausprägung der Kategorien historischen Denkens. Mit dem Fokus auf das historische Ereignis ist auch die histori-

sche Frage in den Texten der Lernenden durchgängig unterschwellig präsent. Sie wird bei der Wiedergabe der Historikeraussagen, aber auch in den eigenen Beurteilungen wiederholt direkt oder indirekt aufgegriffen, wobei ihre Beantwortung jedoch durchgängig als offenkundig erscheint. Ebenso wird Perspektivität vor allem als Persönlichkeitsmerkmal konzipiert, das sich nicht zuletzt in der Omnipräsenz subjektiven Meinens manifestiert. Das korrespondiert mit Geltungsansprüchen, die einerseits durchweg über die Meinung des Subjekts gesichert werden, sich andererseits aber der richtigen historischen Wahrheit stellen müssen. Urteilsbildung bedeutet für diesen Typ somit zuvorderst eine klare Positionierung, die in ihrer oft monokausalen Entschiedenheit auf (einzelnen) historischen Faktenkenntnissen, Alltagskonzepten oder der eigenen Imaginationskraft beruht. Dabei verschwimmen z. T. die Zeitdimensionen über einen unsystematischen Tempusgebrauch in einer diffusen Masse aus Gleichzeitigkeit. Mit dem Vetorecht der im historischen Sachverhalt liegenden Wahrheit und der Gleichstellung meinungshabender Individuen schwindet in diesem Typ die Bedeutung der Historiker*innen als Expert*innen historischer Deutung. Zugespitzt formuliert: Das, was in den Historikertexten an historischem Tatsachenbezug erkenntlich wird, kann grundsätzlich auch ohne Umwege über die Meinungen der Historiker*innen – direkt mit Perspektive auf den historischen Sachverhalt – wiedergegeben werden.

5.4 Typ IV – historisch-diskursorientierte Darstellungskritik

Kognitiv-sprachlich ist dieser Typ durch eine abwägende Darstellungskritik der Historikertexte in Beantwortung der Kriegsschuldfrage charakterisiert, wozu das Vergleichen dienende Funktion einnimmt. Der Vergleich trägt den Historikertexten dabei explizit als von Historiker*innen verfasste, fachspezifische Textgattung Rechnung, die kontrovers deutend auf eine historische Frage antworten. Die Urteilsbildung basiert damit auf der Gegenüberstellung der Historikerargumente und fokussiert ihre Einschätzung in Beantwortung der Kriegsschuldfrage. Auf der sprachlichen Repräsentationsebene fällt zunächst die explizite, einleitende und dann wiederholende Benennung der Kriegsschuldfrage auf, die mit einem expliziten Verweis auf «Historiker» und Verben wie streiten oder kontrovers diskutieren verbunden ist. Bereits zu Beginn der Texte der Lernenden wird der Vergleich durch eine explizite Differenzierung der Zeitdimensionen begleitet: Wie in keinem anderen Typ werden dazu Zeitadverbien (wie damals, heutzutage) gebraucht, Jahreszahlen und Zeiträume genannt und

Präsens und Präteritum systematisch differenziert, sodass Zeit hier als argumentative Darstellungsstruktur markiert wird. Wie in Typ II werden Argumente durch konzessive und adversative Konnektoren gegenübergestellt, allerdings hier stärker und sprachlich explizit als andere Perspektiven oder unterschiedliche Sichtweisen in ihrem variierenden Deutungscharakter gekennzeichnet. Dadurch werden Geltungsansprüche nicht nur auf Quellen bezogen, sondern in ihrer argumentativen Verhandelbarkeit herausgestellt. So stehen nicht Inhalte, sondern «Historiker» im Zentrum, die begründen, argumentieren, belegen, analysieren, vergleichen, zitieren, schlussfolgern usw. So schreibt eine Schülerin:

«Die Kriegsschuld 1914 – wer war es? Viele Historiker diskutieren heutzutage immer noch darüber, wer Schuld trug am ersten Weltkrieg. In den vorliegenden Texten [...] von Christopher Clark (2013) und [...] Gerd Krumeich (2013) lässt sich entnehmen, wie unterschiedlich die Meinungen der beiden zu diesem Thema sind. Clark schreibt in seinem Text, dass nicht wie in den sechziger Jahren die «Fischer-Kontroverse» (Z. 8) annimmt, dass die Deutschen damals schuld waren, sondern [...]. Während Krumeich in dem Interview bekannt gibt, dass er genau davon überzeugt ist, dass die Deutschen die Hauptschuld haben. Er argumentiert damit, dass Deutschland damals versuchte die russische Kriegsbereitschaft zu «testen» [...]. Auch Krumeich geht auf den Vergleich mit den Schlafwandlern ein, aber aus einer anderen Perspektive. Er bezieht sich darauf, dass alle «hellwach» waren [...]» (Schrader, 2021, 317)

Einzig in diesem Typ spiegelt sich ein tendenziell konstruktivistisches Geschichtsverständnis. Die historische Frage stellt in diesem Typ ein Kernelement dar, das nicht nur den offenen und strittigen Ausgangspunkt des Historikertextvergleichs markiert, sondern Letzteren auch in einen syn- und diachronen Diskurskontext bettet. Geltungsansprüche entstehen offenbar durch Quellen- und Darstellungsbezüge und können auch verhandelt werden. Die Zeitdimensionen werden systematisch differenziert und sind dabei erkenntnis- und argumentationsrelevant. In den Beurteilungen stehen die Darstellungen erneut im Fokus, indem z. B. beurteilt wird, inwiefern sich die beiden Historiker empirisch triftig auf «verlässliche Quellen» beziehen, andere Positionen berücksichtigen und damit insgesamt zur Beantwortbarkeit der Kriegsschuldfrage beitragen.

6 Die Typologie als Basis fachspezifischer Diagnostik?

Wie hervorgehoben sind diese Typen eine Verdichtung von unterschiedlichen fachlich-sprachlichen Mustern, die sich in den Texten der Lernenden zeigten. Von Einzelfällen abgesehen, lassen sich die meisten Texte des Samples den Typen damit nicht ausschliesslich, aber tendenziell zuordnen. Dabei ist auffällig, dass Typ I und Typ III als Merkmalsausprägungen im Sample häufiger vertreten sind als Typ II und Typ IV.

Abbildung 2 fasst die Merkmale dieser vier Typen im Überblick zusammen. Sie können als Diagnosekriterien zur Einschätzung von Texten Lernender genutzt werden.

<p>Typ I: Referierende Textwiedergabe Additives Wiedergeben von Inhalten Keine / kaum eigene Strukturierungsmerkmale erkennbar Hohe sprachliche Nähe zu Ausgangstexten Kaum Fachbegriffe</p>	<p>Typ II: Argumentationsorientierte Textanalyse Eigenständige Strukturierung nach allgemeinen Textaufbaukriterien (z. B. Einleitung, Hauptteil, Schluss) Verknüpfungen folgen Ausgangstexten Eigene, analytisch treffende Formulierungen Kaum Fachbegriffe</p>
<p>Typ III: Faktenorientierte Meinungsäußerung Textaufbau und Struktur folgen eigener Meinungsverkündung zum historischen Sachverhalt Ausgangstexte werden selektiv als «richtige» oder «falsche» Aussagen wiedergegeben (historische Wahrheit) Stark mündlich und subjektorientierte Wortwahl</p>	<p>Typ IV: Historisch-diskursorientierte Darstellungskritik Textaufbau und Verknüpfungen sind eigenständig strukturiert, z. B. nach historischer Frage, Perspektivitätskennzeichnungen, empirischen Triftigkeitsmassstäben, zeitlichen und kausalen Zusammenhängen Eigene, analytisch treffende Formulierungen Viele Fachbegriffe</p>

Tab. 2: Typologie fachsprachlichen Handelns von Schüler*innen

Die Typologie gibt diagnostische Anhaltspunkte für die Einordnung von Texten Lernender als fachspezifische Realisierungen von Sprachhandlungen. Die Ermittlung und Beschreibung der Typen hat dabei drei zentrale Zusammenhänge erkennen lassen, die für die sprachbasierte Diagnose historischen Denkens in Empirie und Pragmatik gleich relevant sein dürften.

1. Sprachliche Elaboration und Diagnose historischen Denkens: Folgt man den vier Typen, so hat der sprachliche Elaborationsgrad nur einen begrenzten Aussagewert im Hinblick auf die Repräsentation historischen Denkens.⁴ Zwar steigt mit dem Maß konzeptioneller Schriftlichkeit in den Texten der Lernenden auch der Grad der Differenziertheit v. a. im vergleichenden Umgang mit den Historikeraussagen und auch die Repräsentation elaborierter argumentativer Handlungen nimmt zu, die für historische Denk- und Erzählprozesse besonders wichtig sind. Aber: Es repräsentiert sich hierin vor allem eine allgemeine Textkompetenz (z. B. Becker-Mrotzek & Böttcher, 2020, 126–131; Pohl, 2017) und nicht per se historisches Denken. In den sprachlich elaborierten Texten kommen nicht automatisch auch Operationen und Kategorien historischen Denkens als textstrukturierende Merkmale zum Ausdruck. Dies wird in Gegenüberstellung von Typ II und IV deutlich. Und zudem sind diese Operationen und Kategorien historischen Denkens zu ihrer Repräsentation wiederum nicht zwingend auf sprachliche Elaboration angewiesen. Das zeigen die Typen III und IV. Diese Befunde sind für die Diagnose historischen Denkens insofern relevant, als die Feststellung einer allgemeinen Sprachkompetenz m. E. kein hinreichender Indikator für damit verbundene fachliche Denkleistungen sein kann – dies wäre insbesondere für die Entwicklung quantitativer Diagnoseinstrumente historischer Kompetenzen von Interesse (z. B. Trautwein et al., 2017).
2. Die Rolle einer an Teiloperationen historischen Denkens semantisch orientierten Diagnose: Die für historisches Denken entscheidenden Unterschiede liegen in den Texten des Samples auf der semantischen Ebene der Sprachhandlungsmuster und sprachlichen Ausdrucksmittel zum Aufbau und zur Strukturierung der Texte. Kategorien historischen Denkens, die die Sinnbildungen der Lernenden in Vergleich und Urteil strukturieren, lassen sich über die sprachliche Repräsentation in den Texten der Lernenden ermitteln und z. T. sogar stufen. Aber: Von Bedeutung ist dabei nicht die Repräsentation einer bestimmten Wortart oder grammatikalischen Struktur wie z. B. der Gebrauch indirekter Rede zur Distanzierung von Fremdaussagen (z. B. Steinhoff, 2008), sondern der Bedeutungsgehalt der redeeinleitenden Verben, die dazu im Einzelfall genutzt werden. So gebraucht jeder Typ die indirekte Rede, aber während die Typen I und II dazu Verben wie «sagen», «be-

4 Vgl. ähnlich Bertram et al. in diesem Band.

schreiben», «berichten» semantisch nutzen, stehen in Typ III und IV stärker diskursive, argumentative und fragebezogene Verben im Zentrum. Diagnostisch würde dies v. a. für eine qualitative Interpretation der Äusserungen der Schüler*innen im Zusammenhang historischer Denkoperationen sprechen. Damit ergänzen die Befunde vornehmlich theoriegeleitete sprachbasierte Diagnoseinstrumente, etwa zur Erfassung narrativer Kompetenz (z. B. Pandel, 2015) durch die wichtige Perspektive auf den tatsächlichen Gebrauch sprachlicher Mittel aufseiten der Lernenden.

3. Die diagnostische Bedeutung epistemologischer Überzeugungen: Auf die Realisierung der beiden Sprachhandlungen im Kontext der Studie hatten offenbar die epistemologischen Überzeugungen der Schüler*innen grossen Einfluss (z. B. Mierwald, 2020; Kühberger, 2016). Dies betrifft nicht nur die Frage nach einem positivistischen oder konstruktivistischen Geschichtsverständnis, sondern vor allem das Verständnis des Erkenntniswerts, den eine bestimmte Sprachhandlung in einem fachspezifischen Diskurs- und Aufgabenkontext mit sich bringt (z. B. Steinseifer, 2014). Auch hier unterscheiden sich die Typen im Hinblick darauf, ob ein Urteil oder der textbezogene Vergleich den Schwerpunkt bilden. Während sich Typ I und II, bei divergierender analytischer Komplexität, auf das fremdtextbezogene Vergleichen konzentrieren, richten Typ III und IV bereits das Vergleichen funktional auf den Erkenntnisgewinn zur historischen Fragestellung aus. Auch dies geschieht in unterschiedlicher analytischer Komplexität, ist aber im Kern von einem Fokus auf die historische Urteilsbildung motiviert.

Mit dieser dreidimensionalen Betrachtung grenzen sich die Typen von Graduierungen der schreibdidaktischen Forschung ab bzw. führen diese fachbezogen fort (z. B. Steinhoff, 2008; Becker-Mrotzek & Böttcher, 2020; Pohl, 2017). Damit bringen sie in den geschichtsdidaktischen Diskurs einen qualitativ-empirisch basierten Analyseansatz ein, der die drei Dimensionen erfasst, fachlich zusammenführt und mögliche Graduierungen bietet. Wie bereits reflektiert, basieren die Typen auf einem spezifischen qualitativen forschungsmethodischen Zugang und sind damit zunächst exemplarisch. Inwiefern die Typologie dabei auf verschiedene Sprachhandlungen zu übertragen ist, sich auf Texte von Schüler*innen anderer Schulformen und -stufen beziehen oder in anderen fachlich-diagnostischen Zusammenhängen nutzen lässt, gilt es in weiteren Untersuchungen oder praktischer Anwendung zu prüfen.

Literatur

- Bachmann, Thomas, & Feilke, Helmuth (Hrsg.) (2014). *Werkzeuge des Schreibens: Beiträge zu einer Didaktik der Textprozeduren*. Stuttgart: Fillibach bei Klett.
- Barricelli, Michele (2005). *Schüler erzählen Geschichte: Narrative Kompetenz im Geschichtsunterricht*. Schwalbach/Ts.: Wochenschau.
- Becker-Mrotzek, Michael & Böttcher, Ingrid (2020). *Schreibkompetenz entwickeln und beurteilen (9. Auflage)*. Berlin: Cornelsen.
- Bergmann, Klaus (2016). *Multiperspektivität: Geschichte selber denken (3. Auflage)*. Schwalbach/Ts.: Wochenschau.
- Bernhardt, Markus (2020). Sprache(n) des Geschichtsunterrichts – Sprachliche Vielfalt und Historisches Lernen. Einführung in die Tagung. In Thomas Sandkühler & Markus Bernhardt (Hrsg.), *Sprache(n) des Geschichtsunterrichts: Sprachliche Vielfalt und Historisches Lernen* (S. 29–44). Göttingen: V & R unipress.
- Bernhardt, Markus & Conrad, Franziska (2018). Sprachsensibler Geschichtsunterricht: Sprachliche Bildung als Aufgabe des Fachs Geschichte. *Geschichte lernen*, 31(182), 2–9.
- Bernhardt, Markus & Wickner, Mareike-Cathrine (2015). Die narrative Kompetenz vom Kopf auf die Füße stellen – Sprachliche Bildung als Konzept der universitären Geschichtslehrerbildung. In Claudia Benholz, Magnus Frank & Erkan Gürsoy (Hrsg.), *Deutsch als Zweitsprache in allen Fächern: Konzepte für Lehrerbildung und Unterricht* (S. 281–296). Stuttgart: Fillibach bei Klett.
- Borries, Bodo von / Fischer, Claudia / Leutner-Ramme, Sibylla & Meyer-Hamme, Johannes (Hrsg.) (2005). *Schulbuchverständnis, Richtlinienbenutzung und Reflexionsprozesse im Geschichtsunterricht: Eine qualitativ-quantitative Schüler- und Lehrerbefragung im deutschsprachigen Bildungswesen 2002*. Neuried: ars una.
- Clark, Christopher (2013). *Die Schlafwandler. Wie Europa in den Ersten Weltkrieg zog*. München: DVA.
- Dalton-Puffer, Christiane (2013). A construct of cognitive discourse functions for conceptualising content-language integration in CLIL and multilingual education. *European Journal of Applied Linguistics*, 1(2), 216–253. <https://doi.org/10.1515/eujal-2013-0011>
- 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>
- Drie, Jannet van & Boxtel, Carla van (2018). Historical reasoning: conceptualizations and educational applications. In Scott A. Metzger & Lauren M. Harris (Hrsg.), *The Wiley international handbook of history teaching and learning* (S. 149–176). New York: Wiley-Blackwell.

- Feilke, Helmuth (2015). Text und Lernen – Perspektivenwechsel in der Schreibforschung. In Sabine Schmölzer-Eibinger & Eike Thürmann (Hrsg.), *Schreiben als Medium des Lernens: Kompetenzentwicklung durch Schreiben im Fachunterricht* (S. 47–72). Münster: Waxmann.
- Günther, Katrin (2020). *Historisches Vergleichen: Vergleichsaufgaben in Lehrwerken des Gesellschaftslehre- und Geschichtsunterrichts der Sekundarstufe II: Eine Untersuchung von (Schüler*innen-)Lösungen zu einer Vergleichsaufgabe am Beispiel eines Darstellungstextes zur Industriellen Revolution*. Berlin: Logos Berlin.
- Günther-Arndt, Hilke (2010). Hinwendung zur Sprache in der Geschichtsdidaktik – Alte Fragen und neue Antworten. In Saskia Handro & Bernd Schönemann (Hrsg.), *Geschichte und Sprache* (S. 17–46). Berlin: Lit.
- 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.
- Handro, Saskia (2018). Sprachbildung im Geschichtsunterricht: Leerformel oder Lernchance? In Katharina Grannemann, Sven Oleschko, & Christian Kuchler (Hrsg.), *Sprachbildung im Geschichtsunterricht: Zur Bedeutung der kognitiven Funktion von Sprache* (S. 13–42). Münster: Waxmann.
- Handro, Saskia & Kilimann, Vanessa (2019). Textverstehen im Geschichtsunterricht: Ein Projekt zur Professionalisierung historischer Leseförderung (ProLeGu). In Marion Bönnighausen (Hrsg.), *Universitäre Praxisprojekte in Kooperationsschulen: Lehr-Lern-Konzepte und Module zum Einsatz fachspezifischer und diversitätssensibler Lesestrategien* (S. 165–222). Münster: WTM.
- Hartung, Olaf (2013). *Geschichte Schreiben Lernen: Empirische Erkundungen zum konzeptionellen Schreibhandeln im Geschichtsunterricht*. Berlin: Lit.
- Hartung, Olaf & Memminger, Josef (2017). Schreibend Geschichte lernen: Mehr als ein «mündliches Fach» – die Vielfalt des Schreibens im Geschichtsunterricht entdecken. *Geschichte lernen*, 30(176), 2–11.
- Hodel, Jan / Thünemann, Holger / Waldis, Monika & Zülsdorf-Kersting, Meik (2013). Schülernarrationen als Ausdruck historischer Kompetenz. *Zeitschrift für Didaktik der Gesellschaftswissenschaften* 4 (2), 121–145.
- Husemann, Charlotte (2022). *Geschichte beschreiben, Geschichte erklären: Eine Untersuchung fachsprachlicher Konzepte und fachlicher Sprachhandlungsfähigkeit von Gesamtschüler*innen der Sekundarstufe I*. Göttingen: V & R unipress.
- Köster, Manuel (2013) *Historisches Textverstehen: Rezeption und Identifikation in der multiethnischen Gesellschaft* (Dissertation). Berlin: Lit.

- Kühberger, Christoph (2016). Geschichtsdidaktische Diagnostik. Zum Geschichtsverständnis von Schüler/innen. In Saskia Handro & Bernd Schönemann (Hrsg.), *Aus der Geschichte lernen? Weiße Flecken der Kompetenzdebatte* (S. 139–151). Berlin: Lit.
- Kellerhoff, Sven Felix (2013, 11. September). Das Kaiserreich unterschätzte 1914 Englands Macht. *Welt*. <https://www.welt.de/geschichte/article119906475/Das-Kaiserreich-unterschaetzte-1914-Englands-Macht.html>
- Langer-Plän, Martina (2003). Problem Quellenarbeit: Werkstattbericht aus einem empirischen Projekt. *Geschichte in Wissenschaft und Unterricht*. (5/6), 319–336.
- Mainzer-Murrenhoff, Mirka/Drumm, Sandra & Heine, Lena (Hrsg.) (2021). *Sprachtheorien in der Zweit- und Fremdsprachenforschung: Eine Basis für empirisches Arbeiten zwischen Fach- und Sprachenlernen*. Baltmannsweiler: Schneider.
- 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, 166–187.
- Mayring, Philipp (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12. Aufl.). Weinheim: Beltz.
- Mierwald, Marcel (2020). *Historisches Argumentieren und epistemologische Überzeugungen: Eine Interventionsstudie zur Wirkung von Lernmaterialien im Schülerlabor*. Wiesbaden: Springer VS.
- Monte-Sano, Chauncey (2010). Disciplinary literacy in history: An exploration of the historical nature of adolescents' writing. *Journal of the Learning Sciences*, 19(4), 539–568. <https://doi.org/10.1080/10508406.2010.481014>
- Nitsche, Martin & van Norden, Jörg (2020). Historische Kompetenzen sprachbasiert erfassen: Einführung in die Sektion. In Thomas Sandkühler & Markus Bernhardt (Hrsg.), *Sprache(n) des Geschichtsunterrichts: Sprachliche Vielfalt und Historisches Lernen* (S. 273–276). Göttingen: V&R unipress.
- Oleschko, Sven & Grannemann, Katharina (2018). Fokus Sprachhandlung: Das fachliche Lernen in Geschichte durch die Berücksichtigung fachsprachlicher Besonderheiten unterstützen. *Praxis Geschichte*. (2), 4–9.
- Pandel, Hans-Jürgen (2015). *Historisches Erzählen. Narrativität im Geschichtsunterricht*. Schwalbach/Ts.: Wochenschau.
- Pohl, Thorsten (2017). Sekundarstufe I und II. In Michael Becker-Mrotzek, Joachim Grabowski & Torsten Steinhoff (Hrsg.), *Forschungshandbuch empirische Schreibdidaktik* (S. 89–108). Münster: Waxmann.
- Redder, Angelika/Naumann, Johannes & Tracy, Rosemarie (Hrsg.) (2015). *Forschungsinitiative Sprachdiagnostik und Sprachförderung: Ergebnisse*. Münster: Waxmann.

- Sandkühler, Thomas & Bernhardt, Markus (Hrsg.) (2020). *Sprache(n) des Geschichtsunterrichts. Sprachliche Vielfalt und Historisches Lernen*. Göttingen: V & R unipress.
- Schleppegrell, Mary (2004). *The language of schooling: A functional linguistics perspective*. New York: Routledge.
- 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.). Berlin: Lit.
- Schrader, Viola (2021). *Historisches Denken und sprachliches Handeln* (Dissertation). Berlin: Lit.
- Schrader, Viola (2022). Triftigkeitskonzepte in Schüler*innentexten. Eine Studie zum Zusammenhang von Sprache und historischem Denken. In Saskia Handro & Bernd Schönemann (Hrsg.). *Sprachsensibler Geschichtsunterricht. Geschichtsdidaktische Forschungsperspektiven und -befunde* (S. 202–221). Berlin: Lit.
- Shanahan, Cynthia & Shanahan, Timothy (2008). Teaching disciplinary literacy to adolescents: Rethinking content area literacy. *Harvard Educational Review*, 78(1), 40–59. <https://doi.org/10.17763/haer.78.1.v62444321p602101>
- Steinhoff, Torsten (2008). Kontroversen erkennen, darstellen, kommentieren. Abgerufen von http://www.festschrift-gerd-fritz.de/files/steinhoff_2008_kontroversen_erkennen_darstellen_und_kommentieren.pdf.
- Steinseifer, Martin (2014). Vom Referieren zum Argumentieren: Die didaktische Modellierung von Textprozeduren der Redewiedergabe und Reformulierung. In Thomas Bachmann & Helmuth Feilke (Hrsg.), *Werkzeuge des Schreibens: Beiträge zu einer Didaktik der Textprozeduren* (S. 199–221). Stuttgart: Fillibach bei Klett.
- Thürmann, Eike (2012). *Lernen durch Schreiben? Thesen zur Unterstützung sprachlicher Risikogruppen im Sachfachunterricht*. *dieS-online*, 1/2012. <https://dx.doi.org/10.22029/jlu-pub-2900>
- Trautwein, Ulrich / Bertram, Christiane / Borries, Bodo von / Brauch, Nicola / Hirsch, Matthias / Klausmeier, Kathrin / ... Zuckowski, Andreas (2017). *Kompetenzen historischen Denkens erfassen. Konzeption, Operationalisierung und Befunde des Projekts «Historical Thinking – Competencies in History» (HiTCH)*. Münster: Waxmann.
- Vollmer, Helmut J. (2011). Schulsprachliche Kompetenzen: Zentrale Diskursfunktionen. Abgerufen von <http://www.home.uni-osnabrueck.de/hvollmer/VollmerDF-Kurzdefinitionen.pdf>.
- 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 / 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.
- 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*, 38(1), 73–87.
- Young, Kathleen M. & Leinhardt, Gaea. (1998). Writing from primary documents: A way of knowing in history. *Written Communication*, 15(1), 25–68.

Abstracts

1 Historische Kompetenzen im Umgang mit Medienberichten über den 6. Januar 2021. Eine US-amerikanisch-deutsche Studie mit Lehramtsstudierenden und Lehrkräften

The U.S.-German qualitative empirical study examines the competencies of (prospective) history and social studies teachers in dealing with various and controversial media reports about the storming of the Capitol in Washington, D.C., on January 6, 2021. The study focuses on their methodological competencies, media literacy, and intercultural communication competencies when reading and analyzing newspaper reports. The study reveals remarkable differences in how German and U.S. (prospective) teachers discuss the media reports.

2 Das Repertory-GridInterview als qualitativ empirische Forschungsmethode in der Geschichtsdidaktik

This article discusses repertory grid interviews from a diversity-sensitive perspective and opens up this qualitative empirical research methodology in a discipline-specific way. In the discussion of the resources and barriers of the methodology, the economic and direct elicitation of the personal constructions of meaning of the participants can be identified as a particular strength of the method. Another advantage is the versatile adaptability of the interviews to the research interest and the participants. One barrier is that the methodology is highly structured in terms of form and content, which reduces the amount of knowledge that can be gained and requires certain cognitive abilities on the side of the participants. The method is suitable for participatory research, which generates practically relevant findings for historical learning and teaching processes through the participation of the research participants as co-researchers.

3 Den Umgang mit Quellen und Darstellungen digital beforschen. Eine Pilotstudie

This study used the method of thinking aloud to investigate how people with different levels of expertise deal with sources and accounts when they were asked to develop and answer a historical question in a digital environment. Thinking aloud protocols were analyzed by means of qualitative content analysis in order to determine which operations of narrative competence emerge. In addition, differences in operations between a novice and an expert were identified. Initial results indicate that the operations of narrative competence can be investigated with satisfactory to very good reliability. Further sub-operations could also be differentiated to some extent. With regard to qualitative differences, criteria such as complexity or functionality appear to be suitable for describing different characteristics.

4 Künstliche Intelligenz im Fach Geschichte. Kann ein Computer Texte von Schüler*innen sprachlich und inhaltlich beurteilen?

Open-ended tasks in which students evaluate sources and representations in independently formulated texts are necessary for acquiring competencies in historical thinking. However, objective, valid, and reliable evaluation and assessment of student texts is challenging in research and in the classroom. Machine learning methods, in which algorithms make predictions for unseen data based on a sufficiently large training sample, could be helpful in evaluating student texts. This paper reports how the computational linguistics methods of «automatic content assessment» and «linguistic complexity assessment» were used in an interdisciplinary research project to assess the content accuracy and linguistic complexity of student texts. The results are encouraging, but also show the limitations of AI in the subject of history.

5 «In ihrer Zeitung [...] versuchten sie an alle heranzukommen und sie an ihre tolle Kultur und an sie selbst zu erinnern». (Digitalisierte) historische Zeitungen als Quellen im Geschichtsunterricht

The text focuses on the question of the value of historical newspapers for historical learning. This question has gained importance through numerous digitization projects in recent years, which have made historical newspapers an easily available source. The presented intervention study with four groups of learners in grade ten fits into the context of comparative media studies in history didactics, which aim to explore the value of digital media for history teaching. For this purpose, learners' writing products are examined by means of qualitative content analysis, based on facsimile prints of entire issues of historical newspapers, on digitized facsimiles of the same newspapers, on historical newspapers researched by the learners themselves, and on sources prepared for teaching. The results of an inferential statistical investigation of the quantified results of the content analysis for a subarea of the evaluation are presented.

6 Historisches Erklären lernen? Exploration von Konzepten von Schüler*innen und ihrer Veränderbarkeit

Empirical studies show that students have problems writing adequate historical texts. One can assume that a central reason for this is the students' concepts of the respective text form. Historical explanations, for example, share some very fundamental characteristics with ordinary explanations with which students are familiar from their daily experiences. However, to explain successfully in history, specific epistemic criteria have to be met; ordinary explanations are not considered sufficient. Following the paradigm of conceptual change, these mostly ordinary concepts have to be specified and complemented by specific characteristics of historical explanation. First empirical results hint at group discussions, analogies and cognitive conflict as successful methods to reach this goal. The following paper presents an empirical exploration of two students' concepts of historical explanation and discusses their changes over the course of a quasi-experimental study.

7 Zur sprachbasierten Diagnose historischen Denkens in Texten von Schüler*innen. Eine empirisch gestützte Typologie

Language actions of learners are central to the diagnosis of historical thinking and learning performance. The need for the development of a subject-specific diagnostic tool that is based on the language skills of the learners and that captures historical thinking in a language-based manner was therefore repeatedly emphasized in history didactics. In the article, central results of a qualitative-empirical history-didactic study are put up for discussion from a diagnostic point of view. With a focus on teaching and students, the study is based on the theoretical framework of language acts as an interface between historical thinking and linguistic means of expression. In the study, the students had to compare and evaluate controversial historical arguments. Four types of student texts are presented as the results of the study, and these are reflected as a starting point for the development of a subject-specific diagnostic tool.

Autor*innen

Dr. Christiane Bertram ist Akademische Rätin am Hector-Institut für Empirische Bildungsforschung, Universität Tübingen. Nach mehreren Jahren als Gymnasiallehrerin für die Fächer Deutsch und Geschichte hat sie nach der Promotion 2015 am Hector-Institut die Juniorprofessur Fachdidaktik in den Sozialwissenschaften an der Universität Konstanz bekleidet. Ihre Forschungsschwerpunkte sind u. a. die Kompetenzerfassung mit offenen und standardisierten Instrumenten, Zeitzeug*innen und Oral History im Geschichtsunterricht und die «digitale Souveränität» im Fach Geschichte. Sie arbeitet mit randomisierten Interventionsstudien und nutzt in ihrer Forschung quantitative und qualitative Methoden.

Paul G. Fitchett, Ed. D., ist Professor und Leiter der Abteilung für Curriculum und Unterricht am College of Education der Auburn University (Auburn, AL, USA). Zu seinen Forschungsinteressen gehören die Überschneidungen von Bildungspolitik, Arbeitsbedingungen von Lehrer*innen und Lernergebnissen von Schüler*innen, wobei der Schwerpunkt auf Sozialkunde liegt.

Tobias Flink studierte die Fächer Geschichte, Englisch und Bildungswissenschaften für das Lehramt an Gymnasien und Gesamtschulen an der Universität zu Köln und der University of Aberdeen (2012–2018; 2015: Bachelor of Arts; 2018: Master of Education). Von 2019 bis 2023 wissenschaftlicher Mitarbeiter und Doktorand am Historischen Institut, Abteilung für Didaktik der Geschichte der Universität zu Köln und von 2022 bis 2023 wissenschaftlicher Mitarbeiter am Institut für Didaktik der Geschichte der Westfälischen Wilhelms-Universität Münster.

Carolin Hestler ist als Akademische Rätin an der Pädagogischen Hochschule Ludwigsburg in der Abteilung Geschichte tätig. Zuvor arbeitete sie mehrere Jahre im Schuldienst. Ihre Forschungsschwerpunkte sind u. a. Medien im Geschichtsunterricht, Geschichtskultur und empirische Forschung mit Zeitzeug*innen.

Kristopher Muckel ist seit September 2023 wissenschaftlicher Mitarbeiter am Lehrstuhl für Didaktik der Geschichte an der Universität Augsburg. Zuvor war er seit 2019 am Lehr- und Forschungsgebiet Didaktik der Gesellschaftswissenschaften der RWTH Aachen University tätig und organisierte dort neben seiner Dissertation zum Einsatz historischer Zeitungen im Geschichtsunterricht das gesellschaftswissenschaftliche Lehr-Lern-Labor goAIX! – historische Orte erforschen.

Dr. **Martin Nitsche** ist wissenschaftlicher Mitarbeiter am Zentrum Politische Bildung und Geschichtsdidaktik der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz am Zentrum für Demokratie in Aarau und leitet das SNF-Projekt «Historische Lernprozesse erforschen – Research of Learning Processes in History (RicH)».

Dr. **Franziska Rein** ist seit April 2024 Juniorprofessorin für Geschichtsdidaktik an der Universität Hildesheim. Ihr Studium absolvierte sie in Ludwigsburg und unterrichtete nach ihrem Referendariat mehrere Jahre in Baden-Württemberg. 2021 wurde sie mit einer empirischen Forschungsarbeit zur Bedeutung der eigenen Lebensgeschichte im Kontext inklusiven historischen Lernens promoviert. Franziska Reins aktuelle Forschungsschwerpunkte sind u.a. inklusive Geschichtsdidaktik, Intersektionalität in historischen Lern- und Lehrprozessen und empirische geschichtsdidaktische Professions- sowie Unterrichtsforschung.

Dr. **Viola Schrader** ist seit 2014 Wissenschaftliche Mitarbeiterin am Institut für Didaktik der Geschichte der Universität Münster. Nach dem Studium an der Universität Münster und dem Lehramtsreferendariat in Essen promovierte sie in Münster zum Zusammenhang historischen Denkens und sprachlichen Handelns bei Schüler*innen. Ihre Forschungs- und Lehrschwerpunkte liegen im Kontext sprachsensiblen Geschichtsunterrichts, Heterogenität in historischen Lernprozessen und der Theorie-Praxisverzahnung in der Lehramtsausbildung.

Jan Scheller ist wissenschaftlicher Mitarbeiter am Arbeitsbereich Fachdidaktik des Historischen Instituts der Universität Greifswald. Er arbeitet darüber hinaus als assoziierter Mitarbeiter im SNF-Projekt «Historische Lernprozesse erforschen – Research of Learning Processes in History (RicH)» am Zentrum für Politische Bildung und Geschichtsdidaktik der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz am Zentrum für Demokratie in Aarau.

Jonas Schobinger ist wissenschaftlicher Mitarbeiter am Zentrum für Politische Bildung und Geschichtsdidaktik der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz am Zentrum für Demokratie in Aarau und arbeitet am SNF-Projekt «Historische Lernprozesse erforschen – Research of Learning Processes in History (RicH)».

Prof. Dr. **Monika Waldis** ist Leiterin des Zentrums Politische Bildung und Geschichtsdidaktik der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz am Zentrum für Demokratie in Aarau.

Zarah Weiß ist assoziierte Mitarbeiterin am Lehrstuhl für Theoretische Computerlinguistik an der Universität Tübingen.

Dr. **Lisa Zachrich** ist wissenschaftliche Mitarbeiterin am Hector-Institut für Empirische Bildungsforschung in Tübingen. In ihrer Dissertation mit dem Titel «Empirische Bildungsforschung trifft Geschichtsdidaktik. Eine Untersuchung der Lernprozesse im Lernarrangement mit Zeitzeug:innenberichten» hat sich Lisa Zachrich mit verschiedenen Aspekten der besonderen Lernerfahrung von Schüler*innen mit Zeitzeugenberichten beschäftigt. Aktuell entwickelt sie im Rahmen des BMBF-Verbundprojekts EMPATHIA³ einen standardisierten Test zur Erfassung relevanter Wissensbestände für angehende Lehr- und Polizeikräfte für die Antisemitismusprävention.

Dr. **Ramon Ziai** ist assoziierter Mitarbeiter am Lehrstuhl für Theoretische Computerlinguistik an der Universität Tübingen.

Gutachter*innen der Beiträge

Heinrich Ammerer

Sebastian Barsch

Markus Bernhardt

Christiane Bertram

Andrea Brait

Saskia Handro

Christian Heuer

Charlotte Husemann

Manuel Köster

Georg Marschning

Christian Matthis

Marcel Mierwald

Mario Resch

Wolfgang Wagner

Lale Yildirim

Sabine Ziegler