Learning and Instruction 94 (2024) 101993

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Image interpretation scaffolds support adolescents' historical reasoning

Kevin van Loon [a,*], Monika Waldis [b]

[a] Institute for Educational Sciences, University of Basel, Switzerland
[b] FHNW School of Education, University of Applied Sciences and Arts Northwestern Switzerland, Switzerland

ARTICLE INFO

Keywords:
Image analysis
Image interpretation
Historical reasoning
Critical reasoning
Historical writing

ABSTRACT

Background: Image interpretation scaffolds and mind mapping may support students' historical reasoning. Benefits of such approaches are, until present, not empirically tested.
Aims: This quasi-experimental study compared effects of three conditions: a sequenced image interpretation scaffold, a flexible image interpretation scaffold, and mind mapping on students' historical reasoning *about* and *with* images.
Sample: Participants were adolescents (secondary school students, $N = 145$, M age 13.9 years).
Methods: Historical reasoning was assessed with a writing task. Participants interpreted three photographs over three time points. At T1, students had no support. At T2 and T3, students interpreted an image with the randomly assigned condition. Students' interpretative essays were rated to assess historical reasoning.
Results: Image interpretation scaffolds better supported historical reasoning than mind mapping. Both interpretation scaffolds enhanced historical reasoning. Particularly the flexible image interpretation scaffold appeared to benefit contextualization and the description of the relevance of the image message for the present. However, effects of the flexible scaffold were not stable over time.
Conclusions: Image interpretation scaffolds seem beneficial to support students' interpretation of historical images. Further research needs to investigate how image interpretation scaffolds support historical reasoning for different types of images.

1. Introduction

Adolescents are confronted with images on the internet daily and increasingly produce, inform, and communicate through digitally edited images (Külling et al., 2022). However, images, especially photographs, are often presented uncontextualized and might mistakenly be considered accurate reflections of reality. This can mislead uncritical viewers. To prevent distortion processes, images must be placed in context (Burke, 2008). The time and place of origin and information about the author, like intentions and motivations and how these are shaped by time and culture, need to be considered (Wissner, 2022). Research on students' historical reasoning shows that contextualization seems challenging (Sendur et al., 2021). Particularly when reasoning with images as historical sources, adolescents do not spontaneously apply a thorough image analysis, including contextualization strategies (Bernhardt, 2007; Lange, 2011; Van Boxtel & Van Drie, 2012; Wolfrum & Sauer, 2007). Due to a lack of methodological competencies, students may draw false conclusions. This study aims to investigate how students can be supported to develop competencies to critically analyze and interpret images.

One subject where adolescents are taught about the critical interpretation of images is history education. There, an aim is to develop competencies to connect the past to the present and future by asking questions and analyzing and interpreting sources (Schreiber et al., 2006; Van Drie & Van Boxtel, 2008). Importantly, when applying historical reasoning skills, students have to show that they can reason *about* and *with* sources (Rouet et al., 1996). Reasoning *about* sources refers to students' skills to critically analyze the source and assess the value and the limits of information, including recognition of the author's perspective and aims and the context in which the source was produced. Wineburg (1991) identified three heuristics that historians apply while reading historical texts: Sourcing, contextualization, and corroboration, which can be subsumed to reasoning *about* sources. Reasoning *with* sources refers to the skills involved in selecting information and using this to construct explanations about how the past is connected to the present (Lévesque, 2005; Phillips, 2002). These historical reasoning skills can be linked to critical thinking (Luís & Rapanta, 2020). Critical thinking consists of formulating questions and problems, gathering and

* Corresponding author. FHNW School of Education, Hofackerstrasse 30, 4132, Muttenz, Switzerland.
E-mail address: kevin.vanloon@fhnw.ch (K. van Loon).

https://doi.org/10.1016/j.learninstruc.2024.101993
Received 30 May 2023; Received in revised form 7 August 2024; Accepted 8 August 2024
Available online 26 August 2024

assessing relevant information, and establishing and communicating well-reasoned conclusions and solutions (Paul & Elder, 2019). By applying critical historical reasoning skills, students become aware of the construction of historical narratives and the underlying motives. Images are a type of narrative, as they trigger memories or stories and can indicate movements and actions (Fuhs, 2006). Images contain visual messages, which can be decoded into words. Further, images function as means of communication, as they convey emotions through which stories are being told, for example, with war photography or social media images (Fuhs, 2006; Krammer, 2006). As images are paramount and disinformation in the media and on the internet is common, critical image interpretation skills are becoming more important than ever.

Although the importance of images in history lessons is recognized, and images are used extensively (Bernhard, 2017; Van Nieuwenhuyse et al., 2017), history textbooks and teachers mainly use images for illustrative purposes rather than for discussion and contextualization (Bernhard, 2017). Students often only superficially inspect, analyze, and interpret images (Bernhardt, 2007; Wolfrum & Sauer, 2007) and a prolonged and elaborated examination of images seems to fall short (Bernhard, 2017). Until present, students' historical reasoning skills have mainly been investigated using textual sources (Monte-Sano & De La Paz, 2012). Very few studies investigated students' historical reasoning *about* and *with* images; one of the few studies (Van Loon et al., 2024) showed that when interpreting a historical image, analytical sourcing and contextualization skills were hardly visible, limiting image interpretation quality. Although students' methodological skills when learning with images were lacking, it remains unknown how to support adolescents when analyzing and interpreting images.

1.1. Sequenced image interpretation scaffold

Image interpretation scaffolds were developed to enhance students' examination and reasoning about images (Ormond, 2011). The most common image interpretation scaffolds in history education are based on guidelines by the art historian Panofsky (1939/2018), who differentiates between three sequential steps (see Fig. 1): pre-iconographical description, iconographical analysis, and iconological interpretation (Hamann, 2012). In Panofsky's model, image description focuses on the denotation (what is shown), and iconographical analysis refers to the connotation (how something is shown, considering cultural conventions). Iconological interpretation addresses the image's broader cultural meaning, including the ideas, beliefs, and feelings of that period (Burke, 2008; Pfisterer, 2020; Roeck, 2004; Schmidt-Maiwald, 2018). Panofsky's sequenced scaffold has an easily applicable step-by-step structure. This might be why this approach still occupies a prominent position in education (Schmidt-Maiwald, 2018). Because this scaffold is specifically designed to analyze and interpret historical images, therefore, Panofsky's sequenced scaffold can be seen as a domain-specific approach.

Despite the potential advantages of the Panofsky approach, students report that they are bored with using these steps because it hinders them from creating their own narrative (Lange, 2011). Further, students often skip the description part as they consider this easy and unimportant, and jump directly to the interpretation (Lange, 2011). Bernhardt (2013), Lange (2011), and Oleschko (2013) suggest that Panofsky's sequenced scaffold may hinder a flexible approach to image interpretation. Moreover, Panofsky's model was initially developed to interpret single artworks, mainly Renaissance painting and symbolism (Bernhardt, 2018). Therefore, applying Panofsky's approach to other image types, like contemporary photography, might lead to a limited interpretation, not covering aspects such as digital editing, trustworthiness, or relevance for the present (Paul, 2013).

Although a three-step interpretation scaffold is commonly used to guide students, the question to what extent such scaffolds support image analysis and interpretation has not yet been investigated. The present research is the first to empirically investigate to what extent a sequenced image interpretation scaffold supports students' historical reasoning about images.

1.2. Mind mapping for image interpretation

Another way to enhance historical reasoning may be through making visual representations: Mind mapping (graphically representing associative ideas; see Fig. 2) and concept mapping (representing and connecting concepts by using connective terms or cross-links; Davies, 2011). Mind maps and concept maps are visualizations of "how students

Evaluate an image source in three steps

1 Describe the picture:

- Note all the people, objects, and symbols depicted in the picture.
- How are the people dressed, how are they behaving?
- What is in the foreground and what is placed in the background?
- Are there people or objects that are in a special light?
- What questions do you have when you look at the picture?

2 Explain the content:

- What do you know about the event, the people in the picture, and the artist who created the picture?
- When was the picture created?
- If symbols are shown, what do they mean?
- Where is the picture shown (newspaper, church...), and to whom is it addressed?
- What was the occasion for the picture?

Tip: If you don't already have this information, find it out on the Internet or in the library.

3 Interpret the picture:

- Now combine the information from the first and second step.
- Write down the statement that you wanted to make with this picture.
- What is the meaning of the colors in the picture? What feelings should be conveyed?
- Are there objects or creatures that represent something? Why did you choose these "representatives"?

Fig. 1. Example of a sequenced image interpretation scaffold.

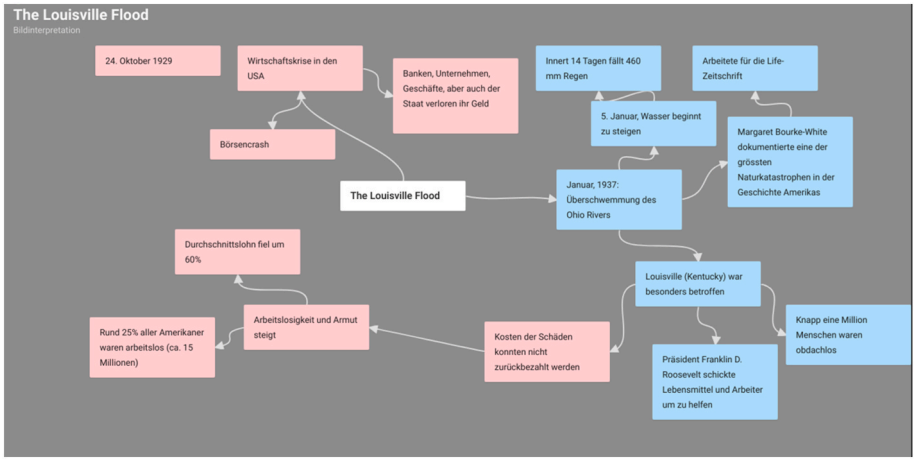


Fig. 2. Example of a students' mind map.

think, understand, organize, and apply knowledge" (Shi et al., 2022, p. 2). By mind/concept mapping, students can be supported to express, explain, and discuss their ideas (Van Boxtel & Van Drie, 2018; Van Drie et al., 2005). When mapping, students are encouraged to represent information about interpreted images in a graphical rather than linear text format (Van Boxtel & Van Drie, 2018). Mind mapping seems beneficial to direct adolescent students' attention and to support memorization of learning materials (Shi et al., 2022). Mind mapping improved students' learning outcomes in different school subjects, promoted creative and critical thinking, and supported writing (Al-Zyoud et al., 2017; Davies, 2011; Shi et al., 2022). Van Boxtel and Van Drie (2018) recommend using mind maps as a prewriting strategy when selecting or organizing information from historical sources before writing an essay.

In history education, Lange (2011) used concept mapping to analyze knowledge structures and learning processes when higher secondary school students interpreted images. Findings showed that meaning-making processes when learning with images seem to happen simultaneously rather than chronologically, contrasting assumptions about the sequenced structure implied by Panofsky (Lange, 2011). Also, mind/concept mapping allows students to take more autonomy than Panofsky's approach because of its unconstrained structure (Davies, 2011). Compared to the sequenced scaffold, no domain-specific terms, questions or hints are leading the mind mapping work, which makes it a more generic approach. This high level of learner autonomy may also be a drawback of mind mapping, as a mind map may only summarize and stay on a general level rather than address critical aspects (Davies, 2011). Although mind mapping seems to promote students' writing, translating a mind map into a text can be challenging (Coirier et al., 1999). Without giving further instructions, students may not engage in higher-order reasoning, particularly when images are complex. Effects of mind mapping have not yet been compared with the effects of structured supports (i.e., the Panofsky approach) when learning with images in history education; the present research aims to do so.

1.3. Flexible image interpretation scaffold

To support image interpretation, it may be necessary to (a) give the opportunity to develop ideas without being forced into a fixed structure (analog to mind mapping) and (b) to guide critical image analysis and interpretation (analog to the steps suggested by sequenced image interpretation scaffolds). To date, no scaffold for image interpretation has been proposed and tested that provides both guidance for image analysis and interpretation while allowing flexibility in structuring the analysis, including which aspects to focus on and in what order. To support self-structured interpretation and critical reasoning, we developed a flexible image interpretation scaffold for the present research (see Fig. 3). This model was based on guidelines for art historians to inspect and analyze images (Bätschmann, 2009). Like the Panofsky-based approach, in this flexible scaffold, students are guided by questions. However, they have the flexibility to choose which questions they address and in what order these are addressed. The developed flexible image interpretation scaffold consists of two circles. The inner *image* circle describes questions as guidelines for visual inspection (i.e., questioning what can be seen in the image). The outer *context* circle describes questions as guidelines to search for additional information (i.e., questioning what cannot be seen directly and requires further research). With the circular structure, the flexible scaffold aims to give students as much autonomy as with mind mapping: They can decide with which questions and at which circle (image or context circle) they start their analysis. Importantly, the flexible scaffold supports critical image interpretation by emphasizing the context of the image. The flexible scaffold aims to engage students in thinking about the connection between an image, a historical event or phenomenon, and its relevance for the present. Further, the flexible scaffold asks for a critical image analysis, by asking about trustworthiness and its closeness to reality, whereas such questions are not asked in the Panofsky-based sequenced scaffold. Due to the specific focus on interpreting historical images, this can be considered a domain-specific approach.

1.4. The present study

We compared the effects of three conditions over time on students' historical reasoning when learning with images: (A) a Panofsky-based sequenced scaffold from a history textbook for adolescent students (Aeby et al., 2017, p. 104; see Fig. 1); (B) mind mapping as used by Buzan and Buzan (1993) (see Fig. 2), and (C) a flexible scaffold, based on the guidelines by Bätschmann (2009) (see Fig. 3). At the first time point, students analyzed and interpreted an image without using any support or specific conditions. For time point two, students were randomly assigned to one of three conditions for image analysis and interpretation. The condition assignment remained the same for time point three.

At each time point, students interpreted an image using an inquiry-based learning task as in Britt and Aglinskas (2002), Rouet et al. (1996), or Wiley and Voss (1999), resulting in an image interpretation essay containing the following elements: A historical question, the main message of the image with a corresponding image description, the results of the image analysis, and a conclusion with a justified interpretation. The quality of the essays was assessed through scoring, in line with the approach by Monte-Sano and De La Paz (2012), Nokes (2017), and Waldis et al. (2020). Specifically, this research addresses the following questions and hypotheses investigating potential benefits of different types of support (conditions) on students' historical reasoning.

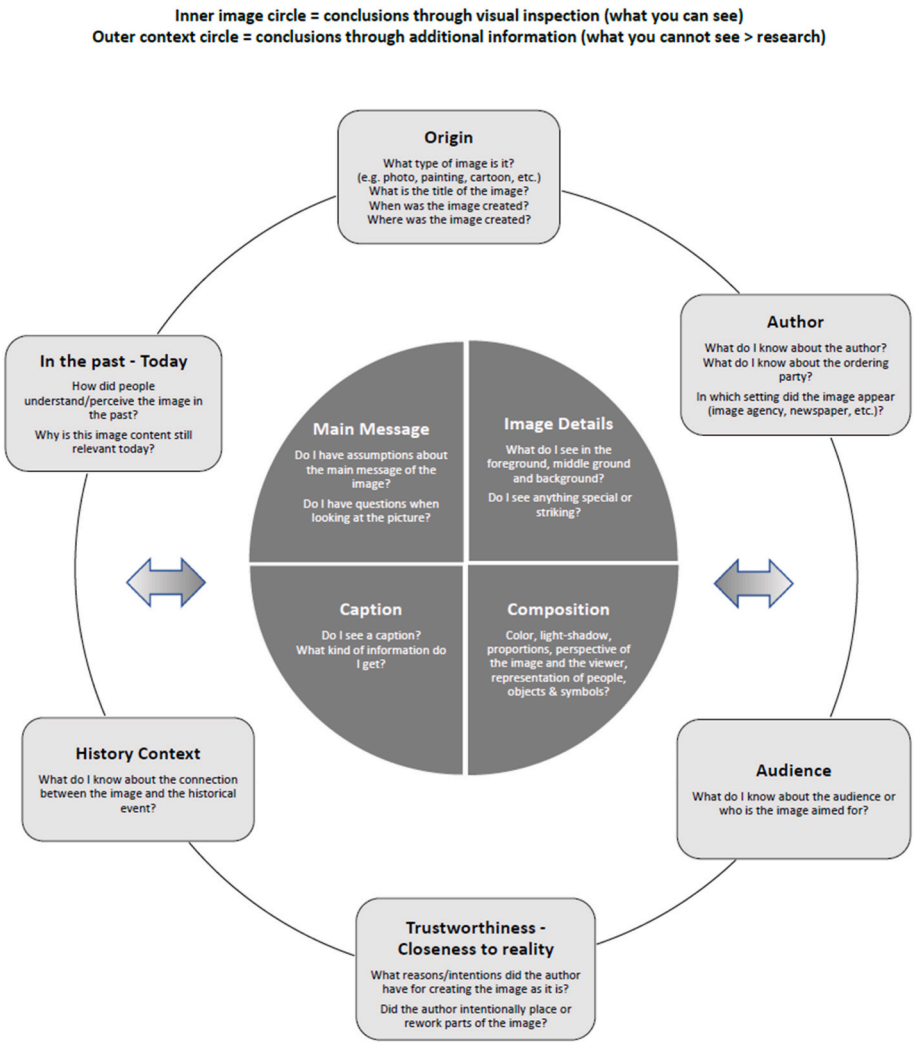


Fig. 3. Flexible image interpretation scaffold.

RQ1 Do students show better historical reasoning (i.e., overall quality of the essay) with the use of support (condition assignment in T2 and T3) than without (no conditions in T1)?

H1. The use of support leads to better historical reasoning at T2 and T3 than at T1, where no support was used.

RQ2 Are there differences in historical reasoning when comparing a generic (mind mapping condition) with a domain-specific approach (sequenced and flexible scaffold condition)?

H2. The domain-specific approach (sequenced and flexible scaffold) supports students' historical reasoning better than generic mind mapping.

RQ3 Are there differences between the sequenced scaffold condition and the flexible scaffold condition in a) students' historical reasoning and b) specific subcomponents of historical reasoning such as "sourcing", "contextualization", and "relevance for the present"?

H3. Because the flexible scaffold guides students to critically analyze the context and the trustworthiness of the image, a) this scaffold leads to the highest scores of historical reasoning, b) particularly for sub-components "sourcing", "contextualization", and "relevance for the present", in comparison to the sequenced scaffold condition.

2. Methods

2.1. Participants and design

In total, 145 secondary school students participated; *M* age = 13.92, *SD* = .65; 50.3 % girls. Participants were 8th graders at the end and 9th graders at the beginning of the school year. Students were from the German-speaking part of Switzerland, and part of nine different school classes from four schools. They attended lower secondary school, which typically spans three to four years, covering grades seven through nine, with an optional tenth grade. These school tracks serve as preparation for further education, commonly in vocational training or potentially leading to enrolment in a gymnasium. On average, students received one to 2 h of history instruction per week. Parents/caretakers gave informed consent, and participants were told that they could drop out without any consequences. Not included in this sample are four students whose parents/caretakers did not give consent, and four students dropped out because they were not present at T1. Further, the sample size got smaller during the study (*n* T1 = 145, *n* T2 = 144, *n* T3 = 140) because students could not participate due to illness or internships.

This quasi-experimental intervention study consisted of three time points to investigate potential changes in historical reasoning *about* and *with* images. Note that another manuscript by Van Loon et al. (2024) explicitly addresses the challenges in historical reasoning at T1, as a baseline measure, when no support was used for image interpretation. This also includes the validation of the scoring rubric to score the essays.

For the present manuscript, the focus was on describing the potential effects of different conditions at T2 and T3 in comparison to each other and to T1; the data from T1 are therefore reported in comparison to T2 and T3. T3 was included into the design to ensure that students worked with the then familiar image interpretation scaffold (at T2, the scaffold was introduced to them for the first time, and thus still new). Moreover, all students worked with the scaffold two times, with two different images, which makes the study more generalizable than when only using one single image and measurement point.

The sequenced scaffold condition consisted of 46 students (25 male and 21 female); the mind map condition consisted of 45 students (21 male and 24 female); and the flexible scaffold condition consisted of 49 students (23 male and 26 female).

2.2. Materials

2.2.1. Handout of the image interpretation task

A handout, shared with students at the start, consisted of an essay structure as writing guideline (Appendix A), the prompt (Appendix B), the image with title and year of origin and a detailed caption, and note paper to write down their image interpretation (paper & pencil). Additional information about the image was printed in the handout below the image, consisting of historical background and an image context. This information was based on history textbooks and did not reveal interpretive elements of the image itself.

2.2.2. Images

For each time point, a different photograph was used; details about the photographs are given in Table 1. The images had to fulfill the following criteria:

- Black and white photographs from the 20th century.
- Humans are in the image's center, representing social grievances or discrimination (e.g., gender, race, class).
- An image is presented in the image (i.e., a poster, an advertisement banner, and a screen with propaganda).
- A relevance for the present can be formulated.
- Photographs should be unknown to students.

The images were checked for suitability by art historians and history educators, and seven teachers were asked to estimate the difficulty of the used images; their ratings are shown in Table 2. The teacher ratings indicated that the images would fit the grade level of the students, and would not be too easy or difficult, and rather comparable in difficulty level (i.e., all mean ratings indicated that the images were perceived as moderately difficult by the teachers). Further, the teachers confirmed that the images were not used in class before. The image rights were purchased for this study from Keystone-SDA and Timeline Images (now SZ Photo).

Table 1
Images used per Time Point.

Time Point	Image Caption	Historical Background	Image Context	Link
1	Unknown photographer, Two women in front of a poster on the occasion of the cantonal vote on women's suffrage in Zurich, 1947, press photo, Keystone SDA.	85 words	129 words	https://tinyurl.com/39eh6n49
2	Margaret Bourke-White, The Louisville Flood, 1937, press photo, 24,6 × 34 cm, New York City, Whitney Museum of American Art.	72 words	78 words	https://tinyurl.com/2cx9edrz
3	Hermann Schröer, *Zaungäste an der Mauer (Berlin Wall)*, 1962, photo, Timeline Images (now SZ Photo).	83 words	129 words	https://tinyurl.com/y898hcsu

2.3. Procedure

Data collection took place over a period of around two weeks. Table 3 shows an overview of the procedure over the three time points. At all time points, students sat alone, or barriers were put up between students. The visiting researcher observed a quiet and concentrated classroom atmosphere, where students read the instructions, inspected the image, searched the internet and took notes about this, and then produced handwritten essays ranging from half a page to two pages in length. No participants chose to withdraw from the study.

2.3.1. Time point 1

At T1, students were instructed about the purpose of image analysis and interpretation and how they could structure an image interpretation essay with an introduction, main part, and conclusion. They were presented with a colored writing guideline showing a high-quality essay structure (as shown in Appendix A). To prepare students for the writing tasks, they were also handed out an image about industrialization and child labor (the photograph by Lewis W. Hine, *John Howell an Indianapolis newsboy*, from 1908) and a corresponding high-quality image interpretation essay (Appendix C). To support a comparison between the writing guideline and the example essay, the same colors were used to refer to the essay structure. The example essay aimed to prepare students for the upcoming image interpretation and writing task. After studying the example essay, it was collected again. Then, students analyzed and interpreted the image individually. Participants used a computer to search the internet for information to analyze the image. They used the handout to write down their image interpretation.

2.3.2. Time point 2 and 3

To ensure that students were prepared to deal with the digital image analysis task at T2 and T3, classes were priorly instructed by their teachers about how to use the online tool Padlet (padlet.com) to create a mind map, to add comments to a PDF-File, and to down- and upload documents on the school server.

After welcoming the students, students were reminded of their participation in T1, when they wrote an image interpretation without any support. They were told that the aim of these lessons was to analyze the image with a support and then write an image interpretation essay. To introduce the image analysis phase, the visiting researcher showed all students a visualization of the three types of support (sequenced scaffold, mind mapping, flexible scaffold), explaining the structure of each. Students were told they would afterwards work with one of these three image interpretation supports. Students were told to make notes on their

Table 2
Difficulty of Images and Additional Image Information rated by Teachers (n = 7).

Images	Difficulty of Images
	M (SD)
T1: Two women in front of a poster on the occasion of the cantonal vote on women's suffrage in Zurich	3.14 (.38)
T2: The Louisville Flood	3.71 (.49)
T3: Zaungäste an der Mauer (Berlin Wall)	3.43 (.53)

Note. Question for teacher was "How do you rate the level of difficulty of the images and the additional information?". They responded on a 5-point Likert scale: 1 = very easy; 2 = rather easy; 3 = just right; 4 = rather difficult; 5 = very difficult.

Table 3
Procedure of each Time Point.

Time Point	Phase (min)
1	-Introduction (5) -Instruction image analysis, image interpretation, and essay structure (20) -Individual study of example essay (5) -Students analyzed image and wrote image interpretation essay (40)
2	-Introduction to the three conditions (sequenced scaffold, mind mapping, flexible scaffold) (5) -Students analyzed image with assigned support condition (40) -Students wrote image interpretation essay based on prior analysis (40)
3	-Introduction (5) -Students analyzed image with assigned support condition (same as in T2) (40) -Students wrote image interpretation essay based on prior analysis (40)

Note. Students' time for image analysis and interpretation at T1 is half as long than at T2 and T3, because they did not have to make notes on image interpretation scaffolds or to create a mind map.

computer when working on their image analysis. The mind map condition had to create a mind map on padlet.com without any further instructions. The sequenced and flexible scaffold condition had to respond to questions posed by the image interpretation scaffolds presented in a PDF file by putting comments in it. Students were told to use keywords when creating the mind map or when writing comments to the questions. They were also reminded to use their image analysis results when writing their image interpretation and to use the writing guideline to structure their essay. The guideline was shown during the entire lesson (as in T1).

Subsequently, students received the handout showing which support (condition) they were assigned to and then started with the task. Students self-paced their internet search, analysis, and writing time. The researcher informed students 20 min before the lesson ended about the time left and recommended them to start finalizing their essays. The procedure for T3 was the same as for T2.

2.4. Scoring of the image interpretation essays

A scoring rubric was developed to assess historical reasoning, combining generic historical reasoning components with specific image interpretation competencies (Kuckartz, 2014). The rubric consists of three components (*asking and answering historical questions, reasoning about images, and reasoning with images*) and ten categories indicating subcomponents of these historical reasoning components (see Appendix D for the categories). Each category could range from level zero (low quality) to three (highest quality) (see Appendix E for an example of a coded students' essay). To investigate the overall quality of historical reasoning, a score was calculated per student by adding up the ratings for the ten subcomponents.

The scoring rubric was based on previous research (as further outlined below) and on scoring data from students in a pilot stage before data collection for the present study. Scoring was discussed in depth by two raters. Before starting the individual scoring process, the two raters double-scored 10% of the essays, and interrater reliability was high (Krippendorffs' $\alpha = .86$). After having scored approximately 50% of the essays, the two raters again double-scored 10% of the essays, and interrater reliability was acceptable ($\alpha = .78$). After both interrater reliability assessments, scoring differences were discussed, and minor additions were made to the scoring rubric where needed.

2.4.1. Component 1: asking and answering historical questions

This component consists of two subcomponents: (1) "historical questioning" and (2) "answering the question". For this component was rated to what extent students could describe "a product or a (potential) start of historical reasoning while trying to put into words a conflict or deficit in prior knowledge about historical constructs, phenomena or developments" (Logtenberg, 2012, p. 91). When answering a historical question, it was rated to what extent students plausibly did this and justified their answer with reasons.

2.4.2. Component 2: reasoning about images

For this component was scored to what extent students were able to identify and use historical information when working with images. Reasoning about images consists of two subcomponents: (1) "sourcing" and (2) "contextualization". "Sourcing" was rated to what extent information from the image caption was included in the written essay, such as the author's name, title, date, technique/type of image (genre), origin, and place of storage (Büttner, 2014). "Contextualization" addresses the context of the image, such as information about the author, publication context, targeted audience, intention/motivation of the author, and tendencies toward media critique (Britt & Aglinskas, 2002; Van Nieuwenhuyse, 2017; Wineburg, 1991).

2.4.3. Component 3: reasoning with images

This component assessed to what extent participants used images and their analysis to describe and explain historical events based on their starting question (Krammer, 2006, p. 26). This component consists of the subcomponents (1) "image description"; (2) "main message of the image"; (3) "image interpretation"; (4) "image reference"; (5) "relevance for the present", and (6) "essay structure". "Image description" was rated to what extent participants described the image or pictorial details in connection to aspects of the image composition (Bernhardt, 2007; Hamann, 2012; Krammer, 2006). To rate the "main message of the image", the quality of the connection between the image, the historical event and the inherent social issues were rated (Baxandall, 1990; Bernhardt, 2011, pp. 37–53). For the "image interpretation" was investigated to what extent students could explain connections between the image elements and the historical event and social issues by justifying their conclusions. To rate the "image reference", the embedding of the image in the written essay was investigated. When rating the component "relevance for the present", it was assessed to what extent participants connected the past to the present by explaining why historical events or social issues might be of today's importance with the use of actual examples (Lévesque, 2005; Partington, 1986, pp. 63–81; Phillips, 2002). With the rating of the "essay structure" was assessed to what extent students could formally structure their essays with an introduction, a main part, and a conclusion. Student essays were scored with MAXQDA version 2020.

2.5. Validation of the scoring rubric

We validated our scoring rubric at T1 with an exploratory factor analysis (for details, see Van Loon et al., 2024). Results showed that the extracted factors largely overlap with the theoretically distinguished competencies of historical reasoning: Factor one indicates the historical reasoning component *asking and answering historical questions* (subcomponents such as "historical questioning", "answering the question", "essay structure", "relevance for the present"), factor two indicates the component *reasoning about images* ("sourcing", "contextualization"), factor three indicates the component *reasoning with images* ("image description", "image reference", "image interpretation"). The subcomponent "main message of the image" loaded on a separate fourth factor.

For T2 and 3, we conducted a confirmatory factor analysis (Figs. 4 and 5; for correlations between the manifest variables see Tables 4 and 5) to assess whether the factor structure of the coded data, as found for T1, was also confirmed for T2 and T3. All analyses were conducted with R, packages lavaan (Rosseel, 2012) and psych (vo.2.3.3; Revelle, 2023). The least squares estimator WLSMV was used because of the ordinal data structure and the small sample size (Kyriazos & Poga-Kyriazou, 2023).

CFAs for both time points confirmed the extracted factors of the EFA and seem in line with the historical reasoning competencies, particularly

asking and answering historical questions, *reasoning about sources*, and *reasoning with sources*. Cut-off values for CFAs, as recommended by Brown (2006), indicate acceptable model fit indices except for the Chi^2 test, which is sensitive to deviations from normality (Newsom, 2023a). T2: Degrees of freedom (24), *a p-value of* Chi^2(.000), CFI (.985), TLI (.978), RMSEA (.079), and SRMR (.069). T3: Degrees of freedom (24), *p*-value of Chi^2 (.001), CFI (.974), TLI (.960), RMSEA (.083), and SRMR (.076).

However, for T2, the subcomponent "sourcing" showed negative variance, which indicated a Heywood case, because the latent variable *reasoning about images* only contained two indicators, "sourcing" and "contextualization" (Newsom, 2023b). As recommended by Newsom (2023b), we set the two loadings to be equal.

Furthermore, the subcomponent "essay structure" correlated strongly with other subcomponents like "historical questioning", "answering the question", and "main message of the image" (Spearman's rank correlations of T2 and T3 ranging between .46 and .67), therefore, this overarching category was excluded from the CFAs.

2.6. Statistical analyses

To test our hypotheses, we conducted multilevel analyses with SPSS 28 to account for the nested data structure (time points, participants, classes, and schools). A random intercept was used for each participant to account for variance in individual observations. Time point was entered as within-person repeated measurement. Repeated covariance type was set as diagonal, and a maximum likelihood estimation was used. The data and analysis code are available on the Open Science Framework (https://osf.io/fq5rm/?view_only=b682bb2a511c42de9a0f2efaa541413c).

3. Results

We firstly test Hypothesis 1, that support (condition assignment in T2 and T3) fosters historical reasoning compared to not having any support (as in T1). Further, we test Hypothesis 2 that a domain-specific approach (sequenced and flexible scaffold condition) has better effects on historical reasoning than a generic approach (mind mapping condition). Finally, we test Hypothesis 3 that a flexible scaffold condition leads to better historical reasoning performance than a sequenced scaffold condition (H.3a), particularly for "sourcing", "contextualization", and "relevance for the present" (H.3b). Considering the directional nature of the hypotheses, statistical significance was assessed using one-sided p-values as indicated.

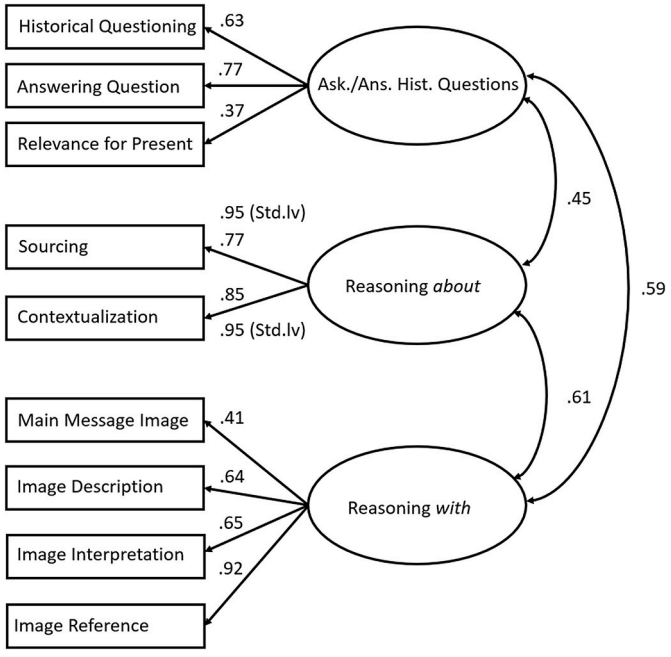


Fig. 4. Confirmatory Factor Analysis: Time Point 2.
Note. Two loadings from latent variable Reasoning *about* was set to be equal, because of a Heywood case (Std.lv = Standardized latent variable).

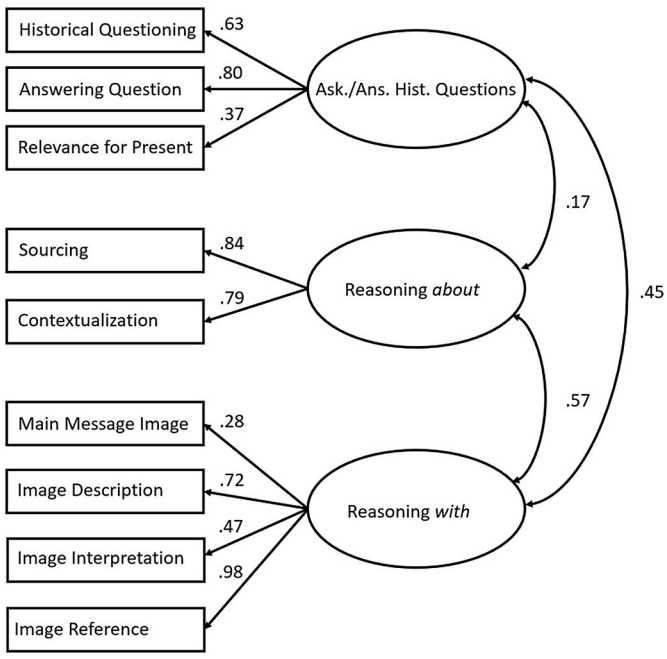


Fig. 5. Confirmatory Factor Analysis: Time Point 3.
Note. At T3 was no Heywood case detected.

To investigate whether students' historical reasoning performance was better at the later time points (T2 and T3, when support was used) as at T1, where no support was used, we conducted a multilevel analysis. The analysis indicated that performance significantly differed between time points, $F(2, 156.42) = 6.51$, $p = .002$. Bonferroni-corrected pairwise comparisons showed that students' historical reasoning at T2 was higher when using support than without at T1, $p = .002$. However, for T3, there was no difference with T1. These results partially confirm Hypothesis 1: Only at T2, but not at T3, overall historical reasoning performance was better than at T1. Follow-up analyses with Bonferroni-corrected pairwise comparisons of the three components of historical reasoning showed that for *asking and answering historical questions*, performance was even lower at T2 ($p = .001$) and T3 ($p = .010$) than at T1. However, for the component *reasoning about images,* participants had higher performance at T2 ($p = <.001$) and T3 ($p = <.001$), whereas for the component *reasoning with images*, performance was better atT2 ($p = <.001$), but not at T3 ($p = 1.00$) in comparison to T1.

Further, we tested whether image interpretation scaffolds have more benefits for historical reasoning than mind mapping at T2 and T3. Figs. 6 and 7 show historical reasoning scores for the three conditions at T2 and T3. There was a significant difference between image interpretation scaffolds (domain-specific approach) and mind mapping (generic approach), $F(1, 142.43) = 4.18$, $p = .043$. Overall performance (estimated marginal means) was for mind mapping $M = .98$ ($SE = .08$; 95% CI .82 to 1.13), and for the students using image interpretation scaffolds $M = 1.17$ ($SE = .05$; 95% CI 1.07 to 1.28). In line with Hypothesis 2, students working with an image interpretation scaffold had higher overall historical reasoning scores than the mind mapping condition, both at T2 and at T3. Further, there was an interaction effect between the type of support and time point, $F(2, 139.76) = 5.01$, $p = .008$, showing that for students using image interpretation scaffolds, performance was lower at T3 than at T2. Follow-up analyses were then conducted for the separate

Table 4
Correlations between subcomponents of historical reasoning for time point 2.

Variable	1	2	3	4	5	6	7	8	9	10
1. Historical Questioning	–	.511[b]	.293[b]	.506[b]	.253[b]	.157	.137	.211[a]	.190[a]	.310[a]
2. Answering the Question	.511[b]	–	.218[b]	.670[b]	.353[b]	.189*	.296[b]	.229[b]	.354[b]	.342[b]
3. Relevance for the Present	.293[b]	.218[b]	–	.261[b]	.170[a]	.142	.202[a]	.239[b]	.288[b]	.238[b]
4. Essay Structure	.506[b]	.670[b]	.261[b]	–	.293[b]	.182[a]	.607[b]	.320[b]	.377[b]	.394[b]
5. Sourcing	.253[b]	.353[b]	.170[a]	.293[b]	–	.685[b]	.228[b]	.408[b]	.332[b]	.622[b]
6. Contextualization	.157	.189[a]	.142	.182[a]	.685[b]	–	.215[b]	.212[a]	.195[a]	.445[b]
7. Main Message of the Image	.137	.296[b]	.202[a]	.607[b]	.228[b]	.215	–	.169[a]	.362[b]	.313[b]
8. Image Description	.211*	.229[b]	.239[b]	.320[b]	.408[b]	.212[a]	.169[a]	–	.523[b]	.593[b]
9. Image Interpretation	.190[a]	.354[b]	.288[b]	.377[b]	.332[b]	.195[a]	.362[b]	.523[b]	–	.578[b]
10. Image Reference	.310[b]	.342[b]	.238[b]	.394[b]	.622[b]	.445[b]	.313[b]	.593[b]	.578[b]	–

Note. Correlations were calculated with Spearman's rho.
[a] $p < .05$.
[b] $p < .01$.

Table 5
Correlations between subcomponents of historical reasoning for time point 3.

Variable	1	2	3	4	5	6	7	8	9	10
1. Historical Questioning	–	.464[b]	.228[b]	.464[b]	.078	.068	.107	.168[a]	.226[b]	.154
2. Answering the Question	.464[b]	–	.257[b]	.566[b]	.095	.081	.144	.278[b]	.361[b]	.261[b]
3. Relevance for the Present	.228[b]	.257[b]	–	.208[a]	.121	−.012	.116	.118	.151	.032
4. Essay Structure	.464[b]	.566[b]	.208[a]	–	.080	.123	.568[b]	.356[b]	.332[b]	.378[b]
5. Sourcing	.078	.095	.121	.080	–	.732[b]	.065	.424[b]	.023	.544[b]
6. Contextualization	.068	.081	−.012	.123	.732[b]	–	.088	.334[b]	.009	.564[b]
7. Main Message of the Image	.107	.144	.116	.568[b]	.065	.088	–	.094	.224[b]	.234[b]
8. Image Description	.168[a]	.278[b]	.118	.356[b]	.424[b]	.334[b]	.094	–	.374[b]	.640[b]
9. Image Interpretation	.226[b]	.361[b]	.151	.332[b]	.023	.009	.224[b]	.374[b]	–	.496[b]
10. Image Reference	.154	.261[b]	.032	.378[b]	.544[b]	.564[b]	.234[b]	.640[b]	.496[b]	–

Note. Correlations were calculated with Spearman's rho.
[a] $p < .05$.
[b] $p < .01$.

components of historical reasoning: *asking and answering historical questions*, $F(1, 143.78) = 2.59$, $p = .11$; *reasoning about images*, $F(1, 142.32) = 2.92$, $p = .045$ (one-sided); and *reasoning with images*, $F(1, 141.79) = 2.57$, $p = .11$. These findings indicate that only for the component *reasoning about images,* there were differences between the mind mapping and the image interpretation scaffolds over time.

Moreover, we address whether there were differences between the two image interpretation scaffolds (sequenced and flexible condition) on overall historical reasoning and on the separate components of historical reasoning. There were no differences between the two conditions using a sequenced and flexible scaffold in overall historical reasoning performance, $F(1, 97.64) = .17$, $p = .678$. However, there was an interaction effect between time point and image interpretation scaffold, $F(2, 94.85) = 5.46$, $p = .006$. Estimated marginal means for the flexible scaffold condition indicated significantly higher performance at T2, $M = 1.30$ ($SE = .09$; 95% CI 1.13 to 1.46) than at T3, $M = 1.09$ ($SE = .08$; 95% CI .93 to 1.25), whereas the sequenced scaffold condition did not differ in performance across T2 and T3, $M = 1.13$ ($SE = .09$; 95% CI 1.00 to 1.35); $M = 1.12$ ($SE = .09$; 95% CI .95 to 1.28). In sum, Hypothesis 3a that historical reasoning would be better for the flexible scaffold condition than the sequenced scaffold condition could not be confirmed. Follow-up analyses for the separate components of historical reasoning did not show main effects of the type of scaffold (sequenced or flexible) on the three components of historical reasoning: *asking and answering historical questions*, $F(1, 99.34) = .42$, $p = .517$; *reasoning about images*, $F(1, 97.68) = .55$, $p = .459$; and *reasoning with images*, $F(1, 97.18) = .01$, $p = .905$. However, there was an interaction effect between time and

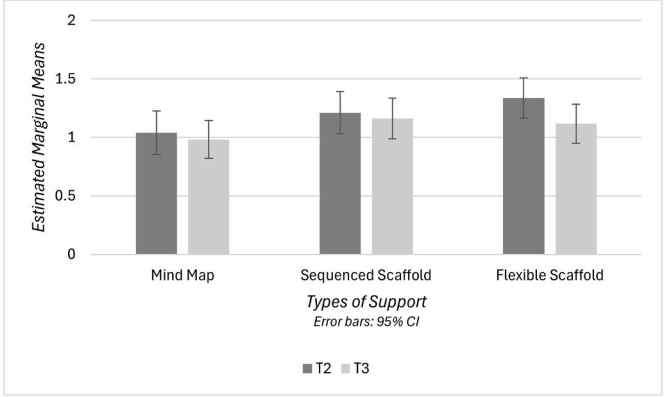


Fig. 6. Effects of Support on Students' Overall Historical Reasoning.
Note. Students' historical reasoning in image interpretation essays was assessed with a scoring rubric consisting of four levels: Level 0 (lowest) to level 3 (highest).

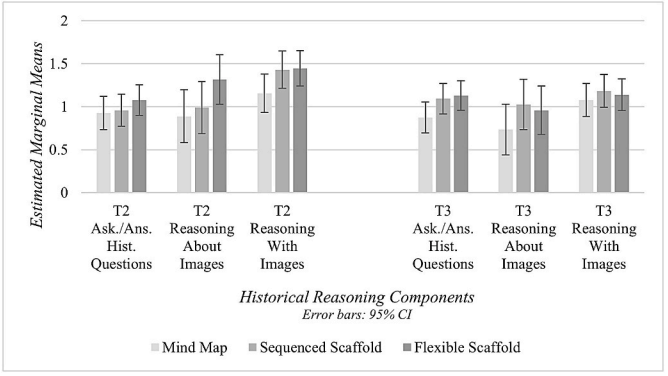


Fig. 7. Effects of Support on Students' Historical Reasoning per Component.
Note. Students' historical reasoning in image interpretation essays was assessed with a scoring rubric consisting of four levels: Level 0 (lowest) to level 3 (highest).

scaffold for the component *reasoning with images*, $F(2, 94.69) = 9.83$, $p = <.001$. Both image interpretation scaffolds showed significantly lower estimated marginal means at T3 than at T2: Sequenced scaffold, $M = 1.18$ ($SE = .10$; 95% CI .99 to 1.38), and $M = 1.43$ ($SE = .11$; 95% CI 1.22 to 1.65); Flexible scaffold, $M = 1.14$ ($SE = .10$; 95% CI .95 to 1.33), and $M = 1.45$ ($SE = .10$; 95% CI 1.24 to 1.65).

In addition, to better understand potential similarities and differences between the sequenced and flexible scaffold condition, we compared the two conditions for each separate subcomponent (i.e., each individually rated score) of historical reasoning at both time points. Table 6 shows the median scores and interquartile range for the subcomponents of historical reasoning. Note that analyses for ordinal data were used when comparing the two conditions, as these were all rated on a 4-point scale. Especially for "sourcing", "contextualization", and "relevance for the present", we expected that both at T2 and T3, the flexible scaffold condition would outperform the sequenced scaffold condition (Hypothesis 3b). Mann-Whitney-U-Tests were conducted to compare the two conditions separately per subcomponent and time point. The flexible scaffold condition did not outperform the sequenced scaffold condition for "sourcing" at T2, U = 1033.00, Z = −1.38, (one-sided) p = .084, and T3, U = 1027.00, Z = −.80, (one-sided) p = .213. For "contextualization", the flexible scaffold condition had higher scores than the sequenced scaffold condition at T2, U = 1001.00, Z = −1.69, (one-sided) p = .046, but not at T3, U = 1017.50, Z = −.89, (one-sided) p = .188. For "relevance for the present", the flexible scaffold condition was significantly better than the sequenced scaffold condition at T2, U = 1005.00, Z = −1.90, (one-sided) p = .029, but not at T3, U = 949.00, Z = −1.56, (one-sided) p = .060. In sum, Hypothesis 3b was partially confirmed; at T2, but not at T3, the flexible scaffold condition showed the highest scores for contextualization and interpreting the relevance of the historical image for the present.

4. Discussion

The present study investigated the effects of three conditions on students' historical reasoning when analyzing and interpreting historical images. An inquiry-based writing task resulting in an image interpretation essay was used to investigate students' historical reasoning *about* and *with* images. The conditions were (A) a sequenced Panofsky (1939/2018)-based scaffold, (B) mind mapping (Buzan & Buzan, 1993), and (C) a flexible scaffold based on Bätschmann (2009) and adapted for this study.

Our first question was whether students show better historical reasoning when interpreting an image using support (the three conditions) than without (no conditions). Using support at T2 improved students' historical reasoning compared to not using any support at T1. Particularly, as follow-up analyses of the components of historical reasoning showed, image interpretation supports improved students' *reasoning about images*, indicating that the image interpretation supports had benefits on students' image analysis, both at T2 and T3. Moreover, support appeared to have benefits on the component *reasoning with images*, addressing skills using information from the analysis to create a meaningful essay; however, only at T2, but not at T3, support improved scores on this component.

Although support improved overall historical reasoning, as indicated by the essay scores at T2, surprisingly, and in contrast to our hypothesis, at T3, even though students used support, overall historical reasoning did not differ from T1 when no support was used. The question arises why this would be the case, as students had more experience with the support at T3. Follow-up analyses for the different historical reasoning components indicated lower scores at T2 and T3 than at T1 for the component *asking and answering historical questions*. It thus seems that for this specific component *asking and answering historical questions*, support in the form of mind mapping or domain-specific scaffolding does not help. Rather than the type of support, the type of image may have affected students' ability to ask and answer historical questions. Possibly, students had higher prior knowledge about the topic of the Swiss vote on women's suffrage (the topic of T1), than about the Louisville Flood in the US (topic T2) or the former Berlin Wall in Germany (topic T3). Furthermore, about performance for the component *reasoning with images* was higher at T2 (when using support) in comparison to T1 (without support), there were no differences between T1 and T3. This may be an indication that also, when reasoning *with* images, for which students need to select and order information found on the internet, the type of image and the type of information they find may affect their skills. At T3, filtering the overload of information on the internet about the Berlin Wall may have been challenging, regardless of the type of support they received. Students may not have filtered sufficiently, but rather seemed to over rely on information from the internet, than on their own inspection of the image. At T2, reading and filtering the information about the Louisville Flood, which was mainly in the English language, may have been challenging for German-speaking students. This may be a reason why they focused more on the analysis of the image itself rather than on the textual information they found at T2. Wolfrum and Sauer (2007) found that secondary school students rated information content found in texts higher than information conveyed by images. Also in the present study, there are indications that students may have been more focused on the information found on the internet as on the image itself, particularly at T3; for instance, the subcomponent "image reference" had a higher number of zero-ratings at T3 (11%, of these four students were assigned to the mind mapping, six to the sequenced scaffold, and six to the flexible scaffold condition) than at T2 (7.6%, of these six students were assigned to the mind mapping, three to the sequenced scaffold, and two to the flexible scaffold condition). When further inspecting the essays with the lowest scores for "image reference", it became evident that students copied and pasted

Table 6
Median and interquartile range of historical reasoning subcomponents per time point and type of support.

Subcomponent	T1			T2			T3		
	M	*S*	*F*	*M*	*S*	*F*	*M*	*S*	*F*
	Mdn (IQR)	*Mdn (IQR)*	*Mdn (IQR)*	*Mdn (IQR)*	*Mdn (IQR)*	*Mdn (IQR)*	*Mdn (IQR)*	*Mdn (IQR)*	*Mdn (IQR)*
Historical Questioning	2 (1–2)	2 (1–2)	2 (1–2)	1 (1–2)	1 (1–2)	2 (1–2)	2 (1–2)	2 (1–2)	2 (1–2)
Answering the Question	1 (1–2)	1 (0–2)	1 (0–2)	1 (0–2)	1 (0–2)	1 (0–2)	1 (0–1)	1 (.75–2)	1 (0–2)
Essay Structure	2 (1–2)	2 (1–2)	2 (0-2	1 (0–2)	2 (1–2)	2 (.25–3)	1 (.5–2)	2 (1–3)	2 (0–2)
Relevance for the Present	1 (0–1)	.5 (0–1)	.5 (0–2)	0 (0–1)	0 (0–0)	0 (0–1)	0 (0–0)	0 (0–1)	0 (0–1)
Main Message of the Image	0 (0–1.5)	1 (0–2)	1 (0–2)	0 (0–1)	1 (0–2)	1 (0–2)	0 (0–1)	0 (0–2)	0 (0–2)
Sourcing	0 (0–1)	0 (0–1)	0 (0–2)	0 (0–2)	1 (0–2)	2 (0–3)	0 (0–1)	0 (0–2)	1 (0–2)
Contextualization	0 (0–1)	0 (0–.75)	0 (0–0)	0 (0–1)	0 (0–2)	1 (0–2)	0 (0–2)	.5 (0–2.25)	0 (0–2)
Image Description	1 (1–2)	1 (.25–2)	1 (1–2)	1 (.5–2)	2 (1–3)	2 (1–2)	1 (1–2)	1 (1–2)	1 (1–2)
Image Reference	1 (1–2)	1 (1–2)	1.5 (1–2)	2 (1–3)	2 (1–3)	2 (1–3)	2 (1–2)	2 (1–3)	2 (1–3)
Image Interpretation	1 (0–1)	1 (0–1)	1 (0–1)	1 (0–2)	1 (0–2)	1 (0–2)	0 (0–1)	1 (0–1)	1 (0–1)

Note. M = Mind Map; S = Sequenced Scaffold; F = Flexible Scaffold. Students' historical reasoning in image interpretation essays was assessed with a scoring rubric consisting of four levels: Level 0 (lowest) up to Level 3 (highest).

information from the internet, confirming that they prioritized textual over visual information. Students appeared to write about the historical topic rather generally instead of connecting it to the image and its meaning. A further explanation for the drop in overall historical reasoning performance from T2 to T3 could be that students were more interested in the task at T2, when the support was new to them, than at T3. Future research could address if feedback would further improve the quality of students' image interpretation essays.

Furthermore, we investigated whether there were differences between generic (mind map condition) and domain-specific approaches (sequenced and flexible scaffold condition) on historical reasoning performance. Students who used a domain-specific approach (sequenced and flexible scaffold) outperformed those who made mind maps, showing higher scores for the component *reasoning about images*. Particularly for the mind map condition, performance declined from T1 to T3, indicating that this support did not enhance historical reasoning about the images at all. Although it has been suggested that mind mapping is helpful when brainstorming about a topic or summarizing content on a general level (Davies, 2011), the present findings may indicate that, when critically reasoning about complex images, more guidance is necessary. Apparently, the image interpretation scaffolds offered more elaborate and structured guidance, including directive questions, which supported students with a more thorough examination of images.

We further investigated whether there were differences in historical reasoning performance between the sequenced scaffold condition and the flexible scaffold condition. The flexible scaffold is more detailed and elaborated than the sequenced scaffold, as it contains guiding questions about the connection between the past and the present. Further, the flexible scaffold leaves more autonomy to students regarding which questions they address, and in which order they do so. Although historical reasoning did not differ between conditions, as hypothesized, for "contextualization" and "relevance for the present" (but not for "sourcing", contrasting our expectations), at T2, the condition using the flexible scaffold had higher scores than the condition using the sequenced scaffold. Although the flexible scaffold had specific benefits for interpretation of the image context and drawing parallels between the past and the present, this effect was only visible at T2; the sequenced and the flexible scaffold condition did not differ on these subcomponents at T3. This may imply that there is no one-size-fits-all solution to support critical reasoning *about* images. That is, only using a specific image interpretation scaffold at all times does not appear to be sufficient to improve image interpretation skills. The present findings draw a more complex picture and indicate that it is necessary to further understand at what times, for what images, and for which complexity levels specific interpretation scaffolds can be beneficial. Future research could investigate how image interpretation scaffolds can support historical reasoning about different types of images, and how students can be trained to select relevant information from internet sources when making meaning of images. Moreover, learner resources should be considered when further investigating the effects of image interpretation scaffolds. For instance, prior knowledge may affect how learners make use of interpretation scaffolds, and learner constraints and resources such as achievement level (Lange, 2011), feelings of autonomy, motivation (Ryan & Deci, 2017), or cognitive load (Mayer & Moreno, 2003) can affect their historical reasoning skills.

Our study aimed to evaluate the effectiveness of three support conditions on students' historical reasoning when learning with images. Our findings provide valuable insights into the potential benefits of support, especially when *reasoning about images*, i.e., image analysis. Most importantly, domain-specific approaches (sequenced and flexible scaffold condition) appear to guide students better than a generic approach (mind mapping condition): students analyze the visual image information more carefully, and have longer-lasting benefits (i.e., over two time points). The choice between a sequenced or a flexible image interpretation scaffold appears to have no significant impact on students' image analysis and interpretation. Although, using a flexible scaffold seems promising when students need to contextualize an image within its historical framework or to relate the image's message to contemporary issues.

However, several limitations should be acknowledged. It was challenging to find adequate images for this study, especially when it comes to the difficulty level and (seemingly impossible) comparability of the photographs. The term "difficulty" seems rather general and unclear when selecting images, because every viewer would evaluate the complexity of an image differently, depending on for instance prior knowledge, image composition and image details, or inherent symbols. Limited research investigated historical image interpretation skills; therefore, we could not make use of images that had been tested before in research with comparable research purposes. Future studies could make use of a rotational study design, to better understand the comparability of images by separating the impact of time point, image type, and type of support (i.e., a design in which the images are differentially assigned to the different time points). Moreover, the study shows low scores of historical reasoning over time (which may be indicated of a bottom effect), which can be seen as a limitation. However, it also shows that when dealing with complex images, students lack the competencies to interpret these, filter information, and reason about these, presumably because they are not trained to do so (Bernhard, 2017). Future studies could investigate how these competencies could be built up and whether they could be maintained in a sustainable way (e.g., with a follow-up test without any support). Further, internet use of students was not tracked. Future research could investigate adolescents' internet research process to obtain more insights into search histories when comparing images or type of supports. Another potential limitation could be, that students had only half the time to complete the task at T1 in comparison to T2 and T3, because at later time points students needed extra time for their image analysis (as they were either creating mind maps or writing comments in the interpretation scaffolds at these later time points). At T1 no support for analysis was given. However, at T1, most students could finish their essays without any time pressure and started with an additional task (finding differences between images, the task was distributed to keep them occupied until the end of the session). Therefore, we do not think that the fact that students had less time at T1 has affected our results.

Despite these limitations, our intervention study is one of the first that has empirically tested effects of image interpretation scaffolds and mind mapping on students' historical reasoning. Although our primary research aim was to investigate benefits of different types of support on students' historical reasoning *about* and *with* images, their improved critical inspection of images may also transfer to other image types. Possibly, not only reasoning about historical images, but also about contemporary images may benefit from support; critical reasoning about such images is crucial, as these are present and manipulated more than ever on the internet. We did not specifically address transfer or long-term effects of using different types of support and different types of images, future research should do so.

5. Conclusion

The unique contribution of this study is that it brings insights into ways students can be supported with their historical reasoning *about* and *with* images. The strength of this design is that three types of support (sequenced scaffold, mind mapping, and a flexible scaffold) were tested over two time points. This design matches actual education practices, where tasks and support are typically repeatedly implemented and practiced, rather than being used only once. Although future research is warranted, our findings can bring a useful starting point for the history education context and for teachers who aim to support students with critical reasoning *about* and *with* images. Results imply that image interpretation scaffolds can be promising to support students with image analysis; these domain-specific scaffolds have more beneficial effects

than generic mind mapping. Both types of scaffolds, the sequenced and the flexible, seem to have the potential to enhance students' historical reasoning. Educational history textbooks commonly use Panofsky's image interpretation scaffold and its adaptations when learning with images (e.g., Schmidt-Maiwald, 2018). Results show that the flexible scaffold can also be helpful to support students, particularly when it is necessary to focus attention on contextualization and connecting the image message with the relevance for the present. This study can provide teachers with a more extensive toolbox. Findings seem to indicate that although interpretation scaffolds help, no scaffold seems best – that is, teachers can flexibly switch between scaffolds and strategically decide which scaffold they like to use at what time. Importantly, students need to recognize the value of a critical and thorough image analysis so they can apply these skills even when no scaffold is available. Additionally, image interpretation scaffolds may not sufficiently help students to interpret all types of images. Particularly when there is an overload of information on the internet, students seem to get easily distracted from the visual information and prioritize textual information instead. In these cases, meaning making from images can be hindered because a critical visual image inspection stops to soon, making it hard to discover controversies or false information. Future research should further investigate how additional support can be offered to students to interpret images and to critically select relevant information about these.

CRediT authorship contribution statement

Kevin van Loon: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Monika Waldis:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization.

Appendix A

Writing Guideline Showing a High-Quality Essay Structure when Interpreting an Image

Question as title
Question related to the image, the historical event and the present day
- What kind of question do I put in the center for my image interpretation that connects the image, the historical event and the present?

Introduction
Entry into the topic
- How do I give the reader a brief overview of the topic? What is it about?
- What is the goal of my image interpretation and how did I arrive at my question?

Main part
The main message of the image and appropriate image description
- What is the main message of the image?
- Image description: Which image details support my main message?

Summarizing the collected information in a meaningful way
- Which of the information collected (research/materials) is relevant to understand and explain the image?

Conclusion
Answer to my question
- What conclusion do I come to, and can I also justify this with the help of the information collected (research/materials)?

Appendix B

Students' Prompt for the Image Interpretation

Give this image a meaning like a historian would do by using your collected information to write a meaningful image interpretation. Write an interpretative essay of at least 300 words, taking the following criteria into account:

- Formulate a relevant question about the image, referring to the history and the present.
- Mention the main message of the image and describe the image accurately.
- Summarize the information you collected (internet research/materials) to analyze the image.
- Write a conclusion in which you present and justify your interpretation of the image.

Appendix C

Example Essay of an Image Interpretation

Has the child labor shown in the photograph by Lewis W. Hine changed today?

In the time of industrialization, workers often earned too little money to be able to feed their families. For this reason, children also had to work hard. Employers took advantage of the children: Long hours, physically demanding jobs, and almost no pay. In addition, the work prevented them from going to school, which worsened their future prospects. With my image interpretation, I want to pursue the question of the extent to which the photographer was successful with his photos against child labor and whether the children are better off today than they were then.

The yellowish-brownish photo by Lewis W. Hine "John Howell, an Indianapolis newsboy" from 1908 documents the dark side of industrialization, child labor in the USA. In the foreground you can see the elongated shadow of the photographer and his camera, which is on a tripod. In the middle, the newspaper boy John Howell can be seen, holding the newspapers he wants to sell under his right arm. These appear huge in comparison to the boy. The newspaper boy looks small, helpless, and lost with his eyes downcast, even though he is in the center of the picture. This impression is reinforced by the photographer's long shadow and the tall streetlamp to his left. In addition, there are no people around him, which makes him even more lonely.

In the background, there is a street corner with people walking.

Hine, a photographer, and teacher was working for the National Child Labor Committee at the time, so the welfare of children was important to him. This organization campaigned against child labor and wanted to make the public aware of the poor working conditions. Through Hine's photos, a larger audience could be made aware of the problem, as these were published by the NCLC in newspapers or their own publications.

The photographer advocated for the children by taking many photos that showed their poor working conditions. Hine and the NCLC hoped that by publishing the photos, child labor could be combated. Unfortunately, however, child labor was not banned in the U.S. until 30 years later, in 1938. Child labor was not only a problem then, but it still is today. In various countries around the world, children continue to be exploited in factories, in mining, or in tourism.

Appendix D

Scoring Rubric to Assess Historical Reasoning Subcomponents

	Level 0	Level 1	Level 2	Level 3
Historical Questioning A historical question relates to the historical event/phenomenon and/ or the inherent social issue present in the image	No question	No historical question **or** no historical question but with 1 time reference point	Historical question with 1 time reference point **or** no historical question but with 2 time reference points **or** question aiming at image comprehension	Historical question with 2 time reference points **or** question aiming at more complex image comprehension with 2 time reference points
Answering the Question	No answer to the question **or** no question has been asked	Answer is given but not supported with an argument **or** answer is false, unlogic	Answer is given and supported with 1 argument	Answer is given and supported with more than 1 argument
Sourcing -Author/employer -Technique/type of image -Title -Date -Place of origin/location	No sourcing	Sourcing with 1 aspect of image formation	Sourcing with 2 aspects of image formation	Sourcing with more than 2 aspects of image formation
Contextualization -Knowledge about the author -Publication context -Targeted audience -Intention/motivation of author -Tendency towards media critique	No contextualization	Contextualization with 1 aspect about the context of the image	Contextualization with 2 aspects about the context of the image	Contextualization with more than 2 aspects about the context of the image
Image Description -Color -Light-shadow -Relations -Perspectives -Fore-, middle-, and background -Facial expressions, gestures	No image description	Image description with just a few image details connected to 0–1 aspect of image composition	Image description with several image details connected to 1–2 aspects of image composition	Image description with many or all image details connected with at least 3 aspects of image composition
Main Message of the Image	No main message of the image	Implicit main message of the image considering the historical event/phenomenon or the inherent social issue **or** explicit main message of the image not considering the historical event/ phenomenon or the inherent social issue **or** main message of the image is false	Explicit main message of the image considering the historical event/ phenomenon **or** the inherent social issue	Explicit main message of the image considering both, the historical event/phenomenon, **and** the inherent social issue
Image Interpretation	No connections between the image and the historical event/ phenomenon and the inherent social issue **or** connections are not explained/supported with arguments **or** connections are false or overinterpreted	1 connection is explained and supported with an argument	Several connections are explained and supported with arguments **or** a contrast/contradiction (image in the image) is noticed, but not explained/justified	Several connections are explained and supported with arguments **and** a contrast/ contradiction (image in the image) is noticed and explained/ justified
Image Reference	No image reference	Tendency towards an image reference is apparent, and refers to less than half of the text	Image reference is partly apparent, and refers to the half of the text	Image reference is totally apparent, and refers to more than the half of the text
Relevance for the Present	No relevance for the present	Tendency towards a relevance for the present is apparent, but is unclear, or without connecting it to the historical event/phenomenon or the inherent social issue	Relevance for the present is partly apparent, but without giving a recent example **or** a recent example was given, but is unclear	Relevance for the present is totally apparent and clearly explained with a recent example
Essay Structure Consisting of 5 elements:	Essay is not structured, consisting of max. 2 elements	Tendency towards a structured essay is apparent, consisting of max. 3–4 elements **or** consisting of	Essay is partly structured consisting of at least 4 elements and essay is clearly structured with	Essay is completely structured consisting of all the 5 elements **and** essay is clearly structured

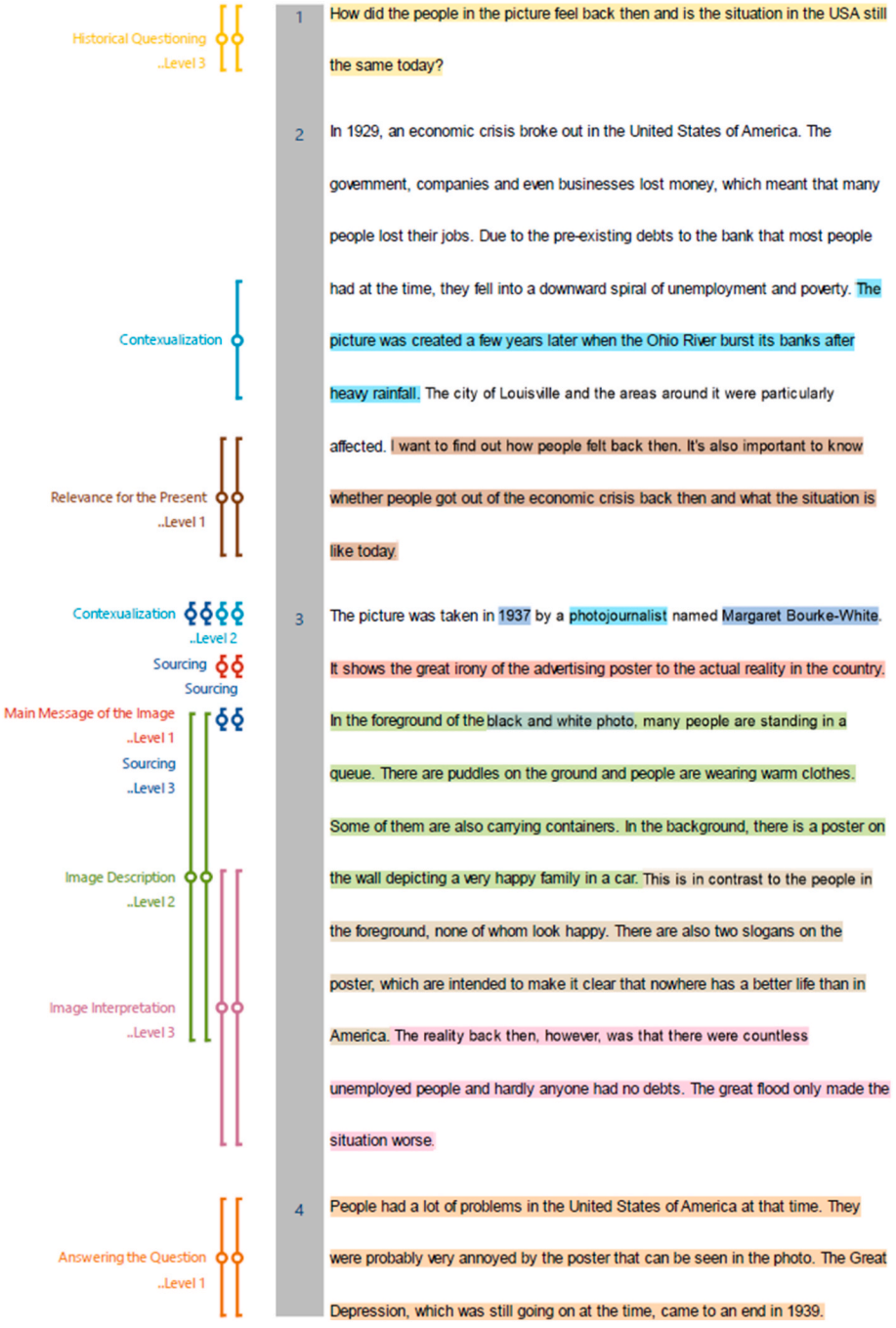
(*continued on next page*)

(*continued*)

	Level 0	Level 1	Level 2	Level 3
-asking a question (1 element) -introduction (1 element) -main part with image description & main message (2 elements) -conclusion/answer (1 element)		4 elements but essay is not structured with paragraphs	paragraphs **or** essay is completely structured consisting of all 5 elements, but essay is not always structured with paragraphs	with an introduction, main part, and a conclusion with paragraphs

Appendix E

Example of a Coded Students' Essay



Note. Additional codes refer to the entire essay: "image reference" (Level 3); "essay structure" (Level 3).

References

Aeby, R., Aebi, R., Baeriswyl, E., Oberholzer, F., Sinik, F., Suter, T., & Davanzo, E. (2017). *1. Durchblick Geschichte Sekundarstufe I* (pp. 6–116). Westermann Schweiz.

Al-Zyoud, A. A., Al Jamal, D., & Baniabdelrahman, A. (2017). Mind mapping and students' writing performance. *Arab World English Journal, 8*. https://doi.org/10.24093/awej/vol8no4.19

Bätschmann, O. (2009). Einführung in die kunstgeschichtliche Hermeneutik. *Die Auslegung von Bildern*. Wissenschaftliche Buchgesellschaft WBG.

Baxandall, M. (1990). Ursachen der Bilder: Über das historische Erklären von Kunst. *Reimer*.

Bernhard, R. (2017). Visual literacy - theoretische Überlegungen und empirische Befunde über Lernaufgaben zu Bildern im Geschichtsunterricht. *Österreichische Pädagogische Zeitschrift. Erziehung und Unterricht, 9–10*, 954–962.

Bernhardt, M. (2007). Vom ersten auf den zweiten Blick. Eine empirische Untersuchung zur Bildwahrnehmung von Lernenden. *Geschichte in Wissenschaft und Unterricht, 58* (7), 417–432.

Bernhardt, M. (2011). Ich sehe was, was Du nicht siehst!"Überlegungen zur Kompetenzentwicklung im Geschichtsunterricht am Beispiel der Bildwahrnehmung. *Visualität und Geschichte*.

Bernhardt, M. (2013). Visual history: Einführung in den themenschwerpunkt. *Zeitschrift für Geschichtsdidaktik, 12*(1), 5–8.

Bernhardt, M. (2018). "Sehen kann jeder!" Zu einem Irrtum der geschichtsdidaktischen Bildinterpretation. In K. Krüger, & K. Kranhold (Eds.), *Bildung durch Bilder. Kunstwissenschaftliche Perspektiven für den Deutsch-, Geschichts- und Kunstunterricht* (pp. 209–230).

Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction, 20*(4), 485–522. https://doi.org/10.1207/S1532690XCI2004_2

Brown, T. (2006). *Confirmatory factor analysis for applied research*. Guilford Press.

Burke, P. (2008). *Eyewitnessing: The uses of images as historical evidence*. Cornell University Press.

Büttner, S. (2014). *Tutorium quellenarbeit*. Retrieved 12.05. from.

Buzan, T., & Buzan, B. (1993). *The mind map book*. BBC Worldwide Limited.

Coirier, P., Andriessen, J., & Chanquoy, L. (1999). From planning to translating: The specificity of argumentative writing. In J. Andriessen, & P. Coirier (Eds.), *Studies in writing: Foundations of argumentative text processing* (Vol. 5, pp. 1–28). University Press.

Davies, M. (2011). Concept mapping, mind mapping and argument mapping: What are the differences and do they matter? *Higher Education, 62*, 279–301. https://doi.org/10.1007/s10734-010-9387-6

Fuhs, B. (2006). Narratives Bildverstehen. Plädoyer für die erzählende Dimension der Fotografie. In W. Marotzki, & H. Niesyto (Eds.), *Bildinterpretation und Bildverstehen: Methodische Ansätze aus sozialwissenschaftlicher, kunst-und medienpädagogischer Perspektive* (pp. 207–225). VS Verlag für Sozialwissenschaften.

Hamann, C. (2012). Bildquellen im Geschichtsunterricht. In M. Barricelli, & M. Lücke (Eds.), *Handbuch praxis des geschichtsunterrichts* (Vol. 2, pp. 108–124). Wochenschau.

Krammer, R. (2006). Historische Kompetenzen erwerben – durch das Arbeiten mit Bildern? In R. Krammer, H. Ammerer, & W. Schreiber (Eds.), *Mit bildern arbeiten: Historische kompetenzen erwerben* (pp. 21–38). Ars Una.

Kuckartz, U. (2014). *Qualitative text analysis: A guide to methods, practice and using software*. Sage.

Külling, C., Waller, G., Suter, L., Willemse, I., Bernath, J., Skirgaila, P., & Süss, D. (2022). *James – jugend, aktivitäten, medien – erhebung schweiz*. Zürcher Hochschule für Angewandte Wissenschaften.

Kyriazos, T., & Poga-Kyriazou, M. (2023). Applied psychometrics: Estimator considerations in commonly encountered conditions in CFA, SEM, and EFA practice. *Psychology, 14*(5), 799–828. https://doi.org/10.4236/psych.2023.145043

Lange, K. (2011). *Historisches Bildverstehen, oder, Wie lernen Schüler mit Bildquellen? Ein Beitrag zur geschichtsdidaktischen Lehr-Lern-Forschung*. LIT Verlag.

Lévesque, S. (2005). Teaching second-order concepts in Canadian history: The importance of" historical significance". *Canadian Social Studies, 39*(2), n2.

Logtenberg, A. (2012). *Questioning the past: Student questioning and historical reasoning [thesis, fully internal*. Universiteit van Amsterdam. Ipskamp.

Luís, R., & Rapanta, C. (2020). Towards (re-) defining historical reasoning competence: A review of theoretical and empirical research. *Educational Research Review, 31*, Article 100336.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Monte-Sano, C., & De La Paz, S. (2012). Using writing tasks to elicit adolescents' historical reasoning. *Journal of Literacy Research, 44*(3), 273–299. https://doi.org/10.1177/1086296X12450445

Newsom, J. T. (2023a). *Some Clarifications and Recommendations on fit indices* [unpublished manuscript]. In *Departement of psychology*. Portland: State University.

Newsom, J. T. (2023b). *Improper solutions in SEM [unpublished manuscript]*. Department of Psychology, Portland State University.

Nokes, J. D. (2017). Exploring patterns of historical thinking through eighth-grade students' argumentative writing. *Journal of Writing Research, 8*(3). https://doi.org/10.17239/jowr-2017.08.03.02

Oleschko, S. (2013). "Ich verstehe nix mehr." Zur Interdependenz von Bild und Sprache im Geschichtsunterricht. *Zeitschrift für Geschichtsdidaktik, 12*(1), 112–127.

Ormond, B. (2011). Enabling students to read historical images: The value of the three-level guide for historical inquiry. *The History Teacher, 44*(2), 179–190. https://doi.org/10.1080/09585176.2011.550755

Panofsky, E. (1939/2018). Studies in iconology I. Introductory. In *Studies in iconology: Humanistic themes in the art of the renaissance*. Routledge (Original work published 1939).

Partington, G. (1986). History: Re-Written to ideological fashion. *The wayward curriculum: A cause for parents' concern*.

Paul, G. (2013). Visual History und Geschichtsdidaktik: Grundsätzliche Überlegungen. *Zeitschrift für Geschichtsdidaktik, 12*(1), 9–26.

Paul, R., & Elder, L. (2019). *The miniature guide to critical thinking concepts and tools*. Rowman & Littlefield.

Pfisterer, U. (2020). *Kunstgeschichte zur Einführung*. Junius Verlag GmbH.

Phillips, R. (2002). *Historical significance-the forgotten 'key element'? Teaching History, 106*, 14–19. https://www.jstor.org/stable/43259875.

Revelle, W. (2023). psych: Procedures for psychological, psychometric, and personality research. https://CRAN.R-project.org/package=psych.

Roeck, B. (2004). *Das historische Auge: Kunstwerke als Zeugen ihrer Zeit: von der Renaissance zur Revolution*. Vandenhoeck & Ruprecht.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Rouet, J.-F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology, 88*(3), 478.

Ryan, R. M., & Deci, E. L. (2017). Self-determination theory. *Basic psychological needs in motivation, development, and wellness*. The Guilford Press.

Schmidt-Maiwald, C. (2018). Bildverstehen und die Methodik der Kunstbetrachtung. *Imago. Zeitschrift für Kunstpädagogik, 6*, 6–16.

Schreiber, W., Körber, A., Borries, B., Krammer, R., Leutner-Ramme, S., Mebus, S., … Ziegler, B. (2006). *Historisches Denken. Ein Kompetenz-Strukturmodell*. Ars una.

Sendur, K. A., van Drie, J., & van Boxtel, C. (2021). Historical contextualization in students' writing. *The Journal of the Learning Sciences, 30*(4–5), 797–836. https://doi.org/10.1080/10508406.2021.1939029

Shi, Y., Yang, H., Dou, Y., & Zeng, Y. (2022). Effects of mind mapping-based instruction on student cognitive learning outcomes: A meta-analysis. *Asia Pacific Education Review*, 1–15. https://doi.org/10.1007/s12564-022-09746-9

Van Boxtel, C., & Van Drie, J. (2012). "That's in the time of the Romans!" Knowledge and strategies students use to contextualize historical images and documents. *Cognition and Instruction, 30*(2), 113–145. https://doi.org/10.1080/07370008.2012.661813

Van Boxtel, C., & Van Drie, J. (2018). *Historical reasoning: Conceptualizations and educational applications*. The Wiley International Handbook of History Teaching and Learning.

Van Drie, J., & Van Boxtel, C. (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

Van Drie, J., Van Boxtel, C., Jaspers, J., & Kanselaar, G. (2005). Effects of representational guidance on domain specific reasoning in CSCL. *Computers in Human Behavior, 21*(4), 575–602. https://doi.org/10.1016/j.chb.2004.10.024

Van Loon, K., Studer, D., & Waldis, M. (2024). Investigating adolescents' historical reasoning skills when analyzing and interpreting an image. *Historical Encounters, 11* (1), 95–112.

Van Nieuwenhuyse, K., Roose, H., Wils, K., Depaepe, F., & Verschaffel, L. (2017). Reasoning with and/or about sources? The use of primary sources in flemish secondary school history education. *Historical Encounters: A Journal of historical consciousness, historical cultures, and history education, 4*(2), 48–70. https://doi.org/10.52289/hej4.200

Waldis, M., Nitsche, M., & Gollin, K. (2020). "Schülerinnen und Schüler schreiben Geschichte" – Eine Interventionsstudie an Deutschschweizer Gymnasien. *Zeitschrift für Geschichtsdidaktik, 19*(1), 90–108.

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*(2), 301.

Wineburg, S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83*(1), 73–87. https://doi.org/10.1037/0022-0663.83.1.73

Wissner, R. A. (2022). Active images: Teaching students critical thinking and analysis skills with picture postcards. In *Innovative approaches in pedagogy for higher education classrooms*. Emerald Publishing Limited. https://doi.org/10.1108/S2055-364120220000042006.

Wolfrum, B., & Sauer, M. (2007). Zum Bildverständnis von Schülern. Ergebnisse einer empirischen Studie. *Geschichte in Wissenschaft und Unterricht, 58*(7/8), 400–416.

Kevin van Loon is a PhD Student at the University of Basel and Lecturer at the FHNW School of Education, University of Applied Sciences and Arts Northwestern, Switzerland.

Monika Waldis, Prof. Dr., is the Head of the Center for Political Education and History Didactics at the Institute for Research and Development, at the FHNW School of Education, University of Applied Sciences and Arts Northwestern Switzerland.