UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

The effects of explicit teaching of strategies, second-order concepts, and epistemological underpinnings on students' ability to reason causally in history

Stoel, G.L.; van Drie, J.P.; van Boxtel, C.A.M.

DOI
10.1037/edu0000143
Publication date
2017
Document Version
Final published version
Published in
Journal of Educational Psychology

Link to publication

Citation for published version (APA):
Stoel, G. L., van Drie, J. P., & van Boxtel, C. A. M. (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

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:23 Sep 2021

Journal of Educational Psychology
2017, Vol. 109, No. 3, 321–337

© 2016 American Psychological Association
0022-0663/17/$12.00 http://dx.doi.org/10.1037/edu0000143

The Effects of Explicit Teaching of Strategies, Second-Order Concepts, and Epistemological Underpinnings on Students' Ability to Reason Causally in History

Gerhard L. Stoel, Jannet P. van Drie, and Carla A. M. van Boxtel
University of Amsterdam

This article reports an experimental study on the effects of explicit teaching on 11th grade students' ability to reason causally in history. Underpinned by the model of domain learning, explicit teaching is conceptualized as multidimensional, focusing on strategies and second-order concepts to generate and verbalize causal explanations and epistemological underpinnings connected to causal reasoning in history. In a randomized pretest–posttest design ($N = 95$), with a treatment and a control condition, effects of explicit teaching were investigated on students' (a) second-order and strategy knowledge, (b) their epistemological beliefs, and (c) their ability to construct a causal explanation, as well as (d) their topic knowledge, and (e) their individual interest. Results show that students in the experimental group scored significantly higher at the posttest on knowledge of causal-reasoning strategies and second-order concepts ($sr^2 = .09$), attributed a significantly higher value to criterialist epistemological beliefs ($sr^2 = .04$), and reported a higher individual interest ($sr^2 = .02$). We found no differences between conditions in the overall quality of students' written explanations. However, the experimental group scored significantly higher on 1 core criterion, that is, the "use of second-order language and causal connections" ($sr^2 = .06$). No differences were found on first-order knowledge. Furthermore, self-reports on learning gains and correlational analysis were applied to explore the interrelatedness of second-order and strategy knowledge, epistemological beliefs, student's ability to construct a causal explanation, topic knowledge, and individual interest.

Keywords: historical reasoning, history instruction, explicit teaching, instructional design, epistemological beliefs

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Over the past two decades, researchers of history education have emphasized the importance of history education as a subject that allows students to develop skills and competencies which are considered important in a democratic and pluralistic society (Barton & Levstik, 2004). As a consequence, historical reasoning has been included in recent years in the national history curricula in many countries (e.g., the Netherlands, Australia, Canada, the United Kingdom; Erdmann & Hassberg, 2011). Among other things, students should learn to reason critically with and about multiple sources; to judge the reliability, usefulness, and representativeness of these sources; and to embed them in their historical context. Furthermore, students should learn to construct and deconstruct historical narratives, which demands understanding that these narratives do not primarily present copies of the past but interpretations and that multiple perspectives can coexist. Finally, students should learn to judge the validity of these interpretations using disciplinary criteria (Seixas & Morton, 2013; VanSledright, 2011; Wineburg, 2001).

Although the importance of teaching historical reasoning skills is widely accepted, relatively little is known about pedagogical principles that foster the development of this reasoning (Levstik & Barton, 2008; van Boxtel & van Drie, 2013). In a previous literature review and an experimental pilot study, several basic principles of a learning environment, intended to foster causal historical reasoning, were defined (i.e., designing open-ended tasks, allowing for social interaction, raising situational interest; Stoel, van Drie, & van Boxtel, 2015). However, our review and pilot study also showed the indispensability of explicit teaching as a design principle in a learning environment intended to develop students' ability to reason causally in history.

Previous studies have shown the effectiveness of explicit teaching in distinct topics such as sourcing strategies in history (Nokes, Dole, & Hacker, 2007; Reisman, 2012), writing historical essays (De La Paz, 2005; De La Paz & Felton, 2010), and epistemological beliefs in science (Khishfe & Abd-El-Khalick, 2002). The current study adds to this research by conceptualizing the focus of explicit teaching in a more integral fashion. We argue from both a theoretical and an empirical standpoint that fostering a causal historical reasoning skill entails teaching explicitly about the strategies and

This article was published Online First June 20, 2016.

Gerhard L. Stoel, Jannet P. van Drie, and Carla A. M. van Boxtel, Research Institute of Child Development and Education, University of Amsterdam.

We kindly thank the history teachers and students from the Keizer Karel College, Amstelveen, the Netherlands, for their involvement. We also thank Terrence Jorgensen, University of Amsterdam, for his help with the statistical analyses.

Correspondence concerning this article should be addressed to Gerhard L. Stoel, Research Institute of Child Development and Education, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS, Amsterdam, the Netherlands. E-mail: g.l.stoel@uva.nl

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

second-order concepts related to historical causation and about the epistemological underpinnings of constructing historical explanations. To investigate the effects of explicit teaching on students' causal-historical reasoning, a randomized controlled trial with two conditions was conducted with 95 eleventh grade students.

Developing Causal Historical Reasoning

Prior to defining pedagogical principles, we conducted a literature review to delineate the cognitive dimensions involved in developing causal-historical reasoning. The model of domain learning (MDL; Alexander, 2003, 2005) provided an appropriate framework toward this goal. Elaborating on this model, we differentiated between (a) knowledge of causal strategies and second-order concepts related to historical causation and (b) epistemological beliefs about the nature of causal interpretations in history, as important aspects underpinning causal historical reasoning. Besides, the MDL conceptualized first-order knowledge and (situational) interest to be important ingredients of this reasoning (Stoel et al., 2015).

The MDL emphasizes that developing expertise in any domain involves acquiring domain-specific, deep-level strategies. These strategies allow a student to construct or critically evaluate new information in ways that are accepted within the given discipline. Within the context of causal-historical reasoning, important strategies to master are, among others, (a) to look for multiple causes; (b) to construct complex—as opposed to simple linear—causal models; (c) to analyze causes along multiple dimensions such as time, content, role; and (d) to analyze individuals' motives and actions in the context of the broader political, economic, cultural, and social context of the time (Chapman, 2003; Coffin, 2004; Halldén, 1997; Seixas & Morton, 2013). In addition to these strategies, students need to develop their knowledge of the second-order concepts, which historians use to construct causal narratives about the past (e.g., categorizing causes requires concepts such as direct, indirect, long term, short term, trigger, catalyst, precondition; contextualizing motives and actions requires concepts such as cultural, political). These second-order concepts give students the vocabulary to verbalize their causal reasoning and, more important, provide them with the conceptual apparatus to reason causally in history (i.e., to engage in deep-level strategies; van Drie & van Boxtel, 2008; VanSledright & Limón, 2006; Woodcock, 2005).

Another aspect that should be addressed in a learning environment aiming at fostering expertise are students' beliefs about the "complexity, sophistication and uncertainty of knowledge" (Alexander, 2005, p. 38). Alexander (2005) stated that students with more nuanced epistemological beliefs "tend to be higher academic achievers, report more strategic processing, and are more persistent in the face of difficulty" (p. 38). VanSledright and Limón (2006) suggested that epistemological beliefs and historical understanding are linked and that teaching historical reasoning involves influencing epistemological beliefs.

Within the field history education, epistemological beliefs have often been conceptualized in three "stances," *copier*, *subjectivist*, and *criterialist*. This stage model is embedded in more general theories about epistemological beliefs (see King & Kitchener, 2002; Kuhn & Weinstock, 2002; Maggioni, Alexander, & VanSledright, 2004; Maggioni, VanSledright, & Alexander, 2009). Operationalized for causal reasoning, students with a copier stance would believe that historical explanations should be a "copy" of the past and that inconclusive or contradictory evidence makes writing history impossible. In this stance, little value is placed on methodology because explanations are either correct or wrong. Students with a subjectivist stance accept the fact that historical explanations are interpretations but lack an understanding and appreciation of the (academic) criteria to judge these interpretations. Often this stance leads to the belief that history is merely a matter of opinion. Only with a criterialist stance do students understand and appreciate both the constructed nature of historical explanations as well as the academic criteria for evaluating these causal statements (cf. Lee & Shemilt, 2009). Based on the theories of Alexander (2005) and VanSledright and Limón (2006), a positive relationship is expected between students' epistemological beliefs, their conceptual and strategy knowledge, and the quality of their historical reasoning.

The MDL emphasizes that students in the early phases of expertise development often rely on the use of generic, surface-level strategies (e.g., rereading, summarizing) when confronted with a problem in a specific domain, whereas experts tend to engage in domain specific, deep-level strategies (e.g., using second-order concepts to categorize causes and embedding the analysis in a broader political, economic, cultural, or social context). Furthermore, the MDL links nuanced epistemological beliefs to higher levels of strategic processing. Therefore, developing expertise in causal-historical reasoning is defined in this study as the acquisition of deep-level strategies and second-order concepts while simultaneously stimulating development of more nuanced ideas on the nature of historical knowledge and the criteria for evaluating and constructing historical explanations.

The Role of Individual and Situational Interest

The MDL conceptualizes interest as an important precondition for developing and engaging students in effortful domain-specific, deep-level strategies. The model differentiates between two types of interest: *situational interest* and *individual interest*. Individual interest can be defined as a relatively stable learner characteristic expected to gradually increase as a student gains more knowledge of the domain and the specific strategies and questions involved. As expertise develops, it becomes easier for a learner to connect new information to the broader domain and to prior knowledge and interests, thus, intrinsic motivation increases. In contrast, learners in the early phases of expertise rely on the teacher and the characteristics of the learning environment in order to increase their situational interest and to connect a new topic to the broader domain as well as to their prior knowledge and interests. Situational interest consists of valuing the relevance of what is to be learned and enjoying the learning activities (Alexander, 2003).

The Role of Explicit Teaching

In a previous literature review and in an experimental pilot study, several basic principles of a learning environment intended to foster causal-historical reasoning were defined (i.e., designing open-ended tasks, allowing for social interaction, and raising situational interest; Alexander, 2005; Collins, Brown, & Holum, 1991; Stoel et al., 2015). However, the MDL also maintains that deep-level strategies "cannot be expected to develop naturally but must be cultivated" (Alexander, 2005, p. 40).

This statement is corroborated by several studies in history education that focused on explicit teaching strategies related to analyzing sources—sometimes in combination with writing strategies (De La Paz, 2005; De La Paz & Felton, 2010; Nokes et al., 2007; Reisman, 2012). In these studies, positive effects were found on the quality and length of students' essays, the use of strategies, and general historical thinking. However, these studies all focused on reasoning with historical sources. No previous studies have focused on causal-historical reasoning. Furthermore, these studies had a quasi-experimental design and limited explicit teaching to instructing strategies.

This study adds to the current research by its randomized-controlled design and by focusing on causal-historical reasoning. The study also expands explicit teaching to include causal-historical strategies and second-order concepts as well as epistemological beliefs. In line with the MDL, effects are analyzed not only by measuring a complex causal historical skill, in the form of an essay task, but also by assessing the underlying aspects of causal-historical reasoning: knowledge of causal strategies, second-order concepts, and epistemological ideas. Including students' epistemological beliefs as a dependent variable is also advocated by Reisman (2012). Because of the centrality of first-order knowledge and individual interest in the MDL, these aspects are also measured.

To summarize, we designed this study to investigate the differences between a condition in which students work together on an open-ended explanatory task, while being explicitly taught about the concepts and strategies involved in causal reasoning and while reflecting on the epistemological aspects of their explanations (*explicit condition*), and a condition in which students work together on the same task but without this explicit teaching of strategies, second-order concepts, and epistemological underpinnings (*implicit condition*).

Research Question

Our central research question is as follows: What is the effect of explicit teaching on second-order concepts, causal reasoning strategies, and epistemological underpinnings (in the context of a collaborative explanatory task) on 11th grade students' (a) second-order and strategy knowledge, (b) their epistemological beliefs and (c) their ability to construct a causal explanation, compared with a control group working on a similar task without explicit attention to causal strategies, concepts, and epistemological beliefs? In addition, the effects of the teaching condition (explicit vs. implicit) on students' first-order knowledge and individual interest was compared.

The study was designed as a pretest–posttest randomized controlled experiment. In our analysis, we not only investigated the effects of explicit teaching on students' knowledge, beliefs, and skills, but we also explored the relationships between these different constructs at the pretest and posttest.

Hypotheses

Based on the theoretical framework, we hypothesized that the explicit teaching of causal reasoning strategies and second-order concepts and epistemological reflection—embedded in an open-ended, explanatory task (explicit condition)—would be an effective learning environment for fostering causal-historical reasoning, compared with a control condition that worked on the same task but without the explicit teaching (implicit condition). While controlling for differences in pretest scores and situational interest, we formed the following hypotheses regarding what would occur at posttest:

Hypothesis 1: In the explicit condition, knowledge of causal reasoning strategies and second-order concepts will be significantly higher, compared with the implicit condition.

Hypothesis 2: A significantly more nuanced epistemological stance (indicated by lesser agreement with subjectivist items and more in accord with criterialist items) will have developed in the explicit condition, compared with the implicit condition.

Hypothesis 3: The ability to construct a causal-historical explanation will be significantly higher in the explicit condition, compared with the implicit condition.

Hypothesis 4: Historical first-order knowledge will not differ between conditions

Hypothesis 5: Individual interest will not differ between conditions.

Method

Participants

In total, 104 eleventh grade preuniversity students from four history classes and two teachers participated in the experiment. The average age of the students was 16.8 years (minimum 16, maximum 19). In the Netherlands, preuniversity education (VWO) is the highest educational track in secondary education. Approximately 20% of the secondary school students are enrolled in this 6-year program (Grades 7 to 12). A preuniversity diploma allows admission to university. The participating school is a public school for higher general secondary and preuniversity education. The school has 1,700 students and is situated in a relatively prosperous, suburban community near Amsterdam—average income is 15% above national average (Central Bureau for Statistics, 2015). The lesson table is comprised of single or double 45-min units. History is a mandatory subject in two of the four predefined profiles from which students choose after the 9th grade. At this school, students receive three history lessons a week. World War I has previously been studied in Grade 9. Our lesson-unit marked the beginning of a module on the "Time of Two World Wars," one of the era's in the framework of orientation knowledge.

Within each of the four classes, students were randomly assigned to a condition, creating four experimental and four control subgroups (see Table 1). The subgroups could not be mixed across classes due to different timetables. Because we wanted students to work in triads, we made minor adjustments to the sample size in the subgroups to ensure that the number of students in the treatment subgroups was divisible by three. (When necessary, we added one or two students per class to the experimental subgroup—which explains the different sample size of the two conditions.) The subgroups from each class were inspected on (a) gender distribution and (b) average achievement, based on stu-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 1
Sampling Design

Class (class teacher) and condition	External teacher	*n* (students)
Class 1 (Teacher A)		
Explicit	1	15
Implicit	2	12
Class 2 (Teacher A)		
Explicit	2	12
Implicit	1	11
Class 3 (Teacher B)		
Explicit	1	15
Implicit	2	13
Class 4 (Teacher B)		
Explicit	2	15
Implicit	1	11

dents' history grades during the school year. This led to some minor exchanges between the subgroups per class. Subsequently, 19 triads were created in the four experimental subgroups. In the control condition, students worked in triads as well, but we allowed for one or two dyads to exist in each of the four subgroups. In total 13 triads and 4 dyads were created in the control condition. All triads were composed of a high scoring, a low scoring, and an average student (based on students' history grades during the school year) to prevent the confounding of outcomes on dependent variables with differences between triads.

One student opted not to participate before the experiment ($n = 1$). After the experiment, we excluded eight students who missed more than one intervention lesson ($n_{\text{exp}} = 3$; $n_{\text{imp}} = 5$). This resulted in a final sample size of 95 students. In our analysis, the explicit condition consisted of 53 students (28 male and 25 female) and the implicit condition consisted of 42 students (21 male and 21 female).

During the lesson-unit, the subgroups from each class worked in two separate classrooms and two external teachers instructed the groups. The first external teacher holds a degree in history and teaching and has taught history at the secondary level for 8 years. The second external teacher holds a PhD in history and a degree in teaching and has taught history at the secondary level for 2 years. We choose two external teachers to teach the intervention lessons—instead of the regular class teachers—in order to prevent the confounding of outcomes with potential teacher effects. To prevent differences between the external teachers (possibly leading to confound learning outcomes), the external teachers switched conditions between classes so that each taught two experimental and two control subgroups. Table 1 presents an overview of the sampling design.

Lesson-Unit and Procedures

The lesson-unit focused on explaining the outbreak of World War I. Before the start of the experiment, the students were tested on (a) knowledge of second-order concepts and causal-reasoning strategies, (b) epistemological beliefs (subjectivist and criterialist), (c) first-order knowledge, and (d) individual interest in history (see Table 2; Pretest I). Subsequently, all students received two preparatory lessons, both lasting 45 min, that focused on developing students' historical knowledge about events, countries, developments, and phenomena in Europe in the run-up to World War I (see Table 2; Lesson 1/2). With this preparation, we sought to provide students with enough first-order knowledge to reduce the confounding of students' reasoning abilities in the pretest essay-writing task with a low level of knowledge about the topic.

After the preparatory lessons, students wrote a history essay (pretest) using several sources and their prior knowledge to explain why Germany became involved in World War I (see Table 2; Pretest II). Subsequently, the actual intervention took place (see Table 2; Lessons 3, 4, & 5). The subgroups (explicit and implicit) worked for three consecutive lessons in separate classrooms on an open-ended collaborative task. At the end of the third lesson, students filled out a short questionnaire, designed to measure their situational interest in the previous lessons. After the experiment, students took one lesson to rewrite their pretest essays (see Table 2; Posttest I). Finally, students retook the tests that measured knowledge, epistemological beliefs, and individual interest (see Table 2; Posttest II).

The experiment was conducted during 2 weeks in March 2014. Pretest I was taken 4 weeks before the start of the intervention. The preparatory lessons and Pretest II took place in the first week. The intervention lesson, as well as Posttest I and II, took place in the second week. Table 2 presents a schematic summary of the design and the measurement instruments.

Designing the Implicit and Explicit Condition

Hereunder both lesson-units are described. Because the basic model and design principals were the same for both groups, we start with the commonalities and then move on to the elaboration

Table 2
Procedure

Lesson	Phase	Measurement
Before preparatory lessons	Pretest I	Knowledge: conceptual & strategic, first order Beliefs: epistemology Interest: individual
Lesson 1 and 2	Preparatory topic lessons	
Before intervention lessons	Pretest II	Skills: essay task
Lesson 3, 4, and 5	Intervention lessons (subgroups work in separate classrooms)	Interest: situational
After intervention lessons	Posttest I	Skills: rewriting essay task
	Posttest II	Knowledge: conceptual & strategic, first order Beliefs: epistemology Interest: individual

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

of the lesson-units in the two conditions. Table 3 provides a summary of both lesson-units. A detailed elaboration of the lesson goals can be found in Appendix A.

Commonalities. Both lesson-units were designed from a constructivist perspective on learning and followed characteristics of a problem based learning environment (Savery & Duffy, 1995) and pedagogical principles described in the MDL (Alexander, 2005) and the model of cognitive apprenticeship (Collins et al., 1991). Based on Merrill's (2002) review of the common characteristics of different instructional theories, we discerned four phases in the learning environment: a preparatory phase, an instructional phase, a phase of application, and a phase of integration.

In both lesson-units, students worked on an authentic, open-ended task, based on an exemplary question in the domain of history: "How can we explain the outbreak of the First World War?" Group work (triads) and whole class discussion was used to stimulate interaction and argumentation. An important characteristic of all learning activities was the aim to make students' thinking visible, thereby allowing the teacher to provide students with scaffolding, coaching, and constructive feedback.

In the preparatory phase, effort was made to raise situational interest and to allow students to understand the relevance of the topic by triggering prior knowledge and interest and connecting the topic to the broader domain. In the instructional phase, the key question and the task were explained and the goals of the lessons were explicated. During the phase of application, students in both conditions worked in triads to coconstruct an explanation. Students worked on card-sorting tasks and graphical representation to select, organize, and connect causes and to construct their explanations. Research has shown that graphical representations allow students to externalize and explicate their thinking, mediate their analysis and discussion, and enhance their historical thinking (Prangsma, van Boxtel, & Kanselaar, 2008; van Drie & van Boxtel, 2003; van Drie, van Boxtel, Jaspers, & Kanselaar, 2005). In the phase of integration, students presented their conclusions and discussed them in a whole-class setting. This phase intended to broaden and deepen students' thinking by allowing them to compare their explanations and to reflect on similarities and differences between their products. Furthermore, these whole-class discussions allowed the teacher to uncover and address possible misconceptions. Previous studies in history education have underscored the effectiveness of whole-class discussions (Havekes, 2015; van Drie & van Boxtel, 2011). A detailed literature study on

Table 3
Summary of the Explicit and Implicit Teaching Environment in the Lesson-Unit "Explaining the First World War"

Learning phase	General design principles	Teacher and learning activities *explicit* condition	Teacher and learning activities *implicit* condition
Preparatory phase(s) Students *connect* the topic to the broader domain and their prior knowledge	Raising situational interest	Teacher fosters a sense of rooted relevance by presenting a funny nonhistorical analogy intended to trigger causal reasoning (Chapman, 2003) Students read, listen, generate ideas, answer	Teacher fosters a sense of rooted relevance by showing a short video-clip about the murder on Franz Ferdinand and asking whether one murder could be responsible for a war to start? Students watch, generate ideas, answer
Instructional phase(s) Students *understand* the task and the goals of the task	Working on (an) open (sub)task(s) Social interaction Making teacher & student thinking visible	Teacher uses the nonhistorical analogy to explicate and model thinking about concepts and strategies connected to historical causation Students read, listen, generate ideas, answer	Teacher uses the video clip to connect students' first-order knowledge with the key question; introduces task and goals Students watch, listen, generate ideas, answer
Phase(s) of application Students *construct* a causal historical explanation		Teacher coaches, scaffolds, provides feedback (focus on strategies and second-order concepts) Students work in triads to coconstruct an explanation Students engage in card-sorting tasks and concept-mapping to select, categorize and connect causes Students develop a vocabulary for categorizing and connecting causes by applying a wordlist (Woodcock, 2005) Students write a mini-essay (supported by the wordlist)	Teacher coaches, scaffolds, provides feedback (focus on first-order knowledge) Students work in triads to coconstruct an explanation Students engage in card-sorting task to select causes and organize causes in a graphical organizer Students synthesize their analysis in a written synopsis that connects causes and answers the question Students prepare a presentation to communicate their analysis
Phase(s) of integration Students *discuss and compare* their explanations		Teacher asks (epistemological) questions, addresses misconceptions, provides feedback; related to causal strategies and concepts Students present concept-maps and essays, and engage in whole-class discussion Students compare and reflect on similarities and differences in products, the concepts used and the strategies engaged Students reflect on epistemological questions	Teacher asks questions, addresses misconceptions, provides feedback; related to first-order knowledge Students deliver presentations and engage in whole-class discussion Students compare and reflect on similarities and differences of products (focus on first-order knowledge) Students exchange perspectives and arguments

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

the principles in both lesson-units can be found in Stoel et al. (2015).

The explicit teaching environment. The major difference between the lesson-units was the explicit attention to strategies, second-order concepts, and epistemological questions related to historical causation. This attention was operationalized for each phase of the learning environment and consisted of teacher-led activities (instruction and scaffolding), student-led activities (group work), and shared activities (whole-class reflection).

In the instructional phases, the teacher explicated relevant second-order concepts and modeled the targeted strategies connected to historical causation. A nonhistorical analogical story was used to start students' thinking about multicausality, causal categories, and connected second-order concepts and to develop a multilayered model of causal relationships. Furthermore, the teacher modeled and discussed different ways to verbalize causal explanations with various degree of causal "nuance."

In the application phases, students practiced the relevant strategies and concepts by working together on causal (sub)tasks. Card-sorting tasks (involving events, people, countries, developments, and phenomena in Europe prior to World War I) and concept maps were used to stimulate analysis of different (types of) causes, to draw causal connections, and to reflect on appropriate causal models and the roles these causes played in their explanations. Students were equipped with a wordlist to scaffold their verbalization of causal connections (Woodcock, 2005). In the final lesson, the groups constructed a miniessay based on their analysis. In this phase, the teacher's role was mainly to scaffold and coach.

In the phases of integration, whole-class discussion was deployed to verbalize, broaden, and deepen students' understanding. Guided by the teacher, students reflected on their categorization of causes, their causal model and connections (as witnessed in their concept maps), and on their miniessays. Furthermore, whole-class discussion was used to reflect on epistemological questions about the constructed nature of students' explanations, the differences between interpretations, and the criteria for assessing the quality of these explanations.

The implicit teaching environment. Students in the implicit condition worked on the same task as did students in the explicit condition. Also in this condition, raising situational interest and collaborative learning were important design principles. In the lesson unit, we differentiated between the same four instructional phases. The crucial difference between lesson-units was that triads in the implicit condition worked on the whole task and constructed a causal explanation without paying explicit attention to causal strategies, concepts, and epistemological questions. The learning activities in this lesson-unit were designed to balance the amount of analysis and synthesis of first-order knowledge in the explicit condition alongside the alternation of writing, discussing, and visualizing, and the total time students interacted with the historical content. Instruction and constructive feedback in all phases focused on first-order knowledge and on supporting task execution.

In the instructional phase, the teacher activated prior knowledge by showing a short video clip about the murder of Archduke Franz Ferdinand to introduce the key question and, subsequently, to explore students' initial ideas. This phase focused on explicating the first-order knowledge that had been the subject of the two preparatory lessons. Afterward, the open-ended task and the goals of the task were introduced to the students.

In the phases of application, a card-sorting task (involving the same cards as in the explicit condition) and a graphical organizer (an empty presentation format) were used to select and organize events, developments, and phenomena connected to the outbreak of World War I. Based on this scheme, the triads subsequently wrote a synthesis text in which they linked their causes together and answered the key question. This synopsis formed the backbone of the PowerPoint presentation, which students designed in the second lesson, and of their subsequent oral presentation.

The final lesson aimed at integrating students' knowledge; the groups presented their causal explanations, compared each other's work, and gave feedback. This activity aimed to broaden and deepen students' (first-order) knowledge. Guided by the teacher, students reflected on each other's explanations, focusing primarily on first-order knowledge. Implicitly, of course, the learning activities confronted the students with several different interpretations and with causal concepts and connections. However, these issues were not explicitly addressed.

Research Instruments

Because our theoretical framework conceptualized causal-historical reasoning to be underpinned by different types of knowledge (knowledge of second-order concepts and strategies, and epistemological beliefs), we not only measured students' ability to construct a causal-historical explanation in a pre- and posttest but also assessed underlying knowledge and beliefs. At the posttest, students received two open prompts asking them to reflect on their learning gains and to provide a heuristic for future causal analysis in history.

Because the MDL considers first-order knowledge to be an important element in the development of expertise, we measured historical-topic knowledge both before and after the experiment. The MDL also conceptualizes a learning environment—focusing on epistemological questions, deep-level strategies, and connected concepts—to stimulate the development of individual interest. This too was measured both before and after the experiment; however, we did not expect individual interest to increase in only three lessons.

Finally, students' situational interest was measured at the end of the third intervention lesson as a control variable. Based on the MDL, we wanted to make sure that (a) both conditions were successful in arousing students' situational interest and (b) would do so to a comparable extend to control for potential differences in motivational quality of both conditions (because this might confound attributing effects to differences in cognitive approach). In our analyses, situational interest was used as a covariate.

Knowledge of causal reasoning strategies and second-order concepts. We administered a 19-item questionnaire twice (as a pretest and a posttest) to measure students' knowledge of second-order concepts and causal-reasoning strategies. Students had to score items on a six-point Likert scale, ranging from 0 (*strongly disagree*) to 6 (*strongly agree*). The questionnaire was based on literature and expert consultation, and a previous version was used in the pilot study (Stoel et al., 2015). Reliability analysis led to the exclusion of four items that lowered scale reliability. Fifteen items yielded a Cronbach's alpha of .64 ($n = 82$) at the pretest and .68

($n = 89$) at the posttest. These items were used in the analysis. The following are examples of items: "in an historical explanation it is important to differentiate between different roles causes might have played"; "an historical explanation is usually constructed as a chain of causes and consequences" [recoded]; and "in an historical explanation you must also explain how causes interact."

Epistemological beliefs. Students' epistemological beliefs were measured twice (in a pretest and a posttest). We used a translated version of the Beliefs About History Questionnaire (BHQ) developed by Maggioni (2010). To explore the translated questionnaire, two datasets were collected prior to the experiment—one in another research project ($N = 140$) and one in the pilot study ($N = 74$). In both datasets, the copier scale yielded unacceptable low reliability. Therefore, we decided to exclude this scale from the current questionnaire. Furthermore, two criterialist items were excluded from the questionnaire because the translated versions did not load with the other criterialist items in both datasets. This could be the result of shifted meaning due to translation, or it could be connected to differences in historical culture between the United States and the Netherlands. The questionnaire used in this study, therefore, includes all subjectivist items from the original BHQ (9-items) and six out of eight items from the criterialist scale. In the results, we report students' (a) subjectivist epistemological beliefs and (b) their criterialist epistemological beliefs as separate dependent variables.

In the questionnaire students had to score items on a six-point Likert scale, ranging from 0 (strongly disagree) to 6 (strongly agree). We calculated scale reliability for the two remaining scales. The Cronbach's alpha for the subjectivist scale was .77 ($n = 90$) at pretest and .85 ($n = 90$) at posttest. Items in this scale related to the supposed subjective nature of historical knowledge. For example, "since there is no way to know what really happened in the past, students can believe whatever story they choose" and "good students know that history is basically a matter of opinion." The six items of the criterialist scale yielded a Cronbach's alpha of .65 ($n = 88$) at pretest and .65 ($n = 92$) at posttest. These items were all related to methodological criteria for constructing and judging historical interpretations. Example items included, "comparing sources and understanding author perspective are essential components of doing history" and "knowledge of the historical method is fundamental for historians and students alike."

Posttest open questions. During the posttest, students were also given two open prompts intended to explore their learning about causal-reasoning strategies, second-order concepts, epistemological reflections, and broader cognitive and motivational learning gains in a more reflective manner.

Heuristic prompt. This prompt asked students to provide a roadmap for when they would engage in future causal-historical inquiry (as an example, the collapse of the Soviet Union was mentioned). We developed a rubric consisting of one criterion (domain specificity) with two levels to code the heuristics. On the first level, students reported no heuristic, a fully generic one, or only very shallow references were made to causal concepts (i.e., only words such as *causes*, *consequences*, or *connections* were used). On the second level, students' answers at least centered on one dimension of causal reasoning (e.g., focusing on historical content, listing second-order concepts or referring to causal-reasoning strategies). Subsequently, we used the rubric to blindly code all the data. After explaining the rubric, the third author scored a subset of 31 random answers (29%). Interrater reliability was $\kappa = .80$.

Report of learning gains. This open prompt asked students to reflect on what they had primarily learned in the previous lessons. Three domain-specific codes were generated in accordance with the lesson goals: "looking for multiple causes" for answers that mentioned this basic causal-reasoning strategy, "drawing causal connections or using causal categories" for answers that referred to more sophisticated second-order concepts and causal-reasoning strategies, and "epistemological reflection" for answers focusing on the interpretative nature of causal analysis and the possibility of multiple valid answers. Furthermore, we defined one generic code ("additional") and gave this code to answers that, among others, referred to motivational aspects, the effectiveness of working with historical-inquiry tasks in general, historical content, and general study skills. Each category could only be scored once per student; although, an answer could be coded in multiple categories. An independent second rater scored a subset of 31 random answers (29%). Beforehand, the codebook was explained, the rater practiced on a subset and the differences were discussed. Interrater reliability for the four categories was multiple causes ($\kappa = .81$), causal connections and categories ($\kappa = .77$), epistemology ($\kappa = .87$), and additional ($\kappa = .67$).

Causal reasoning ability. Research has shown that reading multiple sources and writing argumentative accounts is an effective approach to elicit historical reasoning (Rouet, Britt, Mason, & Perfetti, 1996; van Drie, van Boxtel, & van der Linden, 2006; Wiley & Voss, 1999). Therefore, an explanatory-writing task was designed to measure students' ability to apply their first-order knowledge and their knowledge of causal concepts and reasoning strategies. The writing task was tested in a previous pilot study (Stoel et al., 2015). In the current study, students were asked to rewrite their pretest essay at the posttest. We expected that rewriting would heighten the sensibility of the instrument by lowering the complexity of the task at posttest. (Students were not required to read new sources or create a new written explanation but could revise their essay based on the newly gained knowledge.) Based on Rijlaarsdam, Couzijn, and van den Bergh (2004), who define rewriting as a goal-directed activity, we expected that this would allow students to more easily apply what they had learned in the intermittent lessons.

At pretest, students were asked to construct a 300-word explanation on why Germany became involved in World War I. Students were provided with a set of nine primary historical sources. Also, they were given a factsheet that listed the events, people, developments, countries, phenomena, and dates of the prewar period that had been discussed during the preparatory lessons. The factsheet was designed as a table to prevent a narrative (causal) template. The set of sources was constructed so that students could argue that Germany had provoked the war or had been "pulled in" by actions of other countries. Arguments of different types could be drawn from the sources (e.g., triggers, catalysts, and preconditions; direct and indirect causes; economic, political, and sociocultural causes; and personal agency).

A rubric consisting of four criteria was developed to analyze students' written work (see Appendix B). On each criterion, students received a score ranging between 0 and 2 points.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Students were scored on (a) the number of structural causes presented in their writing, (b) the number of structural causes substantiated by specific historical events (triggers or catalysts), (c) the explanatory model (linear, abstract list, abstract integrated), and (d) the use of nuanced second-order language and causal connections. All essays were blindly coded by two raters. For the first 31 essays, interrater correlation was not yet satisfactory; therefore, a final score for these essays was calculated based on agreement after discussion. For the remaining essays, interrater correlation was Pearson's $r = .71$ ($n = 156$). The mean score between the two raters was used for further analyses. Reliability of the four categories yielded a Cronbach's alpha of .70 ($n = 94$) at pretest and .66 ($n = 92$) at posttest.

First-order knowledge. A 17-item knowledge test was conducted twice (as a pretest and a posttest) to measure historical first-order knowledge. The test was slightly adapted from the test used in the pilot study (Stoel et al., 2015). The items were divided into four categories and measured students' (a) knowledge of prewar alliances (1 item), (b) ability to connect historical concepts to countries (12 items), (c) chronological knowledge (1 item), and (d) ability to generate concrete historical examples of abstract historical concepts (3 items). Two raters scored the three open items in all tests. Interrater reliability on these items was Pearson's $r = .85$ ($n = 182$). In the analysis, we used the scores of the second rater who had not been involved in the implementation of the experiment. A mean score was calculated for each category separately on a scale ranging from 0 to 1, and subsequently we calculated a pooled mean.

Individual interest. An 8-item questionnaire was conducted twice (in a pretest and a posttest) to measure students' individual interest in history. Students scored items on a six-point Likert scale, ranging from 0 (*strongly disagree*) to 6 (*strongly agree*). The individual interest questionnaire was based on an adaptation of the task-value scales from the Motivated Strategies for Learning Questionnaire (MSLQ) developed for mathematics education (Linnenbrink-Garcia et al., 2010; Pintrich, Smith, García, & McKeachie, 1993). Sample items included "I enjoy the school subject of history," "I can use historical knowledge well outside school," and "it is important for me to be able to think historically." Cronbach's alpha for the questionnaire was .89 ($n = 91$) for the pretest and .89 ($n = 93$) for the posttest.

Situational interest. A 12-item questionnaire was conducted at the end of the final intervention lesson to measure students' situational interest. Situational interest was measured to ascertain that differences in learning gains would not be attributable to differences in motivational quality of the lesson-units in both conditions. Furthermore, situational interest was used as a covariate in our model to prevent the confounding of learning outcomes with difference in interest in the learning environment. Students had to score items on a six-point Likert scale, ranging from 0 (*strongly disagree*) to 6 (*strongly agree*). The questionnaire was based on a validated questionnaire for mathematics education (Linnenbrink-Garcia et al., 2010). Sample items included, "what I have learned in these lessons is useful for me to know," "I liked what we learned in these lessons," and "these lessons were so exciting that I could easily maintain my attention." Cronbach's alpha for the questionnaire was .91 ($n = 91$).

Treatment Fidelity, Missing Data, Homogeneity of Triads, Statistical Procedure, Effect Sizes, and Homogeneity of Conditions

Treatment fidelity. The lesson-units were delivered by two external teachers who were both experienced teachers and historians with a firm grasp of the content knowledge, the strategies, and second-order concepts related to causal reasoning as well as the epistemological questions involved in historical explanations. For all lessons, detailed plans were designed—which established learning goals, teacher and student activities, and precise scripts for instruction and whole-class discussions. Both teachers engaged in a 2-day training session in which all lessons were meticulously discussed and prepared, leaving room for adaptation. During these meetings, a shared meaning and approach was negotiated and definitive lesson plans were determined. All lessons were discussed both before and after the execution. No important deviations from the plans were reported; the lessons could be executed as planned. To prevent difference between teachers and to avoid confounding learning outcomes, both external teachers switched conditions between classes—each person taught two explicit and two implicit subgroups. Furthermore, all student products were collected and compared on thoroughness. This comparison showed that students in both conditions were able to complete the tasks and that the quality of their work was satisfactory. Finally, students' engagement and interest in the lessons was measured by the situational interest questionnaire. On a six-point scale, situational interest was rated positive by students in both conditions ($M_{exp} = 4.08$, $SD = .69$, $n = 53$ and $M_{contr} = 3.91$, $SD = .64$, $n = 42$) and did not differ significantly between conditions.

Missing data. Missing value analysis showed that, on average 1.5% of the values were missing. Missing value count on the individual variables ranged between 0 and 5 ($M = 1.5$, $Mdn = 1$, mode $= 1$). The number of absentees varied between 0 and 3 at the different test moments ($M = 1.5$). Full information maximum likelihood approach (FIML; method = FCS; $n_{imputed} = 20$) was used in order to include all students in the analysis (Little, Jorgensen, Lang, & Moore, 2014).

Homogeneity of triads. All triads within the subgroups were composed of a high scoring, a low scoring, and an average student (based on students' history grades during the school year) to prevent the confounding of outcomes on dependent variables with differences between triads. After the experiment, all data was imported into R, and we ran a multilevel model to explore group effects on outcome variables. All of the intraclass correlations were 1% or less, showing that blocking on student achievement worked well and that the groups were indeed homogeneous. Therefore, there was no further need to run a multilevel analysis; a regular regression model would yield similar results.

Statistical procedure. In the analyses, univariate GLM's were used to analyze the mean scores on the six dependent variables (knowledge of causal concepts and strategy, subjectivist epistemological ideas, criterialist epistemological ideas, essay quality, first-order knowledge, and individual interest). The choice for univariate analysis was considered acceptable because of the conceptual distinctions between the measured constructs. This distinction is also reflected in the separate univariate hypotheses of our study. Moreover, the FIML approach for handling missing data does not allow for multivariate analysis. Pooled outcomes on the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

imputed dataset ($n_{\text{imputed}} = 20$) can only be calculated as univariate regression. In the analysis, we controlled for students' pretest scores and for the differences in their situational interest.

Selection and calculation of effect size. Squared semipartial correlations were used as measures of effect size. In regression analysis, squared semipartial effect sizes yield the proportion of variability uniquely predicted by the independent variable when the other independent variables have been controlled (Fritz, Morris, & Richler, 2012). Semipartial square effect size is interpreted as follows: $rs^2 > .01$: small effect; $rs^2 > .09$: medium effect; $rs^2 > .25$: large effect (Cohen, 1988).

Homogeneity of conditions at pretest. To check the homogeneity of the conditions at pretest, a regression analysis was performed on all six dependent variables at pretest means, and we entered condition as a covariate. No significant univariate differences were found at the pretest (see Table 4). Subsequently, the mean scores on students' situational interest in the task were compared to check for potential motivational differences between the lesson-units in both conditions. Differences in motivational quality between conditions might confound the attribution of effects to a difference in cognitive approaches. As reported above, situational interest in both groups was positive and did not differ significantly between the conditions. The checks increased confidence in the homogeneity of both conditions at pretest and the comparability of the motivational qualities of both learning environments.

Results

Effects of Explicit Teaching on Students' Second-Order and Strategy Knowledge and Epistemological Beliefs

Knowledge of causal reasoning strategies and second-order concepts. Two exploratory paired-samples t tests on the nonimputed dataset showed that students' knowledge of causal strategies and second-order concepts increased significantly in the explicit condition, $t(51) = 4.20$, $p < .000$, but no significant change was found in the implicit condition. As we expected, students in the explicit condition scored significantly higher at the posttest than in the implicit condition, while controlling for differences in students' situational interest and pretest scores, $t(91) = 3.33$, $p = .001$, $sr^2 = .09$, post hoc power = .97.

Epistemological beliefs. Two exploratory paired-samples t tests on the nonimputed dataset showed that the level of agreement with subjectivist epistemological ideas increased significantly in the explicit condition, $t(52) = 2.20$, $p = .032$, but no significant change was found in the implicit condition. Students' subjectivist epistemological ideas, controlling for pretest scores and situational interest, differed significantly between conditions at posttest, $t(91) = 2.21$, $p = .027$, $sr^2 = .04$, post hoc power = .74. Students in the explicit condition reported a higher agreement with subjectivist beliefs at posttest compared with students in the implicit condition. This result contradicted the hypothesis that developing more nuanced ideas would lead students to become more critical toward subjectivist beliefs. This expectation was based on the theoretical framework of Maggioni et al. (2009).

Two exploratory paired-samples t tests on the nonimputed dataset showed that the value students attributed to criterialist epistemological ideas decreased significantly in the implicit condition, $t(39) = -2.08$, $p = .044$, whereas a nonsignificant increase was found in the explicit condition $t(52) = 1.85$, $p = .070$. Controlling for pretest scores and situational interest, a significant posttest difference was found for teaching condition on students' criterialist epistemological ideas, $t(91) = 2.60$, $p = .009$, $sr^2 = .04$, post hoc power = .87. Students in the explicit condition reported a higher value on items related to disciplinary criteria for generating historical knowledge compared with students in the implicit condition. This result was in line with our hypothesis.

Open prompts. The open prompts were analyzed for references to causal strategies and concepts and references to epistemological ideas. Because of the qualitative nature of the answers, no covariables were included in the analysis. Chi-square tests of measuring goodness of fit were performed to investigate differences between the explicit and implicit condition. Table 5 lists the descriptive statistics.

Table 4
Pretest and Posttest Means and Standard Deviations, Pooled Across Imputations, and Posttest Differences

Dependent variable	Dependent variable and condition	n	Pretest M	Pretest SD	Posttest M	Posttest SD	p^d	sr^{2d}
Knowledge of causal reasoning strategies[a]	Exp	53	4.16	.38	4.37	.41	.001	.09
	Contr	42	4.05	.32	4.08	.31		
Subjectivist beliefs (epistemology)[a]	Exp	53	3.22	.62	3.38	.69	.027	.04
	Contr	42	3.11	.63	3.03	.69		
Criterialist beliefs (epistemology)[a]	Exp	53	4.39	.49	4.51	.44	.009	.04
	Contr	42	4.41	.49	4.27	.50		
Essay[b]	Exp	53	.89	.39	1.24	.34		
	Contr	42	.76	.36	1.12	.39		
First-order knowledge[c]	Exp	53	.44	.17	.68	.15		
	Contr	42	.40	.17	.68	.13		
Individual interest[a]	Exp	53	3.79	.86	4.10	.67	.008	.02
	Contr	42	3.63	.80	3.72	.79		

Note. Exp = experimental condition; Contr = control condition.
[a] Min = 1, max = 6. [b] Min = 0, max = 2. [c] Min = 0, max = 1. [d] Posttest differences and effect sizes were calculated controlling for pretest scores and situational interest.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Heuristic prompt. The effect of explicit teaching on the domain specificity of the heuristic prompt was significant, $\chi^2(1, N = 95) = 13.20$, $p < .000$. In the explicit Condition 58% of the students referred to at least one dimension of causal-historical reasoning, compared with 21% of the students in the implicit condition.

Report of learning gains. Students in the explicit condition reported significantly more often to have learned about causal concepts or strategies, $\chi^2(1, N = 95) = 5.50$, $p = .019$. Examples of student answers in this category included, "it is important to draw relations between causes"; or "how you can classify historical facts, or events in a narrative, in triggers, catalysts, background causes." Furthermore, students in the explicit condition reported significantly more often to have acquired epistemological understandings, $\chi^2(1, N = 95) = 8.14$, $p = .004$. For instance, students report to have learned that "not everybody will consider the same causes as causes," that "history is mainly about explanation," or that "history can be viewed from multiple perspectives." In contrast, students in the implicit condition reported significantly more "off topic," general, and diffuse learning gains, $\chi^2(1, N = 95) = 18.63$, $p < .000$. In this category, answers were coded that referred to first-order content (e.g., "I remembered almost nothing about the First World War, but have learned a lot about it"), to motivational aspects (e.g., "in history, thorough research can quickly lead to forming an opinion. This is nice, because most people think differently about this"; or "it can become a more interesting and more fun subject if lessons are taught like this"), to the effectiveness of historical inquiry tasks (e.g., "doing historical research is far more complicated and complex than many people think"), or to general study skills (e.g., "you learn a lot by thorough reading and delving into it"). Finally, both groups reported approximately equally as often to have learned that history always involves multiple causes.

Causal Reasoning Ability

Two exploratory paired-samples *t* tests on the nonimputed dataset yielded a significant increase in essay quality in both conditions, $t_{exp}(50) = 7.23$, $p < .000$; $t_{imp}(40) = 6.77$, $p < .000$.

Table 5
Means, Standard Deviations of Open Prompts, and Differences Between Conditions

Prompt and code	Exp ($n = 53$) *M*	Exp ($n = 53$) *SD*	Contr ($n = 42$) *M*	Contr ($n = 42$) *SD*	*p*
Heuristic prompt	.58	.50	.21	.42	<.000
Learning gains					
Multiple causes	.51	.51	.38	.49	
Causal concepts and strategies	.47	.50	.24	.43	.019
Epistemology	.23	.42	.02	.15	.004
Additional, of which	.23	.42	.67	.48	<.000
first-order content	6%		21%		
motivational aspects	2%		7%		
effectiveness of open-ended tasks	4%		19%		
general study skills	8%		10%		

Note. Exp = experimental condition; Contr = control condition.

Contrary to our hypothesis, however, controlling for situational interest and pretest scores no significant difference was found between the mean scores in both conditions at posttest, $t(91) = .57$, $p = .571$ (for descriptives, see Table 4). Subsequently, we explored the four underlying criteria of the rubric on which the essay score was based. Controlling for pretest scores and situational interest, a significant difference was found at posttest on one criterion of the rubric, namely, "use of second-order language and causal connectors." Students in the explicit group ($M_{pre} = .70$, $SD = .48$; $M_{post} = 1.30$, $SD = .54$) scored significantly higher than students in the implicit condition ($M_{pre} = .51$, $SD = .50$; $M_{post} = .96$, $SD = .44$), $t(91) = 2.59$, $p = .010$, $sr^2 = .06$, post hoc power = .95.

First-Order Knowledge

Two exploratory paired-samples *t* tests on the nonimputed dataset yielded a significant increase of first-order knowledge in both conditions, $t_{exp}(52) = 11.05$, $p < .000$; $t_{imp}(40) = 12.01$, $p < .000$. Regression analysis on the posttest first-order-knowledge scores, while entering students' pretest scores and situational interest as covariates, revealed no significant effect of teaching condition on first-order knowledge, $t(91) = -.64$, $p = .520$.

Individual Interest

Two exploratory paired-samples *t* tests on the nonimputed dataset showed that students' individual interest increased significantly in the explicit condition, $t(52) = 4.76$, $p < .000$, but we found no significant change in the implicit condition. Although we expected that individual interest would remain stable, the regression revealed a small significant effect of condition on individual interest when we controlled for pretest scores and situational interest, $t(91) = 2.67$, $p = .008$, $sr^2 = .02$, post hoc power = .87. The effect was tempered by controlling for situational interest because the two constructs correlated very strongly (Pearson's $r = .40$ at pretest; Pearson's $r = .60$ at posttest).

Relationships Between Different Constructs Underlying Causal Historical Reasoning

This study was designed on the premise that, besides first-order knowledge, causal historical reasoning is related to a student's conceptual second-order and strategy knowledge and epistemological ideas. Furthermore, a relationship is expected between the cognitive dimensions of historical reasoning and individual interest. Hereunder, the relationships between the dependent variables at different points of measurement (the pretest and the separate posttests for both conditions) will be presented. The correlation tables can be found in Appendix C.

The correlation tables showed a weak to moderate relationship between second-order and strategy knowledge and students criterialist epistemological beliefs both at the pretest (Pearson's $r = .26$, $p = .026$) as well as at the posttests (explicit, Pearson's $r = .35$, $p = .011$; implicit, Pearson's $r = .28$, *ns*). Students who attributed greater value to criteria for generating historical knowledge also scored higher on their knowledge of causal reasoning strategies and concepts.

At the posttest in the explicit condition, a moderate negative relationship was found between knowledge of causal concepts and

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

strategies, and students' subjectivist epistemological ideas (Pearson's $r = -.39$, $p = .004$). Students in the explicit condition that were more critical about regarding history as "just an opinion," scored higher on knowledge of causal-reasoning strategies and concepts.

The pretest showed that students' criterialist epistemological beliefs were moderately related to their individual interest (Pearson's $r = .39$, $p < .000$). At the posttest, the correlation between criterialist epistemological beliefs and individual interest increased in the explicit condition (Pearson's $r = .66$, $p < .000$), which can be considered a strong relationship. Students in the explicit condition, who were in greater agreement with criterialist epistemological ideas, also reported a higher individual interest in history. The strengthening of this relationship appeared to be caused mainly by the increased level of individual interest reported by the explicit condition at the posttest.

Essay quality at the pretest was found to correlate moderately with interest (Pearson's $r = .30$, $p = .003$). Students who reported a higher interest in history also tended to write better essays. At the posttest in the implicit condition, a strong relationship was found between students' essay scores and their first-order knowledge (Pearson's $r = .41$, $p = .006$) and between essay scores and knowledge of causal-reasoning strategies (Pearson's $r = .43$, $p = .009$). No significant relationships with essay quality were found at the posttest in the explicit condition.

The correlation tables showed a strong relationship between first-order knowledge and individual interest at the pretest (Pearson's $r = .43$, $p < .000$). This relationship was also found at the posttest in both conditions. Students who reported a higher value on individual interest also scored higher on the first-order-knowledge tests.

Discussion

The goal of this study was to investigate the effects of explicit teaching on developing students' ability to construct a causal historical explanation. Based on our theoretical framework, it was expected that learning to reason in a domain not only demands open-ended tasks, social interaction, and the stimulation of situational interest but also requires explicit, well-structured instruction and practice (i.e., explicit teaching environment). An important element in our theoretical framework was the "holistic" definition of explicit teaching. Our model asserted that explicit teaching should attend to students' knowledge of strategies and second-order concepts connected to historical causation as well as to epistemological questions involved in constructing historical explanations.

As we expected, knowledge of causal-reasoning strategies and second-order concepts improved in the explicit condition but not in the implicit condition. At the posttest the explicit condition scored significantly higher than students in the implicit condition. Therefore, we concluded that these strategies and concepts are learnable in an explicit teaching environment but that they do not develop spontaneously in the context of an inquiry task. Analysis of the learner reports did provide additional support for this conclusion. Of the students in the explicit condition, 47% of the students mentioned to have learned about "causal connections and categories" or explicitly mentioned second-order concepts, compared with 24% in the implicit group. The same difference in domain specificity was found in students' heuristics. In the explicit Condition 58% of the students made references to at least one dimension of causal-historical reasoning, versus 21% in the implicit group. Remarkably, students in both conditions mentioned approximately as often to have learned that causal analysis in history always involves looking at multiple causes. At least for 11th grade preuniversity students, this learning goal appears to be attainable even without explicit instruction.

Looking at the value students attributed to the epistemological scales (subjectivist beliefs and criterialist beliefs), a significant difference was found at posttest in students' subjectivist beliefs. Contrary to our expectations, students in the explicit condition rated the subjective nature of historical knowledge at posttest higher than students in the implicit condition. Maggioni et al. (2009), however, associated a higher score on the subjectivist scale with a more naïve position because these items present historical knowledge as an opinion. A possible explanation for this development could be that students in the explicit condition had just received three lessons that focused on the construction of causal interpretations and emphasized the interpretative nature of historical knowledge. Therefore, students' scores may primarily reflect a move away from more absolutist ideas about historical knowledge rather than a strengthening of the idea that historical knowledge is "mere personal opinion." This explanation is supported by the fact that students in the explicit condition valued the items belonging to the criterialist scale significantly higher than the students in the implicit condition. This finding was in line with our expectations. However, effect sizes for the differences between conditions regarding epistemological beliefs were small. The results might indicate that development in epistemological beliefs in history is more adequately conceptualized as a movement along two dimensions—certain or uncertain knowledge and weak or strong emphasis on disciplinary criteria—instead of in three distinct stances (cf. Schommer, 1993). More research is needed to explore this question.

Besides the (small) changes found in students' epistemological beliefs, analysis of the reports on learning gains revealed that 23% of the students in the explicit condition referred to epistemological aspects (e.g., reporting to have learned about history as an interpretation), compared with only 2% in the implicit condition. This difference constitutes a large effect. Based on these results, we conclude that in the explicit condition epistemological aspects shifted more into focus and constituted a tangible part of the learning environment even though epistemological beliefs did not strongly change in these three lessons.

Students' criterialist epistemological beliefs held a weak to moderate correlation with their knowledge of causal strategies and concepts at the pre- and the posttest. Students who attributed a higher importance to disciplinary criteria for constructing historical interpretations also scored higher on the knowledge of causal strategies and second-order concepts needed for these accounts. This finding is in line with the relationship between epistemological beliefs and strategic processing predicated by the MDL (Alexander, 2005).

Furthermore, our data revealed that criterialist epistemological beliefs correlated with students' individual interest in history. Students with a greater interest in history also reported a higher appreciation of the disciplinary criteria for constructing historical knowledge. At the pretest, a strong relationship was found, and it

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

even increased to Pearson's $r = .66$ (a very strong relationship) at the posttest in the explicit condition. This means that about 50% of the variance found in these two variables could be explained by their association. Although no causal inferences can be made, a possible explanation for this result might be that attention to epistemological questions may stimulate students' appreciation of history. Anecdotal support for this was found in the learner reports. One student, for example, stated to have learned that "history is mainly about explanation. Also, it can become a more interesting and more fun subject if lessons are taught like this." Future research should shed more light on the exact (causal) relationship between these constructs.

Students in the implicit condition also mentioned to have learned about conducting research (19%), although these answers were less domain specific and more diffuse—a result that was confirmed in their answers to the heuristic prompt. Judging from the self-reports and the level of situational interest, however, the open-ended task was valued by students in this condition as well. We found it interesting that without explicit attention to the disciplinary concepts, strategies and epistemological underpinnings, students did not appear to regard the inquiry as historical inquiry.

The quality of students' essays developed significantly in both conditions, but contrary to our expectations, no clear difference between conditions was found. Applying understanding of concepts and strategies in an explanatory rewriting task appears to be a difficult step for students. This result was found using an assessment rubric focusing on multiple criteria—by scoring multiple causes, substantiation of causes, text structure, and use of second-order language. When zooming in on a central aspect of the lesson-unit, "use of second-order language and causal connectors," a small but significant effect of explicit teaching was found at the posttest. Students in the explicit condition incorporated a richer vocabulary of causal connectors and concepts to differentiate between different types of causes in their essays. It may be that this aspect of causal-historical reasoning is more readily included in an explanatory text but that more profound changes (e.g., changes in the structure of the explanation) require explicit attention to additional demands (e.g., knowledge of the genre) and perhaps prolonged practice. This is supported by studies focusing on explicit teaching of strategies to analyze historical sources in which effects on students' writing were found (see De La Paz, 2005; De La Paz & Felton, 2010). These studies were longer and included strategies for writing historical essays. However, a study by van Drie, Braaksma, and van Boxtel (2015) found positive effects from a relatively short discipline-based writing instruction on student essays, focusing on historical significance. Future research may investigate the effects of explicit instruction of causal concepts, strategies, and epistemological questions in a longitudinal design and combine a focus on historical reasoning with explicit attention to the causal historical genres.

In line with our expectations, first-order knowledge increased significantly in both groups, without differences between the conditions. This indicates that teaching and engaging students in learning activities while focusing on causal reasoning strategies, use of second-order concepts, and epistemological reflection does not negatively influence students' learning about historical topics—even when considerable time and focus are invested in learning about causal skills. This finding is in line with findings from earlier studies (see Nokes et al., 2007; Reisman, 2012). It is possible that because students in the explicit condition engaged in more deep-level strategies, they more thoroughly processed the first-order knowledge. Because the pretest on first-order knowledge was administered before the two preparatory content lessons, first-order learning gains could also have resulted from these lessons.

A stable correlation in the data was found between first-order knowledge and individual interest. This finding is in line with the conceptualization of expertise in the MDL and provides an impetus to design learning-environments that stimulate epistemological reflection and "[aim] for rooted relevance" (Alexander, 2005). Although it had been 2 years since students had studied World War I, the relationship between first-order knowledge and interest was already found at the pretest. This relationship between prior knowledge and interest is in line with a review study on interest, prior knowledge, and learning (Tobias, 1994), as well as with the MDL.

Correlation analysis did not yield clear support for our model that students' performance on a causal writing or rewriting task is underpinned by students' first-order knowledge, knowledge of causal strategies and concepts, and epistemological beliefs. In the pretest, essay quality was only related to interest but in the posttest of the implicit condition, the relationship shifted to first-order knowledge and knowledge of causal strategies and concepts. In the explicit condition, no clear relationships for essay quality were found. A possible explanation for this may be that our rubric focused on specific aspects of causal-historical reasoning and, for instance, did not include use of first-order knowledge. Therefore, the relationship between first-order, second-order, and strategy knowledge, epistemological beliefs, and "deep historical analysis" (Nokes et al., 2007, p. 503) remains a point to be further explored. In the future, large scale research should further explore the relationships between interest, epistemological beliefs, knowledge (first-order, second-order, and strategy) and causal-historical reasoning skills. Structural equation modeling could be an important step in this direction.

Conclusion

Our study shows that, the explicit teaching of strategies and second-order concepts, within the context of an inquiry task, does constitute a prerequisite for developing students' conceptual and strategy knowledge connected to causal-historical reasoning. The causal reasoning questionnaire, the learner reports and heuristics, and students' essays provided evidence for how this knowledge was more effectively developed and shifted more into focus in the explicit-teaching condition. Based on individual interest and students' learner reports, this explicitness also appears to have contributed to the value that students ascribed to the learning environment. An important finding in this study was the importance of explicitly addressing epistemological beliefs in the history classroom. Although students' beliefs did not strongly change during the lessons, many students referred to the epistemological dimensions in their learner reports. A strong relationship was found between the value students attributed to academic criteria for assessing historical interpretations (criterialist epistemological beliefs) and their individual interest. This relationship increased at the posttest belonging to the explicit condition. These results suggest that addressing these questions may stimulate individual interest. These findings are in line with the MDL (Alexander, 2005). We also found that applying the knowledge in an explanatory (re)writing-task remains a difficult step for students.

This study has several limitations. First, the study was designed to investigate effects of explicit teaching on student learning. We strove to reduce teacher effects by working with two external teachers, who took much time to internalize the pedagogical principles. Although this strengthened treatment validity, it may have reduced ecological validity. This leads to two follow-up questions: To what extent can teachers incorporate the principles in their practices? And what would be the learning effects of integrating such an approach into everyday practice?

Second, the experiment was only conducted among 11th grade preuniversity students and the sample consisted of 95 students from one school. These choices allowed for a tightly organized, randomized setup, but it also limits the generalizability of the results. Future research should explore the effectiveness of explicit teaching of causal strategies, concepts, and epistemological reflection on a wider variety of age groups, schools, and school levels.

Third, we strove to include instruments that had been validated in other studies (Linnenbrink-Garcia et al., 2010; Maggioni et al., 2004, 2009; Pintrich et al., 1993). However, due to the domain-specific nature of our questions, several instruments were designed within the context of this study—most notably the causal-reasoning strategies and concepts questionnaire, and the essay task. We designed these instruments based on a literature review, discussed them with experts in the field of history education, and tested them out in a pilot study. Future studies should provide more robust support for these instruments. This study underscored that writing and rewriting an explanatory essay is a complex task that not only demands causal-historical reasoning ability but also knowledge of the genre and general reading and writing skills. An interesting future study could develop instruments that allow us to measure causal-historical reasoning skills in a more direct manner—rather than through a writing task. Reisman (2012) developed such an instrument, but it focusses on historical reasoning about sources.

We believe that our study holds implications for practice. Based on our results, we suggest that open-ended tasks, social interaction, and students' sense of rooted relevance should be given a more prominent place in history education. These characteristics do not only provide a fruitful context for acquiring historical topic knowledge but also provide a starting point to develop students' historical (causal) reasoning skills. The lesson-unit provides evidence that students value the task, especially when combined with explicit teaching of concepts, strategies and epistemological questions. Such an approach appears to stimulate students viewing inquiry as historical inquiry and history as an interpretative subject. It allows students to acquire intended domain-specific, deep-level strategies and appears to stimulate individual interest—although more support may be needed to apply this knowledge in a writing task. Preservice and in-service training should support history teachers in learning to design and to implement explicit-teaching environments that foster historical reasoning. This support should focus on the content of this explicit teaching and on providing teachers with concrete learning activities and open-ended tasks.

References

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher, 32,* 10–14. http://dx.doi.org/10.3102/0013189X032008010

Alexander, P. A. (2005). Teaching towards expertise (Special Monograph on Pedagogy—Teaching for Learning). *British Journal of Educational Psychology Monograph Series II, 3,* 29–45.

Barton, K. C., & Levstik, L. (2004). *Teaching history for the common good*. Mahwah, NJ: Erlbaum.

Central Bureau for Statistics. (2015, December 8). *Inkomen van particuliere huishoudens met inkomen naar kenmerken en regio* [Data file]. Retrieved from http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=80594ned&D1=2,4&D2=1&D3=0&D4=0,121&D5=1&HDR=G2,G1,T&STB=G3,G4&VW=T

Chapman, A. (2003). Camels, diamonds and counterfactuals: A model for teaching causal reasoning. *Teaching History, 112,* 46–53.

Coffin, C. (2004). Learning to write history: The role of causality. *Written Communication, 21,* 261–289. http://dx.doi.org/10.1177/0741088304265476

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Routledge.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator, 6,* 38–46.

De La Paz, S. (2005). Effects of historical reasoning instruction and writing strategy mastery in culturally and academically diverse middle school classrooms. *Journal of Educational Psychology, 97,* 139–156. http://dx.doi.org/10.1037/0022-0663.97.2.139

De La Paz, S., & Felton, M. K. (2010). Reading and writing from multiple source documents in history: Effects of strategy instruction with low to average high school writers. *Contemporary Educational Psychology, 35,* 174–192. http://dx.doi.org/10.1016/j.cedpsych.2010.03.001

Erdmann, E., & Hassberg, W. (Eds.). (2011). *Facing–mapping–bridging diversity: Foundation of a european discourse on history education* (Vol. 1). Schwalbach am Taunus, Germany: Wochenshau Verlag.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141,* 2–18. http://dx.doi.org/10.1037/a0024338

Halldén, O. (1997). Conceptual change and the learning of history. *International Journal of Educational Research, 27,* 201–210. http://dx.doi.org/10.1016/S0883-0355(97)89728-5

Havekes, H. G. F. (2015). *Knowing and doing history. Learning historical thinking in the classroom* (Doctoral dissertation). Retrieved from http://hdl.handle.net/2066/141364

Khishfe, R., & Abd-El-Khalick, F. (2002). Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science. *Journal of Research in Science Teaching, 39,* 551–578. http://dx.doi.org/10.1002/tea.10036

King, P. M., & Kitchener, K. S. (2002). The reflective judgment model: Twenty years of research on epistemic cognition. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 37–61). Mahwah, NJ: Erlbaum.

Kuhn, D., & Weinstock, M. (2002). What is epistemological thinking and why does it matter? In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 121–144). Mahwah, NJ: Erlbaum.

Lee, P., & Shemilt, D. (2009). Is any explanation better than none? *Teaching History, 137,* 42–49.

Levstik, L., & Barton, K. C. (2008). *Researching history education: Theory, method, and context*. New York, NY: Routledge.

Linnenbrink-Garcia, L., Durik, A. M., Conley, A. M., Barron, K. E., Tauer, J. M., Karabenick, S. A., & Harackiewicz, J. M. (2010). Measuring situational interest in academic domains. *Educational and Psychological Measurement, 70,* 647–671. http://dx.doi.org/10.1177/0013164409355699

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. (2014). On the joys of missing data. *Journal of Pediatric Psychology, 39,* 151–162. http://dx.doi.org/10.1093/jpepsy/jst048

Maggioni, L. (2010). *Studying epistemic cognition in the history classroom: Cases of teaching and learning to think historically* (Doctoral dissertation). Retrieved from http://hdl.handle.net/1903/10797

Maggioni, L., Alexander, P., & VanSledright, B. (2004). At a crossroads? The development of epistemological beliefs and historical thinking. *European Journal of Social Psychology, 2,* 169–197.

Maggioni, L., VanSledright, B., & Alexander, P. A. (2009). Walking on the borders: A measure of epistemic cognition in history. *Journal of Experimental Education, 77,* 187–214. http://dx.doi.org/10.3200/JEXE.77.3.187-214

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development, 50,* 43–59. http://dx.doi.org/10.1007/BF02505024

Nokes, J. D., Dole, J. A., & Hacker, D. J. (2007). Teaching high school students to use heuristics while reading historical texts. *Journal of Educational Psychology, 99,* 492–504. http://dx.doi.org/10.1037/0022-0663.99.3.492

Pintrich, P. R., Smith, D. A. F., García, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement, 53,* 801–813. http://dx.doi.org/10.1177/0013164493053003024

Prangsma, M. E., van Boxtel, C. A. M., & Kanselaar, G. (2008). Developing a "big picture": Effects of collaborative construction of multimodal representations in history. *Instructional Science, 36,* 117–136. http://dx.doi.org/10.1007/s11251-007-9026-5

Reisman, A. (2012). Reading like a historian: A document-based history curriculum intervention in urban high schools. *Cognition and Instruction, 30,* 86–112. http://dx.doi.org/10.1080/07370008.2011.634081

Rijlaarsdam, G. C. W., Couzijn, M., & van den Bergh, H. (2004). The study of revision as a writing process and as a learning-to-write process: Two prospective research agendas. In L. Allal, L. Chanqouy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes* (Vol. 13, pp. 189–207). Boston, MA: Kluwer. http://dx.doi.org/10.1007/978-94-007-1048-1_12

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,* 478–493. http://dx.doi.org/10.1037/0022-0663.88.3.478

Savery, J. R., & Duffy, T. M. (1995). Problem based learning: An instructional model and its constructivist framework. In B. Wilson (Ed.), *Constructivist learning environments: Case studies in instructional design* (pp. 135–148). Englewood Cliffs, NJ: Educational Technology Publications.

Schommer, M. (1993). Epistemological development and academic performance among secondary students. *Journal of Educational Psychology, 85,* 406–411. http://dx.doi.org/10.1037/0022-0663.85.3.406

Seixas, D. P., & Morton, T. (2013). *The big six historical thinking concepts*. Scarborough, Ontario, Canada: Nelson College Indigenous.

Stoel, G. L., van Drie, J. P., & van Boxtel, C. A. M. (2015). Teaching towards historical expertise: Developing a pedagogy for fostering causal reasoning in history. *Journal of Curriculum Studies, 47,* 49–76. http://dx.doi.org/10.1080/00220272.2014.968212

Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research, 64,* 37–54. http://dx.doi.org/10.3102/00346543064001037

van Boxtel, C., & van Drie, J. (2013). Historical reasoning in the classroom: What does it look like and how can we enhance it? *Teaching History, 150,* 44–52.

van Drie, J., Braaksma, M., & van Boxtel, C. (2015). Writing in history: Effects of writing instruction on historical reasoning and text quality. *Journal of Writing Research, 7,* 123–156. http://dx.doi.org/10.17239/jowr-2015.07.01.06

van Drie, J., & van Boxtel, C. (2003). Developing conceptual understanding through talk and mapping. *Teaching History, 110,* 27–32.

van Drie, J., & van Boxtel, C. (2008). Historical reasoning: Towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review, 20,* 87–110. http://dx.doi.org/10.1007/s10648-007-9056-1

van Drie, J., & van Boxtel, C. (2011). In essence I'm only reflecting: Teaching strategies for fostering historical reasoning through whole-class discussion. *International Journal of Historical Learning, Teaching, and Research, 10,* 55–66.

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,* 575–602. http://dx.doi.org/10.1016/j.chb.2004.10.024

van Drie, J., van Boxtel, C., & van der Linden, J. L. (2006). Historical reasoning in a computer-supported collaborative learning environment. In A. M. O'Donnell, C. E. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology* (pp. 265–296). Mahwah, NJ: Erlbaum.

VanSledright, B. A. (2011). *The challenge of rethinking history education: On practices, theories, and policy*. New York, NY: Routledge.

VanSledright, B. A., & Limón, M. (2006). Learning and teaching social studies: A review of cognitive research in history and geography. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 545–570). Hillsdale, NJ: Erlbaum.

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,* 301–311. http://dx.doi.org/10.1037/0022-0663.91.2.301

Wineburg, S. S. (2001). *Historical thinking and other unnatural acts: Charting the future of teaching the past*. Philadelphia, PA: Temple University Press.

Woodcock, J. (2005). Does the linguistic release the conceptual? *Teaching History, 119,* 5–14.

Appendix A

Lesson Goals (Explicit and Implicit Condition)

Experimental phase	Goals
Preparatory topic lessons (1/2)	Students acquire knowledge of several concrete events, concepts, countries and dates in the period leading up to the First World War. Students acquire knowledge of several abstract phenomena (i.e., nationalism, imperialism, alliances, arms race).
Intervention lessons (3/4/5)	Both conditions; general Students improve their ability to construct causal historical explanations by engaging in causal analysis through an open-ended task that prompts them to select and organize possible causes and construct a causal explanation of the outbreak of the First World War. Explicit condition; explicit attention to: Students can explain that historical explanations always involve multiple causes. Students develop a vocabulary related to causal second-order concepts and causal connections, that is, (in)direct, short-term, economic, trigger, catalyst, precondition. When constructing a historical explanation, students can organize and classify causes within the dimensions of time, content and role. Students build a causal model in which causes exert both simultaneous and linear (direct or indirect) influences. Students can explain that causal explanations are never a copy of the past itself (copier stance). Multiple valid explanations can co-exist, but not all explanations are valid (subjectivist stance). There are (academic) criteria for assessing historical explanations, including the use of evidence and arguments (criterialist stance).

Appendix B

Rubric Explanatory Writing Task

Criterion	Beginning	Developing	Adequate	Points
Structural causes	The author mentions no or only one historically correct structural cause	The author mentions two historically correct structural causes	The author mentions three or more historically correct structural causes	Max 2
Substantiation of structural causes[a]	The author doesn't substantiate any structural cause with concrete historical event (incidental causes). OR: The author only superficially elaborates one or two structural causes (e.g., elaboration without using incidental causes, or without making clear the relationship between a structural and incidental cause)	The author substantiates one or two structural cause with concrete historical event (incidental causes). OR: The author superficially elaborates more than two structural causes (e.g., elaboration without using incidental causes, or without making clear the relationship between a structural and incidental cause)	The author substantiates more than two structural cause with concrete historical event (incidental causes)	Max 2
Explanatory model	Concrete. Author describes causality on a linear level.	Abstract. Author describes causality on an abstract level, but this genre is still in development. The structure of the text can be characterized as messy or incomplete	Abstract. The author describes causality on an abstract level and does so in an appropriate and structured manner	Max 2

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix B (*continued*)

Criterion	Beginning	Developing	Adequate	Points
Use of second-order language/causal connections	Author uses no or little causal connectors of second-order language (this category also applies for students that only use "because" and "therefore," unless this is done in a very thorough manner)	Author uses causal connectors and second-order language, but almost completely aimed at organizing the text (i.e. first, multiple) AND/OR: Author makes adequate use of connection words (throughout the text, in a correct way, that makes clear the causal links)	Author uses causal connectors and second-order language, not only to organize but also to describe impact and directness (evaluate) AND/OR: Author uses a rich repertoire of causal connectors that describe nuanced relationships (and differences) between causes (i.e. this reinforced, in the background)	Max 2

[a] On this criterion argumentation is an important element (the relationship between structural and incidental causes must be described.

Appendix C

Correlations Between the Dependent Variables at Pretest and Posttest (Explicit and Implicit Condition)

Table C1
Correlations Between the Dependent Variables at the Pretest

Dependent variables	n	1	2	3	4	5	6
1. 2nd-order/strategy	95	—					
2. Subjectivist beliefs	95	.09	—				
3. Criterialist beliefs	95	.26*	−.01	—			
4. Essay quality	95	.08	−.12	.14	—		
5. 1st-order	95	.22*	.16	.10	.16	—	
6. Individual interest	95	.06	−.04	.39**	.30**	.43**	—

* $p < .05$ (2-tailed). ** $p < .01$ level (2-tailed).

Table C2
Correlations Between the Dependent Variables at Posttest (Explicit Condition)

Dependent variables	n	1	2	3	4	5	6
1. 2nd-order/strategy	53	—					
2. Subjectivist beliefs	53	−.39**	—				
3. Criterialist beliefs	53	.35*	−.01	—			
4. Essay quality	53	.15	−.10	.23	—		
5. 1st-order	53	.07	.04	.14	.24	—	
6. Individual interest	53	.14	.14	.66**	.11	.30*	—

* $p < .05$ (2-tailed). ** $p < .01$ level (2-tailed).

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table C3
Correlations Between the Dependent Variables at Posttest (Implicit Condition)

Dependent variables	*n*	1	2	3	4	5	6
1. 2nd-order/strategy	42	—					
2. Subjectivist beliefs	42	.07	—				
3. Criterialist beliefs	42	.28	.05	—			
4. Essay quality	42	.43**	−.11	.12	—		
5. 1st-order	42	.02	−.19	−.19	.41**	—	
6. Individual interest	42	.15	−.16	.26	.22	.49**	—

* $p < .05$ (2-tailed). ** $p < .01$ level (2-tailed).

Received August 3, 2015
Revision received April 29, 2016
Accepted May 17, 2016 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.