KU LEUVEN

HISTORICAL THINKING IN UPPER SECONDARY EDUCATION

Examination of teachers' beliefs, effective instructional practices and potential civic effects

KU Leuven – Faculty of Arts

Supervisor: Prof. Dr. Karel Van Nieuwenhuyse

Co-supervisor: Prof. Dr. Fien Depaepe

Thesis presented by Maria-Marjolein Wilke

for the degree of Doctor in History

Leuven, December 16th 2022

Acknowledgements

Embarking on a research project at the intersection of history and educational science is not an easy task for someone with no background in educational science. I feel very fortunate to have been surrounded over the years by many people who not only believed in me, but also provided me with the necessary support to bring such an endeavor to a successful conclusion.

In the first place, this research would never have been possible without the great commitment and excellent guidance of my supervisors. Karel: thank you for introducing me to history didactic research during my teacher training, for your great efforts to get me started on a doctoral research after two interesting years filled with various projects, and most of all for the great confidence throughout the entire process. I still don't understand how you managed time and again to provide each document - often within 24h - with the most thorough feedback. This feedback always showed your utmost commitment as a supervisor and your ability to take each text, measurement tool or analysis to the next level. You challenged me considerably in my development as a researcher, while consistently providing the necessary support and encouragement along the way. You also had great empathy for the struggles associated with this process and always freed up time in your busy schedule to get me back on track. I have learned so much over the past few years and owe much of that to you. Fien: thank you for getting on board with this project and for patiently guiding and supporting me in bridging the gap between history and educational science. Your expertise as a co-supervisor was, without a doubt, invaluable in terms of the content of the research. Above all, I am grateful for your persistent optimism and the many positive encouragements throughout the process. In short, I could not have wished for better supervisors to guide me through this research.

I would also like to thank the members of my supervisory committee for their important input throughout the research. Ellen Claes, Geerte Savenije and Michiel Voet: thank you for allowing me to rely on your expertise, critical eye and constructive feedback over the past four years, and for taking the time to reflect on the direction of my research. The meetings of the unofficial 'Dutch-Flemish History Research Group' have also supported me tremendously throughout my

doctoral research. I am very grateful to all colleagues for welcoming me into this 'research group', and for the particularly pleasant, constructive and instructive meetings where I also became familiar with your own research and writing processes. These meetings have been immensely helpful throughout the years.

My deepest gratitude goes out to all the teachers and students who were involved in my doctoral research in various ways. This research would have been impossible without the engagement of history teachers all over Flanders, who shared their experiences and views on history education, opened up their classrooms, trusted me to prescribe their classroom practices for several weeks (even in the middle of a pandemic), and allowed their students (willingly or reluctantly) to participate in my research. The *praktijklectoren* of the history teacher training program, whom I could turn to for all sorts of things during my research, also deserve a special mention. I would also like to thank my former *praktijklector* Guy Putseys, under whose enthusiastic encouragement I took my first steps in teaching and whose expertise I could count on, even several years after my teacher training had been completed.

A doctoral research project can easily feel like a lonely undertaking. Fortunately, this was more than compensated for by the friendly and warm atmosphere among the colleagues at the fifth floor. The many wonderful colleagues with whom I shared an office throughout the years have also made the research process that more enjoyable. In particular, I would like to thank Ruth with whom I shared an office for five whole years, for providing me with good company, fun conversations and, above all, delicious waffles, cake and cookies!

There are also people outside of academia who have been especially supportive over the past few years. I am grateful to my parents and siblings for their unconditional support and unwavering trust. After all these year, I finally have to admit that everything has indeed worked out well. Thank you to my dear friends for always providing me with the necessary distraction and/or encouragement.

Finally, Dries, *liefje*, thank you for being my greatest supporter for over a decade. Thank you for your great intelligence, patience, gentleness and humor which have helped me through many (mental, physical, emotional, and at times even methodological) slumps. I hope that one day I can do the same for you.

Table of contents

Acknowledgements 3

Table of contents 5

Chapter 1 7

General introduction

Chapter 2 35

Teachers and historical thinking. An exploration of the relationship between conceptualizations, beliefs and instructional practices among Flemish history teachers

Chapter 3 71

The interplay between historical thinking and epistemological beliefs. A case study with history teachers in Flanders.

Chapter 4 107

Fostering secondary students' historical thinking. A design study in Flemish history education

Chapter 5 143

Fostering democratic citizenship through historical thinking? A cluster randomized controlled intervention study

Chapter 6 209

Exploring differential effects of an intervention on historical inquiry tasks. A qualitative analysis of 12th grade students' progress.

Chapter 7 251

Exploring the possibility of transfer from historical to contemporary inquiry tasks. A qualitative analysis of the effects of an intervention in 12th grade history education.

Chapter 8 299

General conclusion and discussion

References 329

Output 345

Collaboration agreement 349

Chapter 1

General introduction

INTRODUCING HISTORICAL THINKING AS A PRIMARY GOAL FOR HISTORY EDUCATION

Historical thinking as an educational goal

Throughout the 19th century, history education in Belgium and abroad was largely characterized by a focus on the teaching of single (national) narratives; little attention was paid to the existence of multiple narratives (De Wever et al., 2011; Metzger & McArthur Harris, 2018). Following developments in the academic discipline itself, history education from the 1970s onwards gradually introduced historical thinking as a central objective (Lévesque & Clark, 2018). With this shift towards historical thinking as an educational goal, history education no longer aimed to present students with a single historical narrative, but rather to introduce them to the interpretive and constructed nature of history. Besides providing students with knowledge about the past (i.e., 'knowing history'), historical thinking also aims to introduce students to the ways in which this knowledge has come about and hence, initiate them into the methods of professional historians (i.e., 'doing history') (Havekes et al, 2012). Knowing history is mainly centered around first-order, substantive knowledge, i.e., knowledge about what occurred in the past and of particular historical periods, persons, structures or phenomena such as feudalism or the French Revolution. This substantive knowledge allows students to build a historical frame of reference. Doing history requires students to obtain second-order and procedural knowledge. Second-order knowledge is related to organizing concepts that historians use when studying the past. It contains concepts such as cause and consequence, evidence, or continuity and change. Procedural knowledge describes strategies or heuristics applied by historians when constructing historical representations. Such procedures include, among others, evaluating a source's reliability, assessing an author's perspective or interpreting historical context (Lee, 2004; van Drie & van Boxtel, 2008; VanSledright, & Limón, 2006). Within historical thinking, knowing and doing history are intrinsically intertwined. Students need substantive knowledge in order to reason about history, while second-order and procedural knowledge are necessary to reflect on and broaden this substantive knowledge (Lee, 2005; VanSledright & Limón, 2006).

When students acquire knowledge, not only about what happened in the past but also about how historians construct such knowledge, they are able to develop nuanced epistemological beliefs about history. These beliefs refer to a person's views on the nature of knowledge and processes of knowledge construction in history (Buehl & Alexander, 2001). Nuanced epistemological beliefs occur when students understand that historical knowledge is interpretive and constructed and therefore subject to change. This does not mean that history is merely a matter of opinion, as historical claims can be evaluated using disciplinary criteria (Kuhn et al., 2000; Maggioni et al. 2009; Stoel, Logtenberg et al., 2017). The shift in history education towards this underlying epistemological paradigm runs parallel to changes in history as an academic discipline. For a long time, historians were convinced that it was indeed possible to accurately reconstruct the past. In this 'realist' epistemological paradigm, it was believed that the rigorous application of disciplinary methods would allow historians to uncover and even reconstruct the 'true' and 'authentic' past. A more interpretive, perspectivist, approach, which emphasizes the interpretive and constructed nature of historical knowledge, became prominent in the last quarter of the 20th century (Lorenz, 2002; Munslow, 1997; 2000).

For students in secondary education, this nuanced epistemological stance which underlies historical thinking, does not come about naturally. In fact, students typically hold rather naïve views about knowledge and knowledge construction in history as they tend to assume that historical knowledge is fixed and singular, and that sources are direct reflections of the past (Lee & Shemilt, 2003). Students also experience difficulties in distancing themselves from their own, contemporary viewpoints which is necessary in order for them to understand the past in its own right. Historical thinking is therefore very challenging for students as it directly opposes the way that they spontaneously think about the past. For these reasons, historical thinking has famously been called an 'unnatural act' (Wineburg, 2001). In the same vein, Lee (2011) argued that historical thinking requires students to "substitute counter-intuitive ideas for their common-sense everyday life understandings" (p. 136). Fostering this 'unnatural act' among students is therefore quite an ambitious educational goal which has nevertheless found its way into the history education curricula.

Operationalizing historical thinking

Research on historical thinking starting in the United Kingdom in the 1970s (e.g., Dickinson & Lee, 1978) gave the impetus for significant changes in the English history curricula. This then new approach of history education influenced curricular changes worldwide, with historical thinking and related notions such as 'historical reasoning' or 'historical literacy' becoming embedded in many history curricula to date. In Canada for instance, the concept of historical thinking has been very influential in history and social studies curricula. It also takes on an important position in the United States Common Core State Standards (Reisman, 2015) and in the curricula of several European countries, such as Belgium (in both Flemish and Francophone education), the Netherlands and Finland.

Building on the research initiated in the United Kingdom, various models have been developed that operationalize historical thinking (e.g., Körber & Meyer-Hamme, 2015; Seixas, 2017; Seixas & Morton, 2013; Seixas & Peck, 2004; Van Boxtel & van Drie, 2008; Wineburg, 2001). One influential model is that of Seixas and Morton (2013) who define historical thinking as "the creative process that historians go through to interpret the evidence of the past and generate the stories of history" (p. 2). In the 'Historical Thinking Project', they identify six historical thinking concepts[1] describing "the way historians transform the past into history"

[1] Seixas and Morton describe historical significance as addressing the question of "how do we decide what is important to learn about the past" (p. 5), as particular events can become more or less significant in light of our present circumstances. Evidence relates to the matter of "how do we know what we know about the past" (p. 5). It refers to historians' "finding, selecting, contextualization, interpreting and corroboration of sources for an historical argument" (Seixas & Morton, 2013, p. 5). Continuity and change examine the complex flows of history. Historians examine how, at a specific period in time, particular things can change at different paces, whereas others remain the same. Cause and consequence deal with the question of why historical events happened and what impact they had. This includes a reflection on human choice and action, and the historical circumstances in which they were made, shaped or constrained. Perspective-taking is about understanding people in the past, referring to the different "social, cultural, intellectual and even emotional contexts that shaped people's lives and actions" (Seixas & Morton, 2013, p. 6). Finally, the ethical dimension concerns issues of ethical judgements related to history (Seixas & Morton, 2013).

(Seixas & Morton, 2013, p. 3) which can be used by teachers to help students think about how history is constructed, and to construct history themselves.

In the United States, a strong research tradition has developed around the concept of 'historical literacy', which is closely related to historical thinking. Research on students' historical literacy predominantly focuses on students' disciplinary reading and writing, alongside procedural knowledge, as fundamental elements in developing students' understanding of history (De La Paz et al., 2016; Lévesque & Clark, 2018). Important contributions in this regard came from the work of cognitive and educational psychologist Wineburg (1991; 2001) who studied fundamental differences between novice and expert readings of historical texts. Based on this research, three discipline-specific heuristics were distinguished which are used by professional historians when reading historical texts: sourcing, contextualization, and corroboration[2] (Wineburg, 1991). These heuristics are what distinguishes experts' critical reading of sources from that of students, who are more likely to uncritically approach historical documents as carriers of information and hence read documents in search of historical facts (Lévesque & Clark, 2018; Wineburg, 1991). The use of these heuristics is therefore also connected to differences between experts' and novices' epistemological views (Wineburg, 1991). The heuristics identified by Wineburg (1991) were transformed into tools for educators in history (Wineburg, 2001) and provided the basis for several subsequent studies on students' historical literacy (e.g., De La Paz, 2005; Monte-Sano, 2012; Nokes, 2013).

In the Netherlands, a different approach to operationalizing historical thinking was taken by van Drie and van Boxtel (2008). Rather than using the practices of professional historians as a starting point (as was done by Wineburg), they developed a model of historical reasoning, based on concepts from research literature which they adapted after having analyzed historical thinking in students' essays. They use the term 'historical reasoning' rather than historical thinking to describe the "activity in which a person organizes information about the past in

[2] Sourcing refers to the need to examine information about the source, such as the type of source, the author, and the time and place of its creation. Contextualization refers to the need to situate a historical document in its historical context. Corroboration refers to comparing information across multiple sources (Wineburg, 1991).

order to describe, compare and/or explain historical phenomena" (van Drie & van Boxtel, 2008, p. 89). In operationalizing historical reasoning, they propose a framework consisting of six elements: asking historical questions, using sources, contextualization, argumentation, using substantive concepts and using meta-concepts (i.e., those related to the methods used by historians when investigating and describing the past previously referred to as second-order and procedural knowledge) (van Drie & van Boxtel, 2008).

Despite the variation within the existing operationalizations of historical thinking, most models contain a number of common aspects. They aim to provide students with substantive knowledge about the past in order to build a historical frame of reference. Students also acquire second-order and procedural knowledge which allows them to gain insight into the methods of historians and hence, exposes the interpretative and constructed nature of historical knowledge. To this end, students are supported in applying this second-order and procedural knowledge to analyze and build historical representations by engaging them with historical questions, critical source analysis and the application of historical modes of reasoning, such as causality, continuity and change, and historical contextualization. Thus, an important aspect of historical thinking is the development of students' historical inquiry competences which means they are able to build and substantiate their own historical representations of the past, based on a critical analysis of multiple historical sources and perspectives.

The various operationalizations of historical thinking described above approach historical thinking primarily as a cognitive activity that mirrors the activities of disciplinary history. Increasingly, however, scholars have emphasized the need to also adopt a sociocultural perspective. In 2018, for instance, van Boxtel and van Drie extended their model of historical reasoning by adding individual and sociocultural resources that students use when engaging in historical reasoning. Students' substantive and second-order knowledge, interest in history and epistemological beliefs were identified as individual resources. Sociocultural resources refer to the influence of historical representations in the public sphere and of students' connection to, and identification with, particular social groups (van Boxtel & van Drie, 2018). Such a socio-cultural perspective hence takes into account the fact that students enter the classroom with pre-existing historical narratives that are tied to their personal and social identities, or that are influenced

by public representations of the past (Lévesque et al., 2013; Peck, 2010; van Boxtel & van Drie, 2018; Wineburg, 2000). History teachers need to engage with these pre-existing representations that students bring into the classroom (van Boxtel & van Drie, 2018). It appears, for instance, that students' personal connections may result in certain historical topics becoming controversial or sensitive and may therefore hinder students' ability to reason critically about them (Barton & McCully, 2007; Goldberg et al., 2011; Savenije et al., 2014). Besides the influence of public representations of the past and students' identities, affective dimensions, such as students' interests and motivation, are also believed to be connected to their historical thinking ability (Paul et al., 2017; van Boxtel & van Drie, 2018; van der Eem, 2022). Research on the role of these dimensions within history education research specifically is, however, limited.

Historical thinking in Flemish history education

In Flanders, a curriculum reform is gradually being implemented across secondary education since September 2019[3]. With this reform, historical thinking is firmly and explicitly established as history education's principal goal. While the previous history standards also referred to historical thinking, this was done in a more implicit way (see section on 'research context'). By contrast, the new history standards explicitly put forward a concrete definition of historical thinking:

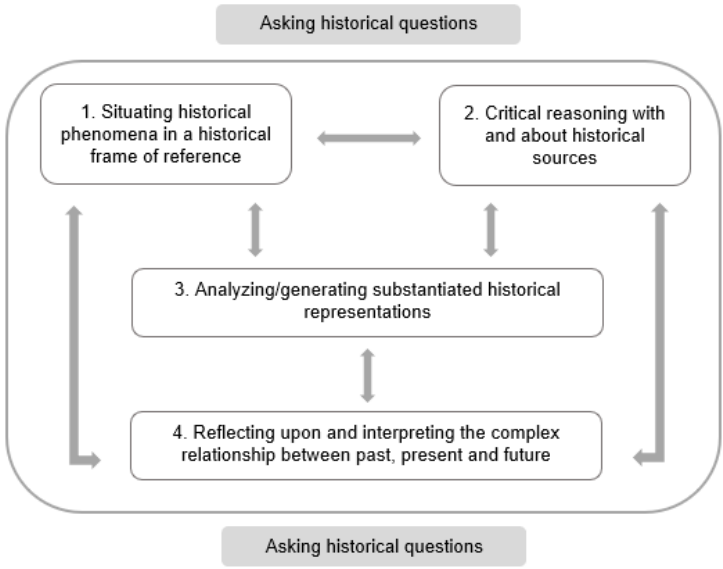
"Historical thinking is first and foremost about understanding and organizing information about the past, with the aim of describing, comparing and explaining historical phenomena (people, groups, events and developments from the past) in their historical context and in a long-term. It is important, in this respect, to understand that past and present are fundamentally different. Therefore, historical thinking is also about an understanding of and a reflection on the complex relationship between past, present and future. This can, among others, be done by drawing analogies between the past and the present, in search

[3] In June 2022, the new educational standards for the second and third stage of secondary education were overturned by the Constitutional Court. An adjusted proposal for 'minimum attainment goals' is being developed and will be presented to the Flemish Parliament in March 2023 (Flemish Ministry of Education and Training, 2022). It is currently unclear how this new proposal will look like for history education.

for similarities and differences. Historical thinking hence requires an understanding of both the past and historical practice, which are inextricably bound up with each other. For one needs to know how knowledge of the past is constructed, and one needs to understand the tentative character of historical knowledge. Only then, one can start thinking critically of (representations and uses of) the past" (Agency for Higher Education, Adult Education, Qualifications and Study Grants, n.d.; translation from: Van Nieuwenhuyse, 2020, p. 371).

In operationalizing historical thinking, the new Flemish history standards state that historical thinking is an act that starts with 'asking historical questions' and which is then developed along four building blocks, following international models of historical thinking. The full operationalization is depicted in Figure 1.

Figure 1. Operationalization of historical thinking in the new Flemish history standards (based on: Van Nieuwenhuyse, 2020, p. 375).



A first building block is concerned with situating historical phenomena, sources and representations in a historical frame of reference. The historical frame of reference contains three dimensions: time, space and societal domains (i.e., the political, economic, cultural and social domain). Within this building block, the history standards do not impose specific historical topics that need to be covered, but do list very general substantive concepts, such as imperialism, migration or

agriculture, and second-order concepts, such as continuity and change, or local-regional.

'Critical reasoning with and about historical sources' (building block 2) expects students to engage in source analysis. This includes both reasoning 'with' and 'about' sources (Rouet et al., 1996). Reasoning with sources means that students should be capable of extracting and selecting information from sources and use it in support of a claim about the pact. Reasoning about sources refers to the ability to critically evaluate the usefulness and value of a historical source in light of a historical question. To this end, students should take into account aspects such as the author's perspective and motivations, the target audience for the source and the context in which a source originated. Reasoning with sources hence requires students to acquire second-order and procedural knowledge related to this critical source analysis.

The third building block prescribes that students should analyze and critically deconstruct existing historical representations and generate their own substantiated historical representations. To this end, students should obtain second-order and procedural knowledge not only related to the critical analysis of historical sources, but also regarding the application of typical historical modes of reasoning. These historical modes of reasoning include, among others, reasoning about cause and consequence, historical contextualization, historical empathy, and reflecting about agency.

The fourth building block 'reflect upon and interpret the complex relationship between past, present and future' explicitly includes the need for epistemological reflection about history. Standards related to this building block state that students should comprehend the interpretive and constructed nature of historical knowledge and the importance of one's positionality. This will allow students to be critical about public uses and misuses of the past, for instance in processes of social identity formation, constructing collective memories, and transmitting certain values (Agency for Higher Education, Adult Education, Qualifications and Study Grants; n.d.; Van Nieuwenhuyse, 2020). When students learn to think about the past in this way, the history standards expect that students will develop 'historically conscious behavior' meaning, among, others that they can reflect on long-term connections between past and present, formulate responses to societal challenges, take into account historical arguments, show

respect for facts and evidence, and be open-minded in their engagement with other value systems and cultures (Van Nieuwenhuyse, 2020). In this way, the new history standards expect history education to contribute to democratic citizenship without instrumentalizing history education solely for this purpose.

Historical thinking and democratic citizenship

The idea that history education has a role to play in shaping future citizens is certainly not new. During the 19th century, in the context of emerging nation states, history education was burdened with the task of shaping national identities and creating national cohesion (Wilschut, 2010). While political and societal demands of history education have shifted over the course of the centuries, the basic notion that history education should fulfill a societal role still prevails. Several history (education) scholars have opposed this instrumental view of history education as these societal goals are often considered at odds with a disciplinary approach to history (Wils et al., 2011; Wilschut, 2010). For instance, a disciplinary focus requires a detachment from the present in order to fully understand the past in its own right, whereas societal expectations often imply a close attachment to the present as they depart from contemporary perspectives and frameworks (Harris, 2011; Wils et al., 2011).

With the introduction of historical thinking into the history curricula, however, new ways of thinking about the connection between history education and democratic citizenship have emerged. Several history education scholars are convinced that, without instrumentalizing historical thinking specifically for the purpose of citizenship, historical thinking may indeed have the ability to contribute to democratic citizenship. They argue that certain knowledge, skills and dispositions that are inherent to historical thinking will also benefit students when they participate in society (e.g., Barton & Levstik, 2004, Cinnamon et al., 2021; Lee & Shemilt, 2007; McCully, 2012; Monte-Sano, 2016; Nokes, 2013; Slater, 1995; Stearns, 1998; Wineburg & Reisman, 2015). In this regard, scholars point to the fact that history education provides students with knowledge of political, social, cultural and economic systems, including contemporary democratic society (Lévesque, 2009). This knowledge, they argue, is essential to the development of citizens who are knowledgeable about institutions, values and issues facing the

contemporary society, who can participate in societal debates, and who can take up an active role in society (Lee & Shemilt, 2007; Lévesque, 2009; Stearns, 1998).

Others allude to the potential democratic benefits of particular historical modes of reasoning, such as reflecting about agency or historical empathy. Agency refers to "the ability to act on decisions in order to bring about desired goals" (Barton, 2012, p. 131). Reflecting on agency in the past not only provides students with a better understanding of historical changes and historical agency but is also expected to encourage them to reflect on their own potential to generate change in contemporary society, and to potentially support a more active societal engagement (Barton, 2012; Peck et al., 2011; Seixas, 2012). Another historical mode of reasoning that is connected to democratic citizenship is historical empathy. Historical empathy can include both a cognitive dimension of understanding the perspective of people in the past, and an affective dimension of connecting to, and caring for, the historical actors involved (Barton & Levstik, 2004; Endacott & Brooks, 2013). When students become familiar with historical empathy, this is assumed to provoke them to also recognize and value the existence of different perspectives in contemporary society, to understand how an individual's perspective is shaped by political, social, cultural and economic contexts, and to motivate them to analyze and reevaluate the actions and beliefs of people in the present (Barton & Levstik, 2004; Endacott & Brooks, 2013).

Another aspect of historical thinking that is expected to potentially yield democratic benefits is the development of students' historical inquiry competence. Several scholars assert that when students are engaged in building substantiated historical representations based on a critical analysis of sources, they will develop knowledge, skills and dispositions which are also applicable to contemporary contexts and which may hence contribute to democratic citizenship (e.g., Barton & Levstik, 2004, Lee & Shemilt, 2007; McCully, 2012; Monte-Sano, 2016; Nokes, 2013; Slater, 1995; Stearns, 1998; Wineburg & Reisman, 2015). These scholars assert that engaging students in historical inquiries will improve their ability to take up a reasoned position based on arguments and evidence, and to consider multiple perspectives, when participating in public debates on societal issues (i.e., democratic skills). Besides improving their ability to do so, it is also expected that these inquiry competences will foster students' inclination to do so (i.e.,

democratic dispositions[4]). Hence, a relationship is assumed between students' historical inquiry competence and certain democratic skills and dispositions.

PROBLEM STATEMENT: HISTORICAL THINKING IN EDUCATIONAL PRACTICE

Considering its complex and 'unnatural' nature, historical thinking constitutes quite an ambitious educational goal. Getting students to think 'historically' requires teachers to apply well thought-out teaching strategies. The design and effectiveness of such strategies have been the subject of various studies (e.g., De La Paz, 2005; De La Paz et al., 2016; De La Paz & Felton, 2010; Huijgen et al., 2018; Monte-Sano & De La Paz, 2012; Nokes et al., 2007; Reisman, 2012; Stoel, van Drie & van Boxtel, 2017). From this research, there is now substantial empirical evidence for the effectiveness of particular instructional strategies, yet research into educational practices - both internationally and in Flanders - shows that these strategies do not always find their way into the history classroom. There is hence still much to be learned about how to foster historical thinking and about the effects this would generate. In this dissertation, we aim to address a number of key issues in this regard.

Issue 1: The role of the history teacher

Teachers have an essential role in implementing historical thinking into educational practice. Various studies have shown that contextual factors, such as the available time, access to didactic resources and the presence of standardized tests affect the extent to which teachers include historical thinking in their practice (Barton & Levstik, 2004; 2003; Hicks et al., 2004; Husbands, 2011; Van Hover & Yeager, 2003; VanSledright & Limón, 2006; Voet & De Wever, 2016; 2019; Wansink et al., 2016). These studies have also shed light on the importance of a number of teachers' beliefs in shaping their educational practice: beliefs about the

[4] Dispositions can be understood as a combination of inclination, sensitivity to occasion and ability (Perkins et al., 1993). In the literature on historical thinking's potential democratic contribution, however, history education scholars (e.g. Barton & Levstik, 2004; Gallagher, 1996; Monte-Sano, 2016) mainly focus on the first element, i.e. the potential effects on students' inclination towards certain behaviors.

goals of history education, beliefs about students' abilities, beliefs about the teaching of history, and epistemological beliefs about history (e.g., Hartzler-Miller, 2001; McCrum, 2013; Virta, 2001; Voet & De Wever, 2016; 2019; Wansink et al., 2016; Yilmaz, 2008).

The available research on the role of contextual factors and beliefs in shaping teachers' instructional practices contains a number of limitations. First, contextual factors and teachers' beliefs do not fully explain the large variation in teachers' historical thinking practices. Some studies, for instance, note a disconnect between teachers' beliefs and their instructional practice (e.g., Hicks et al., 2004; McCrum, 2013; Wils et al., 2011). The existing studies also yield different results regarding the role of teachers' epistemological beliefs. Although it is assumed that nuanced epistemological beliefs are essential for teachers to develop an instructional practice that fosters historical thinking (Wansink et al., 2016; Yilmaz, 2008), this relationship is not always confirmed by concrete studies (Hartzler-Miller, 2001; VanSledright, 1996; Voet & De Wever, 2019). More research is hence needed to further explore how precisely various teachers' beliefs influence their instructional practice.

Second, when examining the relationship between teachers' beliefs and practices, scholars often rely on indirect measures for teachers' practices, for instance through self-reports or via teachers' beliefs about instruction (e.g., Hicks et al., 2004; McCrum, 2013; Voet & De Wever, 2016; 2019). Research studying history teachers' actual teaching practices is scarce and is often limited to a small number of teachers (e.g., Hartzler-Miller, 2001; Martell, 2013; VanSledright, 1996).

Third, existing research has mainly focused on the role of teachers' beliefs and/or contextual factors in shaping their instructional practices but has not yet examined teachers' understanding of the concept of historical thinking, nor its potential influence on teachers' practices. To what extent, for instance, do teachers themselves understand the notion of historical thinking in its entirety? Considering that historical thinking is a broad and complex notion, it is probable that this would vary strongly among history teachers. Although there are studies examining teachers' understanding of particular aspects of historical thinking, such as historical empathy (Bartelds et al., 2020), and historical thinking practices such as inquiry-based learning (Voet & De Wever, 2016), there is little research on

how teachers conceptualize the notion of historical thinking as a whole and how this may affect their instructional practice.

Fourth, scholars in this line of research face conceptual and methodological difficulties, particularly related to measuring teachers' epistemological beliefs. One difficulty is related to the absence of reliable and validated measurement instruments as existing scales trying to grasp epistemological beliefs do not always generate reliable results in follow-up research (e.g., Stoel, van Drie & van Boxtel, 2017; VanSledright & Reddy, 2014; Voet & De Wever, 2019). Another difficulty has to do with the fact that epistemological beliefs can be conceptualized in different ways. Most notably, scholars make a distinction between practical and formal epistemologies. Formal epistemologies are general ideas about the "characteristics of knowledge and its justification in a particular field" (Sinatra & Chinn, 2012, p. 264), whereas practical epistemologies are related to specific epistemic practices (Sinatra & Chinn, 2012). These formal and practical beliefs are not necessarily aligned and measuring them requires different types of instruments (Sandoval, 2005; Sinatra & Chinn, 2012). To date, it is unclear how best to measure both of these types of epistemology, and how they are related to teachers' instructional practice.

Issue 2: Designing instructional practices to foster historical thinking

History teachers face serious challenges with regard to designing instructional practices that address historical thinking in a profound way. Rather than addressing both knowing and doing history in an integrated way, history classes often remain focused predominantly on the content of the past (i.e., knowing history). When teachers include aspects of doing history, their practices often do not accurately reflect the interpretive and constructed nature of historical knowledge (e.g., Hicks et al., 2004; Van Nieuwenhuyse et al., 2015; VanSledright & Limón, 2006; Voet & De Wever, 2016). It is therefore important that teachers are well supported in designing and applying instructional practices that integrate both knowing and doing history, and that are effective in fostering students' historical thinking. Clear guidelines on how to develop such practices are hence much needed.

(Intervention) research on effective instructional practices has already provided evidence for the effectiveness of certain instructional practices. Among

others, design principles such as explicit teaching, guided and independent practice, scaffolding, coaching and interaction have been shown to improve various aspects of students' historical thinking (e.g., De La Paz, 2005; De La Paz et al., 2016; De La Paz & Felton, 2010; Huijgen et al., 2018; Monte-Sano & De La Paz, 2012; Nokes et al., 2007; Reisman, 2012; Stoel, van Drie & van Boxtel, 2017). Nevertheless, inferring concrete guidelines for the design of instructional materials that foster historical thinking in an integrated way (i.e., knowing and doing history) remains difficult due to a number of limitations in the available research.

A first limitation is that the existing studies provide little information about the way that general educational principles were translated to the specific context of history education, although there are exceptions to this (such as Stoel et al., 2015). A second limitation has to do with the fact that these studies typically do not approach historical thinking in an integrated way, but rather focus on aspects of doing history and less so on knowing history. In the reporting of these studies and of the instructional materials that were developed for the study, it is often unclear to what extent the content of these materials is aligned with recent historiographical insights on the topics that are studied. Yet it is important to approach historical thinking in an integrated way, considering that both knowing and doing history are essential for historical thinking (Lee, 2005; van Boxtel & van Drie, 2017), and considering that there is often a disconnect between academic and school history (e.g., Bentrovato & Van Nieuwenhuyse, 2020). A final limitation concerns the lack of information regarding the social validity of the designed materials. Currently, intervention studies primarily report on the effectiveness of the designed materials and pay insufficient attention to their social validity, i.e., to whether the instructional materials are deemed acceptable, relevant and useful by history teachers (Kazdin, 2005). Yet, if interventions are to have an applied value it is crucial that they are considered socially valid by the stakeholders. Aiming for social validity may hence bridge the gap between theoretical guidelines for designing instructional practices to foster historical thinking and their practical implementation.

There is hence much to be learned about how to design instructional practices that are effective in fostering historical thinking among students in an integrated way and that are considered socially valid by teachers.

Issue 3: Students' progress on historical thinking?

Research on students' historical thinking has shown that it is possible to improve various aspects of students' historical thinking through well-designed instructional practices. In particular, intervention studies have managed to advance students' performance on various aspects of doing history, such as historical contextualization (Huijgen et al., 2018), causal reasoning (Stoel et al., 2015; Stoel, van Drie & van Boxtel, 2017) and critical source analysis (e.g., De La Paz et al., 2016; Nokes et al., 2007; Reisman, 2012). Although these studies have overall established progress on students' historical thinking, they do not always unequivocally obtain positive results on all of the targeted variables (e.g., Reisman, 2012; Stoel et al., 2015). Stoel and colleagues (2017), for instance, found positive effects of an intervention on students' knowledge of causal-reasoning strategies and second-order concepts, but not on students' (substantive) topic knowledge, nor on the overall quality of students' causal reasoning essays.

In order to fully understand how and why students (do not) advance in terms of historical thinking, more research is needed in order to overcome a number of limitations. First, intervention studies often offer (too) little insight into the precise process of students' progression in historical thinking due to particular interventions. In the existing studies, essay tasks are often used to examine students' historical thinking, which are then evaluated using rubrics (e.g., De La Paz et al., 2016; Monte-Sano, 2012; Stoel, van Drie & van Boxtel, 2017). When the effects of certain instructional practices are examined, there is in most cases not a detailed analysis of how precisely students progressed on each component of the evaluation rubrics that were used (with certain exceptions, such as Stoel, 2017). These studies hence provide only a general view of students' progress on aspects of historical thinking; more detailed views are often absent. Second, information on whether these interventions potentially yielded differential effects on various types of students is also scarce. Yet, such a detailed analysis of students' progress and of potentially differential effects might provide a better understanding of how precisely students' historical thinking evolves. Third, there is still a lot of unclarity about the role of particular elements in students' learning in history. The role of students' epistemological beliefs, for instance, as well as that of affective aspects, such as students' interest in history or

the extent to which they value particular aspects of historical thinking, has not been extensively studied. Yet, scholars believe these aspects may be of relevance to the development of students' historical thinking.

Issue 4: Fostering democratic citizenship through historical thinking

Despite the caveats discussed above, research suggests that carefully designed learning environments can be effective in fostering students' historical thinking ability. Several scholars have asserted that doing so will also contribute to aspects of democratic citizenship. Of particular interest in this regard are students' historical inquiry competences which are believed to foster certain democratic skills and dispositions (see earlier). Several scholars (e.g. Barton & Levstik, 2004, Cinnamon et al., 2021; Lee & Shemilt, 2007; McCully, 2012; Monte-Sano, 2016; Nokes, 2013; Slater, 1995; Stearns, 1998; Wineburg & Reisman, 2015) describe how engaging students in historical inquiries might improve their ability to make reasoned and rational judgments towards public debates, based on arguments and evidence, a critical evaluation of contemporary arguments and evidence, and an evaluation of different perspectives on the public debate. Besides these democratic skills, engaging students in historical inquiries is also expected to foster democratic dispositions as it might encourage among students an inclination to enter debates from a position of rationality, to question their initial standpoint, and to consider different perspectives (Barton & Levstik, 2004; Gallagher, 1996; Lee & Shemilt, 2007; McCully, 2012; Monte-Sano, 2016). In describing this connection between historical thinking and democratic citizenship, scholars seem to assume that the knowledge (second-order and procedural), skills and dispositions acquired when engaged in historical inquiries will automatically and spontaneously be transferred to contemporary contexts.

Two limitations, however, should be noted regarding this presumed relationship. First, research examining a transfer from historical inquiry competences to democratic skills and dispositions remains very scarce, with the study of Reisman (2012) being one notable exception. Reisman (2012) established that students who were given explicit instruction on the use of discipline specific heuristics when reading historical sources, applied some of these strategies to contemporary sources as well. To date, this potential contribution to democratic citizenship therefore remains mostly based on assumptions, rather than empirical

evidence. Second, research on transfer of learning in other domains indicates that such a 'far' transfer, from one context to another, is not easy to obtain, unless particular instructional conditions are in place (Perkins & Salomon, 2012; van Merriënboer & Kirschner, 2007). Therefore, the expectation of a 'spontaneous', 'far' transfer from a historical to a contemporary context may not be as obvious as several history education scholars seem to assume.

Given the central role that historical thinking has obtained in history education curricula in several countries and given the long-standing assumption that historical thinking may contribute to democratic citizenship, it seems important to gain more insight into whether such a contribution is indeed possible, and what it may look like.

GOALS AND CONTEXT OF THE RESEARCH

Research objectives

Based on the issues described above, seven distinct research goals are outlined which accommodate the various limitations that were identified for each issue.

A first set of research objectives is related to the role of history teachers. Considering the crucial role that teachers play in fostering historical thinking among students, it is essential to gain more insight into how various factors influence teachers' instructional practice, particularly regarding historical thinking. A first aim is therefore to gain a better understanding of the relationship between teachers' beliefs, conceptualizations and practices regarding historical thinking (**research objective 1**). This will also allow us to identify particular difficulties that teachers experience when putting historical thinking into practice. Because several researchers are convinced of the importance of teachers' epistemological beliefs in shaping their classroom practice, and considering that accurately mapping these beliefs is extremely challenging, another goal of this dissertation is to explore the precise role of teachers' epistemological beliefs and compare different methods for measuring them. In this way, we aim to develop a better understanding of (the role of) teachers' epistemological beliefs in their instructional practice and to provide a methodological contribution to the existing research on teachers' epistemological beliefs (**research objective 2**).

Second, it would be beneficial to gain more insight into how to design an instructional practice for history education that is both effective in fostering students' historical thinking in an integrated way (i.e., both knowing and doing history) and that is considered relevant, practical, and applicable by history teachers, and hence socially valid (**research objective 3**).

A third set of objectives is related to examining the effectiveness of such an instructional practice on students' historical thinking and on their epistemological beliefs, taking into account affective variables (**research objective 4**). An additional objective is to gain insight into how precisely students progress in terms of historical thinking (**research objective 5**).

Finally, the potential relationship between historical thinking and democratic citizenship and in particular, the democratic goals related to the development of students' historical inquiry competences, will be studied. In this dissertation we aim to examine whether advancing students' historical inquiry competence can indeed generate a 'spontaneous' transfer to a contemporary context and hence foster particular democratic skills and dispositions related to students' engagement with public debates on societal issues (**research objective 6**). It will also be explored how precisely such a transfer might take shape (**research objective 7**).

In light of these research objectives, three main studies are set up which provide the basis for this dissertation (see section on 'research design'). These studies are conducted in the context of Flemish upper secondary history education. Before describing the three main studies in detail, some essential elements of this particular educational context are discussed.

Research context: a history education in transition

At the time of the studies, Flemish history education was in the middle of a curriculum reform. This particular context created a unique set of circumstances. While the reform of the history standards, explicitly identifying historical thinking as a main goal (see earlier), had been initiated in the first stage of history education, this reform was not yet in effect in the second and third stages. Here, the previous history standards which have been in place since the 1990s, were still in effect (Flemish Ministry of Education and Training, 2000). These standards only touch upon historical thinking in an implicit way. Although they prescribe that history education should offer an introduction to history as a scientific discipline, they offer little guidance on how to do so. They emphasize, for instance, the importance of critical thinking skills and source analysis, yet they do not provide concrete guidelines for how to perform such a critical analysis. The standards also lack a concrete definition or operationalization of historical thinking. In terms of teaching content, teachers are offered a lot of freedom in determining which historical topics to address. The standards only provide general guidelines in this regard, establishing for instance which time period should be addressed at which stage of history education and stipulating that teachers should strive for a balance between political, cultural, economic and social topics (Flemish Ministry of Education and Training, 2000; Van Nieuwenhuyse, 2020). Teachers participating in the studies central to this dissertation were hence in a transitioning phase between the previous and new history standards. Therefore, it was unclear to what extent teachers were already familiar with the concept and teaching of historical thinking, which would become essential in the reformed history education.

Another important characteristic of the educational context of this dissertation are the differences in teachers' initial training. While (history) teachers active in the first and second stage of secondary education are not required to have a master's degree, teachers in the third stage do. For history education, this means that most history teachers are in the first place trained as historians and have subsequently completed an additional teacher training program. This program has changed significantly in the last decade as it increasingly paid attention to historical thinking on a theoretical and practical level.

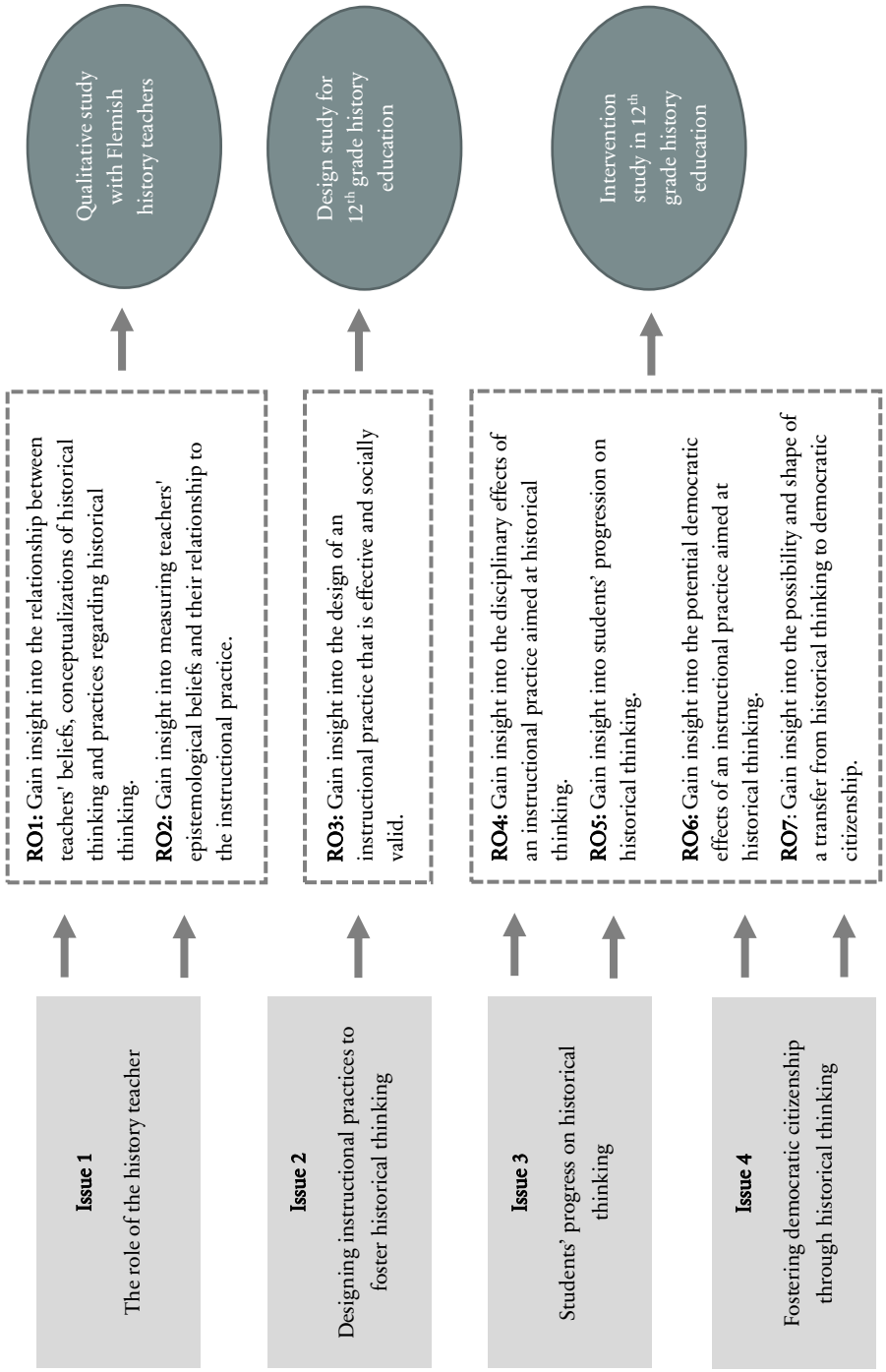
Lastly, the absence of mandatory in-service teacher training is an important contextual factor as it adds to the already substantial variety in teachers' professional background. Although professionalization initiatives organized by universities in the last two decades have focused to a greater extent on the concept and teaching of historical thinking, these initiatives do not equally reach all of the history teachers.

All these elements thus make for an interesting setting at the start of the curriculum reform with history teachers having varying degrees of familiarity with the concept of historical thinking as an educational goal and a highly diversified educational practice.

Research design

In this educational context, three main studies are conducted in order meet each of the research objectives. These studies provide the basis for the various chapters in this dissertation. Figure 2 depicts how each issue is connected to the research objectives and addressed via these main studies.

Figure 2. Overview of the relationship between the issues, research objectives (RO) and the main studies.



In light of the first two research objectives, a qualitative study is set up with 21 Flemish history teachers active in upper secondary education. Data are gathered via a questionnaire and two semi-structured interviews. Teachers are asked about their epistemological views, their own conceptualizations of historical thinking, their views about the desirability of historical thinking (including their beliefs about the goals of history education) and about the feasibility of historical thinking as an objective for history education (including the role of contextual factors). In order to gauge teachers' epistemological views, a questionnaire is combined with a case study with contrasting statements to evaluate both formal and practical epistemological beliefs, and to compare the results of both measurement instruments. Teachers' practices with regard to historical thinking are analyzed as well. To this end, teachers are asked to select and present their best practices regarding historical thinking.

In order to meet the first research objective, data from the questionnaire, teachers' interviews and teaching materials are analyzed qualitatively using a thematic analysis. Based on a within-case and cross-case analysis, the relationship between teachers' conceptualization of historical thinking, their epistemological beliefs, their beliefs about the desirability and feasibility of historical thinking, and their instructional practices are examined. These general patterns are further illustrated and explored via the discussion of three illustrative cases.

Research objective 2 zooms in on the role of teachers' epistemological beliefs and on methodological issues related to measuring them. To this end, teachers' answers to the statements in the questionnaire are analyzed quantitatively; the interviews and instructional practices are analyzed qualitatively using a thematic analysis. Based on this analysis, the relationship between teachers' formal and practical epistemologies, their understanding of historical thinking and their instructional practice is explored. A combined analysis of both formal and practical epistemologies, as well as an in-depth discussion of the results in the interviews, allows for a methodological reflection on how to assess teachers' epistemological beliefs.

The third research objective is addressed by means of a design-based study. Building on the results of the qualitative study with teachers, this study adopts a design-based approach to develop a lesson series for the 12th grade of secondary education, aimed at fostering students' historical thinking in an integrated way.

The aim is to tackle the obstacles identified by teachers with regard to designing instructional practices that foster historical thinking as well as the difficulties that emerged from the analysis of teachers' instructional materials. In addition, insights from research on effective teaching strategies for promoting historical thinking among students are integrated. This design-based study allows for the design to be conducted in close collaboration with teachers while taking into account evidence-based recommendations from existing intervention studies in history education. The design cycle is based on various iterations in which the designed materials are evaluated and adjusted. To this end, the materials are evaluated by history (educational) experts as well as experienced history teachers and teacher trainers. Throughout this iterative process, a 12 to 14-hour lesson series is developed on the theme of decolonization after 1945. This lesson series addresses both knowing and doing history and covers all four building blocks of historical thinking as described in the Flemish history standards. Although it adopts a holistic approach, it puts a particular emphasis on fostering students' historical inquiry competence. The theme 'decolonization after 1945' is chosen as the subject for this lesson series as it is a theme that is commonly addressed by history teachers in the 12th grade of history education and one that has been subject to changes in recent academic historiography. Such a topic hence speaks to teachers, but also provides an excellent opportunity to bridge the gap between school and academic history by ensuring a close connection to recent historiographical insights. The design-based study provides more insight into the design of instructional practices that are both effective in fostering historical thinking and are considered socially valid by teachers.

The lesson series which is developed throughout the design-based study provides the basis for an intervention study in secondary school history education which is designed to meet research objectives 4 to 7. This study is based on a large-scale controlled intervention study with 668 students in the 12th grade of secondary history education. By means of a pretest and posttest the disciplinary effects of the lesson series are evaluated and it is assessed whether improving aspects of students' historical thinking can indeed bring about a 'spontaneous' transfer to contemporary contexts. The pretest and posttest are based on a questionnaire and essay tasks measuring (1) changes in students' historical inquiry

competences and epistemological beliefs about history, (2) a possible transfer to democratic skills and dispositions.

Research objectives 4 and 6 are addressed via a quantitative analysis of the intervention study's data. Using multilevel modeling and linear regression, the effects of the lesson series are examined on students' historical inquiry skills, related procedural knowledge and epistemological beliefs (research objective 4). The quantitative analysis also studies whether improving students' historical inquiry competences contributes to particular democratic skills and dispositions (research objective 6). In this regard, it is examined whether improving students' historical inquiry competences affects their ability to critically examine evidence in order to take a reasoned and rational stance towards contemporary, societal debates (i.e., democratic skills), and whether it fosters an inclination to do so (i.e., democratic dispositions).

Building on the essay data gathered in the intervention study, a qualitative analysis is employed to meet research objectives 5 and 7. Via this qualitative analysis students' progression on historical thinking and the possibility of a transfer to democratic skills is examined via an in-depth exploration of the essays of a purposefully selected sample of students from the intervention group. In order to gain more insight into students' progression on historical thinking (research objective 5), and the role of students' initial historical thinking ability in this regard, a sample of 45 students is selected for whom the intervention was particularly effective in terms of improving their historical inquiry skills. These students' essays are analyzed qualitatively and progression in students' essays is compared between students with low and high initial scores on the essay tasks. This analysis provides insight into patterns of progress in students' historical inquiry skills and into differential progression according to students' initial ability.

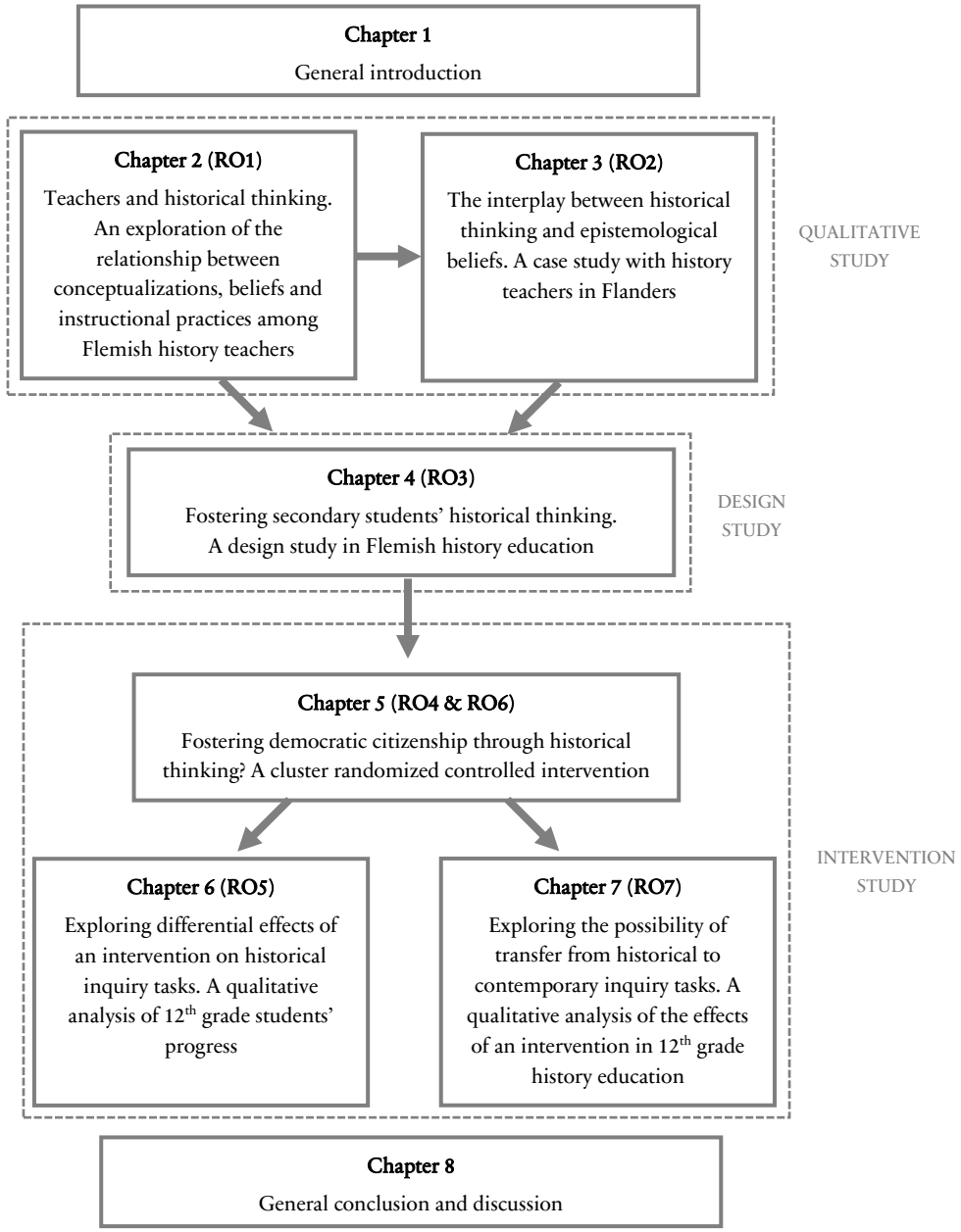
A similar qualitative analysis is employed for research objective 7. In order to explore whether students' historical and topical essays tasks show signs of a transfer from historical to democratic skills, the essays are examined of 23 students who advanced substantially on both historical and democratic skills after the intervention. This analysis describes in a detailed way how the progress on both tasks looks like and relates to each other, providing a better understanding of which aspects of historical skills may or may not transfer to contemporary contexts

as well as which obstacles occur in this regard. The study's findings shed light on the long-standing assumption of a spontaneous transfer from historical inquiry to aspects of democratic citizenship.

Structure of the dissertation

The dissertation is structured in eight chapters which are visualized in Figure 3.

Figure 3. Structure of the dissertation



The first, introductory, chapter sketches the background of the dissertation and the design of the studies. Chapters 2 to 7 report on the results of the qualitative study, design-based study, and intervention study. The results of these studies are analyzed in different ways across the various chapters, addressing each of the research objectives. Figure 3 provides an overview of how each research objective is addressed throughout the various chapters. Because each of those chapters is written as a distinct research article, intended to be understood in its own right, some overlap will occur between them[5]. The final chapter provides a general conclusion to the dissertation and reflects on the implications of the studies' findings.

[5] To improve the dissertation's uniformity, small adjustments were made to the original articles' lay-out, spelling and reference style.

Chapter 2

Teachers and historical thinking. An exploration of the relationship between conceptualizations, beliefs and instructional practices among Flemish history teachers.[1]

Historical thinking has become a central goal for history education, yet teachers' instructional practices are not always aligned with this aim. Teachers' beliefs are considered to be influential in this regard, yet it remains unclear how precisely they are related to their practice. This qualitative study with 21 Flemish upper secondary history teachers examines the relationship between teachers' conceptualization of historical thinking, beliefs about the nature of history and the teaching and learning of history and their instructional practice. The results show that this relationship is not a straightforward one as different beliefs are also related to each other and to teachers' conceptualization of historical thinking. The paper discusses these relationships and reflects on several recommendations for teaching training courses to engage with these results.

[1] This chapter is based on: Wilke, M. & Depaepe, F. (2019). Teachers and historical thinking. An exploration of the relationship between conceptualization, beliefs and instructional practices among Flemish history teachers, *International Journal for History and Social Sciences Education, 4,* 101-135.

INTRODUCTION

Historical thinking has become a main goal for history education in several countries, requiring students to gain both an understanding of the past and an understanding of history as a discipline (Lee, 2004; van Drie & van Boxtel, 2008). This ambition calls for interpretive history teaching, allowing students to become acquainted with the constructed nature of historical knowledge (Martell, 2013).

Research on history teachers' practices shows that they have difficulty doing so and that history lessons often focus predominantly on understanding the past and less so on understanding history (Barton & Levstik, 2004; Martell, 2013; Van Nieuwenhuyse et al., 2015). In explaining these findings, researchers point to teachers' beliefs about the nature and about the learning and teaching of history. In line with educational research in other domains, teachers' beliefs are expected to be influential to their instructional practice (Beswick, 2007; Depaepe & König, 2018; Nespor, 1987; Pajares, 1992; Thompson, 1984). Teachers' beliefs, however, do not function in isolation but are rather connected to each other within a belief system (Pajares, 1992). Furthermore, there is not a direct relation between beliefs and practices. Regarding mathematics teachers, for instance, Thompson (1984) established that contextual factors influenced that relationship. Moreover, teachers' experiences can provoke them to reconsider their beliefs, implying that beliefs and practices are 'dialectically related' (Cobb et al., 1990).

In the field of history education, studies often examined either teachers' beliefs (e.g., Virta, 2001; Yilmaz, 2008) or their instructional practice (e.g., Nokes, 2010; Van Nieuwenhuyse et al., 2015). When the relationship between both is examined, these studies mostly assume a direct relationship between beliefs and practices and rely mostly on self-reported rather than on actual instructional practices, although some small-scale case studies have combined both (e.g., Hartzler-Miller, 2001; VanSledright, 1996). Furthermore, little is known about how teachers conceptualize the notion of historical thinking, even though historical thinking has become a standard within the field of history education (Lévesque & Clark, 2018). Moreover, how teachers' understanding of the concept of historical thinking relates to their instructional practices has seldom been the specific focus. Considering its complex and multifaceted nature, this conceptualization might very well be connected to teachers' beliefs and

instructional practice. It hence remains unclear how these three aspects are related. This study therefore examines that relationship between conceptualization, beliefs and practices regarding historical thinking.

THEORETICAL BACKGROUND

The concept of historical thinking

Since the 1980-90s, historical thinking has gradually become the central goal for history education in the (Western) world (Lévesque & Clark, 2018). Historical thinking aims to introduce students to history as an academic discipline and to give them the tools to understand and critically evaluate accounts of the past.

Several models of historical thinking have been developed. Influential in this regard has been the work of Seixas and Morton (2013) and Wineburg (2001), who developed a conceptualization of historical thinking modelled on the practices of professional historians. Another influential model is that of van Drie and van Boxtel (2008) which was based on literature and adjusted through an analysis of students' actual historical reasoning. Although these models vary, they have in common that they aim to bring students to an understanding of both the past and the interpretive and constructed nature of historical knowledge (i.e., an understanding of history). They rely on several common building blocks characterizing the methods through which historical knowledge is constructed. Key aspects in this regard are the importance of asking historical questions, the elaboration of a historical frame of reference and accompanying selection of factual knowledge of significant historical phenomena, the need for historical contextualization, and the application of characteristic methods of historical reasoning in building historical accounts, such as taking into account multiple perspectives, and reflection on cause and consequence or continuity and change. Another important aspect is that of critical source analysis. This goes beyond extracting information from sources, but also aims at the assessment of the quality of the information, considering aspects such as the author's positionality, the source's context, its reliability and value, thereby gaining more insight in the constructed and interpretive nature of history (Van Nieuwenhuyse, 2020).

It is striking that little research exists on how teachers conceptualize historical thinking. Indirectly, amongst others via studies examining teachers'

conception of inquiry-based learning in history, it can be concluded that teachers hold different conceptions of historical thinking and that these influence their instructional practices to a certain extent (Voet & De Wever, 2016).

Teachers' instructional practices in view of historical thinking

Research on history teachers' instructional practices suggests that they do not regularly engage in interpretive history teaching. Teachers' practices focus predominantly on students' understanding of the past, rather than on their understanding of history (Martell, 2013; Van Nieuwenhuyse et al., 2015; VanSledright & Limón, 2006; Voet & De Wever; 2016, 2019). For instance, regarding source work, several studies reported that teachers used sources particularly to extract information, rather than apply a critical source analysis (Hicks, Doolittle & Lee, 2004; Nokes, 2010; Van Nieuwenhuyse et al., 2017).

Teachers' beliefs about the nature of history and history teaching and learning

To explain the limited attention to interpretive history teaching, researchers have studied teachers' beliefs as these are expected to influence instructional practices (Yilmaz, 2008). A number of studies have looked into teachers' beliefs about the nature of history, specifically those about the nature of knowledge and knowing, i.e., epistemological beliefs. Advanced epistemological beliefs are considered a prerequisite to teach about the interpretive nature of historical knowledge (Wansink et al., 2016; Yilmaz, 2008). In line with general educational research distinguishing between naive and advanced epistemological beliefs (Barzilai & Weinstock, 2015; Kuhn, Cheney & Weinstock, 2000) and working on the influential model of Maggioni, VanSledright and Alexander (2009), Stoel and colleagues (2017) distinguish between objectivist, subjectivist and criterialist ideas on history. In the objectivist (naïve) stance, knowledge is considered as fixed and singular and inherently embedded in sources. The subjectivist stance perceives knowledge to be uncertain and personal; hence historical claims are considered opinions. Criterialists (advanced stance) recognize that knowledge is constructed based on disciplinary methods and criteria. Respondents sometimes position themselves in-between two positions. This is the case with 'procedural objectivists' who believe that true and fixed knowledge can be acquired through a rigorous

application of historical thinking skills in constructing that knowledge (Stoel et al., 2017).

Studies on teachers' epistemological beliefs often take into account beliefs about the learning and teaching of history. Scholars, for instance, examined teachers' views on the learning goals for history education, revealing that teachers hold a diverse set of goals such as fostering civic behavior, clarifying the present or critical thinking. It is striking that in these studies historical thinking, although all the more identified as central aim for history education, is rarely named as such. Rather, the mentioned goals are related to fostering students' understanding of the past (McCrum, 2013; Virta, 2001; Voet & De Wever, 2016). Nevertheless, teachers do acknowledge the value of interpretive history teaching and the development of students' historical thinking skills (Hicks, Doolittle & Lee, 2004; McCrum, 2013; Wansink et al., 2017; Wils et al., 2011). A number of studies demonstrated the connectedness between teachers' beliefs about the nature of history and those about the teaching and learning of history, suggesting the existence of a belief system (McCrum, 2013; Voet & De Wever, 2016; 2019; Wansink et al., 2016).

Other research focused more on what might be labelled as teachers' beliefs about the feasibility of historical thinking and other goals for history education. It appears, for instance, that teachers consider interpretive teaching as requiring greater cognitive abilities from students and they therefore sometimes question their students' ability to engage in interpretive history (Moisan, 2010; Van Hover & Yeager, 2003; Wansink et al. 2016). Furthermore, teachers' perceived need to 'cover the content' and 'control the classroom', inhibits interpretive history teaching (Barton & Levstik, 2004; Van Hover & Yeager, 2003). Contextual factors, such as the available time and resources (Voet & De Wever, 2016), curricular requirements and the presence of standardized tests have also been found to cause teachers to refrain from interpretive history teaching (Hicks, Doolittle & Lee, 2004; VanSledright & Limón, 2006). This points again to the interconnectedness of teachers' beliefs and the complex relation with their practice.

Association between teachers' beliefs and practices

Several abovementioned studies provide evidence for the existence of a relationship between history teachers' beliefs and practices (McCrum, 2013; Voet & De Wever, 2019). However, this relationship is complex, and it remains unclear how instructional practices are affected by teachers' beliefs. For instance, while teachers value goals related to historical thinking, their practices are not always designed to accommodate these goals (Hicks, Doolittle & Lee, 2004; McCrum, 2013; Wils et al., 2011). Contrary to what might be expected, several studies illustrated that advanced epistemological beliefs do not guarantee increased attention for interpretive history teaching (Barton & Levstik, 2003; Hartzler-Miller, 2001; McDiarmid, 1994; VanSledright, 1996; Voet & De Wever, 2016; Wansink et al., 2016). While some studies concluded that teachers' appraisal of their students' cognitive abilities influenced their inclination towards teaching interpretive history (Moisan, 2010; Van Hover & Yeager, 2003; Wansink et al., 2016), a quantitative study by Voet and De Wever (2019) found this not to significantly impact teachers' inclination to engage in inquiry-based interpretive history.

Aims of the present study

Several limitations arise from the existing research. First, while some studies took into account actual teaching practices (e.g., Hartzler-Miller, 2001), conclusions regarding the relationship between teachers' beliefs and practices are often based on self-reported practices without examining teachers' concrete didactic materials or practices. Second, much research has not taken into account teachers' own conceptualizations of historical thinking, which can be assumed to differ greatly, given its complexity.

To overcome these drawbacks, the current study examines (1) teachers' conceptualization, (2) teachers' beliefs and (3) teachers' instructional practices regarding historical thinking to answer the following research question: how are history teachers' conceptualizations, beliefs and instructional practices regarding historical thinking related to each other?

Concerning beliefs, the research examines teachers' beliefs about the nature of history (i.e., epistemological beliefs) and about the teaching and learning of history. Because the latter encompasses a wide range of beliefs, this is limited to

two aspects. First, the research studies how teachers consider the importance of historical thinking in relation to other learning goals, i.e., their beliefs about the desirability of historical thinking as a learning goal. Second, the research examines teachers' beliefs about the feasibility of historical thinking as a learning goal (including perceived contextual hindrances and their perception of students' abilities). In what follows, the examined beliefs will hence be referred to as: epistemological beliefs, beliefs about the desirability of historical thinking and beliefs about its feasibility.

RESEARCH DESIGN AND METHODOLOGY

To gain insight into the relationship between teachers' conceptualization, beliefs and instructional practices, a qualitative study was set up with 21 history teachers, working in the 11th-12th grade of history education or having done so in the past three years. Data were collected between March and June 2019. Teachers were recruited through contacts of the teacher education program and pedagogical counsellors. To avoid a selection bias, the call for participation did not mention historical thinking but presented the research as focusing on general beliefs and practices. To encourage participants to speak freely, it was stressed that we were interested in their personal views and that the data would be anonymized.

Research context

This study was conducted in Flanders, the northern part of Belgium. History standards are defined by the Flemish Ministry of Education. They stress the need for students to gain an understanding of both the past and of history and relate this strongly to source analysis, yet this is not operationalized further. Although the underlying rationale is to focus on historical thinking, this is not explicitly stated as such (Flemish Ministry of Education and Training, 2000; Van Nieuwenhuyse, 2020).

The standards provide teachers with a considerable amount of freedom to choose their teaching content. They prescribe that the period of pre-history and ancient history should be studied in the 7th and 8th grade, the *Ancien Régime* (ca 500-ca 1800) in the 9th and 10th grade, while the 11th and 12th grade should be devoted to the period from ca. 1750 to the present. In each grade, aspects of

political, economic, social and cultural history should be touched upon. The main focus is on Western history, with some specific attention to the national past and the requirement to study at least one non-Western society in depth in each grade. Besides these guidelines, the standards do not prescribe any specific historical content. This freedom for history teachers is further reinforced by the absence of standardized testing or government control on the publication of history textbooks.

At the moment, new history standards are being implemented gradually, beginning with the 7th grade in September 2019. These revised standards identify historical thinking as the main goal for history education. Based on influential international models, the conceptualization of historical thinking in the standards emphasizes the need for students to acquire both an understanding of the past ('knowing history') and of historical practice ('doing history'). Historical thinking has been operationalized in four 'building blocks': situating historical phenomena in a historical frame of reference, selection and critical analysis of sources including reasoning with and about sources, formulating substantiated historical representations from multiple perspectives, and reflecting upon and interpreting the complex relationship between past, present and future (Agency for Higher Education, Adult Education, Qualifications and Study Grants, n.d.; Van Nieuwenhuyse, 2020). Similar to the previous standards, history teachers maintain a lot of freedom to select historical content.

At the time of the study, the new history standards were not yet officially communicated to a broad audience. Although the participating teachers could hence not be expected to be familiar with this model, the concept of historical thinking is addressed in teacher education programs, which pay attention to conceptual models for historical thinking (as mentioned earlier), as well as in professionally oriented journals and professional development training initiatives.

Participants

The sample consisted of 21 teachers with 1 to 38 years of teaching experience (mean = 16, *SD* = 12). Six teachers worked in general education, six in technical education and nine taught in both educational types. History education takes up two hours a week in general education, and one in technical education. Nevertheless, history standards for both are almost identical. All teachers held a master's degree in history and a teaching degree.

Table 1. Respondents' background characteristics

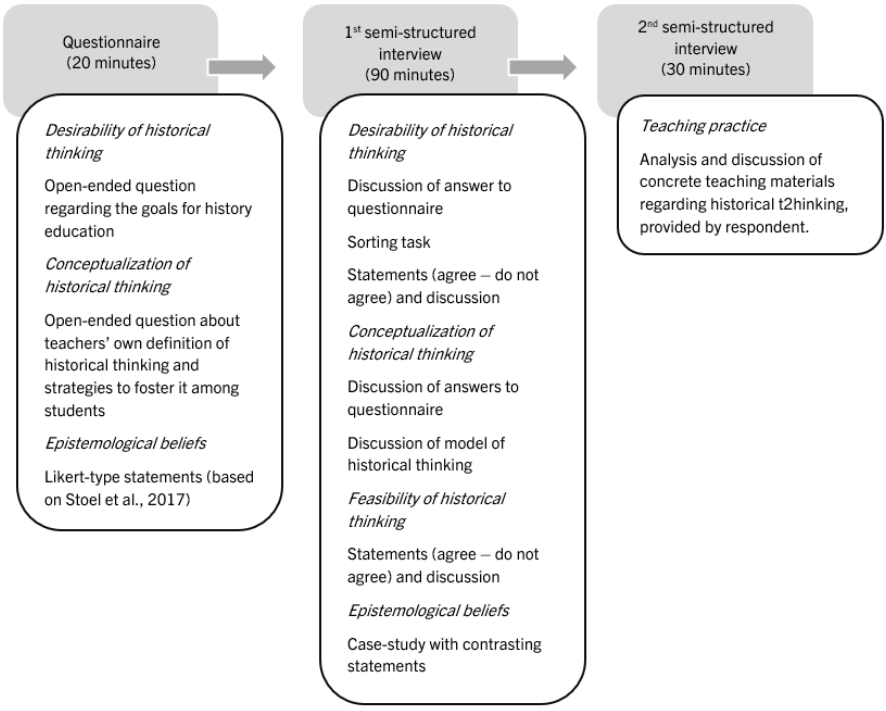
Respondent	Gender	Masters' degree in history + teaching degree?	Years of teaching experience	Currently teaching in the following educational type (general/technical)
R1	M	yes	25	Technical
R2	M	yes	7	Both
R3	M	yes	3	General
R4	F	yes	22	General
R5	M	yes	32	Both
R6	F	yes	15	Both
R7	F	yes	32	General
R8	M	yes	2	General
R9	M	yes	30	Technical
R10	F	yes	5	Both
R11	F	yes	39	General
R12	M	yes	4	Both
R13	M	yes	13	Technical
R14	M	yes	30	Technical
R15	M	yes	2	Technical
R16	M	yes	8	Both
R17	M	yes	1	Both
R18	M	yes	17	Technical
R19	F	yes	17	Both
R20	F	yes	8	General
R21	F	yes	25	Both

Data collection

To overcome the aforementioned limitations in the existing studies, data were collected regarding teachers' conceptualization, their beliefs and their practices related to historical thinking. The instruments were pilot-tested with three respondents, after which small adjustments were made.

Data regarding teachers' conceptualizations and beliefs were gathered through a questionnaire and a semi-structured interview. Data on teachers' instructional practices were collected through document analysis and a second semi-structured interview. An overview is depicted in Figure 1.

Figure 1. Data collection instruments



Data were gathered as follows:

Teachers' conceptualizations of historical thinking: In the questionnaire, participants answered the open-ended questions "how would you describe historical thinking" and "list five strategies you use to foster students' historical thinking abilities". Their answers were further discussed in the first interview.

Teachers' beliefs:

a. Epistemological beliefs: these were assessed through a set of statements in the questionnaire, based on Stoel and colleagues (2017). As those were initially directed at students, some statements were adjusted. To overcome methodological issues associated with the use of Likert-type statements to measure epistemological beliefs (Stoel et al., 2017; VanSledright & Reddy; 2014; Voet & De Wever, 2019), a case study with contrasting statements (about two causes for the spread of agriculture across Europe approx. 6000 BCE) was added as an additional measure (Kuhn et al., 2000).
b. Beliefs regarding the desirability of historical thinking as a learning goal: participants completed an open-ended question in the questionnaire, gauging their spontaneous ideas on the goals for history education. These answers were discussed in the first interview. A sorting task, asking teachers to select and arrange a pre-given set of goals, was used to corroborate findings and further explore the relative importance they assigned to historical thinking, compared to other goals. Lastly, teachers were asked to comment on related statements, such as "Historical thinking can be a central thread throughout the classes".
c. Beliefs about the feasibility of historical thinking (incl. perceived contextual hindrances and their perception of students' abilities): this was examined using an open-ended question and through the discussion of statements such as: "There is too little time to work on historical thinking in the history classes", "Addressing historical thinking implies I can cover less content" and "It is always possible to address historical thinking, regardless of the historical subject that is studied".

Instructional practices on historical thinking: after the first interview, participants were asked to provide concrete examples from their actual teaching practice in which they fostered students' historical thinking skills. In this sense the materials functioned as a best practice portfolio. By asking teachers to select and present these concrete examples additional information was gathered on teachers' practices and how these related to their beliefs. Because the provided materials were used in their classes, we considered them as providing data on teachers' actual

practices. Moreover, the selection of materials by the teachers in itself provided further insight into what they believe constitutes historical thinking and how it can be addressed in class. The provided materials were analyzed and discussed in the second interview to better understand teachers' rationale and purpose underlying these concrete materials (i.e., what learning goals did they try to achieve using this material, what kind of responses did they expect from students, how did this material relate to their overall teaching, and how often were these types of materials used throughout the year?).

Data-analysis

Data were analyzed qualitatively using NVivo software, through the use of a thematic analysis, both inductive, data driven, and deductive, based on a priori codes (Fereday & Muir-Cochrane, 2006). The units of analysis were one or several phrases related to one of the research themes.

The coding process was an iterative process based on multiple re-readings of the data. Data analysis started with an initial coding scheme, based on the research question. The main themes were: conceptualization of historical thinking, beliefs about the feasibility of historical thinking, desirability of historical thinking, epistemological beliefs, and instructional practices. This analysis was applied to the interview data and open-ended questions in the questionnaire. Analysis of the instructional practice was applied to the provided lesson documents and the second interview in which the concrete lesson documents were discussed.

Subcodes were created for several of the main themes. Regarding teachers' conceptualization and practices related to historical thinking, we operationalized historical thinking into five building blocks, based on a synthesis of the models of Seixas and Morton (2013) and van Drie and van Boxtel (2008) as well as the validated instrument designed for observing teaching practices through the lens of historical thinking of Gestsdóttir, van Boxtel and van Drie (2018). The building blocks allowed for a holistic and practice-oriented operationalization of historical thinking and fitted the model for historical thinking which is central in the Flemish history standards starting from September 2019 (see section 3.1). The building blocks were: asking historical questions, situating historical phenomena in a historical frame of reference, selecting and critically analyzing sources,

formulating a substantiated historical representation from multiple perspectives (with sub codes related to various modes of historical reasoning), and reflecting upon the complex relationship between past, present and future (including a reflection on the interpretive nature of historical knowledge) (Van Nieuwenhuyse, 2020). The coding related to the instructional practice was also sensitive to the quality of the teaching materials related to historical thinking. Regarding sources, for instance, it was examined whether they were used in an illustrative way, to extract information from, or for a thorough critical analysis with the latter being regarded as a more profound practice. Subcodes were also created for instances where the material reflected presentism or contained missed opportunities to enhance students' historical thinking.

For the main theme 'feasibility', subcodes such as 'student ability', 'practical' and 'tension content and historical thinking' were created, emerging from the data.

Regarding desirability, subcodes were created relating to various goals for history education, such as 'critical thinking'; 'citizenship education' and 'identity development' and for references to whether and why teachers considered historical thinking a desirable goal.

After testing the initial coding sets on the data from three respondents, some adjustments were made. The revised coding schemes were applied to a second set of interviews, confirming their appropriateness. A complete overview of codes can be found in Appendix 1.

The analysis resulted in the construction of a profile for each respondent, based on a within-case analysis (Miles et al., 2014), providing an in-depth view on each respondent's conceptualization, beliefs and practices and their relationship. A matrix was constructed indicating the quality of their conceptualization and instructional practice for each respondent, and the overall orientation of their beliefs. This was based on an assessment of all coding related to this theme. A cross-case analysis was conducted to look for general patterns throughout the cases and to examine differences (Miles et al., 2014).

RESULTS

First, the general results regarding teachers' conceptualization, beliefs and practice will be discussed. Second, the relationship between teachers' conceptualizations, beliefs and practices will be discussed in general and will then be further explored and illustrated throughout the discussion of three case-studies.

General results regarding teachers' conceptualizations, beliefs and instructional practices regarding historical thinking

Teachers' conceptualization, beliefs and practices were assessed and categorized into three levels, indicated by a plus, minus, or plus-minus sign. Regarding teachers' conceptualization, the categories indicate whether teachers had a profound understanding of the notion. Concerning teachers' beliefs about the desirability and feasibility, they specify to which extent teachers considered historical thinking an important and feasible goal. Teachers' epistemological beliefs were assessed in terms of how naïve or advanced they were; teachers were further grouped into three categories according to the quality of their instructional practice related to historical thinking. The results are summarized in Table 2. The precise meaning of each of these categories, as well as general results, will be discussed at length below the table.

Table 2. Teachers' conceptualizations, beliefs and practice related to historical thinking (HT)

Respondent	Conceptua-lization of HT	Beliefs about the desirability of HT	Beliefs about the feasibility of HT	Epistemo-logical beliefs	Teaching practice related to HT
R1	+	+	+/-	+	+
R2	+/-	+/-	+/-	+/-	+/-
R3	+	+/-	+/-	+	+/-
R4	+	+	+/-	+/-	+/-
R5	+/-	+/-	+/-	+/-	-
R6	+/-	+/-	+/-	+/-	+/-
R7	+/-	+	+/-	+	+/-
R8	+	+	+/-	+	+
R9	+/-	-	-	+/-	-
R10	+	+	+/-	+	+/-
R11	+/-	+	+/-	+/-	+/-
R12	+	+/-	+/-	+	+
R13	+	+	+/-	+	+
R14	+/-	-	-	+/-	-
R15	+/-	+	+/-	+	+/-
R16	**+/-**	+/-	+/-	+	+/-
R17	+	+	+/-	+	+
R18	**+/-**	-	-	+	+/-
R19	+/-	+	+/-	+	+/-
R20	+	+	+/-	+/-	+
R21	+	+	+/-	+	+

Teachers' conceptualizations of historical thinking

Teachers' conceptualizations of historical thinking were grouped into three categories, as shown in Table 2. A minus sign indicates the respondent did not refer to any aspect of the concept (this did not occur in the sample) and a plus sign indicates that the respondent had a profound understanding of the concept. This was the case, for instance, with teacher 10 who explained historical thinking as follows:

> Historical thinking starts with asking questions about the past, studying the past in its own right, thereby making use of sources which are assessed critically. It also means gaining an understanding of the difference between history and the past.

In this conceptualization several of the building blocks of historical thinking are mentioned (e.g., sources, asking historical questions, reflecting on the relationship between past, present and future) and these aspects are connected to each other in a holistic way. Ten teachers were attributed a plus sign.

Most teachers, however, were assigned a plus-minus sign. This indicates their conceptualization was either (1) incoherent, as no connection was made between various aspects of historical thinking, or (2) mentioned only aspects related to one of the five building blocks related to historical thinking, hence revealing a narrow, limited understanding of the concept. Teacher 7, for instance, explained historical thinking as "learning to deal with historical sources, to and analyze them and to deal with them in a critical way". During the interview she elaborated on this, demonstrating she had a well-developed understanding of critical source analysis. However, her conceptualization was limited to this one aspect only of historical thinking. Teacher 11 explained historical thinking in the questionnaire as "making sure students do not just learn factual knowledge but are able to interpret facts from the past". She added as an example that students should be able to explain how Belgian colonization has contributed to the current problems in the Democratic Republic of Congo. During the interview she added three more aspects, establishing a frame of reference, understanding the difference between cause and consequence and using sources. However, these aspects were not connected to each other but rather presented as separate activities. Both of these teachers were assigned a plus-minus, along with nine others whose

conceptualizations were similarly incoherent or limited.

Overall, teachers' conceptualizations were focused predominantly on source work, contextualization and only few historical modes of reasoning: in particular cause and consequence, and drawing analogies between past and present. Each of these were mentioned by over two thirds of the respondents. Other aspects of historical thinking were only included in a limited number of descriptions. For instance, eight teachers referred to the interpretive and constructed nature of historical knowledge, five stressed the importance of multiperspectivity and three referred to the notion of historical empathy. Several other typical historical modes of reasoning, such as reflecting on contingency or agency, deconstructing narratives, and drawing analogies between historical periods were absent in these conceptualizations.

With regard to how historical thinking should be addressed in the classroom, two more things stood out. First, the majority of teachers was convinced that historical thinking is only possible after students master historical content. They considered students' understanding of the past as a precondition for their understanding of history as a discipline and regarded both as separate activities. For instance, teacher 16 stated: "I would never start with historical thinking, I do think that you need to see some factual knowledge first". Second, a number of teachers related historical thinking predominantly to student-centered classroom activities, such as inquiry-based learning, and individual or group work. Teacher 9 for instance explained why he does not spend much time on historical thinking: "I like to have control and working in groups and so on, that is not my thing."

Teachers' beliefs about the desirability of historical thinking as a learning goal

When asked what teachers considered the main goals of history teaching, it is striking that historical thinking as such was rarely mentioned. Only three teachers named it explicitly. Most teachers, however, mentioned at least one related aspect: eight mentioned (historical) source work, seven the importance of the authors' historical positionality and six included aspects of historical reasoning, mostly the examination of (dis)continuity between past and present. Clearly, the relationship between past and present was important for the majority of teachers as the goal

most commonly referred to was to provide students with knowledge about the roots of present-day society and contemporary issues. Contrary to the spontaneous answers, historical thinking was selected by most teachers (except for three) in the card-sorting task, where they were given a set of possible goals. Teachers did not appear to consider it to be contradictory that they had not mentioned historical thinking spontaneously as they connected historical thinking to the development of broader educational goals, such as developing critical thinking, general information skills (e.g., looking up and summarizing information) and citizenship education. As teacher 15 explained: "Critical thinking in my opinion accompanies historical thinking, as does learning to ask questions, or to think and reflect independently about certain issues or conflicts."

Throughout the discussion of teachers' ideas about the goals for history education and the card-sorting task the extent to which teachers considered historical thinking – as they understood it – to be a priority goal in their teaching was assessed. This is indicated in Table 2. In some cases, when teachers stated explicitly that they did not consider historical thinking an important goal for history education, this was indicated with a minus sign. In other cases, teachers prioritized historical thinking, which was indicated with a plus sign. Mostly, however, teachers positioned themselves somewhere in-between: they valued historical thinking but did not consider it the main goal for history education or they conflated historical thinking with other goals to such an extent that historical thinking itself became instrumental to other goals, such as critical thinking or information skills. Teachers fitting this description were indicated with a plus-minus sign. Teacher 5 for instance explained his goals for history education as twofold:

> To allow students to gain a minimal insight into our society, how it functions and how it became this way and as a preparation for skills they will need in higher education, especially learning to deal with long, running text.

Historical thinking in itself is hence barely present in this answer, although it touches upon historicizing and making topical. Upon further discussion throughout the interview and in the card-sorting task, however, the teacher added several other goals such as critical and nuanced thinking and historical thinking.

In terms of historical thinking, he emphasized the need to understand past societies in their own way and to avoid generalizations or moral judgments of the past. In this way he considered historical thinking to be a way to develop students' critical and nuanced thinking. This teacher was hence ascribed a plus-minus sign regarding his beliefs about the desirability of historical thinking: he acknowledged its value as a learning goal, but mainly as a means to foster other goals which he considered to be more important.

Teachers' epistemological beliefs

In Table 2, teachers with a criterialist stance are indicated with a plus sign, while objectivist beliefs are indicated with a minus sign. Although teachers generally held advanced epistemological beliefs, in the in-depth discussion of the case-study, several teachers had difficulty explaining why historians can come up with different explanations for the same historical phenomenon; some revealed a more naïve stance or traces of "procedural objectivism" (Stoel et al., 2017, p. 131). Teacher 4, for instance, explained that "until it has been definitively researched, both explanations can be correct". These teachers' epistemological beliefs were marked with a plus-minus sign.

Teachers' beliefs about the feasibility of historical thinking as a learning goal

Only three teachers did not consider historical thinking to be a feasible goal. This was indicated in Table 2 with a minus-sign. All other cases were assigned a plus-minus sign: these teachers considered it overall feasible yet challenging. Teachers mentioned several hindering factors in this regard. First, contextual factors such as time restrictions were mentioned by all but three of the teachers. To a lesser extent, teachers mentioned the limited availability of didactic materials on historical thinking as a hindering factor. Second, about one third of the teachers expressed doubts about students' interests, motivation and capacity. In particular, respondents mentioned a perceived decline in students' language skills, which they considered necessary to engage with sources, and a lack of general, background knowledge. Teachers' perception of their own capacity constituted a third factor. Nine teachers stated they needed more support and guidance in teaching historical thinking. Almost half of the respondents spontaneously

reflected on aspects of historical thinking they felt could be improved in their teaching, yet they were confident they could achieve this themselves.

Teachers' instructional practices related to historical thinking

Most teachers provided several individual or group assignments for students which they designed specifically to foster students' historical thinking, some provided excerpts from their student syllabus. The materials were aimed at students in the 11th and 12th grade of secondary education (age 16-18), hence covering the period of ca. 1750 to the present. The materials addressed, among others, subjects such as the Industrial Revolution, modern imperialism, the First and Second World War, decolonization and the Cold War. The materials were predominantly designed by teachers themselves, sometimes adapted from textbooks. These were materials that were actually used in class and that provided insight into what kind of lesson texts were provided for students, how sources were used in class, what kind of questions students had to answer and so on. Throughout the materials, source work often occupied a central position. Besides source work the materials covered, to a lesser extent, other aspects of historical thinking, such as multiperspectivity, the deconstruction of historical representations, oral history or analogies between past and present.

The quality of teachers' instructional practices related to historical thinking varied strongly. In three cases, teachers' practices revealed no or only marginal traces of historical thinking. Sources were used sparingly and mostly as an illustration of the teaching content; the materials rarely encouraged students to think historically. These cases were indicated with a minus-sign in Table 2. Seven teachers' practices were assigned a plus. The materials of these teachers addressed several aspects of historical thinking in an in-depth way. In these materials, sources were subjected to a critical source analysis and students were introduced to the interpretive nature of historical knowledge, for instance through a research assignment in which students had to formulate a substantiated answer to a research question based on a critical analysis of historical sources.

The majority of teachers fell in an in-between category (+/-). Their materials addressed aspects of historical thinking but did so in a rather superficial way. For instance, sources were used and questioned but predominantly to extract information. Source analysis was sometimes guided by a fixed set of questions,

generating a rather mechanical approach in which students, for instance, rated the reliability of a source without reference to a particular historical question. In so doing, the source analysis was not really made meaningful as it did not reflect on the consequences of, for instance, the author's background on the source's content. Furthermore, the absence of sufficient historical context information impeded in-depth reasoning about the sources. Another example is an assignment on oral history which did not connect the personal account gathered by the students to more general historical developments and did not address the peculiarity and reliability of oral sources, such as the fact that they are established with hindsight. In an assignment on multiperspectivity, a teacher had students explore different viewpoints on the Israeli-Palestinian conflict but did not include a reflection on why these perspectives differed, how they influenced the source material or whether both perspectives were equally substantiated. In short, although these materials were designed to foster aspects of students' historical thinking, they contained a lot of opportunities to further students' historical thinking skills which were not seized.

The relationship between teachers' conceptualizations, beliefs and instructional practice

Table 2 reveals that it is not possible to identify a decisive or dominant factor to determine the quality of teachers' historical thinking practices. Rather, a belief system appears to be in place within which a complex interaction between various beliefs takes place. The following section therefore discusses the relationship between teachers' conceptualizations, beliefs and instructional practice related to historical thinking, both in a general way and through an in-depth exploration of three cases.

Teachers' conceptualizations of historical thinking

Table 2 suggests that a rich, comprehensive conceptualization of historical thinking is a prerequisite for a profound historical thinking practice. Teachers with a clear and broad understanding of the concept presented material which was diverse in terms of the historical thinking aspects being addressed and which often combined historical knowledge with a reflection on the constructed nature of this knowledge. Historical sources were subjected to an in-depth critical analysis,

allowing students to gain insight into the role which sources play in interpreting and constructing historical knowledge.

In most cases, however, teachers' conceptualizations were rather narrow. This was reflected in teaching practices which mostly addressed historical thinking superficially. For instance, source work was aimed only at extracting information from sources, rather than at critical analysis. Hence, even when teachers valued historical thinking as a goal, a narrow conceptualization could result in limited historical thinking practices (for instance respondent 15).

Teachers' ideas about how to foster historical thinking were also reflected in their practice. Particularly the perception, held by the majority of teachers, that historical thinking is only possible after students have acquired factual historical knowledge was important. It resulted in historical thinking activities that were isolated in separate, supplementary assignments, rather than being integrated throughout the teaching.

Teachers' beliefs about the desirability of historical thinking

Teachers' beliefs about the desirability of historical thinking were related to their practice in a number of ways. Three teachers did not consider historical thinking to be an important goal for history education. These respondents had difficulty providing concrete examples of materials addressing historical thinking.

In six other cases, teachers' beliefs about the desirability were more subtly reflected in their teaching. Although these teachers valued historical thinking, they prioritized other goals. For instance, respondent 16 subordinated historical thinking to goals such as information skills and gaining an understanding of the present. To illustrate his way of addressing historical thinking, he provided an assignment on source analysis regarding the Yugoslavian war. However, the assignment solely focused on extracting information from a set of sources rather than critically assessing them. It was hence more about information skills and making students familiar with a present-day issue. In this case, this was also related to a narrow conceptualization of historical thinking which he understood mostly as being about contextualizing and examining the historical roots of contemporary issues.

Some teachers experienced difficulty in combining their desire to address historical thinking with other goals as they conflicted somewhat. Some, for

instance, wished to provide a complete overview of the past, leaving them in a constant lack of time to work on historical thinking. Teacher 3, for example, valued historical thinking, but also stated that he wanted to provide a complete overview of the past: "Yes, that is the case, when you involve historical thinking, you are studying a topic more profoundly, but then there is less time for other themes." He hence only addressed historical thinking during short lesson moments in which students answered questions regarding a source. The majority of his teaching was dedicated to providing an overview of the past.

Seven teachers were strongly convinced that historical thinking was a priority goal which related positively to their teaching practices. These teachers aspired to address historical thinking regularly and stressed the need to integrate it throughout the classes, which was reflected in their practice. Furthermore, they experienced less time constraints as they emphasized the need to cover less content. Teacher 8 did not consider historical content a prerequisite for historical thinking because "when you are processing the content, you are *de facto* working on historical thinking". However, as can be seen in Table 2, considering historical thinking a desirable goal did not necessarily result in profound practices (e.g., teachers 10 and 15), which points to the influence of other beliefs.

Teachers' beliefs about the feasibility of historical thinking

Only in three cases did teachers consider historical thinking to be an unfeasible goal leading them not to address it in their practice or only barely. These teachers were convinced that their students were unable to think historically within the available amount of time, regardless of the educational type. This is also related to their beliefs about the desirability of historical thinking as a learning goal as all three teachers did not consider it a priority (or even a desirable) goal.

The remaining teachers considered historical thinking a difficult yet feasible goal. In most cases, teachers' concerns related to the feasibility of historical thinking as a learning goal were actually related to other beliefs. For instance, when teachers felt there was too little time to enhance students' historical thinking skills this was often prompted by their belief that historical thinking was only possible when students had acquired factual historical knowledge, so they focused first on teaching the content, leaving little time to address historical thinking. Similarly, some teachers lacked time because they considered it more important

to provide a complete overview of the past. In these cases, their beliefs regarding desirability were playing a role in their appraisal of its feasibility (see for instance teacher 3, as discussed above).

Teachers' epistemological beliefs

A combined analysis of teachers' answers to the questionnaire and discussion of the case-study revealed that some leaned towards more naïve epistemological beliefs or "procedural objectivism" (Stoel et al., 2017, p. 131). This ambiguity was reflected in their practice. Several teachers designed materials in which students had to critically analyze a historical representation (e.g., from a source or historical movie). Yet, this was done naïvely by juxtaposing 'subjective' sources to the 'objective' work of historians, suggesting that 'the (one and only) historical truth' exists and can be found. In another case, a teacher explained how she asked her students to underline 'objective' and 'subjective' aspects within a source, describing the former as occasions where the author expressed facts, the latter as instances where the author provided an opinion. These practices hence reflect and promote objectivist or 'procedural objectivist' views (Stoel et al., 2017).

Teachers with advanced epistemological beliefs and who considered it important to teach students about the interpretive nature of history often addressed historiographical debates or evolutions in historical representations, hence revealing the interpretive and constructed nature of historical knowledge. They also used critical source analysis to this end.

CASE-STUDIES

The following section discusses three cases illustrating the interaction between teachers' conceptualizations, beliefs and instructional practice. Teachers were selected based on the assessment of their teaching practice related to historical thinking. One teacher was selected from each category in Table 2, who most clearly articulated their beliefs and whereby the relationship between their conceptualization, beliefs and instructional practice was very explicit, making them best suited to illustrate the interplay. It will be explored how variations in teachers' practice are related to their conceptualization and beliefs.

Teacher 17 – high quality teaching practice related to historical thinking

This teacher's practice addressed many aspects of historical thinking in a profound manner, using rich and diverse teaching materials. Sources were deployed as a starting point to understand authors' perspectives, reflect on multiperspectivity, critically assess historical representations, practice historical empathy and historical reasoning. In working with sources, he applied critical source analysis to allow his students to generate their own substantiated answers to a historical question and allow them to gain a better understanding of the interpretive nature of history.

The teacher highly valued historical thinking as a central goal and ensured that other goals, such as critical thinking and understanding contemporary issues, reinforced each other rather than competed. He also emphasized the necessity to reduce the amount of subject matter to be covered in order to create time for historical thinking and stressed the need to integrate historical thinking throughout the lessons. Consequently, his concerns related to feasibility were mostly related to the fact that preparing materials focused on historical thinking was time-consuming:

> You need to find sources, for instance, and think about 'how am I going to process this, through which learning activity?'. This takes up a lot of time and effort and that is one of the biggest obstacles.

In describing historical thinking, the teacher stressed the importance of making students aware of the interpretive nature of history. His own advanced epistemological beliefs were in line with his practice. The materials frequently referred to the interpretive nature of history by including examples of conflicting accounts by historians, by addressing the construction and evolution of historical representations and by including critical source analysis.

Teacher 6 – superficial teaching practices related to historical thinking

Teacher 6 provided materials that touched upon various aspects of historical thinking, yet overall, they were characterized by a lack of depth. Several opportunities to engage students in historical thinking were not seized. The materials presented historical knowledge as fixed and singular, supporting

objectivist beliefs. It appears that her beliefs about the desirability and conceptualization of historical thinking, as well as her epistemological beliefs hindered this teacher from addressing historical thinking more profoundly.

The teacher did report that she considered historical thinking to be a desirable goal and actively looked for professionalization opportunities on the subject. However, her views on the goals for history education were marked by a strong present-orientedness. She explained history education as being about raising good citizens, including transmitting moral values, developing a critical attitude specifically regarding media literacy and gaining general knowledge about the present. She hence conceptualized historical thinking from a present-oriented perspective. For instance, analyzing and critically assessing sources was immediately connected to dealing with fake news; researching the past was reduced to gathering and presenting information. Hence, while she asserted that she considered historical thinking to be a desirable goal, she actually prioritized other goals. Furthermore, she regarded historical thinking as a didactic method that could be applied only after students had gained factual knowledge:

> You want to do something fun once in a while (…) I purchased a book on active historical thinking [*a book providing fun activities, promoting historical thinking*] and I would like to apply parts from this book, but the problem is, you need to cover the content first and only then can you apply the learning activities.

These ideas were reflected both in the shape and content of the teaching material. The materials focused mainly on 'what happened'. Aspects of doing history were addressed separately, mainly via source work. Sources were, however, used primarily as illustrations or to extract information from. Only on a few occasions was a source critically questioned, for instance when students were asked to judge a source's trustworthiness. Even then students lacked the necessary contextual information to adequately do so. Furthermore, sources were deployed to enhance present-oriented goals: teaching moral lessons and information skills. For example, one of the materials confronted students with a racist cartoon from the 1930s and asked them to criticize it. No reference was made, however, to the historical context in which this cartoon emerged.

Furthermore, the teacher's epistemological beliefs reflected notions of

'procedural objectivism' (Stoel et al., 2017): in her opinion, a rigorous approach to source analysis, as performed by academic historians, could reveal undisputed truths about the past. She explained, for instance, that "secondary sources are often (…) much more reliable" and that she used a video of an academic because "you can believe him as he is academically trained. He should stick to the facts; he should be objective." This view was reflected in her approach to source analysis, which focused on extracting 'objective information' from subjective sources. Other materials similarly reflected objectivist beliefs. In one example, multiple perspectives on decolonization were addressed, yet, as the teacher explained, the aim was for students to decide which of the sources was true.

Teacher 14 – historical thinking barely addressed

An analysis of the teaching materials revealed very little attention for historical thinking. Sources were used in an illustrative way and opportunities throughout the lessons to engage students in historical thinking were overlooked. The teacher's main goal was to provide a general overview of the past and to use the past to explain present-day phenomena, by pointing out similarities between past and present, while ignoring the differences. For instance, he used Cold War communism to explain the ideology of (extreme)left-wing parties today, without paying attention to the vast differences between both.

At first, this respondent seemed to provide contradictory statements regarding the desirability of historical thinking. He stated that "even if you just bring a chronological story, then you need to engage in historical thinking (…) it is the essence of history". However, he considered history education to be mainly instrumental for understanding current society. He focused on understanding the roots of contemporary issues, citizenship education, nuanced thinking and understanding different perspectives on present-day issues. In discussing the goals for history education, any reference to historical thinking was absent and the past as such was largely ignored.

The absence of historical thinking was influenced by the teacher's one-sided understanding of historical thinking as something that can only be addressed via inquiry-based learning. In his opinion historical thinking requires:

> A number of documents, that relate to each other (…) and a historical research question which you want to investigate and then (…) a provisional conclusion and then you add some more documents and then the conclusion changes, and then you can discuss this.

This particular conceptualization led this teacher to consider historical thinking as completely unfeasible for his students, because of a lack of time and because his students lack the capacity and motivation to do so: "Currently, students need to do this all the time: group work, independent work and 'look it up yourselves' (…) There is an overkill of that." Furthermore, he felt no desire to teach in this way because he liked to maintain control over the classroom and disliked classroom discussions.

The teacher acknowledged that he did not spend much time on historical thinking, yet it was clear that he had, over the years, carefully considered his role as a history teacher. Considering the little time he had with students, and their general disinterest, he pragmatically put forward another goal for history education, namely providing students with a historical frame of reference.

CONCLUSION AND DISCUSSION

This research aspired to gain insight into the relationship between teachers' conceptualizations, beliefs and practices related to historical thinking. To this end, it combined an analysis of teachers' conceptualizations, epistemological beliefs, beliefs about the desirability and feasibility of historical thinking with an analysis of their concrete teaching practices.

Some limitations should be taken into account. First, teachers were selected on a voluntary basis, which may have biased the sample. This was partly prevented by not mentioning the aim of the study in the call, but it remains possible that only highly motivated teachers answered the call. Second, only teachers who were active in upper secondary education were selected, creating a bias as these all held a master's degree in history. It would be interesting to see whether the results differ from teachers in lower secondary education with a professional bachelor's degree. Lastly, although the study corroborated survey and questionnaire data with concrete teaching materials, teachers, in most cases, only shared part of their teaching materials. Even though this was discussed in the second interview it was

not possible to fully determine to what extent the provided materials were representative of their general teaching practice or what relative weight they ascribed to historical thinking in the entirety of their classes.

Regarding the content of teachers' beliefs and instructional practice, the research findings are in line with earlier studies, confirming, most notably, teachers' difficulty in engaging in interpretive history teaching (Hicks, Doolittle & Lee, 2004; Van Nieuwenhuyse et al., 2017), their adherence to non-domain specific learning goals (McCrum, 2013) and the importance of perceived contextual hindrances (Van Hover & Yeager, 2003; Voet & De Wever, 2016). Some practices regarding source enquiry also reflected the findings of Voet and De Wever (2016) who found that teachers' conceptions "misrepresent the practices of inquiry that lie at the core of the discipline" and in that way "give their students the false impression that historical reasoning is mainly a matter of looking up information, or mechanistically assessing the reliability of sources" (p. 65).

The research shows that a relationship can indeed be discerned between teachers' conceptualizations, beliefs and practices regarding historical thinking. However, as the case studies clearly demonstrated, this relationship is complex.

In line with Thompson (1984), it was found that the relationship between history teachers' beliefs and instructional practices is not straightforward, as various beliefs interact with each other within a belief system. This relationship was especially prominent in teachers' beliefs regarding the feasibility of historical thinking. The most cited issue in this regard was a lack of time, often due to an alleged need to cover the content. This supports findings from earlier research (Van Hover & Yeager, 2003; Voet & De Wever, 2016), but is nevertheless surprising considering the freedom provided by the Flemish history standards to select teaching content. Teachers' views on feasibility should therefore be explained not just in light of curricular requirements, but particularly in light of other beliefs, such as teachers' desire to attain other goals, for instance providing a complete historical overview of the past. Besides other beliefs, teachers' conceptualization of historical thinking was also of interest in this regard, especially the idea that historical thinking requires a firm base of factual knowledge. This relates to a naïve distinction between understanding the past and understanding history. The perception that historical thinking is only possible after students have acquired factual historical knowledge was very established and

in a number of cases created a classroom practice which strictly separated the two.

Furthermore, a narrow or superficial conceptualization of historical thinking could sometimes explain why teachers who valued historical thinking as a goal nevertheless had rather superficial practices. Hence, studies examining teachers' beliefs and instructional practice should take into account the possibility that participants might have very different conceptions of historical thinking and related aspects, something which was also established by Voet and De Wever (2016), and should consider its impact on teachers' instructional practice. The addition of teachers' conceptualization of historical thinking therefore provides valuable information for understanding teachers' instructional practice and beliefs.

The research hence shows that there is no single explanatory factor, be it one aspect of teachers' beliefs or their conceptualization of historical thinking, to explain differences in teachers' instructional practice related to historical thinking. It seems imperative for researchers to take into account a multitude of beliefs, rather than studying only one (e.g., Yilmaz, 2008), as well as teachers' conceptualizations.

Overall, our findings make clear that only a minority of history teachers in this study were able to provide teaching materials addressing historical thinking in an in-depth way. Considering the potentially favorable bias in our sample and the fact that teachers were able to share their best practices, this is all the more troubling given the fact that historical thinking occupies a central position in the revised history standards in Flanders. This finding makes it urgent for teacher training courses and professional development activities to reflect on their role in promoting historical thinking as a central teaching aim. Following the suggestions by Van Hover & Yeager (2003), this research implies that any effort to change teachers' instructional practice should address and engage with teachers' underlying beliefs and conceptualizations. Several studies have demonstrated that well developed courses in pre-service teacher training can be successful in altering teachers' beliefs and practices related to the teaching of history, although they were not always effective for all participants and the extent to which these changes transfer to teachers' everyday teaching practice remains unclear (Gibson & Peck, 2020; Fallace, 2007; Monte-Sano, 2011; Voet & De Wever, 2017a).

In line with recommendations made by van Boxtel, van Drie and Stoel

(2020) it appears necessary for teachers to obtain a thorough understanding of historical thinking as a central concept in history education. Considering teachers' overall narrow ideas about what constitutes historical thinking, this seems to be a necessary starting point. Furthermore, teachers need to be supported in adequately operationalizing historical thinking.

A thorough conceptual understanding will enable a discussion and confrontation with teachers' own perceptions of the notion. In this regard, it is important that these are made explicit. It appears, for instance, that several teachers believe that a strong basis of factual knowledge is a prerequisite for historical thinking. Another recurring belief is that history education should provide a complete overview of the past. Both of these beliefs emphasize the importance of 'knowing history' over 'doing history'. If these beliefs are persistent, they might prevent teachers from developing a teaching practice which incorporates both 'knowing' and 'doing' history. Only when teachers' beliefs are identified and made explicit is it possible to engage with them and make clear, in this case, that a one-sided attention to 'knowing history' will leave little room to engage in 'doing history'. Such a dialogue makes it possible to challenge teachers' initial ideas and foster their understanding of how knowing and doing history relate to each other, why historical thinking has become a central goal for history education and how it relates to, and can support, other goals. A similar approach applies to teachers' epistemological beliefs. It is crucial that teachers develop advanced epistemological beliefs, not only in general, but also actively applied to concrete cases. This allows discussion of concrete implications of epistemological stances and enables epistemic change. It is imperative to explicitly discuss and challenge teachers' initial beliefs as this might generate changes in these beliefs and ultimately, classroom practice.

Regarding teaching practice, it needs to be emphasized that our study shows that a thorough understanding of historical thinking, as well as advanced epistemological beliefs appear to be essential and necessary yet insufficient conditions for a high-quality practice regarding historical thinking. In some cases, inconsistency between teachers' beliefs and practice could still be observed, indicating that favorable beliefs do not automatically translate into a more profound teaching practice. Thus, another important task for teacher training programs and professionalization initiatives can be discerned here. They should

not only engage with teachers' beliefs, but effort should also be put into assisting teachers in how to apply theory in practice, and in guiding and supporting them throughout the process of developing good practices with regard to the fostering of historical thinking. Ideally, working on establishing shifts in teachers' beliefs and conceptual understanding goes hand in hand with supporting the process of making changes in their teaching practice.

APPENDIX 1: CODING SCHEME

DESIRABILITY: whether and why the teacher considers historical thinking an (un)desirable goal

Goal: goals for history education:

- explaining the present
- citizenship education: aimed at civic knowledge and/or active engagement in society
- historical thinking: aspects related to historical thinking (e.g., critical source analysis)
- frame of reference
- values, norms, attitudes
- information skills: non-discipline specific (e.g., summarizing information)
- critical thinking
- identity development
- remembrance and knowledge: providing students with basic general knowledge about the past or present
- interest: provoking students' interest about the past
- learning from the past

FEASIBILITY: factors promoting/hindering addressing historical thinking

- students: referring to students' capacities, motivation, interest
- teacher capacity: reflection on teachers' own capacity to teach historical thinking
- practical: time, available resources, infrastructure
- tension content-HT: referring to tension between covering content and historical thinking
- didactic: didactic considerations such as evaluation, learning activities

CONCEPTUALIZATION: statements about what constitutes historical thinking and how to address it in class

- historical questions
- situating in frame of reference
- critical source analysis
- generating historical representations
    - critically assessing historical representations
    - historical modes of reasoning
        - multiperspectivity
        - analogies
        - agency
        - significance
        - causality
        - contextualization
        - continuity and change
        - historical empathy and perspective-taking
        - historicizing-making topical
    - reflecting on the relationship between past, present and future
        - past vs. history
        - interpretive nature of history
        - positionality

EPISTEMOLOGY: beliefs about the nature of history

GENERAL TEACHING APPROACH: statements about preferred teaching approach (e.g., teacher-centered) or regarding teaching practice not specifically related to historical thinking

PRACTICE: used for coding of concrete teaching materials

- historical questions
- situating in frame of reference
- critical source analysis
    - no contextualization
    - sourcing: only basic information about the source (e.g., author, date)
    - contextualization: rich contextual information about source
    - use illustrative
    - use extracting information
    - use critical analysis
- generating historical representations
    - critically assessing historical representations
    - historical modes of reasoning
        - multiperspectivity
        - analogies
        - agency
        - significance
        - causality
        - contextualization
        - continuity and change
        - historical empathy and perspective-taking
        - historicizing-making topical
    - reflecting on the relationship between past, present and future
        - past vs. history
        - interpretive nature of history
        - positionality
    - presentism: material reflects presentism
    - missed opportunity: instances where teacher addresses historical thinking but in a superficial way, small adjustments could further students' historical thinking

Chapter 3

The interplay between historical thinking and epistemological beliefs. A case study with history teachers in Flanders.[1]

History teachers' epistemological beliefs are considered to be greatly influential on their instructional practice and a necessary requirement to foster their students' historical thinking skills. In examining this relationship, two issues arise. First, adequately capturing teachers' epistemological beliefs remains a challenge as existing instruments appear not to be always valid and reliable. Some researchers suggest to distinguish between formal and practical epistemologies, which requires different measuring instruments. Second, it remains unclear how teachers' epistemological beliefs influence their teaching practice as several studies found there is no straightforward relationship due to the influence of other beliefs. At the same time, the role of teachers' own understanding of historical thinking in fostering this thinking among their students has not been extensively studied.

Through a qualitative research with 21 history teachers, this study examines the relationship between teachers' formal and practical epistemologies, their understanding of historical thinking and their instructional practice. It thereby reflects on methodological issues related to mapping teachers' epistemological beliefs.

Data analysis shows that teachers' epistemological beliefs remain difficult to capture, due to inconsistencies in and between measuring instruments. It could be concluded, nevertheless, that while most teachers acknowledge the interpretive and constructed nature of history, they generally do not include this in their own descriptions of historical thinking. The research supports the idea that nuanced epistemological beliefs are required for interpretive history teaching, but are not a sufficient precondition. While other beliefs and contextual factors are indeed at play, it also appears necessary to support teachers' competence in designing materials to foster their students' historical thinking, including epistemological reflection. The article reflects on the implications for teaching training and professional development programs.

[1] This chapter is based on: Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2022). The interplay between historical thinking and epistemological beliefs: a case study with history teachers in Flanders. *Historical Encounters, 9*(1), 190-214. https://doi.org/10.52289/hej9.111

INTRODUCTION

Epistemological beliefs touch upon the core of every discipline and its functioning. They include beliefs about the nature of knowledge, the (un)certainty of knowledge, and the nature or processes of knowing, related to sources of knowledge and the reasoning and justification processes involved (Buehl & Alexander, 2001; Hofer & Pintrich, 1997).

Epistemological beliefs play an important role in the discipline of history. For a long time, a 'realist' epistemological paradigm was the norm whereby historians believed it was possible to reconstruct the past exactly as it happened, by applying disciplinary methods. From the 1980s onwards, a shift occurred towards a perspectivist epistemology, paying more attention to the constructed and interpretive character of historical knowledge (Munslow, 1997). This perspectivist approach also became dominant in history education, which is particularly manifest in the central position occupied by historical thinking in history education in many Western countries (Lévesque & Clark, 2018).

Historical thinking combines 'knowing history' and 'doing history' in order to generate a deep historical understanding. Knowing history refers to the acquisition of substantive knowledge of the past; doing history refers to procedural knowledge in history, to building an understanding of how historical knowledge is constructed (Havekes et al., 2012; van Drie & van Boxtel, 2008). Building on earlier and novel research on historical thinking in the United Kingdom (conducted by scholars such as Denis Shemilt, Peter Lee, Alaric Dickinson and Rosalyn Ashby) various models have been developed that further operationalize historical thinking. One influential model is that presented by Seixas and Morton (2013) who distinguish six historical thinking concepts describing "the way historians transform the past into history" (p. 3). Another model focuses on historical literacy (Wineburg, 2001). In a European context, influential models are those devised by van Boxtel and van Drie (2008) and the model developed by the German FUER group, which distinguishes between four historical competences (Körber & Meyer-Hamme, 2015). Although these models differ in their concrete operationalization of historical thinking, they are based on a common element, namely revealing the interpretive and constructed nature of history by giving

students an insight into the methods of historians, through the use of sources and the application of (characteristic) historical methods of reasoning.

Historical thinking, however, constitutes an 'unnatural act' (Wineburg, 2001), implying that teachers play a particularly important role in students' learning process in this respect. The fact that teachers themselves have nuanced epistemological beliefs, in line with a perspectivist approach to history, is considered a necessary condition to enable them to convey this to their students and enhance their students' historical thinking skills (Wansink et al., 2016; Yilmaz, 2008). Various models exist which capture teachers' (and students') epistemological beliefs, both on a general and discipline-specific level. A number of influential models are presented in Table 1.

Table 1. Generic and domain-specific epistemological models (adapted from Maggioni et al., 2009, p. 196)

	Generic models		Discipline-specific models		
	King & Kitchener, 2002	Kuhn, Cheney & Weinstock, 2000	Lee & Shemilt, 2003	Nitsche, 2016	Maggioni et al. 2004; Maggioni et al. 2009
Naive beliefs: the past is given, knowledge is fixed	Pre-reflective	Realist Absolutist	Pictures of the past Information	Positivism	Copier stance
Knowledge about the past is uncertain, accounts are subjective	Quasi-reflective	Multiplist	Testimony Scissors and paste	Skepticism	Borrower/ subjectivist stance
Nuanced beliefs: the past is constructed and interpretive, based on disciplinary tools	Reflective	Evaluativist	Evidence in isolation Evidence in context	Narrative constructi-vism	Criterialist stance

Table 1 shows that these models use different names to describe epistemological beliefs. In general, however, we can observe a distinction between naive and more sophisticated, nuanced beliefs. People holding naive beliefs consider knowledge as fixed and singular, and ignore the existence of multiple perspectives and interpretations. Another instantiation of naive beliefs is the consideration of historical knowledge being purely subjective, thereby reducing history to a mere opinion. Advanced, nuanced beliefs only occur when history is considered as interpretive and constructed, when it is accepted that knowledge is subject to change and that claims about the past can be assessed based on disciplinary criteria (Kuhn et al., 2000; Maggioni et al. 2009; Stoel et al., 2017). They require a "coordination of the subjective and objective dimensions of knowing" (Kuhn et al., 2000, p. 311).

History teachers' epistemological beliefs are important in order to understand their instructional practice. It is therefore important to map them accurately. In this context, however, two difficulties arise. A first is related to measuring teachers' epistemological beliefs, a second to the relationship between epistemological beliefs and teachers' instructional practice.

Measuring history teachers' epistemological beliefs

A first difficulty is that it is complicated to map history teachers' epistemological beliefs. Maggioni et al. (2009) developed an instrument based on a synthesis of existing (progression) models and literature on the development of historical thinking (Maggioni et al., 2009). The Beliefs about Learning and Teaching History Questionnaire was initially administered to elementary teachers and college history professors. An adapted questionnaire (the Beliefs about History Questionnaire, or BHQ) was developed to assess students' epistemological beliefs (Maggioni, 2010) and has become a widely used instrument (Stoel et al., 2017). Although this is a validated questionnaire, several researchers experienced difficulties in replicating the results, and/or in assigning teachers, and students, to one of the three stances (Stoel et al. 2017; Voet & De Wever, 2019; VanSledright & Reddy, 2014). The items relating to the subjectivist stance, in particular, are problematic, because recognizing subjectivity can suggest both naive and nuanced beliefs (Stoel et al., 2017).

Therefore Stoel et al. (2017) tested an alternative questionnaire based among other things on items from the BHQ (Maggioni, 2010). They tested a simple distinction between naive and nuanced ideas, where items related to subjectivist beliefs would be divided between naive and nuanced ideas. Naive ideas were characterized among other things by statements referring to historical knowledge as fixed and singular, literally embedded in sources, or uncertain and personal. Nuanced ideas were related to statements describing historical knowledge as interpretive yet bound by disciplinary methods and criteria (Stoel et al., 2017). They conducted their research among students in the 12th grade, and also presented their statements to experts in the field of history. They were, however, unable to separate naive and nuanced subjectivist items; the factor analysis resulted in three subscales: on a naïve level, one subscale was labelled 'nature of knowledge – objective' and 'nature of knowing – naïve', both related to objectivist beliefs. On a nuanced level, no distinction was found between beliefs about the nature of knowing and those about the nature of knowledge. Instead, nuanced beliefs were captured in one subscale, labelled 'nature of knowing – nuanced'. Agreement with this subscale meant respondents valued disciplinary "procedures to interpret sources and construct reliable claims" (Stoel et al., 2017, p. 131). For this reason, the subscale was relabeled "historical methodology". Students were sometimes found to agree both with this subscale and with the subscale 'nature of knowing – naïve', indicating a "procedural objectivist" stance in which students "value historical-thinking skills because they believe that these skills make it possible to separate true and false sources and could generate true and fixed knowledge" (Stoel et al., 2017, p. 131). Interestingly, the two subscales related to naive, objectivist beliefs were answered as one factor by experts, and were both strongly rejected. Hence, they hypothesize that an increased epistemological understanding might be characterized both by an acceptance of items related to the importance of historical methods and by the degree to which they reject the naive, objectivist subscales related to the nature of knowing and the nature of knowledge.

Given these methodological difficulties, some researchers have called for the development of different measuring instruments. According to Kuhn and Weinstock (2002), epistemic thinking should be understood as "theory-in-action" (p. 134): peoples' 'tacit theories' about knowledge and knowing are activated when

they are confronted with a specific claim, problem and sources of information. A distinction should hence be made between formal and practical epistemologies (Sandoval, 2005; Sinatra & Chinn, 2012). Formal epistemologies refer to general ideas about the "characteristics of knowledge and its justification in a particular field" (Sinatra & Chinn, 2012, p. 264), while practical epistemologies refer to epistemic practices that are activated, for instance, via inquiry activities. Formal and practical epistemologies are not necessarily aligned. Sinatra and Chinn (2012) remark that beliefs about knowledge in general may be inconsistent with epistemic practice. Within the domain of history, formal epistemologies might be accessed via questions gauging respondents' conceptualization of history as a discipline (e.g., Yilmaz, 2008). Practical epistemologies refer to the actual reasoning and thinking that is at play when someone is engaged in a discipline-specific task, for instance when confronted with a historiographical debate or conflicting sources. Accessing practical epistemologies hence requires instruments that provoke reasoning about concrete problem scenarios (Barzilai & Weinstock, 2015). Kuhn et al. (2000) developed a short paper and pencil assessment based on conflicting statements on various knowledge domains, which they stated corresponded well enough with results from an interview-based assessment to justify its use.

The use of contrasting statements as a measure of epistemic thinking among history teachers has not been broadly researched. Yet several scholars emphasize that this might be beneficial in overcoming the issue related to measuring history teachers' epistemological beliefs, as well as to the need for studies triangulating different instruments (Stoel et al., 2017; Voet & De Wever, 2019). Finding a reliable and valid paper and pencil assessment for measuring epistemological beliefs, however, remains a major challenge to date.

Assessing the relationship between teachers' epistemological beliefs and instructional practice

A second difficulty is that it is unclear exactly how teachers' epistemological beliefs influence their instructional practice. Some studies have found a relationship between epistemological beliefs and instructional practice. Yeager and Davis (1996), for instance, found teachers who perceive history as a construction to be more likely to favor heuristic approaches, such as analysis and interpretation,

whereas teachers who regard history merely as a univocal account were more likely to use sources solely to extract information from. McCrum (2013) found teachers' recognition of the constructed nature of history to be associated with a more student-centered pedagogy and learning activities. However, several studies show that nuanced epistemological beliefs are not a sufficient condition for an instructional practice that pays attention to interpretive history teaching (e.g., Hartzler-Miller, 2001; VanSledright, 1996; Voet & De Wever, 2016, 2019).

Thus, there does not always appear to be a correspondence between history teachers' epistemological beliefs and their instructional practice. In this respect, Maggioni and Parkinson (2008) make mention of a "double epistemic standard" (p. 453) among teachers in different disciplines. While teachers might acknowledge the constructed and incomplete nature of disciplinary knowledge, they sometimes consider 'school knowledge' (the knowledge students encounter in school) as fixed and complete. Hence, they adhere to different formal epistemologies when it comes to disciplinary and school knowledge. Wansink et al. (2017) and McDiarmid (1994) found evidence for the existence of this double epistemic standard among students and prospective history teachers.

This incongruity highlights the importance of other factors that influence this relationship, and their connection in a belief system (Nespor, 1987; Pajares, 1992). Within history education, research has already been carried out into (prospective) teachers' beliefs about the goals of history, the teaching and learning of history and the influence of contextual factors such as the available time, teachers' access to didactic materials, curricular requirements and the presence of standardized tests (e.g. Barton & Levstik, 2004; 2003; Hicks, Doolittle & Lee, 2004; Husbands, 2011; Van Hover & Yeager, 2003; VanSledright & Limón, 2006; Voet & De Wever, 2019; 2016; Wansink et al., 2016). It shows that various beliefs and contextual factors interact with and influence instructional practice. Nitsche (2019) frames the relationship between theoretical and didactic beliefs about history and practice within a set of internal (e.g., beliefs about students) and external (e.g., curricular) factors.

One factor that is often ignored in these studies is teachers' own understanding of historical thinking. Since historical thinking is a broad and complex concept, it can be assumed that this understanding will vary. Research into teachers' conceptions of inquiry-based learning in history confirms this and

suggests that these understandings influence instructional practice (Voet & De Wever, 2016). It therefore seems important to map teachers' understanding of historical thinking as well. To what extent, for instance, do teachers consider epistemological reflection as part of historical thinking, and how does this relate to their epistemological beliefs and instructional practice?

In this latter respect, an additional difficulty arises in adequately capturing teachers' instructional practice. Some small case-studies rely on extensive classroom observations (e.g., Hartzler-Miller, 2001; Martell, 2013; VanSledright, 1996), while others rely on teachers' beliefs about instruction (e.g., Voet & De Wever, 2016) or self-reported practices, gathered through interviews (i.e., McCrum, 2013) or Likert-type statements (i.e., Hicks, Doolittle & Lee, 2004; Voet & De Wever, 2019). While lesson observations are very time-consuming, self-reported practices only provide an indirect insight into teachers' practice. Voet and De Wever (2016) therefore recommend complementing interviews with lesson observations or an analysis of the learning materials actually used by teachers.

Aims of this study

This study examines teachers' epistemologies, their understanding of historical thinking and their instructional practice, thereby taking into account the issues discussed above. The guiding research questions are:

- RQ1: What do teachers' formal and practical epistemologies look like and how do these relate to their understanding of historical thinking? This research question will also address methodological issues related to the measurement of teachers' epistemological beliefs.
- RQ2: How are teachers' epistemological beliefs and understanding of historical thinking reflected in their instructional practice?

RESEARCH DESIGN AND METHODOLOGY

A qualitative study was set up with 21 history teachers teaching in the 11th-12th grade of secondary education. Data collection took place during March-June 2019. The call for participation was distributed through contacts of the teacher training and pedagogical counsellors, but did not mention the specific research topic, so as

to avoid selection bias. The research was presented as an examination of history teachers' opinions and practices.

Research context

The study was conducted in Flanders, the northern part of Belgium. History education is part of the basic curriculum of general (2h/week) and technical (1h/week) secondary education. The standards, issued by the Flemish Ministry of Education, do not differ much between them. They assign teachers a lot of freedom: they do not impose specific content, yet only prescribe which time period should be studied in each stage (e.g., the 11-12th grade deal with the period ca. 1750-present) and require teachers to cover political, economic, social and cultural themes. Also, the government does not organize central examinations or exercise state control on history textbooks (Van Nieuwenhuyse, 2020).

The standards only set historical thinking implicitly as a goal: they do not conceptualize nor operationalize this notion. They try to connect to an academic, disciplinary approach of history yet in so doing bring a rather 'realist' approach to the fore, in which it is assumed that following disciplinary methods (rigorous source analysis) leads to the discovery of 'the' historical truth (Van Nieuwenhuyse, 2020). This is evident, for instance, in the fact that they emphasize the importance of critical source analysis, but do not offer concrete guidelines or didactic support on how to foster this, nor connect it to fostering epistemological reflection (Van Nieuwenhuyse et al., 2017). The standards date from 2000. Since 2019, new ones are gradually being introduced in which historical thinking occupies center stage and in which a perspectivist approach is dominant.

In order to teach history in the 11th-12th grade, both a teaching degree and a Master's degree in history is required. Hence, almost all participating history teachers are trained historians with a good understanding of history as a discipline and acquainted with historical research methods and historical thinking. Furthermore, in teacher training, professional development initiatives and professionally oriented journals, attention is paid to (different models of) historical thinking.

Participants

Twenty-one teachers participated in the research. Six of them worked in general education (2h history education a week), six in technical education (1h/week) and nine in both levels. General education primarily prepares students for higher education in universities or university colleges, whereas technical education prepares students for higher education in university colleges as well as for the labor market. Their teaching experience varied between one and 38 years. All teachers had a Master's degree in history as well as a teaching degree. One teacher also held a PhD in history.

Table 2. Descriptive information on respondents

Respondent	Gender (male/female)	Years of experience	Currently teaching in the following educational type (general/technical)
1.	M	25	Technical
2.	M	7	Both
3.	M	3	General
4.	F	22	General
5.	M	32	Both
6.	F	15	Both
7.	F	32	General
8.	M	2	General
9.	M	30	Technical
10.	F	5	Both
11.	F	39	General
12.	M	4	Both
13.	M	13	Technical
14.	M	30	Technical
15.	M	2	Technical
16.	M	8	Both
17.	M	1	Both
18.	M	17	Technical
19.	F	17	Both
20.	F	8	General
21.	F	25	Both

Data collection

Data on teachers' general and practical epistemologies, their understanding of historical thinking and their instructional practice were collected via a questionnaire and two semi-structured interviews, which had been pilot-tested beforehand and adjusted as necessary.

Teachers first completed an online questionnaire containing two open-ended questions to gauge teachers' understanding of historical thinking: "How would you describe historical thinking" and "List five strategies you use to foster students' historical thinking abilities". The questionnaire assessed teachers' formal epistemologies through statements derived from Stoel et al. (2017). Respondents assessed fifteen statements on a 6-point Likert scale from "strongly disagree" to "strongly agree". Statements included among others "It is not possible to write adequately about history when sources contradict each other" (naive) and "In history education it is important that you learn to support your reasoning with evidence" (nuanced). Considering that the statements assessed teachers' general ideas about knowledge (construction) in history, we considered them to be an adequate measure for their formal epistemologies. As these statements were initially aimed at students, some small adjustments were made. For instance, in the statement "When something is written in your history textbook, you can assume it is true", the phrase "your history textbook" was changed to "a history textbook".

In the first interview, respondents' answers from the questionnaire were discussed. Teachers were asked to elaborate on their answers to the open-ended questions. After discussing their own understanding of historical thinking, they were presented with a model which defined historical thinking as requiring "an understanding of both the past and historical practice". That model (see Appendix A) consisted of five components, based on a synthesis of existing models of historical thinking (Seixas & Morton, 2013; Van Drie & Van Boxtel, 2008; Wineburg, 2001). Respondents were asked to what extent this related to their own understanding of the concept and their instructional practice (e.g., Did they recognize the aspects of historical thinking presented in the model? Were any of them new to them?).

The first interview also gauged teachers' practical epistemologies. To this end, a scenario-based approach was developed using a case study with contrasting

statements. The measurement instrument required teachers to reason about a concrete, discipline-specific task, thereby providing an insight into their practical epistemologies. The measure used contrasting statements based on Kuhn et al. (2000) but added a content-rich scenario with sufficient background information, following the suggestion and work of Barzilai and Weinstock (2015). The case presented two causes for the spread of agriculture across Europe ca. 6000 BC and presented respondents with the following statements: "Only one explanation can be correct" or "Both explanations can be correct". If they selected the latter, they were offered a choice between two additional statements: "Both are equally correct" or "One can be more correct than the other". Respondents' answers were further discussed by asking them to explain their choice and by follow-up questions such as "How can you explain the fact that different scholars come to different conclusions?". If teachers' answers to the statements in the questionnaire contained discrepancies, this was also discussed in the interviews. Teachers with nuanced beliefs were expected to indicate that both explanations could be correct, but that one could be more correct than the other.

After the first interview, teachers were asked to provide concrete, self-developed teaching materials, drawn from their actual teaching practice which they used to promote their students' historical thinking skills. In so doing, we provided teachers the opportunity to present what they considered to be their 'best practices' in terms of historical thinking. Teachers were asked to present the material in the second interview and to explain their rationale and teaching purposes related to the material. Questions addressed aspects such as what learning goals (including historical thinking aspects) teachers hoped to achieve through these materials, what kind of student answers they expected, how often these kinds of materials were used and how the material related to their overall instructional practice. To ensure that we were able to go beyond mere self-reported practices, we applied our own analysis of the materials after the interview. An overview of the analytical concepts in relation to the data collection instruments is presented in appendix B.

Data analysis

Data from the Likert-type items assessing formal epistemologies were analyzed quantitatively. For each respondent a mean score was calculated for each of the three subscales. Data from the interviews as well as the teaching material were analyzed qualitatively, using a thematic analysis. Codes were created inductively, data-driven and deductively. A priori codes were created for the main research themes: understanding of historical thinking, epistemological beliefs and instructional practice. An initial coding scheme was created by the first and third author. The first author conducted an initial analysis of the data and created additional, data-driven codes. The codes were discussed and defined in deliberation with the second and third author. The revised coding scheme was then applied to the data by the first author, after which the analysis was reviewed and discussed with the second and third author, leading to consensual agreement.

For the research theme 'epistemological beliefs', subcodes distinguished between formal and practical epistemologies. In line with the research by Stoel et al. (2017), we employed a simple distinction between naive and nuanced epistemological beliefs in the subcodes, for both the formal and practical epistemologies.

Subcodes were created for the research theme 'understanding of historical thinking', to label various aspects of the concept in teachers' own descriptions. They included, among others, asking historical questions, historical modes of reasoning (e.g., causality, significance, agency, continuity and change) and source analysis. This coding was based on a synthesis of various historical thinking models (Seixas & Morton, 2013; van Drie & Van Boxtel, 2008; Wineburg, 2001) as well as the validated historical thinking observation instrument of Gestsdóttir, Van Boxtel and van Drie (2018).

For the theme 'instructional practice', a first set of codes identified various aspects of historical thinking in the material. Codes included among others: source analysis, multiperspectivity, asking historical questions, analyzing historical representations. As regards source analysis, subcodes were used to analyze how sources were presented and questioned, indicating for instance whether sourcing, contextualization and/or corroboration were present (Wineburg, 2001), as well as reasoning with and/or about sources (Rouet et al., 1996). A second set of codes labelled instances where opportunities to stimulate

epistemological reflection were present or were missed. If the material did address epistemological reflection, subcodes indicated whether this reflection stimulated either nuanced of naive beliefs. For instance, if a teacher designed materials aimed at critical source analysis, but questions in the materials were only aimed at extracting information from the source (i.e., only reasoning with sources), this was coded as naïve. When the questioning of sources fostered students' understanding of history as interpretation and construction, this was labelled as nuanced.

A within-case analysis was applied to create a profile for each respondent, describing their epistemological beliefs, understanding of historical thinking and instructional practice as well as their relationship. General patterns in the data were identified using a cross-case analysis (Miles et al., 2015).

RESULTS

RQ1: What do teachers' formal and practical epistemologies look like and how does this relate to their understanding of historical thinking?

Formal epistemologies

Based on teachers' answers to the statements in the questionnaire, an average score (out of 6) was calculated for each of the three subscales. The results for each respondent are shown in Table 3.

Table 3. Respondents' mean scores on subscales assessing formal epistemologies

	Naive epistemological beliefs		Nuanced epistemological beliefs (advanced)
Respondent	Nature of knowing (naive)	Nature of knowledge (objective)	Nature of knowing – nuanced/historical methodology
1.	1.25	1.40	6.00
2.	1.75	1.80	5.17
3.	2.25	1.60	4.17
4.	3.25	3.00	4.33
5.	2.00	2.60	5.33
6.	3.75	3.00	4.67
7.	1.75	2.20	5.83
8.	1.50	2.20	5.33
9.	3.00	2.40	4.67
10.	2.75	3.00	4.67
11.	3.50	3.00	4.00
12.	2.25	1.00	5.67
13.	2.50	2.20	5.00
14.	2.50	3.00	5.00
15.	1.50	1.20	5.83
16.	2.00	2.00	5.33
17.	2.00	1.80	5.33
18.	1.25	2.20	5.33
19.	3.75	3.20	4.83
20.	2.50	2.20	5.33
21.	2.00	1.80	4.83

A high score on the first two subscales (columns 1 and 2) indicates rather naive beliefs, both with regard to the nature of knowing and the nature of knowledge. A high score on the last subscale (column 3) indicates nuanced beliefs. Stoel et al. (2017) do not provide specific cut-off points indicating when a score can be considered high or low, which makes interpretation somewhat difficult. However, we can assume that a score above 3 indicates that a respondent is more likely to agree with the statements in a certain subscale and a score below 3 that the respondent tends to disagree.

The data in Table 3 show that, broadly speaking, all participants largely agreed with items from the subscale 'historical methodology', which indicates nuanced beliefs. Items related to naive beliefs were generally rejected. When looking in more detail, however, two groups emerge. The first group consists of respondents who assigned a very pronounced score to the nuanced and naive subscales. These respondents strongly agreed with the statements related to nuanced beliefs and strongly rejected statements related to naive beliefs. This is the case with the majority of respondents. In a second, smaller, group the difference was less pronounced. These respondents also agreed with the statements indicating nuanced beliefs, but at the same time gave an above-average score to the statements related to naive beliefs (e.g., respondents 4, 6, 11 and 19).

A second observation is that, although all respondents generally agreed with statements related to nuanced beliefs, some deviated from this pattern for one or more statements. Where this was the case, these statements were discussed in the interview in order to ascertain the reasoning underlying the teachers' responses. This generally revealed two things.

First, the discussion revealed that some teachers merely wanted to nuance the statement or had understood it differently. Teacher 12 indicated that he 'agreed' with the statement "If eyewitnesses disagree, you cannot discover what actually happened", which was striking considering that his answers otherwise very unambiguously suggested nuanced beliefs. However, he explained his answer by referring to a specific topic in ancient history, where the source material is sometimes so limited that it is indeed difficult to find out what happened. This teacher subsequently spontaneously linked this to the need to study the credibility of the authors and to weigh up the arguments, and thus, to the disciplinary methods of historians.

Second, further discussion of certain statements suggested that teachers applied a 'didactic frame' when answering the statements, and kept their students and teaching practice in mind. Teacher 2 explained why he 'rather disagreed' with the statement "A good history account discusses different views of the past". He stated: "I see this from my work as a history teacher. If I make it too complicated, I lose them (the students) completely and then they remember nothing of it". As a history teacher he therefore saw it as his task to consider "what is the most probable at that point in time and then to present it clearly as a univocal, structured account". A similar mechanism was at play in the reasoning of teacher 6. She 'rather agreed' with the statement "You can't write well about the past when sources contradict each other" but clarified in the discussion that she found it especially difficult to teach about the past when sources contradict each other. She pointed to the level of difficulty for her students and the limited teaching time, which makes it difficult to paint a nuanced picture of the past.

On a methodological level, these examples show that caution is needed when interpreting the results of these kinds of statements, as they seem to measure more than epistemological beliefs. First, this is because, when considering these statements, teachers allow other beliefs about the teaching and learning of history, as well as contextual factors, to influence their answer. The finding that, on an academic level, teachers are aware of the constructed nature of historical knowledge, but in their teaching prefer to adhere to a univocal and factual historical account, connects to Maggioni and Parkinson's (2008) notion of a double epistemic standard. Second, caution is needed because it remains unclear precisely how teachers interpret such statements. Although some respondents disagreed with items related to nuanced beliefs, it became clear in the discussion that they did actually hold nuanced beliefs.

To obtain a more accurate picture, therefore, it is important to complement these results with those of the case study involving contrasting statements.

Practical epistemologies

Practical epistemologies were mapped using a case study with contrasting statements. Two theories were presented about the spread of agriculture ca. 6000 BC. Teachers were asked whether both theories could be correct and, if so, whether one could be more correct than the other.

Teachers answered the contrasting statements in two different ways. One teacher indicated that both theories were equally correct. This answer indicates naive beliefs. All the other respondents replied, in line with nuanced epistemological beliefs, that both statements could be correct, but one could be more correct than the other. Thus, based on these answers, almost all respondents seemed to have nuanced epistemological beliefs. However, the discussion of the answers to the case study generated more ambiguous results. Based on the argumentation put forward by the respondents, two categories were distinguished.

A first category consisted of respondents whose explanation confirmed the result of the contrasting statements. These respondents provided an argumentation for their response which clearly pointed to nuanced epistemological beliefs. They directly or indirectly mentioned the constructed and interpretive character of history, as they explained the existence of different theories by referring to differences in the source material which was available or was examined, differences in the interpretation of the source material or in the background or positionality of the researchers. Moreover, these teachers emphasized the importance of argumentation and evidence that could help determine which theory might be more or less plausible.

A second category consisted of teachers whose explanation somewhat questioned the result of the contrasting statements, either because of a lack of good argumentation, or because the argumentation leaned towards rather naive beliefs. Teacher 14, for instance, initially started by saying that "Both could be right; however, these are actually two opposite theories, so I think it is one or the other". Nevertheless, he concluded that both theories could still be right, by theorizing about ways in which both phenomena could occur alongside each other. Rather than explaining why one theory could be more correct than the other, this respondent tried to reconcile both theories and thus somewhat contradicted himself. A similar reasoning was expressed by teacher 2. Other respondents emphasized the importance of historical methodology so strongly that they appeared to lean towards procedural objectivism (Stoel et al., 2017). For example, teacher 4 stated that both theories can be correct “until it has been definitively researched”; teacher 20, who initially did refer to the positionality of historians, subsequently stated that one theory could be more correct because “one method

may be better executed (…): more neutral, objective and scientific". Based on the explanations provided by these teachers, it is questionable whether they do actually hold nuanced epistemological beliefs.

Table 4 shows for each respondent whether the answer to the contrasting statements was in line with nuanced or naive beliefs and whether this was confirmed by the explanations they provided for their answers.

Table 4. Respondents' answers and discussion of the case study measuring practical epistemologies

	Formal epistemologies			Practical epistemologies	
Respondent	Nature of knowing (naive)	Nature of knowledge (objective)	Nature of knowing – nuanced/ historical methodology	Case-study: contrasting statements	Does the discussion of the case-study confirm the result of the contrasting statements?
1.	1.25	1.40	6.00	Nuanced	Yes
2.	1.75	1.80	5.17	Nuanced	No
3.	2.25	1.60	4.17	Nuanced	Yes
4.	3.25	3.00	4.33	Nuanced	No
5.	2.00	2.60	5.33	Nuanced	No
6.	3.75	3.00	4.67	Nuanced	No
7.	1.75	2.20	5.83	Nuanced	No
8.	1.50	2.20	5.33	Nuanced	Yes
9.	3.00	2.40	4.67	Naive	Yes
10.	2.75	3.00	4.67	Nuanced	Yes
11.	3.50	3.00	4.00	Nuanced	No
12.	2.25	1.00	5.67	Nuanced	Yes
13.	2.50	2.20	5.00	Nuanced	Yes
14.	2.50	3.00	5.00	Nuanced	No
15.	1.50	1.20	5.83	Nuanced	Yes
16.	2.00	2.00	5.33	Nuanced	Yes
17.	2.00	1.80	5.33	Nuanced	Yes
18.	1.25	2.20	5.33	Nuanced	Yes
19.	3.75	3.20	4.83	Nuanced	No
20.	2.50	2.20	5.33	Nuanced	No
21.	2.00	1.80	4.83	Nuanced	Yes

Table 4 shows that almost half the respondents gave answers to the contrasting statements that were in line with nuanced beliefs but were unable to substantiate this properly. It seems as if teachers knew on a general, abstract level that history is based on interpretation, but found it difficult to apply this to a concrete case. This only emerged, however, when they were asked to explain their answers. If we compare these results to those of the formal epistemologies, it appears that teachers who have difficulty providing an adequate explanation for their answer are also those who, in addition to a high score on the nuanced subscale, assigned relatively high(er) scores to the other scales (e.g., teachers 19 and 11), although exceptions occur (e.g., teacher 2).

This again raises methodological questions, because, based on these results, it seems that the contrasting statements are insufficient to identify variations in epistemological beliefs. Moreover, it is unclear to what extent practical epistemologies are actually applied when answering the contrasting statements, as teachers' answers seemed to be based primarily on general beliefs about knowledge of the past. The practical epistemologies mainly appeared to come into play when they were explaining their answers.

Teachers' understanding of historical thinking

Introducing students to the practice of historians, and consequently to the interpretive and constructed character of history, is an essential part of various models of historical thinking. The question then arises as to what extent teachers themselves see this as a part of historical thinking, and spontaneously include it in their own descriptions of the concept. Table 5 shows for each respondent whether their description of historical thinking included a reference to epistemological reflection.

Teachers' own descriptions of historical thinking in the questionnaire and in the interviews show that epistemological reflection was spontaneously mentioned by a minority of the respondents. For example, teacher 4 described historical thinking as "making students aware that (...) historians reconstruct the past based on sources and that these sources give 'an image' of the past". Although most teachers, especially at a formal level, seemed to have nuanced epistemological beliefs, these were not strongly reflected in their descriptions of historical thinking. When epistemological reflection was mentioned, this was

usually done by teachers who seemed, based on the statements and the case study, to have nuanced beliefs. A pronounced presence of nuanced epistemological beliefs therefore seems to be consistent with a more complete understanding of historical thinking.

Aspects of historical thinking which were mentioned by the majority of respondents were: working with sources, certain historical modes of reasoning and contextualization. Hence, teachers do actually refer to various aspects of introducing students to the disciplinary methods and in that way involve epistemological reflection in their understanding of historical thinking, albeit in an implicit way.

In order to gain a better understanding of teachers' ideas about historical thinking, respondents were confronted with an existing model and conceptualization of historical thinking, which included epistemological reflection. Teachers indicated that they generally recognized and agreed with this description of historical thinking, even when it was much broader than their own descriptions. This suggests that they have a fairly tacit understanding of the concept. The question therefore arises as to whether this permeates their instructional practice and whether they address epistemological reflection in that practice.

Table 5. Epistemological reflection in teachers' understanding of historical thinking

	Formal epistemologies			Practical epistemologies		Understanding of historical thinking
Respondent	Nature of knowing (naive)	Nature of knowledge (objective)	Nature of knowing (nuanced/historical methodology)	Case-study: contrasting statements	Does the discussion of the case-study confirm the result of the contrasting statements?	Constructed and interpretive nature of history mentioned as part of historical thinking?
1.	1.25	1.40	6.00	Nuanced	Yes	Yes
2.	1.75	1.80	5.17	Nuanced	No	No
3.	2.25	1.60	4.17	Nuanced	Yes	Yes
4.	3.25	3.00	4.33	Nuanced	No	Yes
5.	2.00	2.60	5.33	Nuanced	No	No
6.	3.75	3.00	4.67	Nuanced	No	No
7.	1.75	2.20	5.83	Nuanced	No	No
8.	1.50	2.20	5.33	Nuanced	Yes	Yes
9.	3.00	2.40	4.67	Naive	Yes	No
10.	2.75	3.00	4.67	Nuanced	Yes	Yes
11.	3.50	3.00	4.00	Nuanced	No	No
12.	2.25	1.00	5.67	Nuanced	Yes	Yes
13.	2.50	2.20	5.00	Nuanced	Yes	No
14.	2.50	3.00	5.00	Nuanced	No	No
15.	1.50	1.20	5.83	Nuanced	Yes	Yes
16.	2.00	2.00	5.33	Nuanced	Yes	No
17.	2.00	1.80	5.33	Nuanced	Yes	Yes
18.	1.25	2.20	5.33	Nuanced	Yes	No
19.	3.75	3.20	4.83	Nuanced	No	No
20.	2.50	2.20	5.33	Nuanced	No	No
21.	2.00	1.80	4.83	Nuanced	Yes	Yes

RQ2: How are teachers' epistemological beliefs and understanding of historical thinking reflected in their teaching practice?

Teachers' instructional practice

When asked for concrete self-developed teaching materials regarding historical thinking, teachers mostly presented individual or group assignments for students. A few teachers provided the entire syllabus which they had designed themselves and students used throughout the year. The materials presented were mainly designed to enhance students' skills in source analysis, although some focused on other aspects such as multiperspectivity, or deconstructing historical representations.

When analyzing the material in terms of whether and how historical thinking, including epistemological reflection, was addressed, three types could be distinguished. This is depicted in Table 6, which shows for each respondent whether the material supported epistemological reflection (yes/no) and if so, whether it reflected more naive or nuanced beliefs.

In the first type, the material presented a univocal and closed account of the past, leaving no room for interpretation and paying no attention to the way in which historical representations are created. Sources were for instance used only in an illustrative way, to exemplify the presented historical narrative. This kind of material therefore does not foster students' historical thinking skills and could possibly promote rather naive epistemological beliefs among students.

A second type (labelled as 'yes-rather naïve') did pay attention to the disciplinary methods of historians but reflected rather naive beliefs. These materials were designed by teachers specifically in order to introduce students to disciplinary methods, yet they did not manage to present them in an accurate way. They contained a discrepancy between teachers' (reported) goals and the actual material. Some materials, for example, were aimed at critical source work and urged students to answer a historical question based on sources. The questions related to the sources, however, only asked about the content of those sources. Critical reasoning about the sources, for instance in terms of their trustworthiness and representativeness, was not taken into account. This could implicitly give students the idea that sources are a direct reflection of the past and that historical research is limited to merely collecting and summarizing information obtained

from sources. The role of interpretation and construction by historians was omitted in this type of material.

Even source work in which attention was paid to the trustworthiness of sources sometimes implicitly reflected naive epistemological beliefs, as it was assessed without reference to a concrete research question. The extent to which a source is trustworthy does not however lie within the source itself but depends on the historical question at hand. In this regard, teachers sometimes used a fixed set of questions that students had to answer mechanically for each source. In other cases, source analysis was reduced to identifying objective (described as 'facts') and subjective (described as 'opinions') elements in sources, whereby subjectivity was equated with untrustworthiness. Students were also asked to assess the extent to which a historical representation, for example in a movie, was consistent with historical reality. 'Subjective' sources were then compared to the 'objective' work of historians, without referring to evolving historians' representations or the existence of historiographical debates. Although this type of source work was attentive to the subjective nature of sources and the role of historians, it still reflected rather naive beliefs as it reduced historians' disciplinary methods to the identification of objective elements in sources.

Moreover, material in this category paid little attention to the role of evidence and argumentation. In an assignment on multiperspectivity, in which students were introduced to multiple perspectives on the Israeli-Palestinian conflict, no reflection was encouraged on the quality of the perspectives. The extent to which the perspectives were based on evidence, and whether the arguments were equally strong in both cases, was not assessed. The material emphasized the interpretive and subjective dimensions of historiography, but without acknowledging the role of historical methodology which allows historians to weigh the quality of different claims against each other.

A third type of material (labelled as 'yes-nuanced') did support nuanced epistemological beliefs among students, and explicitly discussed the disciplinary methods of historians, for instance by means of a high-quality critical source analysis through which students answered a historical question. Sources were provided with sufficient contextual information to allow students to make a substantiated and well-reasoned judgement about their trustworthiness in relation to the research question. Other material in this category introduced students to

different views on a historical phenomenon and then, based on an assessment of the sources, asked them to make their own, substantiated claim regarding the topic. For example, teacher 12 presented an assignment in which students weighed the arguments of one historian regarding the responsibility of Christianity for the destruction of classical culture. Based on a critical examination of this historian's evidence, students were encouraged to construct their own argument. This type of material introduced students to interpretation in historiography and to the way in which historians construct historical representations, while at the same time familiarizing them with procedures and methods designed to weigh various claims against each other. This material hence contributed to the development of nuanced epistemological thinking.

Table 6. Epistemological reflection in teachers' instructional practice

	Formal epistemologies			Practical epistemologies		Understanding of historical thinking	Instructional practice
Respondent	Nature of knowing (naive)	Nature of knowledge (objective)	Nature of knowing – nuanced/historical methodology	Case-study: contrasting statements	Does the discussion of the case-study confirm the results of the contrasting statements?	Constructed and interpretive nature of history mentioned as part of historical thinking?	Does the instructional practice support epistemological reflection?
1.	1.25	1.40	6.00	Nuanced	Yes	Yes	Yes - nuanced
2.	1.75	1.80	5.17	Nuanced	No	No	Yes – rather naive
3.	2.25	1.60	4.17	Nuanced	Yes	Yes	Yes – rather naive
4.	3.25	3.00	4.33	Nuanced	No	Yes	Yes – rather naive
5.	2.00	2.60	5.33	Nuanced	No	No	No
6.	3.75	3.00	4.67	Nuanced	No	No	Yes – rather naive
7.	1.75	2.20	5.83	Nuanced	No	No	Yes – rather naive
8.	1.50	2.20	5.33	Nuanced	Yes	Yes	Yes - nuanced
9.	3.00	2.40	4.67	Naive	Yes	No	No
10.	2.75	3.00	4.67	Nuanced	Yes	Yes	Yes – rather naive
11.	3.50	3.00	4.00	Nuanced	No	No	Yes – rather naive
12.	2.25	1.00	5.67	Nuanced	Yes	Yes	Yes - nuanced
13.	2.50	2.20	5.00	Nuanced	Yes	No	Yes - nuanced
14.	2.50	3.00	5.00	Nuanced	No	No	No
15.	1.50	1.20	5.83	Nuanced	Yes	Yes	Yes – rather naive
16.	2.00	2.00	5.33	Nuanced	Yes	No	Yes – rather naive
17.	2.00	1.80	5.33	Nuanced	Yes	Yes	Yes - nuanced
18.	1.25	2.20	5.33	Nuanced	Yes	No	Yes – rather naive
19.	3.75	3.20	4.83	Nuanced	No	No	Yes – rather naive
20.	2.50	2.20	5.33	Nuanced	No	No	Yes – rather naive
21.	2.00	1.80	4.83	Nuanced	Yes	Yes	Yes - nuanced

The relationship between teachers' epistemological beliefs, understanding of historical thinking and instructional practice.

Table 6 allows for an exploration of the manner in which respondents' epistemological beliefs and description of historical thinking relate to their instructional practice.

In a number of cases (e.g., 1, 8 and 12), a clear one-to-one relationship could be discerned between the teachers' epistemological beliefs and the way in which the teaching material presented history. They unambiguously espoused nuanced epistemological beliefs, both in the statements and the discussion of the case study. Moreover, they addressed epistemological reflection in their own description of historical thinking and developed teaching materials which reflected nuanced epistemological beliefs. Conversely, teachers 6 and 9 held epistemological beliefs which could not unequivocally be qualified as nuanced. These teachers scored relatively high on the subscales related to naive and nuanced beliefs. Their results in the case study added to this ambiguity. This was also evident in their instructional practice, which either did not pay attention to epistemological reflection or leaned towards rather naive beliefs. However, for a large number of respondents, the relationship was not so straightforward. Some teachers (e.g., 15 and 18) did have nuanced epistemological beliefs, but did not transfer this into their instructional practice.

In general, it can be concluded from Table 6 that teachers who paid attention to the interpretive and constructed character of history in their instructional practice also held nuanced epistemological beliefs themselves. Conversely, having nuanced beliefs did not necessarily lead to an instructional practice that supported nuanced beliefs.

Two considerations are important in this regard. First, this finding indicates that other beliefs influence the relationship between epistemological beliefs and instructional practice. This was illustrated clearly in the case of teacher 12. Although he himself had nuanced beliefs and addressed this in his instructional practice, he spontaneously indicated that he only rarely did so. Within the amount of time available for history education, he preferred to prioritize other objectives in his teaching. Hence, both his beliefs about the purpose of history education and contextual factors had an influence. Second, methodological issues continue to play a role here. Both the Likert-type statements and the case study appear to be

insufficiently capable of accurately capturing teachers' epistemological beliefs. Moreover, they do not always unequivocally point in the same direction.

CONCLUSION AND DISCUSSION

This study sought to gain an insight into the relationship between teachers' formal and practical epistemologies, their understanding of historical thinking and their instructional practice. Using a questionnaire and interviews, 21 history teachers' epistemological beliefs on a general and practical level were analyzed, along with their understanding of historical thinking and their instructional practice. In so doing, we also sought to contribute to methodological debates about measuring epistemological beliefs.

Teachers' understanding of historical thinking, epistemological beliefs and instructional practice

As regards history teachers' epistemological beliefs, it is difficult to paint an unambiguous picture. Based on a combined analysis of their general and practical epistemologies, it was only possible to unambiguously identify nuanced or naive beliefs in about half the respondents. For the other half, nuanced beliefs seemed to be present only superficially. This was apparent from the absence of an accurate argumentation in the case study and in the combination of nuanced and naive beliefs. In some cases, this appeared to stem from a procedural objectivist stance (Stoel et al., 2017). For these teachers, naive and nuanced beliefs coexisted, and they attached great importance to disciplinary procedures as a means of obtaining absolute knowledge about the past.

As regards teachers' understanding of historical thinking, it appears that most teachers were aware of history being a matter of interpretation and construction. Yet less than half of them spontaneously brought this up. The same applied to other dimensions of historical thinking, which suggests that teachers have only a cursory understanding of the concept. The absence of a clear operationalization of historical thinking in the Flemish history standards, can serve as an explanatory factor here.

This cursory understanding of historical thinking could also be observed in teachers' instructional practice. In the teaching materials we analyzed, teachers

aimed to introduce students to the disciplinary methods of historians, yet only a handful of teachers succeeded in developing materials that adequately reflected the interpretive and constructed nature of historical knowledge. The majority of teachers presented materials which reflected a rather naïve understanding of history by either over-emphasizing the subjective nature of historical knowledge or misrepresenting disciplinary methods. These superficial practices regarding historical thinking might be connected to teachers' rudimentary understanding of the concept, but might also be due to epistemological or other beliefs.

As to the relationship between epistemological beliefs and instructional practice, this study confirms the existence of a such a relationship (Van Hover & Yeager, 2004; Wansink et al., 2016; Yilmaz, 2008). More specifically, nuanced epistemological beliefs among teachers seem to be a necessary condition for a practice that includes interpretive history teaching, although they are not in themselves a sufficient condition. In line with the research by VanSledright (1996), Hartzler-Miller (2001), McDiarmid (1994) and Voet and De Wever (2016; 2019), this study found that teachers' nuanced epistemological beliefs were not always reflected in their instructional practice.

We propose three explanations for this. A first is that teachers sometimes held rather naive practical epistemologies, even though their general epistemologies were more nuanced. As we will argue later, teachers' practical epistemologies were more accurately captured in the discussion of the case study rather than via the contrasting statements themselves. For a number of teachers (e.g., teacher 2, 6, 7 and 11) these discussions revealed rather naive views, which could explain why their teaching materials did not fully support nuanced epistemological beliefs either.

For other teachers, this explanation does not suffice, as both their formal and practical epistemologies (based on the discussion) indicated nuanced beliefs, while their practice did not (e.g., teacher 3 and 10). A second explanation is therefore related to the situated nature of teachers' instructional practice. Several researchers have pointed out the importance of context when assessing epistemological beliefs, as specific conditions might generate different epistemic cognitions and aims (Chinn et al., 2011; Sandoval, 2014). It is for instance possible that a teacher's aim is not to introduce students to the interpretive nature of historical knowledge, but rather to transfer the 'existing' knowledge, because the

teacher does not consider it feasible or desirable to translate their own nuanced epistemic beliefs into their teaching practice. This links to findings in other studies emphasizing the importance of contextual factors in teachers' instructional practice (e.g., Hicks, Doolittle & Lee, 2004; VanSledright & Limón, 2006; Voet & De Wever, 2016).

Lastly, we should consider the possibility that teachers may encounter serious difficulties in designing materials that accurately reflect disciplinary methods and promote a nuanced epistemological understanding in students. Our analysis of the teaching materials showed that several teachers did try to introduce students to the methods of historians, but did not always succeed in developing materials that achieved this adequately. This might indicate that teachers' competence in designing such materials is not sufficiently developed. The fact that teachers cannot rely on a concrete conceptualization or didactic support from the history standards might exacerbate this difficulty.

These observations have implications for teacher training programs and professionalization initiatives. It is particularly important that these pay explicit attention to and support teachers in the development of teaching materials aimed at fostering epistemological reflection. Furthermore, a thorough understanding of the concept of historical thinking should be imparted, given teachers' limited and tacit understanding of the concept.

Methodological findings

On a methodological level, the combined analysis of teachers' formal and practical beliefs in this study revealed two difficulties. A first concerns the question of whether the measurement instruments are capable of accurately mapping teachers' epistemological beliefs. It appeared that the statements based on Stoel et al. (2017) were interpreted in various ways by teachers, and that their answers were not always in line with their actual beliefs. Moreover, the instrument did not succeed in isolating teachers' epistemological beliefs: throughout the discussion of the statements, it became apparent that teachers considered them from a didactic perspective and took into account contextual factors when answering them. Teacher 2, for instance, saw it as his task to weigh up different perspectives on the past for himself, in order to subsequently present a univocal account to his students. The fact that teachers can have a double epistemic stance and thus have

differing views on disciplinary and school knowledge has already been established (Maggioni & Parkinson, 2008), but the discussion of the statements shows that it is difficult to separate the two in teachers' minds and thoughts. This explains inconsistencies in answers and casts doubt on whether general statements are capable of generating a valid assessment of teachers' formal epistemological beliefs. This might also explain why scholars experience difficulties when using statements to capture these beliefs (e.g., Voet & De Wever, 2019).

The finding that teachers do not distinguish their own epistemological beliefs from the context of their teaching only became apparent after we discussed some of the statements with the respondents. We also applied this procedure to the instrument that assessed teachers' practical epistemologies. At first glance, this instrument seemed less sensitive in this regard, as we found no indication that teachers were including their beliefs about school knowledge and contextual factors in their response. However, the discussion of the contrasting statements uncovered a second methodological issue: the inconsistency between teachers' responses and their justification for them. It was only when responses were discussed that variations in teachers' epistemological beliefs emerged, sometimes revealing rather naive views. This suggests that the contrasting statements by themselves were insufficient to identify teachers' practical epistemologies. We suspect this to be the case because although the contrasting statements were connected to a concrete case, they might still primarily gauge formal epistemologies, since they are fairly generic and require little concrete application of epistemological beliefs. Only when teachers were asked to explain their answers did they start to reason about the specific case at hand, which seemingly activated practical epistemologies.

Methodological implications for future research

This study also provides suggestions for how to proceed in further research. First, this study highlighted the profound interconnectedness of teachers' epistemological beliefs with other beliefs and contextual factors. When trying to measure teachers' epistemological beliefs, researchers might therefore consider starting from the assumption that teachers will bring their didactic context into play, making it difficult to measure something as complex as epistemological views using general statements. Precisely because such statements fail to capture

the mediating role of teaching contexts and other beliefs in the translation of teachers' beliefs into their practice, they appear to be poor predictors of that practice.

Second, rather than focusing on general epistemological beliefs, future research might benefit from concentrating more on teachers' practical epistemologies. When teachers were asked to reflect on an authentic discipline-specific case, differences in epistemological views emerged that remained hidden at a formal level and that appeared to correspond better with their teaching practice. However, more research is needed to develop a measurement instrument that can accurately capture these practical epistemologies, as the contrasting statements by themselves still seemed to primarily gauge formal epistemologies. An instrument that requires teachers to provide substantiated reasoning about a discipline-specific task would seem to be more suitable. The scenario-based Epistemic Thinking Assessment by Barzilai and Weinstock (2015), which combines a scenario-based approach for biology and history with Likert-type statements, might be interesting in this regard. Although the history-based scenario was less valid and reliable than the biology scenario, a further examination of this instrument in future research might constitute a valuable addition.

Third, this study showed that teachers' instructional practices revealed a variety of views on disciplinary knowledge. Some teachers presented history as a single, univocal narrative, while others emphasized the subjective nature of historical knowledge or presented history as interpretive and constructed. Considering this variation, it might perhaps be fruitful for further qualitative research to work 'in reverse': rather than trying to understand teachers' practices via their beliefs, teachers' instructional practice might perhaps serve as a fruitful starting point for examining their epistemological beliefs and their understanding of historical thinking. It remains important, however, to simultaneously assess other beliefs and the role of contextual factors and related epistemic aims, so that their influence can be identified.

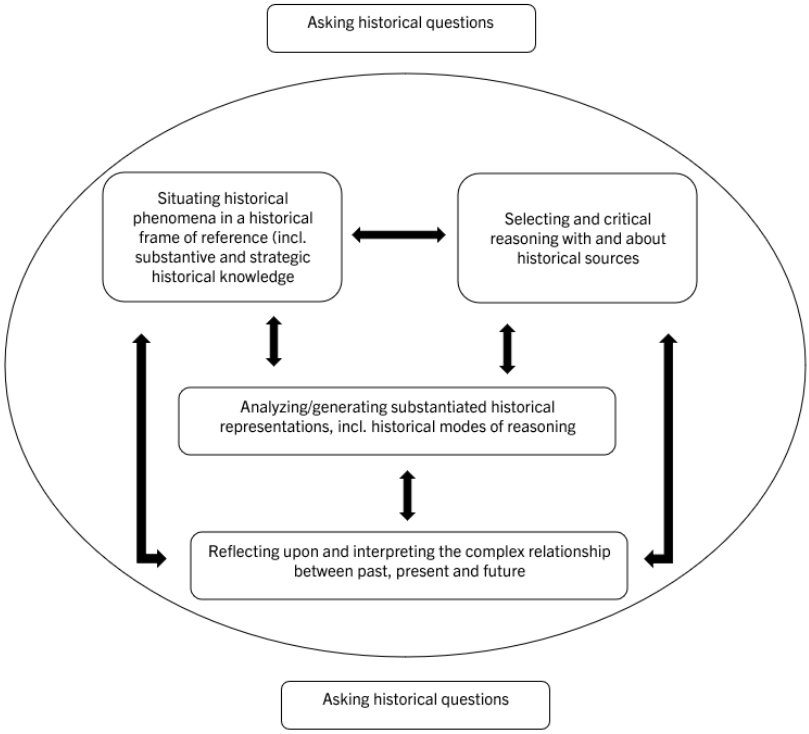
A final implication stems from the finding that a discrepancy sometimes occurs between teachers' intended objectives for their practice and the actual materials used. It therefore seems imperative that research on teachers' epistemological beliefs and practices does not rely solely on self-reported practices

but complements this with an analysis of actual teaching practices, for instance via lesson observations or an analysis of didactic materials.

Finally, two limitations should be mentioned. First, the sample consisted entirely of teachers from the last stage of secondary education. They all had a Master's degree, which can be expected to influence their epistemological beliefs (e.g., Voet & De Wever, 2019). Second, in most cases we only studied a fraction of teachers' instructional practice. Teachers were given the opportunity to make their own selection of 'best' materials, but how these materials relate to their teaching practice as a whole was not thoroughly examined.

APPENDIX A: CONCEPTUALIZATION OF HISTORICAL THINKING AS USED IN THIS RESEARCH

Figure 1. Conceptualization of historical thinking



Note. The conceptualization of historical thinking consists of five building blocks (Based on: Van Nieuwenhuyse, 2020, pp. 373-379).

- Asking historical questions about the past, which constitutes a starting point for historical thinking.
- Situating historical phenomena, sources and representations in a broader historical frame of reference. This includes the acquisition of substantive and strategic historical knowledge.
- Selecting and critically analyzing historical sources, including reasoning with and about sources. This includes, among others, the ability to evaluate a source's usefulness, reliability, and representativeness in light of a specific historical question.
- Analyzing and/or formulating a substantiated answer to a historical question (i.e., constructing a historical representation). This building block includes the ability to deconstruct existing historical representations as well as to construct one. This

is based on a critical source analysis and includes the application of typical historical modes of reasoning such as: cause and consequence (incl. contingency); continuity and change, historical contextualization, historical empathy, multiperspectivity, the attribution of agency, drawing historical analogies, deconstructing collective memory, deconstructing narratives and constructing open and substantiated narratives.

- Critically reflecting on and interpreting the complex relationship between past, present and future. This explicitly includes the need for epistemological reflection and requires students to understand the interpretive and constructed nature of historical knowledge and the importance of one's positionality.

APPENDIX B: OVERVIEW OF ANALYTICAL CONCEPTS IN RELATION TO THE DATA COLLECTION INSTRUMENTS

Table 1. overview of analytical concepts in relation to the data collection instruments

Theoretical concepts	Analytical categories (codes)	Data collection instruments		
		Questionnaire	1st semi-structured interview	2nd semi-structured interview
Teachers' understanding of historical thinking *Which aspects of historical thinking do teachers mention in their own conceptualization?*	Among others: Historical questions Historical modes of reasoning _agency _multiperspectivity _causality _… Epistemological reflection	Open-ended questions: - Teachers' own definition of historical thinking - Strategies to foster it among students	Discussion of answers to questionnaire Discussion of model of historical thinking	
Formal epistemology *What are teachers' general beliefs about the characteristics of knowledge and its justification in history?*	Formal EB _naive _nuanced	Likert-type statements (based on Stoel et al., 2017)	Discussion of inconsistencies within statements gauging formal epistemology	
Practical epistemology *What kind of epistemic reasoning and thinking do teachers display when confronted with a discipline-specific task?*	Practical EB _naive _nuanced		Case-study on a historiographical debate with contrasting statements and discussion	
Instructional practice regarding historical thinking *Which aspects of historical thinking are addressed? How are disciplinary practices and knowledge presented in teachers' instructional practice aimed at fostering historical thinking?*	Among others: Historical modes of reasoning _agency _multiperspectivity _causality _… Source work _reasoning with _reasoning about _… Epistemological reflection _naive _nuanced			Analysis and discussion of concrete teaching materials regarding historical thinking, provided by respondent.

Chapter 4

Fostering secondary students' historical thinking. A design study in Flemish history education.[1]

A recent curriculum reform in Flanders (Belgium) has introduced historical thinking as a central goal for history education. Historical thinking aims to introduce students to the methods of historians and disciplinary ways of thinking. It is a complex act, requiring the application of substantive and second-order knowledge, and is difficult to foster among students. International (intervention) research has provided several guidelines for the design of instructional practices that are effective in promoting specific aspects of students' historical thinking. However, these studies do not approach historical thinking in a holistic manner, are often vague about how general design principles were adapted to history education and rarely report on whether the designed curricula were considered relevant and useful by teachers. Taking into account the many difficulties that teachers encounter in designing practices aimed at historical thinking this design research aims to gain more insight into the design of instructional practices that are both effective in fostering historical thinking in a holistic manner and that are considered socially valid by teachers. The designed artifact is a 12 to 14 hour lesson series on the theme 'decolonization after 1945', for students in the 12th grade. It applies the model of cognitive apprenticeship's (Collins et al., 1991) general design principles to the specific context of history and approaches historical thinking in a holistic manner. The initial lesson series was evaluated and revised in two rounds based on a pilot study, an expert review and an intervention study.

[1] This chapter is based on: Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. Fostering secondary students' historical thinking. A design study in Flemish history education. *Manuscript under review.*

INTRODUCTION

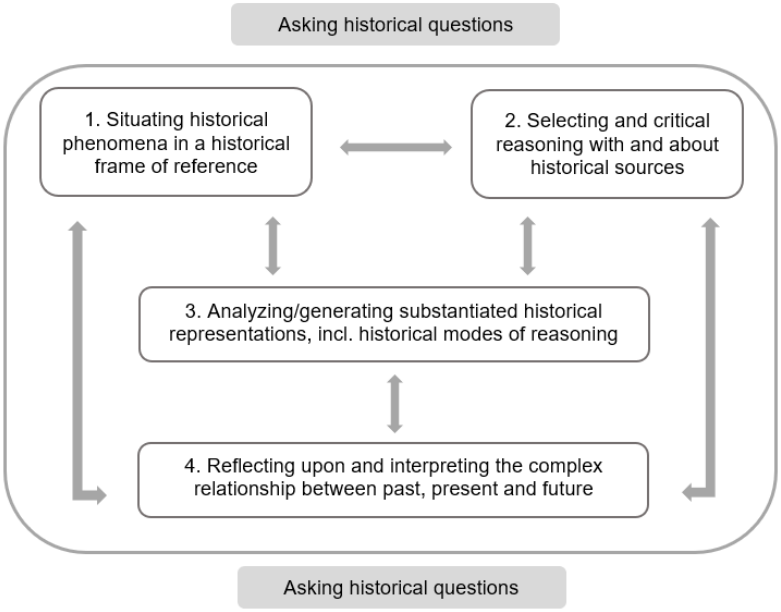
In Flanders (Belgium), a curriculum reform is gradually being implemented in history education since 2018. It includes the introduction of historical thinking as a central goal for history education, in line with a transformation that manifested internationally since the 1990s in which history education connects more closely to the practice of professional historians. Rather than emphasizing a single (national) narrative about the past and simply describing "what history is about", historical thinking offers an insightful and recent historiography-based approach to the past and describes "the way we go about doing history" (Lee & Ashby, 2000, p. 199).

Historical thinking combines 'knowing history' and 'doing history' to generate a deep historical understanding (Havekes et al., 2012). Knowing history refers to the acquisition of substantive knowledge of history and is related to the content of history, describing – in line with recent historiography – in an accurate way specific historical periods, persons or structures, such as feudalism or the enlightenment (Lee, 2004; Van Drie & Van Boxtel, 2008). Doing history points to second-order and procedural knowledge allowing students to gain an understanding of how historical knowledge is constructed (Lee & Ashby, 2000; Lévesque, 2009; van Drie & van Boxtel, 2008). Second-order knowledge refers to the methods of historians (e.g., reflecting on cause and consequence, evidence, sources); procedural knowledge describes strategies and approaches used by historians when studying the past (e.g., how to evaluate the reliability of a source, how to establish causality) (van Drie & van Boxtel, 2008; VanSledright & Limón, 2006). Fostering historical thinking allows students to develop nuanced epistemological beliefs about history: beliefs about the nature of knowledge and processes of knowledge construction in history (Buehl & Alexander, 2001).

In line with international models (Seixas & Morton, 2013; van Drie & van Boxtel, 2008; Wineburg, 2001) the Flemish history standards operationalize historical thinking into four building blocks (see Figure 1): (1) Situating historical phenomena, sources and representations in a broader historical frame of reference. This includes the acquisition of substantive historical knowledge. (2) Selecting and critically analyzing historical sources, including reasoning with and about sources. This includes, among others, the ability to evaluate a source's usefulness,

reliability, and representativeness in light of a specific historical question. (3) Analyzing and/or formulating a substantiated answer to a historical question. This includes the ability to construct a substantiated historical representation, based on critical source analysis and on the application of typical historical modes of reasoning such as: cause and consequence, historical contextualization or multiperspectivity. (4) Critically reflecting on the complex relationship between past, present and future. This explicitly includes the need for epistemological reflection and requires students to understand the interpretive and constructed nature of historical knowledge (Van Nieuwenhuyse, 2022; see also Wilke et al., 2022a). Besides these four building blocks, asking historical questions is considered the starting point for historical thinking: it hence surrounds the four building blocks (Van Nieuwenhuyse, 2020).

Figure 1. Operationalization of historical thinking in the new Flemish history standards (see also Wilke et al., 2022a).



With the introduction of the new standards, Flemish history teachers have to reconsider their practice to align with the new goal of historical thinking, which is considered a complex and unnatural act that goes against the way students

spontaneously look at and engage with the past (Wineburg, 2001). Getting students to think 'historically' is therefore challenging and requires well thought-out instructional materials (van Drie & van Boxtel, 2008). To facilitate the implementation of new curricula, it is essential to provide teachers with good practices aligned with them. In so doing, these practices can function as concrete examples and provide guidelines for the development of new materials (Gouëdard et al., 2020).

Previous (intervention) studies on fostering historical thinking among students have provided several guidelines for designing effective instructional practices, often based on principles of cognitive apprenticeship (Collins et al., 1991). However, it is difficult to use these interventions as blueprints for the design of new instructional materials as descriptions of lesson units in these studies are often very concise. It is not always clear how general educational design principles can be concretely adapted to a specific history education context.

Furthermore, intervention studies on fostering historical thinking often do not adopt a holistic approach but rather focus on particular aspects such as causal reasoning (Stoel et al., 2015), historical contextualization and perspective-taking (Huijgen et al., 2018), critical source analysis (Nokes et al., 2007; Reisman, 2012) or historical writing (De La Paz, 2005; De La Paz & Felton, 2010; Van Drie et al., 2015). Because of the strong focus on specific aspects of doing history, these studies pay little attention to knowing history. It is often unclear to what extent the (content of the) instructional practices adopted in the intervention studies connect to recent historiography, even though this is an important part of historical thinking. As a result, these studies have generated few guidelines on how to integrate multiple aspects of historical thinking, encompassing both knowing and doing history.

A final difficulty concerns the position of the teacher. Reports of intervention studies usually limit themselves to measuring the effects on students. There is often little information about how teachers perceived the materials. Yet, in order for interventions to have an applied value it is important that they have social validity, i.e., that they are considered acceptable, relevant and useful by the actors involved (Kazdin, 2005). This is particularly important given that research has repeatedly shown that historical thinking finds its way only sparingly into educational practice (e.g., Hicks, Doolittle & Lee, 2004; Van Nieuwenhuyse et al.,

2015; VanSledright & Limón, 2006; Voet & De Wever, 2016). This points to a gap between theoretical guidelines for the design of instructional practices aiming to foster historical thinking and their practical application.

The design study

Based on the challenges outlined above, the purpose of this research is to gain more insight into the design of instructional practices that are both effective in fostering historical thinking and are considered socially valid by teachers. A design-based research process was deemed particularly appropriate as it allowed for a close collaboration with teachers while considering evidence-based recommendations regarding the promotion of students' historical thinking. This allows for the development of a maturing intervention that is directly relevant for practice and contributes to a better theoretical understanding of the effectiveness of the applied design principles, hence meeting both the practical and theoretical goals of design-based research (McKenney & Reeves, 2019).

The design team was composed of a professor in educational sciences, a history professor specialized in history didactics and a PhD candidate in history (didactics). The research was part of a broader project examining teachers' beliefs and practices regarding historical thinking and examining the relationship between historical thinking and democratic citizenship. This paper reports on the design process of the lesson series that was developed in this context.

Overview of the design process

The design process was an iterative process throughout which educational practitioners and other experts were actively involved. It started with an analysis and exploration phase, followed by two cycles of design and evaluation and one final design phase (McKenney & Reeves, 2019).

The analysis and exploration phase explored difficulties and solutions regarding the design and implementation of instructional practices to foster historical thinking. It was based on an analysis of research literature and a qualitative study on teachers' beliefs and instructional practices. A first design and construction phase resulted in an initial version of the lesson series, based on principles derived from the model of cognitive apprenticeship (Collins et al., 1991). This version was evaluated via a pilot study and an expert review with

history and (history) educational scholars as well as experienced teachers. The evaluation resulted in a second design phase in which the initial lesson series was adjusted. Due to the Covid19 pandemic, an additional construction phase was needed to modify the lesson series for remote education. The revised lesson series was evaluated via an intervention study in which we collected data from students (via pre- and posttests) and teachers (logbooks and interviews). This resulted in a third design phase, finalizing the lesson series. An overview of the design process is depicted in Figure 2. In what follows, the research context will be described, after which each step in the design process is discussed in detail.

Figure 2. Overview of the design process

Analysis and exploration	Exploration of needs and solutions - Research literature - Qualitative study
First design and construction phase	Development of the initial lesson series - Design principles - Initial lesson series - Teacher and student materials
First evaluation and reflection phase	Evaluating the initial lesson series - Expert review - Pilot study
Second design and construction phase	Revision and digitization of the lesson series - Revised teacher and student materials - Scenario for digitized lesson series
Second evaluation and reflection phase	Intervention study - Pre- and posttests with students - Interviews and online questionnaire with teachers
Third design and construction phase	Finalization of the lesson series

Note. Research phases based on McKenney and Reeves, 2019.

RESEARCH CONTEXT

This design study was conducted in Flanders (Belgium), where history education takes up two hours a week in general education. The new history standards are centered around historical thinking, which is conceptualized as a combination of knowing and doing history and operationalized in four building blocks (see introduction). In terms of historical content, the standards only prescribe which historical periods should be addressed at which stage of secondary education and do not impose concrete factual knowledge.

Teachers in the last stage of secondary history education are required to have a master's degree in history and a teaching degree. In the last decade, significant changes have occurred in the teacher training courses, in particular regarding their attention to historical thinking on a theoretical and practical level. Considering the absence of mandatory teacher professionalization, the curricular freedom for history teachers and the absence of standardized testing, not all teachers have necessarily become acquainted with historical thinking in the same manner, and hence current teaching practices regarding historical thinking vary strongly. Overall, instructional practices by Flemish history teachers are characterized by an emphasis on knowing history, rather than doing history (Van Nieuwenhuyse et al., 2017; Voet & De Wever, 2016). For instance, although literature recommends organizing historical inquiries based on a critical analysis of multiple historical documents to foster students' historical thinking, most Flemish history teachers are not inclined to organize these in their practice (Voet & De Wever, 2016). For the majority of Flemish history teachers, these new history standards hence require a significant reconsideration of their instructional practice, providing an ideal context for this study.

EXPLORATION OF NEEDS AND SOLUTIONS

Difficulties regarding (the teaching of) historical thinking

A first step in the design process was to examine teachers' beliefs and instructional practices with regard to historical thinking as well as general issues related to students' historical thinking ability. This was done via a literature review and via a qualitative study with 21 Flemish history teachers based on two semi-structured

interviews and an analysis of their instructional practices (Wilke & Depaepe, 2019; Wilke et al., 2022a). This allowed to identify concrete difficulties in the Flemish context and to examine to what extent they aligned with international findings.

The literature review revealed several issues with regard to teachers' historical thinking practices. Overall, it showed that these practices often remain focused predominantly on knowing history (Barton & Levstik, 2004; VanSledright & Limón, 2006; Van Nieuwenhuyse et al., 2015). When doing history is addressed, this is rarely done in an in-depth way. Studies established, for instance, that teachers use sources mainly in a purely content-related manner: as illustration or merely as carriers of substantive knowledge rather than subject them to a critical analysis (e.g., Hicks, Doolittle & Lee, 2004; Nokes, 2017; Van Nieuwenhuyse et al., 2015).

Literature on students' historical thinking ability shows that such practices may hinder the development of nuanced epistemological beliefs among students. As students are not (or only superficially) challenged to critically evaluate sources their naïve perceptions that historical knowledge can simply be derived from sources in a straightforward way, are being confirmed instead of questioned (Wilke et al., 2022a). Such naïve epistemological beliefs also hinder the application of typical historical modes of reasoning (e.g., van Drie et al., 2013; Huijgen et al., 2014), as students do not understand that historical knowledge is interpretive and constructed, yet tend to consider historical knowledge to be either fixed and singular, or completely subjective (Maggioni et al. 2009; Stoel, Logtenberg et al., 2017).

In explaining why these difficulties occur, literature points to the importance of teachers' educational beliefs and contextual factors. Teachers sometimes question their students' ability to engage in interpretive history (Van Hover & Yeager, 2003; Wansink et al., 2016) and tend to focus primarily on knowing history (Barton & Levstik, 2004; Van Hover & Yeager, 2003). Contextual factors, such as time and available resources (Voet & De Wever, 2016) and the presence of standardized tests were also identified as hindering factors (Hicks, Doolittle & Lee, 2004; VanSledright & Limón, 2006).

Our own qualitative study confirmed these findings, yet also generated additional points of interest. Teachers reported uncertainty about how to evaluate doing history as well as concerns about the difficulty of historical thinking for

their students, in particular with regard to building block 3 'generating historical representations' (Wilke & Depaepe, 2019). This study also provided possible explanations for these perceived difficulties. First, most teachers only had a limited and one-sided understanding of what historical thinking is. Second, teachers often made a strict distinction between knowing and doing history. This dichotomy goes against literature which stresses the interconnectedness and concurrent nature of knowing and doing history (Lee, 2005; VanSledright & Limón, 2006). Third, several teachers believed that addressing doing history is only possible via student-centered activities, such as inquiry-based tasks, which went against (some) teachers' preference for teacher-guided instructional practices (Wilke & Depaepe, 2019). These (mis)conceptions concerning the teaching and learning of historical thinking fueled teachers' hesitancy to address doing history in their practice.

The needs analysis allowed us to identify areas in which teachers are in need of additional support and which have to be addressed when designing lesson materials that can be considered socially valid (e.g., providing examples on how to integrate knowing and doing history, the need to model good practices, supporting evaluation). It also confirmed the need for lesson materials that foster students' historical thinking in a profound and holistic manner. A next step in the design process was to generate design principles for the construction of such materials.

Instructional design principles to foster historical thinking

To establish design principles a literature study was conducted examining both literature from general educational psychology and history educational literature.

Considering the complexity of historical thinking, requiring a combination of substantive, second-order and procedural knowledge, the model of cognitive apprenticeship (Brown et al., 1989; Collins et al., 1991) seemed an appropriate framework for designing practices to foster historical thinking. This model is directed at the acquisition of complex skills "through processes of observation and guided practice" (Collins et al., 1991, p. 13) and centers around the use of authentic tasks.

In terms of teaching methods, it puts forward the principles of modeling, coaching and scaffolding to help students acquire these skills. In the modelling phase the teacher or expert demonstrates the discipline-specific tasks and provides

explicit instruction on the application of conceptual and procedural knowledge. Subsequently, students practice the application of these skills supported by scaffolds and coaching by the teacher. By fading out the scaffolds, students gradually gain more independence in the application of the intended skills. Articulation and reflection support students to articulate their knowledge and reasoning processes and compare them to others. Finally, students are encouraged through exploration to apply their acquired skills to new (including their own) problems and settings.

History educational literature supports the adequacy of principles from cognitive apprenticeship to promote students' historical thinking (van Boxtel & van Drie, 2018) and provides more concrete guidelines for applying the model's general principles to the specific context of history.

For history, inquiry-based (writing) tasks are considered authentic tasks that connect to the practice of historians (van Boxtel & van Drie, 2018). These tasks allow students to develop their own substantiated historical representations in response to an authentic historical question. Students are required to investigate the historical question via a critical analysis of multiple sources (van Boxtel et al., 2021; Voet & Wever, 2017). This representation then needs to be substantiated by providing arguments and evidence and engaging with counterarguments (van Boxtel et al., 2021; Monte-Sano et al., 2014). In particular, evaluative research questions are recommended for inquiry activities as these are more effective in eliciting historical thinking (van Drie et al., 2006). These inquiry-based activities connect closely to the history discipline and to various aspects of historical thinking. They therefore not only address historical thinking in a holistic manner but are also expected to be able to contribute to the development of nuanced epistemological beliefs about history (van Boxtel et al., 2021).

Several studies have successfully relied on principles of modeling, coaching and scaffolding to enhance aspects of students' historical thinking (see introduction). They emphasize at the same time the importance of explicit teaching. This should be directed at second-order and procedural knowledge and should not be limited to the modeling phase, but embedded throughout the instructional practice (Reisman, 2012; Stoel, van Drie, & van Boxtel, 2017). Moreover, these studies demonstrate that explicit teaching on discipline-specific, rather than general strategies, yielded better results (Reisman, 2012; Van Drie et

al., 2015). In terms of scaffolding, graphical organizers, such as an argumentative diagram, are recommended to support students' reasoning processes (Stoel, van Drie, & van Boxtel, 2017; van Drie et al., 2005) and research suggests that scaffolds for historical writing should contain prompts specifically directed at critical source analysis (Monte-Sano & De La Paz, 2012). These studies also provide concrete examples of scaffolds supporting various aspects of historical thinking, such as argumentative history writing, source corroboration and critical source analysis (e.g., University of Michigan, 2021; Monte-Sano et al., 2014). Concrete examples were also found of materials which could be used for coaching students by providing feedback, particularly through the use of discipline-specific rubrics (e.g., Monte-Sano, 2012; Monte-Sano et al., 2014; Monte-Sano & De La Paz, 2012; Stoel, van Drie, & van Boxtel, 2017).

To support articulation and reflection, discipline-specific literature recommends peer-to-peer interaction, via group work allowing students to explicate their reasoning processes (e.g., Stoel et al., 2015; Stoel, van Drie, & van Boxtel, 2017). However, the importance of whole-class discussions led by the teachers is also emphasized as these provide opportunities for further articulation, teacher feedback and reflection on students' learning (Reisman, 2012; Stoel, van Drie, & van Boxtel, 2017; van Boxtel et al., 2021; van Drie & van de Ven, 2017).

Regarding the integration of knowing and doing history, discipline-specific literature emphasizes the importance of both aspects as part of students' historical thinking (Lee, 2005; van Boxtel & van Drie, 2017). Yet, we found no concrete guidelines on how to adequately integrate both in an instructional practice nor on how to make sure the lesson series corresponds to recent historiographical shifts and insights. Research on the relationship between academic and school knowledge, however, has argued that fundamental changes occur when academic knowledge is transferred to a school context, in the sense that a substantiated scientific historical representation becomes distorted, creating a disconnect between both (Popkewitz, 2004). Within history education such a disconnect has for instance been analyzed for representations of the colonial past (Bentrovato & Van Nieuwenhuyse, 2020). This clearly points at the necessity that lesson materials intended to foster historical thinking will have to align with recent historiographical insights.

DEVELOPMENT OF THE INITIAL LESSON SERIES (1ST DESIGN AND CONSTRUCTION PHASE)

Design principles

Based on the results of the needs analysis and exploration, we outlined a number of design principles that would be implemented in a lesson series. Because we aimed to foster historical thinking in a holistic and in-depth manner (focusing on knowing and doing history and covering all four building blocks of historical thinking) the lesson series was directed at students in the 12th grade (hence towards the end) of secondary education.

In line with recommendations from the literature discussed in the previous section, the following design principles have been distinguished. First, we ensured a close adherence to current historiography by consulting academic literature and an expert professional historian, and by including historiographical debates in the lesson series. These historiographical debates also facilitated an integration of knowing and doing history as they provided challenging questions, functioning as a starting point for inquiry tasks, which constituted the second design principle. These inquiry tasks were used as authentic, discipline-specific tasks in which students constructed a substantiated historical representation. They were connected to authentic historiographical debates and were designed around historical sources that reflected the various perspectives within those debates. Third, the lesson series drew on the principles of modeling, coaching (incl. feedback) and scaffolding to support students in constructing these substantiated historical representation. To support teachers in providing scaffolding, coaching and feedback to their students, scaffolding tools were designed as well as a rubric to evaluate students' ability to construct a substantiated historical representation. Fourth, the lesson series included explicit teaching on second-order and procedural knowledge. Fifth, articulation and reflection were facilitated via group work, alternated with whole-class discussions and with shorter, teacher-guided tasks which provided opportunities for additional teacher feedback. The decision to include whole-class discussions and teacher-guided tasks was in line with research literature, but was also prompted by a concern not to reinforce teachers' preconception that doing history is only possible via time-consuming, student-centered activities.

The initial lesson series

Based on the design principles presented above, a first version of the lesson series was constructed. The lesson series was developed around the theme of 'decolonization after 1945' and included eight lessons, taking up 12 to 14 history classes. Historical thinking was addressed in a holistic way by addressing both knowing and doing history and covering all four building blocks of historical thinking, in line with the new Flemish history standards. To this end, each lesson had a primary focus on either substantive knowledge (knowing history) or second-order and procedural knowledge (doing history). Both aspects, however, were closely interconnected. The lessons with a focus on knowing history provided students with substantive knowledge which they had to apply in the lessons with a focus on doing history. They also included explicit teaching on second-order knowledge and short exercises on aspects of doing history. For instance, lesson 4 centered around the question to what extent the Balfour declaration of 1917 was at the root of the Israeli-Palestinian conflict. This lesson was focused mainly on acquiring substantive knowledge about the origins and characteristics of the conflict. However, students simultaneously compared Israelian and Palestinian narratives on the role of the Balfour declaration which provided opportunities for explicit teaching on the notion of causality (as a historical mode of reasoning), multiperspectivity and the pitfalls of presentism and hindsight. Students had to apply this newly acquired knowledge of causality and multiperspectivity in the inquiry tasks in lesson 5. These lessons thus provided impulses to further students' procedural and second-order knowledge. The lessons with a focus on doing history also provided opportunities to foster substantive knowledge. For instance, the second lesson was directed at doing history and centered around an inquiry task in which students had to provide a substantiated answer to the question who was most responsible for the political decolonization of Belgian Congo. While analyzing the sources, students gained insight into Belgian and Congolese views on independence, important agents in the process towards independence and acquired more in-depth knowledge about the process of decolonization.

An overview of the lesson series is presented in Table 1.

Table 1. Overview of the lesson series

Lesson nr.	Duration	Lesson topic and central historical question	Student and teacher activities	Primary focus	Historical thinking building blocks
1	2 x 50'	Causes and process of decolonization Historical question: Who was most responsible for the decolonization of British-India?	Teachers model the process of generating a substantiated historical representation via historical inquiry. Teachers provide explicit teaching on multiperspectivity and on evaluating sources' usability, reliability, source corroboration and on how to substantiate a historical claim (argumentation and evidence).	Doing history	1, 2, 3
2	2 x 50'	Decolonization of Belgian-Congo Historical question: Who was most responsible for the decolonization of Belgian-Congo?	Students work in groups on a historical inquiry task, supported by scaffolds. Teachers coach students and provide explicit teaching on the concept of the representativeness of sources. Teachers provide feedback and lead class-discussion on the inquiry task.	Doing history	1, 2, 3
3	1 x 50'	Belgian collective memory of the colonial past Historical question: How did the representation of the colonial period evolve over the period 1945-present in Belgian-Flemish history textbooks and what role did the changing context play in this regard?	Students evaluate sources' usability, reliability and representativeness. Teachers provide explicit teaching on different layers of multiperspectivity (in the past, in the present, over time) and on the constructed and interpretive nature of historical knowledge	Knowing history	2, 3, 4
4	2-3 x 50'	The Israeli-Palestinian conflict Historical question: To what extent was the Balfour declaration of 1917 at the root of the Israeli-Palestinian conflict?	Students study and compare multiple perspectives, they evaluate sources' reliability. Students practice formulating a substantiated historical claim, based on evidence. Teachers lead whole class-discussion related to students' activities on evaluating sources and formulating a substantiated historical claim. They provide explicit teaching on presentism and on causality as a historical mode of reasoning, connected to the constructed and interpretive nature of historical knowledge.	Knowing history	1, 2, 3, 4

Table 1. (*continued*)

Lesson nr.	Duration	Lesson topic and central historical question	Student and teacher activities	Primary focus	Historical thinking building blocks
5	2-3 x 50'	Parallel case-study of the Rwandan genocide or the India partition Historical question: To what extent could the negative consequences (being the outbreak of genocide) of the withdrawal of UN troops from Rwanda in 1994 have been foreseen by the actors involved? Historical question: To what extent could the British and Indian leaders involved foresee the negative consequences of the rapid British withdrawal and partition of British India in 1947?	Students work in groups on one of the historical inquiry task, supported by scaffolds. Teachers coach students during the inquiry task and provide feedback on the final products.	Doing history	1, 2, 3, 4
6	1 x 50'	Postcolonial relations between the West and former colonies Historical question: To what extent did the independence of the former colonies after 1945 constitute a radical break in their relationship with the West?	Students evaluate the usability and reliability of sources and substantiate a historical claim. Teachers provide explicit teaching on continuity and change as a historical mode of reasoning.	Knowing history	1, 2, 3
7	1 x 50'	Migration after 1945 Historical question: How did migratory patterns to (Western)-Europe evolve in the period 1945-2000s?	Students evaluate the reliability of sources.	Knowing history	1, 3
8	1 x 50'	Migration and integration Historical question: What was the main cause of integration problems experienced by immigrants in Belgium in the 1970s-1990s?	Students work individually and in groups to evaluate two competing historical representations, based on sources. Teachers coach students and lead class-discussion on the inquiry task.	Doing history	1, 2, 3

Note. Table adapted from Wilke et al. 2022b.

The design principles outlined above were applied in the following ways.

A close adherence to current insights from historical research was ensured via an exploration of recent historiographical literature on decolonization (such as from the historiographical school of new imperial histories). Each lesson was designed around a historical question, connected to an authentic historiographical and/or (historical) societal debate. Lesson 2, for instance, centered around the question who was most responsible for the decolonization of Belgian-Congo. In this lesson, the traditional narrative of independence being granted by the colonizers was questioned: students studied sources demonstrating both Belgian and Congolese agency. In lesson 3, evolving historical representations of the colonial past were examined, connecting to the notion of collective memory and societal debates about the colonial past.

Authentic historiographical debates provided the basis for the inquiry tasks that were used in four of the lessons. These tasks centered around an evaluative question connected to a historiographical debate (see Table 1 for an overview of the historical questions), such as "To what extent could the British and Indian leaders involved foresee the negative consequences of the rapid British withdrawal and partition of British India in 1947?". A number of historical sources were provided representing different answers to the historical question and reflecting different perspectives within that historiographical debate. Students had to compare and critically evaluate the value of these sources to generate an answer to the historical question. They then had to substantiate their claim with arguments and evidence and engage with counterarguments.

In lesson 1, teachers modelled the inquiry task, demonstrating the various steps in providing a substantiated answer to a historical question based on critical source analysis. In subsequent inquiry tasks, teachers provided scaffolds and coaching to the students. Scaffolds were provided for the evaluation and corroboration of sources, the weighing of evidence and the writing of a substantiated answer (modelled after University of Michigan, 2021 and Monte-Sano et al., 2014). They were designed to fade out during the lesson series. For instance, regarding the evaluation and corroboration of sources, the initial scaffold used in lesson 2 provided prompts for each step in the source analysis for each of the sources (e.g., questions regarding various aspects of a source's reliability: the author's positionality, the goal and audience of the source, the societal context,

etc.). Subsequent scaffolds that were used in lesson 5 only reminded students to assess the sources' reliability (see Figure 3). Similarly, initial scaffolds for providing a substantiated answer provided a writing scheme, whereas scaffolds used in the lesson series only provided exemplary sentences on which students could rely.

Figure 3. Scaffolds for source analysis in lesson 2 and lesson 5.

SOURCE 4: MEMOIRES OF GASTON EYSKENS (PUBLISHED IN 1993)	
RELIABILITY	
Consider the following information	
Who was the **author**? What function/position did the author hold? What was their social background, occupation, worldview, beliefs?	Gaston Eyskens was the Belgian prime minister at the time of Congo's independence. He was a member of the Catholic Party.
For what **purpose** did the author create this resource? Who was the author's target **audience**?	The excerpt stems from a memoire, a specific genre in which authors seek to present and preserve their personal views of events. In doing so, authors often portray themselves in a positive light and attempt to justify their own actions and role in the events. In this way, they can counter any criticism from opponents. Memoires are intended to be read by a wide audience.
When and **where** was this source created?	The mémoires were published in 1993, several decades after Congo's independence and after Eyskens' death.
In what **societal context** was the source created?	At the time at which the source was created (1990s) societal views on the colonial past had shifted strongly. There was a much more critical view of the colonial past, compared to the 1960s.
What **information** does the author rely on?	The author draws on his own memories and notes from the period when he was prime minister. He was an eyewitness to these events and closely involved in the events. The mémoires were published several decades after the events: the author therefore already knew how the situation would develop. This may have influenced the way he described the events.
What does this information tell you about the value of this source?	

Figure 3. (*continued*)

SOURCE	USABILITY *What does this source tell you about the consequences of the decision? What arguments does the source use?*	RELIABILITY *How reliable is this source to answer the research question? Consider the following elements: author, goal and audience, time and place, societal context, information*	PERSPECTIVE AND REPRESENTATIVENESS *What perspective does the source reflect? To what extent is the perspective in this source broadly shared? Does the source represent a larger group?*	CRITICAL SOURCE CORROBORATION *Are there other sources that support or refute this source's position?*
Source 1: Nehru				
Source 2: Jinnah				
Source 3: Mountbatten				
Source 4: Wolpert				

Teachers coached students during the lesson series, among others by providing the appropriate level of scaffolding, but also by providing feedback on students' inquiry tasks. To this end, an evaluation rubric was designed. The initial rubric was adapted from Monte-Sano (2012), Monte-Sano and De La Paz (2012), and Monte-Sano, De La Paz and Felton (2014) and included five criteria on a four-level scale: position statement, argumentation based on sources as evidence, evaluating a source's reliability, evaluating a source's representativeness and engaging with counter arguments. Table 2 provides an overview of the different competency levels for the initial criterion 'argumentation based on sources as evidence'.

Table 2. Competency levels for argumentation based on sources as evidence

1	2	3	4
The claim is not supported by evidence from sources.	The claim is supported by evidence from sources, but the evidence is often irrelevant or inaccurate. It is unclear to which source(s) the student refers.	The claim is supported by evidence from sources in a limited way. Selected sources and excerpts from sources are usually accurate and relevant to the claim. In most cases, it is clear which source(s) the student refers to.	The claim is supported by evidence from sources. Selected sources and excerpts from sources are accurate and relevant to the claim. It is clear which source(s) the student refers to.

Explicit instruction was provided on the application of substantive, second-order and procedural knowledge (e.g., reliability and how to assess it, representativeness, source corroboration). In the first lesson for instance, the teacher provided explicit teaching on the second-order concepts of reliability and source corroboration and on procedural knowledge related to critically assessing sources and providing a substantiated answer to a historical question. In lesson 2, explicit teaching on the notion of representativeness was added. In lesson 3, explicit teaching on these notions was repeated during a teacher-guided exercise on source analysis. Explicit teaching was mainly directed at second-order and procedural knowledge related to the processes of constructing a substantiated historical representation and applying historical modes of reasoning (e.g., causality in lesson 4). On two occasions (lesson 3 and 4), this was also used explicitly to teach students about the interpretive and constructed nature of historical knowledge.

To facilitate articulation and reflection, students worked in groups on the inquiry tasks. Whole-class discussions were used to discuss students' findings afterwards and to allow teachers to provide feedback and ask questions to advance students' historical thinking. For instance, in lesson 2, students' inquiry tasks were

discussed in class, and this was used as an opportunity to revise second-order and procedural knowledge addressed in lesson 1.

Exploration was not included in the lesson series as this requires a high level of autonomy which would not be feasible to obtain within the time span of 12 to 14 history lessons, but would rather be achieved over the course of an entire year of history lessons.

Teacher and student materials

To support teachers in implementing the lesson series, the first construction phase included the development of teacher and student materials. For teachers, a document was developed providing an explanation of the various building blocks of historical thinking in the Flemish history standards, and detailing the structure and goal of the lesson series as well as the design principles. A lesson plan was provided for each lesson following a four-part structure. First, an indication of which aspects (building blocks) of historical thinking were addressed. Second, an explanation of the lesson's main objectives and learning goals both with regard to knowing history and doing history, so as to demonstrate the interconnectedness between both. Third, an explanation of the pedagogical approach(es), including a description of teachers' and students' activities and a connection to the main design principles. Fourth, a detailed step-by-step script, divided in lesson phases with a proposed timing, instructions for students, sources, guiding questions for the sources and model answers. For each lesson, furthermore, a PowerPoint presentation was created. The evaluation rubric was also included in the teacher materials.

For students, a sourcebook was designed including the sources being discussed in the lesson series, along with the various scaffolds. In this first construction phase, no additional student materials were created as we expected students to be able to take notes by themselves.

EVALUATION OF THE INITIAL LESSON SERIES (1ST EVALUATION AND REFLECTION PHASE)

Expert review

To evaluate the initial lesson series' connection to current historiographical insights, its adherence to the predetermined design principles, and its feasibility and usability an expert review was set up with three types of experts: a historian, (history) educational scholars and experienced history teachers and history teacher trainers.

In terms of the lesson series' adherence to recent historiography, the historian generally agreed with the proposed content of the lesson series and provided suggestions for additional (very recent) historiographical literature and source material that could be integrated in the lesson series, for instance regarding the lesson on British-India (lesson 1 and 5) and societal debates on migration and integration in Belgian society (lesson 8).

The (history) educational scholars, experienced history teachers and teacher trainers confirmed that the initial lesson series adequately applied the proposed design principles. Although the design principles in themselves were considered valuable, two concerns were raised about how they were implemented in the materials, in particular in light of the material's feasibility and usability. First, regarding the lesson series' use of inquiry tasks, the need was expressed to provide more support for explicit teaching, modeling and coaching on these tasks. In particular, the experts pointed to (variation between) teachers' own second-order and procedural knowledge, for instance regarding (how to assess) reliability as this was not always included in teachers' initial training. The experts therefore suggested that some teachers may need more support to take up their role as experts during modeling, coaching and explicit teaching.

Second, the experts deemed the lessons to be quite challenging for students, among others because of the absence of worksheets or a textbook, the extensive reliance on (textual) sources and the use of inquiry tasks. Although they did not oppose the use of inquiry tasks, they did caution that, as students are mainly accustomed to teacher-directed history classes, this would constitute a new and challenging way of engaging with history. They therefore proposed to elaborate the supporting materials, in line with the explicit teaching done by the teacher on

second-order and procedural knowledge, to provide more guidance during the lesson series and to accommodate differences in students' familiarity with historical thinking.

Pilot study

A pilot study examined the applicability of the evaluation rubric. Two classes in the 12th grade (36 students in total) completed an inquiry task, after which students' answers were evaluated by the design team using the initial evaluation rubric. The pilot test confirmed that a rubric constituted a suitable way of evaluating students' inquiry tasks. However, it became clear that the rubric was quite complex to use, that the evaluation criteria were not clearly delineated, and that more clarity was needed in the descriptions of the various proficiency levels.

REVISION AND DIGITIZATION OF THE LESSON SERIES (2ND DESIGN AND CONSTRUCTION PHASE)

Revision of the lesson series

Based on the evaluation of the lesson materials via the expert review and pilot study, the initial lesson series was adjusted. Table 3 provides an overview of the main issues identified in the initial lesson series and the changes that were made to the lesson series to accommodate them.

Table 3. Revision of the lesson series after the first evaluation phase

Issues identified in expert review	Adaptation in second design phase
• Differences in teachers' expertise regarding historical thinking • Difficulty level for students • Insufficient supporting materials for teachers and students • Reliance on textual sources	Teacher materials • Explanatory texts on second-order and procedural knowledge • Suggestions for coaching students • Additional scaffolds for students Student materials: • Development of a learning text • Development of worksheets • Development of additional scaffolds • Addition of visual sources
Issues identified in the pilot study of the evaluation rubric	**Adaptation in second design phase**
• Criteria not clearly delineated • Ambiguity in descriptions of proficiency levels • Complex to use	• Merging and addition of criteria • Revision of descriptors • Development of a user guide with exemplary essays

Revision of teacher and student materials

Considering that there were no main issues regarding the lesson series' connection to historiography, the revision of teacher and student materials mainly focused on the difficulties identified by the (history) educational scholars and experienced history teachers. In the lesson plans for teachers, additional clarification was provided concerning the second-order and procedural knowledge which required explicit teaching from teachers. Explanatory texts were added, directed at teachers

so as to provide them with further knowledge about the precise meaning of these concepts and procedures (see Figure 4 for an example). This way, we aimed to bolster teachers' confidence and knowledge regarding the teaching of historical thinking and intercept possible differences between them. Furthermore, we provided concrete suggestions for how to coach students (and allow for differentiation), for instance by providing guiding questions that teachers could use to give students additional support (see Figure 5) and by developing additional scaffolds in the student materials (see further).

Figure 4. Example of an explanatory text on source corroboration in the lesson script.

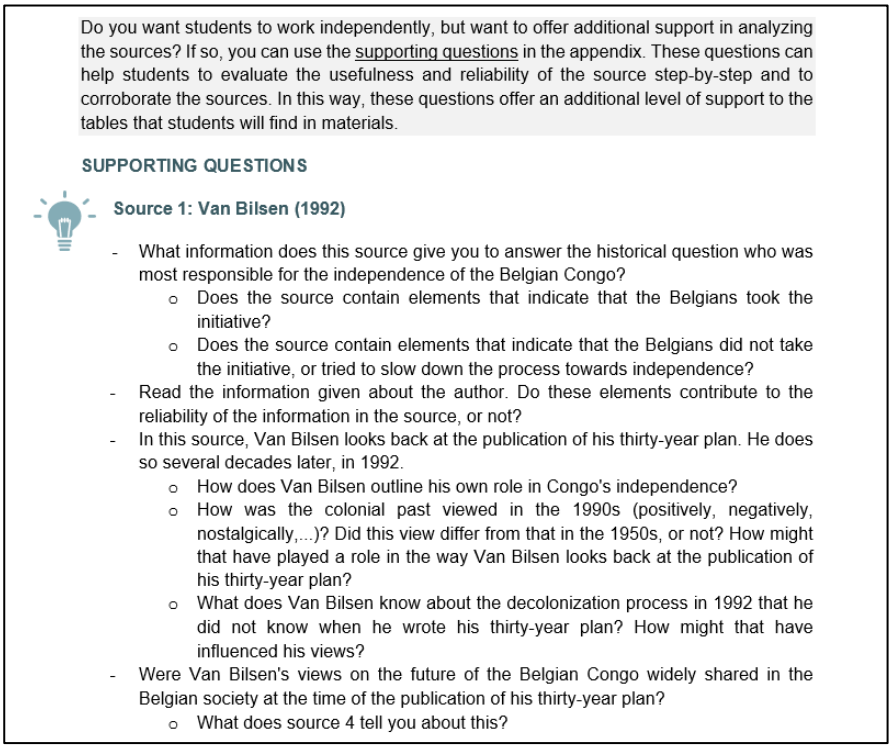
SOURCE CORROBORATION

During source corroboration, multiple sources that contain information regarding a particular historical question are compared to each other. The information in source can complement each other, when sources provide additional information on a topic, or can confirm each other, when several sources contain the same information.

When confronting sources, sources may also contradict each other. This is the case when the information in the sources does not match. This may be because the source is written from a different perspective (multiperspectivity), but may also indicate that the information in one of the sources is less reliable.

Source corroboration is essential to provide a nuanced and substantiated answer to a historical question.

Figure 5. Example of suggestions for coaching during the inquiry task in lesson 2.



Do you want students to work independently, but want to offer additional support in analyzing the sources? If so, you can use the supporting questions in the appendix. These questions can help students to evaluate the usefulness and reliability of the source step-by-step and to corroborate the sources. In this way, these questions offer an additional level of support to the tables that students will find in materials.

SUPPORTING QUESTIONS

Source 1: Van Bilsen (1992)

- What information does this source give you to answer the historical question who was most responsible for the independence of the Belgian Congo?
  - Does the source contain elements that indicate that the Belgians took the initiative?
  - Does the source contain elements that indicate that the Belgians did not take the initiative, or tried to slow down the process towards independence?
- Read the information given about the author. Do these elements contribute to the reliability of the information in the source, or not?
- In this source, Van Bilsen looks back at the publication of his thirty-year plan. He does so several decades later, in 1992.
  - How does Van Bilsen outline his own role in Congo's independence?
  - How was the colonial past viewed in the 1990s (positively, negatively, nostalgically,...)? Did this view differ from that in the 1950s, or not? How might that have played a role in the way Van Bilsen looks back at the publication of his thirty-year plan?
  - What does Van Bilsen know about the decolonization process in 1992 that he did not know when he wrote his thirty-year plan? How might that have influenced his views?
- Were Van Bilsen's views on the future of the Belgian Congo widely shared in the Belgian society at the time of the publication of his thirty-year plan?
  - What does source 4 tell you about this?

Adjustments in student materials were made to provide students with step-by-step guidance during the lesson series. To support them in taking notes, worksheets were developed based on the Cornell note-taking method (Cornell University, n.d., based on: Pauk & Owens, 2010) dividing worksheets in a left-column with keywords that followed the structure of the lesson and a right column for note taking. At the bottom, students could make a short summary or reflection of the lesson. A learning text was created, summarizing the content of each lesson. This functioned as a reference work that students could consult besides their own notes. For the development of this text, we looked for guiding principles supporting to increase text comprehension in history. Based on the work of Land and colleagues (2009) we aimed for this learning text to be coherent, include transition words (e.g., because, however, besides) and be written in a factual and professional style.

Besides the development of worksheets and a learning text, additional scaffolds were added to students' materials which took the form of visual

organizers and summarizing documents for second-order and procedural knowledge, in line with explicit teaching by the teachers. For instance, a summarizing document was made regarding 'reliability of sources' (Figure 6) as well as a visual organizer depicting the process of answering a historical question based on sources (Figure 7). These materials mirrored the explicit teaching provided by the teacher. To balance out the strong reliance on textual sources, additional visual support was provided via the use of visual sources, illustrations and videos, both in the PowerPoint and student materials.

Figure 6. Example of summarizing document on reliability of sources

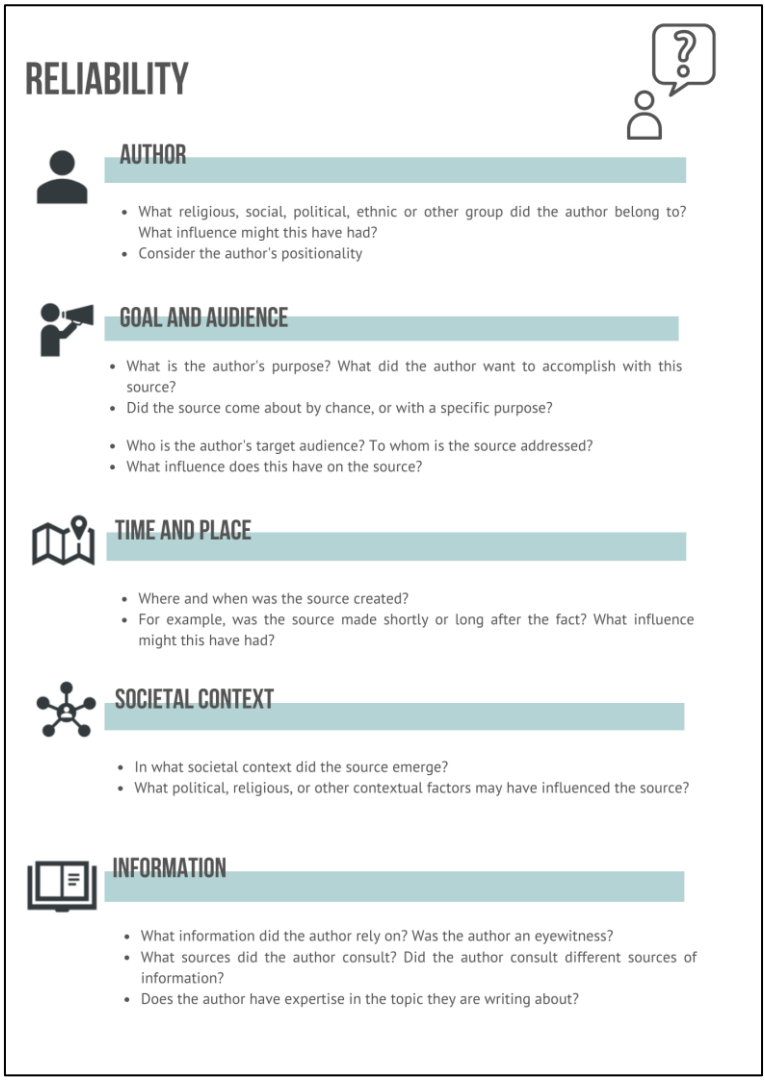
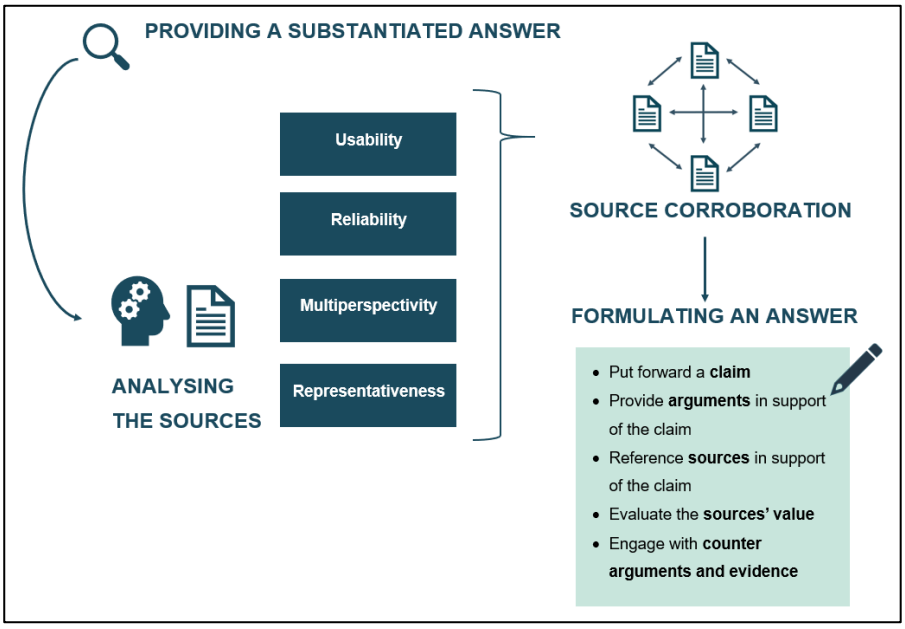


Figure 7. Visual organizer on providing a substantiated answer to a historical question (procedural knowledge)



Revision of the evaluation rubric

To overcome the issues with the evaluation rubric identified in the pilot study, the rubric's criteria and descriptors of proficiency levels were revised. The criterion 'argumentation based on sources as evidence' was broken down into three criteria: argumentation (structure-oriented: number of arguments, whether they were used explicitly or implicitly); quality of the argumentation (content-oriented) and use of sources as evidence (see Table 4 for the revised competency levels). The criteria 'evaluating a source's reliability' and 'evaluating a source's representativeness' were merged into one criterion 'evaluating the sources' value'. To support teachers in applying the rubric, a guide was developed explaining the various criteria and providing examples of evaluated student essays. A simplified version of the rubric was also developed for use as a (peer-)feedback instrument by teachers or students.

Table 4. Competency levels for the revised criterion of argumentation based on sources as evidence

Criterion	1	2	3	4
Argumentation in support of the claim, based on sources	No arguments derived from the sources are used to support the position taken	The position taken is supported only to a limited extent by arguments derived from sources: • One or more arguments are mentioned • But: the argument(s) are not used as such/are not explicitly put forward in support of the position taken (e.g. only descriptive)	The position taken is mostly supported by arguments derived from sources: • One or more arguments are mentioned • But: only some of the arguments are explicitly put forward in support of the position taken	The position taken is fully supported by arguments derived from sources • Multiple arguments are mentioned • And: arguments are explicitly put forward in support of the position taken
Quality of the argumentation in support of the claim	No arguments, derived from the sources, are used to support the position taken	The argumentation derived from the sources is incorrect: One or more arguments are mentioned but none of the arguments are correct.	The argumentation derived from the sources is partially correct: At least one correct argument is mentioned.	The argumentation derived from the sources is correct: At least two correct arguments are mentioned.
Use of sources in support of the claim	The position taken is not supported by concrete sources. Or sources are referred to only in a general way (e.g. 'based on the sources I argue that…')	The position taken is supported only to a limited extent by concrete sources. • Only one concrete sources is used in support of the position taken. • Other sources are not referred to concretely or not mentioned at all..	The position taken is mostly supported by concrete sources: • Multiple sources are mentioned explicitly in support of the position taken. • But: not all sources are referred to concretely (it is not always clear which source is referred to) or sources are not put forward in support of the position taken	The position taken is fully supported by concrete sources Multiple sources are mentioned explicitly in support of the position taken.

Digitization of the lesson series

The Covid19-pandemic prompted an additional construction phase. The corona measures in Flemish education only allowed for half of the lessons to take place in school, the other half had to be organized via (synchronous) online education or via (asynchronous) digital modules. As schools could individually choose for a synchronous or asynchronous policy, and as they use a variety of educational learning platforms, it was not possible to design a completely ready-made digitized version of the lesson series. Rather, we designed a range of materials for remote education that could be used easily and flexibly by teachers.

In designing these materials, maximum adherence to the initial design principles was pursued. In this respect, the corona circumstances mainly created challenges for teachers' role in modeling, coaching and explicit teaching, and for interaction. For synchronous remote education, digitized interactive presentations were created via Nearpod which allowed for whole-class interaction in an online environment. Suggestions were provided for online tools (such as google forms) allowing students to interact and work together on a task. For asynchronous remote education, a collection of short videos was developed by the design team. The videos were developed for substantive knowledge and for the acquisition of second-order and procedural knowledge. They hence provided a substitute for modeling and explicit teaching by the teacher. We also constructed step-by-step guides for the development of digital learning modules, based on the developed videos as well as on online exercises which provided students with immediate feedback on their answers (via edpuzzle).

Concrete guidelines were provided for teachers as to which lesson (segments) to preferably organize in class and which in remote education. It was recommended that whole-class discussions and group work take place in the classroom, in order to facilitate interaction among students and coaching by the teacher.

INTERVENTION STUDY (2ND EVALUATION AND REFLECTION PHASE)

The revised lesson series was evaluated by means of an intervention study aimed at examining the social validity of the lesson series as experienced by teachers, and the effectiveness of the materials in fostering students' historical thinking. This would also allow us to examine the effectiveness of the underlying design principles as well their concrete application in the materials.

A call was launched among Flemish history teachers to find teachers willing to implement the lesson series in their practice. Between January and April 2021 twenty-four teachers (1115 students) tested the materials. Teachers participated in a training session which included an explanation of the design principles, the goals of the lesson series and the lesson plans Another six teachers (266 students) functioned as a control group.

To evaluate the social validity of the lesson series, teachers were asked about their experiences with the materials via online questionnaires and interviews. They completed a questionnaire after each lesson registering how the lesson was administered (in-class or remote (a)synchronous) and general feedback regarding the feasibility and applicability of each lesson. To oversee the implementation of the lesson series, we also asked teachers to indicate whether they had deviated from the provided lesson plans and if so, how and why. Two semi-structured interviews were conducted with each teacher, one after the second or third lesson, another at the end of the lesson series. These interviews elaborated on teachers' evaluations in the online questionnaire and gathered in-depth information on whether they found the materials to be relevant and useful in practice.

To examine the effectiveness of the lesson series on students' historical thinking, students completed a digital pre- and posttest. This pre- and posttest consisted of a historical inquiry task, similar to those of the lesson series. Students also answered two open-ended questions measuring procedural knowledge related to providing a substantiated answer to a historical question. Lastly, the pre- and posttest examined changes in students' epistemological beliefs based on the statements of Wiley and colleagues (2020).

Social validity of the lesson series: teachers' evaluation of the lesson series

Teachers overall evaluated the lesson series in a positive way. Concerning knowing history, they generally appreciated the lesson series' adherence to recent historiography and the inclusion of historiographical debates as it is challenging for them to keep up to date with academic research.

Most outspoken, however, were teachers' comments regarding the way that doing history was integrated throughout the lesson series, which was experienced as a significant change from their conventional practice: not only because doing history was given more attention than in teachers' regular practice, but also because it was covered in a thorough and in-depth way. Teachers particularly appreciated the material's attention to doing history. The use of explicit teaching and modeling on second-order and procedural knowledge was considered a great asset of the material. During the interviews, most teachers reported that the explicit teaching and modeling aspects in the lesson series made them aware that their own practices often touched upon aspects of doing history only briefly and superficially, and implicitly. Furthermore, some concepts included in the lesson series (such as sources' representativeness or usability) were not yet covered in their own teaching practice. The detailed attention in the lesson series to second-order and procedural knowledge was thus not only considered an added value for their students, but also for themselves. Explicit teaching and supporting (scaffolding) documents regarding second-order and procedural knowledge, as well as the explanatory videos on these aspects were mentioned as being particularly valuable in light of the introduction of the new history standards, also outside of this particular lesson series. Teachers also expressed great appreciation for the extensive set of sources provided in the material and in particular in the inquiry tasks as this is one of the challenges they face when designing instructional materials. Moreover, they acknowledged the usefulness of confronting students with inquiry tasks as an effective way to foster and evaluate students' historical thinking. While the evaluation rubric was generally acknowledged as a useful instrument for providing feedback on these inquiry tasks, it was also considered time-consuming. Teachers were therefore looking for additional support to provide feedback or evaluate aspects of doing history in a more concise way.

Overall, teachers' feedback demonstrated a great appreciation for the lesson series' underlying design principles, such as the use of inquiry tasks, explicit teaching and modeling. Teachers, however, did identify a number of issues regarding the practical applicability and usefulness of the lesson series, mainly related to how we gave shape to the inquiry tasks. One difficulty in this regard was that the estimated timing was not always realistic. The majority of teachers spent more time on the inquiry activities than we had anticipated. As a result, some teachers could not complete the entire lesson series. Timing issues were often related to the difficulty of the inquiry tasks. Even though this was already addressed in the first revision of the lesson series, teachers still considered the use of (textual) sources and writing tasks very (in a few cases: too) demanding for their students, certainly in schools where students had lower literacy skills. For these students, the inquiry tasks were only possible with a lot of additional support and coaching by the teacher.

For a few teachers, the difficulty of the inquiry tasks also lead to concerns about students' motivation. They reported that as these tasks required a high cognitive load from students, students became less motivated and expressed a desire to return to 'normal' history classes (i.e. teacher-guided and with an emphasis on substantive knowledge). Although this issue was raised in particular among teachers for whom the lesson series' approach deviated strongly from their conventional practice, the majority of teachers shared the concern that the lesson series in general was rather demanding for students.

From our part, the interviews with teachers made clear that we had underestimated the level of variance in students' historical thinking. Considering that some students did not have a lot of experience with doing history, concerns regarding the difficulty level of the materials were not entirely unwarranted. In light of this, the lesson series' focus on historical thinking in a holistic way as well as the use of multiple inquiry tasks was indeed quite ambitious in the time span of 12-14 lessons.

Because of the Covid19-pandemic, it is difficult to fully separate concerns about the lesson series in particular from the more general impact of the pandemic-related restrictions. During the interviews, teachers repeatedly mentioned that remote education, combined with other corona restrictions in society, took a high toll on students' motivation and learning in general. Despite

the materials which we had developed for remote education, teachers also reported that essential aspects of the lesson series could not always be fully implemented. It remained challenging to adequately coach students, provide appropriate scaffolding and organize whole-class discussions via remote education. These circumstances may thus have added new challenges to the lesson series and amplified existing ones (e.g., regarding students' motivation).

Effectiveness of the lesson series: students' historical thinking ability

To examine the effects of the lesson series on students' historical thinking, students' performance on the pre- and posttest were analyzed and compared to a control group using multilevel analyses (see Wilke et al., 2022b). For these analyses, only students who had completed the entire lesson series were included ($n = 402$).

The multilevel models showed that, when controlling for differences in pretest scores, the lesson series significantly improved students' performance on the historical inquiry task ($B = 4.79$, $p < 0.01$) as well as their procedural knowledge ($B = 2.06$, $p < 0.01$). The lesson series was hence effective in fostering students' historical thinking, in terms of historical inquiry skills as well as related procedural knowledge. Nevertheless, no effects were found regarding students' epistemological beliefs, and significant differences remained in place between students and schools for the inquiry task and for procedural knowledge.

FINALIZATION OF THE LESSON SERIES (3RD DESIGN AND CONSTRUCTION PHASE)

Based on the results of the intervention study, a final revision of the lesson series was implemented. Considering that the lesson series was overall successful in improving students' historical thinking and no issues were raised regarding the content of the lesson series (in terms of knowing history) or the design principles, the final revision of the lesson series focused mainly on accommodating the issues raised by teachers connected to the practical applicability of the material.

To this end, the explanatory document was elaborated to not only detail the design principles and main goals of the lesson series but to also address teachers' concerns regarding the difficulty of the lesson series. It was explained

that the lesson series included multiple, inquiry tasks but that, because of their demanding nature, they should not necessarily be completed over the course of 12-14 history classes. Rather, the inquiry tasks functioned as concrete examples serving as blueprints for the design of tasks on different topics.

To accommodate possible issues with the difficulty of the tasks and students' motivation, the explanatory document emphasized that teachers could adapt them to their students' needs, considering their historical thinking as well as general literacy skills and knowledge. In the lesson plans, concrete suggestions were added for how to do so, among others by adapting the inquiry tasks in terms of length and form. For instance, concrete suggestions were given to reduce source sets from five to three sources and less demanding alternatives were offered for the writing tasks (e.g., additional writing schemes or presentation of the answer via concept maps or short videos on Flipgrid).

Timing estimates for the various lessons and lesson segments were adapted to correspond better with teachers' reported timings, but it was highlighted that these timings were only indicative and vary according to students' prior knowledge.

In the student materials, extra attention was paid to the structure of the materials. A Table of contents was added as well as cross-references between the various materials (e.g., from the learning text to the related inquiry tasks). A visual representation of the structure of the entire lesson series was also added. At the end of each lesson, its specific learning goals were listed, distinguishing between knowing and doing history. The summarizing documents and visual organizers on second-order and procedural knowledge were brought together in a separate 'doing history'-booklet to be also used outside of this specific lesson series.

To support teachers in providing feedback and evaluating aspects of historical thinking, an evaluation booklet was designed containing exemplary questions and tasks based on novel sources connected to the theme of decolonization that could be used for evaluation, including model responses and grading guidelines. The format of these questions could also be used as blueprints for evaluating historical thinking in relation to other historical themes.

REFLECTION & OUTPUT OF THE DESIGN RESEARCH

The aim of this design-based research was to gain more insight into the design of instructional practices that are both effective in fostering historical thinking and are considered socially valid by teachers.

This research first of all shows that the design principles identified in the literature and applied in the lesson series were overall effective in promoting historical thinking, hence providing additional evidence for the effectiveness of principles derived from cognitive apprenticeship. By striving for a close connection to historiography, the lesson materials also managed to address historical thinking in a holistic way and bridge the gap between school and academic knowledge, thus overcoming the disconnect between both.

By involving and supporting teachers in the design and implementation of the lesson series, we were also able to create materials that were overall considered socially valid. However, this study clearly demonstrated the difficulty of translating design principles into concrete materials that teachers consider feasible. The focus in the various iterations throughout the design process was not so much on (the accurate implementation of) the design principles themselves, but rather on their practical application. This study shows that the implementation of the design principles requires multiple evaluations and adaptations in light of various practical and contextual factors, such as teachers' prior knowledge on historical thinking, the available time, students' motivation and literacy skills. In some cases, a trade-off may even be necessary between what is theoretically the best approach and what is concretely feasible given the context. For example, formulating a written answer to an inquiry task is considered extremely useful to foster students' historical thinking (Nokes & De La Paz, 2018). However, as this is both time-consuming and possibly demotivating for at least some students, it may be necessary – in light of contextual and practical factors - to alternate these written answers with other answer formats.

Considering that obtaining social validity is essential to ensure that a curriculum is also used outside the context of one distinct study, this study stresses again the importance of involving teachers closely in the design process and to take their comments seriously.

Despite the overall positive results of the designed materials, several challenges remain in place. First, although the lesson series succeeded in fostering students' historical thinking, it failed to generate changes in students' epistemological beliefs. Considering that these beliefs underlie historical thinking it remains important to consider how such epistemic changes may be achieved. In the design of future interventions, it may be fruitful to include explicit attention to epistemological reflection as this seemed to generate positive results with regard to other aspects of historical thinking. The development of these beliefs may, however, also be more of a goal for history education as a whole, rather than of a single, confined lesson unit. Second, teachers expressed concern about students' motivation, both with regard to working with textual sources and with regard to having to take on an active role in group or independent work. Regarding the former, we made further adjustments to the lesson materials to decrease the reliance on textual sources and to allow for a better balance between visual and textual sources. However, we cannot yet say whether these adaptations are sufficient. Also, further adjustments in this regard are not easy to accomplish, as history is inherently a very linguistic subject in which working with (textual) sources remains a core aspect. Students' motivation for taking on an active role during inquiry tasks may be less affected when these tasks are more dispersed throughout the year, as was recommended in the revised materials. Another way to enhance students' motivation might be to allow them to formulate and investigate their own historical questions. This might increase the authenticity of the inquiry tasks (Freedman, 2015) and offers an opportunity to include the principle of exploration into the materials. This requires, however, a great deal of autonomy from students and may hence be more suited towards the end of secondary education, after several years of immersion in the act of thinking historically. A final challenge is the consolidation of the research's contribution to practice and the dissemination of the materials outside the group of teachers involved in the design process. To facilitate this, the materials will be used in professionalization initiatives for teachers and were made freely available via the website of the university's teacher training program to be used as such in lessons on decolonization, yet also on a more conceptual level, as a design blueprint for the development of new instructional materials.

Chapter 5

Fostering democratic citizenship through historical thinking? A cluster randomized controlled intervention study.[1]

Historical thinking has gradually become a central goal for history education. Its complexity, however, makes it difficult to foster among students and little is known about the role of students' epistemological beliefs. At the same time, several history education scholars have argued that historical thinking may not only contribute to disciplinary goals, but also to civic ones. They assume students to spontaneously transfer historical skills, knowledge and dispositions to contemporary contexts. Although there is some evidence to support such a spontaneous transfer, general educational literature indicates that a transfer of learning is difficult to obtain. This study examines the disciplinary effects of a lesson series aimed at fostering students' historical inquiry competence via a cluster randomized controlled intervention study in Flemish history education (control: $n = 266$; intervention: $n = 402$). It also examines whether a spontaneous transfer to contemporary topics occurs. To this end, a pre- and posttest examines (1) changes in students' historical inquiry competences and epistemological beliefs about history and (2) the transfer to democratic skills and dispositions.

Multilevel and linear regression models showed that the lesson series was effective in enhancing students' historical inquiry skills and procedural knowledge; no effect was found on their epistemological beliefs. The lesson series did not directly improve students' democratic skills and dispositions. However, changes in students' historical inquiry competence did significantly predict their democratic skills at posttest, although in a limited way. These results caution for an excessive optimism regarding the idea that historical thinking automatically contributes to democratic citizenship.

[1] This chapter is based on: Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2022). Fostering democratic citizenship through historical thinking? A cluster randomized controlled intervention study. *Contemporary Educational Psychology, 71, 102-115.* https://doi.org/10.1016/j.cedpsych.2022.102115

INTRODUCTION

Since the 1990s, historical thinking has become a central goal for history education across several Western counties (Lévesque & Clark, 2018). Historical thinking combines 'knowing history' and 'doing history' (Havekes et al., 2012). Knowing history refers to the acquisition of substantive knowledge about the past, related to the content of what historians study. Doing history points to the procedural knowledge which allows students to gain an understanding of how historical knowledge is constructed and allows them to apply discipline-specific ways of reasoning (Lee & Ashby, 2000; Lévesque, 2009; van Drie & van Boxtel, 2008). These disciplinary ways of reasoning include critical reasoning about sources in light of historical questions, reflecting about causality, historical contextualization or continuity and change (Seixas & Morton, 2013; Seixas, 2017; van Drie & van Boxtel, 2008; Wineburg, 2001; Van Nieuwenhuyse, 2020).

Historical thinking is a complex and 'unnatural' act, going directly against the way that students spontaneously look at and engage with the past (Wineburg, 2001). In general, students are more likely to look at the past through a contemporary lens and consider historical sources to be accurate, even exact reflections of the past. Spontaneously, they often struggle to adopt disciplinary ways of thinking which aim to understand the past in its own right and recognize that historical knowledge is interpretive and constructed, based on sources that are inherently biased (Wineburg, 1991; 2001). Students' naïve views about history are clearly reflected in the way they approach essential aspects of historical thinking, such as critical source analysis. As they tend to take sources at face value, students struggle, for instance, with the critical analysis of historical sources and with conflicting accounts in them. When dealing with multiple documents, students experience difficulties when asked to provide a global argument based on multiple documents and tend to summarize historical documents rather than critically compare them (e.g., Britt & Aglinskas, 2002; Nokes, 2017; Reisman, 2010; Rouet et al., 1996; Stahl et al., 1996; Wineburg, 1991; Young & Leinhardt, 1998).

Getting students to think 'historically' is therefore not an easy task and requires the application of specific instructional practices that make disciplinary ways of thinking explicit to students. In this regard, several intervention studies

have succeeded in improving aspects of students' historical thinking, such as historical contextualization (Huijgen et al., 2018), causal reasoning (Stoel et al., 2015; Stoel, van Drie & van Boxtel, 2017); critical source analysis (Nokes et al, 2007; Reisman, 2012), and argumentative writing in history (De La Paz & Felton, 2010; De La Paz et al., 2016). These studies often rely on principles such as explicit teaching, modeling, scaffolding and coaching derived from the general educational model of cognitive apprenticeship (Collins et al., 1991). They also make use of (writing) tasks that are authentic for the discipline of history, based on analyzing and comparing multiple (conflicting) historical documents (van Boxtel & van Drie, 2018). Although these studies have generally yielded positive results, they did not always find positive effects on all outcome measures related to historical thinking (e.g., Reisman, 2012; Sendur et al., 2021; Stoel et al., 2015). Moreover, although scholars point to the potential influence of affective dimensions and of students' epistemological beliefs, these are rarely taken into account in existing intervention studies. There is hence a need for more (encompassing) research into the promotion of students' historical thinking.

Another issue is the relationship between historical thinking and democratic citizenship. History education scholars have long assumed that historical thinking is able to spontaneously contribute to the development of skills and dispositions that are relevant to democratic citizenship. Barton and Levstik (2004), for instance, argue that historical thinking will contribute to democratic citizenship by promoting reasoned judgments, by preparing students to move beyond their own perspective and to take part in collaborative discourses about public issues. Various other scholars have similarly argued that fostering historical thinking will naturally contribute to aspects of democratic citizenship (e.g., Lee & Shemilt, 2007; Lévesque, 2009; McCully, 2012; Monte-Sano, 2016; Nokes, 2013; Reisman, 2012; Stearns, 1998).

The presumed relationship between historical thinking and democratic citizenship is based on the premise that there is an overlap between skills, (procedural) knowledge and dispositions required for historical thinking and those required for democratic citizenship, and that students will spontaneously transfer these from a historical to a contemporary context. Among others, it is expected that students' historical inquiry competences, i.e., their ability to construct substantiated historical representations, based on a critical analysis of

multiple sources and perspectives will transfer to democratic citizenship. Fostering these competences is expected to transfer to contemporary contexts, improving students' ability to make reasoned and rational judgments towards public debates, based on an evaluation of different perspectives, arguments and (contemporary) evidence.

Literature on transfers of learning in other domains, however, warns that establishing a far transfer, i.e., from one context to another, requires specific instructional strategies (Merill, 2002; Perkins & Salomon, 2012; van Merriënboer & Kirschner, 2007). Nevertheless, there is some evidence for such a relationship with regard to students' approach towards historical and contemporary sources (e.g., Reisman, 2012) although studies in this regard remain scarce. There are for instance few large-scale studies based on randomized controlled intervention designs that examine the possibility of such a transfer and studies on the relationship between historical thinking and democratic citizenship often occur in the context of post-conflict or divided societies. There is hence a need for more research to establish whether students spontaneously make a transfer from historical to contemporary contexts.

The brief overview of research on historical thinking brings forward several pressing issues regarding how to advance students' historical thinking, the role of epistemological and affective aspects and the potential benefit to democratic citizenship. In order to gain more insight into how to foster students' historical thinking and in order to evaluate the long-standing assumption that historical thinking will contribute to democratic citizenship, a cluster randomized controlled intervention study was set up in Flemish history education. This study aims to examine the disciplinary effects of a lesson series on historical thinking and to establish whether effectively fostering students' historical thinking may also contribute to aspects of democratic citizenship. Doing so will allow a better substantiated reflection on the potential role of history education in society.

THEORETICAL FRAMEWORKS

Fostering historical thinking among students

Historical thinking is operationalized in various ways (e.g., Seixas & Morton, 2013; Seixas, 2017; van Drie & van Boxtel, 2008; Wineburg, 2001; Van Nieuwenhuyse, 2020), with common aspects including (asking) historical questions, contextualizing and comparing historical phenomena, critically evaluating historical sources, argumentation, and the application of historical modes of reasoning such as causality, agency or continuity and change. It is a complex notion that requires the acquisition of substantive knowledge, second-order and procedural knowledge. Substantive knowledge relates to the content of history, describing historical periods, persons, structures or phenomena. Second-order knowledge refers to organizing concepts that historians use when studying the past, including concepts such as cause or consequence, evidence or continuity or change. Procedural knowledge then describes the strategies and heuristics applied by historians when constructing historical representations. Such procedures include, among others, the evaluation of a source's reliability, how to write an account of the past and how to interpret historical context (Lee, 2004; van Drie & van Boxtel, 2008; VanSledright, & Limón, 2006). Studies have shown that there is not always a direct one-on-one relationship between students' procedural knowledge as such, and students' application of this knowledge in particular tasks (e.g., Stoel, van Drie & van Boxtel, 2017). For instance, even when students are knowledgeable about how to source, they do not necessarily apply this knowledge when engaged in multiple documents tasks (Paul et al., 2017). Van der Eem and colleagues (2022) found that while procedural knowledge about evaluating the reliability of sources was a significant predictor for their performance on a sourcing task, students actually applied more criteria for establishing the reliability of sources than they were able to list. It therefore appears beneficial to consider students' procedural knowledge separately from their application in concrete tasks.

An essential aspect of historical thinking is the ability to construct substantiated representations, based on a critical analysis of multiple sources (De La Paz & Felton, 2010; Monte-Sano, 2010, 2016; van Boxtel et al., 2021; van Boxtel & van Drie, 2017; van Drie & van Boxtel, 2008; Voet & De Wever, 2017b; Van

Nieuwenhuyse, 2020). This ability is often examined using (multiple) documents based questions (DBQs) or inquiry tasks. These tasks invite students to answer a historical question via an analysis of multiple sources expressing different perspectives (Sendur et al., 2021; van Boxtel et al., 2021; Voet, 2017). In line with inquiry tasks in domains outside of history (Chinn & Duncan, 2021), they are considered authentic tasks that closely reflect the nature of the discipline (van Boxtel & van Drie, 2018). DBQs and inquiry tasks are closely related. In fact, van Boxtel, Voet and Stoel (2021) argue that within the field of history education, 'inquiry learning', 'historical inquiry', 'document-based lessons' and 'doing history' are often used interchangeably.

In such tasks, students are required to do far more than integrate the information found in different sources. Rather, to provide a substantiated answer to the historical question, students have to apply discipline specific reasoning strategies to evaluate the value of the available information and sources (Rouet & Britt, 1996). Wineburg's (1991) seminal expert-novice study has been of great influence in this regard. Wineburg distinguished three disciplinary strategies used by professional historians when evaluating the value of multiple documents: sourcing (i.e., examining the author, time and place of the document's creation), contextualization (i.e., placing documents in their historical context) and corroboration (i.e., comparing information across various sources). He also established a connection between the use of these disciplinary ways of reasoning and particular epistemological views about the nature of history and historical inquiry, which differed between experts and novices. The study found that students (novices) tended to view documents as carriers of information, whereas historians regarded them as evidence reflecting one particular point of view and whose content is inherently influenced by that particular view. These views on the nature of sources reflect differences between students' and expert historians' epistemological beliefs (Wineburg, 1991). These refer to beliefs about the nature of knowledge and processes of knowledge construction in the discipline (Buehl & Alexander, 2001). In relation to history, advanced epistemological beliefs occur when one accepts that historical knowledge is constructed, interpretive, and subject to change, and that disciplinary criteria can be applied to assess various claims about the past (Maggioni et al. 2009; Stoel, Logtenberg et al., 2017). The

development of advanced epistemological beliefs is hence an important aspect of students' historical thinking (Lee & Ashby, 2000; Stoel, Logtenberg et al., 2017).

Several studies have elaborated on the disciplinary ways of engaging with sources initially identified by Wineburg (1991), by elaborating the criteria for evaluating a source's value with aspects such as the author's motives, the source type, the information on which the source is based or the source's representativeness (e.g., De La Paz & Felton, 2010; Hicks, Doolittle & Ewing, 2004; Perfetti et al., 1994; Van Nieuwenhuyse, 2020; Wilschut et al., 2012). Scholars have also put forward additional strategies which are adopted in disciplinary reading, such as the need to actively consider the author's arguments, use of language, tone and word choice. This is referred to by Reisman (2012) as 'close reading'. While close reading is not exclusive to historical documents, it is important, Reisman (2012) continues, for students to closely examine the author's wording in order to understand a document's subtext, as opposed to a more 'passive' reading which is often conducted when reading history textbooks. In an overview of discipline specific cognitive processes identified in various studies on historical inquiry, Voet and De Wever (2017) categorized them in five main aspects describing the process of historical inquiry as one of sourcing, appraising, specifying, constructing and arguing.

Studies building on the work of Wineburg (1991) have also used DBQs or inquiry tasks to further examine students' disciplinary ways of reasoning. These studies have confirmed and elaborated on Wineburg's finding, establishing that students tend to take the content of historical sources at face value, rarely evaluate sources and are more inclined to summarize, rather than evaluate sources (e.g., Britt & Aglinskas, 2002; Nokes, 2017; Reisman, 2010; Rouet et al., 1996; 1997; Stahl et al., 1996; Young & Leinhardt, 1998). Rouet and colleagues (1996) for instance, showed that only a minority of undergraduate students referred to the author or document characteristics when assessing the reliability of sources. Studies show that when encouraged to examine source information, students – even at a young age - are able to do so (e.g., Barton, 1997) and can even improve significantly when given adequate support. In a study of Britt and Aglinskas (2002) high school and college students performed poorly on sourcing in a multiple documents based task, although students' performance improved when provided with explicit sourcing instructions and support. Hence, simply presenting

students with multiple documents does not suffice to engage them in disciplinary ways of reasoning; rather students need to be taught "what it means to think like a historian" (Stahl et al., 1996, p. 448).

In this regard, several intervention studies have examined how to effectively foster students' historical thinking using multiple documents inquiry tasks (e.g., Freedman, 2015; Huijgen et al., 2018; Nokes et al., 2007; Reisman, 2012; Sendur et al., 2021; Stoel, van Drie & van Boxtel, 2017). These studies rely on instructional strategies that make disciplinary modes of reasoning explicit to students and provide support in applying them, often based on a cognitive apprenticeship approach (Nokes & De La Paz, 2018; van Boxtel et al., 2021). These studies have found positive results on various aspects of students' historical thinking. Reisman (2012), for instance, examined the effects document-based lessons in which students were supported in the acquisition of disciplinary strategies for analyzing sources through a process of modeling, guided and independent practice. Such an instructional approach improved students' use of sourcing and close reading; although not on contextualization and source corroboration. Adopting a similar instructional approach, De La Paz and colleagues (2016) found positive effects of a curriculum based on a cognitive apprenticeship approach on all aspects of students' historical writing, their overall essay quality and essay length, regardless of students' initial skills.

Studies adopting a cognitive apprenticeship approach are based on several common design principles. First, these studies have designed lesson-units based on modeling by the teacher, whereby the teacher demonstrates discipline-specific approaches to a task, followed by guided practice supported by scaffolds and feedback (e.g., De La Paz, 2005; De La Paz & Felton, 2010; De La Paz et al., 2016; Freedman, 2015; Monte-Sano, 2008; Nokes et al., 2007; Reisman, 2012).

A second, related, teaching strategy is the use of direct, explicit teaching on both second-order and procedural knowledge to make the disciplinary ways of reasoning and underlying procedural strategies visible to students (e.g., De La Paz & Felton, 2010; De La Paz et al., 2016; Nokes, 2017; Reisman, 2012; Stoel et al., 2015; van Boxtel et al., 2021). While explicit teaching is closely related to the principle of modeling, it requires the explicit articulation of discipline-specific ways of reasoning and procedural knowledge modelled by the teachers. Moreover, Stoel and colleagues (2015) argue that explicit teaching should not be limited to

the modeling phase but rather be applied throughout the instructional practice. Explicit teaching also extends beyond the teachers' thinking. It is equally important that students' reasoning becomes explicit as well (van Boxtel et al., 2021).

Third, these studies have emphasized the importance of interaction to facilitate students' historical thinking. In particular, group work is often applied to allow students to articulate their reasoning and make it explicit (Reisman, 2012; Stoel, van Drie & van Boxtel, 2017; van Boxtel & van Drie, 2017). Besides group work, whole-class discussions are considered especially relevant at the final stages of an historical inquiry as they present opportunities for teachers to provide feedback and reflect on learning outcomes (van Boxtel et al., 2021; van Drie & van de Ven, 2017).

Research has also demonstrated that not all inquiry tasks are equally effective in fostering students' historical thinking. In particular, the historical question guiding the task seems to be of importance as evaluative and analytical questions are most likely to foster students' historical thinking (Monte-Sano & De La Paz, 2012; Van Drie & van der Linden, 2006).

These intervention studies, however, have yet to shed light on the role of two influential aspects of students' historical thinking. A first concerns the development of nuanced epistemological beliefs among students which are expected to be essential in students' historical thinking and historical inquiry competence (van Boxtel & van Drie, 2017). Besides Wineburg (1991), several other scholars have proposed a relationship between students' epistemological views and their historical thinking (Havekes et al., 2012; Stoel, Logtenberg et al., 2017; VanSledright & Maggioni, 2016; Wiley et al., 2020). It is expected that when students hold naïve epistemological views, they are more likely to approach historical documents as unbiased literal reflections of the past, and hence less likely to perform a critical analysis. Nuanced epistemological beliefs, i.e., when students understand that historical knowledge is constructed and based on an interpretation of sources that are inherently biased, may correspond to better historical inquiry skills as students holding such views may be more likely to apply disciplinary strategies to compare and analyze the available documents (Wiley et al., 2020). It is unclear, however, how this relationship manifests itself. On one hand, students' epistemological beliefs are expected to influence their historical

inquiry competence (Wiley et al., 2020). On the other hand, scholars assert that engaging students in historical inquiries might foster the development of more nuanced epistemological beliefs as they make students acquainted with the processes of evaluating and constructing historical knowledge (Bråten et al., 2011; Lee & Ashby, 2000; Wiley et al., 2020). These tasks may therefore foster an awareness that multiple interpretations can arise from the available evidence and that the validity of claims can be evaluated, thus connecting to the interpretive and constructed nature of history and hence nuanced epistemological beliefs (Stradling, 2003; van Boxtel et al., 2021). However, little is known about whether and how such changes in students' epistemological beliefs can be supported and intervention studies on fostering aspects of historical thinking rarely examine the effects on students' epistemological beliefs about history (VanSledright & Maggioni, 2016); the little research that does so shows that such a change is difficult to obtain (Stoel et al., 2015; Stoel, van Drie & van Boxtel, 2017). In a review article on inquiry learning in history, van Boxtel and colleagues (2021) therefore state that more research is needed to examine the potential contribution of inquiry learning on students' epistemological beliefs.

Second, scholars argue that the promotion of historical thinking should not be approached as a purely cognitive activity as it inherently touches upon affective dimensions. History classes regularly discuss sensitive and controversial topics which may require teachers to engage with students' emotional connections to the topic (Sheppard & Levy, 2019). Multiple documents (inquiry) tasks usually require students to evaluate multiple perspectives in the sources. This, however, requires a certain detachment from one's initial perspective and a willingness to consider other perspectives (Stradling, 2003). This might decrease when students feel strongly emotionally connected to a subject, for instance when it relates to their social identities (Barton & McCully, 2007; Goldberg et al., 2011). Considering that historical inquiries require students to construct a historical representation based on a rational, critical analysis of the available evidence (i.e., the historical sources) and the various perspectives included in them, a strong emotional connection to the topic at hand may hinder their ability to do so and may hence affect their historical inquiry competence. However, research has yet to shed light on this potential role of students' emotional engagement with a topic on their historical inquiry competence. Moreover, students' motivation, interest

and beliefs about the usefulness of a certain task might also influence their engagement with the task and related learning outcomes (Wigfield & Eccles, 2000). For instance, research outside the domain of history education has shown that students apply more higher-order learning strategies when engaged in a domain that interests them and that students' interest affects their learning from (multiple) texts (List & Alexander, 2017). Scholars believe that similar processes may occur within history education, but research on the role of these affective dimensions within history education is limited (Paul et al., 2017; van Boxtel & van Drie, 2018). One study that offers evidence for such an influence is that of van der Eem and colleagues (2022) who established that the extent to which students value the disciplinary strategies of sourcing and corroboration was a significant predictor of their performance on these strategies. In a study on students' causal reasoning in history (Stoel, van Drie, & van Boxtel, 2017) researchers found that students' interest in history correlated moderately with their causal essay task. These affective dimensions, however, often remain absent in intervention studies aimed at fostering historical thinking.

Considering the potential role of epistemological beliefs and affective aspects in students' historical inquiry competence, they will also be included in this study.

Historical thinking and democratic citizenship

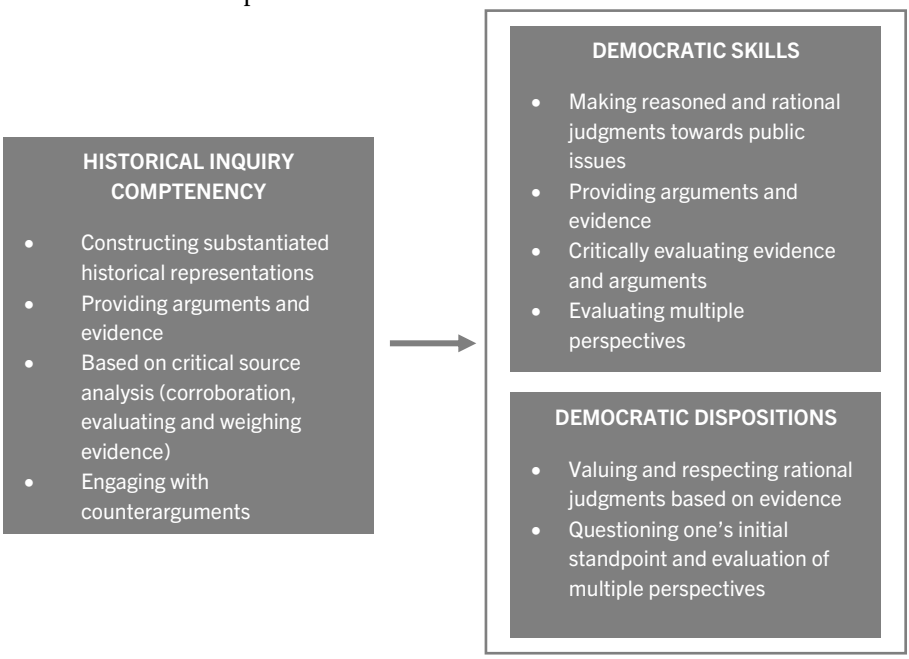
Historical thinking is presumed to contribute to democratic citizenship in various ways. Some scholars, for instance, indicate that historical thinking can provide students with essential knowledge about political, social, cultural or economic systems or that it can improve their understanding of the historical roots of contemporary issues (Barton & Levstik, 2004; Stearns, 1998; Lévesque, 2009). Others have pointed to the importance of studying agency and perspective-taking in encouraging students to reflect on their own role in society, and to value the existence of different societal perspectives (Barton, 2012; Endacott & Brooks, 2013; Seixas, 2012).

Another way in which historical thinking is presumed to contribute to democratic citizenship is through the promotion of students' historical inquiry competence, i.e. their ability to construct substantiated historical representations, based on a critical analysis of multiple sources and taking into account multiple

perspectives is considered to be beneficial for democratic citizenship (Barton & Levstik, 2004, Cinnamon et al., 2021; Lee & Shemilt, 2007; McCully, 2012; Monte-Sano, 2016; Nokes, 2013; Slater, 1995; Stearns, 1998). This competency encompasses the different aspects included in multiple document-based inquiry tasks, as described earlier.

The presumed relationship between historical inquiry competences and democratic citizenship is based on the premise that there is a significant overlap in the underlying dispositions and skills needed to engage in historical inquiry and public debates. Wineburg and Reisman (2015), for instance, argue that the acts of sourcing and contextualization are not merely disciplinary strategies, but more broadly reflect a certain way of looking at information even in contemporary contexts which is essential for citizens in a democracy. In a similar vein, Monte-Sano (2016) states that students learn citizenship skills by learning to construct and critique historical arguments as "attending to evidence that support an argument is key to understanding public issues" and that "thinking about the credibility of evidence and where evidence comes from can help citizens make stronger arguments" (pp. 316-317). Overall, historical inquiry competencies are believed to contribute to students' participation in democratic society via the development of democratic skills and dispositions (see Figure 1).

Figure 1. The presumed contribution of historical inquiry competences to democratic citizenship



In terms of skills, it is expected that historical inquiry competencies will equip students with the necessary skills and procedures to make reasoned and rational judgments on public issues, to participate in contemporary debates based on an evaluation of evidence, to make arguments that are better substantiated, to critically assess the quality of other people's arguments in public debates and to move beyond their own initial perspective (Barton & Levstik, 2004; Cinnamon et al., 2021; McCully, 2012; Monte-Sano, 2016; Nokes, 2013; Slater, 1995; Stearns, 1998).

In terms of democratic dispositions, two interrelated facets are expected to be 'modelled' when supporting historical inquiry competencies. First, a tendency to enter debates from a position of rationality, meaning students will value and respect judgements based on evidence and make their own judgments on the basis of reasonable arguments and reliable evidence, rather than on purely personal opinions or biases. Second, a tendency to question one's own initial standpoint through a consideration and evaluation of different perspectives (Barton & Levstik, 2004; Gallagher, 1996; Lee & Shemilt, 2007; McCully, 2012; Monte-Sano, 2016). These dispositions are also connected to the development of nuanced

epistemological beliefs through historical inquiry. As students become more aware of the existence of multiple interpretations and perspectives and how to assess them, they are expected to obtain more nuanced ideas about the existence of multiple perspectives in society (Barton & Levstik, 2004; Stoel, Logtenberg et al., 2017).

When describing these potential benefits of historical inquiry competence, scholars use a variety of terms such as 'habits of mind', 'moral principles' or 'attitudes' to describe a certain the tendency or inclination to think or behave in a certain way (Sosu, 2013). There is a significant overlap between these tendencies described by history educational scholars and particular democratic dispositions, namely deliberation within, perspective-taking and critical openness. Deliberation within concerns the internal reflection processes a person undertakes to weigh up reasons for and against a certain issue. These internal deliberative processes are considered an essential aspect of more public, interpersonal deliberations required in a democratic society. This internal deliberation encompasses processes such as taking different perspectives and weighing arguments for and against a certain issue and is, within the field of political psychology, expected to precede the deliberation among individuals (Goodin, 2003; Weinmann, 2018). Critical openness refers to the "tendency to be actively open to new ideas and critical in evaluating these ideas and modifying ones thinking in light of convincing evidence" (Sosu, 2013, p. 115). It is a subcomponent of critical thinking which closely relates to the construct of open-mindedness and is considered an essential disposition for critical thinking. Perspective-taking represents the cognitive component of empathy, referring to someone's tendency to imagine or adopt another's perspective (De Corte et al., 2007). Together, these notions closely relate to the dispositions described by scholars in the field of history education.

In short, supporting students' historical inquiry competence is expected to enable students to take up a reasoned position towards public issues based on evidence (democratic skills), and to foster an inclination to do so (democratic dispositions). Scholars argue that, as there is a significant overlap in the underlying dispositions and skills needed to engage in historical inquiry and public debates, students will transfer the skills and dispositions acquired in the historical context to a contemporary context. In describing this relationship, scholars pay little attention to how exactly such a transfer would come about or how history

education could support it. They rather seem to presume that such a transfer occurs spontaneously. While there is some evidence to support such a (spontaneous) transfer within the field of history education, there are also reasons to question whether such a transfer is likely.

Research on the relationship between historical inquiry competence and democratic citizenship

Evidence that fostering students' historical thinking may transfer to democratic skills and dispositions mainly stems from research conducted in post-conflict and divided societies, in which the benefits of an inquiry based and/or multi-perspective approach were tested to challenge dominant, single narratives and facilitate peace building (e.g., Goldberg et al., 2011). A multi-perspective approach was examined, for instance, among Jew and Arab/Palestinian Israeli students writing a joint paper on their common past in response to a research question. A thematic analysis of the essays indicated an increased openness to the perspective of the other (Ben-David Kolikant & Pollack, 2019). Several studies have also been conducted in Northern Ireland, where an inquiry and multi-perspective approach to history education was introduced into the curriculum in the 1990s to overcome communal divisions (McCully, 2012). These studies similarly found that, while students still tended to understand the past mainly from their own perspective and from the perspective of their own social group, there was also a willingness to understand the perspectives of different social groups (Barton & McCully; 2005, 2010; Bell et al., 2010). In line with these findings, Kim's (2019; 2020) study with South Korean and Japanese students found that a textbook offering multiple discourses on their mutual past prompted students' willingness to consider multiple perspectives. While these studies demonstrate that students' views on recent and sensitive history can become more open after inquiry tasks including multiple perspectives, they did not examine whether this increased openness extended to different contemporary topics. Moreover, changes in students' dispositions were mostly examined in an indirect way, for instance through interviews.

More direct evidence for the possibility of transfer apart from the particular context of post-conflict or divided societies is provided in Reisman's (2012) study on the effects of an intervention on primary source instruction. Reisman (2012)

found that students who had received explicit instruction on reading strategies were more likely to apply some of these strategies to contemporary sources, indicating a transfer of sourcing strategies from historical to contemporary topics. However, while students were more likely to apply sourcing and close reading to contemporary sources, no effect was found for contextualization and corroboration. According to Reisman (2012) the difference in transfer between these strategies might be due to the way they were modelled by the teachers. While sourcing and close reading were modelled consistently via concrete actions related to single documents, this was more difficult for contextualization and corroboration as they are intertextual strategies. Reisman therefore proposes the use of whole-class discussions to support students in internalizing these intertextual strategies. While this study provides some evidence for the possibility of a (spontaneous) transfer of disciplinary strategies to contemporary contexts, Reisman's (2012) study was limited to the use of disciplinary reading strategies and did not examine changes in students' dispositions.

These studies hence provide some indications that a transfer from historical to contemporary subject may spontaneously occur, as is assumed by various history educational scholars. By contrast, however, studies outside the field of history education have shown that transfer is not easily obtained (e.g., Tiruneh et al., 2018; van Peppen et al., 2021). The presumed contributions of historical inquiry for democratic citizenship require students to transfer the dispositions and skills needed for historical inquiry to a contemporary context. This calls for the application of skills and dispositions beyond the initial (i.e., historical) context and thus constitutes a 'far transfer' (Perkins & Salomon, 1994). Such a transfer is not easily obtained and requires instructional conditions aimed at generalization (Perkins & Salomon, 2012; Tiruneh et al., 2018; van Peppen et al., 2021). In order for a transfer to occur, students need to recognize similarities between the original context and a novel context, recall the relevant skills and apply them to the novel context (Barnett & Ceci, 2002; van Peppen et al., 2021). This is not to say that a transfer of learning is not possible, but rather that it requires particular instructional conditions. In this regard, scholars emphasize the importance of acquiring knowledge and skills within a discipline-specific context, rather than merely on a general level (Glaser, 1984; Perkins & Salomon, 1989). Instructional conditions that facilitate the mastery of discipline-specific thinking and that may

promote a transfer to different contexts include those adopted in the model of cognitive apprenticeship (Brown et al., 1989) such as the use of authentic learning tasks and explicitly demonstrating solutions and procedures for solving this task (Bransford & Schwartz, 1999; Merill, 2002; van Merriënboer & Kirschner, 2007). These practices may support students to recognize similarities between tasks in different contexts which constitutes the first step in a transfer of learning.

The presumed contribution of historical inquiry to democratic citizenship hence raises an important issue. On one hand, scholars have demonstrated the similarities between dispositions and skills required for historical inquiry and democratic citizenship, and there is some empirical evidence to support transfer from one to the other. On the other hand, such a transfer would be difficult to obtain without explicit attention to these particular similarities. Considering the importance of the democratic skills and dispositions described above, and considering the long-standing assumption that history education will automatically contribute to them, it seems all the more important to gain more insight into whether such a transfer can indeed occur spontaneously among students.

The current study

The central aim of this study is to examine the disciplinary effects of a lesson series, designed specifically to foster students' historical inquiry competences and to examine whether fostering students' historical inquiry competences will spontaneously contribute to democratic skills and dispositions. The study will therefore examine the effects of the lesson series on students' historical inquiry competences, epistemological beliefs about history, and democratic citizenship.

To this end, a cluster randomized controlled intervention study was set up with students in Flemish history education. Students in the intervention group participated in a lesson series aimed at promoting their historical inquiry competence, while students in the control group received conventional classes. Changes related to historical inquiry and democratic citizenship were assessed via a pre- and posttest and compared to the control group. The following research questions were evaluated:

1. What is the effect of a lesson series focused on historical inquiry, on students' (a) historical inquiry skills, (b) procedural knowledge on historical inquiry and (c) epistemological beliefs about history?
2. What is the effect of the lesson series on students' (a) democratic skills and (b) democratic dispositions, in particular: deliberation within, critical openness and perspective taking?

METHOD

Research context

The research took place in the 12th grade (students' age: 17-18 years old) of Flemish (Belgian) history education. In Flanders, history is a mandatory school subject which takes up 2h/week in general education.

The Flemish history standards provide teachers with a lot of freedom regarding both the content of their lessons and didactic methods. They only prescribe which historical period should be addressed at which stage, but do not provide specific topics to be taught. Historical thinking is only very implicitly set as a goal in the history standards; no concrete operationalization of the concept is provided, nor support on how to foster this among students. For instance, while the standards emphasize the importance of critical source analysis, they do not provide guidelines on how to do so (Van Nieuwenhuyse, 2020; Van Nieuwenhuyse et al., 2017). History teachers' freedom is further guaranteed by the absence of centralized state exams and government control on textbooks (Van Nieuwenhuyse, 2020).

History is an autonomous school subject, primarily focused on studying the past and is aimed at fostering students' understanding of the past. Most Flemish history teachers are indeed very past-oriented in their practice (Van Nieuwenhuyse et al., 2015; Wils et al., 2011) and tend to prioritize providing students with a general overview of the past (Voet & De Wever, 2016).

Teaching practices regarding historical thinking differ greatly among teachers. Overall, research on Flemish history teachers' practices has shown that teachers' main focus is on knowing history and that aspects of doing history are only addressed to a limited extent. For instance, while teachers make extensive use of sources, they primarily use them in an illustrative way or to foster students'

substantive knowledge (Van Nieuwenhuyse et al., 2017). Historical inquiry, as part of doing history, is rarely applied. Teachers generally have a limited understanding of what it entails and are hesitant to include it in their practice, although differences among teachers exist (Voet & De Wever, 2016, 2017, 2019). Overall, it can be expected that students' historical inquiry skills and knowledge are limited but could vary strongly across schools.

Participants

In June 2019, a call was launched via various channels among Flemish history teachers, asking them to participate in research on effective strategies for fostering students' inquiry skills in history. We asked for teachers with at least three years of teaching experience and who were active in the 12th grade of general education. Forty teachers in 32 schools entered the study. Schools were assigned randomly to an experimental and control condition. Initially, student numbers were equal in both conditions, however, due to covid-measures in Flemish education, a number of teachers dropped out of the study. This dropout was higher in the control group than in the intervention group. We also applied strict exclusion criteria (see further) throughout the research to ensure treatment fidelity. Eventually, 668 student (266 control group, 402 intervention group) in 47 classes (21 control, 26 intervention) and 17 schools (6 control, 11 intervention) completed the research (see Table 1 for background characteristics).

All schools had similar student demographics both in terms of socio-economic and ethnic backgrounds. In the control group, schools had an average of 28.6% of students with a low socio-economic status[2]. In the intervention group, this was on average 31.5%. Only a small minority of students was born outside of Belgium (2.7% in the control group; 3.7% in the intervention group) or had at least one parent who was born outside of Belgium (10.62% in the control group and 11.44% in the intervention group). Of these students, half had a migration

[2] In Flanders, students are attributed to the category of 'low socio-economic status' when they meet one or both of the following characteristics: being member of a family who receives government aid for education and/or the mother testifying to a low educational level (secondary education not completed).

background from within the EU27 countries and the other half from outside, both in the control and intervention group.

Table 1. Gender and Age Distribution among Intervention and Control Group

	Gender	*n*	Age Mean (SD)	*n*
Intervention group	153 male (38.2%) 247 female (61.6%) 1 other (0.2%)	401	17.04 (0.39)	402
Control group	95 male (35.7%) 170 female (63.9%) 1 other (0.4%)	266	17.06 (0.38)	266

The control and experimental condition

For the experimental condition a lesson series was designed, consisting of eight lessons, taking up twelve to fourteen 50-minute classes. The lesson materials consisted of a detailed lesson script for teachers and a syllabus and supporting materials for students.

The lesson series focused on promoting students' historical inquiry competence. To this end, the lesson series contained several inquiry-based learning activities focused primarily on the development of inquiry skills and procedural knowledge on inquiry (doing history). This was supplemented with lessons which supported the acquisition of substantive knowledge (knowing history) as both are essential in historical reasoning (van Boxtel & van Drie, 2017). These lessons adopted a more conventional approach, in line with 'typical' Flemish history classes based on Socratic dialogue. Both lesson types had a complementary role: the inquiry tasks built further on the substantive knowledge acquired by students and the lessons which focused rather on substantive knowledge contained short exercises on aspects of historical inquiry, for instance on evaluating evidence through an assessment of a source's reliability. 'Decolonization after 1945' functioned as a common theme throughout the lesson series, as this theme easily allowed for the inclusion of multiple perspectives and is commonly studied in the last stage of Flemish history education. An expert panel was composed to validate the lesson series. On multiple occasions during the design process, feedback was gathered from (1) a subject expert on decolonization (2) experienced history

teachers and history teacher trainers and (3) experts in history didactics and educational research. An overview of the lesson unit can be found in Table 2.

In line with the model of cognitive apprenticeship and research on fostering aspects of historical thinking among students, the lesson unit adopted the principles of modelling, coaching and scaffolding to get students acquainted with the procedures of historical inquiry (Collins et al., 1991; Freedman, 2015). This was organized along four major inquiry-based learning tasks that were embedded throughout the lesson series. In what follows, the design of these inquiry tasks will be described, followed by a discussion of students' and teachers' activities during the lessons.

Each inquiry task was designed along the same structure. The task was centered around an evaluative, open-ended historical question with no fixed answer, as these are particularly suited to generate historical reasoning (Monte-Sano, 2016; van Drie et al., 2006). Students were then provided with a number of sources that approached this question from different perspectives and which varied in terms of reliability and/or representativeness. Students had to corroborate sources and critically evaluate their value in order to provide an answer to the historical question. They then had to substantiate their claim through argumentation and evidence, taking into account counter arguments. This took the form of a writing task (with a specific goal and audience as recommended by van Drie et al., 2013) or an oral task.

Students usually worked in groups as interaction can facilitate historical reasoning and allows for articulation and reflection (Collins et al., 1991; van Boxtel & van Drie, 2017). This was supplemented with whole-class discussions in which students discussed their findings and teachers provided feedback and questions to further foster students' reasoning (van Boxtel et al., 2021; van Boxtel & van Drie, 2013).

In the first inquiry task, the teacher modelled the different steps and reasoning processes involved in historical inquiry. In the subsequent inquiry tasks, coaching and scaffolding were the main focus. To this end, the lesson sequence provided scaffolds for each step of the process. These included, among others, guiding questions to support students in evaluating and corroborating sources, visual organizers for weighing evidence and writing schemes and prompts to formulate a substantiated answer (modelled after those of the University of

Michigan (2021) and Monte-Sano et al., 2014). The scaffolds contained prompts regarding sourcing and source analysis to support students' attention to multiple perspectives (Monte-Sano & De La Paz, 2012). The lesson unit provided various degrees of scaffolds and suggested a process to decrease and gradually fade out the level of scaffolding throughout the lesson series; teachers were encouraged to adjust the scaffolds to their students' needs. They coached students during the inquiry activities, in part by providing feedback and selecting the most appropriate level of scaffolding.

Another essential element of the lesson series was the use of explicit teaching. Teachers were instructed to provide explicit teaching on both second-order and procedural knowledge related to the processes of historical inquiry (e.g., reliability and how to assess it, representativeness, source corroboration). On several occasions throughout the lesson series, this was also explicitly connected to the interpretive and constructed nature of historical knowledge, although the main focus was on explicit teaching related to historical inquiry. Explicit teaching on these concepts and procedures was not limited to the modelling phase but took place repeatedly throughout the lesson series. This was further supported in the student materials which contained separate documents and graphical organizers on these concepts and procedures.

Considering that one purpose of this study was to test the assumption that historical inquiry competences would spontaneously contribute to the development of democratic skills and dispositions, explicit teaching in the experimental condition was limited to historical inquiry and epistemological beliefs about history and did not extent to explicit teaching regarding democratic skills and dispositions. Nevertheless, the lesson series included several instructional conditions required for far transfer, as the use of principles from cognitive apprenticeship is expected to allow for generality (Brown et al., 1989). For instance, the use of authentic and relevant learning tasks is expected to facilitate transfer (van Merriënboer & Kirschner, 2007). This is included in the lesson series through inquiry tasks that are based on authentic historical questions and multiple documents. Coaching, scaffolding and the demonstration of solutions and procedures on how to solve a particular task are also considered instructional conditions which are beneficial for transfer (Bransford & Schwartz, 1999; Merill, 2002; van Merriënboer & Kirschner, 2007). The lesson design also

included whole-class discussion, which Reisman (2012) hypothesized might be beneficial to facilitate a transfer of more complex sourcing strategies in history.

In the control group, teachers proceeded with their own scheduled classes between the pre- and posttest. In terms of the teaching content, no restrictions were put in place considering that the study and the experimental condition were aimed at historical inquiry competences, rather than on the acquisition of substantive knowledge about history. Teachers in the control group hence discussed a variety of topics related to 20th century history. Due to the fact that Flemish history teachers usually work chronologically and address all the post-1945 period in the 12th grade, the topics addressed in the control group strongly overlapped with those of the experimental condition. They included decolonization processes, postcolonial relationships (in particular between Belgium and Congo), the Israeli-Palestinian conflict or more general developments in the Middle East. Topics that differed from the lesson series were the founding of the European Union and the Cold War. In terms of pedagogical approaches, no restrictions were provided either, taking into account that most Flemish history teachers overall do not include many inquiry-based activities. However, to account for variance in teaching practices regarding historical inquiry competences, teachers were asked to provide the didactic materials they used between the pre- and posttest, allowing us to monitor their practice and exclude them from the study if necessary (see further).

Table 2. Overview of the lesson-unit

Lesson topic and duration	Lesson type	Teacher and student activities
Lesson 1 (2x50'): Causes and process of decolonization	Inquiry lesson - Historical inquiry task: "Who was most responsible for the decolonization of British-India?"	Teachers model the process of historical inquiry. Teachers provide explicit teaching on multiperspectivity and on evaluating sources' usability, reliability, source corroboration and on how to substantiate a historical claim (argumentation and evidence).
Lesson 2 (2x50'): Decolonization of Belgian-Congo	Inquiry lesson - Historical inquiry task: "Who was most responsible for the decolonization of Belgian-Congo?"	Students work in groups on the historical inquiry task, supported by scaffolds. Teachers coach students and provide explicit teaching on the representativeness of sources. Teachers lead class-discussion on the inquiry task.
Lesson 3 (1x50'): Belgian collective memory of the colonial past	Lesson with focus on substantive knowledge - Changing perspectives on the colonial past (1945-current) and multiple perspectives on the colonial past.	Students evaluate sources' usability, reliability and representativeness. Teachers provide explicit teaching on different layers of multiperspectivity (in the past, in the present, over time).
Lesson 4 (2 or 3 x 50'): The Israeli-Palestinian conflict	Lesson with focus on substantive knowledge - Study of competing narratives on causes of the conflict	Students study and compare multiple perspectives, they evaluate sources' reliability. Students practice formulating a substantiated historical claim, based on evidence. Teachers lead whole class-discussion related to students' activities on evaluating sources and formulating a substantiated historical claim.

Table 2. (*continued*)

Lesson topic and duration	Lesson type	Teacher and student activities
Lesson 5 (2 or 3 x 50'): Case-study of the Rwandan genocide or the India partition	Inquiry lesson - Parallel historical inquiry task: "To what extent could the negative consequences (being the outbreak of genocide) of the withdrawal of UN troops from Rwanda in 1994 have been foreseen by the actors involved?" "To what extent could the British and Indian leaders involved foresee the negative consequences of the rapid British withdrawal and partition of British India in 1947?"	Students work in groups on one of the historical inquiry task, supported by scaffolds. Teachers coach students during the inquiry task and provide feedback on the final products.
Lesson 6 (1x50'): Postcolonial relations between the West and former colonies	Lesson with focus on substantive knowledge - Analysis of aspects of (dis)continuity regarding the relation between the West and former colonies	Students evaluate the usability and reliability of sources and substantiate a historical claim.
Lesson 7 (1x50'): Migration after 1945	Lesson with focus on substantive knowledge - Overview of migratory patterns after 1945	Students evaluate the reliability of sources.
Lesson 8 (1x50'): Migration and integration	Inquiry lesson - Historical inquiry task: "What was the main cause of integration problems experienced by immigrants in Belgium in the 1970s-1990s?"	Students work individually and in groups to evaluate two competing historical representations, based on sources. Teachers coach students and lead class-discussion on the inquiry task.

Procedure, treatment fidelity and exclusion criteria

The study took place between January-April 2021. All students signed an informed consent form and completed a pre- and posttest each spanning two consecutive classes (2 x 50 min.). Students created a unique, anonymized identifier to connect the pre- and posttest. There was a 6–7-week interval (i.e., 12-14 history classes) between the administration of the pre- and posttest.

To ensure treatment fidelity, teachers in the intervention group participated in a training before the start of the study with the aim of making them feel supported and prepared for the implementation of the lesson materials. Due to the corona restrictions, the training had to take place online. Hence, the training session was organized in small groups allowing for interaction among the participants and between the participants and the trainers. It included explicit teaching on the notion of historical thinking to accommodate for possible differences between teachers' prior knowledge on the concept (see: research context). Furthermore, the design principles of the lesson series were explained, and the lesson series was discussed with an explicit connection to the underlying design principles and the way in which they were put into practice in the materials. Concrete examples of students' essay writing were also discussed with the teachers. Teachers were also encouraged to contact the researchers in case of further questions during the implementation of the study so that continuing support could be offered.

Due to the Covid-restrictions only half of the lessons could take place in class. The other half was provided through remote education. To ensure a maximum uniformity in the remote classes, a script was provided for digitized and/or remote classes, based on digital learning tools and instruction videos.

Teachers were also monitored during the intervention. This was planned to be done via lesson observations; however, the corona measures did not allow for the researchers to be present in the classrooms. As an alternative, teachers were asked to complete an online form after each lesson in which they reported on their experiences. They registered, among others, in what form the lesson took place (physical or (a)synchronous remote education), to what extent they were able to reach all of their students when the class was given in remote education and whether they had deviated from the provided script (and if so: why, how and to what extent). These online forms were complemented with two one-on-one online

interviews with one of the researchers during the intervention, to discuss teachers' experiences, potential difficulties and/or deviations from the lesson plan. During those interviews, teachers reported openly about changes they made to the original lesson plans. Teachers' modifications to the lesson materials were overall minor; they included adjusting the scaffolds to their students' needs, providing additional background information in lessons with a focus on substantive knowledge and spending more time than suggested on the inquiry tasks. The interviews were also used as an opportunity to refer back to the training session and verify whether teachers needed additional support in the implementation of the materials.

Several teachers experienced delays due to the difficult combination of in-class and remote education. Because of this, eight teachers reported skipping at least one of the inquiry lessons. Considering that these assignments were essential in the intervention, those teachers' students were excluded from the analysis as their practice differed too much from the designed lesson series. Three more teachers had to delay the posttest by three weeks, due to last-minute changes in the corona measures in Flemish education (early start and extension of the Easter break). Their students were also excluded from the research as this meant there was a gap of four weeks between the last lesson and the posttest, constituting a large deviation from the research design. Students who had missed more than two classes were also excluded ($n = 12$).

In the control group, teachers' instructional materials were screened for the presence of inquiry activities similar to those in the intervention group and for explicit attention to the constructed nature of historical knowledge. Two teachers had included inquiry tasks in line with those of the intervention. Because their instructional practice mirrored those of the experimental condition so closely, these teachers were excluded from the study.

Instrument development and validation

The pretest and posttest assessed historical inquiry competences and epistemological beliefs about history (RQ1) and democratic citizenship (RQ2) and included two affective variables as control variables (see further). A pilot study was conducted in two classes ($n = 13$ and $n = 23$) in secondary history education to test the pre- and posttest which consisted of a questionnaire and two essay tasks. An overview is provided in Table 3.

Table 3. Overview of the pretest and posttest

Part one	Part two

• Informed consent
• Background characteristics
• Historical inquiry skills (historical essay task)
• Sensitivity of historical essay topic (scale)
• Task value (scale)
• Epistemological beliefs (scale)
• Background characteristics
	• Procedural knowledge on historical inquiry (open-ended questions)
	• Democratic skills (topical essay task)
	• Sensitivity of topical essay topic (scale)
	• Democratic dispositions (scales)
	  ○ Deliberation within
	  ○ Critical openness
	  ○ Perspective taking

Historical inquiry competence and related epistemological beliefs (RQ 1)

Historical inquiry skills were assessed via a historical essay task (see appendix B). The historical questions central in the essay assignment were (1) whether police violence at the Sharpeville massacre (1960) was permissible or not[3] and (2) whether Dutch violence during the Indonesian Independence War (1945-1949) was structural or incidental. The essay tasks provided students with a short background sketch, the historical question and a set of sources. Sources were accompanied with basic background information (i.e., author, date, etc.) and offered diverging answers and perspectives on the question at hand. Students were asked to write a short essay of 15-20 sentences in which they took a reasoned stance in response to the historical question. This stance was to be supported by arguments and evidence from the provided sources and based on a critical analysis and corroboration of these sources. The pilot study provided an indication of the feasibility of the essay tasks in terms of essay length, time and difficulty.

To evaluate the essay task, an initial rubric was designed based on research by Monte-Sano (2012); Monte-Sano & De La Paz (2012), and Monte-Sano, De La Paz and Felton (2014) and adapted to better fit the lesson design. Minor adjustments were made after the pilot study. The final rubric (see appendix C)

[3] Due to the phrasing of this particular question, this may appear to be a moral, rather than an evaluative question. However, students had to evaluate whether the police had a legitimate reason to use violence, not whether the use of such violence is morally permissible. The provided sources were clearly aligned with this view (see appendix B).

consisted of six criteria: (1) making a (historical or topical) claim in response to a (historical or topical) question; (2) argumentation in support of the claim (3), quality of the argumentation (4) use of sources as supporting evidence (5) evaluation of the value of the available sources (i.e., reliability and/or representativeness) and (6) discussion of counter arguments. Each criterion in the evaluation rubric was scored on a four-point scale, resulting in a total essay score on 24. Inter-rater reliability was assessed using intraclass correlation coefficients (2-way, absolute agreement, single measures) with 10% of pretest essays being coded by at least two scorers. For each criterion at least a good agreement (> .6) was obtained for each separate criterium (M = 0.82, SD = 0.12) and an excellent agreement (>.75) was obtained for the total score (M = 0.87, SD = 0.08) (Cicchetti, 1994; Hallgren, 2012).

Procedural knowledge on historical inquiry was gauged via two open-ended questions: (1) list aspects of a historical source which could determine its reliability (2) describe which steps you would undertake when asked to provide a substantiated answer to an historical question based on a number of sources. Based on the literature on critical source analysis and substantiating a historical claim (De La Paz & Felton, 2010; Hicks, Doolittle & Ewing, 2004; Monte-Sano & De La Paz, 2012; Nokes, 2013; Reisman, 2012; van Drie & van Boxtel, 2008; Voet & De Wever, 2017b; Wilschut et al., 2012) and in line with explicit teaching on procedural knowledge in the intervention, seven criteria were distinguished for reliability (author, location, time, societal context, purpose, audience, information/expertise) and eight for answering a historical question (evaluating usability of sources, reliability of sources, representativeness of sources, source corroboration, making a historical claim, argumentation, using sources as evidence and engaging with counterarguments and evidence). Students received one point for each correct element in their answer. Ten percent of pre- and posttest answers were evaluated by two scorers. A Cohen's Kappa was calculated resulting in an excellent agreement of κ = .80 for the first question and κ = .86 for the second.

Epistemological beliefs were measured through the scale of Wiley and colleagues (2020). Students evaluated 12 statements on a 6-point Likert scale. This scale specifically assesses students' epistemic beliefs related to multiple-documents inquiry tasks. A first subscale measured students' beliefs about 'the simplicity and certainty of knowledge'. A low score on this subscale indicates more advanced

beliefs. The second measured 'beliefs about the value of engaging in integrating of information across documents' in which case a high score indicated advanced beliefs. The psychometric quality of the scales was examined via their internal consistency. Cronbach's α was .69 for subscale 1 and .75 for subscale 2.

Democratic citizenship (RQ 2)

Democratic skills related to making rational and reasoned judgments on public issues was evaluated via a topical essay task. In order to examine the possibility of a transfer from historical inquiry skills to democratic skills, the topical essay task was designed in parallel to the historical essay task (i.e., background information, societal question/debate, conflicting sources and perspectives). Two topics were selected which were actively debated in Flemish society at the time of the research: (1) Should the voting age be lowered to the age of 16 and (2) Should positioning tests be introduced in higher education. In line with the instructions for the historical essay task, students were asked to write a 15-20 sentence essay in which they took a substantiated stance regarding a current societal debate, supported by arguments and evidence and based on a critical analysis and corroboration of the available sources.

The same evaluation rubric was used to assess historical and topical essay tasks. Two scorers rated 10% of pretest essays. Intraclass correlation for the separate criteria obtained at least a good agreement (M = 0.75, SD = 0.08) and excellent agreement was achieved for the total score (ICC = .87) (Cicchetti, 1994; Hallgren, 2012).

Democratic dispositions were assessed via three validated scales for the notions of deliberation within, critical openness and perspective taking. Deliberation within was assessed using the scale of Weinmann (2018).This scale was directly tied to students' topical essay task, asking them to reflect on the process of determining their own stance regarding the topic. It included items such as "I have thought about arguments for and against my own as well as others' opinions about the topic" and "I have simulated several opinions about the topic". Students evaluated 5 statements on a 5-point Likert scale. Critical openness was measured using the critical openness subscale of the Critical Thinking Disposition Scale (Sosu, 2013). It contained items such as: "It is important to justify the choices I make" and "I use more than one source to find out information for myself". The

scale consisted of 7 items on a 5-point Likert scale. The perspective taking subscale, stemming from the Interpersonal Reactivity Index measuring empathetic tendencies was also included (De Corte et al., 2007). As these items were initially aimed at one-on-one interpersonal interaction, some were slightly adapted so as to represent interactions on a more general, societal level. For instance, in the statement "I sometimes try to understand my friends better by imagining how things look from their perspective", "my friends" was replaced by "others". Students assessed 7 items on a 5-point Likert scale. Cronbach's α for the perspective-taking scale was .73. The critical openness ($\alpha = .63$) and deliberation within scale ($\alpha = .62$) yielded moderate scores.

Affective dimensions

To account for the possible influence of affective variables in both of the research questions, two affective measures were included as control variables. To examine the possible influence of the sensitivity of the essay topics, students evaluated after each essay the following statement on a 5-point Likert scale: "To what extent was the topic of the essay task you just wrote personally sensitive to you?" In addition, the subjective task value subscale of Wigfield and Eccles (2000) was included in the instrument. This scale consisted of 6 statements on a 7-point Likert scale, assessing students' interest and evaluation of the usefulness and importance of history education. Cronbach's α for this scale was .90.

Data analysis

After scoring the open-ended questions and essay tasks, the data were prepared for the statistical analysis. To this end the scores of the two procedural questions were added to generate one total score out of 15 for procedural knowledge in pre- and posttest. As a measure of students' changes in historical inquiry skills and procedural knowledge, two new variables were computed. 'Change in historical inquiry skills' was based on students' historical writing tasks and calculated by subtracting students' pretest scores on the historical essay task from their posttest score. 'Change in procedural knowledge of inquiry' was calculated by subtracting students' pretest score on the open-ended questions on procedural knowledge from the posttest score. Reverse coded items in the perspective taking scale were recoded.

Missing values in the data were limited (less than 1% for each variable). Values were most likely to be missing on the variables 'epistemological beliefs-pretest' (0.7%) and 'task value' (0.4%), which is presumably due to their position at the end of questionnaire. Listwise deletion was applied in case of missing values.

The research questions were answered via multilevel modelling (using IBM SPSS 28), as this allows one to take into account the hierarchical and nested structure of the data. First, a correlation matrix was examined to determine the correlation between the various variables at pretest. For the multilevel analysis, three levels were distinguished: pupils, classes and schools. A bottom-up approach to model building was adopted (Hox, 2010). The analysis started with an intercept-only, three level model which constituted the null model (model 0). From there on, variables were added as fixed effects in ensuing models (first predictors, then covariates and interaction effects). Maximum Likelihood was used to estimate parameters. Assumptions regarding multicollinearity, homogeneity of variances and independence and normal distribution of errors were checked and no deviations were found. In cases where no variance was found on the school and class level, these levels were excluded from the analysis and a hierarchical linear regression was conducted, in which case predictors and covariates were entered in the same sequence as described for the multilevel models.

Separate analyses were conducted for each outcome variable related to either historical inquiry or democratic citizenship. For both research questions, the condition functioned as the predictor, while pretest scores and affective variables functioned as covariates. An interaction effect of condition*pretest score was also examined. For both research questions, the effect of the condition was examined first, and covariates were entered in a subsequent model to check the robustness of this effect. For the second research question, an additional model was tested in which two additional predictors were entered to more closely examine the effect of the intervention: students' changes in historical inquiry skills and procedural knowledge of inquiry. An interaction effect of these variables and the condition was also examined. Following Lorah (2018) and Snijders and Bosker (2012) the proportion of variance explained (R^2) between the full models and the empty, intercept-only model (model 0) was used to calculate Cohen's f^2 as a measure of effect size for the overall models. To examine effect sizes for particular

fixed effects, Cohen's f^2 was calculated based on R^2 of the full model with and without the particular effect.

Model fit of multilevel models was compared using BIC, AIC and -2 log-likelihood values and total explained variance of the best fit model was compared to the null model following Hox (2010). For the linear regression models, the model with the best fit was determined by comparing the total amount of variance explained, based on $R^2_{adjusted}$ and R^2_{change}. Due to the large number of analyses, only the models with the best fit will be discussed in detail.

RESULTS

First, descriptive results will be presented and discussed via an overview of students' scores in pre- and posttest and the discussion of a correlation matrix. Next, the results of the multilevel analyses will be discussed in order to answer the research questions.

Descriptives

An overview of students' scores on pre- and posttest measures is provided in Table 4. One-way ANOVA's were used to examine differences between control and intervention group at pretest level. Significant differences were found only for democratic skills ($F(1, 665) = 4.473$, $p = .03$) and procedural knowledge ($F(1, 653) = 21.132$, $p < .001$). To account for differences in pretest scores, they are included as covariates.

Table 4. Summary of students' scores by condition

Measure	Condition	Pretest *M* (*SD*)	*n*	Posttest *M* (*SD*)	*n*
Historical inquiry skills (essay task)	Intervention	15.20 (3.58)	401	18.14 (4.45)	402
	Control	14.77 (3.62)	265	13.57 (5.17)	266
Sensitivity of historical essay task subject	Intervention	2.95 (1.05)	402	2.11 (0.98)	402
	Control	2.95 (1.04)	266	2.13 (1.01)	266
Procedural knowledge on inquiry	Intervention	4.40 (1.77)*	402	6.42 (3.13)	402
	Control	3.85 (1.51)*	264	3.94 (1.60)	266
Epistemological beliefs (subscale 1)	Intervention	2.31 (0.63)	401	2.10 (0.67)	402
	Control	2.39 (0.60)	262	2.17 (0.61)	266
Epistemological beliefs (subscale 2)	Intervention	4.78 (0.56)	401	4.89 (0.68)	402
	Control	4.77 (0.51)	262	4.78 (0.59)	266
Democratic skills (topical essay task)	Intervention	14.75 (3.44)**	402	16.48 (3.91)	402
	Control	14.15 (3.63)**	265	14.61 (3.36)	266
Sensitivity of topical essay task subject	Intervention	2.56 (1.13)	402	2.93 (1.13)	401
	Control	2.43 (1.14)	266	2.95 (1.16)	266
Democratic dispositions – deliberation within	Intervention	3.72 (0.52)	402	3.83 (0.53)	401
	Control	3.69 (0.51)	266	3.74 (0.56)	265
Democratic dispositions – critical openness	Intervention	3.95 (0.38)	402	3.91 (0.43)	401
	Control	3.94 (0.40)	266	3.92 (0.45)	265
Democratic dispositions – perspective taking	Intervention	3.58 (0.50)	402	3.60 (0.49)	401
	Control	3.53 (0.58)	266	3.55 (0.53)	265
Task value	Intervention	4.60 (1.11)	401		
	Control	4.52 (1.14)	264		

Note: $^{*}p < .001$, $^{**}p < .05$

The correlation matrix (see appendix A) was examined to explore the relationship between the various variables and to rule out multicollinearity. The correlation matrix shows that both epistemological subscales were negatively correlated at pretest and at posttest. Contrary to our expectations based on the literature review, procedural knowledge about inquiry was not significantly correlated with historical inquiry skills at pretest, nor were epistemological beliefs. This indicates that students' procedural knowledge about historical inquiries is not related to

their ability to apply this knowledge. Moreover, it questions the assumption that students' epistemological beliefs are connected to their historical inquiry skills and procedural knowledge. At the posttest, a significant positive correlation did exist between students' historical inquiry skills and procedural knowledge which might point to a positive result of the intervention. With regard to epistemological beliefs, there was still no, or only a very weak correlation at posttest. Again, this indicates that students' historical inquiry skills and procedural knowledge may not be connected to their epistemological beliefs to the extent that researchers expect it to be.

Concerning democratic citizenship, it appears that democratic skills were not or only weakly correlated with democratic dispositions while the three democratic dispositions scales were significantly positively correlated. Interestingly, weak significant correlations existed between epistemological beliefs and democratic dispositions. This is especially the case at posttest and with regard to subscale 2. This epistemological subscale refers to students' ideas about comparing and integrating ideas from multiple documents when constructing (historical) explanations. This indeed connects closely to the democratic dispositions that were measured, assessing how students value the consideration and evaluation of different perspectives towards current issues.

Regarding the affective dimensions, the expected correlation between the affective variables and historical inquiry skills and procedural knowledge did not occur. Task value did not correlate with historical inquiry skills and only weakly with procedural knowledge about inquiry, suggesting it did not affect students' performance on historical inquiry. There was also a negligible correlation between the sensitivity of essay topics and students' performance on related historical skills and democratic skills.

Historical inquiry skills and democratic skills were significantly positively correlated at pretest; this correlation increased in the posttest. There was also a significant positive correlation between procedural knowledge and democratic skills at posttest. Both offer some support for the relationship between historical and democratic skills. Pre- and posttest measures of the same variables were also positively correlated, confirming it is important to take pretest measures into account as covariates.

RQ1: Effects of the intervention on historical inquiry competence and epistemological beliefs

The first research question addressed the effects of the lesson series on students' historical inquiry competence and their underlying epistemological beliefs about history. A multilevel model was built for each outcome variable: historical inquiry skills, procedural knowledge on historical inquiry and epistemological beliefs (two subscales).

Effects of the intervention on students' historical inquiry skills

The results in Table 5 show that a model with both condition and pretest scores on historical inquiry, procedural knowledge, sensitivity of the essay topic, task value and condition*pretest score on historical inquiry skills was the best fit (in comparison to model 1 which only includes the condition as a predictor: $\chi 2(5) = 93.16$, $p < .01$). The total explained variance (R^2) of model 2 compared to model 0 (intercept-only) was 20.8% and Cohen's f^2 was 0.26, indicating a medium effect size for the overall model (Cohen, 1992; Lorah, 2018).

Model 2 shows that pretest scores on historical inquiry skills and procedural knowledge on historical inquiry both significantly predicted the posttest score for historical inquiry (respectively $p < .001$ and $p = .004$). For each unit that students scored higher on procedural knowledge in pretest, the result of the historical essay task in the posttest was 0.3 higher. When controlling for these pretest scores, students in the intervention condition performed significantly better than students in the control condition ($p = .007$, $B = 4.79$). Neither the sensitivity of the topic of the historical essay task, nor task value were significant covariates. There was also no significant interaction effect of condition*pretest score on historical inquiry skills.

The multilevel models showed that the intervention succeeded in significantly improving students' historical inquiry skills. When controlling for the covariates (see model 2) students in the intervention group scored on average 4.79 points higher, out of a total score of 24, on the posttest historical essay task than students in the control group. R^2 for model 2 compared to the same model without the fixed effect of condition was 13%, indicating an effect size of $f^2 = 0.15$ and hence a medium effect size of condition on students' historical inquiry skills at posttest.

Table 5. Multilevel models for historical inquiry skills at posttest

	Model 0 *B* (*SD*)	Model 1 *B* (*SD*)	Model 2 *B* (*SD*)
Fixed effects			
Intercept	15.90 (0.68)*	17.31 (0.60)*	12.24 (1.34)*
Condition (control group)		-3.91 (1.01)**	-4.79 (1.74)**
Pretest score on historical inquiry skills			0.23 (0.06)*
Pretest score on procedural knowledge			0.30 (0.11)**
Sensitivity of historical essay topic			-0.11 (0.17)
Task value			0.11 (0.16)
Interaction condition*pretest score on historical inquiry skills			0.08 (0.10)
Variance components			
Residual (between students)	19.42 (1.097)*	19.46 (1.10)*	18.23 (1.04)*
Between schools	6.44 (2.52)***	2.81 (1.35)***	2.62 (1.19)***
Between classes	0.63 (0.48)	0.62(0.48)	0.12 (0.35)
Number of parameters	4	5	10
Model fit			
AIC	3936.9	3929.275	3846.113
BIC	3954.917	3951.796	3891.051
-2LL	3928.9	3919.275	3826.113

Note. *$p < .001$, **$p < .01$, ***$p < .05$.

Effects of the intervention on students' procedural knowledge on historical inquiry

Model 2 (Table 6) shows that the intervention significantly improved students' procedural knowledge of historical inquiry ($p = .003$) when controlling for pretest score on procedural knowledge, which also significantly predicted the posttest score ($p < .001$, $B = 2.06$). There was no significant effect of students' pretest scores on historical inquiry skills or task value, nor an interaction effect of condition*pretest score on procedural knowledge. This model constituted a better fit than model 1 ($\chi 2(4) = 83.86$, $p < .01$). In comparison to the null model, the total explained variance was 9.5%. R^2 of this model compared to model 0 was 21%,

Cohen's f^2 was 0.27 implying in a medium effect size for the overall model (Cohen, 1992).

Thus, the intervention significantly improved students' procedural knowledge about historical inquiry. In particular, when controlling for the covariates in model 2, students in the intervention group scored at posttest on average 2.06 points higher on procedural knowledge of historical inquiry than students in the control group. As this was out of a total score of 15 and students' average scores at posttest were still rather low (see Table 4), this does not appear to be a large difference. Comparing model 2 with an identical model without the condition revealed that R^2 was 12% and Cohen's f^2 was 0.14. This indeed points to a medium effect of condition on students' procedural knowledge at posttest (Cohen, 1992).

Table 6. Multilevel models for procedural knowledge of historical inquiry at posttest

	Model 0 *B* (*SD*)	Model 1 *B* (*SD*)	Model 2 *B* (*SD*)
Fixed effects			
Intercept	5.39 (0.36)*	6.11 (0.33)*	4.03 (0.67)*
Condition (control group)		-1.997 (0.55)**	-2.06 (0.68)**
Pretest score on procedural knowledge			0.40 (0.07)*
Pretest score on historical inquiry skills			0.01 (0.03)
Task value			0.05 (0.09)
Interaction condition*pretest score on procedural knowledge			0.05 (0.12)
Variance components			
Residual (between students)	5.88 (0.33)*	5.88 (0.33)*	5.51 (0.31)*
Between schools	1.72 (0.69)***	0.78 (0.39)***	0.51 (0.29)
Between classes	0.25 (0.17)	0.26 (0.17)	0.18 (0.15)
Number of parameters	4	5	9
Model fit			
AIC	3140.123	3133.087	3057.224
BIC	3158.040	3155.608	3097.667
-2LL	3132.023	3123.087	3039.224

Note. *$p < .001$, **$p < .01$, ***$p < .05$.

Effects of the intervention on students' epistemological beliefs

Table 7 and 8 report the multilevel models for subscale 1 (beliefs about the simplicity and certainty of knowledge), and subscale 2 (beliefs about the value of engaging in integration of information across documents) respectively. In both cases, model 2 which includes the condition, pretest scores on the same variable and a condition*pretest score interaction effect, constituted the best fit (in comparison to model 1: $\chi 2(2) = 307.097$, $p < .01$ for subscale 1, and 1: $\chi 2(2) = 219.842$, $p < .01$ for subscale 2). Total explained variance compared to the null model was 40% for subscale 1 and 29.6% for subscale 2. R^2 of model 2 compared to model 0 was 39% for subscale one and 30% for subscale to. Cohen's f^2 was respectively 0.64 and 0.43 indicating large effects sizes for these models (Cohen, 1992; Lorah, 2018).

However, the models show that there is no significant effect of condition on students' posttest scores for epistemological beliefs, nor an interaction effect of condition*pretest score. Only students' pretest scores on the related epistemological subscale significantly predicted posttest scores ($p < .001$ for both subscales). In short, the intervention did not significantly affect students' epistemological beliefs.

Table 7. Multilevel models for epistemological beliefs, subscale 1, at posttest

	Model 0 *B* (*SD*)	Model 1 *B* (*SD*)	Model 2 *B* (*SD*)
Fixed effects			
Intercept	2.15 (0.04)*	2.12 (0.05)*	0.63 (0.10)*
Condition (control group)		0.08 (0.08)	0.03 (0.17)
Pretest score on EB_subscale1			0.64 (0.04)*
Interaction condition*pretest score on EB_subscale1			-0.003 (0.07)
Variance components			
Residual (between students)	0.399 (0.02)*	0.399 (0.02)*	0.26 (0.02)*
Between schools	0.01 (0.01)	0.01(0.01)	0.00 (0.00)
Between classes	0.02(0.01)	0.01(0.01)	0.00 (0.00)
Number of parameters	4	5	7
Model fit			
AIC	1317.434	1318.555	1015.458
BIC	1335.452	1341.076	1046.936
-2LL	1309.434	1308.555	1001.458

Note. $*p < .001$, $**p < .01$, $***p < .05$. EB = Epistemological beliefs.

Table 8. Multilevel Models for Epistemological Beliefs, Subscale 2, at Posttest

	Model 0 *B* (*SD*)	Model 1 *B* (*SD*)	Model 2 *B* (*SD*)
Fixed effects			
Intercept	4.82 (0.05)*	4.87 (0.05)*	1.71 (0.24)*
Condition (control group)		-0.11 (0.08)	0.44 (0.40)
Pretest score on EB_subscale2			0.66 (0.05)*
Interaction condition*pretest score on EB_subscale2			-0.11 (0.08)
Variance components			
Residual (between students)	0.39 (0.02)*	0.39 (0.02)*	0.29 (0.02)*
Between schools	0.01 (0.02)	0.01 (0.01)	0.001 (0.01)
Between classes	0.02 (0.01)	0.02 (0.01)	0.005 (0.01)
Number of parameters	4	5	7
Model fit			
AIC	1305.536	1306.033	1030.191
BIC	1323.553	1328.554	1121.668
-2LL	1297.536	1296.033	1076.191

Note. $*p < .001$, $**p < .01$, $***p < .05$. EB = epistemological beliefs.

In answer to the first research question, it can be concluded, based on the results of the multilevel models, that the lesson series was effective in improving students' historical inquiry skills, as well as their procedural knowledge about inquiry. The lesson series did not change students' epistemological beliefs about history.

RQ 2: Effects of the intervention on democratic citizenship

The second research question examined whether the lesson series focused on students' historical inquiry skills resulted in changes in their democratic skills and dispositions. A multilevel model was built for democratic skills, deliberation within, critical openness and perspective taking. In line with the previous research question, the condition functioned as a predictor in the multilevel models. However, while RQ 1 demonstrated that the intervention succeeded in improving students' historical inquiry skills as well as related procedural knowledge, significant variance still occurred between students. Because it is plausible that the effect of the intervention differed among students, the specific role of changes in students' scores on historical inquiry skills and procedural knowledge on inquiry was also examined to account for differences in students' progress.

Effects of the intervention on students' democratic skills

Table 9 shows the multilevel models for students' democratic skills, based on the topical essay task, at posttest. In model 2, the effects of condition, and an interaction effect of both were examined, with the sensitivity of the essay topic and pretest measures for democratic skills functioning as a covariate. In model 3, students' changes in procedural knowledge on inquiry and on historical inquiry skills were added, as well as an interaction effect with the condition. This model generated the best fit (in comparison to model 2: $\chi 2(4) = 46.548$, $p < .01$ and to model 1: $\chi 2(7) = 142.943$, $p < .01$) and provided a total explained variance of 26% compared to model 0. R^2 for this model compared to the null model was 26%, Cohen's f was .35 and hence points to a large effect of the overall model (Cohen, 1992).

Based on model 3, there was no condition effect on democratic skills at posttest, nor a condition*pretest score interaction effect. The sensitivity of the essay task was not a significant covariate. However, while the condition in itself did not significantly predict students' democratic skills as posttest, the change in

students' historical essay skills was a significant predictor (B = 0.16; p < .001). Moreover, a significant interaction effect was established between students' changes in historical inquiry skills and the condition (p = .005), indicating that the influence of changes in students' historical inquiry skills on democratic skills was slightly larger in the intervention group than in the control group. Changes in students' procedural knowledge of inquiry were not a significant predictor, nor was there an interaction effect with the condition.

Even though the condition was not a significant predictor for students democratic skills at posttest the multilevel models indicate that students' progress on the historical inquiry skills was related to an increase in democratic skills. Each one point increase in students' scores on historical inquiry skills, corresponded to an increase of 0.16 on democratic skills. This, however, was only a small effect as R^2 for model 3 compared with a model without this predictor was 6% and Cohen's f^2 was .06.

Table 9. Multilevel models for democratic skills at posttest

	Model 0 *B* (*SD*)	Model 1 *B* (*SD*)	Model 2 *B* (*SD*)	Model 3 *B* (*SD*)
Fixed effects				
Intercept	15.196 (0.46)*	15.68 (0.53)*	10.22 (0.86)*	10.15 (0.82)*
Condition (control group)		-1.34 (0.89)	-0.98 (1.24)	-0.75 (1.19)
Pretest score on democratic skills			0.37 (0.05)*	0.36 (0.05)*
Sensitivity of topical essay topic			0.07 (0.11)	0.04 (0.11)
Interaction condition*pretest score on democratic skills			-0.02 (0.07)	-0.01 (0.08)
Change in procedural knowledge score				0.01 (0.05)
Change in historical inquiry skills				0.16 (0.03)*
Interaction condition*change in procedural knowledge score				0.15 (0.13)
Interaction condition*change in historical inquiry skills				-0.14 (0.05)**
Variance components				
Residual (between students)	11.28 (0.64)*	11.28 (0.64)*	9.96 (0.56)*	9.64 (0.55)*
Between schools	2.84 (1.17)***	2.35 (1.03)***	1.25 (0.62)***	0.89 (0.49)
Between classes	0.33 (0.28)	0.33 (0.28)	0.25 (0.23)	0.21 (0.22)
Number of parameters	4	5	8	12
Model fit				
AIC	3563.785	3563.737	3473.342	3434.794
BIC	3581.796	3586.251	3509.353	3488.737
-2LL	3555.785	3553.737	3457.342	3410.794

Note. *$p < .001$, **$p < .01$, ***$p < .05$.

Effects of the intervention on students' democratic dispositions

To analyze the intervention's effect on students' democratic dispositions, separate models were built for deliberation within, critical openness and perspective taking. However, because variance at school and class level was close to zero for each of these dispositions, these parameters were redundant and reliable parameter estimates could not be generated via the data analysis software. Therefore, linear regressions were used to examine this research question.

Table 10 and 11 show the results of the linear regressions respectively for deliberation within (DW) and critical openness (CO). For both dispositions model 3 accounted for the most variance. These models included the condition, pretest scores on the particular disposition, an interaction of this pretest score and the condition, change in procedural knowledge scores and in historical inquiry skills as well as interaction between these change scores and the condition. These models were significant predictors of students' scores on deliberation within ($F(7,654) = 18{,}562$, $p < .001$) and critical openness ($F(7,654) = 39.940$, $p < .001$) at posttest. For deliberation within, model 3 accounted for 15.7% of variance in the outcome variable, which was a significant increase compared to the second model ($F_{change}(4, 654) = 2.617$, $p = .034$). For critical openness, the third model explained 29.2% of the variance at posttest, which constituted a significant increase from model 2 ($F_{change}(4, 654) = 4.013$, $p = .003$).

The models for these variables show that only pretest measures on these variables significantly predicted their posttest scores ($p < .001$ for DW and CO). The condition was not a significant predictor, nor was there an interaction effect with the pretest score. Changes in procedural knowledge and historical inquiry skills also did not significantly predict posttest scores. Interaction effects between the condition and changes in procedural knowledge and between the condition and changes in historical inquiry skills were also not significant.

Results for linear regressions for perspective taking at posttest are reported in Table 12. Model 2 significantly predicted posttest scores for perspective taking ($F(3,658) = 177.479$, $p < .001$), explaining 45.5% of variance. This model included the condition, pretest score on perspective taking and an interaction between both as predictors. This was a significant increase from model 1 ($F_{change}(2, 658) = 264.630$, $p < .001$) while model 3 did not significantly increase the explained variance ($F_{change}(4, 654) = 1.003$, $p = .405$). In line with findings for deliberation

within and critical openness, there was no condition effect and no interaction effect of the condition and the pretest score on perspective taking. Only the pretest score for perspective taking significantly predicted the posttest score ($p < .001$).

In conclusion, based on the linear regression models the intervention appeared to have no effect on students' democratic dispositions.

Table 10. Linear regression models for deliberation within at posttest

	Model 1		Model 2		Model 3	
	B (*SD*)	*β*	*B* (*SD*)	*β*	*B* (*SD*)	*β*
Constant	3.74 (0.03)*		2.07 (0.23)*		2.07 (0.23)*	
Condition	0.097 (0.04)***	0.09***	0.39 (0.30)	0.36	0.33 (0.30)	0.30
Pretest score on DW			0.46 (0.06)*	0.43*	0.45 (0.06)*	0.43*
Interaction condition*pretest score DW			-0.09 (0.08)	-0.30	-0.08 (0.08)	-0.29
Change in procedural knowledge score					-0.00 (0.02)	-0.02
Change in historical inquiry skills					-0.01 (0.01)	-0.06
Interaction condition*change in procedural knowledge score					0.02 (0.02)	0.12
Interaction condition*change in historical inquiry skills					0.01 (0.01)	0.09
Model fit						
Adjusted R^2	0.006		0.149		0.157	
R^2 change	0.008***		0.145*		0.013***	

Note. * $p < .001$, ** $p < .01$, *** $p < .05$. DW = Deliberation within.

Table 11. Linear regression models for critical openness at posttest

	Model 1		Model 2		Model 3	
	B (*SD*)	*β*	*B* (*SD*)	*β*	*B* (*SD*)	*β*
Constant	3.92 (0.03)*		1.40 (0.23)*		1.38 (0.23)*	
Condition	-0.02 (0.04)	0.02	0.27 (0.30)	0.30	0.20 (0.30)	0.22
Pretest score on CO			0.64 (0.06)*	0.57	0.64 (0.06)*	0.57
Interaction condition*pretest score CO			-0.07 (0.08)	-0.32	-0.07 (0.07)	-0.31
Change in procedural knowledge score					0.02 (0.01)	0.12
Change in historical inquiry skills					-0.00 (0.00)	-0.03
Interaction condition*change in procedural knowledge score					-0.00 (0.02)	-0.01
Interaction condition*change in historical inquiry skills					0.01 (0.01)	0.01
Model fit						
Adjusted R^2	-0.001		0.279		0.292	
R^2 change	0.000		0.282*		0.017**	

Note. *$p < .001$, **$p < .01$, ***$p < .05$. CO = Critical openness

Table 12. Linear regression models for perspective taking at posttest

	Model 1		Model 2		Model 3	
	B (*SD*)	*β*	*B* (*SD*)	*β*	*B* (*SD*)	*β*
Constant	3.55 (0.03)*		1.38 (0.15)*		1.38 (0.15)*	
Condition	0.05 (0.04)	0.05	-0.13 (0.20)	-0.13	-0.13 (0.20)	-0.13
Pretest score on PT			0.62 (0.04)*	0.64	0.62 (0.04)*	0.64
Interaction condition*pretest score PT			0.04 (0.06)	0.16	0.04 (0.06)	0.13
Change in procedural knowledge score					-0.01 (0.02)	-0.04
Change in historical inquiry skills					0.00 (0.00)	0.00
Interaction condition*change in procedural knowledge score					0.02 (0.02)	0.10
Interaction condition*change in historical inquiry skills					0.00 (0.01)	0.01
Model fit						
Adjusted R^2	0.001		0.445		0.445	
R^2 change	0.003		0.445*		0.003	

Note. * $p < .001$, ** $p < .01$, *** $p < .05$. PT = Perspective taking

DISCUSSION

Through a cluster randomized controlled intervention study, we sought to examine the disciplinary effects of a lesson series focused on fostering students' historical inquiry competence and to examine the assumption that this would spontaneously result in civic effects. To this end a lesson series was designed, aimed at fostering students' historical inquiry competencies according to the principles of modelling, scaffolding and coaching (Collins et al., 1991) and taking into account recommendations from research on fostering historical thinking among students. The effects of the intervention on students' historical inquiry skills, procedural knowledge and epistemological beliefs and on related democratic skills and dispositions were examined.

Effects on students' historical inquiry competence and epistemological beliefs (RQ1)

The first research question examined the intervention's effects on students' historical inquiry skills, procedural knowledge on inquiry and epistemological beliefs. With regard to students' historical inquiry competences, the intervention succeeded in significantly improving both students' historical inquiry skills and procedural knowledge on inquiry. The condition was found to be a significant predictor of the posttest scores on historical inquiry skills and procedural knowledge. In both cases a medium effect size was established.

Moreover, a relationship between historical inquiry skills and procedural knowledge was established as students' procedural knowledge at pretest constituted a significant predictor of their inquiry skills at posttest. A weak positive correlation was also established between students' historical inquiry skills and procedural knowledge at posttest, which was not present at the pretest level. It appears that the intervention not only improved students' historical inquiry skills and procedural knowledge separately, but cultivated a connection between the two. This may be due to the fact that the procedural knowledge became meaningful to students as they learned to apply it in concrete tasks.

Despite the intervention's effectiveness in fostering students' historical inquiry competences, no effects were found on students' epistemological beliefs about history.

These results point to the effectiveness of the design principles applied in the intervention that are in line with recommendations from scholars to make disciplinary ways of reasoning and procedural strategies visible to students (e.g. Britt & Aglinskas, 2002; Stahl et al., 1996; van Boxtel et al., 2021) and that support findings from earlier studies using multiple documents based task (e.g. Monte-Sano & De La Paz, 2012; De La Paz et al., 2016; Nokes et al., 2007; Stoel, van Drie, & van Boxtel, 2017). The use of modelling, scaffolding and coaching, as well as explicit teaching on both second-order and procedural knowledge likely helped the lesson series succeed in its primary objective of fostering students' inquiry competence. However, even though the condition was found to be a significant predictor of student scores on historical inquiry and procedural knowledge, significant variance remained in existence between students and, with regard to historical inquiry skills, also between schools, indicating that the intervention yielded varying results. Furthermore, the lesson series was only effective with regard to historical inquiry competence, but did not generate epistemic change. This is a somewhat surprising finding given the expectation that students' epistemological beliefs would be connected to their historical inquiry skills (e.g. van Boxtel et al., 2021; Wiley et al., 2020). It appears that it is possible to improve students' historical inquiry skills and procedural knowledge without affecting their underlying beliefs about knowledge (production) in history. Two explanations for this lack of effect on students' epistemological beliefs can be considered. A first one may be found in the design of the lesson series. Although explicit teaching about the process of historical inquiry implicitly touched upon the constructed and interpretive nature of history, this was not a main focus of explicit teaching during the lesson series. Another is that epistemological beliefs appear to be rather stable beliefs that are not prone to change and are less affected by teachers' instructional practices. This is in line with previous research showing that epistemic change is not easily obtained, even when explicit attention is paid to it (e.g. Stoel et al., 2015). The idea that students' epistemological views are rather stable and less strongly influenced by instructional practices is also supported by the fact that there was no significant variance for these views at either school or classroom level, even though this was the case for historical inquiry competences. In this sense, it seems logical that the relatively short duration of the intervention was insufficient to generate changes in these underlying beliefs.

Effects on democratic citizenship (RQ2)

The second research question studied the effects of the intervention on students' democratic skills and dispositions, thereby gauging a possible transfer between historical inquiry competences and democratic citizenship. The multilevel models showed no effects of the intervention on students' democratic skills and dispositions.

However, with regard to democratic skills, the multilevel models which particularly examined the effect of changes in students' historical inquiry skills and procedural knowledge generated a more nuanced picture. They suggested the existence of a relationship between students' historical inquiry skills and democratic skills. These models took into account the possibility that the intervention had not generated the same effect among all students. This showed that, even though the condition in itself was not a significant predictor for democratic skills at posttest, the changes in students' historical inquiry skills actually were. Students who made more progress on the historical essay task also scored significantly higher on the topical essay task indicating that students' ability to take a substantiated position towards a historical question and towards a societal debate covaried. Moreover, a significant interaction effect was found between these changes and the condition, indicating that the effect of changes in students' historical inquiry skills on democratic skills was slightly larger in the intervention group than in the control group. Although these results establish a correlation between historical skills and democratic skills, they do not provide a definitive answer as to the direction of the relationship and hence regarding the question of transfer. Considering the set-up of the intervention study and the fact that the lesson series deliberately did not address democratic skills, it is plausible that the progress in historical skills affected their democratic skills. It is, however, also possible that students improved slightly on the democratic essay tasks and this affected their historical skills. Students' progress on both tasks may also have been influenced by their general writing abilities, something which was not specifically controlled for in the study. Because the measures for historical and democratic skills were based on a writing task, it is possible that the progress on both was due to a general improvement in students' overall writing ability.

As such, the data hence do not provide clear evidence for a transfer from students' historical inquiry skills to democratic skills, contrary to the expectations

in the literature (e.g. Monte-Sano, 2016; Nokes, 2013; Wineburg & Reisman, 2015). Moreover, the relation between students' progress on historical and democratic skills was very small, as was the interaction effect with the condition. It thus appears that without explicit teaching particularly directed at the development of these democratic skills, the impact of history classes aimed at fostering historical inquiry competences will remain limited.

Furthermore, neither the condition nor changes' in students historical inquiry competence yielded significant changes in students' democratic dispositions. Contrary to what the literature suggested, enhancing students' historical inquiry competence did not affect their democratic dispositions. A number of explanations can be considered in this regard. First, unlike with regard to democratic skills, it appears based on the correlation matrix that there is no relationship between historical inquiry competences and democratic dispositions, or at least not to the dispositions examined in this study. Although the dispositions of deliberation within, perspective taking and critical openness closely adhered to those described by scholars such as McCully (2012), Barton and Levstik (2004) and Monte-Sano (2016) it is possible that other measures would constitute a better fit. A second explanation for the absence of changes in students' dispositions may be found in the design of the intervention. While historical inquiry skills were explicitly modelled and taught, this was not the case for the underlying dispositions. Here as well, the lack of explicit teaching particularly directed at these dispositions may provide an explanation for absence of any change. Lastly, we should consider the possibility that, similar to students' epistemological beliefs, these underlying dispositions are more robust and are simply less prone to change in the course of a few weeks. The absence of variance at school and class level provides support for the robustness of these dispositions. Generating change on these dispositions may thus require more extensive and explicit attention over longer periods of time.

The role of affective variables

For both research questions, the possible influence of affective dimensions in students' performance was examined via a measure of task value and by taking into account the sensitivity of the essay topics. At pretest level, the correlation matrix demonstrated little correlation between affective variables and students'

scores on the essay tasks and procedural knowledge, which already indicated that their influence might be negligible. This was confirmed in the multilevel models for historical inquiry skills, procedural knowledge and democratic skills which found that neither the sensitivity of the essay task topics nor task value were significant covariates.

As far as sensitivity is concerned, literature suggested that high levels of emotional connection with the topic addressed might hinder students' ability to consider multiple viewpoints (Barton & McCully, 2007; Goldberg et al., 2011). Considering that students on average, did not consider the essay topics at posttest to be particularly sensitive, it could be expected that no relationship would be found.

Literature also suggested that higher levels of task value may be related to better learning outcomes (Wigfield & Eccles, 2000) but this was not confirmed in the current study. The finding that task value was not related to students' performance on historical skills and procedural knowledge goes against findings from Paul and colleagues (2017) on sourcing in history. Based on our findings, the importance of affective beliefs in students' performance on historical inquiry may be questioned. However, we would argue that more research is needed to explore the role of these beliefs, especially as they are rarely taken into account in large-scale intervention studies. Moreover, in this study we examined task value for history education in general asking students about their views on history as a school subject. It is possible that there was a mismatch between students' conception of history education and what it entails and our implementation of it in the intervention. This is especially plausible since Flemish history teachers tend to emphasize knowing history over doing history and do not regularly include historical inquiries in their practice (Voet & De Wever, 2016; 2019). Thus, the history lessons that students were familiar with at pretest were likely different from those in the intervention.

Limitations

A number of limitations of the current study should be mentioned. One major limiting contextual factor was the impact of the corona measures that were in place in Flemish history education at the time of the study. During the study, half of the classes in secondary education took place via remote education, either synchronous or asynchronous. Although every effort was made to mitigate the impact on the study, the measures did affect certain aspects of the research. We experienced a larger than expected drop-out of teachers in the study, both before the start of the intervention and during because teachers were unable to complete the intervention as intended. This drop-out lowered the study's sample size and disproportionally affected the control group, resulting in an uneven sample in both conditions. Teachers who were able to complete the intervention also reported that they were unable to provide the same teaching quality through remote education. They felt less connected to their students and reported that essential aspects of the lesson design, such as interaction in whole-class discussions and providing the appropriate scaffolds in line with students' needs, were more difficult to achieve. These circumstances may have weakened the effectiveness of the intervention.

Another limitation of the study is that we have no information about the long-term effects of the intervention. As the study took place at the end of secondary education, we were not able to include a delayed post-test.

Furthermore, it is possible that some of the progress between pretest and posttest may be attributed to a testing effect or differences between the difficulty level of the pre- and posttest essay tasks. Because of changes in corona measures we were unable to conduct the planned full-scale pilot study to examine differences in the difficulty levels between both tests. These were hence only examined in a qualitative way based on a limited pilot study and on a discussion with the expert panel. Although this provided no indications that the difficulty levels of the essay tasks differed, this could not be definitively established. The use of a control group, however, should ensure that the influence of a possible test effect was mitigated.

A possible bias may also have occurred in our selection of participating teachers. Considering that we launched an open call for participants, it is possible that only the most motivated teachers applied to participate in the study.

Alternatively, in the context of the upcoming curriculum renewal, it is possible that the call for participants was of particular interest for teachers who felt unsure about how to implement the renewed history standards. The randomized assignment to the experimental and control condition, however, should ensure that potential differences in teachers' motivation do not affect the study's result.

Another limitation concerns the manner in which treatment fidelity was evaluated. Due to the corona measures in Flemish education, we had to find alternative ways to evaluate treatment fidelity remotely and opted for self-reports by teachers, supplemented with interviews. Based on these data, we were able to effectively identify teachers who had deviated significantly from the scripted curriculum. However, as we were unable to independently establish teachers' adherence to the curriculum based on lesson observations, we cannot say with complete certainty that the teachers in the study strictly adhered to the prescribed curriculum.

This study also employed a very specific lens through which the relationship between historical thinking and democratic citizenship was defined. Our operationalization of democratic skills and dispositions is connected to a view of democratic citizenship as deliberative, pluralistic and participatory (Barton & Levstik, 2004). In this view, education for citizenship should go beyond a transfer of knowledge but should enable students to become critical, responsible and active members of society (European Commission, 2017; Maurissen, 2018). In this study, we focused mainly on history education's contribution to the critical aspect of citizenship, as a result of our specific focus on historical inquiry competences as part of historical thinking. By adopting this particular focus, other possible contributions of historical thinking to democratic citizenship could not be taken into account. For instance, historical thinking may be able to contribute to democratic knowledge about (the background of) current political systems and institutions (Stearns, 1998). Alternatively, students may be encouraged to take on an active engagement in society when systematically studying the agency of various historical actors as part of historical thinking (Barton, 2012; Seixas, 2012). Moreover, in defining the possible contribution of historical inquiry competences to democratic dispositions, we used literature from history educational scholars as a starting point and have tried to connect to existing concepts from literature on citizenship that fit those descriptions as closely as possible. However, an argument

can also be made for the use of other concepts and measurement instruments, especially considering that the validated measurement scales employed in this study did not always generate a high internal consistency.

CONCLUSION

This research makes an important contribution to the quality and effectiveness of history teaching, and on the discussions regarding the relationship between historical thinking and democratic citizenship. Overall, we found that the intervention designed to foster students' historical inquiry competences was successful in improving these competences and, to a limited extent, in fostering students' democratic skills. The results of this study provide further evidence for the effectiveness of teaching strategies derived from the model of cognitive apprenticeship and the extensive use of explicit teaching.

However, the current study also suggests that we need to be careful about the expectations we ascribe to scripted interventions in (history) education. Although this study used a scripted curriculum and strict exclusion criteria to minimize variation between teachers and schools, differences still occurred not only between teachers and schools, but also in the way that students engaged with and reacted to the intervention. The effectiveness of the intervention varied among students, which became even more pertinent when examining the second research question.

The results also indicate that we should not be overly optimistic about the potential spontaneous contribution of historical thinking, and specifically of historical inquiry competences, to democratic citizenship. Despite the effectiveness of the intervention on students' historical inquiry competence, we only established a very small relationship with students' democratic skills suggesting that the two are indeed related, but that a transfer from historical to contemporary contexts will not occur spontaneously, as some scholars seem to have assumed. We should be careful with the expectation that history education, and historical inquiry competences specifically, will automatically contribute to extrinsic goals associated with democratic citizenship. It seems that meeting these democratic aims will require that explicit teaching is directed specifically at fostering these democratic skills and dispositions. This compels us to think about

where precisely these democratic goals should be given a place in education and what role, if any, we can and should expect history education to take on in this regard. Especially considering the fact that historical thinking is considered by various history education scholars to be a valuable goal in itself but one that is difficult to foster among students the question arises as to whether history education should be expected to (also) focus on these democratic aims. This does not negate the fact that history education may be able to contribute to democratic citizenship in other ways, for instance by providing insight into the historical origins of current societal phenomena.

Based on the results of this study, it also seems useful to reconsider the relationship between students' skills and their underlying beliefs and dispositions. Surprisingly, this study did not find a strong relationship between skills (historical and democratic) and epistemological beliefs and dispositions, suggesting that changing students' historical or democratic skills may not suffice to alter underlying epistemological beliefs and dispositions. Although it hence appears that it is possible to change students' historical skills without changing their underlying beliefs, it nevertheless seems important to continue trying to alter both, even if one does not necessarily affect the other. Changing students' more fundamental views of what constitutes historical knowledge and how it can be acquired, might enable more robust and long-term effects in terms of their engagement with historical sources and history in general. To this end, it would be beneficial to examine whether changes on this level can be acquired after more long-term interventions which may be needed to allow for changes in these robust beliefs. Furthermore, future intervention studies may also benefit from more explicit attention to fostering nuanced epistemological beliefs among students. The correlations between epistemological beliefs and democratic dispositions suggests that perhaps an explicit focus on epistemic change may lead to changes in democratic dispositions.

Considering that this intervention study did not yield the same effects for all students, it also seems imperative that future studies pay even more attention to differences between students and allow for differentiation, even in scripted curricula. It may also be fruitful to gather process data both in light of obtaining a better view of treatment fidelity and as a way to potentially explain differences between schools or classes within the intervention group. In examining the

relationship between historical thinking and democratic citizenship, further research might also benefit from more clarity on what this relationship precisely looks like and what kind of effects would be expected, as current literature remains rather vague in this regard.

APPENDIX A: PEARSON'S *R* CORRELATIONS BETWEEN VARIABLES AT PRE- AND POSTTEST

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1.	Hist. inquiry skills – pretest	--																				
2.	Hist. inquiry skills – posttest	.25**	--																			
3.	Sensitivity hist. essay - pretest	.15**	.06	–																		
4.	Sensitivity hist. essay - posttest	.04	-.04	.38**	–																	
5.	Procedural knowledge - pretest	.04	.18**	.02	-.01	--																
6.	Procedural knowledge - posttest	.06	.30**	-.01	.03	.38**	--															
7.	EB_1 - pretest	-.09*	-.13**	.10*	.11**	-.19**	-.12**	--														
8.	EB_1 - posttest	-.09*	-.07	.01	.11**	-.15**	-.06	.61**	–													
9.	EB_2 - pretest	.04	.03	-.03	-.01	.14**	.09*	-.30**	-.27**	--												
10.	EB_2- posttest	.10**	.15**	.06	-.03	.13**	.15**	-.36**	-.47**	.53**	–											
11.	Dem. skills - pretest	.30**	.27**	-.01	-.04	.11**	.20**	-.15**	-.06	.09*	.10*	--										
12.	Dem. skills - posttest	.24**	.44**	-.03	-.06	.13**	.25**	-.13**	-.05	.07	.10**	.43**	--									
13.	Sensitivity topical essay - pretest	-.03	.04	.12**	.14**	.08*	.06	-.12**	-.08*	.09*	.05	.12**	.01	--								
14.	Sensitivity topical essay – posttest	.03	.06	.14**	.17**	.10*	.04	-.06	-.06	.06	.06	.04	.03	.31**	–							
15.	DW - pretest	.09*	.10*	.16**	.03	.06	.04	-.12**	-.15**	.19**	.26**	.13**	.06	.09*	.05	--						
16.	DW - posttest	.07	.11**	.15**	.03	.09*	.16**	-.13**	-.22**	.25**	.35**	.12**	.12**	.05	.08*	.39**	–					
17.	CO - pretest	.05	.02	.18**	.05	.11**	.06	-.21**	-.19**	.32**	.35**	.06	-.01	.14**	.09*	.44**	.32**	–				
18.	CO - posttest	.03	.05	.13**	.10**	.12**	.15**	-.18**	-.20**	.30**	.36**	.07	.05	.09*	.10**	.27**	.43**	.53**	--			
19.	PT - pretest	.03	-.01	.24**	.10*	.11**	.15**	-.12**	-.13**	.18**	.22**	.06	.04	.04	.02	.27**	.26**	.46**	.44**	–		
20.	PT - posttest	.03	.02	.17**	.08	.09*	.16**	-.16**	-.21**	.19**	.26**	.10*	.08*	.03	.03	.20**	.34**	.40**	.50**	.67**	–	
21.	Task value	.04	.04	.15**	.12**	.16**	.10**	-.16**	-.13**	.32**	.25**	.02	.01	.15**	-.04	.23**	.17**	.37**	.29**	.24**	.19**	--

Note. $^{**}p < .01$. $^{*}p < .05$. Hist. = historical, Dem. = democratic, DW = Deliberation within, PT= perspective taking, CO= critical openness, EB_1/2 = epistemological beliefs, subscale 1/subscale 2.

APPENDIX B: HISTORICAL ESSAY TASK AT PRETEST

Historical background

South Africa has a long colonial history. It was colonized in the 17th century by the Netherlands and later by Great Britain. A large number of Dutch and British colonists settled permanently in the area. Compared to other colonized regions in Africa, South Africa gained its independence relatively early (1931). However, the country retained strong ties with Great Britain, as part of the British Empire and later the British Commonwealth. Moreover, power remained in the hands of the colonists and inequality between different ethnic groups persisted.

From 1948, racial segregation and the associated inequality was officially enshrined in the Apartheid regime. Especially the black (indigenous) inhabitants were strongly discriminated against and disadvantaged.

Opposition to the Apartheid system came from various directions. Civil rights organizations, for example, offered resistance through protests, boycotts and strikes. One of these was the demonstration organized by the PAC (Pan African Congress) in Sharpeville on March 21st 1960. This demonstration is known for its violent outcome: a commotion arose at the gate that gave access to the police area, which resulted in dozens of people being shot by the police and almost 200 were wounded.

An important point of discussion after the events was the question of what was at the root of the police violence and whether the violence was justified or not. Some said the police started shooting spontaneously, others claimed that the protesters provoked the violence and yet others pointed to the difficult circumstances the police had to work in that day.

Instruction

In a special issue of the school paper, various aspects of the Apartheid regime in South Africa will be discussed. Write a contribution for this theme issue in which you write about the police violence during the Sharpeville demonstration. On the basis of the sources below, formulate a substantiated answer to the following question: "to what extent was the police violence at the Sharpeville demonstration in 1960 admissible or not? Make sure to take a clear position, present arguments and refer to sources.

Write a text of 15 to 20 sentences, addressed to your fellow students who will read this school paper.

Sources

1. Testimony of journalist Humprey Tyler, March 1960

Humphrey Tyler was a journalist with Drum magazine, a magazine that supported anti-Apartheid movements. He attended the Sharpeville demonstration as a journalist and reported on his experiences afterwards.

> Before the shooting, I heard no warning to the crowd to disperse. There was no warning volley. When the shooting started it did not stop until there was no living thing in the huge compound in front of the police station. The police have claimed they were in desperate danger because the crowd was stoning them. Yet only three policemen were reported to have been hit by stones – and more than 200 Africans were shot down. The police have also said that the crowd was armed with 'ferocious weapons' which littered the compound after they fled. I saw no weapons, although looked very carefully, and afterwards studies the photographs of the death scene; While I was there I saw only shoes, hats and a few bicycles left among the bodies. The crowd gave me no reason to feel scared, though I moved among them without any distinguishing mark to protect me, quite obvious with my white skin. I think the police were scared though, and I think the crowd knew it.

2. Photo of Ian Berry, taken at the Sharpeville-demonstration, March 1960 (Picture removed for copyright reasons)

This picture was taken by Ian Berry, a photographer who worked for Drum Magazine and was present at the demonstration together with Humphrey Tyler.

3. Testimony of Lieutenant-Colonel Pienaar, July 1960.

In an investigation into the events at Sharpeville, Lieutenant Colonel Pienaar was called as a witness. Pienaar was in charge of the police officers present at the demonstration. In his testimony, he explained how, on his arrival in Sharpeville, he found an "angry crowd of 20.000 natives" and ordered the police officers to load their guns. On the events that followed, he recounts the following:

- Pienaar: As far as I can recollect now, three things happened practically simultaneously. I was colonel Spengler again at the gate, and I saw that he was violently flung back into the police yard, and a mob bursting through the fate following him; I then heard two shots, double hosts, coming from the mob, and a heavy shower of stones fell amongst the police where I was. (…)"
- And when the gate flung open and you saw the mod entering and colonel Spengler flung back, what impression did you form as to what was happening?
- Pienaar: I was then satisfied that an attack was being launched.

- (...) what order were you going to give them colonel?
- Pienaar: Well at that stage I was prepared to give the order to fire on the gate where they were bursting through
- Now, before you could give that order, what happened?
- Pienaar: Then two shots rang out form the left flank, i.e. the Southern flank of the line of policemen, followed immediately by the whole line opening fire." (p. 956-957).

In the following excerpt, stemming from the cross-examination, Pienaer is questioned about the many people who died or were injured.

- "I am given to understand that a large number of the wounded were wounded in the back?
- Pienaar: Yes that is true. (…)
- Now is there anything that you can suggest as to why, if these people were fleeing, the firing continued so that these fleeing people were actually wound in the back?
- Pienaar: There are several factors that may have cause that. How far there are effective I don't know, but ricochets may cause wounds, the men may at the start have fired low or fired over their heads; and the crowd was milling and they could have been hit in the back that way. A number of them must have been hit whilst they were in flight. That is unavoidable. (p. 1001-1002)

4. Official statement of the South Africa High Commissioner in London, 1960, shortly after the events in Sharpeville

Shortly after the Sharpeville incident, an official statement was issued by the South Africa High Commissioner. This was the highest representative of the (white) South African government in Great Britain (similar to an ambassador). The statement was issued after fierce criticism from the international community about the events in Sharpeville.

> According to the factual information now available, the disturbances at Sharpeville on Monday resulted from a planned demonstration of about 20 000 natives in which demonstrators attacked the police with assorted weapon including firearms. The demonstrators shot first, and the police were forced to fire in self-defense and avoid even more tragic results.

5. Historian Tom Lodge, 2011

Tom Lodge is a political scientist and Professor of Peace and Conflict Studies at the University of Limerick (Ireland). He previously worked in South Africa. Among other things, he studies political processes in South Africa. The following extract is from his book "Sharpeville: An apartheid massacre and its consequences", published in 2011.

Lodge discusses the events at the fence that marks the entrance to the police station. He explains how chaos ensued and how Colonel Spengler stumbled. People in the front row pushed forward to see what was happening, causing the fence to fold inwards.

> Lieutenant Colonel Pienaar ordered the police to load their weapons, tough he insisted at the Commission of Inquiry that he told the men that they should not fire without a specific order from him. (…) a few people threw stones, conceding that these were the first stones thrown within the crowd that morning. He said at that juncture he was about to give the order to fire, tough what he 'had in mind' was ' a few well directed shots at the leader' not what was to follow. At this critical juncture, one of the few people in the crowd who really was armed decided to begin his own private offensive against the police. (…) One of his neighbors forced his pistol into the air but he fired two shots all the same. (…).
>
> The police fired in two volleys (…) The later shooting, directed from the Saracens found its targets in the mass of people fleeing northwards across the open veldt towards the shopping center; Several of the police fired 12 rounds each: they would have had to reload their weapons to achieve this and indeed Drum's photographer managed to capture on film one constable reloading his revolver. (p. 122-126).

APPENDIX C: EVALUATION RUBRIC USED TO ASSES HISTORICAL AND TOPICAL ESSAY TASKS

	1	2	3	4
Position statement (claim)	The essay does not contain a statement or claim in response to the historical/topical question	The essay contains a statement or claim that is only partly in response to the historical/topical question (e.g. not entirely in line with the question)	The essay contains a statement or claim in response to the historical/topical question, yet only implicitly or vaguely	The essay explicitly puts forward a statement or claim in response to the historical/topical question
Argumentation in support of the claim, based on sources	No arguments, derived from the sources, are used to support the position taken	The position taken is supported only to a limited extent by arguments derived from sources: • One or more arguments are mentioned • But: the argument(s) are not used as such/are not explicitly put forward in support of the position taken (e.g. only descriptive)	The position taken is mostly supported by arguments derived from sources: • One or more arguments are mentioned • But only some of the arguments are explicitly put forward in support of the position taken	The position taken is fully supported by arguments derived from sources • Multiple arguments are mentioned • And: arguments are explicitly put forward in support of the position taken
Quality of the argumentation in support of the claim	No arguments, derived from the sources, are used to support the position taken	The argumentation derived from the sources is incorrect: One or more arguments are mentioned but none of the arguments are correct.	The argumentation derived from the sources is partially correct: At least one correct argument is mentioned.	The argumentation derived from the sources is correct: At least two correct arguments are mentioned.
Use of sources in support of claim	The position taken is not supported by concrete sources. Or sources are referred to only in a general way (e.g. 'based on the sources I argue that…')	The position taken is supported only to a limited extent by concrete sources. • Only one concrete sources is used in support of the position taken. • Other sources are not referred to concretely or not mentioned at all..	The position taken is mostly supported by concrete sources: • Multiple sources are mentioned explicitly in support of the position taken. • But: not all sources are referred to concretely (it is not always clear which source is referred to) or sources are not put forward in support of the position taken	The position taken is fully supported by concrete sources Multiple sources are mentioned explicitly in support of the position taken.

	1	2	3	4
Evaluation of the value of the available sources (i.e. in terms of reliability and/or representativeness)	There is no attention to evaluating the sources' value: aspects of sources which might influence their value (e.g. author, date, goal of the source) are not mentioned at all.	There is superficial attention to evaluating sources' value: Relevant aspects of sources are mentioned (e.g. author, date, goal etc. of the source) but: • This is only done for one source • Or: this is not connected to an evaluation of a source's value • Or: the evaluation of the sources' value is mostly inaccurate/incorrect (e.g. simplistic binary between objective-subjective sources, inaccurate or simplistic reasoning, etc.)	There is limited attention to evaluating sources' value: Relevant aspects of a source are mentioned (e.g. author, date, goal etc. of the source) at least for one source, but: • Only for one sources is this connected to an evaluation of the source's value • And this evaluation is mostly accurate and correct.	There is a lot of attention to evaluating sources' value: Relevant aspects of multiple sources are mentioned (e.g. author, date, goal etc. of the source) and: • This is connected to an evaluation of multiple sources' value • And: this evaluation is mostly accurate and correct.
Discussion of counter arguments	There is no discussion of different perspectives and counter arguments. • Their existence is not mentioned. • Or: their existence is mentioned, but they are not refuted or challenged (no dialogue).	There is a superficial discussion of different perspectives and counter arguments: Their existence is mentioned and partly challenged or refuted, but: • This is done in a predominantly inaccurate or unclear way • And/or sources are only compared to each other in terms of their content (superficial).	There is a superficial discussion of different perspectives and counter arguments: Their existence is mentioned and partly challenged or refuted, not only by comparing their content, but also by comparing the sources' value, but: • This is only done for one source • And this is based only on one aspect that influences the sources' value.	There is a thorough discussion of different perspectives and counter arguments: Their existence is mentioned and challenged or refuted, not only by comparing their content, but also by comparing the sources' value: • This is done for at least two sources • Or (in case of only one source) this is based on multiple aspects that influence the source's value.

Chapter 6

Exploring differential effects of an intervention on historical inquiry tasks. A qualitative analysis of 12th grade students' progress.[1]

Multiple documents based (inquiry) tasks are often used to examine historical thinking as they require students to apply discipline-specific ways of reasoning and writing. Intervention studies using such tasks have often relied on principles from cognitive apprenticeship to make these discipline-specific heuristics explicit to students. While several studies found positive results, they offer little insight into how and where exactly students' progress on historical thinking manifests itself, nor on the differential effects of the intervention. Building on essay data gathered during an intervention study on students' historical inquiry skills, this study explores differential effects of the intervention according to students' initial historical inquiry ability. To this end, a purposeful sample of students was selected for whom the intervention was particularly effective. The qualitative analysis of students' essay tasks (pretest and posttest) revealed remarkable differences between students with high and low pretest scores. Although both groups made progress on all aspects of the essay task, they differed in terms of where and how this progress manifested itself: at posttest, students with a high initial score outperformed others in evaluating sources and rebuttals. This study offers insight into patterns of progress in students' historical inquiry skills which can inform differentiation in instructional practices.

[1] This chapter is based on: Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K.. Exploring differential effects of an intervention on historical inquiry tasks. A qualitative analysis of 12th grade students' progress. *Manuscript under review.*

INTRODUCTION

Since the 1990s, history education across several Western countries has undergone a shift from the teaching of particular (national) narratives to include a focus on "the way we go about doing history" (Lee & Ashby, 2000, p. 199; Lévesque & Clark, 2018). With the introduction of historical thinking into the history curricula, students not only acquire substantive knowledge about the past, but also second-order and procedural knowledge that introduces them into history as an academic discipline, allowing them to gain insight into the way that historical knowledge is constructed (van Drie & van Boxtel, 2008; VanSledright & Limón, 2006; Lee, 2004).

Historical thinking, however, constitutes a complex and 'unnatural' act (Wineburg, 2001) as it is opposite to how students spontaneously tend to look at and engage with the past. Several (intervention) studies have therefore examined effective strategies to foster aspects of students' historical thinking, such as causal reasoning (Stoel et al., 2017), historical contextualization and perspective-taking (Huijgen et al., 2018), critical source analysis (Nokes et al., 2007; Reisman, 2012), and historical writing (De La Paz, 2005; De La Paz & Felton, 2010; De La Paz et al., 2016; Van Drie, et al., 2015). These studies often rely on multiple documents based (inquiry) tasks and adopt principles such as explicit teaching, modelling and guided practice supported by scaffolds, in line with the model of cognitive apprenticeship (Collins et al., 1991).

In these studies, multiple documents based (inquiry) tasks are frequently used during the intervention and as part of the measurement instrument to examine changes in students' historical thinking. These inquiry tasks are usually evaluated using rubrics assessing various aspects of students' historical reasoning and writing (e.g. De La Paz et al., 2016; Sendur et al., 2021; Stoel et al., 2017). Aside from a few exceptions (e.g. Stoel, 2017) intervention studies typically do not include a detailed analysis of students' progress on the various aspects of these rubrics. In most cases, there is hence no in-depth information on the precise nature of students' progress on the multiple documents based (inquiry) tasks. Moreover, these studies usually report on the student sample as a whole and do not address (possible) differential effects of the intervention among students. Thus, while these studies provide important information about effective instructional practices

to foster progress in students' historical thinking, they do not offer as much information about how exactly such progress manifests itself (differently) in students' underlying thinking processes. These insights, however, are equally valuable and necessary in light of informing effective teaching practices.

To overcome this gap, this study builds on data collected in a previous intervention study (Wilke et al., 2022b) to explore the effects of an intervention on students' historical thinking in an in-depth way. Our intervention study was a cluster randomized control intervention study with 668 students (intervention: $n = 402$; control: $n = 266$). The study was based on a lesson series taking up 12-14 history classes aimed specifically at fostering students' historical inquiry competences, or: their ability to generate a substantiated answer to a historical question, based on a critical analysis of sources. The lesson series was based on inquiry tasks and design principles from cognitive apprenticeship (Collins et al., 1991). We examined the lesson series' effects on students' historical inquiry skills via a historical essay task. A first remarkable observation was that students differed significantly in their initial starting points for the historical essay tasks. A second was that, although the intervention was in general successful in improving students' historical inquiry skills, students did not progress equally: significant differences remained in place at the posttest level. Affective variables, procedural knowledge and epistemological beliefs did not provide a sufficient explanation for these large differences in students' scores.

The data from this study provide an excellent opportunity to obtain more insight into how precisely students' progress manifested itself in these historical inquiry tasks and to examine whether this progress differed among students according to their initial scores.

Via a qualitative study of a purposeful selection of students from the intervention group, this study therefore aims to explore students' progress on historical skills in a detailed way, considering differences in students' initial scores. That way, the study aspires not only to gain more insight into (the precise shape of) students' progress on historical skills, but also to reflect on practical implications for the classroom context, for instance, with regard to differentiation.

THEORETICAL FRAMEWORK

Multiple documents based (inquiry) tasks

Although there is no uniform operationalization of historical thinking, existing influential models (Seixas & Morton, 2013; van Drie & van Boxtel, 2008; Wineburg, 2001) aim, through an introduction of students to the methods of historians, to allow students to understand the interpretive and constructed nature of historical knowledge. Core aspects of historical thinking include asking historical questions, constructing and applying a historical frame of reference, critical source analysis, applying historical modes of reasoning (such as agency, causality, continuity and change) and constructing or critically analyzing historical representations.

Multiple documents based (inquiry) tasks are often used to foster and assess students' historical thinking. These tasks combine various aspects of historical thinking as they require students to critically evaluate and compare different sources and to provide arguments and evidence in support of a certain claim (De La Paz & Felton, 2010; Monte-Sano, 2010, 2016; van Drie & van Boxtel, 2008; Voet & De Wever, 2017b; Van Nieuwenhuyse, 2020). These tasks are considered authentic tasks for the discipline of history (van Boxtel & van Drie, 2018).

Providing a substantiated historical claim based on multiple documents requires students to do far more than synthesize the information in the different sources. Considering that not all historical sources are equally valuable, students need to apply discipline-specific heuristics or procedural knowledge to evaluate and compare them (Rouet & Britt, 1996). Wineburg's (1991) influential expert-novice study identified three key heuristics applied by expert historians when dealing with multiple historical documents. Source corroboration refers to comparing (differing accounts in) sources, looking for ways in which the sources confirm or contradict each other (Britt & Aglinskas, 2002; De La Paz & Felton, 2010; Wineburg; 1991). In order to weigh the evidence in (conflicting) sources against each other, experts also applied 'sourcing', i.e. examining the author of the source as well as when and where it was created (Wineburg, 1991). A final heuristic 'contextualization' refers to the need to pay attention to when and where the historical document originated (Wineburg, 1991). Over time, several scholars have elaborated upon these heuristics adding, for instance, 'close reading' referring to

the need to pay attention to aspects such as word choice and language in a document (Reisman, 2012). Scholars have also elaborated upon the heuristic of sourcing with additional criteria for assessing a source's value such as the documents and information on which a source is based, the goal and audience of a source and the need to evaluate a source's representativeness (e.g. Britt & Aglinskas, 2002; De La Paz & Felton, 2010; Hicks, Doolittle & Ewing, 2004; Perfetti et al., 1994; van Drie & van Boxtel, 2008; Van Nieuwenhuyse, 2020; Wilschut et al., 2012).

Scholars have argued that discipline-specific strategies are not only at play in the analysis of multiple documents, but also when formulating a substantiated answer, which is often done through writing. Disciplinary writing practices include presenting arguments, using evidence to support a claim and rebutting opposing evidence or claims (De La Paz & Felton, 2010; Monte-Sano, 2010; Nokes, 2017; Nokes & De La Paz, 2018). Monte-Sano (2010) identified five disciplinary characteristics in students' historical writing based on a multiple documents based task. First, factual and interpretive accuracy of the events and information described in the documents. Second, persuasiveness of evidence referring to the need to substantiate historical claims with relevant and convincing evidence based on the sources. Third, sourcing of evidence indicating that the essay contains references to the author of a document and the bias inherent in the source. Fourth, corroboration of evidence meaning the essays pay attention to the way that different sources work together to support a certain claim but also address counterevidence. Fifth, contextualization of evidence referring to the inclusion of contextual knowledge to evaluate the documents and to the essay's use of sources in a way that respects the contemporary meaning of the source. Monte-Sano (2010) demonstrates that while these characteristics will remain distinct among students with lower historical writing abilities, they will overlap when students become more proficient.

Students' performance on written inquiry tasks is related to their knowledge of historical text structures. Regarding causal historical writing, Coffin (2004; 2006), for instance, makes a distinction between accounts based on narrative (i.e. recording genres) and accounts based on analysis (i.e. explanatory and arguing genres). The former have a narrative structure, present a factual account and contain little argumentation or author voice. The latter are more

analytical and tend to reflect the interpretive nature of history for instance by providing arguments, comparing various causes and drawing evaluative conclusions. Building on Coffin's (2004; 2006) distinction in causal writing genres, Stoel (2017) established a relationship between students' essay structure and their historical reasoning reflected in an inquiry task. In an intervention study on students' causal historical writing he demonstrated that students' initial essay structure mediated the kinds of revisions students' made in their essays after the intervention. The intervention, however, had no effects on the structure of students' essays.

Students' historical thinking

Studies using (multiple) documents based tasks have demonstrated that, spontaneously, students struggle with the application of discipline specific ways of reasoning and writing. As students tend to approach sources as carriers of information, they are not inclined to corroborate or critically evaluate them (Britt & Aglinskas, 2002; Rouet et al., 1996; van der Eem et al., 2022; Wineburg, 1991). When students do evaluate the value of a source, they often do so in a superficial or inaccurate way, for instance by relying on the contents of a source, rather than on the contextual information accompanying it. Harris and colleagues (2016) found that students frequently relied on the content of a source to establish whether it was reliable, for instance referring to whether a source connected to students' prior knowledge or experience, whether it contained facts or statistics or even whether it 'looked official' or 'sounded true' (p. 114). When confronted with writing tasks based on multiple sources, students tend to focus on summarizing the content of sources rather than providing a substantiated claim (e.g. Stahl et al., 1996; Nokes, 2017; Young & Leinhardt, 1998). Students' historical writing is thus more likely to take the form of a 'factual' historical account (van Boxtel et al., 2021).

Although these studies have convincingly demonstrated that applying discipline-specific heuristics is difficult for students, they have also shown that students' performance on these multiple documents based tasks can differ strongly (e.g. Monte-Sano, 2010). Nokes (2017) for instance, described patterns in 8th graders' sourcing and argumentative writing demonstrating the wide variety of students' performance in this regard. Considering this variation, he argues that

teachers must adapt the design of assessments for students' historical thinking to students' incoming abilities.

Fostering students' historical thinking

In order to foster students' ability to provide a substantiated historical representation, intervention studies have successfully relied on design principles which aim to make disciplinary ways of reasoning visible to students and to provide them with support while applying them. In particular, studies have adopted design principles such as explicit teaching on discipline-specific heuristics, modelling, guided practice supported by scaffolding, feedback and (whole-class and peer-to-peer) interaction (Nokes & De La Paz, 2018; van Boxtel et al., 2021), in line with the general educational model of cognitive apprenticeship (Collins et al., 1991). De La Paz and Felton (2010), for instance, found that students who had received instruction on analyzing sources and on writing an argumentative essay created essays that were longer, were of higher quality, contained more advanced claims and rebuttals and cited more sources. Reisman (2012) studied the effects of document-based lessons including explicit teaching on disciplinary strategies and support for students through a process of modelling, guided and independent practice. This approach yielded positive effects on students' evaluation of single sources, yet not on intertextual strategies, such as source corroboration. In explaining these results, Reisman (2012) pointed to the way that these intertextual strategies were approached during the intervention but also hypothesized that perhaps these strategies are more sophisticated as they require a connection between several documents. Positive results of the instructional practices described above have been found across various studies (e.g. De La Paz, 2005; De La Paz et al., 2016; Huijgen et al., 2018; Monte-Sano & De La Paz, 2012; Nokes et al., 2007; Reisman, 2012; Stoel et al., 2017). Several of them have demonstrated that these positive results apply to students regardless of their incoming abilities (e.g. De La Paz & Felton, 2010; De La Paz et al., 2016; Reisman, 2012). These studies, however, provide little insight into how precisely students progress on these inquiry tasks and how the intervention affected students (differently) according to their incoming abilities.

In our own intervention study (Wilke et al., 2022b), we similarly established that a lesson series based on inquiry tasks and design principles derived

from the model of cognitive apprenticeship was effective in improving students' ability to provide a substantiated answer to a historical question. However, we noted significant differences in students' initial level of historical inquiry skills and found that the intervention did not yield the same effect on all students. Based on existing studies as well as our own study, a number of questions remain unanswered. First how precisely do students progress in terms of historical thinking. Do students advance evenly across various aspects of historical thinking, or are some aspects of historical thinking easier to acquire? Second, what is the role of students' initial historical thinking ability in this regard? For instance, does students' initial ability influence how the progress on historical thinking takes shape? Overall, little is known about how interventions aimed at fostering students' historical thinking affect students on an individual level.

The current study

Building further on the historical essay tasks gathered in our previous (intervention) study, this study explores the progress of historical skills of a specific group of students on an individual level. In particular it examines how precisely students' progress on historical inquiry skills takes shape among those students who progressed substantially, and how this progress differs depending on students' starting position. To this end, a purposeful sample of students is selected from the intervention group; and their essay tasks are analyzed qualitatively. For this group of students, the following twofold research question is explored: How does progress on students' historical inquiry skills (a) look like among students who have progressed considerably and (b) differ depending on students' starting positions?

METHOD

Research context and participants

This study was conducted in the 12th grade of general secondary education in Flanders (Belgium). History education is a mandatory subject taking up two hours a week. In the 12th grade, history teachers have both a master's degree (most often in history) and a teacher degree. However, their familiarity with (teaching) historical thinking varies as there is no mandatory professionalization and teacher training courses' attention to this concept has only become more prominent in recent years. Moreover, the Flemish history standards provide teachers with a lot of freedom and only mention historical thinking in an implicit way, providing no concrete guidelines on how to address it in practice (Van Nieuwenhuyse, 2020). This particular context has created a large diversity in teachers' practices regarding historical thinking. Overall, however, these practices are characterized by a focus on substantive knowledge about the past and only limited attention to second-order and procedural knowledge. This is most clearly illustrated by teachers' use of sources which are most often used in an illustrative way or to obtain substantive knowledge (Van Nieuwenhuyse et al., 2017). Even when teachers evaluate the value of sources, they regularly do so in a superficial way and sometimes depart (in line with history textbooks) from a simplistic distinction between objective and subjective sources (Wilke & Depaepe, 2019). In general, teachers are also not inclined to organize full historical inquiries based on multiple sources in their practice and have a limited understanding of what these entail (Voet & De Wever, 2016; 2017).

Between 2019-2025 a curriculum reform is gradually being implemented in Flemish history education which puts forward historical thinking as a central goal. In this context, 40 teachers in 32 schools were recruited to participate in the intervention study. Schools were assigned randomly to a control (n = 266) or intervention group (n = 402). The intervention study (see Wilke et al., 2022) was conducted in the 12th (last) stage of secondary education in which the curriculum reform had not yet been implemented. This study only examines students in the intervention group. Students in this group had a mean age of 17.04 (*SD* 0.39), 38.2% identified as male, 61.6% as female and 0.2% as other.

Procedure and design of the experimental condition[2]

The experimental condition

Students in the intervention group participated in a lesson series of 12-14 history lessons. The lessons were centered around the theme of decolonization after 1945 and were aimed at fostering students' historical inquiry skills. Design principles for the lesson series were derived from the model of cognitive apprenticeship (Collins et al., 1991) and applied to the specific context of history education via existing literature on the promotion of historical thinking.

The lesson series made use of four inquiry based writing tasks, constituting authentic, discipline-specific tasks (van Boxtel & van Drie, 2018). They were centered around an evaluative historical question (van Drie et al., 2006) to which students had to provide a substantiated answer based on a critical analysis of sources. The provided sources differed in terms of perspective towards the historical question and in terms of value. Each source was accompanied by background information which students could use to evaluate its value. Students were taught how to provide a substantiated historical representation in response to these historical questions. A high quality substantiated representation was taught to contain the following elements: a clear response to the historical question (claim), arguments derived from the sources to support the stance, references to the sources in support of the claim, an evaluation of the value of the provided sources in order to argue why they considered the supporting sources to be valuable and to rebut counterarguments based on sources that are less valuable.

Throughout the lesson series, teachers first modelled an inquiry task. They modelled the reasoning process and steps required for the critical analysis of the sources and for the composition of a (written) substantiated stance. In the subsequent lessons, students worked on other inquiry tasks in groups, supported by scaffolds which faded out gradually. Scaffolds (based on those of the University of Michigan, 2021 and Monte-Sano et al., 2014) were provided for each step in the process: corroborating and evaluating sources, weighing evidence and arguments

[2] This section describes the data gathering conducted during the intervention study, which provides the basis for this study. Some overlap may therefore occur between this study and the one described in Wilke et al., 2022b.

and formulating a substantiated answer to the historical question. Teachers coached students during the inquiry tasks, among others by providing feedback on the inquiry tasks and by adjusting the level of scaffolding to their students' needs.

Explicit teaching on second-order and procedural knowledge related to how to construct a substantiated answer to a historical question and related to critical sources analysis was also an essential element throughout the lesson series. Teachers provided, for instance, explicit teaching on the meaning of sources' reliability and representativeness and on how to evaluate it (e.g. by examining the author, the goal and audience of the source, the time in which the source was created) and on the essential elements and structure of a substantiated answer.

To facilitate historical reasoning, students worked in groups on the inquiry tasks. Furthermore, whole-class discussion were used to discuss students' findings and provide additional opportunities to provide feedback and advance students' reasoning (Reisman, 2012; Stoel et al., 2017; van Boxtel & van Drie, 2013; 2017).

Procedure

The study was conducted between January and April 2021. Students completed a pre- and posttest each taking up two lessons (2x50 minutes). Teachers in the intervention group received a training and were monitored during the intervention to ensure treatment fidelity. Online questionnaires were collected after each lesson and two interviews were conducted with each participating teacher to discuss teachers' experiences and possible deviations from the provided lesson plans. Teachers who had skipped one of the inquiry tasks or had to delay the administering of the posttest (due to corona measures in Flemish education) were excluded from the study. Students who had missed more than two classes were also excluded ($n = 12$).

Measurement instrument

To measure students' historical inquiry skills, the pretest and posttest contained a historical essay task. Each task contained a short introduction, a historical question and a set of sources, containing different arguments and differing in perspective and value. Students were required to write a 15-20 sentence essay providing a substantiated answer to these questions, based on a critical analysis of the sources.

The historical essay tasks centered around the questions (1) whether the police at the Sharpeville massacre (1960) had a legitimate reason to use violence and (2) whether Dutch violence during the Indonesian Independence War (1945-1949) was structural or incidental. Students were provided with excerpts from both primary and secondary sources, presenting different views and arguments regarding the subject and accompanied with background information needed to evaluate the source's value (e.g. author, time, etc.). Each task contained five sources which were selected to be similar in type (primary-secondary, textual-visual) and difficulty level between the pretest and posttest. An overview of the essay tasks is presented in Table 1.

Table 1. Overview of historical essay tasks at pretest and posttest

Historical essay task			
Pretest Police violence at the Sharpeville massacre		Posttest Excessive violence during the Indonesian Independence War	
Sources	Content	Sources	Content
A. Anti-Apartheid journalist's report on the events at Sharpeville (1960). B. Picture taken at the Sharpeville demonstration by an anti-apartheid photographer (1960). C. Excerpt from the court trial, interrogating the commanding officer about the events at the demonstration (1960). D. Official statement issued by the South Africa High Commissioner in London, shortly after the events and after strong criticism from the international community (1960). E. Excerpt from a historian's book on the events at Sharpeville (2011).	A. The police started shooting without any warning and without any provocation. Only three policemen got hurt, while 200 native people were hurt by police violence. The journalist states not to have seen any weapons among the demonstrators. B. Picture showing demonstrators running away (unarmed) from the police. C. The commanding officer reports the demonstrators were hostile and was expecting an assault. After a gunshot was heard, the policemen started shooting. He suggests that when demonstrators were hurt in the ricocheting bullets. D. The police was attacked by the demonstrators and acted in self-defense. E. Only a few of the demonstrators were armed and when a shot was fired from the crowd, the police started shooting. They fired in two rounds, including one round affecting the demonstrators running away from the field.	A. Excerpt from a television interview (1969) with a soldier claiming that excessive violence was used structurally during the war. B. Picture (1947) of Dutch soldiers who had taken Indonesian soldiers captive. C. Letter from a veteran corporal during the Indonesian war in response to the television interview. D. Excerpt from a government report (1969) investigating the nature of the violence during the war. The report had to be finished withing three months. E. Statement from the Dutch government declaring the start of a new investigation into the existence and nature of excessive violence (2016).	A. Excessive violence was conducted structurally. The soldier had unsuccessfully approached various media outlets immediately after the war to tell his story. B. The picture was deliberately kept from the Dutch general public during the war in an attempt to shape their perception of the war. C. The television interview was a lie. Violence is part of any war, but there was no excessive violence during this war. D. Dutch soldiers behaved correctly during the war; there were no cover-ups of excessive violence and there even exist accounts of positive acts committed by Dutch soldiers in Indonesia. E. After the publication of new historical research about excessive violence during the war, the government issues a new investigation.

Students' essays were evaluated with a rubric containing six criteria. The initial rubric was based on research by Monte-Sano (2012); Monte-Sano and De La Paz (2012), and Monte-Sano, De La Paz and Felton (2014) but adapted to better correspond with the explicit teaching in the lesson series. It was further adjusted after a pilot study in two classes ($n = 36$), which also evaluated the difficulty and feasibility of the essay tasks. The final rubric contained the following criteria: (1) making a claim in response to the historical question; (2) argumentation in support of the claim: this criterion evaluated the amount of arguments that were used to support the essay's claim; (3) quality of the argumentation, assessing the quality of students' arguments in support of their claim; (4) use of sources as supporting evidence: this criterion assessed the extent to which students' had referenced specific sources in support of their claim; (5) evaluation of the value of the supporting sources, indicating to what extent students demonstrated that they had evaluated the value (in terms of reliability and/or representativeness) of the arguments they used in support of their claim and (6) engagement with counter arguments and evidence: this criterion indicated to what extent students had acknowledged the existence of arguments and evidence that went against their claim and whether they had provided suitable rebuttals.

For each criterion, four competency levels were distinguished ranging from poor (1) to excellent (4) resulting in a total score on 24. Ten percent of pretest essays was scored by two raters and intraclass correlation (2-way, absolute agreement, single measures) was calculated to establish inter-rater reliability. This yielded at least a good agreement (>.6) for each separate criterion (ICC ranged from .64 to 1; $M = .82$, $SD = .12$) and an excellent agreement (>.75) for the total score (ICC ranged from .75 to .97; $M = .87$, $SD = .08$) (Cicchetti, 1994; Hallgren, 2012).

Data sampling

In line with the methodology as applied by Stoel (2017), this study builds on the essay data collected as part of the intervention study and uses purposeful sampling to study changes in students' essays in an in-depth manner. In our previous study, we established large differences in students' initial essay scores as well as in the progress they made. These differences are illustrated in the descriptive results in Table 2.

Table 2. Descriptive results on historical essay at pre- and posttest in the intervention group (N = 402).

Historical essay task (on 24)		
Pretest *M* (*SD*)	Posttest *M* (*SD*)	Change *M* (*SD*)
15.20 (3.58)	18.14 (4.45)	2.95 (4.93)

We opted for a critical case purposeful sampling (Patton, 2015) as we were particularly interested in those students for whom the intervention was very effective, i.e. students who had made substantial progress on historical inquiry skills. This was operationalized as an increase in students' historical essay tasks of at least six points. Students with a six point increase between pre- and posttest progressed at least two times more than average and the increase allows progression to be noticed on each of the rubric's criteria. In light of the research question, we then opted to divide students within this sample into two homogenous groups according to their pretest score on historical inquiry as we were interested in differential effects of the intervention according to students' starting positions. We therefore distinguished between students with high versus low initial scores, whereby a pretest score of 12 points or more (out of 24) was considered a high starting point.

This initially yielded a sample of 112 students who had made at least six points progress on the historical essay tasks. Students who had obtained a pretest score of 19 or more on this task (n = 65) had been excluded as they would have been unable to progress six points. Of the 112 students, 45 had a low starting position and 67 had a high starting position. We selected a subsample of both groups for further qualitative analysis: students with a pretest score of 8 and 9 (n = 19) and those with a pretest score of 15 and 16 (n = 26). These subsamples were selected because these students' essays were not the weakest nor the very best, still had room for improvement, yet differed significantly (6 to 8 points) from each other.

Data analysis

First, descriptive quantitative results of the obtained samples, based on the evaluation rubric's score, were explored to examine general patterns in students' essays and differences according to their starting positions. These findings were then further examined through a qualitative analysis. The qualitative analysis was conducted using Nvivo and was an iterative process departing from an a priori set of codes, stemming from the evaluation rubric, supplemented by additional codes emerging from the essay tasks. The initial coding scheme was developed by the first author and reviewed and discussed with the second and third author until a final coding scheme was established, including clear descriptions for each of the codes.

The predefined set of codes was derived from the criteria in the evaluation rubric and contained five main themes: claim; argumentation (incl. quality of argumentation); use of sources as evidence; evaluation of sources' value (i.e. reliability and/or representativeness); and rebuttal of counter arguments and evidence. For each of these main themes, subcodes were created allowing to analyze students' essay tasks in a fine-grained way.

'Claim' contained subcodes to indicate whether the essay contained an answer to the historical question, whether this answer was in line with the question, and whether the claim was explicitly or implicitly put forward. For the main code 'argumentation', subcodes were created to distinguish between students' own arguments and those derived from the sources, to examine the quality of the arguments (high-mediocre-low) and to indicate whether they were used implicitly or explicitly. 'Use of sources as evidence' evaluated whether the essay contained references to specific sources and whether they were used explicitly in support of the essay's stance. 'Evaluation of sources' value' contained subcodes indicating whether students merely mentioned information about the source, such as who the author was (i.e. superficial attention to the value of a source) or whether they also evaluated the source's value. On occasions where students evaluated the value of a source, the quality of the evaluation (high-mediocre-low) was coded as well as which criteria they had used to do so (e.g. author, date, goal) and whether this was done in an implicit or explicit way (i.e. explicit mentioning of reliability, value, representativeness, etc.). The theme 'engagement with counter arguments and evidence' contained subcodes to

indicate whether students had acknowledged the existence of counter arguments or evidence, whether they had engaged with them and how this was done (superficially or by evaluating the sources' value). If the engagement with counter argumentation included an evaluation of the sources' value, subcodes indicated whether this was done implicitly or explicitly, the quality of the evaluation (high-mediocre-low) and which criteria were used.

During the coding process, some new codes were created stemming from the data. For instance regarding the evaluation of a source's value, a code 'nuanced evaluation' was added to highlight instances where students had evaluated a source's value in a nuanced way, paying attention both to aspects that could increase and to aspects that could decrease the source's value.

Besides the five main coding themes stemming from the evaluation rubric, an additional coding theme was established to describe various types of students' essays as it became clear that these took on distinctive forms and transformed between pretest and posttest. Based on students' essays, we distinguished four main codes for this theme, describing the main focus of students' writing: descriptive, evaluative, argumentative and argumentative-evaluative depending on the essay's main focus. A descriptive essay's main focus was on describing what happened and what the various sources mentioned about the event. An argumentative essay's main focus was on arguing a certain answer to the historical question by use of arguments and/or evidence (sources) but without an evaluation of the sources. Evaluative essays were mainly centered around the evaluation of the sources but lacked a connection to the essay's claim. Finally, an argumentative-evaluative essay did both. It substantiated a certain answer to the historical question with arguments and evidence (sources) and included an evaluation of the sources.

As the coding for the five main themes derived from the evaluation rubric stayed close to the initial rubric for which an acceptable interrater agreement had been reached, no additional measures for interrater agreement were calculated for the coding related to these themes. For the additional coding theme 'essay type' the initial coding by the first author was reviewed and discussed with the third author until consensual agreement was reached.

RESULTS

In order to examine changes in the historical inquiry tasks among students who had made a lot of progress, the quantitative results based on the evaluation rubric were first examined and compared for students with a low and high starting position. Table 3 shows for both groups their main total score on the historical essay task at pretest and posttest and their mean score on each criterion of the evaluation rubric.

Table 3. Mean score on students historical essay tasks at pretest and posttest, based on the evaluation rubric.

	Low pretest score on historical essay task (n = 19)		
	Pretest	Posttest	Change
Criterion	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Claim	1.79 (1.10)	3.95 (0.23)	2.16 (1.12)
Argumentation	1.05 (0.23)	3.37 (0.76)	2.32 (0.82)
Quality of argumentation	1.00 (0.00)	3.47 (0.61)	2.47 (0.61)
Use of sources as evidence	1.00 (0.00)	3.74 (0.65)	2.74 (0.65)
Evaluation of sources' values	1.84 (0.61)	2.68 (0.95)	0.84 (1.12)
Engagement with counter-arguments and evidence	1.58 (0.51)	2.47 (0.84)	0.89 (0.99)
Total score	8.26 (0.45)	19.68 (2.56)	11.42 (2.61)
	High pretest score on historical essay task (n = 26)		
	Pretest	Posttest	Change
Criterion	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Claim	3.23 (0.77)	3.92 (0.39)	0.69 (0.74)
Argumentation	2.73 (0.78)	3.77 (0.43)	1.04 (0.92)
Quality of argumentation	3.50 (0.65)	3.85 (0.37)	0.35 (0.80)
Use of sources as evidence	2.81 (0.94)	4.00 (0.00)	1.19 (0.94)
Evaluation of sources' values	1.73 (0.45)	3.58 (0.50)	1.85 (0.78)
Engagement with counter-arguments and evidence	1.50 (0.51)	3.38 (0.57)	1.88 (0.77)
Total score	15.50 (0.51)	22.50 (1.03)	7.00 (0.94)

Based on Table 3, a number of general trends can be distinguished in students' progress on historical inquiry skills. At pretest, students with a high pretest score achieved remarkably higher scores on the criteria of claim, argumentation, quality of argumentation and evidence. On the criteria of evaluating the value of sources and engaging with counterarguments and evidence, both groups performed more or less equally low at pretest.

Both groups progressed on all six criteria between pretest and posttest. At posttest, students with a low pretest score seemed to catch up in terms of claim, argumentation, quality of argumentation and evidence as the differences between both groups decreased. Regarding evaluating a source's value and dialogue, however, differences between both groups increased. Whereas students in both groups obtained more or less equal scores on these criteria at pretest, students with a high initial score clearly outperformed those with a low initial score at posttest.

While both groups progressed on all criteria, this progress was unevenly distributed among the various criteria and manifested itself differently in each group. In the group with a low pretest score, progress appears to have happened predominantly in terms of stance, argumentation, quality of argumentations and use of evidence. Progress also occurred in the other criteria but was far less outspoken. In the high pretest group, strong progress was made with regard to use of evidence, evaluating a source's value and engagement with counter arguments.

The results of the qualitative analysis allow for an exploration of how precisely this progress manifested itself in the essay tasks and how it differed between both groups. For both groups, the frequency of codes within each main coding theme is depicted in Table 4. These results will be discussed for each of the main coding themes, first for the group with a low pretest score on historical skills, followed by the group with a high pretest score.

Table 4. Frequency of coding and number of essays in which each code was used.

	Students with a low initial score on historical skills $n = 19$				Students with a high initial score on historical skills $n = 26$			
	PRETEST		POSTTEST		PRETEST		POSTTEST	
	Files	Frequency	Files	Frequency	Files	Frequency	Files	Frequency
Claim								
None	6		0		0		0	
Explicit	3		18		14		26	
Implicit	6		1		11		0	
Partial	4		0		1		0	
Argumentation								
Own argument	4	4	0	0	3	4	0	0
Explicit argument derived from sources	5	7	12	21	17	27	21	45
Implicit argument derived from sources	6	8	7	9	14	18	10	18
Quality								
High	6	11	15	24	25	41	26	59
Mediocre	4	4	4	4	3	3	3	3
Low	0	0	2	2	1	1	1	1
Use of sources as evidence								
Vague source reference	2	2	1	1	4	8	1	1
Source reference	11	25	7	10	14	38	8	12
Source reference in support of the claim	6	10	19	36	14	22	25	53

Table 4. (*continued*)

	Students with a low initial score on historical skills *n* = 19				Students with a high initial score on historical skills *n* = 26			
	PRETEST		POSTTEST		PRETEST		POSTTEST	
	Files	Frequency	Files	Frequency	Files	Frequency	Files	Frequency
Evaluation of sources' value								
No evaluation	7	8	14	16	14	22	12	13
Superficial evaluation	8	12	3	3	15	26	4	4
Evaluation of a source's value								
Form								
Explicit evaluation	4	8	14	20	6	8	23	42
Implicit evaluation	5	7	7	7	2	4	6	6
Quality								
High	2	2	7	8	2	2	16	22
Mediocre	6	10	9	11	5	9	18	23
Low	3	3	8	8	1	1	3	3
Nuance	0	0	6	6	0	0	10	11
Criterion								
Author	9	13	6	6	6	13	6	10
Societal context	0	0	1	1	1	1	9	11
Goal and audience	1	1	2	2	3	3	6	6
Information	5	7	11	15	8	9	23	33
Representativeness	0	0	5	8	0	0	10	15
Time and place	2	3	3	3	2	2	11	12
None	0	0	1	1	0	0	0	0

Table 4. (*continued*)

	Students with a low initial score on historical skills *n* = 19				Students with a high initial score on historical skills *n* = 26			
	PRETEST		POSTTEST		PRETEST		POSTTEST	
	Files	Frequency	Files	Frequency	Files	Frequency	Files	Frequency
Engagement with counterarguments and evidence								
Mentioned without dialogue	6	7	1	1	11	12	1	1
Superficial dialogue	4	5	4	5	9	11	1	1
Rebuttal								
Form								
Explicit evaluation	0	0	10	16	1	1	23	35
Implicit evaluation	1	1	9	11	0	0	10	12
Quality								
High	0	0	2	2	0	0	2	2
Mediocre	1	1	12	16	1	1	12	14
Low	0	0	7	9	0	0	22	31
Criterion								
Author	1	1	7	9	1	1	18	22
Goal and audience	0	0	6	6	0	0	13	17
Information	0	0	14	17	0	0	25	25
Representativeness	0	0	4	5	0	0	13	16
Time and place	0	0	0	0	0	0	4	4
Societal context	0	0	0	0	0	0	1	1
None	0	0	0	0	0	0	0	0

Table 4. (*continued*)

	Students with a low initial score on historical skills *n* = 19				Students with a high initial score on historical skills *n* = 26			
	PRETEST		POSTTEST		PRETEST		POSTTEST	
	Files	Frequency	Files	Frequency	Files	Frequency	Files	Frequency
Essay type								
Descriptive	14		0		11		0	
Evaluative	3		6		2		8	
Argumentative	1		4		13		2	
Argumentative-evaluative	1		9		0		16	

Progress on historical inquiry skills among students with a low initial score

Claim

Students at pretest had great difficulty in providing an answer to the historical question. Only three essays included an explicit answer to the historical question. Other students either only referred to the historical question implicitly (6); partially (4) or not at all (6). At posttest all but one of the essays contained an explicit claim in response to the historical question. Students thus exhibited great progress on this aspect of the essay tasks.

In their pretest essay, student B3 for instance provided an elaborate report on the events at the Sharpeville demonstration, concluding the essay with "we conclude that the Apartheid regime was a very dark chapter in the history of colonialization", which did not provide an answer to the historical question at hand. At posttest, this student's essay contained a very different type of conclusion: 'After a considered evaluation and source corroboration, I came to the following conclusion: The violence of the Dutch military was structural in nature'.

Argumentation

At pretest, several difficulties can be observed with regard to students' use of arguments. First, some of the students' essays (8) did not contain any argument derived from the sources. Second, students did not always base their arguments on the information provided by sources but provided arguments based on their own ideas. Regarding the police violence at the Sharpeville demonstration, student B10, for instance, explained why the police did not have a legitimate reason to use violence using arguments that were not derived from the sources:

> But still I stand by the fact that what those police officers did could have been avoided and that the decision to fire dozens of shots and to injure 200 people should not and can never be approved. Those protestors already had fewer benefits and already started the protest from a disadvantaged position. They were therefore frustrated, but an escalated protest should not resemble a scene from the First World War! If I study the sources further, there are lies and falsehoods to be found along both sides of the story, but the most can be found at the side of the policemen.

> It can be said that those policemen were not happy with the [protest's] turnout and because it got out of hand, they were very angry and heard a shot which made them think, subconsciously, that it was best to shoot at those, in their eyes, filthy people. This is all due to the Apartheid regime which made those police officers look at those black people as inferior.

When students supported their answer with their own arguments, this sometimes reflected presentism. For instance, student B18 argued against the police violence by explicitly referring to contemporary regulations for police enforcement:

> According to Belgian law, a policeman must avoid using his gun at all costs. It also states that an officer must first fire a warning shot before firing a bullet. Finally, the force used must always be proportionate to the situation and the crime. So the police should only fire when it is unavoidable. We can apply these standards to the situation in Sharpeville. (...) By today's standards, the police officer must himself be in danger, in mortal danger even, before he or she shoots. Whether that was the case in South Africa I leave in the open, as I have too few sources to make a judgment. If the indigenous masses actually became aggressive, perhaps a better solution would have been to arrest the offenders with or without force.

Both students hence argued, not based on the information provided by the sources, but rather based on contemporary arguments or their own views on the issue. A third issue regarding students' argumentation at pretest was that in cases where arguments derived from the sources were used, half of them was only mentioned implicitly (eight out of 15) although they were usually of high quality (11 out of 15).

Between pre- and posttest students made a lot of progress regarding argumentation. Whereas at pretest 10 essays contained not a single argument derived from the sources, all but three of the essays at posttest contained at least one and students no longer used arguments based on their own views. The number of arguments derived from the sources doubled (from 15 to 30) and students connected the arguments more explicitly to their answer (21 out of 30 were used explicitly). At posttest, student B10 for instance, substantiated his

answer explicitly with arguments derived from the sources, rather than provide his own explanation for the events:

> I am convinced that these [violent acts] are rather of a structural nature because of the following arguments. First, these atrocities committed by Dutch soldiers have been confirmed by the study of Dr. Limpach (see source 5). This is an extensive study on the use of excessive violence by soldiers of both the Dutch and Indonesian side of these acts during the war of independence. In addition, Dr. Limpach also mentions that these Dutch atrocities were part of a larger scheme, of a larger structural nature.

The overall increase in the number of arguments and in the number of arguments used explicitly in support of a claim show that, after the intervention, students in general appeared to have gained more awareness of the need to (explicitly) substantiate their claim with arguments derived from the sources.

Use of sources

Regarding the use of sources, the majority of students referred to specific sources during their pretest essays. However, in most cases, students merely mentioned sources but did not explicitly put forward these sources as evidence to support their claim (25 out of 35 sources were not connected to the essay's claim). At posttest, students increasingly referenced sources in their essays (from 35 at pretest to 46 at posttest). Significant progress was also made in the way that students used these sources. At posttest, most of the sources that were mentioned were connected explicitly to students' claim and hence, used as evidence in support of this claim (36 out of 46 sources were used explicitly). Whereas at pretest only six essays used specific sources to support their claim, all posttest essays contained at least one source that was explicitly used as such.

This change is apparent in the excerpts from student B10 (see earlier). At pretest, this student only vaguely referred to the sources ("When I study the sources further, there are lies and falsehoods to be found along both sides of the story, but most of them are found in those of the police officer") whereas at posttest, this student connected specific sources to his essay's claim.

Evaluating the sources' value

At pretest, students' attention to the value of sources was rather poor. For the majority of sources, students either did not mention anything that may have influenced the sources' value (eight times), or merely mentioned something about the source, such as the author, but did not connect it to an evaluation of the source's value (12 times). Student B13 for instance mentioned important aspects of the sources, but did not relate this to the source's value:

> From source 3 we can infer that civilians were hit in the back, we can corroborate this from source 2. Source 3, a lieutenant corporal in charge of the police officers present [at the demonstration], states that a police officer was abruptly pushed backwards by the crowd. This, in turn, is confirmed by source 5, the historian.

On 15 occasions, students evaluated the value of a source. This evaluation, however, was only done explicitly about half the time (8 out of 15) and the quality of students' reasoning was mostly low (3) or mediocre (10). This was most often due to simplistic reasoning, based on a naïve distinction between 'objective' and 'subjective' sources. Other common issues with regard to students' evaluation of the value of sources were unsubstantiated evaluations of a source's value, or reasoning errors. Student B9, for instance, mentioned in the pretest 'The picture in source 4 shows the harsh reality in the Dutch East Indies. Because it is a photograph, this source is reliable'. Another example of a reasoning error can be found in essay B12 where the student claimed the following: "In 2011 an excerpt from the book by the historian Tom Lodge was published. In it he tells of a private attack by someone in the crowd, which has never been effectively proven. In addition, this was not published until 2011, which makes this source less reliable."

A similar reasoning error was apparent in essay B13 when the student evaluated the statement of the Dutch government declaring the start of a new investigation into excessive violence during the war.

> In the last source, the Dutch government decides to open another investigation following the publication of a book by Rémy Limpach who also states that the violence was structural in nature. This confirms again that there is sufficient evidence to start an investigation. It is again confirmed that there are indications of structural violence during the

> decolonization of the Dutch East Indies. However, it must be noted that this source was created in 2019, 65 years after the facts. This again makes it less reliable due to the large time difference.

While these students clearly tried to evaluate the sources' value by reflecting on the timeframe between the historical event and the source, their reasoning is in this case rather poor.

At posttest, students' attention to the value of sources advanced in parallel with the increased use of sources as evidence. While students still regularly neglected to evaluate the value of a source (16 times) or did so only superficially (3 times), the majority of sources at posttest was subjected to an evaluation of their value indicating an increased attention to the need to evaluate the value of sources. Compared to the pretest, students increasingly evaluated the sources' value explicitly (20 out of 27 times) and this was done in a more accurate way. Whereas at pretest students referred mostly to 'objectivity' or 'subjectivity', students at posttest referred to the more accurate concepts of reliability and representativeness. Here, a clear influence can be distinguished of the intervention which paid explicit attention to these second-order concepts.

While the quality of students' reasoning about the value of sources remained somewhat the same, varying between low (8), mediocre (11) and high (8), students did exhibit a broader understanding of the concept of reliability. At posttest, five students evaluated the representativeness of a source and six students included a nuanced evaluation of a source's value, with attention both to aspects that increased the sources' value and aspects which may decrease it. Student B11 wrote at posttest about the television interview with Joop Hueting:

> You could say that the source is unreliable because it was written 20 years later. But Joop Hueting was Dutch himself and confessed to the terrible deeds carried out by the Dutch. He also had to go through a lot of trouble to be able to tell his story, because many newspapers had rejected him.

Overall, students at posttest demonstrated increased and more explicit attention to the value of sources. This progress, however, should not be overstated as students still did not systematically evaluate the sources' value and the quality of students' evaluations still varied strongly.

Engagement with counter arguments and evidence

Students' pretest essays were characterized by a lack of thorough attention to counter arguments and rebuttals. In the rare cases (four essays) where some attention to counter arguments was present, this only happened in a superficial way. Student B11 for instance mentioned that the sources disagreed with regard to the question whether the demonstrators throwed rocks at the police but offered no explanation as to why these sources' accounts may have differed, nor on which of the sources she deemed more reliable: "In response, the police fired bullets at the crowd. According to source 4, the police fired in self-defense. I can understand that if everyone had been throwing rocks at them, but that was not the case according to sources 1 and 5".

At posttest, most essays exhibited some attention to the existence of a counter argument and in line with students' increased explicit attention to evaluating the value of sources, these rebuttals were more often linked explicitly to the source's value. On 26 occasions did students attempt to rebut a counterargument by comparing the value of the sources. These evaluations, however, were usually mediocre (16) or poor (2) in quality although eight students managed to demonstrate high quality reasoning in this regard.

Essay type

Pretest essays predominantly took on a descriptive focus (14) with students describing the events, based on information they had extracted from the provided sources. Between pretest and posttest, strong shifts occurred in the nature of students' essays as these were no longer descriptive at posttest. Rather, students' essays became evaluative (6), argumentative (4) or, in the majority of cases (9), both. This indicates an increased attention to substantiating a claim and evaluating the value of sources, which is in line with changes that were established in the previous coding themes.

Progress on historical inquiry skills among students with a high initial score

Claim

Concerning the statement provided in students' essay tasks, all essays at pretest contained an answer to the historical question, although 11 out of 26 did so implicitly. At posttest, all essays contained an explicit answer to the historical question.

Argumentation

Students at pretest already preformed relatively well regarding argumentation. Their essays contained arguments which were mainly of high quality (41 out of 45) and the majority of which was used explicitly in support of the claim (27 out of 45 were used explicitly).

At posttest, students' argumentation improved further on a number of aspects: students provided even more high-quality arguments (on average from 1.7 arguments at pretest to 2.4 at posttest) which were predominantly connected explicitly to students' claim (45 out of 63). Despite these positive results, however, ten essays still contained at least one argument which was not connected to the essay's claim.

Use of sources

At pretest, students' essays contained a total of 60 references to specific sources. However, four essays contained only vague references to the sources and the majority of source references (38 out of 60) was not connected to students' claim.

At posttest, the number of specific source references remained more or less stable (from 60 to 65), yet the most striking progress was how these were used in the essays. At posttest the vast majority of sources (53 out of 65) were referenced specifically in support of the answer that was argued. This progress, however, did not occur across all essays. In some cases, students' use of sources became less instead of more explicit. These students had pretest essays which contained several source references in support of their claim, yet at posttest their use of sources became less explicitly connected to their essay's stance. Eight posttest essays

predominantly contained references to sources which were not used in support of the stance.

Evaluating the sources' value

Students' pretest essay overall showed only low levels of attention to the value of sources. Students either did not include any information that may be relevant to the source's value (22 times), or only mentioned it superficially (26 times). In some essays, however, students did already go beyond the superficial level and included an evaluation of the value of a source. This was done mostly in an explicit way (8 out of 12 times) and based on mediocre reasoning.

Between pretest and posttest, significant progress could be observed both in students' attention to the value of sources and in the quality of their reasoning. Regarding the attention for the value of sources, students at posttest were more inclined to include an evaluation of the value of the source. Whereas at pretest, students only evaluated the value of a source 12 times, this happened 48 times at posttest. In almost all instances (42 out of 48 times) students explicitly referred to the notion of reliability or representativeness when evaluating the value of a source. The quality of students' reasoning also improved, with 22 out of 48 evaluations being of high quality (22). A high-quality reasoning is illustrated in the posttest of student A8. This student explicitly evaluated the value of several sources, including the television interview of Hueting about which she wrote:

> This concerns an interview with a veteran who was at the scene himself and is therefore an eyewitness and even participated in the extreme violence himself. This source is therefore highly reliable. However, this source alone does not give a complete picture because it is only about the experiences of one person in the area where he fought. There is also another war veteran who contradicts this in a reader's letter.

Students, however, still struggled with the evaluation of the value of sources and encountered similar difficulties as students with a low pretest score. Student A18's reasoning, for instance, depicted a naïve, rather simplistic evaluation of the value of the television interview of Joop Hueting. She stated:

> After the testimony of an ex-soldier Joop Hueting in 1969 it became clear that the violence occurred on a regular basis. He described how he himself

> saw that militarily unnecessary actions were carried out, such as the assault on the *kampongs* (local villages) and revenge actions. Because Joop was an eyewitness we can conclude that the source is reliable.

Whether the author was an eyewitness or not is definitely relevant in evaluating the source's value but does not by itself guarantee that a source is reliable, especially considering that the source set contained another eyewitness who provided an entirely opposite view. Another student (A19) inaccurately used the criterion of 'time' to question the reliability of a historian's work. She argued that "This source is very reliable because the research deals specifically with this issue. However, we must take into account that the book was written sixty years after the date of the event".

Despite the variance in the quality of students' reasoning about the value of sources, students did apply a more diverse set of criteria to evaluate the sources and showed an increased attention for the representativeness of sources (15 times) and for nuanced evaluations (11 times).

Engagement with counter arguments and evidence

Pretest essays demonstrated only limited attention to counterarguments with counterarguments being mentioned (11 times), but rarely rebutted (only once at pretest). This changed drastically in the posttest where students exhibited a strongly increased attention to counterarguments (47 times). In these instances, students not only mentioned the existence of a counterargument or evidence but rebutted it by evaluating the various sources' value. On 35 occasions, this was done explicitly and on 12 occasions implicitly. Not only did students' attention to counterarguments and evidence increase, their reasoning at posttest was mostly of high quality (31 out of 47 times). In her essay about the Indonesian Independence War, student A4, for example, wrote:

> In contrast to this study, the *Excessennota*, a study commissioned by the government, states: "The government regrets that excesses occurred, but it maintains its view that the armed forces as a whole behaved correctly in Indonesia. The collected data confirm that systematic brutality did not occur". However, this does not strike me as a reliable source because it was written at the behest of the government, and they may indeed have

> an interest in covering up the situation. In addition, it was not a thorough study because it had to be completed in a short period of time.

This student explicitly rebutted the counterargument provided by the government report by accurately arguing why this source was less reliable than the other provided sources.

When rebutting counter arguments by referring to the value of sources, the same difficulties sometimes occurred as described earlier regarding the evaluation of the value of sources. Students sometimes expressed rather simplistic reasoning, or their reasoning was not well supported. In both cases, their reasoning was categorized as 'mediocre quality'.

Essay type

Pretest essays of students with a high pretest score were most likely to take on a descriptive (11) or an argumentative (13) focus. At posttest, no descriptive essays were identified. Rather students' essays had shifted predominantly to an evaluative (8) or an argumentative and evaluative focus (16), indicating an increased attention to evaluating the value of sources.

CONCLUSION AND DISCUSSION

Building on data collected in a large-scale intervention study, this research aimed to explore differential effects of an intervention of students' historical inquiry skills according to their initial starting positions. The qualitative analysis of essays of a purposeful sample of students who had progressed substantially after the intervention revealed remarkable differences between students with high and low pretest scores. Although both groups had made progress on all criteria evaluated in the rubric, both groups differed in terms of where this progress was most outspoken.

At pretest, students with low pretest scores typically struggled with the need to formulate an explicit answer to the historical question and substantiate it with arguments and evidence. Their pretest essay either did not contain a claim or their claim was only weakly substantiated with evidence and arguments from the sources. Progress in this group was situated mostly with regard to these aspects. Between pretest and posttest these students overall seemed to become more aware

of the need to make a claim and support it, using the available sources. Besides these notable changes, there also appeared to be an emerging recognition among most students of the need to address the sources' value and engage with counterarguments. Students, however, did so only sporadically and when they did, their reasoning was often not of high quality. These changes are reflected in the changing essay structures where a shift occurred away from descriptive essays which lacked a substantiated stance. Instead, students wrote essays which were centered around substantiating a stance, evaluating sources or both.

Students with a high pretest score predominantly wrote pretest essays in which they took a certain position towards the historical question and substantiated it with arguments and sources. This, however, did not always happen in an explicit way. At pretest, students' essays exhibited a limited attention to the value of sources, but usually in a superficial way. The same can be said for their engagement with counterarguments, which occurred in a limited and usually superficial way. In the posttest, these students made especially strong progress with regard to evaluating the value of sources and engaging with counterarguments. This was done more consistently and explicitly and based on high quality reasoning. They hence demonstrated an increased awareness of the need to argue why they considered some sources to be more valuable than others, as well as an increased ability to do so. With regard to argumentation and using sources as evidence, the change between pretest and posttest among these students was more ambiguous. Although there was an overall increase in the number of arguments and source references, these were in some cases not clearly connected to the position taken. Students hence not always progressed in a straightforward way across all criteria. An unwanted side-effect of students' increased attention to the value of sources appeared to be that students failed to use the content of the sources to provide arguments for their answer to the historical question and overlooked the need make a clear connection between the sources they discussed, and the position taken. In some cases, students' increased focus on evaluating the value of sources thus seemed to go at the expense of the argumentative nature of the essay. Rather than adding the evaluation of sources to their already argumentative essays, these students shifted from an argumentative to an evaluative essay type.

The qualitative analysis of the essay tasks allowed for a much more detailed mapping of changes in students' pretest and posttest essays and, in so doing, provided insight into patterns of progress. These analyses show that there seems to be some sort of sequence in students' development with certain aspects of historical inquiry only developing fully when students are better acquainted with other aspects. In particular, providing a claim, arguments and evidence in support of the claim seem to be the basic aspects of historical inquiry that students master first. Only then do students seem to progress in more advanced aspects of historical inquiry, namely critically evaluating sources and engaging with counter arguments and evidence. This is evidenced by the differences in progress made by students with high and low initial scores. These clearly show that progression on the more advanced aspects only occurred substantially among those students whose pretest essays showed that they already mastered to a great extent the more basic elements of historical inquiry.

While the analyses show a certain progression in students' historical inquiry skills on the separate criteria of the evaluation rubric and qualitative coding scheme, they also reveal that progress on the various aspects of historical inquiry is not separated from each other. It seems that as students become more proficient in these (written) inquiry tasks, the progress that students make on one aspect of the inquiry task is inextricably linked to progress on other aspects. For instance, students' increased attention to substantiating a claim resulted in improvements with regard to argumentation, which became more elaborate and more extensively based on the provided sources, as well as with regard to the use of evidence. This is in line with Monte-Sano's (2010) claim that characteristics of students' historical writing will overlap at higher performance levels. Yet, this connection did not exist between all aspects of students' writing but rather among the more basic aspects and the more advanced.

The student essays also make clear that historical thinking remains an unnatural act and demonstrates in particular how difficult certain discipline-specific ways of reasoning are for students. Most notably and in line with earlier studies (e.g., Harris et al, 2016; Rouet et al, 1996; van der Eem et al.; 2022) evaluating sources and engaging with counter arguments and evidence remained a challenge, even among those students for whom the intervention had worked particularly well.

As in Stoel's (2017) research we observed a relationship between text structure and certain characteristics of students' historical writing. In our analysis, however, four, rather than two, essay types emerged. In contrast to Stoel's (2017) research, we observed strong shifts in these text structures between pretest and posttest as a result of the intervention. After the intervention, students' essays more often adopted a text structure that paid more attention to the evaluation of sources. It appears that the intervention's explicit attention to procedural knowledge on how to formulate a substantiated answer to a historical question based on sources affected students' essays. These evolutions nevertheless demonstrate that progress on historical inquiry is not a strictly linear process. Changes in the essay types showed that in some cases students' increased attention to the value of sources decreased their attention to other aspects, such as argumentation and using evidence. In particular, evaluative essays appeared to function in some cases as an intermediate position when students evolved from a descriptive essay to a truly argumentative-evaluative essay.

An important limitation of this study is that it is based on a purposeful sample of precisely those students who had progressed substantially on historical inquiry. This sample allowed us to closely examine how such progress looked like and which role students' starting position played in this regard but provides no insight into the way in which progress manifested itself in pupils for whom the intervention yielded a more moderate effect. The study is therefore limited in terms of its generalizability. In addition, the study provides no explanation as to why the intervention worked particularly well for these students. The use of written inquiry tasks constituted another limitation. While disciplinary writing in history has distinct characteristics (e.g., Monte-Sano, 2010), students' historical writing is also connected to their general literacy skills (e.g., Reisman, 2012) and students are not always able to accurately demonstrate their historical reasoning in a written task (Stoel et al, 2015; van Drie et al., 2013).

Despite these limitations, the study provides some important implications for history education and future studies as it reveals patterns in students' progress in historical inquiry and offers possibilities for differentiating according to students' starting positions. Based on our findings, it seems beneficial to focus first on the argumentative elements of a written historical inquiry tasks (i.e., claim, arguments and evidence) before adding the need to include the evaluative

elements (i.e., source evaluation and rebuttal of counter arguments and evidence). In line with Nokes (2017) we also recommend that assessments of students' historical thinking as well as teachers' instructional practices should be adapted to these differences in students' initial abilities which may be present within a single group of students. Students with lower incoming abilities may need to focus predominantly on substantiating their claim, while students who have mastered these aspects may be encouraged to provide more attention to accurately evaluating the value of sources and engaging with counter arguments. Several intervention studies, including our own, have demonstrated the benefits of using scaffolds when asking students to engage with documents based tasks (e.g., De La Paz & Felton, 2010; De La Paz et al., 2016). These scaffolds can be used as a tool differentiation in the classroom, when adapted to students' individual needs.

For the reporting of future intervention studies, this study demonstrates the value of including a more in-depth analysis and description of the effects of the designed intervention. It also highlights the importance to pay attention to ingroup differences in intervention studies, especially in light of their practical implications for the instructional practice.

Main coding themes and subcodes	Description	Example
CLAIM (POSITION STATEMENT)		
No statement	The essay does not contain a statement or claim in response to the historical/topical question	There has never been any clarity on who shot first and whether this was a matter of police brutality or pure self-defense. The police violence was both legitimate and illegitimate.
Explicit statement	The essay explicitly puts forward a clear statement or claim in response to the historical/topical question	Based on the analysis of these sources, I conclude that the acts of violence committed by Dutch troops in Indonesia were structural acts.
Implicit statement	The essay contains a statement or claim in response to the historical/topical question, but is not explicitly stated as such.	In my opinion, this was a misunderstanding where there were bad apples in both groups. However, the police should have given a warning and stopped shooting once the crowd fled. I can well imagine that the police panicked because of this attack. But shooting at innocent people is never the solution, I think.
Partial statement	The essay contains a statement or claim that is only partly in response to the historical/topical question (e.g. not entirely in line with the question)	What exactly do I think about it? Well, I haven't been able to do enough research yet, but I think it's easy to conclude that killing people, for whatever cause and certainly in their backs, is something terrible and you should be punished for it.
ARGUMENTATION IN SUPPORT OF THE CLAIM		
Own argumentation	The students provides an argument in support of their answer to the historical question but this is not based on the sources	In addition, a lot of innocent people were injured. Because these people were fleeing, they were shot in the back. If a demonstration undertaken by a civil rights organization itself violates the civil rights of others, I personally think that strict intervention is allowed. I think the police felt threatened by the large crowd, so they reacted in an extreme way. They killed everyone because they were afraid.
Argument derived from the sources		
Form_implicit Form_explicit	Implicit: the argument is mentioned, but not clearly presented in support of the claim Explicit: the argument is clearly put forward in support of the claim (e.g. by use of phrases such as 'a first reason is that' or 'therefore')	*Implicit* One may ask: being pelted with some stones and one protester drawing a weapon, is this a good reason to start shooting at the protesters? Looking at the interview with the war veteran Joop Hueting in 1969, this man talks about the atrocities that he has seen happen. Namely: villages that were ruined, horrible interrogations that took place without military necessity. *Explicit* Another argument for systematic violence is the interview with Joop Hueting in 1969. In it he describes that that he and his fellow soldiers did many horrible things. It was normal, he says. A photograph taken in Malang in East Java in 1947 shows us how Dutch soldiers held Indonesian civilians at gunpoint with dead people lying next to them. This terrible image was seen as an undesirable one by the Dutch government and thus was not shown in the Dutch press during the period in which the troops were active in the Dutch East Indies. From this we can conclude that unnecessary violence was used and that the Dutch government tried to hide this from the population. Moreover, no warning shots were given and the police were pelted with stones (see sources 1 and 3).

Main coding themes and subcodes	Description	Example
Argument derived from the sources (*continued*)		
Quality_high Quality_mediocre Quality_low	High: the argument is derived from the sources, based on an accurate understanding of the source(s), supports the claim and is accurately/clearly presented/explained Mediocre: the argument is derived from the sources and supports the claim but is not clearly presented/accurately explained Low: the argument is derived from the sources but based on an inaccurate understanding of the source and/or does not support the claim	*High* In 2016, the government launched a new investigation in response to a book that had been published. This book relies on Limpach's research. He confirmed that the violence was structural in nature and the new investigation shows that the government itself has its doubts about the real nature of the violence. According to journalist Humprey Tyler, the crowd was not warned to disperse. The police accused the crowd of being heavily armed, however, the journalist claims that he did not see these heavy weapons. *Mediocre* I think the biggest problem here is with the mentality of the police before the demonstration and the violence. The police were not only prepared for violence, they expected violence (source 3). *Low* It is noteworthy, that source 5 also came about after an accusation towards the Dutch. A comprehensive investigation, conducted by Limpach, once again states that the Dutch violence was of a structural nature. In 2016, the Dutch government insists on conducting a second investigation, in order to prove their innocence after all. *(Student misinterpreted the source).*
USE OF SOURCES		
Source reference	A specific source is mentioned (e.g. by referencing the source's number or author), but not connected to the claim	Historian Tom Lodge describes the situation from a historical perspective. He states that the police shot at fleeing people, which can also be proven by the many bullet wounds in the back.
Source reference in support of the claim	A specific source is mentioned (e.g. by referencing the source's number or author), and presented in support of the claim	A first reason is that extreme violence was the normal course of events and was therefore a structural fact. This is said in the television interview with Joop Hueting (source 1). A second reason *(for the students' claim)* is that the Dutch government wanted to cover up the incidents. I established this through the 'unwanted' photograph that did later get exhibited in the Resistance Museum in Amsterdam. This also showed how the Dutch government manipulated the people by misrepresenting the events and making them look better.
Vague source reference	Sources are mentioned in general, without reference to one or more specific sources.	Several sources indicate that the violence was illegitimate and that the police abused their power.
EVALUATION OF THE SUPPORTING SOURCES' VALUE, IN TERMS OF THEIR RELIABILITY AND/OR REPRESENTATIVENESS		
No evaluation of a source's value	No mention of any element or aspect of the source which may influence its value	Of these *[protestors]*, many were unarmed. A few, however, had brought firearms (source 5), which eventually caused a misunderstanding and thus led to the massacre. The picture of Ian Berry confirms this, we see the crowd running away. Also we see none of the so-called heavy weapons.
Superficial evaluation of a source's value	One or more relevant aspects of a source is mentioned (e.g. author, date, goal etc. of the source) but this is not connected to an evaluation of its value.	Humphrey Tyler, a reporter, and Ian Berry, a photographer, both worked for Drum Magazine. This magazine expressed their support for the anti-Apartheid movement. The two men were present at the Sharpeville demonstration. The historian states that the demonstrators did start firing first but that the police reacted much too violently.

Main coding themes and subcodes	Description	Example
EVALUATION OF THE SUPPORTING SOURCES' VALUE, IN TERMS OF THEIR RELIABILITY AND/OR REPRESENTATIVENESS (*continued*)		
Evaluation of a source's value	The student evaluates the value of a source/explains why they consider the source to be valuable	
Quality_high Quality_mediocre Quality_low Quality_nuanced (additional/double coding)	High: Reasoning regarding a source's value makes sense/is accurate and clear Mediocre: Reasoning regarding a source's value touches upon a relevant aspects but is not entirely clear or accurate, for instance in the case of oversimplification Low: Reasoning regarding a source's value is missing or completely inaccurate. Nuance: Reasoning is presented both in favor of the source's value, but also against.	*High* A study by Dr. Limpach shows that the violence was indeed structural in nature. Dr. Limpach is himself a historian and has done extensive research to reach this conclusion, which leads me to believe that this source is reliable. *Mediocre* One can assume that the statement of the South Africa High Commissioner is subjective. The people in this commission were white and therefore had no idea what it is like to be black, in 1960 there was also very much racism. Because of this we see that the media manipulated the Dutch people by showing them only the pictures they wanted. This source is very reliable as it is an original photo from the war. Because Joop was an eyewitness we can conclude that the source is reliable. *Low* In this research under supervision of dr. Limpach, it was shown that unnecessary violence was indeed used on a structural level. Although dr. Limpach is an educated man we can question the reliability of this source as the research was carried out long after the facts. Source 4 shows a photo of Dutch soldiers standing next to wounded and deceased Indonesian soldiers. This is a reliable source because photoshop did not exist at that time and it reflects reality very well. *Nuance* I consider this to be a reliable source since he himself was present in the Dutch East Indies as a soldier, and therefore experienced the facts himself. One could say that he perhaps does not tell the facts 100% accurate as there is a 20 year time span between the incidents and the interview. In my opinion, this source is a reliable source to answer this question because his goal is to talk about his experiences, he also tried to tell his story in the 1950s but was always rejected. However, we do not know whether Hueting's experiences are representative of all the other soldiers since no other accounts were given.
Criterion_author Criterion_goal Criterion_audience Criterion_societal context Criterion_information Criterion_time/place Criterion_representativeness	Specification of which aspect of the source is mentioned in relation to its value	*Time, information, goal:* Hueting's testimony is indeed reliable. He was present at the moment itself (although this was 20 years ago, so this is not a sufficient condition). In addition, his goal appears to be simply to tell the truth to the Dutch population. He has nothing to gain from this interview, does not defend himself and only tells what he saw.

Main coding themes and subcodes	Description	Example
Evaluation of a source's value (*continued*)		
Form_implicit Form_explicit	Implicit: the student evaluates the source's value but does not explicitly mention this by referring to notions such as reliability, value, representativeness, neutrality, objectivity, etc. Explicit: the student makes explicit that they are evaluating the source's value, (e.g. by referring to reliability, representativeness, neutrality or subjectivity of the source)	*Implicit* Joop Hueting talked about the war crimes in an interview and indicated that these were not just a few incidents but were daily occurrences. He was a veteran of the war so an eyewitness. *Explicit* We can assume that Joop Hueting is a reliable source as the man was an eyewitness during the war. One can assume that the statement of the South Africa High Commissioner is subjective. The people in this commission were white and therefore had no idea what it is like to be black, in 1960 there was also very much racism.
ENGAGEMENT WITH COUNTER ARGUMENTS AND EVIDENCE		
Mentioned without dialogue	The essay contains a possible counterargument or source, but this is not presented as such (no confrontation with other sources or arguments), or the student only refutes it based on their own input.	Investigations have shown that a large number of people were shot in the back. According to Pienaar, this had two reasons: ricocheting bullets may have hit the protesters in the back. Also, some of them may have been hit in the back while fleeing. I find the first reason a bit unlikely.
Superficial dialogue	The existence of counterarguments or sources is mentioned, but they are not refuted.	According to the police, their unit was pelted by a large number of stones, and the crowd of protesters was also armed with heavy weapons. Humphrey on the other hand had not seen any weapons himself. Sources 2 and 3 contradict my position and thus state that everything was done correctly and the Dutch are not to blame.
Rebuttal	The existence of a counterargument or source is mentioned and refuted by comparing the sources' value	
Quality_high Quality_mediocre Quality_low	High: Reasoning regarding differences between the sources' value makes sense/is accurate and clear Mediocre: Reasoning regarding differences between the sources' value touches upon a relevant aspects but is not entirely clear or accurate, for instance in the case of oversimplification Low: Reasoning regarding differences between the sources' value is missing or completely inaccurate.	*High* In the reader's letter from a former veteran we can read that he accuses Hueting of telling incorrect facts. To his knowledge no prisoners of war were ever killed, they were always released. At first sight this seems to be a reliable source, he has seen the facts happening just like Hueting and is therefore an eyewitness. But we may assume that after the television interview with Hueting, the authorities from this period will have received much criticism. So there is a possibility that this Corporal just wanted to clear his name. It is also possible that both Hueting and the Corporal were active in a different area. And thus that they did not witness the same acts of violence. *Mediocre* Another source on this subject is the 1969 government report. However, I do not find this reliable (…) They did not have the time to properly investigate both sides of the story and therefore I do not find the source representative. *Low* Source 3 is a reader's letter. This is probably just an old man wanting to clear his conscience. Thus, not so reliable.

Main coding themes and subcodes	Description	Example
Rebuttal (*continued*)		
Criterion_author Criterion_goal Criterion_audience Criterion_societal context Criterion_information Criterion_time Criterion_place Criterion_representativeness Criterion_medium	Specification of which aspect of the source is mentioned in relation to its value	*See earlier*
Form_implicit Form_explicit	Implicit: the student evaluates the source's value to rebut the counter argument but does not explicitly mention this by referring to notions such as reliability, value, representativeness, neutrality, objectivity, etc. Explicit: the student makes explicit that they are evaluating the source's value to rebut the counterargument, (e.g. by referring to reliability, representativeness, neutrality or subjectivity of the source)	*Implicit* A possible counter-argument could be that a study was already issued by the Dutch government in 1969. However, this study was only allowed to last three months (which is not enough time to comprehensively investigate everything). In addition, it came from the government itself. They could present the facts in a more positive way in order to present themselves in a more favorable way. *Explicit* J.G. Schuitema, a corporal from 1946 to 1948 in Central Java, issued a reader's letter eight days after the television interview, in which he writes that what Joop Hueting said is a big lie. Personally, I think this is a less reliable source, since he was a corporal and thus in charge of several soldiers. I think that with this letter he mainly wanted to save his own skin.
ESSAY TYPE		
Descriptive	The essay's main focus is on describing what happened or summarizing the content of the various sources. The essay provides a report of the events, as described in the sources. The essay may contain a claim in response to the historical question but no arguments or sources are clearly put forward to support the claim.	
Argumentative	The essay's main focus is on arguing why a certain claim has been made. The essays provides a claim and substantiates it by providing arguments and/or by referring to sources that support the claim. Arguments may be derived from the sources, but can also be based on students' own ideas. There is no attention for the value of sources.	
Evaluative	The essay's main focus is on (systematically) evaluating the different sources that were provided. Although the essay may include a claim, there is no/very little attention to (explicitly) substantiating the claim with arguments or sources (lack of connection between the claim and the sources).	
Argumentative-evaluative	The essay both substantiates a claim with arguments and sources *and* demonstrates attention to the value of sources. The essay contains arguments and source references in support of a claim *and* argues why certain sources may be considered more/less valuable in light of the historical/topical question.	

Chapter 7

Exploring the possibility of transfer from historical to contemporary inquiry tasks. A qualitative analysis of the effects of an intervention in 12th grade history education.[1]

Several history education scholars argue that historical thinking may not only increase students' familiarity with the practice of historians but will also contribute to skills that are relevant for democratic citizenship. In particular, fostering students' historical inquiry skills is expected to improve their ability to critically evaluate contemporary sources, weigh evidence and substantiate a position towards public debates. These scholars assume that students will spontaneously transfer skills and procedural strategies acquired via historical inquiries to contemporary contexts. However, such a 'far transfer' from one context to another is difficult to obtain; also, there is little research examining whether such a transfer from historical inquiry skills to democratic skills is possible. This study builds on data stemming from an intervention study examining the effects of a lesson series specifically aimed at fostering students' historical inquiry competences on (a) students' historical inquiry skills and (b) democratic skills, using historical and topical essays based on multiple documents tasks. It explores in a qualitative way changes in students' historical and topical essay tasks to explore whether and on which particular aspects these tasks potentially showed signs of a transfer from historical to democratic skills. To this end, a purposeful sample of students is selected who had progressed substantially on both tasks. The study shows limited evidence for a transfer from historical to democratic skills, among others in students' use of sources and attention for their reliability. It discusses how this transfer manifests itself and reflects on the implications for our understanding of the relationship between historical thinking and democratic citizenship.

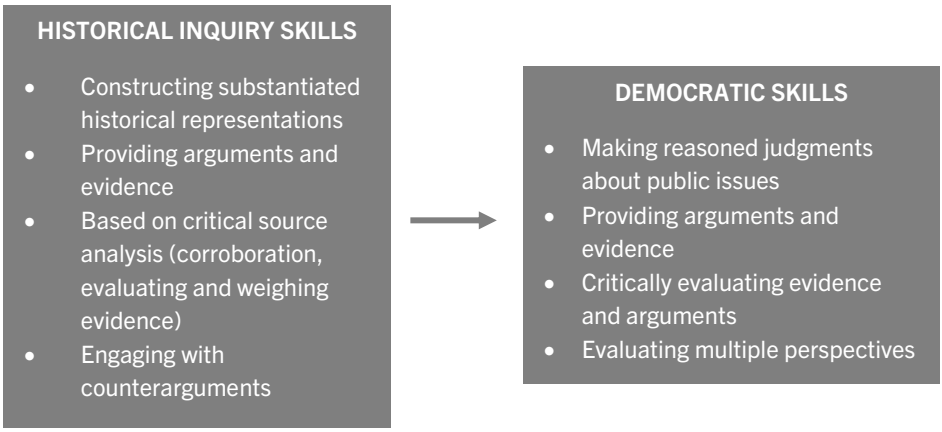
[1] This chapter is based on: Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K.. Exploring the possibility of transfer from historical to contemporary inquiry tasks. A qualitative analysis of the effects of an intervention in 12th grade history education. *Manuscript under review.*

INTRODUCTION

With the introduction of historical thinking as its primary goal, history education has moved closer to history as an academic discipline. Historical thinking not only instills students with knowledge about the past, but also with insight into the methods that historians use to acquire such knowledge (Lee & Ashby, 2000). While historical thinking is first and foremost a way to meet history education's disciplinary goal of introducing students to the practice of historians, several scholars are convinced that it may also contribute to civic goals. This potential contribution of history education to citizenship is connected to various aspects of historical thinking. For instance, analyzing agency, or "the ability to act on decisions in order to bring about desired goals" (Barton, 2012, p. 131), is expected to help students understand their own role in generating change in society and to shape civic attitudes (Barton, 2012; Seixas, 2012). Developing students' historical empathy, i.e., understanding and connecting to the perspective of people in the past, is assumed to stimulate them to recognize and value the existence of different perspectives in society, to foster a willingness to change their own values, attitudes and/or behavior and even to inspire them to prevent similar wrongdoings in present-day society (Barton & Levstik, 2004; Endacott & Brooks, 2013).

Another way in which history education is expected to contribute to democratic citizenship is through the promotion of historical inquiry skills. Scholars have argued that historical inquiries foster certain skills which may spontaneously transfer to contemporary settings and may hence contribute to the development of democratic citizenship (e.g., Lee & Shemilt, 2007; Lévesque, 2009; McCully, 2012; Monte-Sano, 2016; Nokes, 2013; Reisman, 2012; Stearns, 1998; Wineburg & Reisman, 2015). Figure 1 depicts the presumed relationship between these historical inquiry skills and a set of skills which are relevant for democratic citizenship. In particular, the above-mentioned scholars argue that by engaging students in historical inquiries, students will become better at making reasoned judgments about public debates on societal issues, based on arguments and evidence, a critical evaluation of contemporary sources (evidence) and arguments and an evaluation of different perspectives on the issue. In what follows, we will refer to this set of skills as 'democratic skills'.

Figure 1. The presumed contribution of historical inquiry skills to democratic skills



This assumption may, however, be questioned from a general educational perspective as the presumed contribution of historical thinking to democratic citizenship would require students to apply skills acquired in the initial context (historical) to another (contemporary) context. Such a 'far transfer', from one context to another, is not easily obtained (Perkins & Salomon, 1994). To date, research examining the possibility of transfer from historical thinking to democratic citizenship remains scarce.

Rather, research so far has focused predominantly on how to foster historical thinking among students, as this is a challenging goal in itself. Historical thinking is considered an 'unnatural act' (Wineburg, 2001) that runs counter to the way that students spontaneously look at history. Instead of looking at the past from contemporary perspectives and considering sources to be direct reflections of the past, historical thinking requires students to study the past in its own right and to understand how historians construct the past by interpreting and analyzing sources (Wineburg, 2001). While several studies have managed to improve aspects of students' historical thinking (e.g., De La Paz, 2005; De La Paz & Felton, 2010; De La Paz et al., 2016; Huijgen et al., 2018; Nokes et al., 2007; Reisman, 2012; Stoel, van Drie & van Boxtel, 2017), very few studies have examined whether these skills do indeed transfer to contemporary settings and hence, contribute to democratic citizenship. Thus, little is known about whether a transfer from historical to democratic skills is possible and how such a transfer may look like.

In a previous study (Wilke et al., 2022b) we set up a cluster randomized controlled intervention study to examine the disciplinary effects of a lesson series aimed at fostering students' historical thinking and to examine whether the acquired skills would indeed spontaneously transfer to contemporary contexts. We examined the effects on a lesson series spanning 12-14 history lessons on students' historical inquiry skills via a historical essay task. In line with previous intervention studies (e.g., De La Paz et al., 2016; Reisman, 2012), the study adopted principles from the model of cognitive apprenticeship (Collins et al., 1991), including modeling, scaffolding coaching and interaction, to foster students' historical inquiry skills. To this end, the lesson series included several multiple documents based essay tasks in which students critically analyzed multiple historical documents in order to provide a substantiated answer to a historical question. A possible transfer to the democratic skills as described above (see figure 1) was examined via a topical essay task, designed in parallel to the historical task. In particular, this task examined whether students improved their ability to take up a reasoned position towards a public debate based on a critical evaluation of contemporary sources, a weighing of the available evidence and a consideration of different perspectives. The intervention was successful in improving students' historical inquiry skills but did not in itself significantly predict students' democratic skills at posttest. However, it became clear that students' historical inquiry skills and democratic skills covaried and that in some cases, students' historical and topical essay tasks improved simultaneously after the intervention. This result may hence indicate that some degree of transfer might have occurred but does not offer a clear-cut answer as to the direction of the relationship between students' historical inquiry skills and democratic skills.

The results of this study raise questions regarding how precisely students' progress on the historical and topical essay tasks looked like and whether changes in students' essay tasks show signs of a transfer from historical to democratic skills. Analyzing these tasks in a more detailed way may provide more insight into whether and how such a transfer can occur, and precisely which aspects of historical inquiry may and may not transfer to contemporary contexts. To this end, this study qualitatively analyzes the essays tasks of a purposefully selected group of students from the intervention group. In particular, students are selected who

progressed substantially on both the historical and topical essay tasks and therefore showed the most promising signs of a potential transfer.

THEORETICAL FRAMEWORK

Historical inquiry skills and democratic citizenship

Several frameworks for historical thinking have been developed, describing what it means to introduce students to the practice of historians (e.g., Seixas & Morton, 2013; Seixas, 2017; van Drie & van Boxtel, 2008; Wineburg, 2001; Van Nieuwenhuyse, 2020). These frameworks, although differing in their concrete operationalizations, often include the need for students to generate their own historical representations, based on a critical analysis of historical sources.

Multiple-documents based or inquiry tasks (both terms are used interchangeably, see van Boxtel et al., 2021) are often used to foster and evaluate these skills among students. These tasks are based on a historical question and a number of (conflicting) historical documents which students need to critically analyze and corroborate in order to provide a substantiated answer to the question (van Boxtel et al., 2021; Voet & De Wever, 2017b). Such tasks cultivate students' mastery of discipline-specific procedures regarding source analysis and the construction of a substantiated answer. The procedures include, for instance, comparing (conflicting) accounts in sources (i.e., source corroboration) and evaluating their reliability, among others by taking into account the author's background and motives, the time and place of the source's creation and its representativeness (i.e., sourcing) (De La Paz & Felton, 2010; Hicks, Doolittle & Ewing, 2004; Wineburg, 1991). Students should also consider the time and place in which a source was created (i.e., contextualization) and pay close attention to the author's word choice and language (i.e., close reading) (Reisman, 2012; Wineburg, 1991). In substantiating an answer, students need to provide arguments, use evidence and rebut counterarguments and evidence (De La Paz & Felton, 2010; Monte-Sano, 2012, 2016; Nokes & De La Paz, 2018; Reisman, 2012; van Drie & van Boxtel, 2008; Van Nieuwenhuyse, 2020; Voet & De Wever, 2017b; Wineburg, 1991). These multiple documents inquiry tasks touch upon the very core of the historical discipline, bringing together various aspects of historical thinking.

The historical inquiry skills promoted through these tasks are not only considered essential for obtaining history education's disciplinary goals but are also expected to be beneficial for democratic citizenship. In particular, they are presumed to influence students' critical engagement with contemporary sources of information and societal debates. Monte-Sano (2016), for instance, argues that students learn citizenship skills by learning to construct and critique historical arguments as "attending to evidence that support an argument is key to understanding public issues" and that "thinking about the credibility of evidence and where evidence comes from can help citizens make stronger arguments" (pp. 316-317). Barton and Levstik (2004) state that engaging students in historical inquiries (incl. an examination of evidence) can prepare them to make reasoned judgments, move beyond their own perspective and take part in collaborative discourses about public issues. This connection between historical inquiry skills and democratic skills is supported by various other scholars who argue that fostering students' ability to provide a substantiated answer to a historical question based on a critical analysis of sources, will improve their ability to make reasoned judgments about public issues, based on a critical evaluation of contemporary sources and to provide better substantiated arguments (Barton & Levstik, 2004; Cinnamon et al., 2021; McCully, 2012; Monte-Sano, 2016; Nokes, 2013; Slater, 1995; Stearns, 1998; Wineburg & Reisman, 2015). Research on civic reasoning indeed presents some clear similarities to historical inquiry skills. For instance, in order to effectively evaluate (online) information students should, among others, examine aspects such as the author of the source, the author's motivations and the evidence presented and should compare the source's information to that in other sources (McGrew et al., 2018; Wineburg & McGrew, 2019). These aspects strongly overlap with discipline-specific heuristics applied to historical sources (e.g., Wineburg, 1991).

This potential contribution of historical thinking to democratic citizenship is based on the idea that students will spontaneously transfer skills and knowledge acquired through historical inquiries and apply them to contemporary sources of information and topics. If students are expected to apply skills and procedures acquired in the historical context to a novel, contemporary context, this requires a 'far transfer' (Perkins & Salomon, 1994). However, such a transfer, from one context to another, is not easily obtained and is unlikely to occur spontaneously,

as this type of transfer requires specific instructional practices that trigger active, deep processing. A transfer of learning can only happen if students recognize that certain previously acquired knowledge is relevant in a new context, for instance, because they see similarities between the tasks; if students are able to recall the relevant knowledge, and if they are able to apply it to the new context (Barnett & Ceci, 2002; van Peppen et al., 2021). Yet, scholars arguing for a transfer from historical to democratic skills rarely acknowledge this need for specific (instructional) conditions, implying that such as transfer may occur spontaneously. Another reason to question whether a transfer from historical to democratic skills can indeed be expected stems from the research of Wineburg and McGrew (2019). In comparing historians' and professional fact checkers' strategies for evaluating online information Wineburg and McGrew (2019) established that historians used different, less effective strategies than professional fact checkers. They showed that historians were often misguided by website features such as an official-looking logos or domain names. Historians also tended to stay within a website when judging its reliability whereas fact-checkers opened additional websites to gather more information to assess the initial website's credibility (i.e., 'lateral reading'). Although this study was only directed at online sources, it does raise the question whether heuristics applied to historical sources can be applied to contemporary contexts in a straightforward way. Despite these difficulties and the scarcity of research on a potential relationship between historical thinking and democratic citizenship there is also some evidence to suggest that a transfer from a historical to a contemporary context may indeed be possible (see further).

Research on historical inquiry skills and on their potential transfer to democratic skills

Several studies have examined how students engage with (multiple) historical sources. These have shown that students experience great difficulty with the application of discipline-specific procedures related to analyzing sources and constructing historical representations. Spontaneously, students are more likely to summarize rather than evaluate sources as they tend to take sources at face value. When they do evaluate the value of sources, they often express superficial or inaccurate reasoning (e.g., Britt & Aglinskas, 2002; Harris et al., 2016; Nokes, 2017; Rouet et al., 1996; Stahl et al., 1996; van der Eem et al., 2022; Wineburg,

1991; Young & Leinhardt, 1998. For instance, Britt and Aglinskas (2002) established that high school and college students spontaneously paid little attention to source information but rather focused on the content of a source. Nokes' (2017) large-scale study of eight-graders historical writing based on multiple historical documents showed vast differences between students' attention to sourcing with over 40% of students paying no attention to the reliability of sources, even when prompted to do so. Students also expressed difficulty in reasoning about the reliability of sources in an accurate way. They judged documents based on their attention to detail, their clarity or the amount of information present, or dismissed sources based on potential biases or when an author expressed uncertainty. Similarly, Harris and colleagues (2016) found that high school students struggled to evaluate the reliability of sources. Besides basing their judgments on source information, the students also assessed the reliability of sources based on their content (e.g., whether it connected to students' prior knowledge, or the types of details present in the document) or even based on emotional responses.

In exploring how students engage with multiple documents (inquiry) tasks, scholars have described how students create task and activity models which contain a representation of both the end goal of a task and of the activities needed to achieve it (Britt & Rouet, 2012; Rouet et al., 2017; Wiley et al., 2020). These representations influence how students approach a task and which activities are evoked. Hence, in order for students to engage in historical inquiry tasks in an advanced way, they need to be aware of what such tasks require, i.e., applying discipline-specific procedures, such as sourcing, contextualizing and corroboration to analyze and compare sources in order to generate a substantiated answer.

Intervention studies using multiple document-based inquiry tasks have yielded positive results from design principles that make these activities explicit to students and provide them with support while applying them. In particular, instructional strategies from the model of cognitive apprenticeship (Collins et al., 1991) have proven their effectiveness. These include explicit teaching on discipline-specific procedures, guided practice supported by coaching and scaffolds and interaction in the form of group work and whole-class discussions (e.g., De La Paz, 2005; De La Paz et al., 2016; De La Paz & Felton, 2010; Freedman,

2015; Huijgen et al., 2018; Nokes et al., 2007; Reisman, 2012; Sendur et al., 2021; Stoel, van Drie & van Boxtel, 2017).

Currently, research examining whether improving students' historical inquiry skills might foster a transfer to democratic skills is very limited. One exception is the study of Reisman (2012) who examined the effects of an intervention based on document-based lessons and design principles derived from cognitive apprenticeship, including explicit instruction on historical reading strategies. This study also included a 'transfer of historical thinking' measure to examine whether students transferred the discipline-specific reading strategies to contemporary topics. The intervention found significant effects on historical thinking and on transfer with regard to two strategies: sourcing and close reading. No effect, however, was found for intertextual strategies, such as corroboration. Although this study provides some evidence that students may become more critical of contemporary sources after receiving instruction on critical analysis of historical sources, it only found a transfer effect on some aspects and does not provide insight into how students use these sources when taking up a reasoned position towards a public debate. Considering the widespread assumption among history education scholars that historical inquiry skills will affect how students analyze and use contemporary sources in light of public debates it seems imperative to gain more insight into whether a transfer from historical to democratic skills is indeed possible and how such a transfer may take shape.

Our own intervention study (Wilke et al., 2022b) sought to remedy this gap and examined the disciplinary effects of a lesson series aimed at students' historical inquiry skills, as well as a possible transfer to democratic skills (see figure 1 in the introduction). The lesson series made use of multiple documents inquiry tasks and design principles form the model of cognitive apprenticeship (Collins et al., 1991) to foster these skills among students. In the pretest and posttest, students' historical inquiry skills were measured using a historical inquiry task. Democratic skills, in particular students' analysis and use of contemporary sources in light of public debates, were assessed using a similar task based on contemporary sources. In both tasks, students were instructed to critically analyze a set of sources in order to take up a reasoned position towards a historical and a societal issue (for a detailed description: see further). Although the intervention successfully improved students' historical inquiry skills, there was no direct effect on students'

democratic skills. However, changes in students' historical inquiry skills corresponded to a limited degree to progress on students' democratic skills at posttest ($B = 0.16$; $p < .001$) and this effect was slightly larger in the intervention group than in the control group. ($B = -0.14$, $p = .005$). By itself, this result only provides evidence for a correlation between both, rather than for a direct (transfer) effect. Considering the design of the study, however, which was aimed specifically at fostering students' historical inquiry skills, it appears possible that the changes in students' democratic skills were due to progress on students' historical inquiry skills. This finding, however, needs further exploring in order to establish whether students' essays indeed show signs of a transfer and on which aspects precisely this transfer may have occurred.

The current study

The current study builds on the historical and topical essay tasks gathered in our previous intervention study to examine whether and how precisely a transfer from historical to democratic skills occurred. The study focuses on those students who progressed substantially on both the historical and topical essay task after the intervention. To this end, a purposeful sample of students is selected whose essay tasks will be analyzed to answer the following two folded research question: (a) How does progression on these particular students' historical and topical essay tasks manifest itself after the intervention and (b) does this progression show signs of a transfer from historical inquiry skills to democratic skills?

METHOD[2]

Research context

This study took place in the 12th grade of secondary history education in Flanders (Belgium). History education is an autonomous school subject that takes up two hours a week in general education. The Flemish school curriculum does not include a social studies course, nor a similar subject discussing societal issues. Educational goals related to citizenship do not have a formal place in a subject-specific curriculum (such as of 'civics education'); they are rather meant to be achieved in a cross-curricular manner.

At the time of the study, a curriculum reform had been initiated in history education. This reform firmly establishes historical thinking as a central goal for history education, requiring a significant change in teachers' educational practice. Before this curriculum reform, teachers' practices with regard to historical thinking and regarding the use of multiple documents based (inquiry) tasks varied strongly. Overall, teachers tended to focus predominantly on the acquisition of substantive knowledge about the past, and less so on introducing students to the methods of historians. (Van Nieuwenhuyse et al., 2017; Voet & De Wever, 2016; Wilke & Depaepe, 2019). At the time of the study, however, this reform had only been implemented in the 7th and 8th grade of secondary education. A strong diversity was hence still present in 12th-grade-teachers' practices.

Procedure and participants

The intervention study (Wilke et al., 2022b) took place between January and April 2021 in 17 schools which were randomly assigned to a control or intervention group. Students in the intervention group participated in a lesson series of 12 to 14 lessons of 50 minutes (see further) and completed a pretest and posttest during two consecutive lessons (2x50 minutes). To ensure treatment fidelity, teachers in the intervention group had participated in a training and filled out an online

[2] This section is based strongly on the method section described in a previous article (Wilke et al., under review) as both articles build further on data gathered during the intervention study (Wilke et al., 2022b), therefore some overlap may occur between the method section of these articles.

questionnaire after each lesson. This was complemented with two interviews with each participating teacher to discuss their experiences with the materials and establish whether they had deviated from the scripted lesson plans. Schools in which one of the inquiry lessons had been skipped or who had experienced a delay in the administering of the posttest (due to corona measures in Flemish education at the time) were excluded from the sample, just as students who had missed more than two classes.[3] The final sample consisted of 266 students (six schools) in the control group and 402 students (11 schools) in the intervention group. In the intervention group, which will provide the basis for this study, students' mean age was 17.04 ($SD = 0.39$), 38.2% of students identified as male, 61.6% as female and 0.2% as other.

Design of the experimental condition

The lesson series which was implemented in the intervention group took up 12-14 history lessons and was designed around the theme of decolonization after 1945 (see Wilke et al., 2022b). The lesson series' aim was to improve students' historical inquiry skills. To this end, it included four historical inquiry tasks which were designed around an evaluative historical question (van Drie et al., 2006) and a number of conflicting sources, offering different perspectives and varying in terms of their reliability. Students were instructed to generate a substantiated answer to the historical question by critically evaluating the available sources.

Throughout the various lessons, principles from the model of cognitive apprenticeship (Collins et al., 1991) and recommendations based on previous intervention studies in history education (e.g., De La Paz & Felton, 2010; De La Paz et al., 2016; Nokes, 2017; Reisman, 2012; Stoel et al., 2015) were adopted to familiarize students with the discipline-specific procedures required for such historical inquiries. This included explicit teaching about how to evaluate and compare sources and how to construct a substantiated answer. Students were taught that a substantiated answer should include the following elements: an explicit response to the historical question (claim), arguments based on the sources in support of the claim, references to the sources to substantiate the claim, an

[3] For a more detailed discussion, see: Wilke et al., 2022b.

evaluation of the value of the sources to demonstrate why students deem the supporting sources to be valuable and to rebut counterarguments stemming from less valuable sources.

Inquiry tasks were taught via a process of modeling, scaffolding and coaching. Teachers modelled the first inquiry task explicating the various steps and reasoning processes that were needed to critically evaluate and compare the various sources and to construct a substantiated answer in writing. In subsequent inquiry tasks, students were supported by (gradually fading) scaffolds. These scaffolds were modelled after those of the University of Michigan (2021) and Monte-Sano and colleagues (2014) and supported students in every aspect of the inquiry tasks: comparing and evaluating sources, weighing evidence and arguments and writing a substantiated answer. Teachers provided feedback and coaching to students and adjusted the level of scaffolding according to students' needs.

In line with recommendation from earlier studies on fostering historical thinking, interaction was included in the lesson series via group work and whole-class discussions which were used as an opportunity for feedback and to foster students' reasoning (Reisman, 2012; Stoel et al., 2017; van Boxtel & van Drie, 2013; 2017).

Because the study was set up to examine the possibility of a 'spontaneous' form of transfer form historical to contemporary contexts, the lesson series specifically did not include any explicit attention to evaluating contemporary sources or engaging in current societal debates.

Measurement instrument

The pretest and posttest each contained a historical and topical essay task to measure students' historical inquiry and democratic skills (see Figure 1). Each of these tasks started with a brief introduction followed by a historical or topical question and a set of sources which contained different viewpoints and arguments in light of the essay's question. The sources varied in terms of perspective and value. Students wrote a 15-20 sentence essay that provided a substantiated answer to the historical or topical questions, based on a critical analysis of the provided sources.

For each essay task, a similar set of excerpts from five sources (in terms of the ratio of primary-secondary and textual-visual sources and difficulty level) was provided. Contextual information was provided for each source, elaborating for instance on the background of the author (also shedding light on the aims (s)he pursued) and the origins of the source. This information could be used to evaluate the sources' value. Topics for the historical essay tasks were (1) whether the police had a legitimate reason to use violence at the Sharpeville demonstration (1960) and (2) whether Dutch violence during the Indonesian Independence War (1945-1949) was structural or incidental. The questions for the topical essay tasks were connected to two debates that were actively discussed in Flanders at the time of the study: (1) Should the voting age be lowered to the age of 16 and (2) Should mandatory positioning tests be introduced in higher education. An overview of the sources in each essay task is presented in Table 1.

Table 1. Overview of sources in the historical and topical essay tasks at pretest and posttest

Historical essay task					
Pretest: Police violence at the Sharpeville massacre			Posttest: Excessive violence during the Indonesian Independence War		
Sources		Content	Sources		Content
A.	Anti-Apartheid journalist's report on the events at Sharpeville (1960).	A. The police started shooting without any warning and without any provocation. Only three policemen got hurt, while 200 native people were hurt by police violence. The journalist states not to have seen any weapons among the demonstrators.	A.	Excerpt from a television interview (1969) with a soldier claiming that excessive violence was used structurally during the war.	A. Excessive violence was conducted structurally. The soldier had unsuccessfully approached various media outlets immediately after the war to tell his story.
B.	Picture taken at the Sharpeville demonstration by an anti-apartheid photographer (1960).	B. Picture showing demonstrators running away (unarmed) from the police.	B.	Picture (1947) of Dutch soldiers who had taken Indonesian soldiers captive.	B. The picture was deliberately kept from the Dutch general public during the war in an attempt to shape their perception of the war.
C.	Excerpt from the court trial, interrogating the commanding officer about the events at the demonstration (1960).	C. The commanding officer reports the demonstrators were hostile and was expecting an assault. After a gunshot was heard, the policemen started shooting. He suggests that when demonstrators were hurt in the ricocheting bullets.	C.	Letter from a veteran corporal during the Indonesian war in response to the television interview.	C. The television interview was a lie. Violence is part of any war, but there was no excessive violence during this war.
D.	Official statement issued by the South Africa High Commissioner in London, shortly after the events and after strong criticism from the international community (1960).	D. The police was attacked by the demonstrators and acted in self-defense.	D.	Excerpt from a government report (1969) investigating the nature of the violence during the war. The report had to be finished withing three months.	D. Dutch soldiers behaved correctly during the war; there were no cover-ups of excessive violence and there even exist accounts of positive acts committed by Dutch soldiers in Indonesia.
E.	Excerpt from a historian's book on the events at Sharpeville (2011).	E. Only a few of the demonstrators were armed and when a shot was fired from the crowd, the police started shooting. They fired in two rounds, including one round affecting the demonstrators running away from the field.	E.	Statement from the Dutch government declaring the start of a new investigation into the existence and nature of excessive violence (2016).	E. After the publication of new historical research about excessive violence during the war, the government issues a new investigation.

Table 1. (*continued*)

Topical essay task			
Pretest: Lowering of the voting age to 16.		Posttest: Mandatory positioning tests in higher education.	
Sources:	Content:	Sources:	Content:
A. Advisory note of the Flemish Youth Council. The official advisory board for the Flemish government for all issues related to children and youth.	A. In favor of lowering the voting age as young people are inherently part of society and able to make an informed decision.	A. Opinion piece of a student representative and member of a left-wing party.	A. Research demonstrates the negative impact of a positioning test on students' motivation. Students should be allowed to 'fail' an education.
B. Opinion piece of the chair of the youth section of a Flemish nationalist party.	B. Against lowering of the voting age considering young people do not want this themselves and are too young and inexperienced to know their ideological preference. 18 is a more logical voting age as adolescents then become adults.	B. Opinion piece of lecturer and director of a University college.	B. Positioning tests would restrict access to higher education for a lot of students. Research shows that these tests are unreliable and do not generate higher success rates. Moreover, they cannot be applied to all fields of study.
C. Opinion piece of a professor specialized in democratic participation youth, based on existing research.	C. International research demonstrates young people's willingness to be politically active.	C. Opinion piece of a professor in cognitive psychology in response to opinion piece B.	C. Positioning tests only measure basic minimal competences requires for higher education. The author refers to a specific study demonstrating that the positioning test accurately identifies those students who will not pass their first year.
D. Opinion piece of a professor specialized in voting behavior and political interest among young people.	D. 16-year-olds are able to choose a political party that aligns with their ideological preferences but are more likely to have extreme ideologies. The impact of lowering the voting age on electoral results would be small.	D. Opinion piece of a lecturer at a university college.	D. Positioning tests are only snapshots of a students' ability. They cannot measure maturity, interest or motivation. These tests could demotivate potentially talented students.
E. Results from an large-scale survey conducted by several universities about 14 to 25-year-olds' views on lowering the voting age.	E. 52.3% of respondents voted against lowering the voting age, 46% was in favor only when voting would not be not mandatory.	E. Research report written by several university professors about the effects of positioning tests in STEM courses.	E. Positioning tests only have very limited impact on students' enrolment choices. Existing positioning tests for STEM are accurate predictors of students' study achievements.

Note. The first part of this table (historical essay tasks) is based on Wilke et al., *under review.*

Students' essays were evaluated with a rubric which was initially based on the work of Monte-Sano (2012), Monte-Sano and De La Paz (2012), and Monte-Sano, De La Paz and Felton (2014). It was adjusted after a pilot study with 36 students and modifications were made to better fit the explicit teaching during the lesson series. The pilot study was also used to evaluate the difficulty level of the essay tasks. The rubric consisted of six criteria: (1) making a (historical or topical) claim in response to the historical or topical question; (2) argumentation in support of the claim, which evaluated the number of arguments that were used to support the essay's claim; (3) quality of the argumentation, which assessed the quality of arguments used by students; (4) use of sources as supporting evidence, evaluating the extent to which students' referenced specific sources to support the essay's claim; (5) evaluation of the value of the supporting sources; this criterion assessed to what extent students demonstrated that they had evaluated the value (in terms of reliability and/or representativeness) of the arguments they used in support of their claim and (6) engagement with counter arguments and evidence, evaluating whether students acknowledged the existence of arguments and evidence that opposed their claim and whether they had included rebuttals.

Each criterion was graded on a 4-point scale, yielding a total essay score on 24. To establish inter-rater reliability; intraclass correlation (2-way, absolute agreement, single measures) was calculated for 10 percent of the pretest essays based on two raters. At least a good agreement (>.6) was established for each separate criterion (ICC: $M = .82$, $SD = .12$ for historical tasks and $M = .75$, $SD = .08$ for the topical task) and excellent agreement (>.75) was obtained for the total scores ($M = .87$, $SD = .08$ for the historical tasks and ICC = .87 for the topical task) (Cicchetti, 1994; Hallgren, 2012).

Data sampling

Building on the essay data collected in the intervention study, we used critical case purposeful sampling (Patton, 2015) to select students for this study. In light of the research question, we were particularly interested in those students in the intervention group who had made a substantial progress on the historical essay task and on the topical essay task as these students' essays were most likely to show signs of a potential transfer from the historical to the topical task. Table 2 depicts the average scores of students on the historical and topical essay task at pretest and

posttest, demonstrating the large degree of variation in students' progress on both tasks.

Table 2. Descriptive results on historical and topical essay at pre- and posttest in the intervention group (N = 402).

Historical essay task (on 24)			Topical essay task (on 24)		
Pretest	Posttest	Change	Pretest	Posttest	Change
M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
15.20 (3.58)	18.14 (4.45)	2.95 (4.93)	14.75 (3.44)	16.49 (3.91)	1.74 (3.96)

To determine the study's sample, a substantial progress on the historical and topical essay tasks was operationalized as an increase in both tasks of at least six points (out of 24). A progress of six points enables that students could have been able to progress on all six criteria of the evaluation rubric and that they had progressed considerably more than the average student (i.e., two to three times more, see Table 2). Students with a pretest score of 19 points or more on either or both of the pretest tasks were excluded as they would be unable to progress at least six points (n = 108). This yielded a sample of 23 students in eight intervention schools who had progressed at least six points on both the historical and topical essay task. These students' essays were selected for further analysis.

Data analysis

First, general patterns in students' pretest and posttest essays were explored based on descriptive quantitative results of the evaluation rubric's score. Then, a qualitative analysis via NVivo was used to explore these findings in an in-depth way. This qualitative analysis was an iterative process with both data driven and theory driven codes (Fereday & Cochrane, 2006). The coding process initially started from an a priori set of main coding themes in line with the evaluation rubric, which was based on criteria established in previous studies (Monte-Sano, 2012; Monte-Sano and De La Paz, 2012; Monte-Sano et al., 2014). One main coding theme was established for each of the rubric's criteria. However, the criteria 'argumentation' and 'quality of argumentation' were merged into one main coding theme 'argumentation'. Moreover, one main coding theme 'essay type' was added during the coding process as it became clear that students' essays adopted

particular structures. The final coding scheme thus consisted of six main coding themes: claim, argumentation, use of sources as evidence, evaluation of sources' value (i.e., reliability and/or representativeness), rebuttal of counter arguments and evidence, and essay type. For each main coding theme, subcodes were defined which allowed for a more detailed analysis of students' essays. These subcodes started from an a priori set of codes; new (additional) subcodes were added during the coding process[4]. The subcodes in each of the main coding themes will be briefly discussed below. Table 3 provides an overview of the coding scheme. An elaborate description for each subcode, as well as examples from students' essays can be found in Appendix A.

Table 3. Overview of the main coding themes and related subcodes

Main coding theme	Subcodes
Claim	No statement Explicit statement Implicit statement Partial statement
Argumentation in support of the claim	Own argumentation Argument derived from the sources _Form: Implicit _Form: Explicit _Quality: High _Quality: Mediocre _Quality: Low
Use of sources	Source reference Source reference in support of the claim Vague source reference

[4] Regarding criteria to evaluate a source's value, for instance, the initial coding scheme included six subcodes to identify criteria students used to evaluate a sources' value, based on the literature and in correspondence with the explicit teaching in the lesson series. A new subcode 'medium' was created as students in the topical essay tasks occasionally referred to the medium in which an opinion piece was published when evaluating a source's value. Another subcode that emerged from the data was 'nuanced evaluation' to identify instances where a student had evaluated a source's value in a nuanced way, discussing both aspects that may increase the value of the sources as well as aspects which could decrease it.

Table 3. (*continued*)

Main coding theme	Subcodes
Evaluation of sources' value	No evaluation of a source's value Superficial evaluation of a source's value Evaluation of a source's value _Form: Implicit _Form: Explicit _Quality: High _Quality: Mediocre _Quality: Low _Quality: Nuanced (additional/double coding) _Criterion: author _Criterion: goal _Criterion: audience _Criterion: societal context _Criterion: information _Criterion: time/place
Rebuttal of counter arguments and evidence	Mentioned without dialogue Superficial dialogue Rebuttal _Form: Implicit _Form: Explicit _Quality: High _Quality: Mediocre _Quality: Low _Quality: Nuanced (additional/double coding) _Criterion: author _Criterion: goal _Criterion: audience _Criterion: societal context _Criterion: information _Criterion: time/place _Criterion: representativeness _Criterion: medium
Essay type	Descriptive Argumentative Evaluative Argumentative-evaluative

The coding theme 'Claim' included subcodes indicating whether the essay contained an answer to the historical or topical question, whether this answer was in line with the question, and whether the claim was explicitly or implicitly put forward.

'Argumentation' contained subcodes to identify students' own arguments and those derived from the sources. When arguments were derived from the sources subcodes specified whether these arguments were used implicitly or explicitly and distinguished the quality of the arguments (high-mediocre-low). A high-quality argument was one that correctly reflected the contents of the source, was accurately presented and supported the essay's claim. 'Mediocre quality' was used when an argument reflected the content of a source and supported the essay's claim but was not presented clearly or accurately. A low-quality argument indicated that a student had misunderstood the content of a source or that the argument did not support the essay's claim.

For the coding theme 'Use of sources as evidence' subcodes were created to evaluate whether or not the essay contained references to specific sources and whether they were used explicitly in support of the essay's claim.

The coding theme 'Evaluation of sources' value' included subcodes to establish whether students evaluated the sources' value or not, and whether this was done in a superficial or more profound way. If students merely mentioned information about the source, such as who the author was, this was considered superficial. If a student evaluated the value of a source, subcodes signaled whether students had evaluated a source implicitly or explicitly (i.e., explicit mentioning of reliability, value, representativeness, etc. or not). The quality of the evaluation (high-mediocre-low) was also coded, as well as which criteria they had applied to do so (e.g., author, date, goal). The quality of the evaluation was indicated as high, mediocre or low, depending on the accuracy and clarity of a students' reasoning. For instance, if a student stated that a source was completely reliable because it stemmed from an eyewitness, this was coded as 'mediocre quality' because the student's reasoning was based on a relevant aspect of the source (i.e., eyewitness) but was oversimplified.

Subcodes in the coding theme 'Rebuttal of counter arguments and evidence' specified whether students addressed the existence of counter arguments or evidence, whether they had engaged with them and how this was done

(superficially or by evaluating the sources' value). When the engagement with counter argumentation included an evaluation of the sources' value, subcodes indicated whether this was done implicitly or explicitly, the quality of the evaluation (high-mediocre-low) and which criteria were used.

The coding theme 'Essay type' described particular structures which became evident in students' essays, and which changed between pretest and posttest. Research on students' causal historical writing has similarly found that students' essays can take on distinct forms (Coffin, 2004; 2006; Stoel, 2017). Four subcodes were created that referred to the main focus of students' writing: descriptive, evaluative, argumentative and argumentative-evaluative. In a descriptive essay, students were mainly occupied with describing the events discussed in the sources and giving an overview of what each sources mentioned about it. Argumentative essays were focused on arguing a certain claim by providing arguments and/or evidence (sources) but without any attention to the value of these sources. In evaluative essays, students' main focus was to evaluate the various sources, but these essays did not establish a connection between the discussion of the sources and the essay's claim. Lastly, argumentative-evaluative essays combined both. They argued for a certain claim, substantiating it with arguments and evidence (sources) while also paying attention to the value of the sources.

For the coding theme 'Essay type', the initial coding conducted by the first author was reviewed and discussed with the third author to reach consensual agreement. For the five main coding themes that were based on the evaluation rubric we did not include additional measures for interrater agreement as this coding remained close to the initial rubric for which an acceptable interrater agreement had already been established.

In what follows, the quantitative results based on the evaluation rubric will be discussed after which the results of the qualitative analysis will be presented for each of the main coding themes. In search for changes between pretest and posttest, general patterns in students' historical and topical pretest essays will first be explored. This will provide a baseline against which the posttest essays can be compared to distinguish changes in students' essays.

RESULTS

Table 4 shows the mean scores on each of the criteria of the evaluation rubric as well as the total scores on the historical and topical essay task at pretest and posttest.

Table 4. Mean score on students' historical and topical essay tasks at pretest and posttest, based on the evaluation rubric ($n = 23$).

Historical essay task	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Change *M* (*SD*)
Claim	3.00 (1.09)	3.83(0.49)	0.83 (1.27)
Argumentation	2.39 (0.84)	3.30 (0.70)	0.91 (0.79)
Quality of argumentation	2.39 (1.16)	3.35 (0.78)	0.96 (0.88)
Use of sources as evidence	1.91 (1.13)	4.00 (0.00)	2.09 (1.12)
Evaluation of sources' values	1.57 (0.51)	3.22 (0.90)	1.65 (0.88)
Engagement with counter-arguments and evidence	1.30 (0.47)	3.04 (0.88)	1.74 (0.92)
Total score	12.57 (3.27)	20.74 (2.42)	8.17 (2.27)
Topical essay task	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Change *M* (*SD*)
Claim	3.26 (0.86)	3.91 (0.29)	0.65 (0.83)
Argumentation	2.43 (0.95)	3.74 (0.45)	1.30 (0.76)
Quality of argumentation	2.26 (0.92)	3.83 (0.39)	1.57 (0.90)
Use of sources as evidence	1.65 (0.88)	3.83 (0.58)	2.17 (0.94)
Evaluation of sources' values	1.13 (0.46)	1.87 (0.87)	0.74 (0.75)
Engagement with counter-arguments and evidence	1.17 (0.49)	1.78 (0.67)	0.61 (0.58)
Total score	11.91 (2.31)	18.96 (2.12)	7.04 (1.33)

An exploration of the descriptive results in Table 4 allows for an exploration of general trends in students' pretest essays and in the progress made by students on the historical and topical essays between pretest and posttest. This will provide a first, tentative view regarding the question of how students progressed on both tasks and whether students' essays show signs of a transfer.

Table 4 demonstrates that, at pretest, students obtained, on both the historical and topical essay tasks, higher scores on the first three criteria (claim, argumentation and quality of argumentation) than on the other three (using sources as evidence, evaluating the value of sources and engaging with counter arguments and evidence). When it comes to the progress made between pretest and posttest, the change scores show that students progressed on all of the criteria of both essay tasks, but progress was distributed differently between the topical and historical essay tasks. For the historical essay tasks, students progressed the most on the criteria for which the initial (pretest) scores were the lowest. In particular, they progressed mostly on 'use of sources as evidence', 'evaluation of sources' values' and 'engagement with counterarguments and evidence' thereby closing the gap with the remaining criteria. For the topical essay task, progress was situated mostly in other criteria, namely 'argumentation' and 'quality of argumentation'. For these criteria, the pretest scores were already relatively high compared to other criteria. In line with the progress on the historical tasks, students' topical essay tasks also improved with regard to 'use of sources of evidence' which obtained low pretest scores in both the historical and topical essay tasks. Students progressed remarkably less on two criteria. For the criterion 'claim', students' pretest scores were already high, for the criterion 'engagement with counterarguments and evidence' students started out with low pretest scores which improved only slightly at posttest.

Standard deviations across the essay tasks were relatively high, indicating large differences between individual students' essay tasks. The qualitative analysis allows for a more in-depth exploration of these differences. Table 5 shows the frequency of each code at the historical and topical pretest and posttest. The results will be discussed for each of the main coding themes. In order to establish how changes in students' historical and topical essay task manifested themselves, the qualitative analysis will first examine general patterns in students' pretest essays.

Table 5. Frequency of coding and number of essays in which each code was used ($n = 23$)

Code	Historical essay				Topical essay			
	Pretest		Posttest		Pretest		Posttest	
	Files	Frequency	Files	Frequency	Files	Frequency	Files	Frequency
Claim								
None	5		0		0		0	
Explicit	13		21		19		21	
Implicit	4		1		3		2	
Partial	1		1		1		0	
Argumentation								
Own argument	4	5	0	0	18	28	7	9
Explicit argument derived from sources	11	18	18	39	17	22	21	44
Implicit argument derived from sources	7	11	7	10	4	4	9	11
Quality								
High	12	21	21	38	14	17	23	48
Mediocre	4	5	8	8	3	3	4	5
Low	2	3	3	3	6	6	2	2
Use of sources as evidence								
Vague source reference	4	6	0	0	2	3	0	0
Source reference	12	24	10	10	3	3	7	12
Source reference in support of the claim	8	13	21	43	13	17	21	48

Table 5. (*continued*)

Code	Historical essay				Topical essay			
	Pretest		Posttest		Pretest		Posttest	
	Files	Frequency	Files	Frequency	Files	Frequency	Files	Frequency
Evaluation of sources' value								
No evaluation	10	18	11	13	14	16	15	32
Superficial evaluation	7	11	8	9	3	4	6	11
Evaluation of a source's value								
Form								
Explicit evaluation	3	4	15	27	0	0	8	12
Implicit evaluation	1	2	4	4	0	0	2	2
Quality								
High	1	1	10	14	0	0	5	6
Mediocre	3	3	14	14	0	0	5	8
Low	2	2	3	3	0	0	0	0
Nuance	0	0	8	9	0	0	1	1
Criterion								
Author	4	5	9	9	0	0	4	6
Societal context	0	0	2	3	0	0	0	0
Goal and audience	0	0	5	5	0	0	2	2
Information	2	3	12	16	0	0	7	8
Representativeness	0	0	5	8	0	0	3	3
Time and place	2	2	8	9	0	0	0	0
Medium	0	0	0	0	0	0	3	3
None	1	1	1	1	0	0	0	0

Table 5. (*continued*)

Engagement with counterarguments and evidence								
Mentioned without dialogue	10	10	3	3	11	13	9	10
Superficial dialogue	6	6	4	4	1	1	4	4
Rebuttal								
Form								
Explicit evaluation	0	0	17	27	2	2	5	6
Implicit evaluation	0	0	9	9	0	0	3	5
Quality								
High	0	0	15	20	2	2	2	4
Mediocre	0	0	13	14	0	0	4	4
Low	0	0	2	2	0	0	3	3
Criterion								
Author	0	0	12	16	0	0	0	0
Goal and audience	0	0	6	7	0	0	0	0
Information	0	0	19	21	0	0	5	7
Representativeness	0	0	4	4	2	2	1	1
Time and place	0	0	1	1	0	0	0	0
Societal context	0	0	0	0	0	0	0	0
None	0	0	0	0	0	0	0	0
Essay type								
Descriptive	9		0		0		1	
Evaluative	2		5		0		0	
Argumentative	12		2		23		15	
Argumentative-evaluative	0		16		0		7	

Claim

At pretest, students' topical essay tasks were more likely to include an explicit answer to the topical question than the historical essay tasks. Nineteen out of 23 topical essays already contained an explicit answer to the topical question, whereas for the historical essay tasks this was only 13. Hence, it appeared that students take up a position towards a topical question more easily than towards a historical question.

Between pretest and posttest, students' historical essay tasks increasingly took on an explicit stance. At posttest, almost all of the essays (21 out of 23) contained an explicit answer to the historical and topical question.

Argumentation

Students' historical and topical pretest essays differed strongly with regard to argumentation in three aspects. First, in terms of the number of arguments included in the essay. Students' topical essay tasks contained far more arguments than the historical tasks (34 in the historical task vs. 54 in the topical tasks). Second, with regard to the ratio of arguments derived from the sources and those based on students' own ideas. Argumentation in the topical essay tasks was far more likely to be based on students' own views, rather than on the sources: in total, 28 out of 54 arguments were not derived from the sources and a vast majority of students (18 out of 23) included their own arguments in the essays, most often based on students' own experiences or beliefs. Student 7, for example, explained why he opposed the lowering of the voting age, based entirely on his own views on the matter as none of his arguments were mentioned in the sources (see Table 1):

> By lowering the age limit, I think we will create too much discussion: young people will get into arguments with their parents because they have a different opinion and views and they themselves are not even sure what those views are. (…) I also believe that young people at the age of 16 are still too much influenced by their parents. They have been spoon-fed their political views and may not have had time to do their own research.

Student 15, on the other hand argued, again based on her own experiences, in favor of lowering the voting age to 16, by stating that 'From my own experience I know that at that age I was already following current affairs closely'. In the historical task, only four students used their own arguments to support their answer to the historical claim; most arguments (29 out of 34) were based on the provided sources. A third difference concerned the way in which these arguments were used. Arguments derived from the sources were more likely to be connected explicitly to the essay's claim in the topical essay task (22 out of 26), than in the historical tasks (18 out of 29).

Between pretest and posttest, similar changes can be observed in both the historical and topical essays. The number of arguments increased strongly, and the vast majority of these arguments was used explicitly in support of students' stance. Moreover, students' argumentation was based almost exclusively on the sources rather than on their own arguments. Student 7 for instance, whose topical pretest was based solely on his own views and experiences, now substantiated his posttest essay on positioning tests by arguments derived from the sources. None of the historical posttest contained references to students own arguments and only seven topical essay tasks did (as opposed to 18 at pretest). Hence, students' historical and topical essay task demonstrated a similar progress towards more, higher quality and more explicit argumentation, based on the provided sources rather than on students' own views.

Use of sources as evidence

At pretest, most of students' historical and topical essay tasks contained references to sources although not always explicitly in support of the essay's claim. Particularly in the historical essay tasks some students tended to summarize the sources, rather than use them to support their answer to the historical question. This was most often the case in essays which did not contain a clear answer to the historical question. In the historical essay tasks, however, students did include more specific source references than in the topical essay tasks (37 vs. 20) as students were more likely to substantiate their answer to the topical question with their own views, rather than with the provided sources.

Between pretest and posttest, there was an increase in the number of sources that was referenced. This occurred in both the topical and historical essays

but was most striking in the topical essay tasks (from 20 to 60 specific sources references). In the vast majority of cases, sources were now used explicitly to support the essay's claim and vague references to sources disappeared entirely. Student 12's historical pretest essay, for instance, was well supported by sources but his topical pretest essay contained not a single reference to a source. At posttest, this student explicitly referred to several sources in support of his claim in the topical essay task, writing among others:

> I agree with the opinion piece by Johan De Wilde and Ann Martin in *De Standaard*. "Research by the University of Hasselt and the VUB has shown that positioning tests at the start are unreliable and thus not useful in practice". This further demonstrates that the tests are irrelevant and cannot predict success rates. (...) As Beno Schraepen mentioned in source 3, a positioning test can only assess basic knowledge.

Overall, both essay types thus demonstrated a(n even) better use of sources in support of students' claim.

Evaluating the sources' value

Historical and topical pretest essays differed in their attention for the value of sources. While students paid almost no attention to the value of sources in topical pretest essays (only three essays contained superficial attention for it); the historical pretest essays did demonstrate some attention for this. In seven essays, there was at least superficial attention for the value of sources meaning that students had mentioned an aspect of the source which was relevant in light of its value (such as the author's background) but had not made a connection to the value of the source. In six cases, students had not only mentioned a relevant aspect of one of the sources but had also connected this to an evaluation of the value of that source. This, however, still occurred only to a limited extent and students' reasoning in this regard was mostly of mediocre quality as they tended to assess the value of sources in a simplified or unsubstantiated way, stating for instance that they considered one of the sources 'more neutral' (student 17) without further clarification.

The historical essay tasks demonstrated remarkable progress in terms of students' attention to the value of sources. In 31 (vs. six at pretest) instances did a

student evaluate the value of a source beyond a superficial evaluation. This was, moreover, predominantly done in an explicit way (27 out of 31). Most of students' reasoning in this regard was of high (14 out of 31) or mediocre quality (14 out of 31) and in nine instances did students include a nuanced evaluation.

Similar progress in terms of students' attention to the value of sources could be observed in students' topical essay tasks where there was an increase in the instances in which a student evaluated the value of a source, in a superficial (from four to 11) or more profound way (from zero to 14). This also occurred mostly in an explicit way (12 out of 14 instances). Students referred most often to information about the author of a source or about information on which a source was based to evaluate the value of contemporary sources. Student 20, for instance, argued that the research report about the effects of positioning tests in STEM courses (source E) was reliable based on the following reasoning:

> In fact, the result of one's positioning test is strongly related to one's study success in the first year, as can be read in the research report on the usefulness of positioning tests. This report is based on a study by a group of authors from different universities. Thus, it is representative of a larger group. Also, it contains specific data showing that only for 12.1% of students the result [of the test] played a role in their enrollment.

However, this progress was less pronounced than in the historical essay tasks. Students' attention to the value of sources was less consistent and their reasoning did not always achieve the same quality. Although some students successfully evaluated the value of these contemporary sources, others failed to accurately argue why they considered a source in the posttest essay on positioning tests, student 10 wrote:

> Beno Schraepen is a remedial educationalist and a lecturer at the AP University College, so he does not strike me as someone who does not know what he is writing about. In addition, his opinion was published in the newspaper *De Standaard*, which is regarded as a high-quality newspaper. So, normally, no unprofessional opinions will be featured here. I therefore consider this source to be rather reliable.

Later this student argued why she considered the opinion piece of a student representative to be a reliable source, stating: “I consider this source rather reliable since she is a student representative for language and linguistics students at UGent and is active in Comac (the student movement of the [extreme left-wing party] PVDA)". Although the value of these sources is explicitly evaluated, it is not entirely clear why the fact that the source was written by a student representative means it is a reliable source as the student does not elaborate on her reasoning in this regard. Another student (18) argued that “the articles that appear in TH&MA can absolutely be considered truthful, as this is a magazine for executives of universities and university colleges". While the audience of the magazine is of course not entirely irrelevant, the fact that the article is based on extensive research conducted by specialists in the field would constitute a better reason to argue that the source could be considered reliable.

On a few occasions, students also used different criteria than those taught with regard to historical sources. The medium in which a source was published was, for instance, mentioned by three students as an aspect impacting the sources’ value as was illustrated in the excerpt from students 10’s essay.

In both the historical and topical essay tasks, students hence exhibited an increased awareness of the need to (explicitly) evaluating the value of the sources that they are using to support their claim.

Rebuttal of counter arguments and evidence

At pretest, most students’ essays contained no attention to the existence of arguments or evidence that went against their essay’s stance. Students arguing against the lowering of the voting age, for instance, argued that ‘a large number of young people do not or cannot have an opinion about politics at the age of 16. When I was 16, I wasn't concerned with this at all’ (student 12) even though the available source material clearly contradicted this position. This student (as well as several others putting forward similar arguments) valued their own experience above the evidence provided in the sources (e.g., source D which clearly states that research shows that 16-year-olds are able to choose a political party that aligns with their personal views). In the historical essays, students were equally unlikely to engage in dialogue with arguments and evidence that contradicted their own

claim, although some essay did contain limited attention to counter evidence or sources.

Students' historical and topical essays exhibited similar progress between pretest and posttest. At posttest, students' essay contained more (explicit) attention to counter arguments, most often including an evaluation of the value of sources. Student 10, for instance, referred to both the reliability and the representativeness of a source to rebut its content in the historical and topical essay task. In the historical essay task, she wrote:

> The 1969 *excessennota* [government report on Dutch violence] shows us the conclusion of an investigation into the acts of violence committed by the Dutch military, commissioned by the Dutch government. This source is, in my opinion, not reliable at all since it does not constitute a thorough investigation. The report had to be ready within three months. Also, an investigation would never have been initiated if Hueting had not spoken about the violence. This source also mentions that a lot of good deeds were done by the Netherlands in that context, but there is no evidence recorded of this. So, in my opinion this is a big lie. So, it seems very clear to me that the Dutch government wants to condone the violence. This source is also certainly not representative, because it concerns the position of the Dutch government.

In the topical essay task, this student similarly included a rebuttal of counter evidence by arguing why the source was less reliable:

> Source 4 "Research report on the usefulness of positioning tests in TH&MA (2020)" contradicts De Cooman. The research shows that such a positioning test does not stop most [students] from enrolling and so it would not affect their self-confidence as much as De Cooman describes. I have no doubt that the authors conducted their research well, but the research was only carried out in STEM-disciplines. In many of these fields of study, positioning tests are already mandatory and I agree that in these fields of study, they are indeed necessary. What about all the other fields of study? No research has been conducted on this and so I consider this source to be rather unreliable and unrepresentative in answering the historical [*sic*] question.

As was the case for evaluation of the value of sources, students on average advanced more on the historical than on the topical essay task. Their historical tasks contained more rebuttals, and these rebuttals were more explicit. Moreover, the quality of students' rebuttals was higher in the historical essay tasks. In this regard, the same difficulties could be discerned as described with regard to evaluating of the value of sources. For instance, student 9 clearly tried to argue why some of the sources were more reliable than others but in doing so, put forward some questionable reasoning. She argued that sources based on opinion pieces were written 'voluntarily' and therefore more reliable than the research article which was – according to her - a 'commercial piece'.

> The sources that support my position are all individuals who are in or related to education. These are the closest to students and have written their [opinion] pieces voluntarily. The first opposing source comes from a cognitive psychology professor who talks about the success of a test that should help with study choices and relates it to the [positioning] tests. He talks about the usefulness of these [positioning tests] and not about the impact on the students themselves. The last source mentioned also stems from individuals close to the students, but it is a commercial piece that only talks about STEM-fields.

In line with the findings for the previous coding theme students' essays demonstrate an increased consideration of the need to include rebuttals in their topical essay tasks, but also show that students struggle to do so in an adequate way.

Essay structure

The essay structure of the historical and topical essay tasks differed strongly at pretest. Topical essays were in all cases argumentative in nature whereas students' historical essays were mostly descriptive (nine essays) or argumentative (12 essays).

At posttest, the historical essays demonstrated a strong shift towards essay types that included more attention to the value of sources as the majority of historical essays (16 out of 23) took on an argumentative-evaluative stance. This shift also occurred in the topical essays, although it was less prominent. Most topical essays (15 out of 23) remained argumentative in nature, implying they did

not pay a lot of attention to the value of sources, although seven did shift to an argumentative-evaluative type. These changes reflect an increased attention to the value of sources, similar to that in the historical essays, but also show that this change was not as common in the topical essays as in the historical ones.

CONCLUSION AND DISCUSSION

This study explored changes in a purposefully selected group of students' historical and topical essay tasks to explore whether and on which particular aspects these students' tasks potentially showed signs of a transfer from historical to democratic skills. In particular, it explored whether students' progress on a historical inquiry task improved their ability to make reasoned and rational judgments towards public debates on societal issues, based on arguments and evidence, a critical evaluation of contemporary sources (evidence) and arguments and an evaluation of different perspectives. To this end, it qualitatively analyzed a purposeful sample of students' historical and topical essays stemming from a previously conducted intervention study that found a connection between progress on students' historical and topical essay tasks. Although students' essays in this study may indeed suggest that a limited transfer of learning occurred, in principle we could not be sure about whether the changes in students' topical essays were due to changes in students' historical essays, or vice versa. However, considering the set-up of the intervention and the nature of changes in students' essays, these indications seem to suggest that a transfer did indeed occur – to a limited extent - from the historical to the topical essays. Moreover, as the students selected for this study were part of eight different intervention schools, this minimizes the likelihood that students' progress on the topical essays was caused by factors outside of the intervention.

The analysis of students' essay tasks revealed strong differences at the pretest level between students' historical and topical essays with regard to claim, argumentation and reference to sources, which was confirmed by the differences in their pretest essay structures. In the topical essays, students were more likely to explicitly provide an answer to the topical question and substantiate it with arguments, most often based on their own views rather than on the sources. Historical essays did not always contain a clear statement but were more likely to

reference the available sources, whether or not in support of the essay's claim. This is evidenced by the predominance of the descriptive essay type in the historical pretests. These descriptive essays lack a clear answer to the historical statement but do describe the content of the provided sources.

It thus appeared that students were more comfortable taking up and arguing for a certain position in response to a societal debate (i.e., the topical essay question) than they were in response to a historical question. This may not seem all that surprising considering that students are more likely to spontaneously have an opinion on societal issues, rather than on historical issues. These differences, however, can also point to variations in students' task and activity models (see Britt & Rouet, 2012; Rouet et al., 2017; Wiley et al., 2020). It is possible that, based on students' experiences with similar tasks in the pasts, students had different ideas about what was expected of them, and which activities were needed to achieve that. This would also explain the large variation in students' historical essay types. Considering the large diversity in teachers' current practices regarding historical thinking and historical inquiry, it is likely that students started these tasks with very different initial views.

Besides the existing differences at pretest, both essay types demonstrated, to a certain extent, a similar evolution between pretest and posttest, as they relied more on the sources to substantiate their claim and to derive arguments from and as they did so more explicitly. Considering that some of these elements were already present in students' topical essays at pretest, we should be cautious to assume that this necessarily indicates a transfer from the historical to the topical essay tasks. Nevertheless, students' historical and topical essays clearly reflected an increased reliance on, and use of, the provided sources. The finding that students' essays became more similar at posttest indicates that students are at least able to recognize the similarities between the historical and topical essay tasks. This is a necessary step to allow for a transfer of learning (Barnett & Ceci, 2002; Van Peppen et al., 2021). For other aspects of students' progress, namely the increased attention to the value of sources and to engaging with counterarguments and evidence, a transfer effect seems to be more likely. These elements were not or only barely present in the topical pretest and occurred in the topical essay tasks only after the intervention which was directed explicitly at students' historical inquiry

skills. Only at posttest did students appear to develop an awareness of the need to evaluate the value of sources and engage with counter arguments and evidence.

It should be noted, however, that this particular progress on the historical essay tasks only occurred in the topical essays in a limited way. The discussion of these coding themes made clear that this increased attention to the value of sources and to rebuttals was less prominent in the topical essays. Shifts in students' essay types also showed that only seven students evolved to an argumentative-evaluative topical essay type which focused both on substantiating a certain claim and on evaluating the value of sources. Students' initial pretest scores may provide an explanation in this regard as these varied strongly among students in the sample (as evidenced by the relatively large standard deviations in Table 4). When taking a closer look at those students whose essays showed the most obvious signs of a transfer, it seems that these were predominantly students whose pretest essays already obtained relatively high scores. This may indicate that the elements of evaluating a source's value and rebuttal require more advanced abilities.

An in-depth analysis of how precisely this increased attention to the value of sources and to counterarguments and evidence appeared in the topical essay tasks, highlights an important issue. Although students exhibited an increased attention to these aspects of the essays, this did not necessarily result in higher quality essays. In particular, the quality of students' reasoning about the value of contemporary sources and with regard to rebutting counter arguments and evidence was not of the same quality as with regard to historical sources. This shows that, even if history education is able to increase students' inclination to evaluate the value of sources, the criteria that students learn to apply to evaluate the value of historical sources are not straightforwardly applicable to contemporary sources. Therefore, it seems necessary that students are not only explicitly taught how to assess the value of historical sources, but also of various types of contemporary sources.

Although the essay data used in this study allowed for a detailed analysis of students' writing, they offer little insight into students' underlying reasoning processes. Ideally, students' answers towards the topical and historical essay tasks would be shaped through a weighing of the available evidence and sources, rather than on their own personal opinions or preconceptions. We cannot be sure, however, whether changes observed in students' essays truly reflect changes in

their reasoning processes. This is of particular importance for the topical essays which centered around topics that are relevant to students' personal lives and towards which they can be expected to already have a particular point of view. When students' topical essays show an increase in students' attention to the value of source and counter sources, we cannot be sure to what extent this consideration of source material truly shaped their point of view and potentially compelled them to (re)evaluate their initial position. At the same time, it is possible that a lack of attention to the value of sources and counterarguments and evidence in a student's posttest essay does not indicate that these skills did not transfer, but rather that the student chose to only consider and refer to sources that supported their initial position. In order to fully determine why a transfer effect was not always observed on these aspects, we would need to gather data on students' underlying reasoning processes when completing these tasks. This could be achieved in follow-up research, for instance, via interviews or think-aloud methods which would allow a better understanding of students' reasoning when confronted with a multiple-documents inquiry task.

This study allows us to further reflect on the possible relationship between historical thinking and democratic citizenship. It provides evidence that fostering students' historical inquiry skills may indeed affect, to a small degree, how students approach and use contemporary sources to argue their position on contemporary debates but mainly highlights that such a transfer is difficult to obtain and – in the few cases where it does occur - remains very limited in scope. This is particularly relevant when taking into account the fact that this study is based on a small, purposeful sample of only those students whose essay scores indicated that a transfer may in fact have occurred. Rather, when taking into account general educational literature on conditions for a transfer of learning, it appears that fostering a substantial transfer among a general population of students would require explicit attention to be dedicated to the development of these democratic skills. This, however, takes time and therefore urges to (re)consider what role is expected from history education to take up in this regard. Can historical thinking, which is complex and difficult to foster, be considered valuable enough as an educational goal in itself, or does history education also need to contribute directly to democratic citizenship? Considering the importance

of developing these democratic skills, it may perhaps be more worthwhile to give them a designated place within the educational curriculum.

Main coding themes and related subcodes	Description	Example
CLAIM (POSITION STATEMENT)		
No statement	The essay does not contain a statement or claim in response to the historical/topical question	There has never been any clarity on who shot first and whether this was a matter of police brutality or pure self-defense. The police violence was both legitimate and illegitimate.
Explicit statement	The essay explicitly puts forward a clear statement or claim in response to the historical/topical question	Based on the analysis of these sources, I conclude that the acts of violence committed by Dutch troops in Indonesia were structural acts. Thus, in my opinion it is not a good idea to lower the voting age.
Implicit statement	The essay contains a statement or claim in response to the historical/topical question, but is not explicitly stated as such.	In my opinion, this was a misunderstanding where there were bad apples in both groups. However, the police should have given a warning and stopped shooting once the crowd fled. I can well imagine that the police panicked because of this attack. But shooting at innocent people is never the solution, I think.
Partial statement	The essay contains a statement or claim that is only partly in response to the historical/topical question (e.g. not entirely in line with the question)	What exactly do I think about it? Well, I haven't been able to do enough research yet, but I think it's easy to conclude that killing people, for whatever cause and certainly in their backs, is something terrible and you should be punished for it.
ARGUMENTATION IN SUPPORT OF THE CLAIM		
Own argumentation	The students provides an argument in support of their answer to the historical question but this is not based on the sources	I think the police felt threatened by the large crowd, so they reacted in an extreme way. They killed everyone because they were afraid. A lot of 16-year-olds are too immature to take this seriously. If they are required to vote, a lot of votes would be cast randomly or as a joke.
Argument derived from the sources		
Form_implicit Form_explicit	Implicit: the argument is mentioned, but not clearly presented in support of the claim Explicit: the argument is clearly put forward in support of the claim (e.g. by use of phrases such as 'a first reason is that' or 'therefore')	*Implicit* One may ask: being pelted with some stones and one protester drawing a weapon, is this a good reason to start shooting at the protesters? Looking at the interview with the war veteran Joop Hueting in 1969, this man talks about the atrocities that he has seen happen. Namely: villages that were ruined, horrible interrogations that took place without military necessity. *Explicit* Another argument for systematic violence is the interview with Joop Hueting in 1969. In it he describes that that he and his fellow soldiers did many horrible things. It was normal, he says. Moreover, no warning shots were given and the police were pelted with stones (see sources 1 and 3). I derived a very important argument from source 5. Namely, when students fail a positioning test, it can have a very negative impact on their motivation at the beginning of the academic year.

Main coding themes and related subcodes	Description	Example
Argument derived from the sources (*continued*)		
Quality_high Quality_mediocre Quality_low	High: the argument is derived from the sources, based on an accurate understanding of the source(s), supports the claim and is accurately/clearly presented/explained Mediocre: the argument is derived from the sources and supports the claim but is not clearly presented/accurately explained Low: the argument is derived from the sources but based on an inaccurate understanding of the source and/or does not support the claim	*High* According to journalist Humprey Tyler, the crowd was not warned to disperse. The police accused the crowd of being heavily armed, however, the journalist claims that he did not see these heavy weapons. Looking at source 3 (results from large scale survey in August 2018) it appears that 52.3% of 14 to 25-year-olds who participated in this survey do not think that the 16-year-olds should be required to vote (…) about half of them prefers the right to vote, without obligation. *Mediocre* I think the biggest problem here is with the mentality of the police before the demonstration and the violence. The police were not only prepared for violence, they expected violence (source 3). *Low* It is noteworthy, that source 5 also came about after an accusation towards the Dutch. A comprehensive investigation, conducted by Limpach, once again states that the Dutch violence was of a structural nature. In 2016, the Dutch government insists on conducting a second investigation, in order to prove their innocence after all. *(Student misinterpreted the source).*
USE OF SOURCES		
Source reference	A specific source is mentioned (e.g. by referencing the source's number or author), but not connected to the claim	Historian Tom Lodge describes the situation from a historical perspective. He states that the police shot at fleeing people, which can also be proven by the many bullet wounds in the back.
Source reference in support of the claim	A specific source is mentioned (e.g. by referencing the source's number or author), and presented in support of the claim	A first reason is that extreme violence was the normal course of events and was therefore a structural fact. This is said in the television interview with Joop Hueting (source 1). Moreover, research from the University of Hasselt in source 1 mentions that the positioning tests are unreliable and therefore not useful in practice.
Vague source reference	Sources are mentioned in general, without reference to one or more specific sources.	Several sources indicate that the violence was illegitimate and that the police abused their power.

Main coding themes and related subcodes	Description	Example
EVALUATION OF THE SUPPORTING SOURCES' VALUE, IN TERMS OF THEIR RELIABILITY AND/OR REPRESENTATIVENESS		
No evaluation of a source's value	No mention of any element or aspect of the source which may influence its value	The picture of Ian Berry confirms this, we see the crowd running away. Also we see none of the so-called heavy weapons. Source 1 and 4 clearly show that a positioning test cannot predict how well a student will perform in higher education.
Superficial evaluation of a source's value	One or more relevant aspects of a source is mentioned (e.g. author, date, goal etc. of the source) but this is not connected to an evaluation of its value.	Humphrey Tyler, a reporter, and Ian Berry, a photographer, both worked for Drum Magazine. This magazine expressed their support for the anti-Apartheid movement. The two men were present at the Sharpeville demonstration. Johan De Wilde, lecturer and researcher at Odissee University College and Ann Martin, director or Education at Odisee University College consider this test to provide little information.
Evaluation of a source's value	The student evaluates the value of a source/explains why they consider the source to be valuable	
Form_implicit Form_explicit	Implicit: the student evaluates the source's value but does not explicitly mention this by referring to notions such as reliability, value, representativeness, neutrality, objectivity, etc. Explicit: the student makes explicit that they are evaluating the source's value, (e.g. by referring to reliability, representativeness, neutrality or subjectivity of the source)	*Implicit* Joop Hueting talked about the war crimes in an interview and indicated that these were not just a few incidents but were daily occurrences. He was a veteran of the war so an eyewitness. *Explicit* We can assume that Joop Hueting is a reliable source as the man was an eyewitness during the war. Source 3, 'opinion piece of Beno Schraepen' mainly states that positioning tests do not provide insight into the talent or potential of a student. Considering that this is merely an opinion piece and there is no information about who shares this opinion, I consider this source to be unrepresentative.

Main coding themes and related subcodes	Description	Example
Evaluation of a source's value (*continued*)		
Quality_high Quality_mediocre Quality_low Quality_nuance (additional/double coding)	High: Reasoning regarding a source's value makes sense/is accurate and clear Mediocre: Reasoning regarding a source's value touches upon a relevant aspects but is not entirely clear or accurate, for instance in the case of oversimplification Low: Reasoning regarding a source's value is missing or completely inaccurate. Nuance: Reasoning is presented both in favour of the source's value, but also against.	*High* A study by Dr. Limpach shows that the violence was indeed structural in nature. Dr. Limpach is himself a historian and has done extensive research to reach this conclusion, which leads me to believe that this source is reliable. An article in the journal 'TH&MA' mentions that positioning tests currently only have a limited effect on students' enrollment. This conclusion is based on extensive research which makes this a reliable source. *Mediocre* One can assume that the statement of the South Africa High Commissioner is subjective. The people in this commission were white and therefore had no idea what it is like to be black, in 1960 there was also very much racism. Because Joop was an eyewitness we can conclude that the source is reliable. *Low* Source 4 shows a photo of Dutch soldiers standing next to wounded and deceased Indonesian soldiers. This is a reliable source because photoshop did not exist at that time and it reflects reality very well. *Nuance* In my opinion, this source is a reliable source to answer this question because his goal is to talk about his experiences, he also tried to tell his story in the 1950s but was always rejected. However, we do not know whether Hueting's experiences are representative of all the other soldiers since no other accounts were given.

Main coding themes and related subcodes	Description	Example
Evaluation of a source's value (*continued*)		
Criterion_author Criterion_goal Criterion_audience Criterion_societal context Criterion_information Criterion_time/place Criterion_representativeness	Specification of which aspect of the source is mentioned in relation to its value	*Time, information, goal:* Hueting's testimony is indeed reliable. He was present at the moment itself (although this was 20 years ago, so this is not a sufficient condition). In addition, his goal appears to be simply to tell the truth to the Dutch population. He has nothing to gain from this interview, does not defend himself and only tells what he saw.

REBUTTAL OF COUNTER ARGUMENTS AND EVIDENCE		
Main coding themes and related subcodes	Description	Example
Mentioned without dialogue	The essay contains a possible counterargument or source, but this is not presented as such (no confrontation with other sources or arguments), or the student only refutes it based on their own input.	Investigations have shown that a large number of people were shot in the back. According to Pienaar, this had two reasons: ricocheting bullets may have hit the protesters in the back. Also, some of them may have been hit in the back while fleeing. I find the first reason a bit unlikely. Matilde De Cooman claims that a positioning test can negatively impact a student's motivation. I think the opposite is true: (…) you know where you stand and are offered additional support which in my opinion can only be helpful.
Superficial dialogue	The existence of counterarguments or sources is mentioned, but they are not refuted.	According to the police, their unit was pelted by a large number of stones, and the crowd of protesters was also armed with heavy weapons. Humphrey on the other hand had not seen any weapons himself. Sources 2 and 3 contradict my position and thus state that everything was done correctly and the Dutch are not to blame.
Rebuttal	The existence of a counterargument or source is mentioned and refuted by comparing the sources' value	
Form_implicit Form_explicit	Implicit: the student evaluates the source's value to rebut the counter argument but does not explicitly mention this by referring to notions such as reliability, value, representativeness, neutrality, objectivity, etc. Explicit: the student makes explicit that they are evaluating the source's value to rebut the counterargument, (e.g. by referring to reliability, representativeness, neutrality or subjectivity of the source)	*Implicit* A possible counter-argument could be that a study was already issued by the Dutch government in 1969. However, this study was only allowed to last three months (which is not enough time to comprehensively investigate everything). In addition, it came from the government itself. They could present the facts in a more positive way in order to present themselves in a more favorable way. A final source that opposes my statement is the research report in the journal 'TH&MA' (…) which stated that the positioning tests only had a very limited effect on students' enrollment (…) However, this study only applies to STEM-studies (…), thus it is not applicable to a whole range of other fields of study. *Explicit* J.G. Schuitema, a corporal from 1946 to 1948 in Central Java, issued a reader's letter eight days after the television interview, in which he writes that what Joop Hueting said is a big lie. Personally, I think this is a less reliable source, since he was a corporal and thus in charge of several soldiers. I think that with this letter he mainly wanted to save his own skin.

Main coding themes and related subcodes	Description	Example
Rebuttal (*continued*)		
Quality_high Quality_mediocre Quality_low	High: Reasoning regarding differences between the sources' value makes sense/is accurate and clear Mediocre: Reasoning regarding differences between the sources' value touches upon a relevant aspects but is not entirely clear or accurate, for instance in the case of oversimplification Low: Reasoning regarding differences between the sources' value is missing or completely inaccurate.	*High* In the reader's letter from a former veteran we can read that he accuses Heuting of telling incorrect facts. To his knowledge no prisoners of war were ever killed, they were always released. At first sight this seems to be a reliable source, he has seen the facts happening just like Hueting and is therefore an eyewitness. But we may assume that after the television interview with Hueting, the authorities from this period will have received much criticism. So there is a possibility that this Corporal just wanted to clear his name. It is also possible that both Hueting and the Corporal were active in a different area. And thus that they did not witness the same acts of violence. A final source that opposes my statement is the research report in the journal 'TH&MA' (…) which stated that the positioning tests only had a very limited effect on students' enrollment (…) However, this study only applies to STEM-studies (…), thus it is not applicable to a whole range of other fields of studies. *Mediocre* Another source on this subject is the 1969 government report. However, I do not find this reliable (…) They did not have the time to properly investigate both sides of the story and therefore I do not find the source representative. *Low* Source 3 is a reader's letter. This is probably just an old man wanting to clear his conscience. Thus, not so reliable.
Criterion_author Criterion _goal Criterion _audience Criterion _societal context Criterion _information Criterion _time/place Criterion _representativeness Criterion _medium	Specification of which aspect of the source is mentioned in relation to its value	

Main coding themes and related subcodes	Description	Example
ESSAY TYPE		
Descriptive	The essay's main focus is on describing what happened or summarising the content of the various sources. The essay provides a report of the events, as described in the sources. The essay may contain a claim in response to the historical question but no arguments or sources are clearly put forward to support the claim.	
Argumentative	The essay's main focus is on arguing why a certain claim has been made. The essay provides a claim and substantiates it by providing arguments and/or by referring to sources that support the claim. Arguments may be derived from the sources, but can also be based on students' own ideas. There is no attention for the value of sources.	
Evaluative	The essay's main focus is on (systematically) evaluating the different sources that were provided. Although the essay may include a claim, there is no/very little attention to (explicitly) substantiating the claim with arguments or sources (lack of connection between the claim and the sources).	
Argumentative-evaluative	The essay both substantiates a claim with arguments and sources *and* demonstrates attention to the value of sources. The essay contains arguments and source references in support of a claim *and* argues why certain sources may be considered more/less valuable in light of the historical/topical question.	

Note: Parts of this coding scheme are based on Wilke et al., *under review.*

Chapter 8

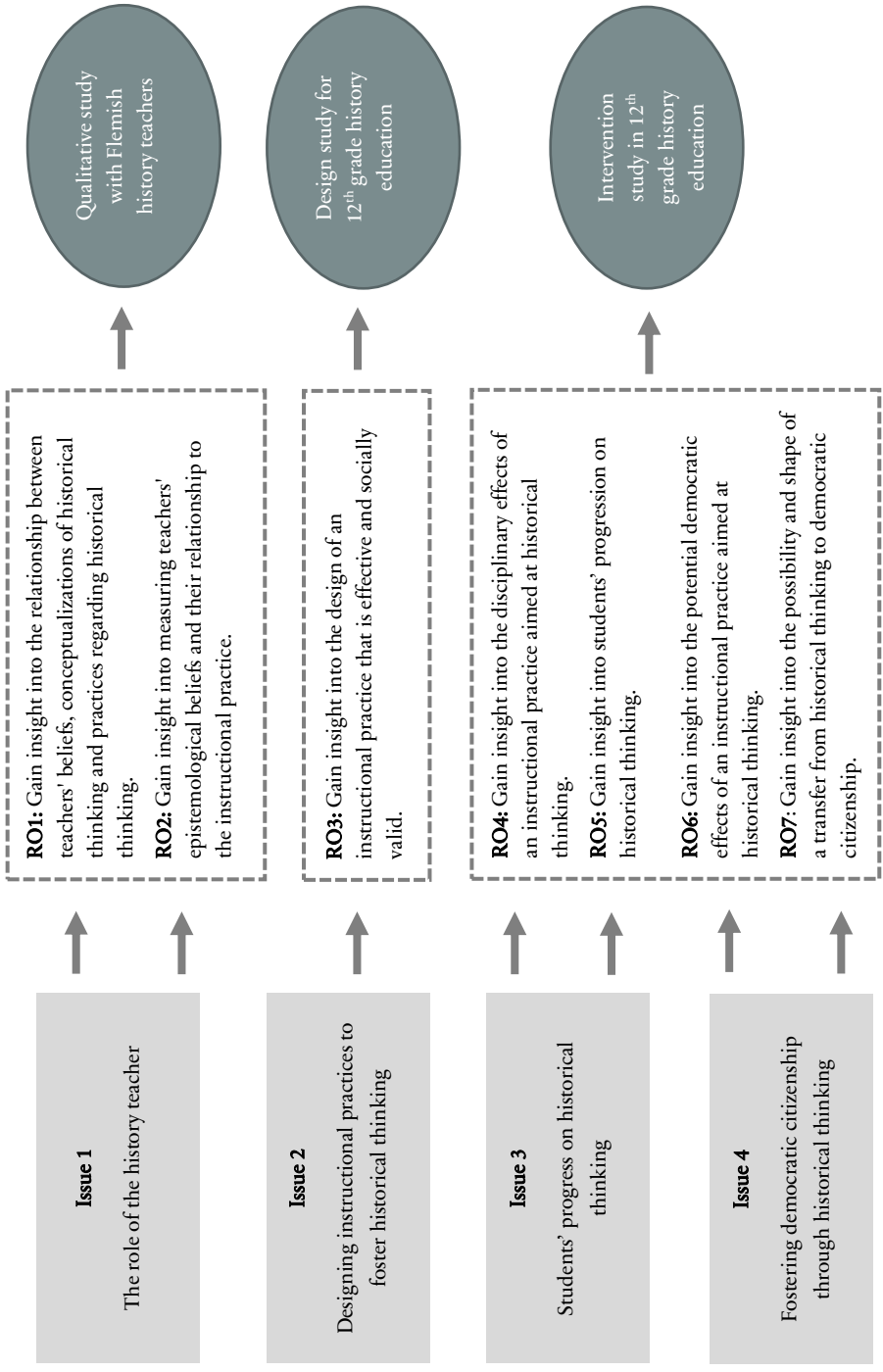
General conclusion and discussion

INTRODUCTION

The prevalence of historical thinking as an educational goal has increased strongly over the last decades. In several western countries history education has shifted closer to the academic discipline of history with the introduction of historical thinking as a central goal for history education. When students learn to think historically, they are expected to gain a better understanding of the past and of the discipline of history. Moreover, it is presumed that historical thinking might also provide a contribution to the development of democratic citizenship.

In light of the introduction of historical thinking as an educational goal, previous studies have looked at the role of history teachers' beliefs and contextual factors in shaping their instructional practice. Scholars have also examined effective teaching practices to foster historical thinking among students. Yet, more research is needed in order to better understand how teachers' beliefs and other factors influence their instructional practices. Additional empirical research is also needed regarding the implementation of historical thinking into educational practice, and regarding its disciplinary and potentially democratic effects. Four main issues are identified and addressed in this dissertation: (1) the role of the history teacher (2) the design of instructional practices that foster historical thinking, (3) students' progress on historical thinking and (4) the potential contribution of historical thinking to democratic citizenship. In light of these broader issues, seven distinct research objectives (RO) were determined which were examined via three main studies (see Figure 1 for an overview).

Figure 1. Overview of the issues, research objectives (RO) and main studies in this dissertation



This final chapter first provides a brief overview of the main findings in each of these studies in response to the proposed research objectives. Subsequently, it discusses the overall meaning and importance of the research findings in light of the four main issues this dissertation aimed to address, taking into account the particular context and limitations of the study. It also reflects on perspectives for future studies and implications for educational practice.

RECAPITULATING THE MAIN RESEARCH FINDINGS

Issue 1: The role of the history teacher

Issue 1 concerned the need to gain a better understanding of the role of the history teacher in the implementation of historical thinking into educational practice. A first set of research objectives was therefore related to obtaining insight into how various factors, including teachers' epistemological beliefs, influence teachers' instructional practice. These objectives were addressed via a qualitative study with 21 Flemish history teachers, active in upper secondary education. A questionnaire and two semi-structured interviews were developed to gather data on teachers' conceptualizations of historical thinking, their epistemological beliefs, and their beliefs about the desirability and feasibility of historical thinking as a (primary) educational goal. In order to overcome the limitations of earlier studies which relied on indirect measures to assess teachers' practices (e.g., Hicks et al., 2004; McCrum, 2013; Voet & De Wever, 2016, 2019), teachers were also asked to present their 'best instructional practices' regarding historical thinking, stemming from their actual practice.

In order to gain a more in-depth understanding of the relationship between teachers' beliefs, conceptualizations of historical thinking and instructional practices regarding historical thinking (research objective 1), the data were analyzed qualitatively through a thematic analysis. Based on a within-case analysis, a profile was constructed for each respondent describing how their various beliefs, conceptualizations and instructional practices were interconnected. A cross-case analysis was applied to distinguish general patterns of similarities and variations across the cases.

In line with findings on particular aspects of historical thinking, such as teachers' understanding of historical empathy (Bartelds et al., 2020) and of

inquiry-based learning (Voet & De Wever, 2016), the study revealed that teachers' conceptualizations of historical thinking strongly varied across participants. Overall, the majority of teachers testified to a rather narrow or one-sided understanding of the concept which only touched upon one of the building blocks described in the new history standard's operationalization. In these cases, historical thinking was often reduced to source work or a particular historical mode of reasoning (i.e., only aspects related to building block 2 or 3 of the new Flemish history standards were mentioned, and often only partially). Teachers also varied in their beliefs about the desirability and feasibility of historical thinking as an educational goal, although they overall considered it worthwhile and feasible. Teachers' narrow understanding of historical thinking was reflected in their practice. Although their instructional practices touched upon the various building blocks of the Flemish history standards, most teachers struggled to design instructional materials that did so in a profound way, and that could further students' historical thinking. The interviews and qualitative analysis identified specific difficulties in this regard, such as that teachers' questioning of sources rarely went beyond extracting the content of sources and did not adequately represent the interpretive and constructed nature of historical knowledge.

Differences in the quality of teachers' instructional materials regarding historical thinking could be connected to particular beliefs and (mis)conceptions about the notion of historical thinking or about its desirability and feasibility, and to teachers' epistemological beliefs. Importantly, the study showed that it is not possible to distinguish a clear-cut, straightforward connection between these beliefs, teachers' conceptualization of historical thinking and their practices as these factors are interconnected. For example, some teachers did not address historical thinking in an elaborated way because they felt pressured into providing students with a lot of substantive knowledge about the past, and into covering various historical topics. However, this perceived pressure was also connected to their understanding of the concept, their beliefs about its (un)importance as an educational goal, and views on how to foster it. The in-depth analysis of teachers' beliefs and practices thus highlighted the interconnectedness of teachers' beliefs in a belief system, the complexity of this belief system, and its connection to practice. With this study, we also brought to light the importance of including teachers' conceptualizations of historical thinking as an important factor in

shaping their instructional practices, something which had been overlooked in previous studies.

The qualitative study with history teachers also contributed to research objective 2. As previous scholars had encountered difficulties when trying to measure teachers' epistemological beliefs and connect them to teachers' practice (e.g., VanSledright & Reddy, 2014; Voet & De Wever, 2019), a second research objective was to develop a better understanding of how to measure teachers' epistemological beliefs and their relationship to the instructional practice. To this end, the qualitative study with history teachers contained both a measure for teachers' formal and practical epistemological beliefs. Formal epistemological beliefs concern general ideas about knowledge and knowledge construction in the discipline. Practical epistemological beliefs refer to the reasoning that is activated when engaged with concrete, discipline-specific tasks (Sinatra & Chinn, 2012). Teachers' formal epistemological beliefs were measured using statements (Stoel, Logtenberg et al., 2017); a case study with contrasting statements (based on Barzilai & Weinstock, 2015 and Kuhn et al., 2000) was used to evaluate teachers' practical epistemologies. The results of both measures were compared and connected to teachers' instructional practices and their conceptualizations of historical thinking. It appeared that teachers overall had rather nuanced epistemological beliefs, although variations occurred depending on how these were measured (see further). The study also indicated that while nuanced epistemological beliefs are a necessary requirement for a profound historical thinking practice that reflect the interpretive and constructed nature of history, they do not guarantee that such a practice is developed. This can be explained by the fact that teachers' instructional practices are not merely shaped by their epistemological beliefs, but also by other factors.

In comparing different measures for teachers' epistemological beliefs in the study, it was possible to evaluate their usefulness for capturing teachers' epistemological beliefs. In this regard it appeared that the statements used to measure formal epistemological beliefs did not succeed in isolating teachers' epistemological beliefs from more general educational ones. Rather, teachers' answers to the statements were interpreted differently and were sometimes influenced by more general educational ones, for instance about what would be feasible for students to learn. As such, teachers responded to these statements not

necessarily from their expertise as trained historians, but rather from their position as history teachers. This measurement therefore seemed less appropriate to study teachers' epistemological beliefs. The case-study with contrasting statements that was used to measure teachers' practical epistemological beliefs appeared to be better suited to capture variations of these beliefs among teachers. These variations, however, did not come to the fore by the contrasting statements themselves, but only in the ensuing discussion of teachers' responses. We therefore hypothesized that the contrasting statements still primarily gauged formal, rather than practical epistemological beliefs and that an adjusted measurement instrument is needed. Nevertheless, based on the study's findings, it seems that future studies might benefit from focusing on teachers' practical, rather than formal, beliefs in relation to their practice.

Issue 2: Designing instructional practices to foster historical thinking

A second issue which was identified in this dissertation was related to the design of instructional practices to foster historical thinking among students. In particular, more insight is needed into the design of practices that are both effective in fostering students' historical thinking in an integrated way and that are considered socially valid by teachers (research objective 3). In order to meet research objective 3, a lesson series of 12 to 14 lessons, aimed at the 12th grade of history education, was developed via a design-based study. In the lesson series, historical thinking was addressed in an integrated, holistic way, including both knowing and doing history and addressing the four building blocks of historical thinking as defined in the new Flemish history standards (Agency for Higher Education, Adult Education, Qualifications and Study Grants, n.d.). There was, however, also a particular focus on the development of students' historical inquiry competence. The lesson series adopted design principles derived from the general educational model of cognitive apprenticeship (Collins et al., 1991) which were adjusted to the specific context of history education. They consisted of the use of multiple documents inquiry tasks, explicit teaching on second-order and procedural knowledge, modeling, scaffolding, interaction, coaching, and feedback. Moreover, because the lesson series intended to foster historical thinking in a holistic way it aimed to adhere closely to recent historiography on the theme of decolonization.

The design process consisted of six phases based on the framework of McKenney and Reeves (2019) in which a close collaboration with educational practitioners and other experts was ensured. The design process started with an analysis and exploration phase in which the educational needs were defined, and design principles were established (phase 1). An initial version of the lesson series was created (phase 2), which was evaluated via a pilot study and an expert review with history and (history) education scholars as well as experienced teachers (phase 3). Based on this evaluation, an adapted version of the lesson series was designed (phase 4). This revised lesson series was evaluated via an intervention study in which data were gathered on the effectiveness of the lesson series as well as its social validity, based on teachers' reports (phase 5). Taking into account the results of this evaluation, the lesson series was subjected to a final revision (phase 6).

Throughout the various iterations, the appropriateness of these design principles was overall confirmed, but adjustments to the materials were needed to ensure that the materials were usable and feasible in practice. For instance, additional supporting materials were required to accommodate differences in teachers' prior knowledge related to historical inquiries and to support them to take on their role as experts during modeling, coaching and explicit teaching. New supporting materials were also developed for students as the lesson materials were deemed to be very challenging, among others due to the heavy reliance on textual sources and writing tasks. The lessons were also considered challenging because they introduced second-order and procedural knowledge that students were unfamiliar with. Even after these revisions, the materials were in some cases still considered very demanding for students, especially for those with lower literacy skills, and therefore difficult to apply in practice. This prompted further adjustments to the materials, for instance by way of reducing the number of sources in the inquiry tasks or offering alternatives to the writing tasks.

The design study demonstrated how to effectively translate general educational principles to the specific context of history education and managed to design a curriculum in which historical thinking was addressed in a holistic way. Via the iterations, however, the unavoidable impact of practical and contextual considerations was also exposed. The available time, students' motivation or literacy levels were elements that sometimes urged a trade-off between what would theoretically be considered a best practice and what is practically feasible. The

design study also illustrated the importance of collaborating with educational practitioners. Doing so allowed for the creation of materials that were useful and applicable in practice and which would therefore be considered socially valid.

Issue 3: Students' progress on historical thinking

Issue 3 was related to the need to gain more insight into the effects of a designed curriculum to foster students' historical thinking, and into how precisely such a curriculum affects students' progress in terms of historical thinking.

To this end, a cluster randomized controlled intervention study was set up with 668 students ($n = 266$ in the control group, $n = 402$ in the intervention group). This intervention study assessed the disciplinary effects of the lesson series that was developed through the design-based study, and of its underlying design principles (research objective 4). These principles were derived from the model of cognitive apprenticeship (Collins et al., 1991) and included, among others, explicit teaching, modeling, coaching and scaffolding (see earlier). More specifically, it was examined how the designed intervention affected 12th grade students' historical thinking, in particular their historical inquiry skills, procedural knowledge about inquiry and epistemological beliefs. To measure the intervention's disciplinary effects, a pre- and posttest was developed which was administered during two lesson hours. It consisted of a questionnaire and writing task. Historical inquiry skills were measured via a historical essay task in which students had to take a reasoned stance regarding a historical question, based on historical sources. Procedural knowledge on inquiry was measured via open-ended questions; students' epistemological beliefs were measured using statements to be evaluated on a Likert scale (Wiley et al., 2020). Multilevel modeling revealed that the lesson series was indeed effective in increasing students' historical inquiry skills and procedural knowledge on inquiry. This study thus provided further evidence for the effectiveness of teaching strategies derived from the model of cognitive apprenticeship, explicit teaching and multiple documents inquiry tasks to foster historical inquiry skills and procedural knowledge. It thus confirmed recommendations and results from earlier studies that emphasized the importance of making disciplinary ways of reasoning visible and explicit to students (e.g., Britt & Aglinskas, 2002; Stahl et al., 1996; van Boxtel et al., 2021) and of supporting students in applying them (e.g., De La Paz et al., 2016; Monte-Sano & De La Paz,

2012; Nokes et al., 2007; Stoel, van Drie & van Boxtel, 2017). However, despite the fact that the study was based on a scripted curriculum, the intervention's effects still varied across schools and students. Moreover, no effect was found with regard to students' epistemological beliefs, indicating that students' initial epistemological beliefs remained stable after the lesson series, and that students did not develop more nuanced beliefs. A more explicit and sustained effort appears to be required to change these rather robust beliefs.

The intervention study's findings were examined in a more in-depth way to meet research objective 5. Existing intervention studies often use essay tasks to examine aspects of students' historical thinking (e.g., De La Paz et al., 2016; Monte-Sano, 2012; Stoel, van Drie & van Boxtel, 2017) but do not always provide insight into the precise nature of students' progress; nor into potentially differential effects of certain interventions. In order to gain more insight into students' progression on historical thinking, a qualitative analysis was conducted to explore how precisely students' progression in terms of historical inquiry skills took shape (research objective 5). Via this qualitative analysis, we also explored whether differential effects could be discerned according to students' initial starting positions. A purposeful sample was selected of those students who had made substantial progress on the historical essay task. Their essay tasks were analyzed, comparing students with high (n = 26) and low (n = 19) initial scores. Differences in students' essay scores on the essay task were first explored quantitatively to distinguish general patterns. Subsequently, a qualitative analysis was conducted to provide a detailed view of (changes in) students' essays. The analysis was based on six main coding themes: claim, argumentation, use of sources as evidence, evaluation of sources' value (i.e., reliability and/or representativeness), engagement with counter arguments and evidence, and essay type. The analysis revealed that students' pretest essays had distinct features depending on whether they had obtained a high or low pretest score. Essays with a low pretest score usually did not contain an explicit answer to the historical question and failed to adequately support their claim with arguments and sources. Essays with a high pretest were typically better substantiated with arguments and sources, although not always in an explicit way. The changes that occurred in students' historical essays also differed strongly according to their initial scores. Students with a low initial score exhibited the most progress in substantiating

their claim with arguments and evidence, and using the available sources to do so. They also showed an emerging awareness of the need to consider the sources' value, although this occurred only sparingly. Students with a high initial score improved strongly in their attention to the value of sources and rebuttal of counterarguments and evidence, which occurred in a more consistent, explicit way and reflected a high(er) quality of reasoning.

By paying attention to differential effects of the intervention study according to students' initial ability, this analysis provided insight into developmental patterns regarding students' historical inquiry skills. In particular, providing a claim with arguments and evidence in support of it, preceded students' ability to critically evaluate sources and engage with counter arguments and evidence. By analyzing these differential effects of the intervention study, this qualitative study provided relevant insights for differentiating according to students' needs and abilities.

Issue 4: Fostering democratic citizenship through historical thinking

A final issue concerned the long-standing assumption among history education scholars that historical thinking might contribute to democratic citizenship. In particular, the studies in this dissertation were designed to examine the possibility and shape of a spontaneous transfer from students' historical inquiry competence to particular democratic skills and dispositions.

The intervention study described above (see issue 3) therefore not only examined the disciplinary effects of the lesson series developed during the design study (see issue 2), but also explored its potential democratic effects. We aimed to evaluate the largely untested assumption that fostering students' historical inquiry competences might be able to spontaneously contribute to aspects of democratic citizenship (research objective 6) (e.g., Barton & Levstik, 2004; McCully, 2012; Lee & Shemilt, 2007; Monte-Sano, 2016; Nokes, 2013; Slater, 1995; Stearns, 1998, Wineburg & Reisman, 2015). In light of research objective 6, the intervention study examined whether fostering students' historical inquiry competences affected students' ability to take up a reasoned position towards public debates on societal issues based on a weighing of evidence (democratic skills), and their inclination to do so (democratic dispositions). To this end, the pretest and posttest also contained measures to evaluate students' democratic skills and dispositions.

Democratic skills were measured via a topical essay task, parallel to the historical task, in which students took a substantiated stance towards a current societal debate, based on a critical evaluation and weighing of sources. Three democratic dispositions were evaluated using statements: deliberation within (Weinmann, 2018) critical openness (Sosu, 2013) and perspective-taking (De Corte et al., 2007).

Multilevel and linear regression models were used to examine the effects of the lesson series. The multilevel and linear models showed that there was no direct effect of the lesson series on students' democratic skills, nor dispositions. A closer look at the data did reveal a connection between students' progression on historical inquiry skills and their democratic skills. Additional analyses showed that, while the condition was not a significant predictor for democratic skills at posttest, changes in students' historical inquiry skills were. Students who had made more progress on the historical essay task also scored significantly higher on the topical essay task at posttest. This finding demonstrated that students' historical and democratic skills were connected and could improve simultaneously, yet a direct transfer effect from one to the other could not be established. We hence did not find conclusive evidence to support the expectation shared by various history education scholars (e.g., Monte-Sano, 2016; Nokes, 2013; Wineburg & Reisman, 2015) that fostering students' historical inquiry competence will affect certain democratic skills and dispositions. Although only one aspect of historical thinking's presumed contribution to democratic citizenship was examined in this study, the findings do caution for an overly optimistic view on history education's 'spontaneous' contribution to democratic citizenship and in particular, that of students' historical inquiry competences.

An additional qualitative analysis was conducted, based on the essay data gathered during the intervention study. In this analysis, it was explored whether students' essays tasks showed signs of a transfer from historical inquiry skills to democratic skills in order to gain insight into whether and how a transfer from historical inquiry skills to democratic skills might take shape (research objective 7). The analysis was based on the historical and topical essays tasks of 23 students who had progressed substantially on both tasks after the intervention. In line with the qualitative analysis for research objective 5, students' essay scores were compared quantitatively to identify common patterns after which the essays were analyzed qualitatively. To this end, an identical coding scheme was applied to both

the historical and topical essay tasks. This coding scheme again consisted of six main themes: claim, argumentation, use of sources as evidence, evaluation of sources' value (i.e., reliability and/or representativeness), rebuttal of counter arguments and evidence, and essay type. The analysis revealed that students' historical and topical essay tasks showed noticeable differences at pretest. Topical essays were more likely to contain a clear claim but often lacked appropriate substantiation with arguments and evidence. Historical essay tasks were more likely to use the provided sources but did not always include a clear response to the historical question. When examining the changes occurring between pretest and posttest, students' essay tasks indeed showed some signs of a transfer, as certain changes occurring in the historical essay tasks also occurred in the topical essay tasks at posttest. Most notably, an increased reliance on the available sources and an increased attention to their value was observed in students' topical essay tasks, in parallel to developments in students' historical essay tasks. Moreover, both essay types demonstrated an increased attention to the value of sources and the need to engage with counter arguments and evidence although students' reasoning in this regard was not always of a high quality. Considering that the study was only based on those students whose historical and topical essays showed the most promising signs of simultaneous improvement due to the intervention, and considering that evidence for a transfer remained very limited, this study mainly exposed that it is difficult for history education to contribute to democratic citizenship via students' historical inquiry competences when it does not explicitly dedicate time and effort to this goal. Again, the findings of this study go against the presumed relationship between historical inquiry skills and democratic citizenship that is asserted by various scholars (e.g., Barton & Levstik, 2004; Monte-Sano, 2016; Nokes, 2013; Wineburg & Reisman, 2015).

OVERARCHING LIMITATIONS OF THE DISSERTATION

Besides the specific limitations of each of the studies which are described throughout the various chapters, the generalizability of the studies' overall results are impacted by a number of overarching limitations and contextual factors.

Limitation 1: The potential bias among participants

A first limitation concerns the potential bias among the studies' participants. Research in educational contexts strongly relies on the willingness of educational stakeholders to participate in the research. In the studies that are central to this dissertation, these stakeholders consisted of history teachers and, by extension, their students. As these participants were recruited via open calls on a voluntary basis, it is possible, even likely, that our sample consisted of history teachers that were particularly motivated to support the implementation of historical thinking in educational practice. In order to overcome such a potential bias, the open call for the qualitative study did not mention historical thinking specifically but was presented as a study on history teachers' beliefs and practices. Even if we assume that a bias was then still present in the sample, this qualitative study revealed substantial differences in teachers' beliefs, understanding of historical thinking and their instructional practices. While these teachers may hence have been particularly motivated to participate in research on history education in general, there was no evidence to suggest that their views, understanding or practices regarding historical thinking were particularly skewed.

In the intervention study, this potential bias was also mitigated. First because of the context of the open call, which occurred at a time when history teachers were anticipating the introduction of the new history standards with a more outspoken focus on historical thinking. Hence, teachers, no matter their personal motivations or inclination towards historical thinking, would be required to make modifications to their practice to accommodate these new history standards. In this sense, the call may have invited both teachers motivated to include historical thinking more centrally in their practice but may also have spoken to teachers who mainly regarded the study as an opportunity to gain ready-made lesson materials and a training in accordance with these new history

standards. A second way in which this potential bias was mitigated was because teachers were randomly assigned to the control and intervention study.

Limitation 2: The focus on upper secondary history education

Another limitation that is of particular importance is the research's focus on upper secondary history education. This focus had implications for the teachers involved in the studies. Participating teachers had a master's degree in history and could hence be expected to be familiar with the discipline of history. This is a notably different educational background from teachers in the second and especially the first stage of secondary education who generally do not have a master's degree in history, and are hence not trained historians.

In the design and intervention study, the studies' focus was further limited to the 12th grade of history education in general secondary education. We opted for this grade because the lesson series aimed to address historical thinking in a holistic and in-depth way and although the call for participants in the intervention study was open to teachers in technical education, only classes in general education participated. In general education, history education takes up two hours a week, whereas only one hour a week is dedicated to it in technical education. Although the new history standards for the general educational track are somewhat more extensive than those for the technical track, the general goal of fostering students' historical thinking is central to the history standards in both educational tracks. As the studies in this dissertation have shown, however, getting students to think historically not only calls for well thought-out instructional practices, but also takes time. Considering that teachers both in the qualitative study and in the design study regularly asserted that time constraints hindered them from doing educational activities that foster historical thinking, this poses an even greater challenge to history teachers in technical education.

Limitation 3: The Covid-19 pandemic

Another (contextual) limitation that should be taken into account is the context of the covid-19 pandemic, in which the design and intervention study were conducted. This context caused significant challenges in a number of ways.

First, the original plans for pilot testing some of the materials fell through because of last minute changes in covid restrictions. This was accommodated by organizing a more restricted pilot test, with a small number of students.

Second, the early start (one week earlier than initially planned) and extension (from two to three weeks) of the Easter break caused, in certain classes, significant delays between the implementation of the lesson series and the completion of the posttest. In order to avoid this significant deviation from the research design to impact the results of the study, these classes were excluded from the study retrospectively. This did, however, cause a substantial loss of respondents.

Third, the covid restrictions impeded us from attending the classes in which the intervention took place, which was originally planned to ensure treatment fidelity. We hence gathered treatment fidelity data in an alternative way. Two interviews were conducted with each teacher during the course of the intervention study and teachers were asked to fill in a questionnaire after each lesson. Although it would have been informative to get a first-hand view on the implementation and students' reception of the materials, the interviews and self-reports provided rich information that allowed us to evaluate teachers' adherence to the provided curriculum and provided valuable information on the materials' practical applicability.

Finally, and most importantly, the circumstances of the pandemic created an atypical teaching context in which half of the lessons had to be given remotely. To support teachers in doing so, as well as to ensure a maximal uniformity and adherence to the designed curriculum, additional lesson materials were developed for implementing the curriculum in a remote context. Teachers reported that the context of remote education caused major challenges with regard to some key aspects of the intervention, such as interaction and coaching. Moreover, teachers noted that students' learning was negatively impacted by the context of remote education alongside the general circumstances of the pandemic. Nevertheless, the design of additional materials to support teachers during these remote classes ensured that the intervention study could still proceed in this atypical context.

Limitation 4: The focus on particular aspects of historical thinking and democratic citizenship

A final overarching limitation relates to the fact that only particular aspects of historical thinking and democratic citizenship were studied in this dissertation. Although historical thinking is connected to democratic citizenship in various ways (see Chapter 1), only the democratic effects related to students' historical inquiry competences were examined. This particularly affected the intervention study, where we focused in on evaluating the effects of the lesson series on students' historical inquiry competence. The limited amount of time available for pre- and posttest did not allow for a more integrated approach which was nevertheless present in the designed curriculum. The lesson series developed during the design study addressed historical thinking in an integrated way. The lesson series was developed in a way that all of the building blocks of historical thinking, as prescribed in the new Flemish history standards, were included. This also implied that both knowing and doing history were given a substantial place throughout the lessons. Nevertheless, in light of the intervention study, we did choose to include a distinct focus on fostering students' historical inquiry competence. A more holistic approach to historical thinking was also present in the qualitative study with history teachers, which examined teachers' conceptualizations and practices regarding historical thinking in an integral way.

DISCUSSION

Taking into account the limitations described above, the findings throughout the various studies allow for a broader reflection on their overall meaning in light of the main issues introduced in the introduction of this dissertation. The following section discusses for each of these issues (1) the importance and meaning of the findings in this dissertation, also in relation to the existing literature (2) perspectives for future research, and (3) implications for actors involved in history education practice, mainly (student-)teachers, pre-service and in-service teacher trainers.

The role of the history teacher

The studies in this dissertation first and foremost reiterate the essential role of history teachers in developing an educational practice that fosters historical thinking among students. Although this observation is not surprising in itself, the studies do offer more insight into a number of aspects that are essential in order for teachers to develop a practice that systematically fosters students' historical thinking. First, it is of great importance that teachers themselves have a deep and complete understanding of what the concept of historical thinking entails and how to promote it among students. This is in itself not an easy task, given the complexity of the notion. For the Flemish educational context, the introduction of the new history standards is a first step in ensuring that teachers are provided with a clear and uniform framework which can support them in further implementing historical thinking in their instructional practice. Second, our research shows that a profound understanding of historical thinking is, by itself, not sufficient. In order for teachers to develop an educational practice aimed at fostering historical thinking, their educational and epistemological beliefs must also be aligned with this goal. In terms of educational beliefs, the development of instructional practices that foster historical thinking is more likely to occur when teachers are convinced of the merits of fostering historical thinking and when they consider it an attainable goal for (their) students. As regards epistemological beliefs, nuanced epistemological beliefs are essential, not only on a formal level but also, or even especially, on a practical level. After all, teachers themselves must be skilled in historical thinking in order to take on the role of 'expert' when modeling and teaching explicitly about disciplinary ways of thinking. Only then can they create instructional practices that accurately reflect the constructed and interpretative nature of history, and coach students closely in becoming proficient in the act of historical thinking. However, in line with findings of earlier studies in Flanders and beyond (i.e., Hartzler-Miller, 2001; McDiarmid, 1994; VanSledright, 1996; Voet & De Wever, 2016; 2018) this study showed that these educational and epistemological beliefs are not always fully developed among history teachers.

For future studies, our findings imply that more research is needed on the development of reliable measurement instruments to capture teachers' epistemological beliefs. Based on the qualitative study, it seems particularly

important to develop measurement instruments that are able to capture teachers' practical, rather than formal, epistemological beliefs. It is also of great importance that when conducting research on factors influencing teachers' instructional practices, such as epistemological beliefs, scholars do not look at these factors in isolation. Rather, teachers' practices appear to be shaped by a multitude of beliefs and contextual factors, and by teachers' understanding of historical thinking.

For educational practice, our findings mainly contain implications for pre-service and in-service teacher training as they highlight the importance of familiarizing (student-)teachers with the concept of historical thinking, both on a theoretical and on a practical level. This means that (student-)teachers should obtain a thorough knowledge of what historical thinking entails and how to design lesson materials that effectively foster this among students. In-service and pre-service teacher trainers therefore have a vital role to take up in getting teachers 'on board' with historical thinking as an educational goal. This requires them to make explicit and engage with pre-service teachers' pre-existing beliefs and conceptions about the goals of history education and the feasibility of historical thinking. Yet, it also means that (student-)teachers should master these disciplinary ways of thinking in order to take up their role as experts. In the next section, we will argue that this is not only a task for actors involved in the formation of prospective teachers (i.e., pre-service teacher training), but also for those involved in the disciplinary training of historians (i.e., the initial disciplinary training in history).

Designing instructional practices to foster historical thinking: bridging the gap between academia and education

The studies in this dissertation emphasize the importance and value of well-designed instructional practices to foster students' historical thinking. In this regard we adopted particular design principles that were used in earlier intervention studies (e.g., De La Paz, 2005; De La Paz & Felton, 2010; De La Paz et al., 2016; Huijgen et al., 2018; Nokes et al., 2007; Reisman, 2012; Stoel, van Drie & van Boxtel, 2017), derived from the model of cognitive apprenticeship (Collins et al., 1991). The results of the intervention study confirmed the effectiveness of these principles. A first important principle is to make second-order and procedural knowledge explicit to students. By doing so, students are introduced

to the language of the discipline and are given a vocabulary to talk not only about the past, but also about history. Another effective design principle that was confirmed in this research is to design practices that encourage students to work concretely with this knowledge by having them participate - in a guided and supported manner - in the historical practice. In line with the model of cognitive apprenticeship, a process of modeling, scaffolding and coaching successfully supports students in constructing their own historical representations. When applying these principles in a lesson series, we were able to foster both students' procedural knowledge about historical inquiry and their concrete skills with regard to constructing historical representations. Moreover, these effects were found in a context of remote education and other covid-related restrictions and hence, in a far from optimal educational situation. Under normal circumstances, it is quite possible that even more substantial effects for these design principles would have been found. The studies in this dissertation, however, not only adopted design principles put forward by earlier studies, and confirmed their effectiveness, but contributed to research on the design of effective practices by identifying two additional issues that require particular attention when designing and implementing instructional practices to foster historical thinking. A first is the need to take into account practical and contextual (including teacher- and student-bound) considerations when translating the theoretical principles into the actual educational practice. These considerations include the available time for history education, students' interests, general literacy skills and motivation, but also differences in teachers' and students' prior knowledge, both in terms of knowing and doing history. These considerations occasionally make it necessary to deviate from or adapt general theoretical principles to make them workable in a specific educational context. Such adjustments are valuable and indispensable elements of the design process as they allow to translate effective design principles into educational practice. A second issue that our studies highlighted with regard to the design of instructional practices that foster historical thinking is the importance of addressing historical thinking in an integrated way, with a combined focus on both knowing and doing history. Doing so is essential because historical thinking relies on a close connection between students' substantive, second-order and procedural knowledge (Lee, 2005; VanSledright & Limón, 2006). The qualitative analysis with history teachers showed that teachers often

consider both aspects as separate and address them accordingly in their practice. So far, however few studies have been conducted which have adopted such an integrated approach. The lesson series which was developed through the design study did include a combined focus on knowing and doing history, but also highlighted how difficult it is for teachers to do so on their own. Developing a practice that fosters students' historical thinking in a profound and integrated way requires teachers to obtain a broad expertise, combining insight into the (methods of the) discipline, general pedagogy and history didactics with a profound subject matter knowledge. This is difficult to acquire, among others because recent historiographical insights are not always easily accessible for a broader audience outside of academia. For this reason, among others, a gap sometimes occurs between school and academic knowledge (e.g., Bentrovato & Van Nieuwenhuyse, 2020).

These findings contain important perspectives for future research. They clearly demonstrate that, considering the complexity of designing effective history educational practices, we need a collective effort from history education scholars and teachers. It does not seem reasonable to rely solely on the efforts of individual history teachers to conduct, on their own, all necessary efforts with regard to both knowing and doing history for particular historical topics. A joint effort therefore seems more advisable, and realistic. History education scholars can take on an active role in bridging the gap between academia and education, especially when conducting intervention studies which contribute to the design of effective history educational practices. First, history education scholars can contribute to the design of instructional practices that address historical thinking in an integrated way if they collaborate with academic historians. Such a collaboration could benefit the designed materials both in terms of knowing history (i.e., ensuring an accurate representation of recent historiographical insights) and in terms of doing history (i.e., support in terms of selecting relevant sources or historical questions). Second, history education scholars need to invest in a close(r) collaboration with history teachers. They can make an important contribution to supporting teachers in developing concrete didactic materials if they pursue social validity in their research. This can be achieved by co-creating instructional materials with teachers, and by being receptive to teachers' genuine concerns, for instance related to practical and contextual factors. Prioritizing social validity may bridge the gap

between theoretical guidelines for designing instructional practices to foster historical thinking and their practical implementation. In this way, it may ensure that history education studies have a more sustained impact on the history educational practice.

These observations also contain implications for actors involved in educational practice. First, bridging the gap between academia and education requires a combined effort from academic historians and teacher trainers. These actors are responsible for making sure that teachers are capable of historical thinking, acquire nuanced epistemological beliefs, and are knowledgeable about how to convey this to their students. Considering that historical thinking involves a combination of knowing and doing history, both of which are intrinsically connected (Lee, 2005; van Boxtel & van Drie, 2017) and taking into account the importance of teachers' own disciplinary expertise (and related epistemological beliefs), it seems imperative that teachers' initial (disciplinary) training as historians and their subsequent pre-service teacher training should work in close collaboration towards this goal. In line with recommendations from van Boxtel and colleagues (2020) it seems important that prospective history teachers are engaged in historical thinking themselves. In this regard it can be useful to consider whether the instructional practices that have been shown to be effective in promoting historical thinking among students, are sufficiently present in teacher's initial (disciplinary) training as historians. For example, explicit teaching and guided practice supported by (fading) scaffolds are practices that can be embedded both in the initial disciplinary training and in the pre-service teacher training. Second, continuing professional development via in-service teacher training is essential in order to continue to bridge the gap between academia and education. Building on the experience of the design and intervention study it seems imperative that in-service teacher training both informs teachers about instructional practices that have proved to be effective in fostering historical thinking, as well as concretely support them in implementing these practices. To this end, teachers should be coached and supported to adapt these practices and theoretical recommendations to fit the complex reality of their concrete educational contexts. Such a training, that is situated in a concrete context, cannot be obtained in a single session. Rather, in-service teacher training initiatives should aim for a regular interaction with teachers over a longer period of time.

These programs, however, should not only focus on (the implementation of) effective instructional practices, but also on elaborating teachers' subject matter knowledge. After all, high quality teaching materials that foster historical thinking in an integral way require a continued connection with recent historiography. In this regard, in-service teacher training can be supported by professional historians as well, if they make an effort to disseminate recent historiographical insights to a broader audience. Third, the implementation of instructional practices that foster students' historical thinking cannot be achieved without effort from history teachers to be open to trying out instructional approaches that may deviate from their familiar practice. In particular, the use of inquiry-based tasks, explicit teaching, modeling, scaffolding, coaching and interaction can be adopted by teachers to enhance their students' historical thinking. In this regard, the studies in this dissertation not only provide further insight into which educational principles are effective, they also offer guidelines for how to implement them. The design process which is outlined in the design study provides insight into the decisions underlying the design of the curriculum, and into how to translate and adapt these general educational principles to the particular context of history education. The lessons which were created as part of this dissertation are both directly applicable by teachers and can function as a blueprint for the design of similar materials on different topics or in different educational contexts, such as technical education or lower secondary education.

Students' progress on historical thinking

The studies in this dissertation demonstrated – in line with earlier studies – that instructional practices based on principles such as the use of multiple documents inquiry tasks, explicit teaching, modeling, scaffolding, interaction, coaching and feedback can improve aspects of students' historical thinking (e.g. De La Paz, 2005; De La Paz & Felton, 2010; De La Paz et al., 2016; Huijgen et al., 2018; Nokes et al., 2007; Reisman, 2012; Stoel, van Drie & van Boxtel, 2017). A more detailed exploration of the overall positive results of the intervention study, however, contributed to a better understanding of precisely which aspects of students' historical thinking were affected, how they were affected differently, and why this was the case. First, similar to findings of previous intervention studies (e.g., Sendur et al., 2021; Stoel, van Drie & van Boxtel, 2017), we found that certain aspects of

students' historical thinking were only affected in a limited way. For example, while students made significant progress on procedural knowledge, their final scores on procedural knowledge were on average still relatively low. Moreover, in line with the study of Stoel and colleagues (2015), no changes occurred in the area of students' epistemological beliefs. These beliefs, it appears, are particularly robust and may require more extensive and explicit attention. Hence, some aspects of students' historical thinking still remain difficult to change substantially or require attention over a longer period of time. Second, a more detailed exploration of the studies' results revealed that the lesson series did not generate the same positive effects among all students. Previous studies have already established that interventions sometimes yield differential effects according to particular student characteristics, such as students' reading skills (e.g., De La Paz et al., 2016). In our study, we found that progress on students' historical inquiry skills differed strongly, as was evidenced by the large standard deviations on students' posttest scores. In this regard, the qualitative analysis of students' progress on the historical essay tasks revealed one important element that influenced this progress, namely students' initial historical inquiry ability. Through the qualitative analysis, it was established that students progressed on different aspects of historical inquiry (e.g., substantiating a claim with arguments vs. evaluating the value of sources), depending on their initial ability. Other factors, however, such as students' general literacy skills or their motivation, may also have caused differential effects. Finally, the results of the intervention study differed at the school level. This suggests differences in the implementation of the scripted curriculum, something which was also established in other intervention studies (e.g., De La Paz et al., 2016; Reisman, 2012). This variation again points to the complexity of implementing the instructional principles that supported the lesson series. Although the scripted curriculum was developed to support teachers as extensively and concretely as possible in the implementation of the lesson series, they were still the ones responsible for putting the lessons into practice. This required them, among others, to model historical ways of reasoning, to adapt scaffolds to individual students' needs as appropriately as possible, and to provide adequate coaching and feedback throughout the lessons in order to advance their students' learning. The context of remote learning made this even more challenging for teachers, but these

instructional activities should not be underestimated even in a regular teaching context.

Based on these observations, several perspectives for future research can be discerned. First, these observations reiterate the importance of designing instructional materials that are socially valid, but also the need to gain a better understanding of how to do so. To this end, it is advisable that history education scholars work together closely with teachers when developing teaching materials and offer concrete support to teachers when implementing scripted curricula, for instance in the context of intervention studies. Second, more research is needed on how to generate epistemic change among students. Considering the importance of developing nuanced epistemological beliefs as part of historical thinking, additional studies are needed to better understand how instructional practices could support the development of such nuanced epistemological beliefs. In this regard, a first step might be to further explore whether more extensive explicit teaching on these epistemological underpinnings might indeed support students in developing nuanced epistemological beliefs about history, as this was only present in the lesson series to a limited extent. The research of Stoel, van Drie and van Boxtel (2017) shows that such explicit attention might indeed generate (small) changes in students' epistemological beliefs. Third, it seems important to gain a more in-depth knowledge of how precisely each of the various design principles adopted in the intervention study affects students' learning as well as how these effects are mediated by various student attributes. More research is needed to determine which instructional practices are more or less effective for particular groups of students, for instance according to their cognitive abilities, motivation or other traits. This could be achieved by conducting more in-depth (qualitative) analyses of intervention data to explore differential effects according to particular student characteristics, for instance via qualitative follow-up studies as was done in this dissertation. Another way to achieve this is by conducting aptitude-treatment interaction studies. These types of studies systematically evaluate how individual differences among students influence the effects of various instructional treatments (Salomon, 1974; Snow, 1974; 1989). Considering that there is little research that examines the influence of particular student characteristics on their ability to think historically, such studies seem all the more important. Doing so would allow researchers to develop targeted interventions

and instructional practices that are tailored to students' different needs, and might hence generate more optimal learning results for all students.

For educational practice, a first implication is that fostering historical thinking demands a purposeful effort which needs to be continued over an even longer period of time. Although an intervention of twelve to fourteen hours is quite substantial, it does not suffice. Moreover, such studies always function as a somewhat artificial, scripted environment. Therefore, it would be desirable that the insights and successful practices used in this, as well as in earlier intervention studies, are scaled up to history education in general and implemented in practice. In this real-life context, it will be even more important to consider differences between educational contexts, classes and individual students. This ties in to a second implication, namely the need to design instructional practices that enhance the learning of all students. This calls for a differentiated practice that accommodates students' individual needs. The studies in this dissertation provide concrete suggestions for the development of such a differentiated historical thinking practice as they provide a better understanding of students' development of historical inquiry skills. They provide concrete materials (in particular the evaluation rubric) which can be used to assess students' historical inquiry skills in a fine-grained matter and, hence, offer a better understanding of how students advance in terms of historical inquiry. This is useful by itself, as teachers' understanding students' progression in historical thinking is important in shaping their practice (Husbands, 2011). Additionally, these insights can be used to inform differentiation practices as they allow teachers to identify areas of strengths and weaknesses in student's historical inquiry competence. In particular, it can be inferred from the qualitative analysis of students' historical essay tasks that certain aspects of historical inquiry are more difficult to acquire than others. A first step, it seems, is for students to realize that they must use sources to substantiate their answer, among other things, by deriving arguments from them. Only mastering these first steps, will enable students to pay more attention to evaluating the reliability of sources and to engage with counter arguments and evidence, which are aspects that appear to be more difficult to master. While it hence may seem logical to add these more advanced elements only after students have mastered the basic ones, this is not advisable. Allowing students to use sources only to support a certain historical representation without paying attention to their value, might

reinforce pre-existing, naive views of sources as mirrors of the past. A more appropriate way might therefore be to build up the use of multiple document inquiry tasks in a step-by-step way, with students first working on smaller multiple document assignments. These tasks could, for instance, only contain two sources that clearly differ in terms of their value. From there it is possible to gradually move on to more extensive source assignments in which there is more variation and nuance in sources' value. In addition to implementing a general progression of students' historical inquiry competence across the various years of secondary education, scaffolds can be used for the purpose of within-class differentiation. This differentiation can be applied in a very targeted way, supporting students in distinct aspects of historical inquiry. Depending on the needs of the student, they can be supported in determining the usefulness and reliability of sources, extracting arguments from the sources, corroborating sources, or in formulating an answer to a historical question. Of course, a fully differentiated historical thinking practice should also account for other aspects of historical thinking, besides students' historical inquiry skills, but these fall outside the scope of this dissertation.

Fostering democratic citizenship through historical thinking

This dissertation empirically examined one of the potential contributions that historical thinking can make to democratic citizenship, namely that of students' historical inquiry competences. The results of this dissertation mainly offer an invitation to reconsider the long-standing assumption among history education scholars that fostering students' historical inquiry competences might be able to 'spontaneously' contribute to democratic goals (e.g., Barton & Levstik, 2004; Monte-Sano, 2016; Nokes, 2013; Slater, 1995; Stearns, 1998; Wineburg & Reisman, 2015). The studies in this dissertation showed that there are certainly similarities to be found between the knowledge and skills that are part of historical inquiries and those that students need when participating in societal debates. Nevertheless, it seems clear on the basis of the studies in this dissertation that we should not assume that fostering these historical inquiry competences will automatically lead to the development of certain democratic skills, nor dispositions. Evidence for such a relationship was very small and limited to a distinct group of students. This finding is not particularly unanticipated: if one

does not explicitly focus on developing certain dispositions and on the acquisition of (democratic) skills, it is unlikely that students will develop them spontaneously, as doing so would require a 'far' transfer of learning, from one context to another. This is unlikely to occur in a spontaneous way, but rather calls for instructional practices that are specifically designed to encourage such a transfer (Perkins & Salomon, 2012; van Merriënboer & Kirschner, 2007).

A first potential avenue for future studies is therefore to examine the design of history classes that would facilitate a transfer of learning to contemporary contexts. This would most likely require instructional practices that explicitly discuss the connection between historical inquiry competences and democratic skills and dispositions, and that support students in making the transfer from the former to the latter. Such research would provide valuable insights into what kind of history education, and what kind of explicit teaching and instruction, would be needed to foster democratic citizenship. Another fruitful avenue for future studies is to empirically examine other aspects of historical thinking's potential contribution to democratic citizenship. Historical thinking and democratic citizenship are broad and complex notions. In this dissertation, only one specific aspect of historical thinking's potential democratic contribution was examined. Scholars have, however, described various other ways in which historical thinking might be able to support democratic goals. It might hence be fruitful for future researchers to empirically examine these other presumed contributions of historical thinking to democratic citizenship, such as those related to agency or historical empathy (e.g., Barton, 2012; Barton & Levstik, 2004; Endacott & Brooks, 2013; Seixas, 2012). Doing so would also open up possibilities to study potential effects of historical thinking on different aspects of democratic citizenship, such as students' democratic knowledge or their active participation in society.

For educational practice, the findings in this dissertation imply that in order for historical inquiry competences to contribute to democratic goals, explicit attention and time needs to be dedicated to them. Considering that the similarities between students' historical inquiry skills and particular democratic skills are there, one could argue that history education might as well do so. While a history practice which explicitly focuses on fostering a transfer to democratic skills will likely succeed in doing so, a number of issues need to be taken into consideration. A first issue is of a practical nature. Just as explicit teaching, modeling, scaffolding

and coaching are needed in order for students to become familiar with disciplinary ways of reasoning in history, developing particular democratic skills would equally require a lot of dedicated time and effort. This time, however, is absolutely vital in order to be capable of focusing on the core objective of history education, namely developing students' historical thinking. A second issue relates to what is to be considered the main purpose of history education. It should be contemplated upon whether history education, if it does not spontaneously contribute to the development of these democratic skills and dispositions, should be tasked with the objective of developing them. The results of this study therefore not only invite us to reconsider the role of history education in the development of democratic citizenship, but also urge us to think about the status that we want to ascribe to these democratic goals in education in general. Should we continue to expect that these democratic skills and dispositions will be developed implicitly throughout the educational system, or should they be entitled to their own designated time and place, and hence a particular course, in education? Giving these vital democratic goals their own place in the educational system does not imply that history education does not have anything to offer in light of fostering democratic citizenship. However, it may be more fruitful to look for history education's contribution in aspects that are fundamental and unique to historical thinking. Historical thinking, for instance, allows students to explore alternative ways of thinking about societies and offers a chronological perspective that is crucial when trying to understand the roots and complexities of contemporary issues. Familiarizing students with the practice of historians and providing them with the skills to critically examine historical representations also enables them to be mindful of the way that history is (mis)used in contemporary debates. In this way, history education may be able to take on a complementary role in preparing students for their participation in a democratic society, while still staying close to its disciplinary nature.

References

Agency for Higher Education, Adult Education, Qualifications and Study Grants (n.d.). *Onderwijsdoelen, Uitgangspunten.* Flemish Government. https://onderwijsdoelen.be/uitgangspunten/4826

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Bartelds, H., Savenije, G. M., & van Boxtel, C. (2020). Students' and teachers' beliefs about historical empathy in secondary history education. *Theory and Research in Social Education, 48*(4), 529-551. https://doi.org/10.1080/00933104.2020.1808131

Barton, K. (2012). Agency, choice and historical action: How history teaching can help students think about democratic decision making. *Citizenship Teaching and Learning*, *7*(2), 131–142. https://doi.org/10.1386/ctl.7.2.131_1

Barton, K. C. (1997) "I just kinda know": Elementary students' ideas about historical evidence. *Theory & Research in Social Education, 25*(4), 407-430. https://doi.org/10.1080/00933104.1997.10505821

Barton, K. C. (2012). Agency, choice and historical action: How history teaching can help students think about democratic decision making. *Citizenship Teaching and Learning*, *7*(2), 131–142.

Barton, K. C., & Levstik, L. S. (2003). Why don't more history teachers engage students in interpretation? *Social Education, 67*(6), 358–361.

Barton, K. C., & Levstik, L. S. (2004). *Teaching history for the common good*. Lawrence Erlbaum Associates.

Barton, K. C., & McCully, A. (2007). Teaching controversial issues… where controversial issues really matter. *Teaching History*, *127*, 13–19.

Barton, K. C., & McCully, A. (2010). Secondary students' perspectives on school and community history in Northern Ireland. *Teachers' College Record, 112*(1) 142–181.

Barton, K. C., & McCully, A. W. (2005). History, identity and the school history curriculum in Northern Ireland: an empirical study of secondary students' ideas and perspectives. *Journal of Curriculum Studies, 37*(1), 85–116. https://doi.org/10.1080/0022027032000266070

Barzilai, S., & Weinstock, M. (2015). Measuring epistemic thinking within and across topics: A scenario-based approach. *Contemporary Educational Psychology, 42*, 141–158. https://doi.org/10.1016/j.cedpsych.2015.06.006

Bell, J., McCaffery, N. & Hansson (2010). *The troubles aren't history yet: Young people's understanding of the past.* Community Relations Council. https://www.community-relations.org.uk/sites/crc/files/media-files/The-Troubles-arent-History-Yet.pdf

Ben-David Kolikant, Y., & Pollack, S. (2019). Collaborative, multi-perspective historical writing: the explanatory power of a dialogical framework. *Dialogic Pedagogy: An International Online Journal*, *7*. https://doi.org/10.5195/dpj.2019.245

Bentrovato, D., & Van Nieuwenhuyse, K. (2020). Confronting "dark" colonial pasts: a historical analysis of practices of representation in Belgian and Congolese schools, 1945–2015, *Paedagogica Historica, 56*(3), 293-320. http://doi.org/10.1080/00309230.2019.1572203

Beswick, K. (2007). Teachers' beliefs that matter in secondary mathematics classrooms. *Educational Studies in Mathematics, 65*, 95–120. https://doi.org/10.1007/s10649-006-9035-3

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24*, 61–100. https://doi.org/10.2307/1167267

Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J.-F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist, 46*, 48–70. https://doi.org/10.1080/00461520.2011.538647

Britt, M. A., & Rouet, J.-F. (2012). Learning with multiple documents: Component skills and their acquisition. In M. J. Lawson, & J. R. Kirby (Eds.). The quality of learning: Dispositions, instruction, and mental structures (pp. 276–314). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139048224.017.

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

Brown, J. S., Collins, A. & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42. http://doi.org/10.3102/0013189X018001032

Buehl, M. M., & Alexander, P. A. (2001). Beliefs about academic knowledge. *Educational Psychology Review, 13*(4), 385–418. http://dx.doi.org/10.1023/A:1011917914756

Chinn, C. A., & Duncan, R. A. (2021). Inquiry and Learning. In R. G. Duncan, & S. A. Chinn (Eds.), *International Handbook of Inquiry and Learning* (pp. 1-14). Routledge. https://doi.org/10.4324/9781315685779

Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist, 46*(3), 141–167. https://doi.org/10.1080/00461520.2011.587722

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284-290. https://doi.org/10.1037/1040-3590.6.4.284

Cinnamon, S. A., Ribera, M. O., & Dial Sellers, H. K. (2021). Maurissen, L. (2018). Teaching disciplinary literacy through historical inquiry: Training teachers in disciplinary literacy and historical inquiry instructional practices. *The Journal of Social Studies Research*, *45*(4), 241–252. https://doi.org/10.1016/j.jssr.2021.03.001

Cobb, P., Wood, T., & Yackel, E. (1990). Classrooms as learning environments for teachers and researchers. In R. Davis, C. Maher, & N. Noddings (Eds.), *Constructivist views*

on the Teaching and Learning of Mathematics, Monograph 4 (pp. 125–146). National Council of Teachers of Mathematics. https://www.jstor.org/stable/pdf/749917.pdf

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

Coffin, C. (2006). *Historical discourse: The language of time, cause and evaluation.* Continuum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159. https://doi.org/10.1037/0033-2909.112.1.155

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, *6*(11), 1-18. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.124.8616&rep=rep1&type=pdf

Cornell University. *The Cornell note taking system. https://lsc.cornell.edu/how-to-study/taking-notes/cornell-note-taking-system/*

De Corte, K., Buysse, A., Verhofstadt, L. L., Roeyers, H., Ponnet, K., & Davis, M. H. (2007). Measuring empathic tendencies: Reliability and validity of the Dutch version of the Interpersonal Reactivity Index. *Psychologica Belgica*, *47*, 235–260. https://doi.org/10.5334/pb-47-4-235

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*(2), 139–156. http://dx.doi.org/10.1037/0022-0663.97.2.139

De La Paz, S., & Felton, M. (2010). Reading and writing from multiple source documents in history: Effects of strategy instruction with low to average high school writers. *Contemporary Educational Psychology*, *35*(3), 174–192. https://doi.org/10.1016/j.cedpsych.2010.03.001

De La Paz, S., Monte-Sano, C., Felton, M., Croninger, R., Jackson, C., & Piantedosi, K. W. (2016). A historical writing apprenticeship for adolescents: Integrating disciplinary learning with cognitive strategies. *Reading Research Quarterly, 52*(1), 31–52. https://doi.org/10.1002/rrq.147

De Wever, B., Vandepitte, P., & Jadoulle, J.-L. (2011). Historical education and didactics of history in Belgium. In E. Erdmann & W. Hasberg (Eds.), Facing, mapping, bridging diversity. Foundation of a European discourse on history education (Vol. 1, pp. 45-90). Wochenschau Wissenschaft.

Depaepe, F. & König, J. (2018). General pedagogical knowledge, self-efficacy and instructional practice: Disentangling their relationship in pre-service teacher education. *Teaching and Teacher Education, 69,* 177–190. https://doi.org/10.1016/j.tate.2017.10.003

Dickinson, A. & Lee. P. (1978). Understanding and research. In A. Dickinson, & P. Lee. (Eds.), *History teaching and historical understanding* (pp. 94-120). Heinemann.

Endacott, J., & Brooks, S. (2013). An updated theoretical and practical model for promoting historical empathy. *Social Studies Research and Practice*, *8*(1), 41–58. https://doi.org/10.1108/SSRP-01-2013-B0003

European Commission, European Education and Culture Executive Agency (2017). *Citizenship education at school in Europe. Eurydice report.* Publications Office. https://data.europa.eu/doi/10.2797/612559

Fallace, T. D. (2007). Once more unto the breach: trying to get preservice teachers to link historiographical knowledge to pedagogy. *Theory & Research in Social Education, 35*(3), 427–446. https://doi.org/10.1080/00933104.2007.10473343

Fereday J, & Muir-Cochrane E. (2006). Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods, 5*(1), 80–92. https://doi.org/10.1177/160940690600500107

Flemish Ministry of Education and Training (2000). *Secundair onderwijs, derde graad ASO: Uitgangspunten bij de vakgebonden eindtermen geschiedenis.* Flemish Government.

Flemish Ministry of Education and Training (2022, September 19). *Akkoord over aanpak ontwikkeling nieuwe minimumdoelen.* Flemish Government. https://onderwijs.vlaanderen.be/nl/nieuwsoverzicht/akkoord-over-aanpak-ontwikkeling-nieuwe-minimumdoelen

Freedman, E. (2015). "What happened needs to be told": Fostering critical historical reasoning in the classroom. *Cognition and Instruction*, *33*(4), 357–398. https://doi.org/10.1080/07370008.2015.1101465

Gallagher, C. (1996). *History teaching and the promotion of democratic values and tolerance. A handbook for teachers.* Council of Europe. https://rm.coe.int/1680494432

Gestsdóttir, S., van Boxtel, C., & van Drie, J. (2018). Teaching historical thinking and reasoning: Construction of an observation instrument. *British Educational Research Journal, 44*(6), 960–981. http://doi.org/10.1002/berj.3471

Gibson, L. & Peck, C. (2020). More than a methods course: teaching preservice teachers to think historically. In C. Berg & T. Christou (Eds.), *The Palgrave Handbook of History and Social Studies Education* (pp. 213–251). Palgrave Macmillan.

Glaser, R. (1984). *Education and Thinking The role of knowledge. American Psychologist, 39* (2), 93-104. https://doi.org/10.1037/0003-066X.39.2.93

Goldberg, T., Schwarz, B. B., & Porat, D. (2011). "Could they do it differently?": Narrative and argumentative changes in students' writing following discussion of "hot" historical issues. *Cognition and Instruction*, *29*(2), 185–217. https://doi.org/10.1080/07370008.2011.556832

Goodin, R. E. (2003). *Reflective democracy.* Oxford University Scholarship Online. https://doi.org/10.1093/0199256179.001.0001

Gouëdard, P., Pont, B., Hyttinen, S., & Huang, P. (2020). *Curriculum reform: A literature review to support effective implementation* (OECD Education Working Paper No.

239). Organisation for Economic Co-operation and Development. https://dx.doi.org/10.1787/efe8a48c-en

Hallgren, K. A. (2012). Computing inter-rater reliability, for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Harris, L. M., Halvorsen, A.-L., & Aponte-Martínez, G. J. (2016). [My] family has gone through that": How high school students determine the trustworthiness of historical documents. *Journal of Social Studies Research, 40*(2), 109–121. http://dx.doi.org/10.1016/j.jssr.2015.06.007

Harris, R. (2011). Citizenship and history. Uncomfortable bedfellows. In I. Davies (Ed.), *Debates in history teaching* (pp. 186–196). Routledge.

Hartzler-Miller, C. (2001). Making sense of "best practice" in teaching history. *Theory & Research in Social Education*, *29*(4), 672–695. https://doi.org/10.1080/00933104.2001.10505961

Havekes, H., Arno-Coppen, P., Luttenberg, J., & van Boxtel, C. (2012). Knowing and doing history: A conceptual framework and pedagogy for teaching historical contextualisation. *International Journal of Historical Learning, Teaching and Research*, *11*(1), 72–93.

Hicks, D., Doolittle, P. E., & Ewing, T. (2004). The SCIM-C strategy: Expert historians, historical inquiry, and multimedia. *Social Education, 68*(3), 221–225.

Hicks, D., Doolittle, P., & Lee, J. K. (2004). Social studies teachers' use of classroom-based and web-based historical primary sources. *Theory & Research in Social Education*, *32*(2), 213–247. https://doi.org/10.1080/00933104.2004.10473253

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67*(1), 88–140. http://dx.doi.org/10.3102/00346543067001088

Hox, J. (2010). *Multilevel Analysis. Techniques and Applications.* Routledge.

Huijgen, T., van de Grift, W. van de, Boxtel, C. van, & Holthuis, P. (2018). Promoting historical contextualization: The development and testing of a pedagogy. *Journal of Curriculum Studies*, *50*(3), 410–434. https://doi.org/10.1080/00220272.2018.1435724

Huijgen, T., van Boxtel, C., van de Grift, W. (2014). Testing elementary and secondary school students' ability to perform historical perspective taking: the constructing of valid and reliable measure instruments. *European Journal of Psychology of Education, 29*, 653–672. https://doi.org/10.1007/s10212-014-0219-4

Husbands, C. (2011). What do history teachers (need to) know? A framework for understanding and developing practice. In I. Davies (Ed.), *Debates in history teaching* (pp. 84–95). Oxon, England: Routledge.

Kazdin, A. (2005). Social validity. In B.S. Everitt, & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science,* (pp. 1875-1876), Wiley. http://doi.org/10.1002/0470013192.bsa632

Kim, J. (2019). Beyond national discourses. South Korean and Japanese students "Make a better social studies textbook". In B. C. Rubin, E. B. Freedman, & J. Kim (Eds.), *Design research in social studies education. Critical lessons from an emerging field* (pp. 225-246). Routledge. https://doi-org/10.4324/9780429024382

Kim, J. (2020). Educating citizens who are open to the discourse of others: "the last 10 feet" project and the "making a better Hiroshima textbook" project. *E-Journal of Philosophy of Education: International Yearbook of the Philosophy of Education Society of Japan, 5*. https://pesj.sakura.ne.jp/english/vol_05_1-101.pdf

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). Mahway, NJ: Lawrence Erlbaum Publishers.

Körber, A., & Meyer-Hamme, J. (2015). Historical thinking, competences and their measurement. Challenges and approaches. In K. Ercikan & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 89–101). New York & London: Routledge.

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: Lawrence Erlbaum Publishers.

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development, 15*(3), 309–328. http://dx.doi.org/10.1016/S0885-2014(00)00030-7

Land, J., Sanders, T., & van den Bergh, H. (2009). Duidelijke leerteksten in het vmbo. *Levende Talen Tijdschrijft, 10*(4), pp. 3–13.

Lee, P. & Ashby, R. (2000). Progression in historical understanding among students ages 7–14. In: P. Stearns, P. Seixas, & S. Wineburg (Eds.), *Knowing, Teaching, and Learning History: National and International Perspectives* (pp. 199–222), New York University Press.

Lee, P. (2004). Understanding history. In P. Seixas (Ed.), *Theorizing Historical Consciousness* (pp. 129–164), University of Toronto Press.

Lee, P. (2005). Putting principles into practice: Understanding history. In: M. S. Donovan, & J. D. Bransford (Eds.), *How Students Learn: History in the Classroom* (pp. 31-77), The National Academies Press.

Lee, P. (2011). Historical literacy and transformative history. In L. Periklous, & D. Shemilt (Eds.), *The future of the past: Why history education matters* (pp. 129-167). University of Toronto Press. https://issuu.com/ahdr/docs/low_ahdr_volume_a5_en

Lee, P., & Ashby, R. (2000). Progression in historical understanding among students ages 7–14. In P. Stearns, P. Seixas, & S. Wineburg (Eds.), *Knowing, teaching, and learning history: national and international perspectives* (pp. 199–222). New York University Press.

Lee, P., & Shemilt, D. (2003). A scaffold, not a cage: Progression and progression models in history. *Teaching History (113)*, 13–23.

Lee, P., & Shemilt, D. (2007). New alchemy or fatal attraction? History and citizenship. *Teaching History, 129,* 14–19.

Létourneau, J., Lévesque, S., & Gani, R. (2013), " Giant with Clay Feet' Québec Students and their Historical Consciousness of the Nation. *International Journal of Historical Learning*, 11(2), 156-172.

Lévesque, S. (2009). *Thinking historically. Educating students for the twenty-first century.* University of Toronto Press.

Lévesque, S., & Clark, P. (2018). Historical thinking: definitions and educational applications. In S. A. Metzger, & L. M. Harris (Eds.), *The Wiley international handbook of history teaching and learning* (pp. 119–148). Wiley Blackwell.

List, A., & Alexander, P. A. (2017) Cognitive affective engagement model of multiple source use. *Educational Psychologist, 52*(3), 182-199. https://doi.org/10.1080/00461520.2017.1329014

Lorah, J. (2018). Effect size measures for multilevel models: definition, interpretation, and TIMSS example. *Large-scale Assessments in Education,* 6. https://doi.org/10.1186/s40536-018-0061-2

Lorenz, C. (2002). *De constructie van het verleden. Een inleiding in de theorie van geschiedenis.* Boom.

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

Maggioni, L., & Parkinson, M. (2008). The role of teacher epistemic cognition, epistemic beliefs and calibration in instruction. *Educational Psychology Review, 20*(4), 445-461. http://doi.org/445–461. 10.1007/s10648-008-9081-8

Maggioni, L., Alexander, P., & VanSledright, B. (2004). At a crossroads? The development of epistemological beliefs and historical thinking. *European Journal of School Psychology, 2*(1–2), 169–197.

Maggioni, L., VanSledright, B., & Alexander, P. (2009). Walking on the borders: A measure of epistemic cognition in history. *The Journal of Experimental Education, 77*(3), 187–213. http://doi.org/10.3200/JEXE.77.3.187-214

Martell, C. (2013). Learning to teach history as interpretation: A longitudinal study of beginning teachers. *The Journal of Social Studies Research, 37*(1), 17-31. http://dx.doi.org/10.1016/j.jssr.2012.12.001

Maurissen, L. (2018). *Civic engagement of adolescents. A quantitative study of the relation between citizenship education, democratic attitudes, and political participation* [Unpublished doctoral dissertation]. KU Leuven.

McCrum, E. (2013). History teachers' thinking about the nature of their subject. *Teaching and Teacher Education, 35*(1), 73–80. http://dx.doi.org/10.1016/j.tate.2013.05.004

McCully, A. (2012). History teaching, conflict and the legacy of the past. *Education, citizenship and social justice, 7*(2), 145–159. https://doi.org/10.1177/1746197912440854

McDiarmid, G. W. (1994). Understanding history for teaching: A study of the historical understanding of prospective teachers. In M. Carretero & J. F. Voss (Eds.),

Cognitive and instructional processes in history and the social sciences (pp. 159–185). Hillsdale, NJ: Lawrence Erlbaum.

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education, 46*(2), 165-193. https://doi.org/10.1080/00933104.2017.1416320

McKenney, S., & Reeves, T.C. (2019). *Conducting educational design research.* Routledge.

Merrill, D. M. (2002). First principles of instruction. *Educational Technology Research and Development, 50*(3), 43–59. https://doi.org/10.1007/BF02505024

Metzger, S.A., & McArthur Harris, L. (2018). Introduction: history education in (and for) a changing world. In: In S. A. Metzger, & L. M. Harris (Eds.). *The Wiley international handbook of history teaching and learning* (pp. 1-12). Wiley Blackwell.

Miles, M., Huberman, M., & Saldaña, J. (2014). *Qualitative data analysis. A methods sourcebook.* Thousand Oaks: SAGE publications.

Moisan, S. (2010). *Fondements épistémologiques et représentations sociales d'enseignants d'histoire du secondaire à l'égard de l'enseignement de l'histoire et de la formation citoyenne* [Doctoral dissertation, Université de Montréal]. Papyrus. http://hdl.handle.net/1866/4771

Monte-Sano, C. (2010). Disciplinary literacy in history: an exploration of the historical nature of adolescents' writing. The Journal of the Learning Sciences, 19(4), 539–568. https://doi.org/10508406.2010.481014

Monte-Sano, C. (2011). Learning to open up history: preservice teachers' emerging pedagogical content knowledge. *Journal of Teacher Education, 62*(3), 260–272. http://doi.org/10.1177/0022487110397842

Monte-Sano, C. (2012). What makes a good history essay? Assessing historical aspects of argumentative writing. *Social Education, 76*(6), 294–298.

Monte-Sano, C. (2016). Argumentation in history classrooms: a key path to understanding the discipline and preparing citizens. *Theory Into Practice*, *55*(4), 311–319. https://doi.org/10.1080/00405841.2016.1208068

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

Monte-Sano, C., De La Paz, S. & Felton, M. (2014). *Reading, thinking, and writing about history. teaching argument writing to diverse learners in the common core classroom, grades 6-12.* Teachers College.

Munslow, A. (1997). *Deconstructing history.* London & New York: Routledge.

Munslow, A. (2000). *Historical Studies.* Routledge.

Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies, 19*(4), 317–328. http://doi.org/10.1080/0022027870190403

Nitsche, M. (2016). Geschichtstheoretische und -didaktische Überzeugungen von Lehrpersonen. Begriffliche und empirische Annäherungen an ein Fallbeispiel. In

M. Buchsteiner & M. Nitsche (Eds.), *Historisches Erzählen und Lernen* (pp. 159–196). Wiesbaden: Springer Fachmedien.

Nitsche, M. (2019). *Beliefs von Geschichtslehrpersonen – eine Triangulationsstudie* [Doctoral dissertation, University of Basel]. Hep-verlag. http://doi.org/10.36933/9783035516005

Nokes, J. (2017). Historical reading and writing in secondary school classrooms. In M. Carretero, S. Berger, & M. Grever (Eds.), *Palgrave handbook of research in historical culture and education* (pp. 553–571). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-52908-4

Nokes, J. (2017). Historical reading and writing in secondary school classrooms. In M. Carretero, S. Berger, & M. Grever (Eds.), *Palgrave handbook of research in historical culture and education* (pp. 553–571). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-52908-4

Nokes, J. D. (2010). Observing literacy practices in history classrooms. *Theory & Research in Social Education, 38*(4), 515–544. https://doi.org/10.1080/00933104.2010.10473438

Nokes, J. D. (2013). *Building students' historical literacies: learning to read and reason with historical texts and evidence*. Routledge.

Nokes, J. D. (2017). Historical reading and writing in secondary school classrooms. In M. Carretero, S. Berger, & M. Grever (Eds.), *Palgrave Handbook of Research in Historical Culture and Education* (pp. 553–571). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-52908-4_29

Nokes, J. D., & De la Paz, S. (2018). Writing and argumentation in history education. In S. A. Metzger & L. McArthur Harris (Eds.), The Wiley international handbook of history teaching and learning (pp. 551–578). Wiley Blackwell.

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*(3), 492–504. https://doi.org/10.1037/0022-0663.99.3.492

Pajares M. F. (1992). Teachers' beliefs and educational research: cleaning up a messy construct. *Review of Educational Research, 62*(3), 307–332. https://doi.org/10.3102/00346543062003307

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice.* SAGE.

Pauk, W., & Owens, R. (2010). *How to Study in College (10 ed.).* Wadsworth.

Paul, J., Macedo-Rouet, M., Rouet, J.-F., & Stadtler, M. (2017). Why attend to source information when reading online? The perspective of ninth grade students from two different countries. *Computers & Education, 113*, 339–354. https://doi.org/10.1016/j.compedu.2017.05.020

Peck, C. (2010). 'It's not like [I'm] Chinese and Canadian. I am in between': Ethnicity and students' conceptions of historical significance. *Theory & Research in Social Education, 38*(4), 574-617. https://doi.org/10.1080/00933104.2010.10473440

Peck, C., Poyntz, S. and Seixas, P. (2011). "Agency" in Students' Narratives of Canadian History. In D. Shemilt & L. Perikleous (Eds.), *The Future of the Past: Why History*

Education Matters (pp. 253-280), Association for Historical Dialogue and Research. file:///C:/Users/u0096281/Downloads/thefutureofthepast.pdf

Perfetti, C. A., Britt, M. A., Rouet, J.-F., Georgi, M. C., & Mason, R. A. (1994). How students use texts to learn and reason about historical uncertainty. In M. Carretero & J. F. Voss (Eds.), *Cognitive and instructional processes in history and the social sciences* (pp. 257-283). Lawrence Erlbaum.

Perkins, D. & Salomon, G. (2012). Knowledge to go: A motivational and dispositional view of transfer. *Educational Psychologist*, *47*(3), 248–258. https://doi.org/10.1080/00461520.2012.693354

Perkins, D., & Salomon, G. (1989). Are Cognitive Skills Context-Bound? *Educational Researcher, 18*(1), 16-25. https://doi.org/10.3102/0013189X018001016

Perkins, D., & Salomon, G. (1994). Transfer of learning. In T. Husen, & T. Postlethwaite (Eds.), *The international encyclopedia of education: Vol 11* (2nd ed., pp. 6452–6457). Pergamon Press.

Perkins, D.N., Jay, E., & Tishman, S. (1993). Beyond Abilities: A Dispositional Theory of Thinking, *Merrill-Palmer Quarterly, 39* (1), 1-21. https://www.jstor.org/stable/23087298

Popkewitz, T. (2004). The Alchemy of the Mathematics Curriculum: Inscriptions and the Fabrication of the Child. *American Educational Research Journal, 41*(1), 3–34. http://www.jstor.org/stable/3699383

Reisman, A. (2012). Reading like a historian: A document-based history curriculum intervention in urban high schools. *Cognition and Instruction*, *30*, 86–112. https://doi.org/0.1080/07370008.2011.634081

Reisman, A. (2015). The difficulty of assessing disciplinary historical reading. In K. Ercikan, & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 29-39). Routledge.

Rouet, J. F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology*, *88*(3), 478–493. https://doi.org/10.1037/0022-0663.88.3.478

Rouet, J.-F., Britt, M. A., & Durik, A. M. (2017) RESOLV: Readers' representation of reading contexts and tasks, *Educational Psychologist,* 52(3), 200-215. https://doi.org/10.1080/00461520.2017.1329015

Rouet, J-F., Favart, M., Britt, M.A., & Perfetti, C. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction, 15*(1), 85-106. https://doi.org/10.1207/s1532690xci1501_3

Salomon, G. (1974). ATI research: For better psychological insights or for better educational practice? In H. Crombag & D. De Gruijter, *Contemporary issues in educational testing* (pp. 248-255). De Gruyter Mouton. https://doi.org/10.1515/9783111557045

Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education, 89*, 634–656. http://doi.org/10.1002/sce.20065

Savenije, G., van Boxtel, C., & Grever, M. (2014). Sensitive 'Heritage' of Slavery in a Multicultural Classroom: Pupils' Ideas Regarding Significance, *British Journal of Educational Studies, 62*(2), 127-148. https://doi.org/10.1080/00071005.2014.910292

Seixas, P. & Peck, C. (2004). Teaching Historical Thinking. In A. Sears & I. Wright (Eds.), *Challenges and Prospects for Canadian Social Studies* (pp. 109-117). Pacific Educational Press.

Seixas, P. (2017) A model of historical thinking, *Educational Philosophy and Theory, 49*(6), 593-605. https://doi.org/10.1080/00131857.2015.1101363

Seixas, P. (2017). A model of historical thinking. *Educational Philosophy and Theory, 49*(6), 593-605. https://doi.org/10.1080/00131857.2015.1101363

Seixas, P., & Morton, T. (2013). *The Big Six Historical Thinking Concepts*. Nelson Education.

Seixas, P., (2012). Historical agency as a problem for researchers in history education. *Antíteses*, *5*(10), 537–553. http://doi.org/10.5433/1984-3356.2012v5n10p537

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

Sheppard, M., & Levy, A. A. (2019). Emotions and teacher decision-making: An analysis of social studies teachers' perspectives. *Teaching and Teacher Education, 77*, 193-203. https://doi.org/10.1016/j.tate.2018.09.010

Sinatra, G. M., & Chinn, C. A. (2012). Thinking and reasoning in science: Promoting epistemic conceptual change. In K. R. Harris, S. Graham, T. Urdan, A. G. Bus, S. Major & H. L. Swanson (Eds.), *APA educational psychology handbook* (Vol. 3, pp. 257–282). Washington, DC: American Psychological Association

Slater, J. (1995). *Teaching history in the new Europe.* Cassell Council of Europe series.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Sage Publishing.

Snow, R. E. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences: Advances in theory and research* (pp. 13–59). W H Freeman/Times Books/ Henry Holt & Co.

Snow, R.E. (1974). Aptitude-treatment interactions in educational research and evaluation. In H. Crombag & D. De Gruijter, *Contemporary issues in educational testing* (pp. 229-247). De Gruyter Mouton. https://doi.org/10.1515/9783111557045

Sosu, E. M. (2013). The development and psychometric validation of a Critical Thinking Disposition Scale. *Thinking Skills and Creativity, 9*, 107–119. http://dx.doi.org/10.1016/j.tsc.2012.09.002

Stahl, S. A., Hynd, C. R., Britton, B. K., McNish, M. M., & Bosquet, D. (1996). What happens when students read multiple source documents in history? *Reading Research Quarterly*, *31*(4), 430–456. https://www.jstor.org/stable/748185

Stearns, P. (1998). Why study history? *American Historical Association.* https://www.historians.org/about-aha-and-membership/aha-history-and-archives/historical-archives/why-study-history-(1998)

Stoel, G. L. (2017). *Teaching towards historical expertise. Developing students' ability to reason causally in history.* Doctoral dissertation, University of Amsterdam. https://hdl.handle.net/11245.1/1cfaaed8-9929-462e-b2e0-ad1945f1f034

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*(1), 49–76. https://doi.org/10.1080/00220272.2014.968212

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

Stoel, G., Logtenberg, A., Wansink, B., Huijgen, T., van Boxtel, C., & van Drie, J. (2017). Measuring epistemological beliefs in history education: An exploration of naïve and nuanced beliefs. *International Journal of Educational Research, 83*, 120–134. http://dx.doi.org/10.1016/j.ijer.2017.03.003

Stoel, G.L., van Drie, J., & van Boxtel, C. (2015). Teaching towards historical expertise. Developing a pedagogy for fostering causal reasoning in history. *Journal of Curriculum Studies, 47*(1), 49-76. https://doi.org/10.1080/00220272.2014.968212

Stradling, R. (2003). *Multiperspectivity in history teaching: a guide for teachers.* Council of Europe. https://rm.coe.int/1680493c9e

Thompson, A. (1984). The relationship of teachers' conceptions of mathematics and mathematics teaching to instructional practice. *Educational Studies in Mathematics,15*(2), 105–127. https://doi.org/10.1007/BF00305892

Tiruneh, D. T., Gu, X., De Cock, M., & Elen, J. (2018). Systematic design of domain-specific instruction on near and far transfer of critical thinking skills. *International Journal of Educational Research, 87*, 1–11. https://doi.org/10.1016/j.ijer.2017.10.005

University of Michigan (2021). *Disciplinary literacy tools structure, a process for inquiry and argument writing*. https://readinquirewrite.umich.edu/tools-structure/

Van Boxtel, C. & van Drie, J. (2018). Historical reasoning: Conceptualizations and educational applications. In S. A. Metzger & L.M. Harris (Eds.), *The Wiley international handbook of history teaching and learning* (pp. 149–176). Wiley Blackwell.

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 Boxtel, C., & van Drie, J. (2017). Engaging students in historical reasoning: the need for dialogic history education. In M. Carretero, S. Berger, & M. Grever (Eds.), *Palgrave handbook of research in historical culture and education* (pp. 573–589). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-52908-4

Van Boxtel, C., & van Drie, J. (2018). Historical reasoning: conceptualizations and educational applications. In S. A. Metzger, & L. M. Harris (Eds.). *The Wiley international handbook of history teaching and learning* (pp. 149–176). Wiley Blackwell.

van Boxtel, C., van Drie, J., & Stoel, G. (2020). Improving teachers' proficiency in teaching historical thinking. In C. Berg & T. Christou (Eds.), *The Palgrave Handbook of History and Social Studies Education* (pp. 97–117). Palgrave Macmillan.

van Boxtel, C., Voet, M., & Stoel, G. (2021). *Inquiry Learning in History*. In R. G. Duncan, & S. A. Chinn (Eds.), *International Handbook of Inquiry and Learning* (pp. 296–310). Routledge. https://doi.org/10.4324/9781315685779

van der Eem, M., van Drie, J., Brand-Gruwel, S., & van Boxtel, C. (2022). Students' evaluation of the trustworthiness of historical sources: Procedural knowledge and task value as predictors of student performance. *The Journal of Social Studies Research.* https://doi.org/10.1016/j.jssr.2022.05.003

van Drie, J., & van Boxtel, C. (2008). Historical reasoning: Towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review, 20*(2), 87–110. http://dx.doi.org/10.1007/s10648-007-9056-1

van Drie, J., & van Boxtel, C. (2008). Historical reasoning: towards a framework for analysing students' reasoning about the past. *Educational Psychology Review, 20*(2), 87–110. http://dx.doi.org/10.1007/s10648-007-9056-1

van Drie, J., & van de Ven, P.-H. (2017). Moving ideas: an exploration of students' use of dialogue for writing in history, *Language and Education, 31*(6), 526-542. https://doi.org/10.1080/09500782.2017.1326504

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*(1), 123–156. https://doi.org/10.17239/jowr-2015.07.01.06

van Drie, J., van Boxtel, C. & Stam, B. (2013). 'But why is this so important?': Discussing historical significance in the classroom. *History Education Research Journal, 12*(1), 146-168. http://doi.org/0.18546/HERJ.12.1.11

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. 266–297). Routledge. https://doi.org/10.4324/9780203826843

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

Van Hover, S. D., & Yeager, E. A. (2003). Challenges facing beginning history teachers: An exploratory study. *International Journal of Social Education*, *19*(1), 8–21. https://eric.ed.gov/?id=EJ718724

Van Merriënboer, J. J. G., & Kirschner, P. (2007). *Ten steps to complex learning: A systematic approach to four-component instructional design*. Routledge. https://doi.org/10.4324/9781410618054

Van Nieuwenhuyse, K. (2020). From knowing the national past to doing history. History (teacher) education in Flanders since 1918. In C. Berg, & T. Christou (Eds.), *The Palgrave handbook of history and social studies education* (pp. 355–386). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-37210-1

Van Nieuwenhuyse, K., & Wils, K. (2015). Historical narratives and national identities. A qualitative study of young adults in Flanders. *Journal of Belgian History, 45*(4), 40-73. https://www.journalbelgianhistory.be/en/system/files/article_pdf/005_VanNieuwenhuyse_Wils_2015_4.pdf

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

Van Nieuwenhuyse, K., Wils, K., Clarebout, G. & Verschaffel, L. (2015). The present in Flemish secondary history education through the lens of written history exams. *McGill Journal of Education / Revue des sciences de l'éducation de McGill, 50*(2-3), 433–451. https://doi.org/10.7202/1036440ar

van Peppen, L. M., van Gog, T., Verkoeijen, P., & Alexander, P. A. (2021). Identifying obstacles to transfer of critical thinking skills, *Journal of Cognitive Psychology, 34*(2), 261-288. https://doi.org/10.1080/20445911.2021.1990302

VanSledright, B. A. (1996). Closing the gap between school and disciplinary history? Historian as high school history teacher. In J. Brophy (Ed.), *Advances in Research on Teaching: Teaching and Learning History* (Vol. 6, pp. 257–289). JAI Press.

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). Lawrence Erlbaum Publisher. https://psycnet.apa.org/record/2006-07986-023

VanSledright, B. A., & Reddy, K. (2014). Changing epistemic beliefs? An exploratory study of cognition among prospective history teacher. *Revista Tempo E Argumento, 6*(11), 28–68. http://dx.doi.org/10.5965/2175180306112014028

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

VanSledright, B., & Maggioni, L. (2016). Epistemic cognition in history. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of Epistemic Cognition (*pp.128-146). Routledge. https://doi.org/10.4324/9781315795225

Virta, A. (2001). Becoming a history teacher: observations on the beliefs and growth of student teachers. *Teaching and Teacher Education, 18*(6), 687–698. https://doi.org/10.1016/S0742-051X(02)00028-8

Voet, M., & De Wever, B. (2016). History teachers' conceptions of inquiry-based learning, beliefs about the nature of history, and their relation to the classroom context. *Teaching and Teacher Education, 55*, 57–67. http://dx.doi.org/10.1016/j.ijer.2017.03.003

Voet, M., & De Wever, B. (2017a). Preparing pre-service history teachers for organizing inquiry-based learning : the effects of an introductory training program. *Teaching and Teacher Education, 63*, 206–217. https://doi.org/10.1016/j.tate.2016.12.019

Voet, M., & De Wever, B. (2017b). History teachers' knowledge of inquiry methods: an analysis of cognitive processes used during a historical inquiry. *Journal of Teacher Education*, *68*(3), 312–329. https://doi.org/0.1177/0022487117697637

Voet, M., & De Wever, B. (2019). Teachers' adoption of inquiry-based learning activities: The importance of beliefs about education, the self, and the context. *Journal of Teacher Education, 70*(5), 423–440. http://10.1177/0022487117751399

Wansink, B. G. J., Akkerman, S. F., & Wubbels, T. (2016). The certainty paradox of student history teachers: Balancing between historical facts and interpretation. *Teaching and Teacher Education, 56*, 94–105. http://dx.doi.org/10.1016/j.tate.2016.02.005

Wansink, B. G. J., Akkerman, S. F., Vermunt, J. D., Haenen, J. P. P., & Wubbels, T. (2017). Epistemological tensions in prospective Dutch history teachers' beliefs about the objectives of secondary education. *The Journal of Social Studies Research, 41*(1), 11–24. http://dx.doi.org/10.1016/j.jssr.2015.10.003

Weinmann, C. (2018). Measuring political thinking: Development and validation of a scale for "deliberation within", *Political Psychology, 39*(2), 365–380. *https://doi.org/10.1111/pops.12423*

Wigfield, A., & Eccles, J. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*, 68–81. https://doi.org/10.1006/ceps.1999.1015

Wiley, J., Griffin, T., Steffens, B., & Britt, M. A. (2020). Epistemic beliefs about the value of integrating information across multiple documents in history. *Learning and Instruction, 65,* 101–266. https://doi.org/0.1016/j.learninstruc.2019.10126

Wilke, M., & Depaepe, F. (2019). Teachers and historical thinking. An exploration of the relationship between conceptualization, beliefs and instructional practices among Flemish history teachers. *International Journal for History and Social Sciences Education, 4,* 101-135.

Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2022a). The interplay between historical thinking and epistemological beliefs: a case study with history teachers in Flanders. *Historical Encounters, 9*(1), 196-219. https://doi.org/10.52289/hej9.111

Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2022b). Fostering historical thinking and democratic citizenship? A cluster randomized controlled intervention study. *Contemporary Educational Psychology, 71*, 102-115.. https://doi.org/10.1016/j.cedpsych.2022.102115

Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. *(under review).* Exploring differential effects of an intervention on historical inquiry tasks. A qualitative analysis of 12th grade students' progress.

Wils, K., Schampaert, A., Clarebout, G., Cools, H., Albicher, A. & Verschaffel, L. (2011). Past and present in contemporary history education. An exploratory empirical research on prospective history teachers. *Yearbook of the International Society for History Didactics, 32*, 217–236. https://jhec.wochenschau-verlag.de/wp-content/uploads/sites/23/2021/05/issue_2011.pdf#page=217

Wilschut, A. (2010). History at the mercy of politicians and ideologies: Germany, England, and the Netherlands in the 19th and 20th centuries. *Journal of Curriculum Studies*, *42*(5), 693–723. https://doi.org/10.1080/00220270903049446

Wilschut, A., Van Straaten, D., & Van Riessen, M. (2012). *Geschiedenisdidactiek. Handboek voor de vakdocent*. Coutinho.

Wineburg, S. & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record, 121* (11), 1-40. https://stacks.stanford.edu/file/druid:yk133ht8603/Wineburg%20McGrew_Lateral%20Reading%20and%20the%20Nature%20of%20Expertise.pdf

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

Wineburg, S. (2000). Making historical sense. In P. Stearns, P. Seixas, & S. Wineburg (Eds.), *Knowing, teaching and learning history. National and international perspectives* (pp. 306-325). New York University Press.

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

Wineburg, S., & Reisman, A. (2015). Disciplinary literacy in history: a toolkit for digital citizenship. *Journal of Adolescent & Adult Literacy, 58* (8), 636-639. https://www.jstor.org/stable/44011195

Yeager, E., & Davisz, O. (1996). Classroom teachers' thinking about historical texts: An exploratory study. *Theory & Research in Social Education*, *24*(2), 146–166. http://doi.org/10.1080/00933104.1996.10505774

Yilmaz, K. (2008). Social studies teachers' conceptions of history: Calling on historiography. *Journal of Educational Research, 101*(3), 37–41. http://doi.org/10.3200/JOER.101.3.158-176

Young, K. M., & Leinhardt, G. (1998). Writing from Primary Documents: A Way of Knowing in History. *Written Communication, 15*(1), 25–68. https://doi.org/10.1177/0741088398015001002

Output

Output integrated in this dissertation

International, peer-reviewed publications

- Wilke, M. & Depaepe, F. (2019). Teachers and historical thinking. An exploration of the relationship between conceptualization, beliefs and instructional practices among Flemish history teachers, *International Journal for History and Social Sciences Education, 4*, 101-135.
- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2022a). The interplay between historical thinking and epistemological beliefs: a case study with history teachers in Flanders. *Historical Encounters, 9*(1), 190-214. https://doi.org/10.52289/hej9.111
- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2022b). Fostering democratic citizenship through historical thinking? A cluster randomized controlled intervention study. *Contemporary Educational Psychology, 71*, 102-115. https://doi.org/10.1016/j.cedpsych.2022.102115
- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. Fostering secondary students' historical thinking. A design study in Flemish history education. *Manuscript under review.*
- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K.. Exploring differential effects of an intervention on historical inquiry tasks. A qualitative analysis of 12th grade students' progress. *Manuscript under review.*
- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. Exploring the possibility of transfer from historical to contemporary inquiry tasks. A qualitative analysis of the effects of an intervention in 12th grade history education. *Manuscript under review.*

Academic conference contributions

- Wilke, M., Van Nieuwenhuyse, K., & Depaepe, F. (2022). The interplay between teachers' epistemological beliefs, conceptualization of historical thinking and historical thinking practices: a case-study with history teachers in Flanders. *History teachers' epistemic considerations.* Umeå University, Sweden.
- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2022). Fostering historical inquiry competences and democratic citizenship. An intervention study. *EARLI*

SIG 20 and 26: Dialogue, inquiry and argumentation: shaping the future(s) of education. Utrecht, The Netherlands.

- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2022). Can historical thinking foster democratic citizenship? A cluster randomized controlled intervention study. *EARLI SIG 13: Creating Democratic Societies – Making Democracy Sustainable*. University of Agder, Norway.
- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2020). Exploring the interplay between teachers' beliefs and practices regarding historical thinking. *Didactifen 2020: Identifier, modéliser et surmonter les obstacles liés à l'apprentissage*. Luik, Belgium.
- Wilke, M., & Van Nieuwenhuyse, K. (2019). The interplay between historical thinking and epistemological beliefs: A case-study with history teachers in Flanders. *EARLI SIG 20 and 26: Enhancing historical reasoning: Measuring and developing students' epistemological understanding.* Amsterdam, Netherlands.
- Wilke, M., & Van Nieuwenhuyse, K. (2019). Fostering students' historical thinking through teaching migration history: teachers' beliefs and practices. *International Society for History Didactics: Migration and History Education*. Tutzing, Germany.

Professionally oriented output and presentations

- Castro Montes, E., Goddeeris, I., Rodríguez García, M., Schreurs, J., Van Nieuwenhuyse, K., & Wilke, M. (Forthcoming). *Dekolonisering in verleden en heden. Postkoloniale en subalterne perspectieven voor geschiedenisonderwijs.* Leuven University Press.
- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2022). Historisch denken in het geschiedenisonderwijs: een kwalitatief onderzoek naar de relatie tussen opvattingen en praktijken van leraren. *Hermes: Tijdschrift van de Vlaamse Leraren Geschiedenis, 26*(3), 6-11.
- Wilke, M. & Van Nieuwenhuyse, K. (2021). Lespakket "Historisch denken over dekolonisatie na 1945". https://www.arts.kuleuven.be/outreach/geschiedenis/lespakket-dekolonisatie-na-1945 (pdf)
- Wilke, M., & Van Nieuwenhuyse, K. (2018). Historisch denken over migratiegeschiedenis: opvattingen van leerkrachten en effecten op disciplinair denken en burgerschap. *Dag van de Belgische migratiegeschiedenis.* Kadoc, Leuven.

Other output

International, peer-reviewed publications

- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2021). Students' understanding and perception of agency in past and present. *International Journal for History and Social Sciences Education*, *2018*(3), 39-62.
- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2019). Teaching about historical agency. An intervention study examining changes in students' understanding and perception of agency in past and present. *International Journal of Research on History Didactics, History Education, and Historical Culture*, *5*(1), 53-80.

Academic conference contributions

- Wilke, M., & Van Nieuwenhuyse, K. (2019). Fostering students' historical thinking through teaching migration history: teachers' beliefs and practices. International Society for History Didactics: *Migration and History Education*. Tutzing, Germany.
- Wilke, M., Van Nieuwenhuyse, K., & Depaepe, F. (2019). Een empirisch onderzoek naar het begrip van leerlingen over agency. *Historicidagen - Inclusieve geschiedenis*. Groningen, The Netherlands.
- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2019). Teaching historical agency. Examining changes in students' perception of agency in past and present. *EARLI: Thinking Tomorrow's Education: Learning from the past, in the present and for the future.* Aachen, Germany.

Professionally oriented output and presentations

- Van Nieuwenhuyse, K., & Wilke, M. (2020). Une refonte des programmes d'histoire en Flandre. *Éduquer*, *158*, 23-24.
- Van Nieuwenhuyse, K., & Wilke, M. (2020). History education in Belgium/Flanders since 1945. Between a national and a global scope: Whose past, what for and for whom?. *Bulletin du CREAS*, *7*, 65-76. https://www.usherbrooke.ca/creas/fileadmin/sites/creas/documents/Publications/Bulletin_du_CREAS/7/09-CREAS_Bulletin_7_Van_Nieuwenhuyse-et-Wilke.pdf
- Wilke, M., & Van Nieuwenhuyse, K. (2021). Lesmodule: Armoede in België sinds 1800. Een agency-perspectief. https://www.arts.kuleuven.be/outreach/geschiedenis

- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2020). Agency in geschiedenisonderwijs (deel 3): effecten van expliciete instructie op leerlingen en hun visie op agency in verleden en heden. *Hermes: Tijdschrift van de Vlaamse Leraren Geschiedenis*, *24*(1), 24-29.
- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2019). Agency in geschiedenisonderwijs (deel 2): spontane opvattingen van leerlingen over agency in verleden en heden. *Hermes: Tijdschrift van de Vlaamse Leraren Geschiedenis*, *23*(3), 27-32.
- Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2019). Agency in geschiedenisonderwijs: belang en conceptualisering. *Hermes: Tijdschrift van de Vlaamse Leraren Geschiedenis*, *23*(2), 28-32.
- Wilke, M., & Van Nieuwenhuyse, K. (2019). Design van een wetenschappelijke onderzoekssetting: casestudie "agency in geschiedenisonderwijs". Geel.
- Wilke, M., & Van Nieuwenhuyse, K. (2019). De plaats van burgerschapsvorming in het geschiedenisonderwijs. (Hoe) kan historisch denken bijdragen tot burgerschapsvorming?. In *School ♡ onderzoek. Interactieve inspiratiedag voor leraren en onderzoekers*. Brussel: Departement Onderwijs en Vorming, Vlor.

Collaboration agreement

Contribution of authors in the various articles included in the dissertation:

Chapter 2

Based on: Wilke, M. & Depaepe, F. (2019). Teachers and historical thinking (in press). An exploration of the relationship between conceptualization, beliefs and instructional practices among Flemish history teachers', *International Journal for History and Social Sciences Education, 4,* 101-135.

Marjolein Wilke reviewed the literature, designed the initial data collection instrument, developed the coding scheme, gathered, coded and analyzed the data and wrote the initial manuscript, as well as revisions. Fien Depaepe and Karel van Nieuwenhuyse supervised the research and provided input on the conceptualization of the research, the data collection instrument and data analysis. They provided feedback on earlier versions of the manuscript.

Chapter 3

Based on: Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2022). The interplay between historical thinking and epistemological beliefs: a case study with history teachers in Flanders. *Historical Encounters, 9*(1), 190-214. https://doi.org/10.52289/hej9.111

Marjolein Wilke reviewed the literature, designed the data collection instrument and coding scheme, gathered, coded and analyzed the data and wrote the initial manuscript. Fien Depaepe and Karel Van Nieuwenhuyse supervised the research, providing feedback and input on the design of the study and the data collection instrument and reviewed the data analysis. They also provided feedback on various versions of the manuscript.

Chapter 4

Based on: Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. Fostering secondary students' historical thinking. A design study in Flemish history education. *Manuscript under review.*

Marjolein Wilke reviewed the literature, designed and revised the lesson series, with support from various educational and disciplinary experts, and wrote the initial manuscript. Fien Depaepe and Karel Van Nieuwenhuyse supervised the research and the design of the lesson series. They provided input on the design principles, the structure and content of the lesson series and on interim versions of the lessons. They also reviewed earlier versions of the manuscript.

Chapter 5

Based on: Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2022). Fostering democratic citizenship through historical thinking? A cluster randomized controlled intervention study. *Contemporary Educational Psychology, 71*, 102-115. https://doi.org/10.1016/j.cedpsych.2022.102115

Marjolein Wilke reviewed the literature, designed the intervention, study and the measurement instruments and organized the implementation of study. She collected and coded the data, assisted by students' in the educational master's program. She analyzed the data and wrote the initial manuscript. Fien Depaepe and Karel Van Nieuwenhuyse supervised the research, providing essential input on the design of the study, the measurement instruments and the data analysis. They also provided feedback on the interpretation of the results of the data analysis and on earlier versions of the manuscript. Wim van Den Noortgate provided advise on the statistical analyses.

Chapter 6

Based on: Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K.. Exploring differential effects of an intervention on historical inquiry tasks. A qualitative analysis of 12th grade students' progress. *Manuscript under review.*

Marjolein Wilke reviewed the literature, developed the coding scheme, coded and analyzed the essay data and wrote the manuscript. Fien Depaepe and Karel Van Nieuwenhuyse supervised the research. They provided feedback on the design of the study, the coding scheme, the data analysis and on various versions of the manuscript. This study is a follow-up study of the intervention study described in chapter 5. Some parts of the method section therefore occur both in this chapter and in chapter 5.

Chapter 7

Based on: Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K.. Exploring the possibility of transfer from historical to contemporary inquiry tasks. A qualitative analysis of the effects of an intervention in 12th grade history education. *Manuscript under review.*

Marjolein Wilke reviewed the literature, developed the coding scheme, coded and analyzed the essay data and wrote the manuscript. Fien Depaepe and Karel Van Nieuwenhuyse supervised the research. They provided feedback on the design of the study, the coding scheme, the data analysis and on various versions of the manuscript. This study is a follow-up study of the intervention study described in chapter 5. Some parts of the method section therefore occur both in this chapter and in chapter 5.