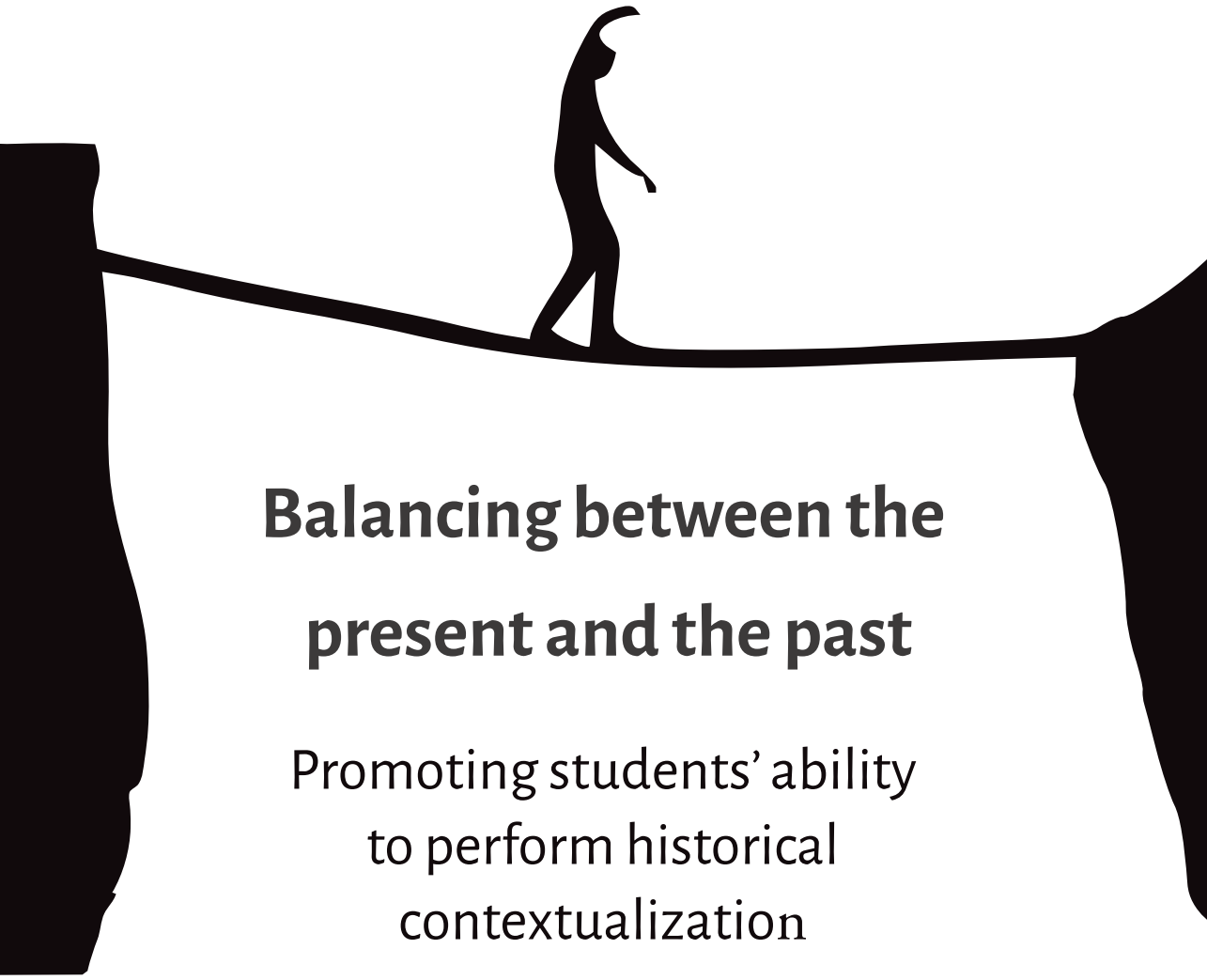


Balancing between the present and the past

Promoting students' ability to perform historical contextualization

Tim Huijgen

Balancing between the present and the past

Promoting students' ability to perform historical contextualization

Timothy David Huijgen

ISBN (printed version): 978-94-034-0921-4
ISBN (digital version): 978-94-034-0920-7

Lay-out: Ferdinand van Nispen, citroenvlinder-dtp.nl, *my*-thesis.nl
Print: ProefschriftMaken, Vianen

This research was funded by the Netherlands Organisation for Scientific Research (NWO) under grant number 023.001.104.

university of groningen

Balancing between the present and the past

Promoting students' ability to perform historical contextualization

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Thursday 18 October 2018 at 11.00 hours

by

Timothy David Huijgen
born on 24 December 1985
in Smallingerland

Supervisors
Prof. W.J.C.M. van de Grift
Prof. C.A.M. van Boxtel

Co-supervisor
Dr. P. Holthuis

Assessment Committee
Prof. J.A. van der Schee
Prof. K. van Veen
Prof. D.J. Wolffram

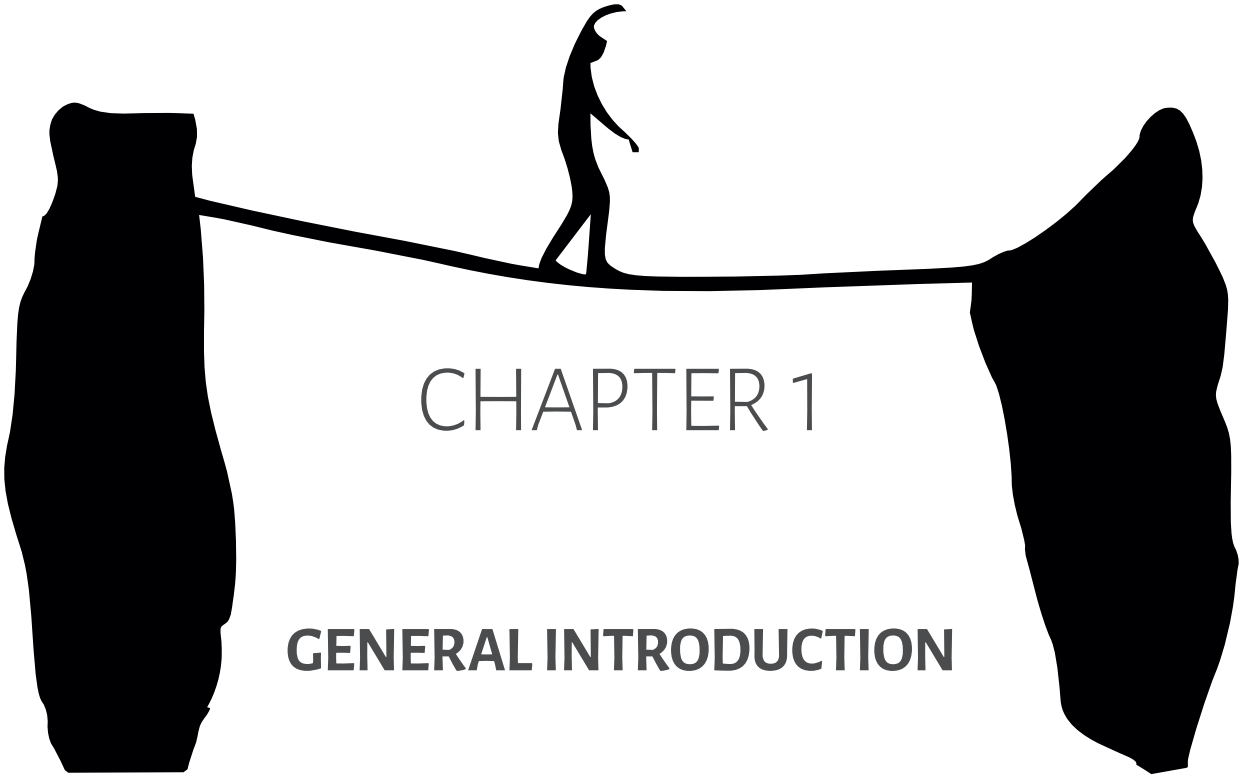
TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

An alert-looking boy, apparently at the head of the class, asked me the obligatory question: 'But how come you didn't escape?' I briefly explained to him what I have written here; not quite convinced, he asked me to draw a sketch of the camp on the blackboard indicating the location of the watch towers, the gates, the barbed wire, and the power station. I did my best, watched by thirty pairs of intent eyes. My interlocutor studied the drawings for a few instants, asked me for a few further clarifications, then he presented to me the plan he had worked out: here, at night, cut the throat of the sentinel; then, put on his clothes; immediately after this, run over there to the power station and cut off the electricity, so the search lights would go out and the high-tension fence would be deactivated; after that I could leave without any trouble. He added seriously: 'If it should happen to you again, do as I told you; you'll see that you'll be able to do it'.

Primo Levi, Holocaust survivor, in the book *The Drowned and the Saved* (1988), page 177-178.



CHAPTER 1

GENERAL INTRODUCTION

This chapter first presents the aim and theoretical framework of the thesis. Next, the research questions of the thesis are presented, and the chapter concludes with an overview of the subsequent chapters.

1.1 Aim of the thesis

"That is just inhuman. Such a girl should be at school and not be forced to marry someone she does not know and love." This response was said by Emma, one of my 12-year-old students, when I provided my students with a historical source describing the medieval marriage of a 13-year-old girl to a 36-year-old knight. After this lesson, I noticed similar reactions from my students when teaching other historical topics. Students could not understand why Germans in the 1930s voted for "a man who killed millions of people and loved violence" or that people in the 19th century thought they would suffocate when travelling in the first trains at speeds of more than 20 miles per hour. When I told students that the Dutch Republic exchanged the colony of New Netherland (currently New York City) for Suriname in the 17th century, they responded, "Man, giving up a world class city; that is just stupid."

Scholars worldwide agree that history education should aim at promoting students' ability to perform historical thinking and reasoning rather than training students to memorize as many historical facts as possible (e.g., Lévesque, 2008; Levstik & Barton, 2011; Van Drie & Van Boxtel, 2008). However, Wineburg (2001) noted that historical thinking is an "unnatural act" since people automatically tend to view the past from a present-oriented perspective. This *presentism* often results in misunderstanding historical phenomena and historical agents' actions (Barton & Levstik, 2004; Lee & Ashby, 2001). This misunderstanding is similar to what occurred in the case of Emma and with many of my other students: they were not able to explain and interpret the historical events and historical agents' decisions under study because they viewed the past from their own current beliefs, values, and knowledge.

Historical contextualization can help students such as Emma to become aware of their present-oriented perspectives. Historical contextualization is the ability to situate historical phenomena or historical agents' actions in a temporal, spatial, and social context to describe, explain, compare, or evaluate them (Van Boxtel & Van Drie, 2012). Without this ability, students often misunderstand historical phenomena and historical agents' actions (Reisman & Wineburg, 2008; VanSledright, 2001; Wineburg, 2001); therefore, the ability is considered a key component of historical thinking and reasoning (Seixas & Morton, 2013; Van Drie & Van Boxtel, 2008).

Despite the importance of historical contextualization, history education professionals are faced with three major problems: (1) the lack of standardized instruments that can provide insight into how students of different ages and educational levels perform historical contextualization, (2) the absence of domain-specific observation instruments that can provide insight into how teachers promote historical contextualization in classrooms, and (3) the lack of classroom materials that can help teachers to improve their students' ability to perform historical contextualization.

The first problem focuses on measuring students' ability to perform historical contextualization. Students often view the past from a contemporary or present-oriented perspective, for example, when debating the issue if the United States will always remain the most powerful country in the world (Foster, Ashby, & Lee, 2008) or when trying to understand involuntary teenage marriages during the 15th century (Angvik & Von Borries, 1997). Research in social psychology indicates that youngsters especially suffer from the *curse of knowledge*, a cognitive bias that makes it difficult for students who have more knowledge to think from the perspective of less informed people (Birch & Bloom, 2007). Despite the fact that research has been conducted on how historians and students perform historical contextualization (e.g., Van Boxtel & Van Drie, 2012; Wineburg, 1998), there are not many standardized instruments available to test their ability to do so. The lack of such instruments could result in a shortage of systematic assessments of students' progression in historical reasoning competencies (Peck & Seixas, 2008). Several scholars therefore argue for new assessment formats to make sense of how students learn history and how they improve in learning history (e.g., Breakstone, 2014; Ercikan & Seixas, 2015; Reich, 2009; VanSledright, 2013). The first challenge is therefore to develop and test instruments that can examine student differences in the ability to perform historical contextualization.

The second problem is that not much is known about how history teachers promote historical contextualization in their classrooms. Since research indicates that teachers seem to struggle to develop meaningful and activating learning tasks that promote students' historical reasoning competencies (e.g., Barton & Levstik, 2003; Reisman, 2015; Saye & Social Studies Inquiry Research Collaborative, 2013; VanSledright, 2008), an observation instrument that can examine teachers' instructions with regard to historical contextualization is essential. Therefore, a second challenge is the construction of such an instrument and the use of it to explore how history teachers promote historical contextualization in classrooms.

The third problem is the need for practical tools that can help teachers promote historical contextualization among their students. Classroom materials that are based upon research are missing because design and intervention studies on teaching and learning historical reasoning competencies such as historical contextualization are scarce. This situation has resulted in a call from scholars to develop and test domain-specific practices and tools that can promote historical thinking and reasoning (e.g., Fogo, 2014; Grant & Gradwell, 2010; Reisman & Fogo, 2016; Reisman et al., 2018). The third and final challenge is therefore to develop and to test classroom materials that promote historical contextualization among students.

1.2 Theoretical framework

The central topic of the thesis is historical contextualization. This section, therefore, first defines this ability and describes its important role in historical thinking and reasoning processes. Subsequently, this section presents a theoretical framework for teaching historical contextualization. That theoretical framework is used in the different studies of this thesis.

1.2.1 Defining historical contextualization

Historical contextualization is about understanding the differences between past and present (Seixas & Peck, 2004). It requires an understanding of the social, political, and cultural values of the time period under investigation and knowledge of the events leading up to the historical situation as well as other relevant events occurring at the same time (Endacott & Brooks, 2013). By no means should historical contextualization lead to relativism among students, such as the justification of controversial actions by people in the past. Rather, it should help students to make reasoned ethical judgements and to understand and explain historical phenomena and the actions of people (Seixas & Morton, 2013).

Some studies define historical contextualization as one of the heuristics that can be applied (in addition to corroboration, close reading, and sourcing) to examine historical sources (e.g., Baron, 2016; Nokes, Dole, & Hacker, 2007; Reisman, 2012a; Wineburg, 1998). However, in history education, it is possible to contextualize historical events, developments, sources, and agents' actions (Havekes, Coppen, Luttenberg, & Van Boxtel, 2012). Therefore, in this study, we use the definition of Van Boxtel and Van Drie

(2012) and visualize historical contextualization as an activity in which one situates phenomena and agents' actions in the context of time, historical locations, long-term developments, and specific events to give meaning to these phenomena and actions. When historical contextualization is used to compare, evaluate, interpret, or examine historical agents' actions, the term *historical perspective taking* (HPT) is often used (e.g., Hartmann & Hasselhorn, 2008; Seixas & Morton, 2013; Yeager & Foster, 2001). HPT focuses on understanding the perspectives of historical agents by considering their historical contexts (Seixas & Morton, 2013).

1.2.2 The role of historical contextualization in history education

Recently, several books on history education have been published (e.g., Carretero, Berger, & Grever, 2017; Chapman & Wilschut, 2015; Counsell, Burn, & Chapman, 2016). These books display a general view that the transmission of historical content knowledge should not be the sole aim of history education but students in history classrooms should also use this knowledge, for example, to evaluate historical sources, determine causes and consequences, and perform historical contextualization. In the literature, these competencies are often described as students performing *historical thinking* or *historical reasoning* (e.g., Lévesque, 2008; Van Boxtel & Van Drie, 2018; Wineburg, 2001). Van Drie and Van Boxtel (2008) presented an analytical framework of historical reasoning that comprises six interrelated components:

- asking historical questions,
- historical contextualization,
- using substantive concepts (concepts referring to historical phenomena, structures, persons, and periods),
- using meta-concepts and related strategies (concepts and strategies referring to the methods used by historians to investigate and describe historical processes),
- using sources,
- and argumentation.

These six components should enable students to reach justifiable conclusions about (1) processes of change and continuity, (2) causes and consequences of historical phenomena, and (3) differences and similarities between historical phenomena or periods (Van Boxtel & Van Drie, 2018).

The ability to perform historical contextualization is needed to apply all three types of historical reasoning. First, continuity and change are very difficult to interpret without considering the historical context of the different periods or historical events under study. For example, the shift from a preindustrial society to an industrial society in England in the 18th century can only be explained when the historical contexts of both periods are reconstructed and compared. Second, interpreting the causes and consequences of historical developments and actions of people is not possible when events or actions of people are not placed in a broader historical context. For example, the shot fired by Gavrilo Princip in 1914 loses all meaning when this action is not placed in the context of rising nationalism, alliances, and the imperialism of European countries of the 19th and 20th century. Third, to examine and compare differences and similarities in the past, a historical context of developments and phenomena should first be created. For example, when examining and comparing the concept of trade throughout history, it must be understood that trade within the Roman Empire had a different meaning (e.g., monetary economy, large-scale trading) than trade in the early Middle Ages in Western Europe (e.g., manorialism, self-sufficiency). Teachers should therefore teach students to consider particular policies, institutions, worldviews, and circumstances that shape a given moment in time to identify enduring themes and patterns (Reisman & Wineburg, 2008).

Apart from historical contextualization being considered a key component of historical thinking and reasoning in secondary history education (e.g., Seixas & Morton, 2013; Van Drie & Van Boxtel, 2008) and an essential skill for historians (e.g., Gaddis, 2002; Tully, 1988), it is also considered a possible contributor to instilling democratic citizenship in students. For example, it can provide background and context for democratic debate in post-conflict societies (McCully, 2012) and prepares students to participate in a pluralistic society in which people hold differ opinions (Barton & Levstik, 2004). Moreover, the ability to perform historical contextualization is important in other school subjects, such as in science classrooms when discussing the scientific development of the atomic bomb, in English literature classrooms when discussing Mark Twain's novel *Huckleberry Finn*, or in art classrooms where historical contextualization is needed to examine and interpret artworks (e.g., Nikitina, 2006; Pauly, 2003). Furthermore, students in Dutch language classrooms need to consider the historico-literary context when reading and interpreting texts (Witte, 2008).

1.2.3 A framework for teaching historical contextualization

In their review of history education research, VanSledright and Limón (2006) outlined that in an average history classroom, the teacher does most of the talking. Lecturing and story-telling often dominate the classroom. Recent research seems to confirm this finding (e.g., Saye & Social Studies Inquiry Research Collaborative, 2013; VanSledright, 2011). In such history classrooms, historical reasoning might not be promoted, since reasoning requires active participation and input from the students (Van Boxtel & Van Drie, 2017). Based on this view, and with regard to the knowledge that students learn most when actively engaged in learning tasks (Bransford, Brown, & Cocking, 2000), we view teaching historical contextualization as an interactive process in which both students and teachers participate and engage in dialog.

Based on a review of the literature on historical contextualization, teaching historical contextualization in this thesis is conceptualized as four interrelated components: (1) reconstructing a historical context, (2) raising awareness of present-oriented perspectives, (3) enhancing historical empathy, and (4) creating opportunities for students to practice historical contextualization to enable historical reasoning. All components should occur in interactions between teachers and students. The components are presented in Figure 1.

The first component is reconstructing the historical context. To perform historical contextualization successfully, the historical context of a phenomenon must be reconstructed including knowledge of chronological, spatial, socio-political, socio-economic, and socio-cultural frames of reference (De Keyser & Vandepitte, 1998). A frame of reference is a knowledge base for interpreting and dating historical phenomena (Lee & Howson, 2009). Without background knowledge of historical phenomena, students cannot grasp the "sense of a period," as Dawson (2009) noted. Teachers can explore and explain different frames of reference with students and teach them to use these frames to reconstruct a historical context. It is important to consider all frames when examining historical phenomena and agents' actions. The chronological frame includes knowledge of time and period and chronological knowledge of significant events and developments. Students must situate phenomena and historical agents' actions in time to be able to explain, compare, or evaluate these phenomena and acts (Van Boxtel & Van Drie, 2012). The spatial frame focuses on knowledge of (geographical) locations and scale. For example, when students lack spatial context knowledge and think that Suriname is a country in

Africa, the concept of triangular trade in the early modern period is misunderstood. Social frames include knowledge of human behavior and social conditions of life as well as knowledge of economic and political developments. Without this knowledge, students are not able to interpret historical phenomena and agents' actions (Reisman & Wineburg, 2008).

The second component is raising awareness of students' possible present-oriented perspectives. Avoiding presentism in history education is necessary to point out to students the differences and connections between the past and the present (Seixas & Morton, 2013). When students view the past from a present-oriented perspective, they do not succeed in explaining and understanding historical phenomena or agents' actions (Lee & Ashby, 2001). A promising approach to raising students' awareness of their possible present-oriented perspectives is by creating a *cognitive conflict*. These conflicts occur when incompatible ideas exist simultaneously in a person's mind or when information that is received does not seem consistent with what one already knows (Johnson & Johnson, 2009). In history education, for example, this could be done by presenting a historical situation that students consider "strange" (e.g., Havekes et al., 2012; Logtenberg, 2012). Teachers could, for example, ask students to explain why there was child labor in the Netherlands in the 19th century.

The third component is enhancing historical empathy. Different scholars agree that historical empathy and historical contextualization are closely related (e.g., Cunningham, 2009; Endacott & Brooks, 2013). Historical empathy can help students to see and judge the past on its own terms by attempting to understand the historical agents' frames of reference and actions (Yilmaz, 2007). We centralize two approaches to promote historical empathy: (1) using affective connections and (2) examining the role and position of the historical agent. Affective connections are considerations of how historical agents' experiences, situations, or actions may have been influenced by their affective response based on a connection made to students' own similar yet different life experiences (Endacott & Brooks, 2013). Seixas and Morton (2013) talked about *universals*: using commonalities in students' and historical agents' emotions to infer how people in the past thought and felt. A more cognitive approach investigates the role and the position of a historical agent, which includes understanding another's prior lived experience, principles, positions, attitudes, and beliefs. This method also provides more insight into how a historical agent might have thought and behaved in a particular situation (Bermúdez & Jaramillo, 2001; Endacott & Brooks, 2013; Hartmann & Hasselhorn, 2008).

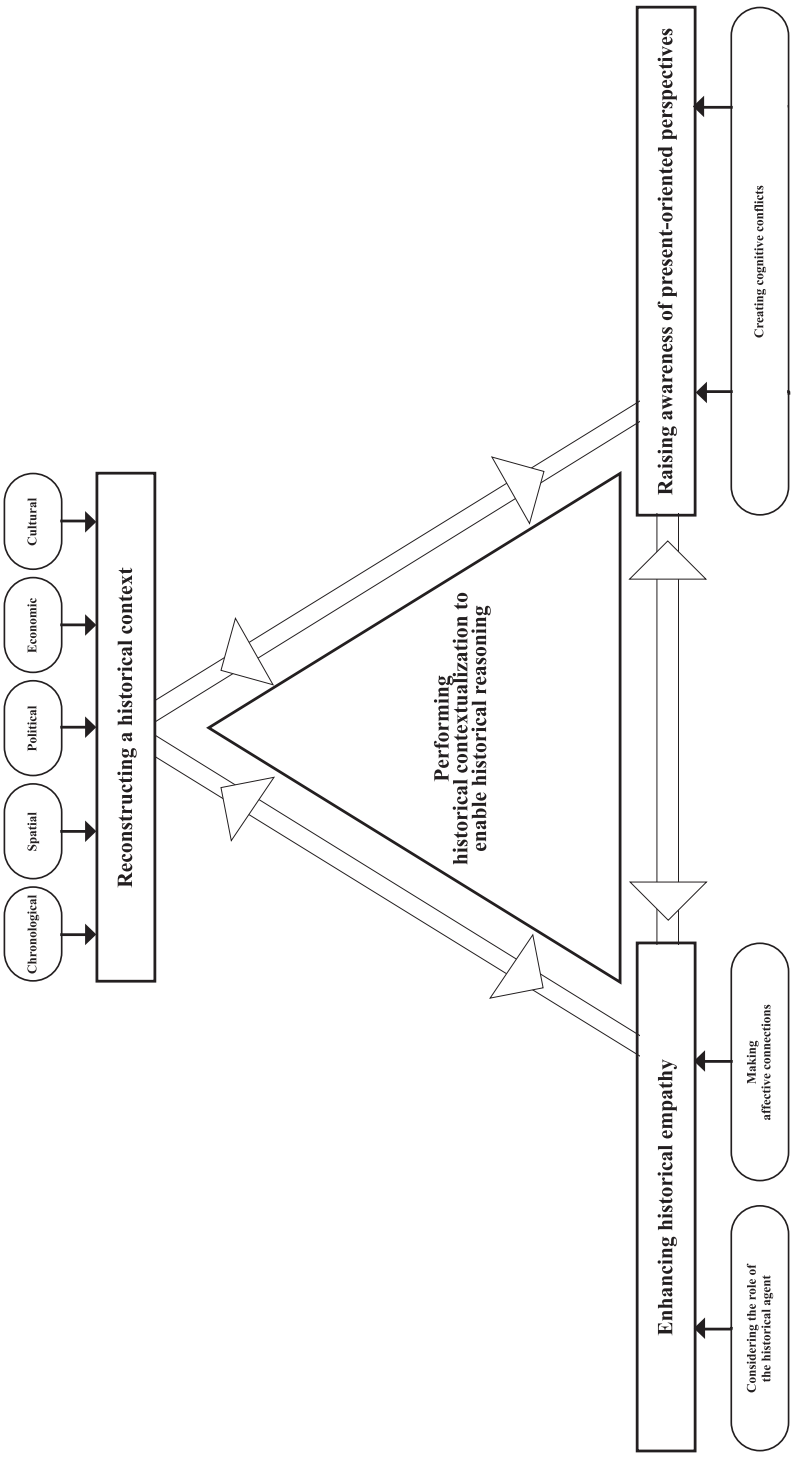


Figure 1. Framework for teaching historical contextualization

However, it is not enough to promote students' awareness of their present-oriented perspectives, reconstruct a historical context, and enhance historical empathy. The fourth and final component is that teachers should create opportunities for students to practice historical contextualization to enable historical reasoning. An example of such a task is asking students to explain why a particular German person in 1930 might have voted for the Nazi Party of Hitler or why the Dutch Republic exchanged New Netherland for Suriname in the 17th century. When historical contextualization is used to examine such questions, it becomes meaningful because it helps to explain and interpret historical phenomena (Van Boxtel & Van Drie, 2018).

1.3 Research questions

Based on the three problems that we described in the first section of this chapter, we formulated the following main research question for this thesis: *How can students' ability to perform historical contextualization be promoted?* To answer this question, we formulated five research questions.

The first challenge was to develop and test instruments that can examine students' ability to perform historical contextualization. This task was examined in two studies that focused on students' ability to contextualize historical agents' actions. The following questions were researched:

1. How can we measure elementary and secondary school students' ability to contextualize historical agents' actions?
   (Study 1)
2. How successfully can secondary school students contextualize historical agents' actions?
   (Study 2)

The second challenge was to develop an observation-instrument to observe how history teachers promote historical contextualization in classrooms. Two studies were therefore conducted with the following questions:

3. What instrument can be used to observe how history teachers promote historical contextualization in classrooms?
   (Study 3)

4. How do history teachers promote historical contextualization in their classrooms?
   (Study 4)

The third and final challenge was to develop and test classroom materials that promote historical contextualization. This aspect resulted in two intervention studies answering the following question:

5. What are the effects of a lesson unit designed to promote secondary school students' ability to perform historical contextualization?
   (Studies 5 and 6)

1.4 Structure of the thesis

After the introduction to the thesis in this chapter, the **second chapter** discusses the benefits and limitations of two instruments intended to measure students' ability to contextualize historical agents' actions. Moreover, the second chapter presents information on how students of different ages and educational levels exhibit this ability. In this chapter, we view three components of the framework for teaching historical contextualization (reconstructing a historical context, avoiding presentism, and historical empathy) as necessary to explain and interpret historical agents' actions (historical perspective taking). This chapter is based on an article in the *European Journal of Psychology of Education* (Huijgen, Van Boxtel, Van de Grift, & Holthuis, 2014).

The **third chapter** examines how students contextualize historical agents' actions. How successful are they? What knowledge and strategies do students use to explain those actions? Which frames of reference are used the most? As in the second chapter, the three components of the framework are considered essential to achieve historical perspective taking. This chapter is based on an article in *Theory & Research in Social Education* (Huijgen, Van Boxtel, Van de Grift, & Holthuis, 2017).

The **fourth chapter** focuses on how teachers' instructions with regard to historical contextualization could be observed. In this chapter, the four components of the framework (reconstructing a historical context, avoiding presentism, historical empathy, and practicing historical contextualization to enable historical reasoning) are used to develop and test a subject-specific observation instrument called the

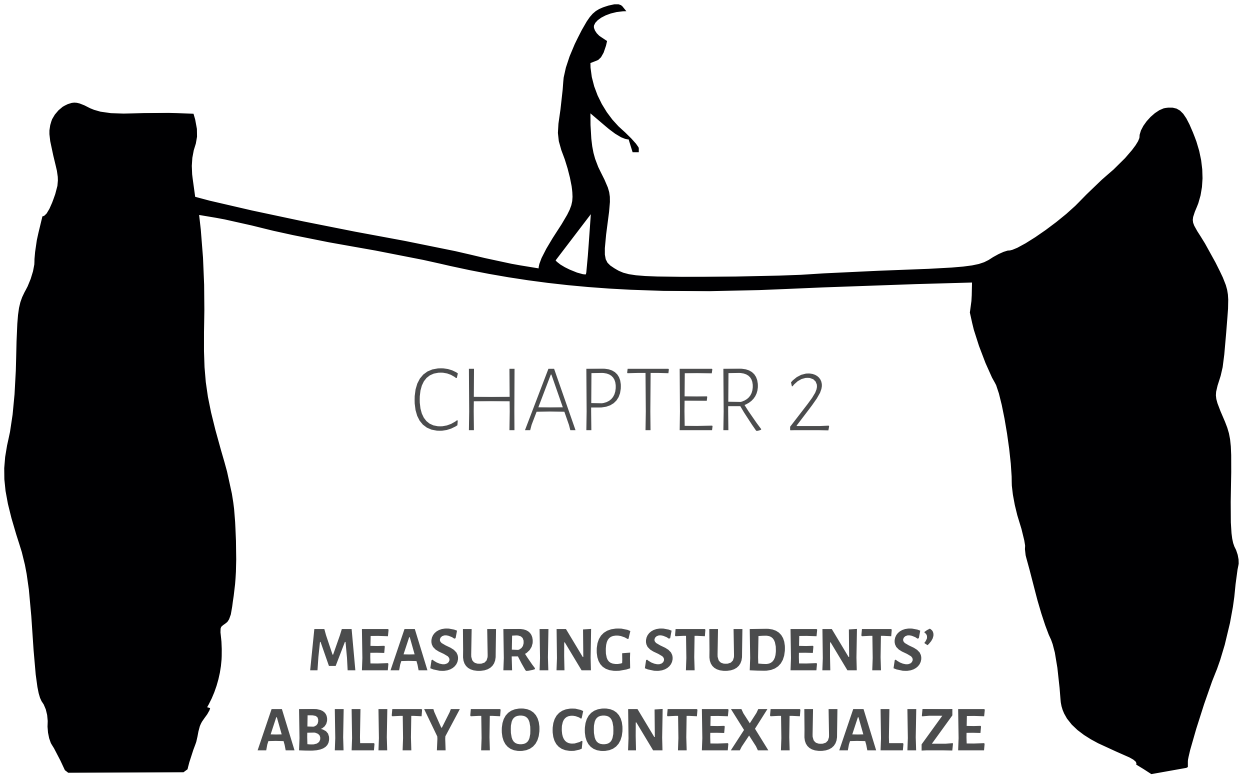
Framework for Analyzing the Teaching of Historical Contextualization (FAT-HC). The chapter also presents information on how many raters and lessons are needed to obtain a reliable image of how history teachers promote historical contextualization. This chapter is based on an article in the *European Journal of Psychology of Education* (Huijgen, Van de Grift, Van Boxtel, & Holthuis, 2017).

The **fifth chapter** explores how history teachers promote students' ability to perform historical contextualization in classrooms. Using the FAT-HC, eight history teachers were observed twice by five raters. Historical contextualization examples from these teachers are provided to examine teachers' behavior with regards to historical contextualization. What kind of instructions do teachers use and do not use? The four components of the framework are used to provide more specific insights into how teachers promote historical contextualization. This chapter is based on an article in *Educational Studies* (Huijgen, Holthuis, Van Boxtel, & Van de Grift, 2018).

The **sixth chapter** uses the four components of the framework to formulate four design principles that are used to develop a lesson unit on the 17th and 18th centuries. The effects of the lesson unit on students' ability to perform historical contextualization are explored through a quasi-experimental pre- and post-test design. Using additional qualitative methods, the strengths and weaknesses of the design principles are further explored. This chapter is based on an article in the *Journal of Curriculum Studies* (Huijgen, Van de Grift, Van Boxtel, & Holthuis, 2018).

Elaborating on the findings of the sixth chapter, the **seventh chapter** uses three components of the framework (reconstructing a historical context, avoiding presentism, and practicing historical contextualization to enable historical reasoning) to present a three-stage framework. This framework is used to develop a lesson unit on Cold War events. The effects of the lesson unit on students' ability to promote historical contextualization are explored through a quasi-experimental pre- and post-test design. This chapter is based on an article in the *British Journal of Educational Studies* (Huijgen, Holthuis, Van Boxtel, Van de Grift, & Suhre, 2018).

In the **eighth chapter**, the general conclusions of this thesis are discussed. Moreover, limitations are elaborated, directions for future research are provided, and practical implications for teachers and other education professionals are presented. Finally, the **ninth chapter** presents the Dutch summary of the thesis.



CHAPTER 2

MEASURING STUDENTS' ABILITY TO CONTEXTUALIZE HISTORICAL AGENTS' ACTIONS

This chapter considers two instruments for measuring students' ability to perform historical perspective taking (HPT) as a historical reasoning competency. The instruments have been tested for validity and reliability among 1,270 Dutch upper elementary and secondary school students, ranging in age from 10 to 17 years. One instrument offers effective validity and reliability and can map HPT performance among a large and heterogeneous student population. The results show that even upper elementary school students are capable of performing HPT. However, as students age, their ability to perform HPT increases. Differences regarding the ability to perform HPT were also found between educational levels. Pre-university students performed HPT more successfully compared to students at lower educational levels. The results of this study can be used to gain insight into the construct of HPT and into how historical reasoning competencies such as HPT can be measured. Furthermore, the results provide insight into how differences between students, such as age and educational levels, influence the performance of HPT.

This chapter is based on: Huijgen, T. D., Van Boxtel, C. A. M., Van de Grift, W. J. C. M., & Holthuis, P. (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. doi:10.1007/s10212-014-0219-4.

2.1 Introduction

During one observation of classroom practice, we heard a history teacher asking his students the following question: "Can you explain why people in Germany voted for Hitler in the 1930s?" Most students answered that they could not understand why anyone would vote for such a terrible and evil leader, who was responsible for the deaths of millions. Just one student in this class described the historical context of Germany in the 1930s, coming to the conclusion that some people may well have voted for Hitler in response to the poor economic circumstances, German anger over the Treaty of Versailles, and widespread calls for a strong leader. This last one student was the only one to display *historical perspective taking* (HPT).

Historical reasoning competencies including HPT have become increasingly important for learning history (Barton & Levstik, 2004; Haydn, Arthur, Hunt, & Stephen, 1997; Haydn & Counsell, 2002; Lévesque, 2008; Maggioni, Alexander, & VanSledright, 2004; Osborne, 2006; O'Reilly, 1991; Perfetti, Britt, & Georgi, 1995; Seixas & Morton, 2013; Spoehr & Spoehr, 1994; Van Drie & Van Boxtel, 2008; Wineburg, 2001; Yeager & Foster, 2001). Historical reasoning competencies therefore have been incorporated in the history curricula of many countries such as the USA, Canada, the UK, Australia, New Zealand, Belgium, Germany, Finland, and the Netherlands.

Despite the growing importance of historical reasoning competencies, valid and reliable large-scale measurement instruments for assessing these competencies are scarce. Rothstein (2004) noted that history teachers often assess only the factual background of history and not students' ability to perform historical reasoning. The reason for this, according to Rothstein, is the difficulty of constructing valid and reliable standardized tests. This difficulty is emphasized by Reich (2009), who was one of the few to attempt to measure historical reasoning competencies using multiple-choice items. However, he concluded that multiple-choice items merely tested history content, literacy, and test-wiseness but not important discipline-based thinking, such as HPT. Peck and Seixas (2008) noted that the focus of classroom assessment relies on factual recall and that, as a result, there is a lack of systematic assessment of students' progression in historical reasoning competencies. Students, teachers, and educational professionals might therefore have an uncertain grasp on what progress in history education means, as Haydn (2011) noted. Recently, Fordham (2013) and VanSledright (2013) also argued for new assessment formats, if educational professionals wish

to make sense of how students learn history and how they improve in it. Increasing numbers of research studies, projects, conferences, and books concentrate on the assessment of history education to gain insight into its benefits and problems (e.g., Breakstone, Smith, & Wineburg, 2013; Davies, 2011; Harris & Foreman-Peck, 2004; Martin, Maldonado, Schneider, & Smith, 2011; Seixas & Colyer, 2012; SERVE, 2006).

Our study should be placed in this context, and we took up the key challenge of constructing a reliable and valid measure instrument that could assess historical reasoning competencies within a large and heterogeneous student population and which was also time- and cost-effective. We focused on HPT because this student ability is crucial to learning history. Failing to perform HPT leads to important misunderstanding about the past (Barton & Levstik, 2004; Davis, 2001; Husbands, 1996; Lee & Ashby, 2001; Leinhardt, Beck, & Stainton, 1994; Lévesque, 2008; Seixas & Morton, 2013; Van Boxtel & Van Drie, 2012; Wineburg, 2001; Wineburg & Fournier, 1994; Yeager & Foster, 2001). Scholars also have argued that HPT can contribute to citizenship competencies because recognizing other people's views is necessary in a multicultural democracy (e.g., Barton & Levstik, 2004; Den Heyer, 2003).

Hartmann and Hasselhorn (2008) designed a measurement instrument that offers positive indicators for assessing students' ability of performing HPT. However, they tested their instrument only among a homogenous group of 170 tenth-grade German students (16 years old) and focused on only one historical topic. Our study focuses on testing the instrument format among students in a larger and more heterogeneous student population and with two different historical topics to map possible differences between students. In this study, we first present the theoretical framework, starting with the conceptualization of HPT and how it relates to historical reasoning. Subsequently, we look at what is already known about students' ability to perform HPT and focus on the opportunities and difficulties that exist for measuring HPT. Then, our research questions, method, results, conclusions, and discussion will be presented.

2.2 Theoretical framework

2.2.1 Historical perspective taking: a conceptualization

Without the ability to perform HPT, it is impossible to achieve historical reasoning and thinking (Lévesque, 2008; Seixas & Morton, 2013; Van Drie & Van Boxtel, 2008). Seixas and Peck (2004) conceptualize HPT as an understanding of the social, cultural, intellectual, and emotional setting that shaped people's lives and actions, and they emphasize the importance of being aware of the difference between the past and present. Hartmann and Hasselhorn (2008) follow the definition of Lee and Ashby (2001) and define HPT as the application of the knowledge that historical agents had particular perspectives on their world that affected their actions. Van Boxtel and Van Drie (2012) and Yeager and Foster (2001) talk about the application of the knowledge and understandings of the historical context and chronology.

Based on a review of the literature, we distilled three elements necessary for performing HPT successfully. First, the ability to perform historical contextualization was identified (e.g., Britt & Aglinskas, 2002; Doppen, 2000; Havekes, Coppen, Luttenberg, & Van Boxtel, 2012; Leinhardt & McCarthy Young, 1996; Nokes, Dole, & Hacker, 2007; Rouet, Favart, Britt, & Perfetti, 1997; Van Boxtel & Van Drie, 2012; Wineburg, 1998). Historical contextualization refers to building a context of circumstances or facts that surround the particular historical phenomenon to describe, compare, explain, or evaluate it (Van Drie & Van Boxtel, 2008; Wineburg, 1991). In history, it is possible to contextualize historical sources or historical phenomena, including persons, events, developments, or structures. In HPT, the focus is the contextualization of actions of people and groups in the past. Students can therefore use chronological, spatial, and socio-cultural frames of reference (De Keyser & Vandepitte, 1998).

Second, students need to exhibit historical empathy (e.g., Davis, 2001; Endacott, 2010; Lee & Ashby, 2001; Skolnick, Dulberg, & Maestre, 2004). Without the ability to imagine oneself in a situation that he or she is not likely to experience, the past remains an unopened book. However, historical empathy is not sympathy, as Eisenberg (2000) notes. Sympathy is compassion, sorrow, or concern for another person. Historical empathy focuses on identifying with people in the past based on historical knowledge to explain their actions in the past.

Third, students have to avoid *presentism*, the bias by which people assume that the same goals, intentions, attitudes, and beliefs existed in the past as they exist today (e.g., Barton, 1996; Barton & Levstik, 2004; Lee & Ashby, 2001; Seixas & Morton, 2013; Shemilt, 1983; Stahl, Hynd, Britton, McNish, & Bosquet, 1996; VanSledright & Afflerbach, 2000; Wineburg, 2001). The failure to perform HPT—and, therefore, the failure to explain, evaluate, or describe the past—often stems from this type of reasoning (Lee & Ashby, 2001; Wineburg, 2001). Its danger is explicitly mentioned in the American National Standards for History, which demands that students "avoid present-mindedness, judging the past solely in terms of the norms and values of today" (National Center for History in the Schools, 1996).

History education research has debated the extent to which HPT is an affective or cognitive achievement (e.g., Barton & Levstik, 2004; Davis, 2001; Endacott, 2010; Foster & Yeager, 1998). Some researchers claim that it is predominately a cognitive function (e.g., Foster, 1999; Lee & Ashby, 2001; Stern, 1998), and others claim that it is more an affective process (Riley, 1998; Skolnick et al., 2004). Although affective processes, such as connecting with known and familiar emotions of people in the past, may be at work during HPT, we consider it to be predominately a cognitive process in which students, based on historical evidence, perform historical contextualization and historical empathy and avoid presentism.

2.2.2 Addressing the different needs of students

Unfortunately, we know relatively little about which students suffer from presentism and which students can perform HPT successfully. In accordance with Piaget's theory of the stages of cognitive development, researchers, such as Hallam (1970), have concluded that historical thinking is not possible for people younger than 16 years of age. These students cannot be expected to cope with abstract concepts or investigation, analysis, and interpretation—all of which are elements required to perform HPT successfully. However, Brophy and VanSledright (1997) argue that fifth graders (ages 10–11 years) can overcome their tendencies toward presentism and other biases to identify and empathize with people from the past. A general consensus among scholars concurs that children are capable of historical reasoning and HPT much earlier than Hallam suggested (e.g., Barton, 1997; Foster & Yeager, 1999; Levstik & Smith, 1996; VanSledright, 2002).

Specific information about which students perform HPT successfully is still lacking, however. This is a great concern with regard to the tendency for classrooms and schools to become increasingly diverse (Forsten, Grant, & Hollas, 2002; McCoy & Ketterlin-Geller, 2004; Subban, 2006; Tomlinson, 2002; Tomlinson & Kalbfleisch, 1998). Teachers should therefore know their students' competency levels, such as for HPT, to adapt their teaching and to reshape history curricula to fit it to students' needs (Jonassen & Grabowski, 1993). However, one of the most important conclusions in the annual report of the Dutch Inspectorate of Education (2012) was that most teachers do have the basic skills to offer good teaching but are not able to provide teaching tailored to the different needs of students. The use of reliable and valid measurement instruments can help teachers and other educational professionals gain insight into student performance and can assist them in achieving the important ability of addressing the different needs of students.

2.2.3 Measuring the ability to perform historical perspective taking

Measuring historical reasoning competencies is a very difficult challenge (e.g., Haydn, 2011; Peck & Seixas, 2008; Reich, 2009; VanSledright, 2013). HPT can be measured through semi-structured interviews (e.g., Berti, Baldin, & Toneatti, 2009; Lee & Ashby, 2001; Shemilt, 1987) and think-aloud assignments (e.g., Van Boxtel & Van Drie, 2004; Wineburg, 2001; Wooden, 2008), but these methods are time- and cost-ineffective. Hartmann and Hasselhorn (2008) have recently developed an instrument using a hypothetical scenario with an item-rating format. Their study offers positive indicators for measuring HPT among a large and heterogeneous student population.

The scenario refers to the rise of the Nazi Party in Germany in the 1930s. The central historical agent is a young man who is deciding which political party to vote for in the next election. In relation to the historical story, the authors formulated nine items, corresponding to three stages of HPT: the present-oriented perspective, the role of the historical agent, and the historical contextualization (Hartmann & Hasselhorn, 2008). The three present-oriented perspective items display contemporary views on the past, whereas the three items pertaining to the role of the historical agent refer to his personal situation: What is his family like? Is he a member of the elite? This category is marked by the authors as an intermediate category between the present-oriented perspective and the historical contextualization items. These latter items display historical contextualized thinking. The student's assignment is to place himself or herself in the historical context of this agent and decide if Hannes is willing to vote for the Nazi Party.

Hartmann and Hasselhorn (2008) found positive initial results for their instrument's reliability and validity. Their instrument is also a time- and cost-effective measurement instrument that can easily be implemented by, for example, teachers and test administrators. However, no study has tested the instrument in a large, heterogeneous population of students. Hartmann and Hasselhorn (2008) also raise the question about the instrument's reliability and validity should it incorporate a different historical topic. In this study, we took up these challenges. We tested the instrument in a different country among both upper elementary and secondary school students and developed a second version of the instrument to test the reliability and validity effects when a different historical topic was used.

2.3 Research questions

Despite the importance of historical reasoning competencies, almost no reliable and valid instruments exist to measure HPT among upper elementary and secondary school students. This results in little knowledge about the differences between students in terms of this capability. Therefore, we specify three research questions:

1. Does the instrument developed by Hartmann and Hasselhorn (2008) have positive reliability and validity outcomes when it is used to measure the ability to perform HPT among a large, heterogeneous student population in a different country?
2. What are the reliability and validity outcomes when the instrument format developed by Hartmann and Hasselhorn (2008) focuses on a different historical topic?
3. Which differences arise among students of different ages and educational levels regarding their ability to perform HPT?

2.4 Method

2.4.1 Constructing and adjusting the instruments

The first step was translating the hypothetical scenario and the accompanying items of the Nazi Party instrument developed by Hartmann and Hasselhorn (2008) into Dutch without affecting the instruments' interpretative framework. Hartmann

and Hasselhorn (2008) excluded one instruments' item (ROA1) from their analysis because factor analysis showed that it violated the two-dimensional structure of their conceptualization of HPT. We included this item in our instrument because our study has been conducted in a larger and more heterogeneous student population and therefore might fit in our conceptualization of HPT.

As a second step, to investigate the effect of topic choice on a student's ability to perform HPT, we developed three other hypothetical scenarios and items about different historical topics, with the same item-rating format Hartmann and Hasselhorn (2008) used. The first scenario was about medieval witchery, the second scenario was about the Nazi occupation of the Netherlands from 1940 to 1945, and the last scenario focused on 19th century-slavery. Constructing the scenarios and items was a difficult challenge because every historical topic has its own historical context with different related historical phenomena. HPT was embedded in different ways into the scenarios and with different student tasks. In the medieval witchery scenario, students had to explain the burnings of witches; in the Nazi occupation scenario, students had to decide what Dutch policemen would have done when asked to sign a document of collaboration with the Nazis. In the slavery scenario, we triggered HPT in the context of a question to evaluate information from a historical source. All three newly developed scenarios and items intentionally were designed to give rise to students' emotions and their present values and beliefs just as Hartmann and Hasselhorn's (2008) instrument did, because we wanted to examine whether students could set aside their first emotional reaction, create a historical context, and explain people's actions in the past.

To decide which additional instruments were the most suitable for use in our research and whether such instruments would be practically used by teachers in the classroom, we organized an expert panel composed of four history teacher educators from two universities (two with more than 4 years' work experience; two with more than 14 years' work experience), six secondary school history teachers (all six with more than 22 years' work experience), and two elementary school teachers (both with more than 16 years' work experience). The meeting took place in the context of a 1-day teacher-training program at the University of Groningen, and all teachers and teacher educators participated voluntarily.

All secondary school teachers and teacher educators were optimistic about the use of these instruments in classroom practice, not only for assessing the ability to

perform HPT but also as a practice and training instrument for their students. The secondary school teachers noted that history textbooks do not provide these types of assessment formats but focus more on assessing factual knowledge. The teachers also noted that using these instruments also supports other historical thinking and reasoning competencies, such as a critical evaluation of historical sources or providing solid argumentation. Furthermore, the secondary school teachers were optimistic about the use of the instruments as starting point for a whole-classroom discussion about, for example, the rise of Hitler in Nazi Germany.

The elementary school teachers were more restrained because they did not explicitly see the relevance of the instruments regarding the government's goals for elementary history education. However, they were positive about the "empathy" aspect of the instruments and expected that such assignments would help students developing a better understanding of decisions made by a historical agent. The experts concluded that the topics of the Nazi occupation of the Netherlands and medieval witchcraft needed too detailed historical content knowledge, which would result in comprehension difficulty for upper elementary and young secondary school students. Therefore, we excluded these two scenarios and selected the slavery-related instrument as the second instrument.

The third and final step was shaping the two final instruments (see Appendix A for the Nazi Party instrument and Appendix B for the slavery instrument) in a manner that would make them suitable for both upper elementary and secondary school students. Therefore, we first conducted a qualitative pilot study among upper elementary ($n = 6$) and secondary school students ($n = 9$) to test the comprehension difficulty of the two instruments' hypothetical scenarios. Specifically, while students performed the assignment and thought aloud, their answers were transcribed and analyzed to examine comprehension difficulty. We also asked the students to highlight difficult words in the scenarios and the accompanying items. The analysis of the pilot study showed that some abstract concepts in the hypothetical scenarios and question items were too difficult for upper elementary children. For example, the word master as a designation for a plantation owner in the slavery scenario caused confusion. In the hypothetical scenario of the Nazi Party, some upper elementary and secondary students also experienced difficulties with abstract concepts such as conservative. Second, we asked elementary school teachers ($n = 4$) and secondary school history teachers ($n = 6$) in an expert panel to review both hypothetical scenarios and items

for their levels of comprehension difficulty. All teachers involved in the expert panel had more than 15 years' work experience. The experts noted concerns about a few substantive concepts in the hypothetical scenarios that were found to be too difficult, especially for children in upper elementary schools, such as conservative, policy of appeasement, and the name of the German political party DVNP.

The results of the qualitative pilot study and the expert panel meeting showed that both instruments needed minor revisions. We replaced difficult concepts with more specific terms or else removed them without affecting the interpretive framework of the hypothetical scenarios. In a second session with different upper elementary ($n = 4$) and secondary ($n = 5$) school students, we noticed that there was no more comprehension difficulty.

2.4.2 Sample and procedure

The study was conducted on 1,383 students in elementary ($n = 178$) and secondary ($n = 1,205$) schools—specifically, four elementary and 18 secondary schools in the northern part of the Netherlands. Missing data led us to exclude 113 cases, leaving 1,270 cases for further analysis. In the Dutch educational system, students begin their elementary education around the age of four and continue in elementary education for 8 years. In the last 2 years of their elementary education, students are advised about their further (secondary) education, including pre-vocational secondary education (4 years), senior general secondary education (5 years), or pre-university education (6 years). We included students undertaking elementary education, senior general secondary education, and pre-university education, as described in Table 1. Pre-vocational secondary education was not included in the research sample because of the different history curriculum of this type of education in which the ability to perform HPT played a far less substantial role compared to senior general secondary education and pre-university education.

Table 1. Participants by age, educational level, and gender (N = 1,270)

Age	10	11	12	13	14	15	16	17	Total
Elementary male	20	43	14	*	*	*	*	*	77
Elementary female	34	50	17	*	*	*	*	*	101
Senior general secondary male	*	1	34	16	39	60	58	59	267
Senior general secondary female	*	2	45	13	42	63	61	62	288
Pre-university male	*	1	28	9	39	61	37	59	234
Pre-university female	*	2	37	17	53	76	44	74	303
Total	54	99	175	55	173	260	200	254	1,270

Note. *No students of this age occur in the educational level.

The mean student age was 14.2 years (SD = 2.2). In terms of gender, the distribution in the research sample was 45% boys and 55% girls; in the Netherlands, overall, the distribution between male and female students is 48% and 52%, respectively (Statistics Netherlands, 2012). The participating schools generally matched the total population in terms of the number of students and graduation rates (Statistics Netherlands, 2012).

The data collection took place during March and April 2012. Participating schools and teachers received hard copies of the instruments. Students were instructed at the beginning of a lesson to complete the instruments individually, in silence and without asking the teacher or other students for help. No time limit was given, but they all completed each instrument within 15 minutes. To assess students' prior knowledge about a topic, we included four multiple-choice items for each instrument. The multiple-choice items focused on historical content knowledge. For example, we asked for the year in which Hitler came to power in Germany and in which year the great worldwide economic depression was. Related to the slavery instrument, we asked them to define the triangular trade and in which part of America slavery was most prominent in the 19th century.

Furthermore, we asked for the students' ages, history grades, genders, and scores on a Dutch standardized final test (Citotoets) that is administered to upper elementary students. This optional test, commissioned by the Dutch Ministry of Education, Culture, and Science and developed by the Dutch National Institute for Educational Measurement, aims to measure pupils' attainment of certain standards in elementary education. The test contains 290 multiple-choice items in the fields of language (100 items), mathematics (60 items), learning skills (40 items), and world orientation (90

items). World orientation is a combination of history and geography multiple-choice items and forms a substantial part of the test. The history items focus on content knowledge and historical reasoning competencies. For example, students have to date historical pictures and choose periods in which there was war in the Netherlands (Dutch National Institute for Educational Measurement, 2013).

2.4.3 Data analysis

To answer the first two research questions, we began by examining the psychometric quality of the Nazi Party instrument and the slavery instrument. To be able to do this, we needed a coding system. In contrast with Hartmann and Hasselhorn (2008), who worked with latent class analysis, we used student mean scores on both instruments. Hartmann and Hasselhorn (2008) conducted their research on a small and homogeneous population. In our study, working with student mean scores showed the best results regarding the large and heterogeneous research population.

The present-oriented perspective items of both instruments used the following coding system from left to right for the answer columns (see Appendix A and Appendix B for the four columns). Selecting the first column yielded 4 points, the second column 3 points, the third column 2 points, and the last column 1 point. The role of the historical agent and historical contextualization items had the opposite coding system from left to right. Selecting the first column yielded one point, the second column two points, the third column three points, and the last column four points. A mean category score was calculated by summing the category items' scores and dividing this score by three (because each category has three items). A total mean score of HPT was calculated by adding up the different mean category scores and dividing this score by three (because the instrument has three categories).

To test the content validity of both instruments, we asked 10 teachers from the teacher network of the Department of Teacher Education of the University of Groningen as an expert panel. The results of two teachers were deleted due to procedural mistakes when conducting the assignment. The eight teachers varied in work experience from 2 to more than 30 years. We also asked 10 historians who held a position at a university or at a university of applied sciences as a second expert panel. Because they are accustomed to taking historical perspectives, they ought to score consistently high on the role of the historical agent and historical contextualization items and low on the present-oriented perspective items. All historians held university degrees in the

field of history and participated voluntarily.[1] The instruments' content validity was tested on both expert panels. Furthermore, we performed a principal component analysis (PCA) and a reliability analysis using the Cronbach's alpha coefficient to explore the data structure and internal consistency of both instruments. Finally, we examined the predictive validity and calculated correlations between the scores of both instruments. To answer the third research question, we used the different mean category scores, plotted this by age and calculated correlations between the students' HPT scores and different student characteristics (viz., age and educational level).

2.5 Results

The first two research questions focus on the reliability and validity of the instrument format developed by Hartmann and Hasselhorn (2008) when used in a different country, among a far larger and more heterogeneous student population and with a different historical topic. To answer both research questions, we looked at the instruments' content validity, dimensionality (i.e., whether the three categories of each instrument form one or multiple factors), internal consistency, and predictive validity.

2.5.1 Content validity of both instruments

Eight teachers sorted the nine items of each instrument into the three categories (viz., the present-oriented perspective, the role of the historical agent, and the historical contextualization) to confirm the categories' and items' face validity. A brief description of each category was provided, and they were instructed to place the items in the appropriate category. For both instruments, we calculated the agreement among the eight experts using the jury alpha and Fleiss's kappa, which we preferred to Cohen's kappa so that we could calculate the agreement among more than two raters. Fleiss's kappa values above .61 indicate substantial agreement; values greater than .81 are almost perfect agreement (Landis & Koch, 1977). For the Nazi Party instrument, the jury alpha was .96, and Fleiss's Kappa was .64. The jury alpha for the slavery instrument was .98, and the Fleiss' kappa was .71.

1 The published article (Huijgen, Van Boxtel, Van de Grift, & Holthuis, 2014) stated that 10 teachers were randomly selected from a list of 52 teachers and had more than 10 years' work experience. Instead, 10 teachers were asked to examine the content validity of both instruments. The results of two teachers were deleted due to procedural mistakes when completing the task. The eight teachers varied in work experience from 2 to more than 30 years. The article stated also that 10 historians were randomly selected from a list of 44 historians. Instead, 10 historians were asked to conduct the task of both instruments to further examine the content validity.

Beyond face validity, we wanted to test the instruments for accuracy, so we invited 10 professional historians to complete the measures. We calculated mean item scores for all three categories using a 4-point scale.[2] The expert scores on the historical contextualization items were 3.93 (Nazi Party) and 3.87 (slavery); those for the role of the historical agent items were 3.87 (Nazi Party) and 3.63 (slavery). The scores on the present-oriented perspective items (using a reverse-coding scheme, in contrast to the role of the historical agent items and historical contextualization items) were 3.83 (Nazi Party) and 3.83 (slavery). As we expected, the experts scored the role of the historical agent and historical contextualization items high and did not reason from a present-oriented perspective.

In accordance with these findings and to refine our content validity results, we derived two hypotheses, in which we predicted higher HPT scores among (1) older students and (2) students with more topic knowledge. The mean student score (on a 4-point scale) for the Nazi Party prior-topic knowledge test was 2.77 compared to 2.10 for the slavery prior-topic knowledge test. We calculated the correlation of students' total HPT scores with their ages and their prior topic knowledge scores. The results appear in Table 2.

Table 2. Correlations of student HPT scores with age and prior knowledge (N = 1,270)

Instrument	Age	Prior knowledge
Nazi Party	.35*	.27*
Slavery	.21*	.24*

Note. *Correlations are significant at the .01 level.

2.5.2 Dimensionality and internal consistency of both instruments

The principal component analysis (PCA) served to examine the structure of our data collected using our instruments. In line with Hartmann and Hasselhorn (2008), we expected to find two dimensions: one representing the two poles of a present-oriented perspective vs. a historical contextualization and the other representing the role of the historical agent. The results of the PCA for the Nazi Party instrument in Table 3 reveal two factors extracted with eigenvalues greater than 1. They accounted for 42% of the variance (factor 1: 28%, factor 2: 14%). The factor loadings after Varimax rotation

2 The published article (Huijgen, Van Boxtel, Van de Grift, & Holthuis, 2014) described the wrong expert scores. The scores on the historical contextualization items were not 3.88 (Nazi Party) and 3.77 (slavery) but 3.93 and 3.87, respectively. The scores on the role of the historical agent items were not 3.56 (Nazi Party) and 3.23 (slavery) but 3.87 and 3.63, respectively. The scores on the present-oriented perspective items were not 3.93 (Nazi Party) and 3.89 (slavery) but 3.83 and 3.83, respectively.

with Kaiser normalization also indicate that the present-oriented perspective items and historical contextualization items constituted one factor. The three items pertaining to the role of the historical agent constituted the second factor. In contrast with Hartmann and Hasselhorn (2008), our item ROA1 did not violate the simple structure.

Table 3. Principal component analysis results (rotated), Nazi Party instrument

Items	Factor 1	Factor 2
POP1	**.59**	.06
POP2	**.71**	-.03
POP3	**.72**	.01
ROA1	.22	**.71**
ROA2	-.09	**.70**
ROA3	-.21	**.47**
CONT1	**-.53**	.24
CONT2	**-.52**	.10
CONT3	**-.66**	.04

Note. POP = present-oriented perspective, ROA = role of the historical agent, and CONT = historical contextualization.

Table 4. Principal component analysis results (rotated), slavery instrument

Items	Factor 1	Factor 2	Factor 3
POP1	**.71**	-.11	.06
POP2	**.77**	.02	.09
POP3	**.73**	.01	.05
ROA1	-.04	-.16	**.58**
ROA2	.08	.04	**.77**
ROA3	.20	.12	**.67**
CONT1	.09	**.69**	.02
CONT2	-.07	**.78**	-.03
CONT3	-.02	**.67**	-.01

Note. POP = present-oriented perspective, ROA = role of the historical agent, and CONT = historical contextualization.

The PCA results for the slavery instrument data (see Table 4), however, highlight three factors extracted with eigenvalues greater than 1. They accounted for 52% of the variance (factor 1: 21%, factor 2: 18%, and factor 3: 13%). The factor loadings after Varimax rotation with Kaiser normalization indicate that the present-oriented perspective items constituted one factor, the historical contextualization items represented another factor, and the items pertaining to the role of the historical agent constituted a third factor.

Furthermore, we performed a reliability analysis using Cronbach's alpha coefficient to determine the internal consistency of both instruments (see Table 5). The slavery instrument showed a very low internal consistency score ($\alpha = .25$), compared with the Nazi Party instrument ($\alpha = .62$). Further analysis of the data showed that the historical

agent items for both instruments were primarily responsible for this low internal consistency. Excluding these items from the analysis resulted in higher internal consistency scores for both instruments (slavery: α = .49, Nazi Party: α = .69).

Table 5. Internal consistency of two instruments (N = 1,270)

Instrument	(α) POP	(α) ROA	(α) CONT	(α) POP, ROA, and CONT	(α) POP and CONT
Slavery	.60	.42	.54	.25	.49
Nazi Party	.62	.30	.51	.62	.69

Note. POP = present-oriented perspective, ROA = role of the historical agent, and CONT = historical contextualization.

2.5.3 Predictive validity

To assess the predictive validity of the instruments, we tested two hypotheses: namely, that the highest HPT scores would come from students with (1) high scores on the Dutch standardized final test for upper elementary students (Citotoets) and (2) high grades in history. Because historical reasoning and historical content knowledge form a substantial part of the final test, high scores on this test should be successful predictors for HPT performance. In line with Hartmann and Hasselhorn (2008), we also used history grades as a predictor for HPT performance. In Table 6, we present these correlation coefficients; the missing data are due to the non-obligatory nature of the Citotoets, such that not every Dutch elementary school (approximately 15%) has implemented this test (Dutch National Institute for Educational Measurement, 2013). The missing data regarding students' history grades exist because elementary school students do not have separate grades for history. We found a small but significant correlation between students' HPT scores and their Citotoets scores for the Nazi Party instrument but not for the slavery instrument. In addition, in contrast with Hartmann and Hasselhorn (2008), we did not find a significant correlation between students' history grades and their HPT scores.

Table 6. Correlations between HPT scores and student characteristics

Instrument	Citotoets	Students' history grades
Nazi Party	.17*	-.02
Slavery	.06	-.01
n	659	885

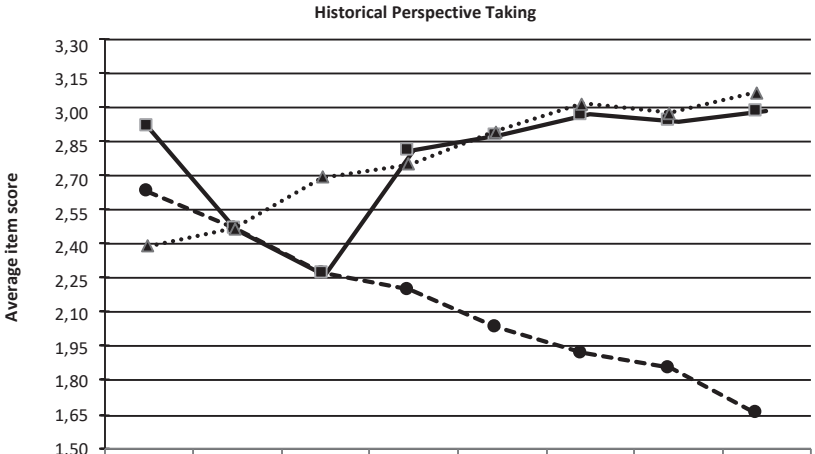
Note. *Correlation is significant at the .01 level.

Because we assume that both instruments test the same abilities of students, we calculated correlations between the category scores of the two tests across all students. The correlation coefficient between the present-oriented perspective category scores was .24, which was significant at the .01 level. We did not find a significant correlation between the two instruments category scores for the role of the historical agent. Between the historical contextualization category scores, there was a significant correlation of .23 at the .01 level. The correlation coefficient between the total HPT scores of all students was .23, which was significant at the .01 level.

2.5.4 Differences among the students when executing HPT

The third research question focuses on possible differences between students regarding their ability to perform HPT. The data obtained from the slavery instrument offered too low of an internal consistency to support the reliability of the data; therefore, we decided to work only with the Nazi Party instrument's data. Using these data, we investigated student mean scores for the three different categories (viz., the present-oriented perspective, role of the historical agent, and historical contextualization), plotted by age and educational level. Figure 2 presents the three mean category scores for students between the ages of 10 and 17 years. Both the declining trend for the present-oriented perspective category and the ascending trend for the historical contextualization category are notable. Starting at approximately eleven years of age, students began scoring higher in the historical contextualization category than in the present-oriented perspective category. With regard to the role of the historical agent, a decline occurred between the ages of 10 and 12 years, then after the age of 12, the line began to ascend, similar to the historical contextualization scores.

We calculated correlations for further analysis. Between 13 and 17 years (secondary education), the students showed a small but significant correlation of .11 (at the .01 level) between their scores in the category measuring the role of the historical agent and in the historical contextualization category. We did not find such a significant correlation (at the .01 or .05 level) when students were between 10 and 12 years of age (elementary education). Both senior general secondary and pre-university education showed the same trend (as plotted in Figure 2) between the ages of 12 and 17.



	10	11	12	13	14	15	16	17
pop	2,65	2,48	2,28	2,21	2,04	1,93	1,86	1,66
roa	2,94	2,48	2,28	2,83	2,90	2,99	2,96	3,01
cont	2,40	2,48	2,71	2,77	2,92	3,04	3,00	3,09

Figure 2. Historical perspective taking, plotted by age, Nazi Party instrument (N = 1,270)

Note. POP = present-oriented perspective, ROA = role of the historical agent, and CONT = historical contextualization.

When compared with students in other educational levels, the pre-university students scored the highest on HPT. A one-way analysis of variance-based post hoc multiple comparison with assumed Scheffé equal variance was used to test for any significant differences across the different educational levels. The difference between senior general secondary education (total score of 2.44, SD = 0.51) and pre-university education (total score of 3.15, SD = 0.50) was significant at the .05 level. The comparison of elementary education with both senior general secondary education (total score of 2.90, SD = 0.54) and pre-university education showed significant differences at the .01 level.

2.6 Discussion and conclusions

Our study focused on the reliability and validity of the instrument of Hartmann and Hasselhorn (2008) when tested among a large and heterogeneous student population in a different country and when applied to a different historical topic. Furthermore, we explored possible differences between students on HPT performance. In this section, we discuss our findings, outline limitations of our study, and present suggestions for further research.

Regarding the first research question, we found, in line with Hartmann and Hasselhorn (2008), positive indicators for the reliability and validity for the Nazi Party instrument. We also concluded that HPT is a two-dimensional construct consisting of (1) a dimension characterized by present-oriented perspective and historical contextualization poles and (2) items pertaining to the role of the historical agent. A PCA performed on our data from the Nazi Party instrument confirmed this. The reliability analyses indicated very acceptable (nearly characterizable as good) internal consistency for the Nazi Party instrument when we excluded the items tapping the role of the historical agent. We do not know the implications of the role of the historical agent items and its relation with the historical reasoning competency of HPT. Thinking-aloud methods could provide more insight if the role of historical agent items can contribute to students' ability to perform HPT.

To examine the second research question, we used a different historical scenario about 19th-century slavery with the same item-rating format. When examining the psychometric qualities of this slavery instrument, we did find positive evidence for content validity but not for predictive validity or internal consistency. In line with Hartmann and Hasselhorn's (2008) conclusions, our findings using the data obtained from the Nazi Party instrument showed HPT emerging as a two-dimensional construct. However, with the slavery instrument, our PCA identified three dimensions that were separately associated with each perspective (viz., present-oriented perspective, role of the historical agent, and historical contextualization).

Regarding the third research question, using the data obtained from the Nazi Party instrument, we found that upper elementary school students, starting at the age of 10 years, successfully performed some historical contextualization efforts. This is in line with research conducted by Barton (1997), Brophy and VanSledright (1997), and Field (2001). However, they also displayed more presentism, which resulted in higher scores on the present-oriented perspective items. Older students achieved higher scores for historical contextualization than younger students, and pre-university students held the highest HPT scores compared with students in senior general secondary and elementary education groups.

There may be several reasons for the differences in reliability and validity observed between the slavery instrument and the Nazi Party instrument. Because we embedded testing students' ability to perform HPT into determining the usefulness

of a source for making statements about the past, the observed differences might have stemmed from the specific instructions provided for the slavery instrument. For this instrument, students had to approach the story about how the enslaved people were treated as historical source. In addition to performing HPT, students also had to execute other historical thinking and reasoning competencies related to the use of historical evidence (e.g., assessing the reliability of the source) when completing the slavery instrument successfully. This dimension is missing from the Nazi Party instrument.

The differences also might be explained by the students' having less prior knowledge of slavery. Students scored lower on the slavery prior knowledge questions compared with the topic knowledge questions related to the Nazi Party. Van Boxtel and Van Drie (2012) concluded that knowledge of key historical concepts and dates plays an important role in a student's ability to contextualize a historical source. Thinking-aloud methods could be a valuable addition for gaining insight about whether students use knowledge (and what knowledge they do use) when responding to the slavery items and whether they notice differences in how the items are constructed.

Although we found a significant correlation between students' scores between the Nazi Party instrument and the slavery instrument, the results show that the slavery instrument did not meet our reliability and validity criteria. The secondary and elementary school teachers who were consulted were encouraging about the use of these instruments in classroom practice as both an assessment format and as a training exercise to stimulate HPT. Still, we do not exactly know if it is possible to test the ability to perform HPT in a reliable and valid way using items reflecting a present-oriented perspective, the role of the historical agent, and historical contextualization in the context of different historical topics. Our results illustrate the difficulties that are encountered when trying to construct a new instrument with this item-rating format using the same types of items used in the Nazi Party instrument.

We must take into account the limitations of our study. Our instruments focus on a student's ability to consider the historical agents' personal situations (i.e., the role of the historical agent) and the broader historical context (i.e., the present-oriented perspective and historical contextualization). This is a narrow view of HPT because scholars also refer to students' awareness of the differences between past and present (e.g., Seixas & Peck, 2004), the sense of a period (Dawson, 2009), and the application

of different frames of reference, specific (prior) knowledge, and understanding of the historical context and chronology (e.g., De Keyser & Vandepitte, 1998; Van Boxtel & Van Drie, 2012). For example, if students have little prior knowledge about a topic, they might refer more to specific characteristics of the historical agent to perform HPT (Berti et al., 2009). These are difficult abilities to measure using only the instruments described in this study.

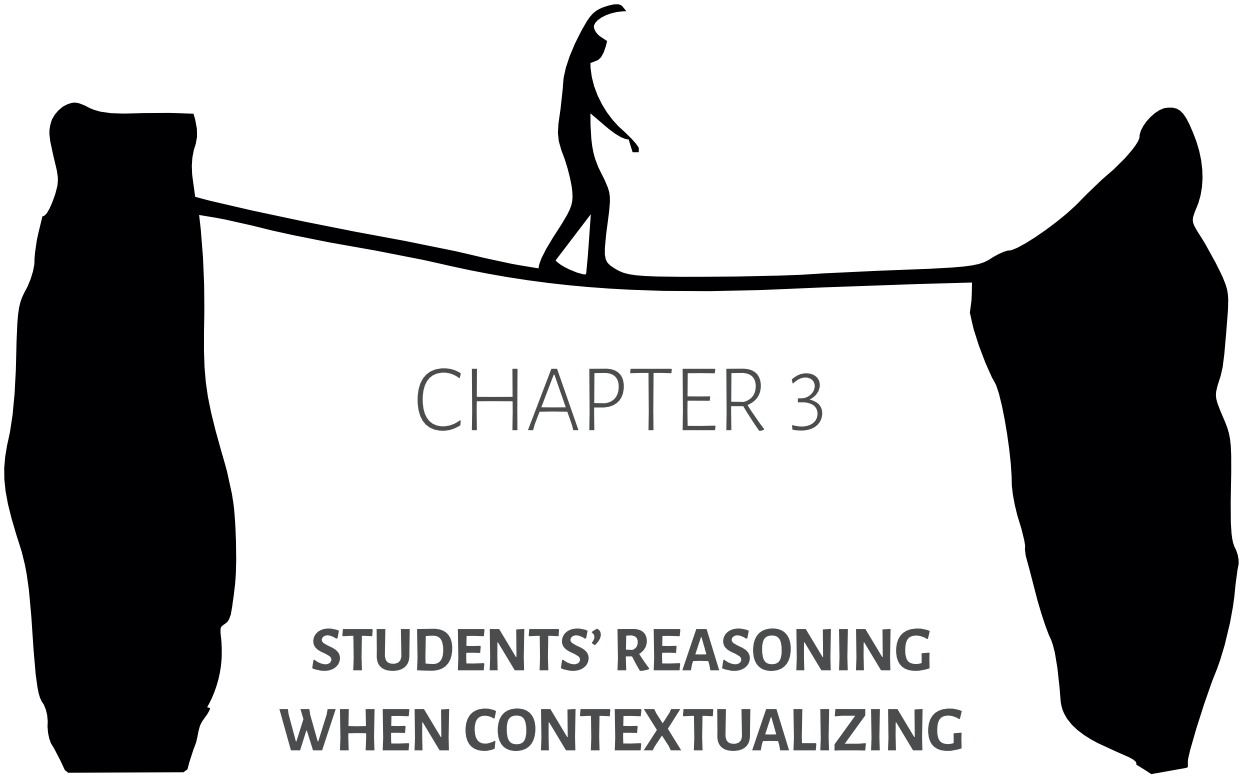
A more comprehensive measurement procedure might be necessary if we want to include the measurement of students' underlying knowledge and understanding. Constructing items that take into account, different frames of knowledge might provide insight into which different frames of reference are used by students when performing HPT (e.g., De Keyser & Vandepitte, 1998). The addition of thinking-aloud methods could also facilitate improved insight into whether students apply specific knowledge of topics and whether they combine this with knowledge about the specific characteristics of the historical agent. Combining the instruments with related historical empathy tasks and historical content tests might also provide insight into the roles played by distinguished elements (viz., historical contextualization, historical empathy, and avoiding presentism) when students perform HPT.

Another limitation is that both instruments focused purposefully on topics that give strong rise to students' emotions, such as anger and compassion, and these emotions may hinder efforts to better understand the past (Von Borries, 1994). It would be interesting to see how students perform HPT with respect to historical topics that do not explicitly give rise to emotions such as the invention of the steam engine. Furthermore, the items and the scenarios do not represent the whole historical context of the historical phenomena. The instruments had to be suitable for elementary school students; therefore, the items might consist of more simple functional explanations about the past (e.g., Bermúdez & Jaramillo, 2001). Constructing more items for each category or using different instruments focusing on the same historical topic might tackle this problem.

Further research should focus on the question of whether it is possible to construct a reliable and valid measurement of the ability to perform HPT, without the dependency of a specific historical topic and without being embedded in different tasks, such as historical empathy tasks in which students are asked to take the perspective of a fictional or genuine historical person or to examine the trustworthiness and

usefulness of a historical account. More research is also needed to investigate how students perform when the central historical agent of the instrument is, for example, a child or a politician. Students might identity themselves more with other children or heroic figures than politicians, and this might affect their ability to perform HPT (Brophy & VanSledright, 1997).

Additionally, the differences between taking the historical perspective of a group vs. taking the perspective of an individual should be further elaborated, following an interesting question raised by Berti et al. (2009). Furthermore, we only used one type of source: a textual story and its accompanying items. Textual sources play a very important role in history education, but so do visual sources. Further examination needs to be made of the differences that exist in HPT performance when the source is non-textual. Finally, but not less important, further research should focus on the role of the teacher. What types of instruction do teachers use to stimulate HPT in elementary and secondary education? Can the role of the historical agent be used as a scaffold for stimulating HPT? Such research could provide more insight into how to stimulate the important ability to perform HPT.



CHAPTER 3

STUDENTS' REASONING WHEN CONTEXTUALIZING HISTORICAL AGENTS' ACTIONS

This chapter examines students' reasoning when contextualizing historical agents' actions. Therefore, we first assessed a sample of 15- and 16-year-old pre-university students (n = 143) to determine their ability to contextualize the actions of people in the past. Subsequently, we explored, using thinking aloud methodology, students' reasoning (n = 36) to uncover their contextualization process. The results of this mixed methodology study indicate that most of the students in the sample performed well when engaging in HPT. Moreover, protocol analysis identified the different reasoning strategies that students employed to successfully perform HPT.

This chapter is based on: Huijgen, T. D., Van Boxtel, C. A. M., Van de Grift, W. J. C. M., & Holthuis, P. (2017). Toward historical perspective taking: Students' reasoning when contextualizing the actions of people in the past. *Theory & Research in Social Education*, *45*, 110-144. doi:10.1080/00933104.2016.1208597.

3.1 Introduction

In his book *Logics of History*, social historian William Sewell Jr. (2005) noted that historians should respect the differences that separate one period from another. He argued, "We cannot know what an act or utterance means and what its consequences might be without knowing the semantics, the technologies, the conventions, in brief, the logics, that characterize the world in which the action takes place" (p. 10). Other historians also stress the importance of considering the contextual circumstances when interpreting historical phenomena (e.g., Bevir, 2002; Gaddis, 2002; Tully, 1988). Accordingly, as student ability to contextualize historical phenomena is considered an important component of historical thinking, such conceptualization is being incorporated into history education worldwide (e.g., Lévesque, 2008; Seixas & Morton, 2013; Van Drie & Van Boxtel, 2008).

In history education, it is possible to contextualize historical sources and phenomena, including persons, events, and developments (Havekes, Coppen, Luttenberg, & Van Boxtel, 2012). When history education researchers discuss the contextualization of the actions of people and groups in the past, they often use the term *historical perspective taking* (HPT; e.g., Davis, Yeager, & Foster, 2001; Doppen, 2000). Though people in the past lived under different circumstances and viewed the world through different belief systems, many students might assume that people of the past had the same goals, intentions, attitudes, and beliefs as people in today's society, and as such, this *presentism* might result in misunderstandings about the past (Barton & Levstik, 2004; Lee & Ashby, 2001). For example, without the ability to perform HPT, students could not explain that Julius Caesar could not have breakfasted in Rome and dined in the Gaul region of France on the same day, as the transportation necessary for such a trip was not available during Caesar's time (Lévesque, 2008). Engaging in HPT could avoid presentism and help students understand and explain historical agents' decisions and historical phenomena (Van Boxtel & Van Drie, 2012). Some scholars also argue that HPT could contribute to citizenship in multicultural societies as it promotes the recognition and understanding of other people's views (e.g., Barton, 2012; Den Heyer, 2003; Rüsen, 2004). For example, Seixas and Peck (2004) argued that to promote students' social and political orientation and moral judgment, they must engage in HPT assignments.

Despite the importance of HPT in enhancing students' historical thinking and promoting citizenship among students, recent research has indicated that students may struggle when asked to perform thinking skills, such as HPT (e.g., Beyer, 2008; Huijgen, Van Boxtel, Van de Grift, & Holthuis, 2014; Reisman & Wineburg, 2008; Van Boxtel & Van Drie, 2012), and that history teachers may lack the requisite knowledge to promote historical reasoning competencies, such as HPT (e.g., Achinstein & Fogo, 2015; Bain & Mirel, 2006; Grant & Gradwell, 2010). Moreover, valid assignments and measurement instruments to assess students' historical reasoning competencies, such as HPT, are scarce (Breakstone, 2014; Reich, 2009; Rothstein, 2004; VanSledright, 2013). Therefore, to understand how students learn history and how they improve as a result of such learning, more information is needed regarding how students reason when performing historical reasoning competency tasks and regarding the development of instruments that operationalize this type of reasoning (Hartmann & Hasselhorn, 2008; Huijgen et al., 2014).

In this study, which uses an HPT instrument developed by Hartmann and Hasselhorn (2008), the ability of 15- and 16-year-old pre-university students (n = 143) to contextualize the actions of people in the past was assessed. Furthermore, we explored, using thinking-aloud methodology, how a sample of 15- and 16-year-old students (n = 36) reasoned to uncover their contextualization process when working with the HPT instrument. The results of this study provide insights into the difficulties students experience when engaging in HPT and into the validity and reliability of HPT classroom assignments, thereby helping teachers to promote their students' ability to perform HPT.

3.2 Theoretical framework

3.2.1 HPT: a conceptualization

Because of the critical role HPT plays in students' understanding of history and in promoting the competencies students need to successfully participate in civic life, the ability to perform HPT is incorporated into the formal K-12 history curricula of, for example, the United Kingdom (Cooper & Chapman, 2009; Department for Education, 2013), Australia (Australian Curriculum, Assessment and Reporting Authority, 2014), Canada (Peck & Seixas, 2008), Germany (Hartmann & Hasselhorn, 2008), Finland (Rantala, 2011), Belgium (Wils & Verschaffel, 2012), and the Netherlands (Van Boxtel

& Grever, 2011). Though in many states of the United States, HPT and similar reasoning competencies have appeared to play only a marginal role in the formal curricula (e.g., Evans, 2011; VanSledright, 2008; Wineburg, 2001), with the recent development of The College, Career, and Civic Life (C3) Framework for Social Studies State Standards (National Council for the Social Studies, 2013), more attention, in the near future, may be given to implementing reasoning competencies, such as HPT, in state curricula. For example, two objectives of the C3 Framework are that, by the end of Grade 12, students will be able to "analyze complex and interacting factors that influenced the perspectives of people during different historical eras" and "analyze how historical contexts shaped and continue to shape people's perspectives" (p. 47).

In the literature, different definitions of HPT exist. For example, Seixas and Morton (2013) defined HPT as an attempt to see through the eyes of people who lived in other times and circumstances that are sometimes far removed from our present-day lives. Levstik (2001) defined HPT as the ability to see how people acted in the past and understand why they acted as they did. To achieve HPT, scholars stress the importance of understanding the social, cultural, intellectual, and emotional settings that shaped people's lives and actions (e.g., Lee & Ashby, 2001; Seixas & Peck, 2004). Moreover, it must be emphasized that knowledge and understanding of chronology are important for successful HPT (Yeager & Foster, 2001).

Accordingly, HPT is a complex historical reasoning competency that consists of several components. From the extant literature, we identify three interrelated components needed to successfully perform HPT. These include applying the awareness that a present-oriented perspective might hinder the understanding of people's actions in the past, demonstrating historical empathy, and reconstructing an adequate historical context.

The first component is to be aware of a possible present-oriented perspective and the consequences of this perspective when examining the past. Present-oriented thinking, or presentism, is the bias by which people assume that the same goals, intentions, attitudes, and beliefs that exist in the present day existed in the past (Barton & Levstik, 2004). Forms of displaying a present-oriented perspective include viewing people in the past as stupid or assuming that people in the past had the same knowledge available to them that we currently have (Lee & Ashby, 2001; Shemilt, 1983). This perspective could cause misconceptions that lead to incorrect conclusions about

the past and thus hinder successful HPT (Reisman & Wineburg, 2008). Although we can never be perfectly non-presentist (e.g., VanSledright, 2001; Wineburg, 2001), students must understand that the past differs from the present when interpreting historical phenomena and the decisions of historical agents (Seixas, 1996; Seixas & Morton, 2013). Students who are aware of the difference between the past and the present and acknowledge their present-oriented perspective might demonstrate this awareness by explaining that people in the past did not know what we now know or that people thought differently in the past.

The second component is to exhibit historical empathy. Historical empathy refers to placing oneself in the position of people in the past to understand their motives and values regarding their decisions and actions (e.g., Cunningham, 2009; Endacott & Sturtz, 2014). Although some scholars have argued that historical empathy can never be fully achieved and is idealistic because it is impossible to put oneself in the shoes of a historical agent (e.g., Kitson, Husbands, & Steward, 2011; Riley, 1998; Wineburg, 1998), many scholars have concluded that historical empathy contributes to insights about historical agents' decisions (e.g., Brooks, 2011; Endacott & Brooks, 2013; Kohlmeier, 2006). However, though history education research has debated the extent to which historical empathy is an affective or cognitive achievement (e.g., Virja & Kouki, 2014), we consider historical empathy as a combination of affective and cognitive processes, following the conceptualization of scholars such as Endacott and Brooks (2013). It is further posited that connecting with known and familiar emotions of people in the past as an affective process might promote historical empathy and understanding of historical agents' decisions (Riley, 1998; Skolnick, Dulberg, & Maestre, 2004). Furthermore, considering the roles and positions of different historical agents in society and how such positions may have affected their views on historical phenomena as a more cognitive process could also contribute to historical empathy and to the understanding of historical agents' actions (Bermúdez & Jaramillo, 2001). In this study, we use the concept of historical empathy as putting oneself in the shoes of a historical person by considering his or her emotions, role, and position. For analytical reasons, we consider the reconstruction of a historical context as a distinct component.

The third component is the reconstruction of the historical context. Yeager and Foster (2001) argued that students must possess historical context knowledge, which includes knowledge about chronology, before they can interpret historical phenomena

and historical agents' actions. To reconstruct a historical context, students can use different frames of reference, including a chronological frame of reference, a spatial frame of reference, or a social frame of reference. The chronological frame includes knowledge about the time and the period as well as the sequence of significant events and developments (e.g., Dawson, 2009; Wilschut, 2012). For example, when attempting to understand why people in Germany in the 1930s voted for the Nazi Party, it is important to know the sequence of the First World War, the economic crisis of 1929, and the rise of Hitler. In contrast, the spatial frame focuses on knowledge about geographical locations and scale (e.g., De Keyser & Vandepitte, 1998; Havekes et al., 2012), such as knowledge of where Germany is located in Europe, what countries share boundaries with Germany, and what countries are near Germany. The social frame includes not only knowledge about human behavior and the social conditions of life but also knowledge about socio-economic, socio-cultural, and socio-political developments (e.g., Pontecorvo & Girardet, 1993; Shemilt, 2009; Van Boxtel & Van Drie, 2012), such as knowledge of the poor German economic circumstances and the anger Germans had regarding the Treaty of Versailles. Some studies (e.g., Berti, Baldin, & Toneatti, 2009) contended that if students do not possess sufficient knowledge to reconstruct a historical context, they may use historical empathy (by referring more to specific characteristics of the historical agent) to perform HPT.

3.2.2 Students' ability to perform HPT

Are secondary school students cognitively capable of taking a historical perspective? Using short historical stories and classifying students' answers to questions related to those stories, Hallam (1970) and Kennedy (1983) concluded that students under the age of 16 years lack historical reasoning competencies, such as HPT. Compared with adults, elementary and secondary school students do indeed experience greater difficulty taking another person's perspective, particularly when that other person does not possess the same knowledge that they have (Bloom & German, 2000; Perner, 1991; Wellman, Cross, & Watson, 2001). Birch and Bloom (2007) discussed the *curse of knowledge*, which is a cognitive bias that makes it difficult for students who have more knowledge to think from the perspective of lesser-informed people. This inability hinders the successful implementation of HPT in history education, as students must be aware that much of the information and knowledge they possess was not available to people in the past.

However, studies on students' ability to perform HPT have shown that even upper elementary school students are capable of some form of HPT and can overcome tendencies of presentism (e.g., Barton, 1997; Davis et al., 2001; Foster & Yeager, 1999; VanSledright, 2002). In their Concepts of History and Teacher Approaches 7 to 14 project, Lee, Dickinson, and Ashby (1997) examined how students between the ages of 7–14 understand the nature and status of different historical claims. They found that some students between the ages of 11–14 were beginning to distinguish between what they know and what the historical agent knew at that time. Berti et al. (2009) interviewed a total of 150 students aged 8–25 years about the concept of the ordeal during the Middle Ages and concluded that nearly every student understood that the ordeal involved the intervention of God and was related to religious beliefs that differ from the beliefs held in the present. Hartmann and Hasselhorn (2008) investigated how 170 German 10th graders (mean age of 16) performed on an HPT instrument. They found that approximately 90% of the participants in the study successfully performed HPT. Huijgen et al. (2014) used the same instrument to test the ability of 1,270 elementary and secondary students aged between 10–17 years to perform HPT. Their results showed that even upper elementary school students are capable of performing some elements of HPT, though older students performed HPT more successfully than younger students.

3.2.3 Task approaches and the ability to perform HPT

Research has indicated that not only domain-specific knowledge, understanding, and strategies are important for solving problems, but also that more generic task approaches are important, such as carefully analyzing a problem and evaluating decisions (e.g., Alexander, 2003; Bransford, Brown, & Cocking, 2000). Only a few studies have focused on the use of task approaches in combination with contextualizing historical sources and historical agents' actions. When investigating how students contextualize and date historical images and documents, Van Boxtel and Van Drie (2012) found that students who rushed to a conclusion or ignored information regarding the source more often failed to contextualize the source compared with students who approached the task systematically and used many clues provided by the source to generate alternative hypotheses.

Wineburg (1998) investigated how two historians created a historical context from a historical text noting that *specification of ignorance* could promote the ability to create an adequate historical context. This specification of ignorance can refer to expressing

puzzlement, asking questions, or specifying gaps in knowledge. Though research has been conducted on how certain tasks, such as class discussions (Kohlmeier, 2006), source work (Brooks, 2011), and writing assignments (Brooks, 2008), can support components of HPT, important questions regarding the process of HPT and the difficulties students experience when performing HPT remain. For example, do students who rush to conclusions or who do not display their specification of ignorance perform more poorly on HPT than students who express doubt, ask questions, and understand the consequences of what they do not know?

3.3 Research questions

Teachers, educators, and researchers are still missing relevant information about why some students successfully perform HPT while others fail. In this study, we answer the call of previous research that argues for the use of think-aloud methods to identify students' reasoning when performing HPT and to further validate instruments that assess students' ability to perform HPT. We therefore specify the following two research questions:

1. What are the HPT abilities of 15- and 16-year-old pre-university students?
2. How do 15- and 16-year-old students reason when completing an HPT instrument?

3.4 Method

3.4.1 Research design

To answer our research questions, we used a mixed-method research design incorporating an HPT instrument as a student task. First, we conducted quantitative research to examine 15- to 16-year-old pre-university students' general level of ability to perform HPT. Next, we conducted qualitative research using the think-aloud methodology to explore students' underlying reasoning processes when performing HPT. In other words, we investigated how these students solve the assignment of the HPT instrument. The think aloud methodology, which has been widely used to capture students' reasoning processes (Van Someren, Barnard, & Sandberg, 1994),

is used as surveys and experiments would be unable to provide the rich and deep information about students' reasoning processes that is necessary to answer our research questions (Creswell, 2009; Macpherson, Brooker, & Ainsworth, 2000). We chose a mixed-method design because combining quantitative and qualitative research provides a better understanding of a research problem or issue than does the use of either research approach alone (Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 2010). Moreover, we focused on students aged 15–16 because, based on previous research (e.g., Berti et al., 2009; Hartmann & Hasselhorn, 2008), we concluded that students are able to perform HPT at this age, thus enabling us to investigate the reasoning that underlies one's ability to take a historical perspective.

3.4.2 The HPT instrument

An HPT instrument developed by Hartmann and Hasselhorn (2008) and translated into Dutch by Huijgen et al. (2014) was selected, as this instrument is suitable for research on a large group of students and refers to a historical topic that has been taught to the students participating in this study, thus resulting in sufficient prior knowledge. The HPT instrument consists of a hypothetical scenario referring to the rise of the Nazi Party in Germany in the 1930s (see Appendix A). The central historical agent in the scenario is a young man (Hannes) who struggles to decide which political party to vote for in the next election. An authentic historical source was not included in the instrument because Hartmann and Hasselhorn (2008) did not want to conflate students' HPT ability with their ability to understand historical sources. The students' central assignment was to decide if Hannes is willing to vote for the Nazi Party. In relation to the scenario, Hartmann and Hasselhorn (2008) formulated nine items that corresponded to three categories: the present-oriented perspective (POP), the role of the historical agent (ROA), and historical contextualization (CONT).

These three POP items may trigger possible forms of presentism in the students. For example, the first item, "He definitely will not vote for the NSDAP [National Socialist German Workers' Party, or Nazi Party]. No one approves of what this party has done to the world," illustrates knowledge that contemporary society possesses, but the German people living in 1930 did not possess this level of knowledge regarding the Nazi regime. This category aligns with our first conceptualized component of HPT, specifically, applying awareness that a present-oriented perspective might hinder the understanding of people's actions in the past.

The three ROA items refer to the historical agents' personal situation, such as the agents' family life. For example, the item "Because his father's business is almost bankrupt, he might vote for a party that protects small business owners" may trigger possible affective connections between the students and the historical agent through, for instance, recognizable emotions, such as protecting family members, thus aligning with our conceptualized affective processes of historical empathy, or trigger considerations of the position of Hannes' family in society, such as wealth and influence, thus aligning with our conceptualized cognitive processes of historical empathy.

In contrast, the three CONT items display historical contextualization and form the opposite of the POP items. For example, the item "Hannes has little experience with democracy. He probably does not know the risks associated with the NSDAP and thus will probably vote for the NSDAP" should trigger the reconstruction of the socio-political context of Germany in the 1930s. In this scenario, students would have to know that Germany was an empire led by one strong leader for a long time and that the German people may want to return to this state, in which case, they would view Hitler as the new strong emperor. The CONT category aligns with our conceptualization of the third HPT component, namely, reconstructing the historical context.

Hartmann and Hasselhorn (2008) tested their instrument among 170 German 10th graders (mean age of 16). In a confirmatory factor analysis, they found that the POP and CONT items constituted one factor and that the two ROA items constituted the second factor. One item in this category (ROA1) displayed loadings above .40 on both factors and was excluded from further analysis. Huijgen et al. (2014) translated the instrument into Dutch and tested 1,270 Dutch upper elementary and secondary school students, ranging in age from 10–17 years. Their confirmatory factor analysis also indicated that the POP items and CONT items constituted one factor and that the three ROA items pertaining to the role of the historical agent constituted the second factor. In contrast to Hartmann and Hasselhorn's (2008) finding, the item ROA1 did not violate the simple structure.

To assess the instrument's face validity, Huijgen et al. (2014) asked 10 expert history teachers to sort the nine items on the instrument into the three categories. To determine the level of agreement among these experts, the authors calculated the Fleiss kappa, which at .64 indicated substantial agreement (Landis & Koch, 1977).

Additionally, Hartmann and Hasselhorn found good inter-coder consistency (κ = .83) when four coders sorted the items into the three categories (POP, ROA, and CONT). To calculate an HPT score, we used the same scoring system and 4-point scale as Hartmann and Hasselhorn (2008) and Huijgen et al. (2014). The selection of responses in the instrument's first column of the POP items (*Does not fit his situation at all*; also see Appendix A) receives 4 points. Second column responses receive 3 points, third column responses are awarded 2 points, and fourth column responses receive 1 point. The role of the historical agent and historical contextualization items had the opposite coding system, i.e., from left to right, as these items reflected good HPT ability. Selecting first column responses yield 1 point, second column responses receive 2 points, third column responses are awarded 3 points, and fourth column responses receive 4 points.

A mean category score was calculated by summing the scores of the items in the category and dividing the total by three (because each category has three items). A total mean HPT score was calculated by adding the different mean category scores and dividing this score by three (because the instrument has three categories). Mean HPT scores < 2.50 denote inadequate ability to perform HPT. The 2.50 limit was chosen because it is the middle of the instrument's 4-point scale. Mean HPT scores ≥ 2.50 < 3.00 denote adequate ability to perform HPT, scores ≥ 3.00 < 3.50 denote good ability, and scores ≥ 3.50 denote excellent ability.

To test students' prior chronological knowledge about the historical topic, we included four multiple-choice items in the instrument (see Appendix C). These items focused on important German historical events for the period 1900–1950, such as the year of the great worldwide economic depression and the year Hitler came to power in Germany. We did not ask for more detailed knowledge about significant events and developments during this period because we did not want to reference too much topic knowledge before students were asked to complete the HPT instrument. The four historical events presented in the questions were chosen because of their importance and their relationship to the scenario in the instrument. Each correct answer to a question yielded 1 point, resulting in a maximum score of 4.00.

3.4.3 Research context

In the Netherlands, all children receive elementary education between the ages of 4–12 years. They receive education in, for example, writing, reading, geography,

history, math, and English. Around age twelve, the children transition to secondary education. This is when the first differentiation among three educational levels occurs. Approximately 60% of the students go to pre-vocational schools (duration of 4 years), 20% receive senior general secondary education (duration of 5 years), and 20% receive pre-university education (duration of 6 years). The determination of a student's level of education is based on the advice of the elementary school and supported by a mandatory standardized test that measures the student's attainment of certain standards (e.g., language, world orientation, mathematics) in elementary education. Only a pre-university degree allows access to Dutch universities. Furthermore, the educational quality of all elementary and secondary schools is monitored by the Dutch Inspection of Education. The ability to perform HPT is included only in the formal history exam program of senior general secondary education and pre-university education. A total number of 323,498 students attend upper secondary education between the ages of 15–18, of which 49% are placed in senior general secondary education and 51% receive pre-university education. The gender distribution for senior general secondary education is 49% male and 51% female, and for pre-university education it is 47% and 53%, respectively (Statistics Netherlands, 2014).

3.4.4 Quantitative research sample and data analysis

Using the HPT instrument, we tested 143 tenth-grade pre-university students from seven schools (four urban, three rural) to examine their ability to perform HPT.[3] The mean student age was 15.1 years, and the gender distribution of the sample was 54% female and 46% male. The participating schools generally matched the total population in terms of student enrollment and graduation rates (Statistics Netherlands, 2014). History was a compulsory subject for all 143 students, and students received two history lessons, each ranging from 50–60 minutes per week. Approximately one year prior to the study, the students had studied the history of Germany. The foci of the course included the First World War, the rise of Hitler, the Second World War, and Germany's role in the Cold War. To examine how the 143 students performed, we calculated students' mean HPT score, mean category scores (POP, ROA, and CONT), and mean prior knowledge scores.

3 The published article (Huijgen, Van Boxtel, Van de Grift, & Holthuis, 2017) described that 170 students were included in the sample but due to missing data the analyses were conducted among 143 students (HPT instrument) and 139 students (prior knowledge test). This chapter, including the results in Table 8 and Table 9, displays the correct information.

3.4.5 Qualitative research sample and data analysis

For our qualitative research sample, we used non-probability sampling to select 10 history teachers from 10 schools, five urban and five rural. The participating schools generally matched the total population in terms of student enrollment and graduation rates (Statistics Netherlands, 2014). We asked the 10 selected teachers to randomly select four 10th-grade students from the two highest Dutch educational tracks.[4] Four students did not agree to participate in the study, resulting in a sample of 36 students. In this sample, the mean student age was 15.6 years, and the gender distribution was 19 female students (53%) and 17 male students (47%). The students' answers were videotaped and transcribed for further data analysis (Ericsson & Simon, 1993). The interviewer was instructed to encourage students to think aloud and to read the instrument's items aloud to trigger students' reasoning processes. The mean time that students spent on the instrument was 13.8 minutes. The protocols were coded by one of the authors using the software program ATLAS.ti (Muhr, 1991), and two expert secondary school history teachers, both of whom held a Master's degree in history, reviewed the coding. Coding categories were based on our theoretical framework, and we formulated four primary categories with subcategories, as displayed in Table 7.

3.5 Results

3.5.1 Students' scores on the HPT instrument

In Table 8, we present the students' mean prior knowledge score, mean category scores (POP, ROA, and CONT), and mean HPT score. We consider mean HPT scores < 2.50 as denoting inadequate ability, scores ≥ 2.50 < 3.00 as denoting adequate ability, scores ≥ 3.00 < 3.50 as denoting good ability, and scores ≥ 3.50 as denoting excellent ability. The students' individual mean HPT scores ranged from 1.56 to 3.89 on a 4-point scale, with a mean score of 3.21. The mean prior knowledge score was 2.21 on a 4-point scale (see Appendix C for the four prior knowledge questions). The best overall student performance was observed on question one, which asked about the First World War, with 91% of the students answering correctly. The second question, which asked about the rise of Hitler, was answered correctly by 55% of the students, whereas the third question, which asked about the Wall Street Crisis, was answered correctly by only 26% of the students. The last question, which asked about the Treaty of Versailles, was answered correctly by 48% of the students.

4 The published article (Huijgen, Van Boxtel, Van de Grift, & Holthuis, 2017) stated that the sample comprised of pre-university students but also senior general secondary education students were included.

Table 7. Categories, subcategories, and examples of students' reasoning

Category	Subcategory	Example
Displaying presentism	Viewing people in the past as stupid	Hannes is just acting stupid when he votes for Hitler.
	Assuming that people in the past had the same knowledge that we have today	Hannes would definitely not vote for Hitler, because his Party was responsible for the Second World War.
Historical empathy	Making affective connections	If my own father was going to be broke, I would also help him.
	Involving the position of the historical agent in the society	Hannes was a member of the bourgeoisie.
Reconstructing a historical context	Using chronological knowledge	The Second World War has not begun.
	Using spatial knowledge	The location of the scenario is Germany.
	Using socio-economic knowledge	There were poor economic circumstances.
	Using socio-political knowledge	The Germans did not have much experience with democracy.
	Using socio-cultural knowledge	There was a lot of anger among many Germans regarding the Treaty of Versailles.
Task approaches	Referring to text	The text stated..., in the text...
	Specification of ignorance	I do not know if the Germans had much experience with democracy / When did the Second World War begin?

Using Tukey's honestly significant difference (HSD) post hoc test, we found no significant differences between the average HPT performances of female and male students. Next, we calculated a mean HPT score for students from the same school and used Tukey's HSD post hoc test to determine whether some schools outperformed other schools or scored exceptionally low compared to other schools. The test displayed no significant differences among schools. To examine the possible correlation between students' mean prior knowledge scores and their mean HPT scores, we calculated a Pearson correlation coefficient and found a small but statistically significant correlation of .19 at the .05 level. In contrast to Hartmann and Hasselhorn (2008), we did not find a significant correlation between students' mean HPT scores and their history grades. Table 9 breaks down the sample by student HPT ability. A mean HPT score ≥ 3.50 was achieved by 32 students (22.4%), which indicated excellent ability to perform HPT, while only seven students (4.9%) obtained a mean HPT score < 2.50, which indicated inadequate ability to perform HPT. Most students (n = 82, 57.3%) achieved mean HPT scores ≥ 3.00 < 3.50, which indicated good ability to perform HPT.

Table 8. Students' mean prior knowledge score, mean category scores, and mean HPT score

Mean score	Mean score	Mean score	Mean score	Mean score
prior knowledge (n = 139)	POP (n = 143)	ROA (n = 143)	CONT (n = 143)	HPT (n = 143)
2.21 (SD = 1.01)	3.28 (SD = 0.63)	3.08 (SD = 0.49)	3.26 (SD = 0.51)	3.20 (SD = 0.37)

Note. POP = present-oriented perspective, ROA = role of the historical agent, and CONT = historical contextualization.
The mean POP score was calculated using the opposite scoring system of the ROA and CONT categories. The maximum score of 4.00 for the POP category shows a very low level of presentism.

3.5.2 Student reasoning on the HPT instrument

To explore how the 15- to 16-year-old students arrived at their answers, we asked 36 students to think aloud as they solved the HPT instrument. Table 10 displays the individual mean HPT scores combined with students' reported use of the various components of HPT. The highest mean HPT score achieved in this sample was 3.89 on a 4-point scale (David, Eva, and John), and the lowest was 1.89 (Bas). The mean HPT score for this sample was 3.39, and only two students (6%) received mean HPT scores < 2.50 (Sophie and Bas), while 17 students (47%) achieved mean HPT scores ≥ 3.50.

Table 9. Students categorized by their mean HPT score (n = 143)

Mean HPT score		*n*	% of total students
≥ 3.50		32	22.4%
≥ 3.00 < 3.50		82	57.3%
≥ 2.50 < 3.00		22	15.4%
< 2.50		7	4.9%
Total		143	100%

Table 10. Students' mean HPT score and use of components of HPT (n = 36)

Student*	Mean HPT score	Displaying presentism		Reconstructing a historical context					Historical empathy		Task approaches	
		Viewing people in the past as stupid	Assuming that people in the past have the same knowledge we have today	Using chronological knowledge	Using spatial knowledge	Using socio-economic knowledge	Using socio-political knowledge	Using socio-cultural knowledge	Making affective connections	Involving agents' position	Referring to text	Specification of ignorance
David	3.89	0	0	2	0	0	9	3	2	0	6	2
Eva	3.89	0	0	2	0	5	7	4	0	0	4	0
John	3.89	0	0	10	0	7	10	0	0	0	3	0
Nina	3.78	0	0	3	1	5	8	2	2	0	3	1
Sean	3.78	0	1	7	0	6	15	6	1	0	2	1
Robin	3.78	0	0	9	0	7	8	1	0	0	3	5
Bob	3.67	0	0	4	0	8	14	1	2	0	1	1
Jim	3.67	0	0	6	0	6	7	2	0	2	2	0
Kevin	3.67	0	0	3	0	2	10	4	0	0	1	1
Maya	3.67	0	0	1	0	4	7	0	0	0	6	0
Ashley	3.67	0	0	5	0	6	14	4	2	1	4	1
Kim	3.67	0	0	7	0	8	8	1	3	0	6	2
Rose	3.67	0	1	4	0	5	6	0	0	0	2	3
Paul	3.67	0	0	6	0	7	10	5	1	0	5	4
Lauren	3.56	0	0	5	0	2	7	2	0	0	3	7
Maria	3.56	0	0	2	0	1	5	3	0	0	1	0
Jaimy	3.56	0	0	6	0	7	10	5	0	0	5	4
Ryan	3.44	0	0	7	0	2	7	0	0	0	2	2
Rachel	3.44	0	0	1	0	6	5	2	0	0	2	2
Stella	3.44	0	0	9	0	5	12	8	5	0	1	2
Tom	3.33	0	0	10	1	8	16	3	6	0	3	2
Emma	3.33	0	0	0	1	5	7	1	1	0	1	1
Judith	3.33	0	0	3	0	9	8	1	1	0	2	2
Ben	3.33	1	0	6	1	12	5	3	1	1	7	1
Emmy	3.33	0	0	1	0	8	8	3	3	0	4	1
Sarah	3.33	0	0	6	0	4	5	1	1	0	0	0
Kim	3.33	0	0	4	0	5	7	2	2	1	1	2
Lisa	3.33	0	0	5	0	3	7	3	2	0	2	7
Peter	3.22	0	2	4	0	2	7	4	1	0	0	0
Evan	3.11	0	0	3	0	1	4	2	3	0	1	3
Anna	3.00	0	0	5	1	6	8	3	3	0	2	1
Mark	3.00	0	1	0	0	4	5	0	5	0	0	0
Tim	3.00	0	0	5	0	6	6	0	0	0	8	2
Amy	2.78	0	0	3	0	3	4	2	1	0	2	0
Sophie	2.11	0	1	0	0	4	1	0	1	0	3	0
Bas	1.89	0	0	0	0	4	2	0	0	0	0	0
Total		1	6	154	5	183	279	81	49	5	98	60

Note. *Names are pseudonyms.

o = not observed in the protocols, x = observed in the protocols.

3.5.2.1 Viewing the past from a present-oriented perspective

Only one student (Ben) in the sample viewed Hannes as stupid or ignorant. Moreover, this perspective was only apparent when Ben thought about the second item of the HPT instrument ("He will see that only in a democracy can people take part in decision making. He will decide wisely and not choose NSDAP."). Ben applied his historical economic knowledge (high rate of unemployment) but did not include in his reasoning that democracy was uncommon in Germany in the 1930s, which caused Ben to perceive Hannes as being naïve:

> Honestly, I think that he is too naïve to understand that only in a democracy can people take part in decision making. He is only afraid that his business, more specifically, his fathers' business, is going to be bankrupt. Nobody had a job, and he only wants economic welfare. (Ben, reasoning about Item 2)

Five other students (Sean, Rose, Peter, Mark, and Sophie) also exhibited a present-oriented perspective when working on the HPT instrument. Though they did not view Hannes as stupid or ignorant, none of them included in their reasoning that the knowledge we have now was not available to people in the 1930s. For example, Rose appeared unaware that Hannes could not have known the outcome (e.g., the beginning of the Second World War) of Hitler's political rise:

> Rose: He will definitely not vote for the NSDAP. No one can approve of what this party has done to the world. Hitler was responsible for the Second World War.
> Interviewer: Is Hannes going to consider this?
> Rose: Yes, I think so. Hannes might vote for the NSDAP because he is not satisfied with the current government, but I think that he will not vote for the NSDAP because Hitler murdered thousands of people. (Rose, reasoning about Item 1)

In addition, Sophie, Mark, Sean, and Peter indicated in their reasoning that Hannes knew that voting for the NSDAP would result in violence and terror. Therefore, these students concluded that Hannes could not vote for the NSDAP. Out of the six students who displayed a present-oriented perspective, two students (Rose and Sean) recorded mean HPT scores > 3.50, three students (Ben, Peter, and Mark) achieved mean HPT scores ≥ 3.00 < 3.50, and one student (Sophie) had a mean HPT score < 2.50.

However, most of the students (*n* = 30) were aware of their possible present-oriented perspective when attempting to explain Hannes' actions. Many students applied their chronological knowledge to emphasize that the information we have now was not available to Hannes at that time. For example, Ryan noted that the scenario was set before the start of the Second World War, and thus, Hannes could not have known the consequences of Hitler's rise to power:

> The source states that the scenario is set in 1930. Hitler became the political leader of Germany in 1933? I do think so. Hannes is living in 1930, and Hitler became the leader a few years later so he could vote for the NSDAP in 1930, right? Because he does not know what Hitler has done to the world. (Ryan, reasoning about Item 1)

Another example of the awareness of a present-oriented perspective through the use of chronological knowledge was detected when students reasoned about Item 3 ("He will not vote for the NSDAP as their ideas are highly transparent. It is clear that this party wants war.") and concluded that we now know the outcome of the political rise of Hitler, but that people in the past did not have access to this knowledge in the 1930s:

> I do not think that he knows that the NSDAP might want a Second World War because this scenario is set in 1930. I think that in 1930 he easily could not know that the NSDAP wanted a war. He really could not know it. (Paul, reasoning about Item 3)

3.5.2.2 Historical empathy

Of the sampled students, 22 students made 49 affective connections with the historical agent (Hannes) as they explained his actions. In their reasoning, these students included arguments based on recognizable situations and emotions. They seemed to interpret or translate the historical situation into a situation that they could experience today. For example, Mark attempted to explain Hannes' decision by describing a more contemporary situation that he himself could experience as he reasoned about Item 4 ("As a member of a wealthy family, he would like to return to the German Empire as his family was better off. Therefore, he will vote for an anti-democratic party."):

> I think that this fits his situation because his father had told him that the time of the German Empire was far better compared to the contemporary circumstances. And yes, most of the time, I believe what my parents are telling me. So if Hannes had the opportunity, he would vote for the NSDAP. (Mark, reasoning about Item 4)

With respect to Item 6, "Because his father's business is almost bankrupt, he might vote for a party that protects small business owners," many students used affective connections to explain why Hannes might vote for the NSDAP. For example, Paul imagined that he himself had financial troubles and thus considered what he would do in a similar situation:

> I think that this is legitimate. I think that he will vote for the NSDAP. He is going to consider…I have a feeling that looking at Hannes' situation, the most important goal for him is that the family business is going well. Looking at myself, I would be happy if my business was making a profit, so I think that this could be the case for Hannes, too. (Paul, reasoning about Item 6)

Three students (Stella, Tom, and Mark) made five or more affective connections in their reasoning. These students often tried to personalize the historical situation when deciding whether Hannes would vote for the NSDAP. When reasoning about different items, these students used phrases such as "If I were in his shoes" and "I would decide what the best option is for me." Furthermore, 13 of the 49 affective connections (27%) were made by students with mean HPT scores ≥ 3.50. Most affective connections (n = 34, 69%), however, were made by students with mean HPT scores ≥ 3.00 < 3.50, and two affective connections (4%) were made by students with mean HPT scores < 3.00 (see Table 10). Interestingly, though some students indicated that they did not know the specific historical context of Germany, they nonetheless succeeded in answering items correctly by making affective connections. For example, Stella used affective connections to reason that Hannes might long for the period of the German Empire:

> I think most Germans were better off during the German Empire period, but I do not know the specific circumstances of that period, and if this might have resulted in better economic conditions for his family. Personally, I can understand that you might want the German Empire back because your personal wealth might be higher, and I personally can imagine that is a good thing for everybody. (Stella, reasoning about Item 4)

Only four students (Jim, Ashley, Ben, and Kim) referenced Hannes' position or his family's position in society in their reasoning. For example, Jim reasoned that Hannes' family was wealthy and respected, and therefore, he might have voted for the NSDAP if this party could ensure the prestige of Hannes' family. However, no student reasoned that Hannes' family might be part of the bourgeoisie and, thus, might long for the German Empire period (1871–1918) when most of these families had far greater political influence. In addition to making affective connections and considering the role of the historical agent, the protocols revealed yet another type of historical empathy. In particular, one student (Stella) used her knowledge of current values and beliefs of different places in the world. Stella reasoned that in other parts of the world that had a one-party political system, there could be economic welfare and people could be satisfied:

> Yeah, but maybe he does not feel the need to take part in political decision making. Why would you take part in decision making if you think the government makes wise and good decisions? Decisions that are also good for you. The only thing I then could say every 4 years when there are elections: You are doing a great job, keep up the good work. You still see this in parts of Asia, where people think that they do not need political influence because it is going very well within their own country. (Stella, reasoning about Item 2)

3.5.2.3 Reconstructing the historical context

The protocols further revealed that the 36 students used different types of knowledge to reconstruct the historical context as they engaged in HPT. Most references were made to socio-political knowledge, with a total of 279 references, followed by 183 references to socio-economic knowledge, and 154 references to chronological knowledge. Far fewer references were made to socio-cultural knowledge (n = 81) and spatial knowledge (n = 5). See Table 10 for detailed information. The protocols revealed that 32 students made references to chronological knowledge. Four students (Emma, Mark, Sophie, and Bas) did not display any form of adequate chronological knowledge. Three of these students obtained mean HPT scores ≤ 3.00, while one student (Emma) obtained a mean HPT score of 3.33. Two of these students (Mark and Sophie) also viewed the past from a present-oriented perspective. Only five students made references to spatial knowledge. Of these five, one student had a mean HPT score of 3.78 (Nina), three students (Tom, Emma, and Ben) had mean HPT scores of 3.33, and one student (Anna) had a mean HPT score of 3.00. These students, for

example, referred to the geographical size of Germany during the period of the German Empire. Socio-economic knowledge was referenced by 35 students in their reasoning when taking a historical perspective. Kim and Anna, for example, both noted that the economic circumstances in Germany in the 1930s resulted in people being attracted to a strong leader who promised increased economic welfare by creating jobs:

> There [in Germany] was much unemployment, and it is all very bad. The country was first doing alright, but at that time, the economic circumstances were poor and people were dissatisfied with this. Because Hitler was a strong leader and the NSDAP would stimulate the economy by creating jobs, Hannes might vote for the NSDAP. (Kim, reasoning about Item 1)

> I think that he could vote for the NSDAP because he says that he is desperate. He is close to being unemployed, and I think that the NSDAP tried to create more jobs and they will provide a job for Hannes. They [the NSDAP] promised more jobs, and that might result in Hannes voting for the NSDAP. (Anna, reasoning about Item 1)

All students in the sample displayed socio-political knowledge in their reasoning. For example, Kevin stressed the political statements of the NSDAP to explain why Hannes might vote for such a political party:

> Before the Second World War, the NSDAP was known as a very good political party. Hitler promised many things, and many people believed him because they wanted a better future. I do not know if Hitler was already against the Jews, but he promised a lot more jobs. I think that was very clever, and it resulted in many votes for the NSDAP. (Kevin, reasoning about Item 8)

Most students in the sample (n = 28) also displayed socio-cultural knowledge in their reasoning. These students referenced their knowledge about German cultural behaviors and beliefs in the 1930s. For example, Ashley noted that Hannes might have been influenced by the propaganda spread by Hitler and his political party:

> Hannes could not see what was really going on in Germany because the German people were getting a very subjective image due to all the Nazi media and propaganda. Therefore, he could not see that the NSDAP wanted a war. (Ashley, reasoning about Item 3)

Other students referred to the "unfair" Treaty of Versailles, which resulted in anger among Germans toward a democratic government, while still others mentioned that Hannes was not used to living in a democracy with more freedom but poor economic circumstances. Therefore, he might be skeptical about this type of government. The two students (Sophie and Bas) with the lowest mean HPT scores (2.11 and 1.89, respectively) demonstrated far less knowledge than did students with higher mean HPT scores. Together with Mark (HPT score of 3.00), they were the only students who combined just two different types of knowledge, specifically, socio-economic and socio-political knowledge. They did not use chronological, spatial, or socio-cultural knowledge. All other students combined at least three types of knowledge in their reasoning. For example, Sean and Stella combined chronological knowledge (e.g., Treaty of Versailles in 1919), socio-economic knowledge (e.g., poor economic circumstances), socio-political knowledge (e.g., foreign policy of the Nazis), and socio-cultural knowledge (e.g., the Germans' anger regarding the Treaty of Versailles) when completing the assessment.

We also calculated a mean score for the use of different knowledge components by totaling the number of references to knowledge and then dividing this sum by five (the number of different knowledge components). For example, Bas made six references to knowledge and obtained a mean score of 1.20, whereas Tom made 38 references to knowledge and obtained a mean score of 7.60. When dividing our sample by the mean HPT score of 3.20 (based on the 143 students' mean HPT score), students with a mean HPT score above 3.20 had an average of 4.26 references to knowledge in the protocols, whereas students with mean HPT scores less than 3.21 made, on average, just 2.34 references to knowledge.

3.5.2.4 Task approaches

Most of the students (n = 32) explicitly referenced the text about Hannes when working on the instrument, as displayed in Table 10. These students re-read parts of the text or referenced specific information when reasoning about individual items. Only four students read the text once, did not look at it again, and did not explicitly refer to it in their reasoning (see Table 10). Furthermore, most of the students (n = 25) displayed their specification of ignorance, i.e., they doubted their conclusions or indicated that they did not possess the knowledge. These students, for example, were not familiar with the specific political viewpoints of the Nazi Party and did not know how Hannes would react or respond to the instrument's items. Consequently, they had to speculate:

> Lauren: He will not vote for the NSDAP. Their ideas are easy to see through. It is clear that this party wants a war, but I do not know if Hannes could see this.
> Interviewer: Why not?
> Lauren: I do not know the ideas of the NSDAP. Was it obvious that Hitler wanted to start a war? I do not know this. (Lauren, reasoning about Item 3)

Three students, Kim, Rachel, and Anna, explicitly stated in their reasoning that they could not identify the answer because the source did not provide the specific information. One student, Tom, explicitly stated how he was going to approach the task without any encouragement from the interviewer:

> First, I always examine the assignment before looking at the source. What do I have to do? I see a fill-in assignment with statements that I have to score. Next, I am going to look at the source. Okay, we have a source about Germany in the 1930s. (Tom, before beginning the assessment)

3.6 Conclusions and discussion

In this mixed-method study, we tested 143 pre-university students' abilities to perform HPT as evidenced by their performance on an HPT instrument developed by Hartmann and Hasselhorn (2008), and we explored the underlying contextualization processes of 36 students. In the late 1990s, Angvik and Von Borries (1997) conducted a cross-national survey that aimed to examine 15- to 16-year-old students' views on history education in Europe. One of the questions asked that students place themselves in the position of a young man or woman living in the 15th century who was being forced into marriage. The students were asked what they would do in such a circumstance if they had lived during that time period. Most students participating in the study found it difficult to reconstruct, accept, and acknowledge the concept of a forced marriage, and thus, they often expressed a present-oriented perspective. Nonetheless, only seven of the 143 students (4.9%) participating in our study had a mean HPT score < 2.50 out of a maximum 4.00 score, indicating inadequate ability to perform HPT. Most students (n = 82, 57.3%) achieved a mean HPT score ≥ 3.00 < 3.50, indicating good ability to perform HPT. This finding is consistent with Hartmann and Hasselhorn (2008), who also examined 15- to 16-year-old pre-university students' abilities to perform HPT.

Furthermore, our analysis of the verbal protocols of 36 students indicated that five students viewed the past from a present-oriented perspective in that they did not realize that people in the past did not have the same knowledge that we have today. One student viewed Hannes as naïve and reasoned at the lowest level of the Lee and Ashby (2001) taxonomy such that people of the past are regarded as ignorant or stupid. In contrast, the other 30 students were aware of the consequences of their present-oriented perspective when explaining historical agents' decisions, a finding consistent with that of Berti et al. (2009).

Various studies and handbooks on teaching and learning history emphasize that presentism restricts historical understanding and that many students might view the past from a present-oriented perspective (e.g., Haydn, Stephen, Arthur, & Hunt, 2015; Lévesque, 2008; Seixas & Morton, 2013). However, the causes of presentism exhibited by students are typically not described. Based on our findings, we posit that historical knowledge plays a critical role in preventing presentism. Scholars, such as Endacott and Brooks (2013), Van Boxtel and Van Drie (2012), VanSledright (2001), and Wineburg (2001) have suggested a relationship between historical content knowledge and students' ability to perform HPT, and our study seems to confirm this association. Specifically, we found a small but significant correlation (.19) between students' prior chronological knowledge and their performance on the HPT instrument.

Furthermore, students' protocols indicate that students who displayed good or excellent ability to perform HPT (mean HPT score ≥ 3.00) used more historical topic knowledge, particularly chronological and socio-political knowledge but also socio-cultural and socio-economic knowledge, in their reasoning than did students with mean HPT scores < 3.00. Compared to the lowest-performing students (Sophie and Bas), students who demonstrated good and excellent abilities to perform HPT also employed more types of knowledge in their reasoning. Making affective connections with a historical agent (e.g., if the students' own fathers had money problems) could also facilitate individuals as they engaged in HPT (Endacott & Sturtz, 2014; Virja & Kouki, 2014), and our data seem to confirm this. However, five of the six students who displayed a present-oriented perspective also made affective connections. This suggests that making affective connections alone might not prevent presentism but that, to prevent such presentism, affective connections must include the role of the historical agent and the broader historical context.

Based on the protocols, we also contend that students are skipping specific scaffolds that include the role of the historical agent when working on HPT tasks, as few students explicitly considered the role of the historical agent (e.g., Hannes' wealthy family influenced his preference for conservative political parties). Teachers who focus their instruction on teaching students to combine affective connections while also considering the role of the historical agent and the broader historical context might find that their students' performance on HPT tasks improves rather significantly.

Consistent with a relevant point raised by Berti et al. (2009), we found evidence that students used the affective element of historical empathy when they did not succeed in reconstructing the historical context. Some students explicitly noted that they did not know the specific historical circumstances but could understand Hannes' decision to vote for the Nazi Party because they, too, would not want to be unemployed. Because we observed this in the reasoning of only a few students, historical empathy as a fallback rationale and the interaction between affective and cognitive processes of historical empathy when performing HPT require further research.

Future research could also focus on whether affective connections are more difficult to make when the historical topics or issues are more distant, such as dating back to ancient Rome or the Middle Ages. Again, further research is needed to examine the extent to which students can perform HPT by evaluating current beliefs and values of different parts of the world, as we found one student in our study who applied this strategy.

Another finding of our study regards the instrument itself. While testing students about their knowledge of historical facts is rather easy, valid and reliable instruments that measure students' historical reasoning competencies are scarce. Consequently, scholars have argued for new assessment formats (Ercikan & Seixas, 2015; Reich, 2009), and the development of the History Assessments of Thinking (HATs) is a good example (Breakstone, 2014; Breakstone, Smith, & Wineburg, 2013). That said, we used an instrument validated by Hartmann and Hasselhorn (2008) and Huijgen et al. (2014). However, some limitations with regard to the instrument's validity and practical improvements must be noted. First, students may have misread or misunderstood two of the instruments' items. We noted in the protocols that some students explicitly struggled with answering the instruments' first and sixth items. For example, when working on the first item, students had to check the first box

(*Does not fit his situation at all*) to receive the maximum score. However, these students reasoned that Hannes could not have known the outcome of the Second World War, thus displaying a non-present-oriented perspective, and therefore ticked the last box (*Fits his situation very well*), which yielded a score of 1. The same thing occurred when students answered the sixth item. Rewriting these two items might reduce the potential for misunderstanding, and more detailed instruction on the terminology of the scoring boxes (such as the inclusion of a test item) before beginning the instrument might resolve this problem. Second, we observed in the data protocols that students answered the instruments' items after closely reading and investigating the source. Testing students' reading comprehension levels could better identify the impact student reading ability has on students' scores on the HPT instrument.

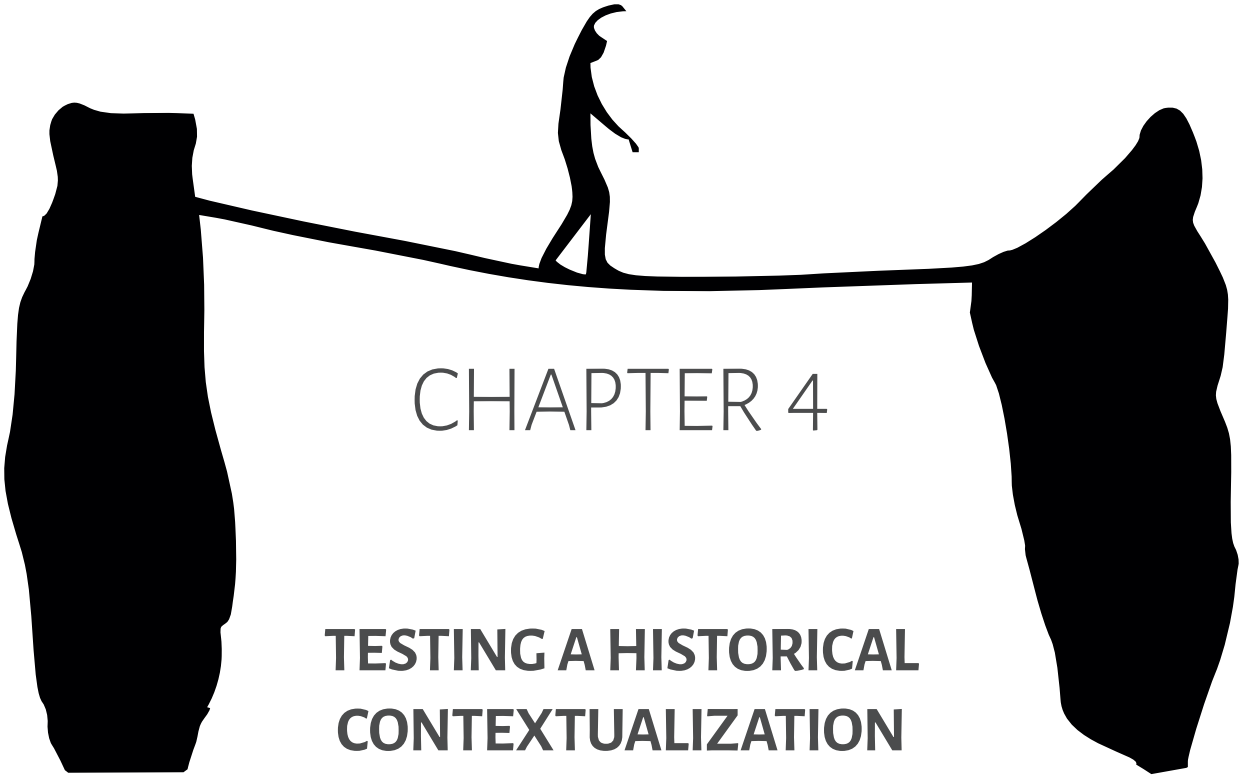
Furthermore, the ROA items on the instrument require examination. Originally, Hartmann and Hasselhorn (2008) conceptualized the ROA items as an intermediate stage between presentism and historical contextualization. Students could refer to roles or institutions that they know from their own lives (e.g., the role of a father or businessman). However, we did not find evidence that the ROA items represented an intermediate stage between presentism and historical contextualization. To further investigate the relationships between the cognitive and affective elements of historical empathy and the instrument's ROA items, the ROA items could be divided into two categories, specifically, items that might trigger more affective processes of historical empathy (e.g., "If my own father would be fired, I could vote for the Nazi Party in the 1930s.") and items that might trigger more cognitive processes (e.g., "Hannes belongs to a wealthy family. Therefore, he could vote for the Nazi Party.").

Though we did not find a strong relationship between generic task approaches (e.g., evaluating decisions, expressing doubt) and domain-specific strategies, such as performing HPT, quasi-experimental studies that focus on promoting HPT and include generic task approaches could provide valuable insights for the teaching and learning of history. Furthermore, more quasi-experimental research involving the spatial context and the position of a historical agent in society is needed as only a few students in our study displayed this in their reasoning. Thus, it would be interesting to see whether teacher instruction focused explicitly on the spatial context and the historical agents' position results in better HPT performance.

One study limitation is that we conducted exploratory research that included only 36 students in our thinking aloud protocols and focused on only one historical topic with one related assignment. The instrument's scenario was also fictional. Thus, more research on how students perform HPT when addressing real historical sources or other tasks about agents' decisions is needed. Furthermore, in the quantitative portion of our study, we included only four questions about students' chronological knowledge to measure prior knowledge. As this is a further limitation, future research should focus on the relationship between one's ability to perform HPT and one's prior knowledge and should include more questions on different types of historical knowledge to confirm the relationship we found when analyzing the thinking-aloud protocols. Another limitation is that our mean HPT score for the qualitative sample was slightly higher (3.39) than the mean HPT score for the quantitative sample (3.20). Furthermore, as we only included pre-university students in our quantitative sample, it would be interesting to compare their HPT ability with students' HPT abilities at other educational levels and to examine possible differences in students' specific needs to successfully perform HPT.

Finally, we discuss some practical implications for the teaching of history. Although the majority of the students in our study did not view the past from a present-oriented perspective, six students did do so. To decrease students' presentism, Huijgen and Holthuis (2015) presented a sample lesson about the rise of Hitler that was shaped by the theory of constructive controversy (Johnson & Johnson, 2009). Teachers could use these types of exercises to evaluate students' presentism and to scaffold historical contextualization.

Furthermore, our study, similar to those of Levstik (2011) and Havekes et al. (2012), indicated that building extensive and different frames of reference could help students perform HPT. However, as Reisman and Wineburg (2008) noted, it does not result in the automatic application of historical knowledge, as HPT also requires a deep understanding of the difference between past and present. Accordingly, this is not an easy task for teachers, as Levstik and Groth (2002) noted. However, lessons combining historical contextualization with historical empathy tasks could promote this understanding. Recently, Endacott and Pelekanos (2015) presented a good example of such a lesson when teaching a unit on ancient Athens. As shown in this study, HPT is a complex process, but structural attention and classroom practice can promote students' understanding of the past and help them prepare to participate in a civic society.



CHAPTER 4

TESTING A HISTORICAL CONTEXTUALIZATION OBSERVATION INSTRUMENT

This chapter describes the development and testing of the Framework for Analyzing the Teaching of Historical Contextualization (FAT-HC). This high-inference observation instrument focuses on history teachers' competency in promoting historical contextualization in classrooms. Generalizability studies were conducted to assess the instrument's dimensionality and reliability by decomposing the instrument's variance. A large proportion of the variance was explained by differences between observed teachers, and a small proportion of the variance was explained by lessons and observers, demonstrating the instrument's reliability. Furthermore, a decision study was conducted to determine the optimal number of observers and lessons needed for a reliable scoring design.

This chapter is based on: Huijgen, T. D., Van de Grift, W. J. C. M., Van Boxtel, C. A. M., & Holthuis, P. (2017). Teaching historical contextualization: The construction of a reliable observation instrument. *European Journal of Psychology of Education*, *32*, 159-181. doi:10.1007/s10212-016-0295-8

4.1 Introduction

Since the 1970s, increasing attention to the evaluation of teachers' generic competencies has resulted in the development of a variety of observation instruments that are widely used to assess elementary and secondary education, such as the Stallings Observe System (Stallings & Kaskowitz, 1974), the Framework for Teaching (Danielson, 1996), the International System for Teacher Observation and Feedback (Teddlie, Creemers, Kyriakides, Muijs, & Yu, 2006), the International Comparative Analysis of Learning and Teaching (Van de Grift, 2007), and the Classroom Assessment Scoring System (Pianta, La Paro, & Hamre, 2008). Other instruments used to examine teacher behavior include, for example, teachers' self-reports, (semi-structured) interviews, and student questionnaires (e.g., Kyriakides, 2008; Kyriakides, Campbell, & Christofidou, 2002; Maulana, Helms-Lorenz, & Van de Grift, 2015; Muijs, 2006). However, despite its labor-intensive nature, classroom observation is viewed as a more unbiased form of data collection to examine teacher behavior (Pianta & Hamre, 2009; Wragg, 1994).

The development and implementation of observation instruments can be very useful in more effectively shaping teacher education and professional development programs and in evaluating classroom-based interventions (e.g., Darling-Hammond, 2012; Lavigne & Good, 2015; O'Leary, 2014; Yoder & Symons, 2010). However, most of these instruments focus on teachers' generic competencies rather than teachers' subject-specific competencies. Therefore, scholars such as Grossman and McDonald (2008), Desimone (2009), and Schoenfeld (2013) emphasized the importance of adding subject-specific observation instruments to research on teaching and teacher education.

Although some recently developed observation instruments focus on more specific teacher competencies, such as classroom talk (Mercer, 2010), project-based learning (Stearns, Morgan, Capraro, & Capraro, 2012), and the reform of learning and instruction (Sawada et al., 2002), only a few observation instruments focus on teachers' subject-specific strategies, such as English reading (Gertsen, Baker, Haager, & Graves, 2005; Smit, Van de Grift, De Bot, & Jansen, 2017), content- and language-integrated learning (De Graaff, Koopman, Anikina, & Westhoff, 2007), English language arts (Grossman et al., 2010), and mathematical instruction (Hill, Charalambous, & Kraft, 2012; Matsumura, Garnier, Slater, & Boston, 2008; Schoenfeld, 2013).

To date, however, there are no validated and reliable observation instruments that evaluate secondary-school history teaching. This is unfortunate, especially because, as noted by Achinstein and Fogo (2015), Bain and Mirel (2006), and Grant and Gradwell (2009), current teacher education and professional development programs may not meet history teachers' needs so that they can achieve the aims set by history curricula. Observation instruments that evaluate history teachers' subject-specific strategies could identify history teachers' specific needs and, thus, further improve teacher education and professional development programs for history teachers.

Van Hover, Hicks, and Cotton (2012) attempted to construct a validated observation instrument to evaluate secondary-school history teaching. Their Protocol for Assessing the Teaching of History (PATH) is promising, but information about the measure's reliability is lacking. In contrast to PATH, the observation instrument that we developed focuses on a single but highly important history teacher competency; promoting students' ability to perform historical contextualization. Historical contextualization is considered an important component of historical thinking and reasoning and is incorporated into history curricula worldwide (Lévesque, 2008; Seixas & Morton, 2013; Van Drie & Van Boxtel, 2008). In previous research, we examined how students performed on a historical contextualization task and found that secondary-school students of different ages experience difficulties in performing historical contextualization tasks (Huijgen, Van Boxtel, Van de Grift, & Holthuis, 2014). Therefore, we must gain greater insight into how history teachers promote students' ability to perform historical contextualization in classrooms. The purpose of the present study is, therefore, to construct a reliable high-inference observation instrument and scoring design to assess history teachers' competency in promoting historical contextualization in classrooms. In this study, we first present the theoretical framework and our research questions. Then, we present our methodology and results. Finally, we discuss our findings and present the practical implications of the results and directions for future research.

4.2 Theoretical framework

4.2.1 Teaching historical reasoning competencies

Scholars and other educational professionals widely agree that secondary-school history education should involve more than the simple learning of facts (e.g.,

Lévesque, 2008; Van Drie & Van Boxtel, 2008; Wineburg, 2001). Therefore, historical reasoning competencies, such as determining causality, investigating sources, asking rich historical questions, and performing historical contextualization, have become increasingly important in Western history education over the last two decades (Erdmann & Hasberg, 2011; Seixas & Morton, 2013). Some scholars also stress the importance of historical reasoning competencies for promoting students' democratic citizenship (e.g., Barton, 2012; Saye & Brush, 2004). To achieve historical reasoning competencies, students in history classes must be involved in engaging learning tasks and activities (Gerwin & Visone, 2006; Grant & Gradwell, 2010; Levstik & Tyson, 2008) and history lessons should extend beyond factual recall to achieve deep subject understanding (Bransford, Brown, & Cocking, 2000; Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009). However, both novice and experienced history teachers seem to struggle when they are asked to develop engaging learning tasks and teach students historical reasoning competencies (Monte-Sano, 2011; Van Hover & Yeager, 2004; VanSledright, 2010; Virta, 2002). Many history lessons might, therefore, have a strong focus on historical content knowledge (Saye & Social Studies Inquiry Research Collaborative, 2013; VanSledright, 2011).

4.2.2 Observing history education

To explore the challenges and problems that history teachers face, qualitative research studies have been conducted (e.g., Bain & Mirel, 2006; Fogo, 2014; Monte-Sano & Cochran, 2009; Virta, 2007). However, few quantitative research studies using standardized instruments have been conducted to explore history teachers' competencies (Adler, 2008; Ritter, 2012). For example, only two studies used observation instruments to examine how teachers actually teach historical content knowledge and historical reasoning competencies. Thus, the use of standardized observation instruments in research on history education is an underexamined topic, as Van Hover et al. (2012) noted:

> While the field of history education elucidates a clear and ambitious vision of high-quality history instruction, a current challenge for history educators (including teacher educators, curriculum specialists, and school-based history and social science supervisors) becomes how to illuminate and capture this when observing classrooms to research history instruction or to provide useful discipline-specific feedback to preservice (and inservice) history teachers. (p. 604)

Nokes (2010) used an observation instrument and focused on history teachers' literacy-related decisions about the types of texts they used and how students were taught to learn with these texts. Eight secondary-school history teachers were observed over a 3-week period using two frequency counting observation instruments, one instrument to record the type of texts and one to record teachers' activities and instruction; however, detailed information about the instruments' validity and (inter-rater) reliability is lacking. The other study was conducted by Van Hover et al. (2012), the only researchers who attempted to construct a subject-specific observation instrument, called the PATH, with the goal of evaluating and improving history instruction. PATH has the same structure as Pianta and Hamre's (2009) Classroom Assessment Scoring System-Secondary (CLASS-S) and consists of six dimensions: (1) lesson components, (2) comprehension, (3) narrative, (4) interpretation, (5) sources, and (6) historical practices. Each dimension includes indicators and behavioral markers that are scored "high," "middle," and "low" by observers. The authors tested the inter-rater reliability for PATH and found positive indicators, but detailed information about the instrument's validity and reliability is lacking.

4.2.3 Historical contextualization: a conceptualization

Rather than constructing an observation instrument for all historical reasoning competencies, we focus on how history teachers promote historical contextualization in classrooms. This focus provides us with the opportunity to spend sufficient time on item development and to test whether it is possible to observe history teachers' subject-specific strategies using an observation instrument. We chose historical contextualization because it is considered a key competency of historical reasoning (Davies, 2010; Lévesque, 2008; Seixas & Morton, 2013) and is, therefore, included in the formal history curricula of many countries, such as Australia, Belgium, Canada, Finland, Germany, the Netherlands, Spain, and the UK (Huijgen, Van Boxtel, Van de Grift, & Holthuis, 2014).

In history education, it is possible to contextualize historical sources and phenomena, including persons, events, and developments (Havekes, Coppen, Luttenberg, & Van Boxtel, 2012). Historical contextualization is the ability to situate a historical phenomenon or person in a temporal, spatial, and social context to describe, explain, compare, or evaluate it (Van Boxtel & Van Drie, 2012). Wineburg and Fournier (1994) defined historical contextualization as building a context of circumstances or facts

that surround a particular historical phenomenon to render it more intelligible. Endacott and Brooks (2013) viewed historical contextualization as:

> A temporal sense of difference that includes deep understanding of the social, political, and cultural norms of the time period under investigation as well as knowledge of the events leading up to the historical situation and other relevant events that are happening concurrently. (p. 43)

Historical events and historical agents' decisions must be placed in the specific socio-spatial and socio-temporal locations in which they emerged. For example, students must know that in ancient Roman times, Julius Caesar could not have had breakfast in Rome and dinner in the Gaul region of France on the same day because the transportation modes needed for such a trip was not available (Lévesque, 2008).

4.2.4 Teachers' strategies for promoting historical contextualization

Research has been conducted to conceptualize the instructional practices that effective teachers employ to promote historical contextualization in classrooms (e.g., Doppen, 2000; Rantala, 2011; Van Boxtel & Van Drie, 2012). To teach historical reasoning competencies such as historical contextualization, teachers must not only possess expert levels of subject content knowledge but also activate students to acquire knowledge and help them apply this knowledge to gain different historical reasoning competencies (Haydn, Stephen, Arthur, & Hunt, 2015). Additionally, Hattie's meta-analysis (2008) indicated that effective teachers activate student learning. Other meta-analyses on effective teaching seem to confirm this finding (e.g., Kyriakides, Christoforou, & Charalambous, 2013; Seidel & Shavelson, 2007). Exposure to information alone is not sufficient for students to gain deep subject-specific understanding and historical reasoning competencies. Based on research that focused on historical contextualization, we identified four main teaching strategies for promoting historical contextualization in classrooms: (1) reconstructing the historical context, (2) fostering historical empathy, (3) performing historical contextualization to explain the past, and (4) raising awareness of present-oriented perspectives when examining the past.

First, the historical context of a phenomenon must be reconstructed to perform historical contextualization. Foster (1999) argued that students must possess

historical context knowledge, including knowledge about chronology, before they can perform historical contextualization. Reisman and Wineburg (2008) also stressed the importance of background knowledge for the performance of historical contextualization. To reconstruct the historical context, students and teachers can use different frames of reference such as the chronological frame of reference, spatial frame of reference, or social frame of reference (e.g., De Keyser & Vandepitte, 1998; Pontecorvo & Girardet, 1993; Van Boxtel & Van Drie, 2012). The chronological frame includes knowledge of time and period, significant events, and developments (Dawson, 2009; Wilschut, 2012). The spatial frame focuses on knowledge of (geographical) locations and scale (Havekes et al., 2012). The social frame includes not only knowledge of human behavior and the social conditions of life but also knowledge of socio-economic, socio-cultural, and socio-political developments (Van Boxtel & Van Drie, 2004). To reconstruct the historical context, teachers and students should explore the different frames of reference. For example, in previous research, we found that most students who used and combined different types of knowledge (e.g., chronological, spatial, economic, political, and cultural knowledge) obtained higher scores on a historical contextualization task than students who used a single type of knowledge (Huijgen, Van Boxtel, Van de Grift, & Holthuis, 2017). Teachers could use different sources to build the different frames of knowledge, such as movies (Marcus, 2005; Metzger, 2012), written documents, objects, and images (Fasulo, Girardet, & Pontecorvo, 1998; Van Drie & Van Boxtel, 2008).

Second, although some scholars claim that historical empathy is idealistic and can never be fully achieved because most historical agents are dead (e.g., Kitson, Husbands, & Steward, 2011; Riley, 1998; Wineburg, 2001), most scholars agree that historical empathy could promote historical contextualization (e.g., Cunningham, 2007; Davis, 2001; Endacott & Brooks, 2013; Lee & Ashby, 2001; Skolnick, Dulberg, & Maestre, 2004). Historical empathy focuses on empathizing with people in the past based on historical knowledge that explains their actions. Colby (2008) noted that the primary purpose of historical empathy is to enable students to transcend the boundaries of *presentism* by developing a rich understanding of the past from multiple viewpoints. In history lessons, teachers could focus on a historical agent to gain insight into the views and values of people who lived in the past (e.g., Foster, 1999; Wooden, 2008) or discuss historical agents' decisions with a group of students (Kohlmeier, 2006). Teachers could also promote historical empathy by promoting the formation of affective connections with the historical agent based on students'

own similar yet different life experiences (Endacott & Pelekanos, 2015; Kitson et al., 2011) and focusing on understanding historical agents' prior knowledge and positions (Berti, Baldin, & Toneatti, 2009; Hartmann & Hasselhorn, 2008; Huijgen et al., 2017).

Third, students should be able to explain the past based on their historical context knowledge (Lévesque, 2008; Seixas & Morton, 2013; Wineburg, 2001). For example, students must explain why the Great Depression of 1929 spread to Europe or the differences between governance in ancient Greece and governance in the Middle Ages. To answer such historical questions, students must link the Great Depression and the different types of governance to their historical context (Seixas, 2006). Furthermore, the successful performance of different historical reasoning competencies, such as identifying indirect and direct causes (Stoel, Van Drie, & Van Boxtel, 2015), understanding change and continuity (Haydn et al., 2015), reasoning with historical sources (Reisman & Wineburg, 2008), and asking historical questions (Logtenberg, Van Boxtel, & Van Hout-Wolters, 2011), requires an analysis of the broader historical context. Teachers should, therefore, create opportunities for students to practice these competencies with these types of questions. Halldén (1997) suggested that teachers should focus their instruction on the relationship between historical factual details (lower-level context) and large historical developments (larger context). Kosso (2009) also noted that "Individual events and actions are understood by being situated in the larger context. However, the larger context is understood by being built of individual events. It is a hermeneutic circle and perhaps the only way to understand other people" (p. 24). Presenting and evaluating historical phenomena from different perspectives is also considered an effective approach (e.g., Ciardiello, 2012; Levstik, 1997; McCully, 2012; Stradling, 2003). For example, to understand and explain the Cuban Missile Crisis of 1962, students should examine this phenomenon from not only a capitalist Western perspective but also a communist Soviet perspective.

Finally, teachers should raise awareness of students' present-oriented perspective and the consequences of this perspective when examining the past (Barton & Levstik, 2004; Huijgen et al., 2014; Wineburg, 2001). Students must know that the past differs from the present (Seixas & Peck, 2004); however, social psychology research illustrates that young students especially find it very difficult to take another persons' perspective, particularly when that other person does not have the same knowledge that the students have (Bloom & German, 2000; Wellman, Cross, & Watson, 2001). This inability could cause problems in history education, as students must be aware

that much of the information that they know was not available to people in the past. Students' present-oriented thinking or presentism is considered one of the main reasons why they fail to achieve historical contextualization and could cause misconceptions among students, leading them to reach incorrect conclusions about historical phenomena (Huijgen et al., 2014; Lee & Ashby, 2001; VanSledright & Afflerbach, 2000).

Although we can never be perfectly non-presentist (e.g., Pendry & Husbands, 2000; Wineburg, 2001), teachers should foster students' awareness of their own contemporary values and beliefs and the consequences of this perspective when explaining the past. To achieve this goal, teachers could present the past as tension for students (e.g., Savenije, Van Boxtel, & Grever, 2014; Seixas & Morton, 2013), present conflicting historical sources (Ashby, 2004), not present the past as progress (Wilschut, 2012), and promote intellectual conflict regarding historical phenomena that might be difficult for students to understand and explain (Foster, 2001; Huijgen & Holthuis, 2015). Furthermore, to prevent students from viewing the past from a present-oriented perspective, teachers should explicitly model or scaffold how historical contextualization can be performed successfully, for example, by providing learning strategies. Explicit teaching of domain-specific strategies, such as how to perform historical contextualization, could promote students' ability to explain historical events (Stoel et al., 2015). Reisman and Wineburg (2008) stressed the importance of explicitly providing students with an illustration of contextualized thinking, for example, by providing videos of good examples of professional historians who scaffold their contextualization processes.

4.3 Research questions

A subject-specific observation instrument could provide insight into the instructions and methods that history teachers employ to promote students' ability to perform historical contextualization. Therefore, we aimed to construct a reliable subject-specific observation instrument and scoring design that measures how history teachers promote historical contextualization in classrooms. To address this central aim, we specify the following three research questions:

1. What is the observation instruments' dimensionality when used to observe how history teachers promote historical contextualization?
2. What are the reliability outcomes when the observation instrument is used to observe how history teachers promote historical contextualization?
3. How many lessons and observers are necessary to establish a reliable and optimal scoring design?

4.4 Method

4.4.1 Structure of the observation instrument

To design and construct our observation instrument, we used the guidelines described by Colton and Covert (2007), which focus on the development of valid and reliable instruments in social sciences. Our instrument could be characterized as a high-inference observation instrument. In contrast to low-inference instruments (such as time sampling and time logs), high-inference instruments provide a more qualitative verdict (Chávez, 1984). However, these instruments are more susceptible to subjectivity; therefore, thorough inter-rater reliability procedures are necessary.

We modeled our instrument on Van de Grift's (2007, 2009) International Comparative Analysis of Learning and Teaching (ICALT) observation instrument. We chose this instrument's format because it also seeks to observe teachers' professional strategies and calculate scores based on these strategies. Similar to the ICALT instrument, our instrument utilizes a 4-point Likert scale to score the items. In our instrument, scores 1 and 2 represent a negative verdict, while scores 3 and 4 represent a positive verdict. Score 1 should be used only if teachers do not use a particular strategy in their lessons.

4.4.2 Formulating and refining the items

Based on the four main strategies identified in our theoretical framework (reconstructing the historical context, fostering historical empathy, performing historical contextualization to explain the past, and raising awareness of a present-oriented perspective) and a review of literature on teaching historical contextualization, we formulated observable items to assess classroom teachers' behavior in regards to historical contextualization. Furthermore, during two national teacher professionalization conferences, we asked 25 history teachers (after an

introduction of the concept of historical contextualization) to each formulate 20 items that assess classroom teachers' behavior in regards to historical contextualization. Combining these items with the items that we formulated resulted in a total of 121 items.

Meta-analyses on effective teaching illustrate that promoting different types of interactions in classrooms (i.e., student-student interactions and teacher-student interactions) could promote student learning (Kyriakides et al., 2013; Seidel & Shavelson, 2007). Therefore, we formulated three items (the teacher asks evaluative questions; the teacher uses classroom discussion; the teacher uses group work), focusing on more generic teacher strategies and different (social) interactions in the classrooms. We included these three generic items because history education research shows that these types of interaction could promote historical reasoning competencies (e.g., Brooks, 2008; Van Drie & Van Boxtel, 2008; Van Drie, Van Boxtel, & Van der Linden, 2006; Stoel et al., 2015). Therefore, the total list included 124 items. By excluding double items and items that might be very difficult to evaluate, we shortened the list to 82 items. For example, we first included individual items for all time indicators (e.g., year, period, and century), but we then incorporated these items into one item, "the teacher gives time indicators." Another example is that we excluded items focusing on the specific economic, political, and social circumstances (e.g., form of government, welfare, scientific knowledge, wars, and laws) of historical phenomena. Because these specific circumstances are difficult to observe in one history lesson, we included only items such as "appoints political/governance characteristics at the time of phenomena" and "appoints socio-cultural characteristics at the time of phenomena." This method might result in a less nuanced image of a lesson, but we preferred to develop an instrument that allows us to observe all behavior indicators in a single history lesson.

Next, we organized two expert panel discussions to further shorten the list of 82 items and ensure the instrument's face and content validity. The first panel discussion was held with two history teacher educators and seven secondary-school history teachers. The second panel discussion was held with one history teacher educator and four secondary-school history teachers. All experts had more than 7 years of work experience. The experts were asked to (1) remove unnecessary items that did not measure history teachers' competency in terms of promoting historical contextualization, (2) remove possible multiple items that might cover the same

teacher behavior, (3) reformulate unclear items, and (4) formulate new items that they thought were missing. In total, the experts excluded 24 items, reformulated 12 items, and created no new items, resulting in a list of 58 items.

Subsequently, we trained 10 student history teachers on the use of the observation instrument, and they observed one videotaped history lesson using the instrument. We calculated Cronbach's alpha (jury alpha) for their observation scores to explore the instrument's internal consistency. This jury alpha was .58 (poor internal consistency). After deleting 10 items that threatened internal consistency, the jury alpha increased to .81 (good internal consistency). Examples of the deleted items are "appoints relations between historical phenomena," "uses substantive concepts when explaining historical phenomena," and "uses general schemas to explain historical phenomena." We asked the experts in the first panel session to determine whether the 10 deleted items could jeopardize the instrument's face and content validity; they found no threats. The same experts were also asked to observe three videotaped history lessons taught by three different history teachers using the 48 items. After discussing each lesson, three items ("explains the importance of placing phenomena in a chronological framework," "explains the importance of placing phenomena in a spatial framework," and "explains the importance of viewing phenomena from different dimensions") led to strong disagreement among the experts; thus, we deleted these items. This resulted in a total list of 45 items in the first version of the Framework for Analyzing the Teaching of Historical Contextualization (FAT-HC).

4.4.3 Research design

Following Hill et al. (2012), we adopted generalizability theory to explore the instrument's dimensionality and to determine its reliability (Brennan, 2001; Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Shavelson & Webb, 1991). Compared to the classical test theory, generalizability theory is more informative and useful in educational systems because the classical test theory considers only one source of measurement error at a time. Additionally, it does not result in specific information on how many forms, items, occasions, or observers are required (Shavelson, Webb, & Rowley, 1989). A generalizability study (G-study) can accommodate any observational situation and is restricted by only the practical limitations of data collection and software (Lei, Smith, & Suen, 2007). A G-study views a behavioral measurement (for example, an observed score) as a sample from a universe of admissible observations. Each aspect (called a facet) in the measurement procedure is considered a possible

source of error. A G-study provides estimates of the variance contributed by persons, observers, occasions of measurement, and each of the possible interactions between these facets. Generalizability theory distinguishes a decision study (D-study) from a G-study. A D-study uses information from a G-study to construct a scoring design that minimizes error for a particular purpose (Shavelson & Webb, 1991). In addition to a G-study, a D-study can identify the optimal data collection for a desired score reliability (Hill et al., 2012).

4.4.4 Sample and data collection

Non-probability sampling was used to select five teachers to observe and five observers (see Tables 11 and 12 for the teachers' and observers' characteristics). In the Netherlands, the average age of male teachers is 46 years and that of female teachers is 42 years. The gender distribution of the teachers was 48% female and 52% male. In total, there are 1785 history teachers with a Master's degree and 3944 history teachers with a bachelor's degree working in the Netherlands (Dutch Ministry of Education, Culture, and Science, 2011). The teachers in the sample worked at different schools, and these schools did not differ significantly from the total population in regards to student enrollment, location (rural or urban), or graduation rate (Statistics Netherlands, 2014). The national students' mean score on the formal history exam for general secondary education and pre-university education was 6.35 on a 10-point scale.

Table 11. Teachers' characteristics

Teacher	Gender	Age	Educational qualification	Years' work experience	Nationality	Historical expertise	Students' performance*
A	Male	60	Bachelors	37	Dutch	Modern History	=
B	Male	34	Masters	8	Dutch	Modern History	>
C	Male	43	Masters	17	Dutch	Early Modern History	>
D	Male	63	Masters	41	Dutch	Middle Ages	>
E	Male	41	Masters	14	Dutch	Early Modern History	>

Note. *Students' mean score on the formal history exam compared to the national mean score on the formal history exam.

Table 12. Observers' characteristics

Observer	Gender	Age	Educational qualification	Years' work experience	Nationality
1	Female	29	Masters	7	Dutch
2	Female	32	Masters	7	Dutch
3	Male	32	Masters	8	Dutch
4	Male	33	Masters	7	Dutch
5	Male	29	Masters	8	Dutch

We videotaped two different lessons for each teacher ($n = 5$), and all lessons were taught in the two highest tracks of secondary education in the Dutch educational system. We observed only the lessons for upper secondary school students in the two highest tracks because the Dutch formal exam program considers the ability to perform historical contextualization to be an important aim for these students (Board of Tests and Examinations, 2015). A total of 267 students, with a mean age of 16.2 (SD = 0.7) years old, were involved. The mean duration of analyzed lessons was 39 min (SD = 2.4). Each observer individually evaluated the 10 videotaped lessons using the developed observation instrument, yielding a total of 50 observations.

4.4.5 Training observers to use the instrument

All observers received a 4-hour training. In this training, we used three videotaped history lessons taught by three history teachers (one female teacher with more than 15 years of work experience, one male teacher with 4 years of work experience, and one male teacher with more than 25 years of work experience) from three different schools as training materials. One lesson was about the Ancient Roman period, one was about the Middle Ages, and one was about the Second World War. These three lessons were not used in our data analyses. The observers received an explanation of the 45 items and evaluated the videotaped lessons using a training version of the observation instrument that included more in-depth explanations of the items. After the observers observed each videotaped lesson, their results were discussed, and some items were clarified by the trainers to minimize inter-rater bias.

4.4.6 Data analysis

To explore the instrument's dimensionality, we conducted a G-study at the item level with seven facets in a crossed design. To estimate the reliability of our instrument and produce a composite of scores with maximum generalizability, we conducted a new G-study and employed multivariate generalizability using a "t × I × o" design, where t represents the observed history teachers, I represents the number of observed

lessons, and o represents the number of observers. To determine the optimal number of observers and lessons needed in a scoring design to achieve acceptable reliability, we conducted a D-study using the information from the earlier conducted G-study that estimated the reliability of our instrument.

4.5 Results

4.5.1 The instrument's dimensionality

Based on our theoretical framework, we consider our instrument to be one-dimensional because all items should measure teachers' ability to promote historical contextualization. The first data analysis indicated that five items ("the teacher asks evaluative questions," "the teacher uses classroom discussion," "the teacher uses group work," "the teacher compares phenomena with the present," and "the students compare phenomena with the present") displayed a low correlation (< .30) with the other items. These five items also obtained a standard deviation above 1.00 and were excluded from further data analysis, resulting in a total list of 40 items in the final version of the FAT-HC observation instrument (see Appendix D).

To further explore the instrument's dimensionality, we conducted a G-study at the item level with seven facets in a crossed design using the collected data of the five observers who each evaluated two lessons taught by five teachers (50 observations in total). If our instrument is, in fact, one-dimensional, the item facet should explain the main part of the overall variance and the other facets (including the interaction effects) should explain a lesser part of the variance (e.g., Brennan, 2001; Shavelson & Webb, 1991). As shown in Table 13, the item facet was responsible for most of the variance (47.25%), indicating that our instrument is one-dimensional in regards to observing how history teachers promote historical contextualization in classrooms.

4.5.2 The instrument's reliability

To determine the reliability of our instrument, a new G-study was conducted using the same data set (50 observations). The analysis was conducted on the final version of our observation instrument, which consisted of 40 items (see Appendix D). Table 14 displays the results of this G-study and presents the variance decomposition to assess the instrument's reliability. A reliable instrument should have a high proportion of the variance explained by differences between the observed teachers and a low proportion of the variance explained by lessons and observers.

Table 13. Variance decomposition for the item level

Variance components	Estimated variance	Percentage of variance
Item	0.43	47.25
Item*Teacher	0.10	10.99
Item*Observer	0.03	3.30
Item*Lesson	0.00	0.00
Item*Teacher*Observer	0.02	2.20
Item*Teacher*Lesson	0.08	8.79
Item*Lesson*Observer	0.00	0.00
Residual	0.25	27.47
Total	0.91	100.00

Note. * = interaction effect.

Table 14. Variance decomposition for the observation instrument

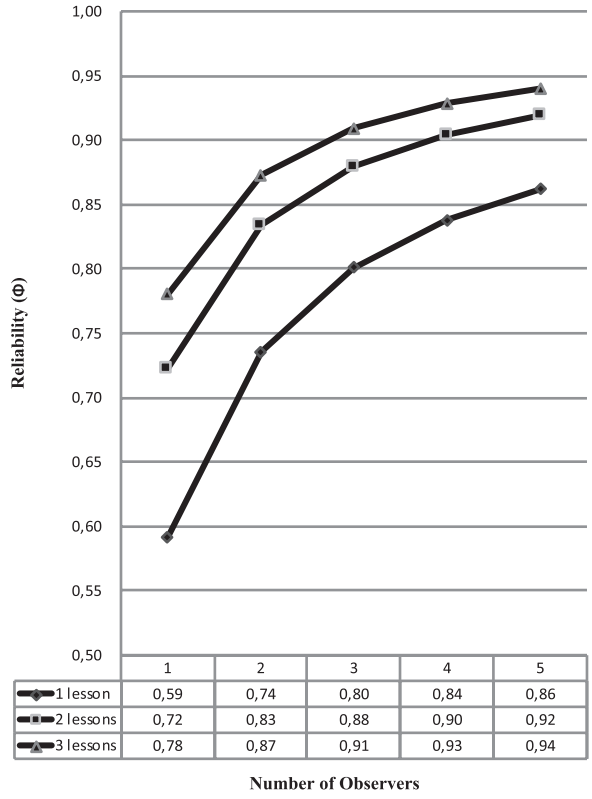
Variance Components	Estimated variance	Percentage of variance
Teachers	30.05	59.12
Observers	2.33	4.58
Lessons	0.83	1.63
Residual	17.62	34.67
Total	50.83	100.00

The difference between the observed teachers accounted for 59.12% of the variance, the difference between the observers accounted for 4.58% of the variance, and the difference between the lessons accounted for 1.63% of the variance. The residual was 34.67%. The results show that the influence of the observers and lessons was very low, indicating that the observers and lessons can be considered to be inter-changeable and that the observers understood the observation items. Interaction effects between the different facets (observers*lessons, observers*teachers, and teacher*lessons) were also calculated and did not display any variance, indicating small differences between the observers' observations of the different teachers and lessons.

4.5.3 The optimal reliable scoring design

To identify the optimal number of observers and lessons needed for a reliable scoring design, we conducted a D-study based on the results of our G-study, which estimated the instrument's reliability. Because we are interested in the absolute level of an individual's performance independent of others' performance, we calculated the index of dependability coefficient (Φ) to identify the optimal number of observers (Shavelson & Webb, 1991). The Φ should be $\geq .7$ for research purposes, $\geq .8$ for formative evaluations, and $\geq .9$ for summative evaluations (Brennan & Kane, 1977).

The results of our D-study can be found in Figure 3. A scoring design with one observer evaluating one lesson taught by a teacher yields a Φ of .59 (poor reliability), and this value increases to Φ = .72 when one observer evaluates two lessons taught by the same teacher. Because we are interested in research purposes and formative evaluations, the optimal scoring design would use two observers who each evaluate two different lessons taught by the same teacher (Φ = .83) or three observers who each evaluate the same lesson taught by a teacher (Φ = .80).



	1	2	3	4	5
1 lesson	0,59	0,74	0,80	0,84	0,86
2 lessons	0,72	0,83	0,88	0,90	0,92
3 lessons	0,78	0,87	0,91	0,93	0,94

Figure 3. Results of the D-study

4.6 Conclusion and discussion

The aim of the present study was to develop a reliable observation instrument and scoring design to assess how history teachers promote historical contextualization in classrooms. This study resulted in the FAT-HC observation instrument. Using

expert panels, we found positive indicators of the instrument's content validity. Furthermore, generalizability theory analysis provides indicators that the instrument is one-dimensional when used to evaluate how history teachers promote historical contextualization. Generalizability theory analysis also showed that a large proportion of the instrument's variance was explained by the differences between the observed teachers and a small proportion of the variance was explained by the differences in lessons and observers, which demonstrates the instrument's reliability (Brennan, 2001; Hill et al., 2012; Shavelson & Webb, 1991). Our D-study showed a reliable scoring design, with one observer evaluating two lessons as the most effective method for research purposes. For formative teacher evaluations, a reliable scoring design in which two observers each evaluate two lessons or three observers each evaluate one lesson is most effective.

Van Hover et al. (2012) noted that instruments that provide "useful discipline-specific feedback to preservice (and inservice) history teachers" (p. 604) are lacking. Additionally, Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein (2012) emphasized that most current teacher evaluation programs do little to help teachers improve their teaching. The FAT-HC instrument could provide insight into teachers' subject-specific needs, resulting in a valuable addition to existing generic observation instruments (Grossman & McDonald, 2008). For example, if a teacher obtains low scores on the instrument, attention could be devoted to specific items of the instrument in teacher education or professional development programs. The pre-observation and post-observation interviews also could be structured based on the instrument's items, resulting in more concrete feedback for the observed teacher.

The instrument could also help researchers examine the instructions and methods that teachers employ to promote historical contextualization in classrooms. In the history education literature, there is a clear view of high-quality teaching and learning of history; however, research instruments that capture this view when observing history teachers while they work do not exist (Van Hover et al., 2012). Furthermore, our instrument could be used to gain more insight into the association between history teachers' instructions and methods and student achievement. Do teachers who activate their students to reconstruct a historical context better promote students' historical understanding than teachers who do not? The instrument could also be used to evaluate intervention studies, for example, to examine the effects of training teachers in the use of instructions incorporated into the observation instrument.

In addition to the function of the research instrument and feedback instrument, the instrument could be used as a framework for teachers who want to reshape and improve their instruction on historical contextualization. Slavin (1996) noted that teachers who explicitly model and scaffold their instructions contribute to their students' academic success. The instrument's strategies and items could provide direction for designing meaningful learning tasks and scaffolds for students. This is important, especially because, as noted by Grant and Gradwell (2010), many history teachers focus on recalling factual knowledge despite the fact that the teaching and learning of history includes far more activities, such as investigating sources and evaluating the past (VanSledright, 2008). Bain and Mirel (2006) and Saye and Social Studies Inquiry Research Collaborative (2013), therefore, argued that instruction models that help teachers learn how to promote students' ability to perform historical contextualization or other historical reasoning competencies are needed. In a post-observation interview, one of our observed teachers noted that he now uses the instrument as a checklist when designing his lessons. Prior to the study, he would forget the spatial context of historical phenomena. However, he now structurally includes the geographical context in his lessons when reconstructing the historical context of phenomena.

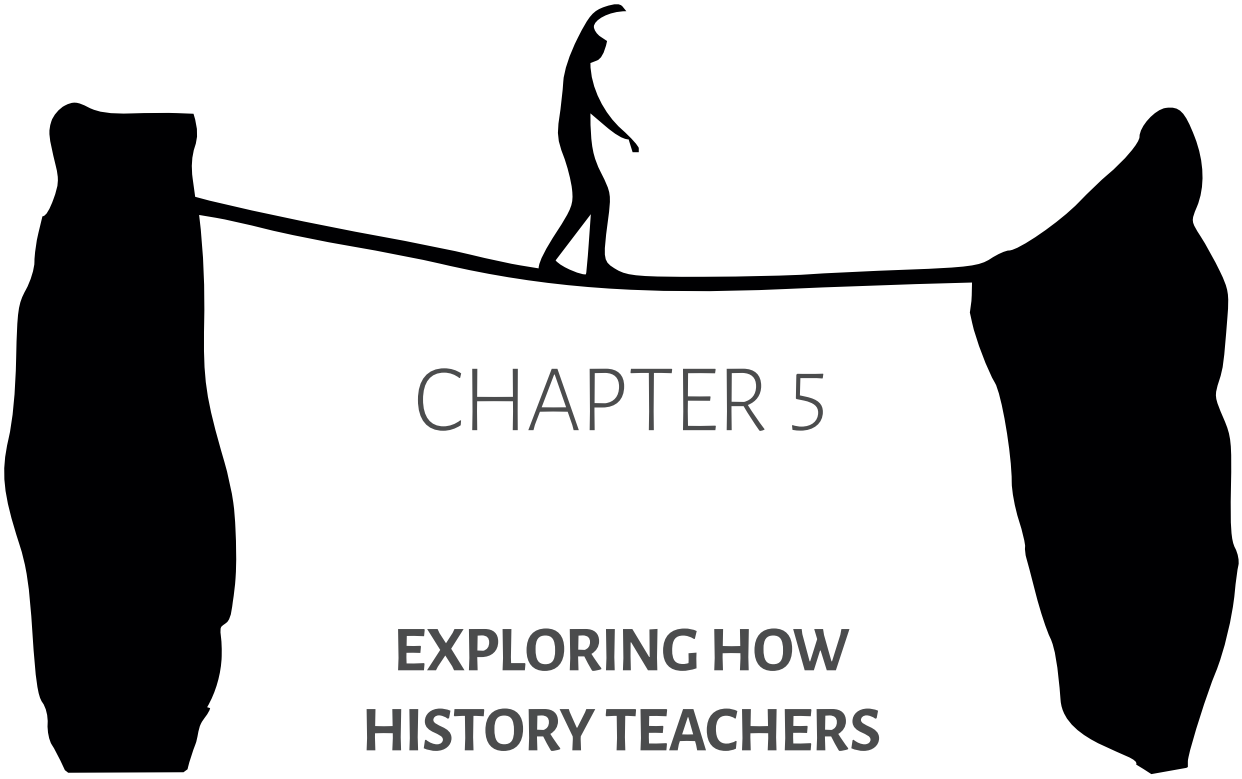
Despite the positive indicators of the instrument's reliability, some limitations must be acknowledged. We used a research design with only five observers and five teachers, who participated voluntary and, thus, might be more eager to learn (Desimone, 2009; Desimone, Smith, & Ueno, 2006). More observers, teachers, and lessons (cf. Hill et al., 2012) are needed to provide greater insight into the instrument's dimensionality, reliability, and optimal scoring design. Including teachers and observers with more varied backgrounds (e.g., differences in gender, student performance, age, and educational qualification) might also provide useful insights to further strengthen the instrument and scoring design. Furthermore, when examining the instrument's reliability, nearly 35% of the variance (residual) could not be explained by teacher, observer, or lesson variance. Future research and analyses must be conducted to decrease the residual variance and achieve greater reliability.

The observers also noted that it is difficult to evaluate 40 items when observing one history lesson. Because the observation instrument must be practical and suitable for observing a single lesson, more research is needed to decrease the number of items while maintaining good reliability. A larger G-study including a D-study, which

focuses on how many items are necessary to achieve reliability, could provide these insights (Brennan, 2001). We also used videotaped lessons. Although videotaped lessons have many benefits and are widely used for constructing and validating observation instruments (e.g., Yoder & Symons, 2010), they differ from "live" classroom observations. Future research should include live observations to assess possible differences in the instrument's reliability for live vs. videotaped sessions. Live video classroom observations (e.g., Liang, 2015) could also be an interesting method to examine possible differences in reliability.

To further assess the instrument's construct validity, intervention studies with a quasi-experimental design and pre- and post-tests to further test the framework's efficiency for promoting historical contextualization are needed. The use of other methods to assess teacher factors, such as student questionnaires and teachers' self-reports on historical contextualization, could also provide important insights into the instrument's construct validity (e.g., Kyriakides, 2008; Muijs, 2006). Additionally, Rasch modeling could provide information on the instrument's reliability, which items history teachers find more difficult to perform and which items they consider easier to perform (e.g., Fischer & Molenaar, 1995; Maulana, Helms-Lorenz, & Van de Grift, 2014; Van de Grift, Helms-Lorenz, & Maulana, 2014).

In conclusion, Ball and Forzani (2009) noted that current teacher education programs are often centered on teachers' beliefs and knowledge and argued that teacher education programs should mainly focus on the task and activities of teaching. They concluded that far more research is needed to gain insight into the tasks and activities of teaching across different subjects. We hope that our instrument can contribute to further insights into teachers' subject-specific activities for the teaching and learning of historical contextualization. Our instrument is not designed to assess history teachers; rather, it should function as a tool used to improve history instruction. Marriott (2001) noted that "Teachers seldom have a clear idea about their strengths and weaknesses. This is often because they have not been systematically observed and constructively debriefed" (p. 6). History teachers could observe each other using the instrument, discuss their lessons and findings, and collaboratively design new lessons with the instrument as framework, which might result in a giant step forward in the teaching and learning of history.



CHAPTER 5

EXPLORING HOW HISTORY TEACHERS PROMOTE HISTORICAL CONTEXTUALIZATION

This chapter explores how history teachers promote historical contextualization in their lessons. Using the Framework for Analyzing the Teaching of Historical Contextualization (FAT-HC), five trained raters observed eight history teachers twice. To further analyze the observation scores, the FAT-HC items were divided into eight categories while distinguishing between items that demonstrate historical contextualization and items focusing on engaging students in historical contextualization processes. The results indicate that the teachers in the sample did not explicitly promote historical contextualization in their lessons. No teacher obtained a mean FAT-HC score > 2.00 on a 4-point scale. The teachers mainly demonstrated historical contextualization, while engaging students in historical contextualization processes was observed far less often.

This chapter is based on: Huijgen, T. D., Holthuis, P., Van Boxtel, C. A. M., & Van de Grift, W. J. C. M. (2018). Promoting historical contextualization in classrooms: An observational study. *Educational Studies*. Advance online publication. doi:10.1080/03055698.2018.1509771

5.1 Introduction

An important goal of modern Western history education is the teaching of historical reasoning competencies, such as examining change and continuity, asking historical questions, and performing historical contextualization (Counsell, Burn, & Chapman, 2016; Seixas, 2015; Wineburg, 2001). Students not only have to possess historical content knowledge but also need to reason with this knowledge. In many countries, historical reasoning competencies therefore comprise a large part of the formal history curriculum (Erdmann & Hasberg, 2011). To acquire historical reasoning competencies, students need to be actively engaged in domain-specific learning processes, such as working with historical sources, determining causes and consequences, and engaging in historical contextualization (e.g., Lévesque, 2008; Seixas & Morton, 2013; Van Drie & Van Boxtel, 2008). History teachers therefore play a key role in teaching students how to examine historical phenomena within the confines of the discipline (Bain & Mirel, 2006; VanSledright, 2011).

In this study, we focus on how historical contextualization is promoted in classrooms. Historical contextualization is considered an essential skill for historians (e.g., Gaddis, 2002; Sewell Jr., 2005), a key component of historical thinking and reasoning (e.g., Seixas & Morton, 2013; Van Drie & Van Boxtel, 2008; Wineburg, 2001), and a possible contributor to instilling democratic citizenship in students (e.g., Barton, 2012; Barton & Levstik, 2004; McCully, 2012). The Dutch formal history curriculum therefore considers the ability to perform historical contextualization important (Board of Tests and Examinations, 2017). Moreover, Nikitina (2006) argues that the ability to perform historical contextualization is also important in other school subjects (e.g., when teaching the scientific development of the atomic bomb in science classrooms or when discussing Mark Twain's novel *Huckleberry Finn* in English classrooms).

Despite the importance of historical contextualization, different studies indicate that students experience difficulties when asked to perform historical contextualization tasks (Foster, Ashby, & Lee, 2008; Huijgen, Van Boxtel, Van de Grift, & Holthuis, 2014; Wineburg, 2001). Students may be inclined to view the past from a present-oriented perspective, and this is considered one of the main reasons that students fail to achieve historical contextualization, resulting in the misunderstanding of historical phenomena (Lee & Ashby, 2001; Seixas & Peck, 2004). For example, some students cannot explain why someone voted for the Nazi Party of Hitler in the 1930s (Hartmann

& Hasselhorn, 2008) or why forced marriages took place in the 15th century (Angvik & Von Borries, 1997) due to a present-oriented perspective. History teachers should therefore explicitly teach students historical contextualization (Lévesque, 2008).

However, little is known about the extent to which history teachers demonstrate historical contextualization themselves in their history lessons and how they engage students in historical contextualization processes. Previous observational studies focused more on general history teachers' classroom behavior. For example, Van Hover, Hicks, and Cotton (2012) included general history teachers' instructional practices, such as writing, simulations, and discussion, in their developed observation instrument. Nokes (2010) developed and used an observation instrument to examine history teachers' practices but focused on their literacy-related decisions, such as the texts they used as well as activities and instruction they provided in association with various types of texts. Huijgen, Van de Grift, Van Boxtel, and Holthuis (2017) developed the Framework for Analyzing the Teaching of Historical Contextualization (FAT-HC), which is a more specific observation instrument. However, they focused on the reliability of the instrument and did not present any results on how history teachers promoted historical contextualization in classrooms. The aim of this study is therefore to build upon the work of Huijgen, Van de Grift, et al. (2017) and to explore how teachers promote historical contextualization in their classrooms using the FAT-HC.

5.2 Theoretical framework

5.2.1 Historical contextualization

The ability to perform historical contextualization has become important in Dutch history education (Van Boxtel & Van Drie, 2012). Since the implementation of a framework of overview knowledge (consisting of 10 historical periods with associated key features) in the Netherlands in 2007, students have to use this framework to contextualize historical events, agents' actions, and sources to explain, compare, or evaluate them (Board of Tests and Examinations, 2017; Wilschut, 2012).

Some studies define historical contextualization as one heuristic that can be applied (in addition to corroboration and sourcing) to examine historical sources (e.g., Britt & Aglinskas, 2002; Wineburg, 1998). However, in history education, it is possible

to contextualize historical agents' actions, historical events, or historical sources (Havekes, Coppen, Luttenberg, & Van Boxtel, 2012). Therefore, in this study, we use the definition of Van Boxtel and Van Drie (2012) and conceptualize historical contextualization as an activity in which one situates phenomena and people's actions in the context of time, historical locations, long-term developments, or specific events to explain, compare, or evaluate these phenomena and actions. Huijgen, Van de Grift, et al. (2017) distinguished four interrelated components of historical contextualization: (1) reconstructing the historical context, (2) enhancing historical empathy, (3) using knowledge of the historical context to explain historical phenomena, and (4) enhancing the awareness of present-oriented perspectives among students when examining the past.

Reconstruction of a historical context needs to consider chronological, spatial, socio-political, socio-economic, and socio-cultural frames of reference (De Keyser & Vandepitte, 1998). The chronological frame includes knowledge of time periods and chronological knowledge of significant events and developments. The spatial frame focuses on knowledge of (geographical) locations and scales, and the social frames include knowledge of human behavior and the social conditions of life as well as knowledge of economic and political developments. When students do not consider these frames of reference, they are often not able to explain, compare, or evaluate historical phenomena and historical agents' actions (Reisman & Wineburg, 2008). For example, to understand and explain the Valais witch trials between 1428 and 1447, students need to situate these witch hunts in the isolated and mountainous border region of France and Switzerland during the late Middle Ages (chronological and spatial context). Furthermore, students have to consider that this region endured a civil war from 1415 to 1419, that the clans of the nobility fought each other, and that society was in a state of heightened tension (political, economic, and cultural context).

When historical empathy is used to promote historical contextualization, it can be seen as an interplay between an affective and cognitive element. The affective element is that students need to consider how historical agents' lived experiences, situations, or actions may have been influenced by their affective response based on a connection made to one's own similar yet different life experiences (Endacott & Brooks, 2013). A more cognitive element is that students need to examine the role and position of a historical agent, which includes understanding another's prior lived experience, principles, positions, attitudes, and beliefs (Hartmann & Hasselhorn, 2008).

Students should not only reconstruct the historical context of a historical phenomenon, but this context should be used to construct or evaluate a historical reasoning (Van Drie & Van Boxtel, 2008). Historical contextualization becomes meaningful when it helps to explain historical phenomena, make comparisons, or understand processes of change and continuity (Van Boxtel & Van Drie, 2016). Students should therefore be engaged in tasks in which historical contextualization is needed to explain, compare, or evaluate historical phenomena and historical agents' actions.

A final component of historical contextualization is raising awareness of students' present-oriented perspectives or *presentism*. Viewing the past from a present-oriented perspective leads to the misunderstanding of historical phenomena and agents' actions (Lévesque, 2008; Wineburg, 2001). Students therefore have to become aware of the differences between the past and present and evaluate the past on its own terms (Seixas & Morton, 2013).

5.2.2 Students' ability to perform historical contextualization

Compared to adults, elementary and secondary school students experience difficulty adopting a perspective that is different from their own, especially when this perspective is not consistent with the knowledge they have (Birch & Bloom, 2007). In history education, where students must be aware that people in the past may not have had the same information that the students possess now, this may lead to a misunderstanding of historical events (Seixas & Peck, 2004). For example, this could result in viewing historical agents as "stupid" or "that they did not know any better" (cf. Lee & Ashby, 2001).

Different studies have focused on how students perform historical contextualization. Hartmann and Hasselhorn (2008) examined how 170 German 10th graders performed historical contextualization to explain a historical agent's decision. Most students (66%) in their sample obtained a moderate score on the ability to explain a historical agent's decision, 24% obtained a very high score, and 10% obtained a very low score. Huijgen et al. (2014) used the same task to examine how 1,270 Dutch upper elementary and secondary school students (ranging in age from 10 to 17 years) performed historical contextualization. They concluded that older students achieved higher scores than younger students. This finding also appeared in a study by Berti, Baldin, & Toneatti (2009), who interviewed a total of 150 students (8 to 25 years old) to examine the concept of ordeals among children and young adults. Recently, studies

also have focused on how students' ability to perform historical contextualization can be advanced. Huijgen, Van Boxtel, Van de Grift, & Holthuis (2017) found indicators that secondary school students (15- and 16-year-olds) who combined different frames of reference were more successful in explaining historical agents' decisions. Baron (2016) concluded that a visual coding system based on the use of reliable visual cues to establish a historical time period may help students contextualize historical documents. Van Boxtel and Van Drie (2012) found that students between the ages of 14 and 17 who connected images or textual elements with key historical concepts or knowledge of landmarks were able to create a historical context of historical images and documents with greater success.

5.2.3 Teaching historical contextualization

Not much is, however, known about how history teachers promote historical contextualization in classrooms. Seixas (1998) found that pre-service history teachers incorporated documents in their lesson plans that showed that thinking in the past differed to present thinking. However, different studies on history teacher classroom behavior convey the general image of a teacher who mostly uses the history textbook narratives and focuses on the transmission of historical content knowledge (Barton & Levstik, 2003; Saye & Social Studies Inquiry Research Collaborative, 2013). This "traditional" approach of history education appears to focus on students' ability to memorize (nationally) significant figures, events, and narratives (Carretero, Asensio, & Rodriguez-Moneo, 2012; Symcox & Wilschut, 2009).

Huijgen, Van de Grift, et al. (2017) developed and tested a domain-specific observation instrument focusing on historical contextualization called the Framework for Analyzing the Teaching of Historical Contextualization (FAT-HC). Their instrument was based on four teaching strategies on historical contextualization. The first strategy is reconstructing the historical context. Students need to possess historical context knowledge, including knowledge about chronology and spatial, and socio-economic, socio-cultural, and socio-political developments before they can perform historical contextualization successfully. The second strategy is increasing historical empathy—for example, by selecting a historical agent relevant to the topic under study and focusing on the role and position of the historical agent in society and promoting students' affective connections with the historical agent. The third strategy is enhancing the use of historical context knowledge. Not only do students have to reconstruct a historical context, they also must use it, for example, to determine

causes and consequences, compare historical phenomena, and understand different perspectives on phenomena. The final strategy is enhancing the awareness of present-oriented perspectives among students when examining the past. Without the awareness of the differences between past and present, students are not able to compare, explain, or evaluate the past. These teaching strategies can be applied in different (chronological) sequences in classrooms.

The FAT-HC can be used to examine how students are engaged in historical contextualization processes since it makes a distinction between items focusing on teachers *demonstrating* historical contextualization (e.g., the teacher gives time indicators) and items focusing on teachers *engaging* students in historical contextualization processes (e.g., the students give time indicators). When the teacher gives time indicators, the teacher mentions, for example, the year or historical period in which a historical event took place. When the students give time indicators, teachers ask students, for example, in which year or historical period a historical event took place. The focus of the FAT-HC is therefore on teacher lectures and behavior. Huijgen, Van de Grift, et al. (2017) used generalizability theory (e.g., Brennan, 2001; Shavelson & Webb, 1991) to test the observation instrument for reliability. They calculated an index of dependability coefficient (Φ) to determine the number of observed lessons and raters needed for a reliable observation score. Brennan and Kane (1977) argued that the Φ should be ≥ .7 for research purposes; in the Huijgen, Van de Grift, et al. (2017) study, the Φ was .74 when one lesson was observed by two raters and the Φ increased to .86 when one lesson was observed by five raters. However, Huijgen, Van de Grift, et al. (2017) did not use the instrument to examine how history teachers promote historical contextualization in classrooms, leaving important questions for educational professionals unanswered, such as "Which teaching strategies from the instruments do teachers use the most?" and "Do teachers only demonstrate historical contextualization or do they also engage students in historical contextualization processes?"

5.3 Research question and hypotheses

For this explorative study, we formulated the following research question: how do history teachers promote historical contextualization in their lessons? We focus in this observational study on the two highest tracks in the Dutch educational system (senior general secondary education and pre-university education) since the formal Dutch

history exam program of these educational tracks demands that students examine the differences between past and present and create a historical context when interpreting historical events (Board of Tests and Examinations, 2017). Moreover, we focus on students aged 14 to 17 years old because these students possess the historical content knowledge necessary to perform historical contextualization successfully (e.g., Hartmann & Hasselhorn, 2008).

To examine the research question, we formulated the following two hypotheses:

1. Since historical contextualization is considered a key component of historical thinking and reasoning and is included in the Dutch formal history curriculum, we expect that the teachers in our sample will demonstrate historical contextualization in their lessons.
2. Since research indicates that many teachers focus on the transmission of historical content knowledge, we expect that the teachers in our sample seldom encourage their students to engage in historical contextualization processes themselves.

5.4 Method

5.4.1 Research design

To answer our research question, we used systematic observational measurement (Suen & Ary, 2014; Yoder & Symons, 2010). This approach allowed us to examine the data within the situation in which the activities took place (i.e., the classroom). Other methods, such as interviews, student and teacher questionnaires, or self-reports, did not offer this option (George & Bennett, 2004). Moreover, despite its labor-intensive nature, classroom observation is viewed as a more unbiased form of data collection to examine teacher behavior compared to other methods (Pianta & Hamre, 2009). This is stressed by VanSledright, Kelly, and Meuwissen (2006), who argue that teachers in interviews often talk about "idealized versions of practice" (p. 220) instead of what actually happens in their classrooms.

5.4.2 Research context

In the Netherlands, students receive elementary education from ages 4 to 12. They are educated in, for example, history, writing, reading, geography, and science. Around

age twelve, children transition from elementary education to secondary education. There are three educational tracks in secondary education. Approximately 60% of the students continue on to pre-vocational schools (duration of 4 years), 20% receive a senior general secondary education (duration of 5 years), and 20% receive a pre-university education (duration of 6 years). The determination is based on the advice of the elementary school and is supported by a mandatory standardized test. For our research, we focus on senior general secondary education and pre-university education since the ability to perform historical contextualization is not explicitly mentioned in the pre-vocational history exam program. History is a mandatory subject in the first three years of senior general secondary education and pre-university education. After 3 years, history becomes an elective subject. Generally, in senior general secondary education, approximately 65% of the students take the final history exam, and in pre-university education, approximately 50% of the students take the exam (Netherlands Institute for Curriculum Development, 2016). The educational quality of all elementary and secondary schools is monitored by the Dutch Inspection of Education.

5.4.3 Sample

We asked eight history teachers from our professional network to participate in our study. To explore the possible differences between teachers, we wanted the sample to be as varied as possible with respect to gender, age, and work experience as a history teacher. The teachers participated voluntary in the study, and all had Dutch nationality. The teachers were not informed of the purpose of the research but were only asked for permission to videotape two of their lessons. The gender distribution in the Netherlands of teachers is 48% female and 52% male (Statistics Netherlands, 2014). Each teacher was from a different school (six schools are in the northern part of the Netherlands, and two schools are in the central part of the Netherlands). Table 15 presents an overview of the teachers' characteristics.

5.4.4 Observation instrument

For each teacher, two lessons were videotaped, yielding a total of 16 different lessons. We used videotaped records because this allowed for stop-and-go coding and repeated viewing of key scenes (Yoder & Symons, 2010). All lessons were given in the two highest educational tracks of the Dutch educational system. We chose to use the Framework for Analyzing the Teaching of Historical Contextualization (FAT-HC) to observe the videotaped lessons. The FAT-HC is developed and tested for reliability by Huijgen, Van de Grift, et al. (2017) and focuses on observing how history teachers

promote historical contextualization in classrooms. The FAT-HC is modeled on Van de Grift's (2007) International Comparative Analysis of Learning and Teaching (ICALT) high-inference observation instrument.

Table 15. Teachers' characteristics

Teacher*	Gender	Educational qualification	Age	Years of work experience	Schools' student enrollment
John	Male	Bachelors	60	37	< 250
Mark	Male	Masters	34	8	>1000
Nick	Male	Masters	43	17	>1000
Bob	Male	Masters	63	41	>1000
Dylan	Male	Masters	41	14	>1000
Lisa	Female	Masters	45	22	>1000
Anna	Female	Masters	30	5	>1000
Kim	Female	Masters	27	1	>1000

Note. *Names are pseudonyms.

The FAT-HC comprises 40 items and utilizes a 4-point Likert scale to score the items, where 1 = weak, 2 = more weak than strong, 3 = more strong than weak, and 4 = strong. Observers have to provide a qualitative verdict of an item based on the whole lesson. Similar to the ICALT instrument, scores of 1 and 2 represent a negative verdict, while scores of 3 and 4 represent a positive verdict. To analyze the lesson observations more specifically, we divided the 40 FAT-HC items into eight categories (see Table 16). The categories were based on four teaching strategies: (1) reconstructing an adequate historical context, (2) enhancing historical empathy, (3) using the historical context to explain historical events, and (4) raising awareness of the consequences of a present-oriented perspective when examining the past.

Table 16. Categories and accompanying FAT-HC items

Categories	FAT-HC items
Reconstructing the historical context (teacher demonstrates)	1-13
Reconstructing the historical context (teacher activates)	14-21
Enhancing historical empathy (teacher demonstrates)	22-24
Enhancing historical empathy (teacher activates)	25-27
Using the historical context to explain historical events (teacher demonstrates)	28-31
Using the historical context to explain historical events (teacher activates)	32-35
Not using anachronisms and presenting the past as progress	36-37
Raising awareness of present-oriented perspectives	38-40

Note. See Appendix D for the FAT-HC of Huijgen, Van de Grift, et al. (2017).

To examine our hypothesis that history teachers might not engage students in the process of contextualization, we made a distinction for the "Reconstructing the historical context," "Enhancing historical empathy," and "Using the historical context to explain historical events" strategies between items that demonstrate historical contextualization and items that engages students in historical contextualization processes. This distinction can be seen in the FAT-HC of Huijgen, Van de Grift, et al. (2017) because the items starting with "The teacher…" demonstrate historical contextualization, while items starting with "The students…" implies that students are engaged in historical contextualization. An example of an item that demonstrates historical contextualization is "The teacher discusses the economic circumstances at the time of the phenomena." This item only includes the explaining of the economic circumstances by the teacher, but there is no classroom interaction with the students. An example of an item that engages students in historical contextualization is "The students explain the economic circumstances at the time of the phenomena." This item does include an interaction between the teacher and students because the teacher, for example, asks students to describe or research the economic circumstances.

We created a separate category for the items "The teacher does not use anachronisms" (FAT-HC Item 36) and "The teacher does not present the past as progress" (FAT-HC Item 37) since the mean scores of these items were very high (indicating that teachers almost never used anachronisms and presented the past as progress) and therefore did not display a representative and nuanced image of the category "Raising awareness of present-oriented perspectives." The categories "Not using anachronism and presenting the past as progress" and "Raising awareness of present-oriented perspectives" only focus on demonstrating historical contextualization according to the FAT-HC (all these items start with "The teacher…"). No distinction could therefore be made for these categories between items that demonstrate historical contextualization and items that engage students in historical contextualization.

5.4.5 Observers

We trained five observers (three male and two female history teachers ranging in age from 29 to 33 years and having 7 to 8 years of work experience as history teachers) to each observe the videotaped history lessons. We used multiple observers because research indicates increased reliability when using two or more observers for the same lesson when using the FAT-HC (Huijgen, Van de Grift, et al., 2017). The observers were selected from the professional network of the authors and participated

voluntary in the study. They all held the Dutch nationality and a Master's degree in history education. All observers received 4 hours of training in the use of the FAT-HC. Three videotaped history lessons taught by three different history teachers were used as training material. These lessons were not used in our data analyses. The observers received an introduction and explanation of the FAT-HC items and evaluated the videotaped lessons using a training version of the observation instrument that included more in-depth explanations of the items. After the observers observed each videotaped lesson, their results were discussed, and some items were clarified by the trainers to minimize inter-rater bias. The items "The teacher creates historical tension (the past as different)," "The teacher moves the self into the past (if I…)," and "The teacher outlines a recognizable role for students to foster historical empathy (as a businessman / like a father)" needed the most clarification.

5.4.6 Data analysis

First, to examine the extent to which the history teachers promoted historical contextualization, we calculated the observers' mean FAT-HC score for each lesson. Next, based on two lessons, we calculated a category mean score for each teacher to examine the differences between the different categories. This also provided an opportunity to examine the extent to which the history teachers demonstrated historical contextualization and engaged students in historical contextualization processes. Finally, we analyzed the videotaped lessons to identify examples that illustrate our findings.

5.5 Results

5.5.1 FAT-HC scores

To examine how the eight history teachers promoted historical contextualization in their lessons, we present the observers' mean FAT-HC scores in Table 17. Most teachers obtained similar FAT-HC scores in their different lessons except Kim, Anna, and Nick. If FAT-HC scores > 2.00 denote a positive verdict and scores < 2.00 denote a negative verdict, no teacher in the sample obtained a positive mean FAT-HC score, which was the opposite of what we expected.

5.5.2 Category scores

To examine possible differences between the categories, we present the observers' mean category scores (based on two lessons) in Table 18. The highest scores were achieved in not using anachronisms and presenting the past as progress (Category 7, mean score = 3.59) and reconstructing a historical context (Category 1, mean score = 2.18). The observers almost never noticed the use of anachronisms (i.e., something or someone that is not in its correct historical or chronological context) or that the past was presented as progress. The lowest mean category scores were achieved in enhancing the use of the historical context to explain historical events (Category 6, mean score = 1.24) and enhancing historical empathy among students (Category 4, mean score = 1.25). Interestingly, as displayed in Table 18, Bob and Kim achieved the highest scores in all categories, which demonstrate the engagement of the students in historical contextualization (Categories 2, 4, and 6). Compared to the other teachers, they seemed to engage students more when reconstructing a historical context, promoting historical empathy, and explaining historical events. Lisa is also interesting because she obtained the highest scores for the categories 1, 3, and 5 (demonstrating historical contextualization) but the lowest scores in the same categories when engaging students in historical contextualization (Categories 2, 4, and 6).

5.5.3 Differences between demonstrating and engaging

For the categories "Reconstructing the historical context," "Enhancing historical empathy," and "Using the historical context to explain historical events," we made a distinction between the items focusing on demonstrating historical contextualization by a teacher (Categories 1, 3, and 5) and items focusing on engaging students in historical contextualization processes (Categories 2, 4, and 6). Table 19 present the differences between demonstrating and engaging students in historical contextualization processes. As expected, the teachers paid less attention to engaging students in historical contextualization processes in the lessons.

5.5.4 Promoting historical contextualization

For each category, we use examples from the videotaped lessons to illustrate our findings in more detail. The examples provide more qualitative insights into how historical contextualization was promoted by the teachers, in the missed opportunities of the teachers, and in the differences between high and low scoring teachers.

Table 17. Mean FAT-HC scores of the observed lessons (maximum score = 4.00)

Teacher*	Educational track	Lesson topic	Students (*n*)	Students' age	Mean FAT-HC score** (SD)
Lisa	1. Senior general secondary education	Renaissance	26	16-17	1.98 (0.14)
	2. Senior general secondary education	Western Colonialism	27	15-16	1.97 (0.26)
Bob	1. Pre-university Education	Cold War	30	14-15	1.89 (0.17)
	2. Pre-university Education	Cold War	29	14-15	1.87 (0.26)
Nick	1. Senior general secondary education	Second World War	28	16-17	1.93 (0.16)
	2. Senior general secondary education	Monotheism	26	15-16	1.83 (0.15)
Kim	1. Senior general secondary education	Enlightenment	25	15-16	1.96 (0.42)
	2. Senior general secondary education	Democratic Revolutions	25	15-16	1.80 (0.35)
Anna	1. Senior general secondary education	Dutch Republic	26	15-16	1.87 (0.19)
	2. Pre-university Education	Ancient Greece	29	15-16	1.71 (0.19)
John	1. Senior general secondary education	Investiture Controversy	16	15-16	1.76 (0.14)
	2. Senior general secondary education	Late Middle Ages cities	13	15-16	1.73 (0.07)
Mark	1. Senior general secondary education	Second World War	28	15-16	1.75 (0.07)
	2. Pre-university Education	Alexander the Great	23	15-16	1.74 (0.10)
Dylan	1. Senior general secondary education	Enlightenment	25	15-16	1.66 (0.13)
	2. Senior general secondary education	Slavery	31	15-16	1.65 (0.11)

Note. *Names are pseudonyms. **Mean score of five observers.

Table 18. Observers' mean scores based on two lessons (maximum score = 4.00)

Teacher Age Years' work experience	Lisa (SD) 45 22	Bob (SD) 63 41	Nick (SD) 43 17	Kim (SD) 27 1	Anna (SD) 30 5	John (SD) 60 37	Mark (SD) 34 8	Dylan (SD) 41 14	Mean category score (SD)
1. Reconstructing context	2.72 (0.39)	2.02 (0.25)	2.52 (0.28)	2.17 (0.46)	2.16 (0.34)	1.81 (0.22)	2.25 (0.33)	1.77 (0.25)	2.18 (0.33)
2. Reconstructing context (students)	1.18 (0.30)	1.85 (0.22)	1.45 (0.30)	1.90 (0.38)	1.61 (0.33)	1.68 (0.25)	1.50 (0.20)	1.33 (0.36)	1.56 (0.25)
3. Enhancing empathy	2.63 (0.58)	1.47 (0.39)	1.43 (0.39)	1.47 (0.48)	1.23 (0.27)	1.30 (0.29)	1.40 (0.21)	1.50 (0.59)	1.55 (0.45)
4. Enhancing empathy (students)	1.07 (0.14)	1.27 (0.47)	1.13 (0.17)	2.10 (1.28)	1.07 (0.14)	1.23 (0.35)	1.00 (0.00)	1.13 (0.23)	1.25 (0.35)
5. Contextualize to explain	2.35 (0.52)	1.75 (0.43)	2.08 (0.43)	1.90 (0.49)	1.88 (0.43)	1.80 (0.31)	1.48 (0.32)	1.53 (0.30)	1.85 (0.28)
6. Contextualize to explain (students)	1.08 (0.17)	1.50 (0.49)	1.18 (0.27)	1.55 (0.55)	1.23 (0.28)	1.13 (0.21)	1.10 (0.18)	1.13 (0.27)	1.24 (0.18)
7. No anachronisms / past as progress	3.30 (0.63)	3.95 (0.16)	3.95 (0.16)	2.50 (1.05)	3.50 (1.00)	3.85 (0.24)	4.00 (0.00)	3.65 (0.94)	3.59 (0.50)
8. Raising awareness of presentism	1.50 (0.28)	1.27 (0.26)	1.30 (0.25)	1.43 (0.50)	1.67 (0.69)	1.17 (0.24)	1.27 (0.21)	1.23 (0.35)	1.36 (0.17)
Mean FAT-HC score	1.98 (0.20)	1.88 (0.21)	1.88 (0.16)	1.88 (0.37)	1.79 (0.20)	1.75 (0.10)	1.75 (0.08)	1.66 (0.11)	1.82 (0.10)

Table 19. Differences between teacher demonstrates and teachers engages

Category	Teacher demonstrates (I)	Teacher engages (J)	Difference (I – J)
Reconstructing context	2.18	1.56	0.62
Enhancing empathy	1.55	1.25	0.30
Contextualize to explain	1.85	1.24	0.61
Mean	1.86	1.35	0.51

5.5.4.1 Reconstructing the historical context

Lisa and Nick obtained the highest scores in demonstrating the reconstruction of the historical context. These teachers considered the different frames of reference (i.e., chronological, spatial, socio-political, socio-economic, and socio-cultural) in each lesson when reconstructing the historical context of a historical event. For example, Lisa addressed the different frames of reference when discussing western European colonies in the 20th century. To explain the colonies' struggle for independence, she reconstructed the historical context at the beginning of the lesson:

> It started 400 years ago; you should go back 400 years to understand the colonies' struggle for independence. Around 1600, different European countries wished to buy cheap spices. At first, the European countries would make economic agreements with the locals. An example is the Dutch East India Company, which traded often with Dutch India and other Asian countries. However, the merchants stayed on the coast and did not try to change, for example, the locals' religion or government. So, what you see [points at a world map] is that the Dutch traveled to Asia but they stayed along the coast and not inland. But around 1800, there was a change due to the Industrial Revolution in Europe. Different European countries needed more colonies for their raw minerals and to sell their products. In order to do so, they needed more political and economic influence in the colonies.

Interestingly, although the teachers in the sample used time indicators several times (FAT-HC Item 4), they almost never showed historical events on a timeline (FAT-HC Item 6). Moreover, despite the fact that Lisa used a world map in the example, most teachers did not use geographical (historical) maps to reconstruct a spatial dimension. For example, John could have shown a map of Western Europe in the Middle Ages when discussing medieval trade to illustrate the different sizes and names of countries compared to the present. Bob and Kim encouraged the students the most to reconstruct the historical context. For example, Kim asked the students to reconstruct

the historical context of the democratic revolutions instead of reconstructing the historical context herself:

> Kim: The Dutch Revolution. If you look at the specific time when it happened, why is that name strange?
> Student A: The Netherlands did not yet exist at that time.
> Kim: What was the name of the Netherlands back then?
> Student A: The Dutch Republic.
> Kim: Excellent. And what was the relationship between the Enlightenment and the Democratic Revolutions?
> Student B: They started thinking about the best type of government, and they wished to be independent in the case of the American Revolution.
> Kim: And what is the relationship with the Enlightenment? How did the people of the Enlightenment view society?
> Student C: They wanted equality between people.

Lisa, who obtained the highest score in reconstructing the historical context by herself (demonstrating), obtained a far lower score in encouraging students to reconstruct the historical context. This may have been caused by the fact that she did ask questions in her lessons but often answered these questions herself. For example, she asked in one lesson: "Why did the Netherlands and other European countries want so many colonies? What were the reasons?" She, however, answered these questions herself instead of asking the students to provide an answer. The other teachers in the sample also answered their own questions. Moreover, Lisa could have asked the students to create a timeline with historical events relating to Western colonialism from 1600 to 1800 to create a chronological context instead of providing the chronological context herself.

5.5.4.2 Enhancing historical empathy

Lisa and Dylan used historical empathy the most in their lessons, particularly by presenting historical agents relevant to the historical topic under study. For example, when talking about the consequences of the French Revolution, Dylan explained and described the life and role of Napoleon. When explaining 18th-century slavery, he described the life of a 14-year-old slave who worked on a plantation to illustrate the contextual circumstances. Compared to the other teachers, Lisa moved into the past often, for example, to explain why the Netherlands needed colonies:

> If I had a textile factory and I made a lot of coats, then I needed, first of all, a lot of cotton. So where did I get my cotton? Secondly, if I produce 5,000 coats a day and almost everybody in my own country already wore my coats, where could I sell my coats?

Bob also provided an interesting option for using historical empathy to explain historical phenomena. Instead of explaining the differences between communism and capitalism himself when talking about the beginning of the Cold War, Bob asked his students to imagine that they were blindfolded and dropped into an unknown country. Next, he asked his students to remove the imaginary blindfold and asked them to describe how they would know if they were in a communist or capitalist country:

> Bob: What do you have to notice? Where do you look?
> Student A: The buildings. In a communist country, the buildings look very similar.
> Student B: Maybe the differences between people?
> Student C: Communism does not focus on making profit; capitalism does.
> Bob: And how could you see this?
> Student C: The cars, the communist countries might drive the same car, often Ladas.
> Bob: And why is that?
> Student C: The government owned the factories and why does the government need to produce different cars?

Kim encouraged her students to practice historical empathy the most. She was the only teacher in the sample who explicitly used a historical empathy task. When explaining 18th-century child labor, she divided her class into dyads, and each dyad was instructed to empathize with a different historical agent living in the 18th century, for example, an 8-year-old child, a factory owner, and a politician. The central task was to reason whether the historical agent was against or in favor of child labor. Mark did not engage his students in historical empathy at all. When he taught his students about the Second World War and the rise of Hitler, he could have, for example, described a young German man who was unsure as to which political party he would vote for in 1930 and asked his students to empathize and reason if the man had voted for the Nazi Party. In his other lesson, he centralized a historical agent (Alexander the Great),

but he never asked his students to reason as to how Alexander the Great's motivation and beliefs affected his actions.

5.5.4.3 Contextualize to explain

Compared to the other teachers, Lisa and Nick made more use of the historical context to explain historical events. They not only reconstructed the historical context but also used historical context knowledge to compare phenomena or presented different perspectives on a historical event. For example, Nick used historical context knowledge of the Ancient Period, such as time indicators (e.g., 63 B.C., the first century), a geographical context (e.g., map of the Middle East and the Roman Empire), and the political and socio-cultural circumstances (e.g., the differences between monotheistic religions such as Christianity and Judaism and the polytheistic Roman religion) to explain the Roman persecution of Jews and Christians.

Bob and Kim encouraged their students to use their historical context knowledge the most. Interestingly, this happened the most when presenting and discussing historical sources. For example, Bob presented a 1950 Russian cartoon displaying American insects that were raiding Soviet Union territory. He asked his students to use their historical context knowledge (e.g., the Cold War climate, the Marshall Plan) to examine and interpret the cartoon.

> Bob: All right, who knows when this cartoon was made?
> Student A: The source states 1950.
> Bob: Yes, 1950. Which important historical events took place around 1950?
> Student B: The Korean War.
> Bob: Correct but think again. Which historical event could be related to the source?
> Student C: The Marshall Plan?
> Bob: Yes, but how is that related to the source? How did the Americans experience the Marshall Plan?
> Student C: As something good. They wanted to help other people who needed help after the Second World War.
> Bob: All right. And how could the Russians have viewed the Marshall Plan?
> Student C: As something negative.
> Bob: But it was something good, was it not? You cannot be angry at something that is good, can you?

> Student C: Yeah, but the Soviet-Union viewed it as unwanted interference. The Russians thought that the United States tried to influence European countries.
> Bob: Perfect. And who knows what this cartoon means?
> Student D: I see insects that are eating all the Soviet Union's food. I think the insects represent the Americans plundering the Soviet Union. I think the creator of the cartoon might be Russian.
> Bob: Why?
> Student D: The Americans are negatively displayed as imperialists who try to enlarge their influence in Europe and the Soviet-Union.

Lisa and Nick obtained high scores for demonstrating historical contextualization in this category; however, they did not engage students much in using knowledge of the historical context to explain historical phenomena. For example, instead of explaining how the Roman persecution of Jews and Christians originated, Nick could have provided the students with historical sources addressing the different frames of reference to reconstruct a historical context and formulate an answer to how the Roman persecution of Jews and Christians originated.

5.5.4.4 Anachronisms and the past as progress

The observers never observed the use of anachronisms (i.e., something or someone that is not in its correct historical or chronological context) by the teachers. Moreover, the teachers generally did not present the past as progress (i.e., the present is better than the past). Compared to the other teachers, Kim did make remarks a few times (e.g., "Nowadays we have it a lot better") indicating that the present is better than the past. However, this category obtained by far the highest observation scores, indicating that the teachers in the sample could not improve much in this category.

5.5.4.5 Raising awareness of presentism

Anna and Lisa paid the most attention to preventing presentism among the students. An important item in this category is that teachers present learning strategies for historical contextualization to prevent presentism (FAT-HC Item 40). Anna was the only teacher who explicitly taught her students a learning strategy by guiding them to consider different frames of reference when examining a historical event. For example, she taught her students to examine the political, economic, and socio-cultural circumstances of prehistoric hunter gatherers step-by-step. By teaching her students to approach a historical phenomenon this way, the chance that they view historical events

and historical agents' actions from a present-oriented perspective decreases because the students interpret and examine phenomena in their own time and circumstances.

Moreover, an important FAT-HC item of this category is "The teacher uses historical tension." A teacher could present a problem or case that students find difficult to explain due to their possible present-oriented perspectives. It was striking that none of the teachers in the sample explicitly used historical tension to trigger possible present-oriented perspectives among the students. There were often missed opportunities to do so. For example, when Mark discussed the rise of Hitler in Germany in the 1930s, he could have asked his students to explain why so many Germans voted for his political party. This gave him the opportunity to evaluate their answers: could they explain the rise of Hitler (using historical context knowledge) or were they not able to explain this because they viewed the past from a present oriented-perspective (e.g., Hitler killed millions of people)?

5.6 Conclusions and discussion

The aim of this study was to explore how history teachers promoted historical contextualization in their classrooms. Using the Framework for Analyzing the Teaching of Historical Contextualization (FAT-HC), two lessons from eight history teachers were observed by trained raters, yielding 16 different lessons in total. Our first hypothesis was that teachers demonstrate historical contextualization in their lessons because an important aim of the Dutch history curriculum is for students to be able to use their acquired historical overview knowledge to perform historical contextualization (Board of Tests and Examinations, 2017). In contrast to our expectations, the overall results indicated that most teachers did not often demonstrate historical contextualization in their classrooms. None of the teachers in the sample obtained a mean FAT-HC score > 2.00. The highest scores could be found in the categories focusing on not using anachronisms and presenting the past as progress (mean score = 3.59) and the category focusing on reconstructing the historical context (mean score = 2.18). All other categories obtained mean scores < 2.00, with the category focusing on promoting the use of historical empathy among students (mean score = 1.25) and the category on promoting the use of the historical context (mean score = 1.24) displaying the lowest scores.

The second hypothesis was that the teachers did not often engage students in historical contextualization processes. As expected, we found a mean score of 1.35 in the categories focusing on engaging students in historical contextualization compared to a mean score of 1.86 in the categories focusing on demonstrating historical contextualization. This finding is in line with research, which illustrates that history teachers focus on covering content knowledge and less on creating opportunities to promote historical thinking and reasoning (Barton & Levstik, 2003; Saye & Social Studies Inquiry Research Collaborative, 2013).

Research suggests three possible reasons for the differences between the desired instruction methods (i.e., engaging students in historical reasoning competencies) and daily classroom practice (i.e., focusing on the transfer of historical content knowledge). Scholars such as Grant and Gradwell (2010) and Meuwissen (2017) argue that the first reason may be contextual factors, such as state tests and history textbooks. A second reason may be an ineffective classroom climate (Martell, 2013; Virta, 2002), and a final reason may be the *problem of enactment* (Kennedy, 2016) since research indicates that (student) teachers want to teach historical reasoning competencies but do not know how to transfer their beliefs into classroom action (Wansink, Akkerman, & Wubbels, 2016). Since contextualization plays an important role in the Dutch formal exam program and since most teachers in the sample had an effective pedagogical classroom climate, the problem of enactment appears the most relevant.

If our findings also appear in studies with more participants, future research should focus on helping teachers overcome the problem of enactment, for example, by developing and testing effective and activating instructional tools to teach historical contextualization. To examine the problem of enactment in more detail, future research should also include the relationship between teachers' beliefs and their historical contextualization practices. Using belief interviews (e.g., Richardson, Anders, Tidwell, & Lloyd, 1991; Tuithof, 2017) or surveys (e.g., Stipek, Givvin, Salmon, & MacGyvers, 2001) in combination with FAT-HC observations can provide useful insights for developing teacher professionalization programs for historical contextualization.

An important limitation of our study is that we conducted exploratory research among only eight history teachers and observed only two lessons from each teacher. Future research should therefore examine whether the findings of this study also

appear among larger samples of teachers and lessons. Moreover, we only used classroom observations. Using other methods, such as student questionnaires and teachers' self-reports (e.g., Muijs, 2006), could also contribute to increasing insights as to how teachers promote historical contextualization in classrooms. The unit of analysis was also the whole lesson with a focus on teacher behavior and lectures. Comparison of teacher lectures, teacher-student interactions, and student discussion lesson fragments could provide more insight into how historical contextualization is promoted during different lesson activities. Furthermore, beyond the scope of our research was the relationship between lesson topics and forms of historical contextualization. Further research is needed to answer the following question: Do teachers use different historical contextualization teaching strategies depending on the historical topic? Stimulated recall (e.g., Lyle, 2003) where teachers think aloud when watching their own lessons could provide insights into the relationship between historical topics and historical contextualization.

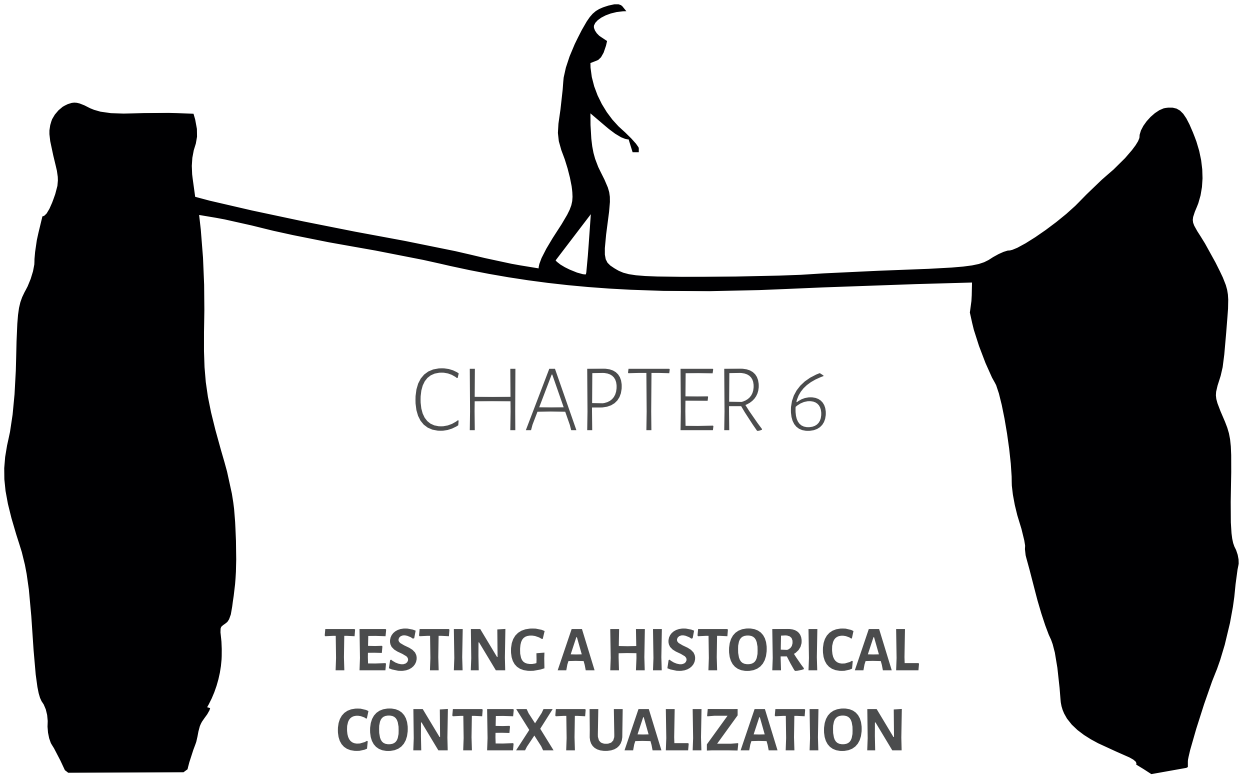
Despite these limitations, our study showed the possibilities of using the FAT-HC to operationalize history teachers' specific professionalization needs since it provides domain-specific insights into teachers' strengths and weaknesses. For example, an optimistic finding is that the observers almost never noticed the use of anachronisms or presenting the past as progress by teachers. By contrast, teachers can, for example, engage students more in historical contextualization by creating opportunities where students use their historical context knowledge to explain, compare, or evaluate historical phenomena. Mariott (2001) and Ball and Forzani (2009) noted that these insights are important to educating and professionalizing (history) educators.

We conclude with some practical implications. Our findings illustrate that teachers often answered questions themselves. Instead, teachers could create opportunities for students to answer questions. Furthermore, despite the fact that teachers provided time indicators (e.g., year, century, period) when explaining historical phenomena, they almost never displayed a timeline to establish a chronological context or encouraged the students to create timelines. Additionally, geographical maps were rarely used to establish a spatial context. To enhance historical empathy, teachers should not only present a historical agent but also consider, for example, the agents' motives, beliefs, and knowledge (Endacott & Pelekanos, 2015). This was often not the case in the observed lessons.

To engage students more in historical contextualization, teachers have to remember not to "showcase" their own knowledge and skills. It is important to let the students do the work and make mistakes and to help them in the processes of historical contextualization. For example, it is suggested to not only display a timeline but also instruct students to create (different) timelines themselves. It is also important to provide historical sources that address the different frames of reference and ask students to reconstruct a context on their own to answer evaluative and explanatory questions. The History Assessments of Thinking on historical contextualization, which are developed by the Stanford History Education Group, are promising tools to engage students more in historical contextualization and can be used for formative assessment and feedback on this historical reasoning competency (Breakstone, Smith, & Wineburg, 2013). Discussing historical sources in classroom discussions might also be an effective strategy since we found that this often engaged students in historical contextualization processes. Moreover, teachers could focus more on triggering possible present-oriented perspectives among students. Presenting the past as strange (e.g., child labor and the poor working conditions in the 18th century compared to the daily life of a child currently) could promote awareness of the differences and connections between the past and present (Huijgen & Holthuis, 2015; Seixas & Morton, 2013).

Furthermore, the teachers in our sample did not explicitly teach students how to perform historical contextualization. To improve in this area, teachers could use the scaffolds developed by Reisman and Wineburg (2008) and Havekes et al. (2012). To help history teachers promote historical contextualization, teachers could participate in professional development programs, including pre- and post-observation interviews and opportunities to collaboratively develop lesson activities guided by experts. *Lesson study*, including the use of the FAT-HC, which focuses on collaborative planning, teaching, observing, and discussion of lessons (cf. Lewis, Perry, & Murata, 2006), could help teachers design effective learning tasks. Recently, Korthagen (2017) described an interesting approach called "Professionalization development 3.0," which might help to overcome the problem of enactment. This is a bottom-up approach that centralizes the teachers' potential where the teacher sets relevant (personal) learning goals instead of dealing solely with expert knowledge (top-down approach). As the results of this study show, the teaching of historical contextualization is a complex process, but if teachers, teacher educators, and researchers work together to design

effective instructional tools and specific professionalization programs on historical contextualization, this might result in an increase in students' ability to perform historical contextualization.



CHAPTER 6

TESTING A HISTORICAL CONTEXTUALIZATION PEDAGOGY

This chapter describes the development and testing of a pedagogy aimed at promoting students' ability to perform historical contextualization. Teaching historical contextualization was conceptualized in terms of four pedagogical design principles: (1) making students aware of the consequences of a present-oriented perspective when examining the past, (2) enhancing the reconstruction of a historical context, (3) enhancing the use of the historical context to explain historical phenomena, and (4) enhancing historical empathy. The effectiveness of these principles was explored in a lesson unit focusing on the 17th and 18th centuries. In a quasi-experimental pre-test–post-test design with experimental and control conditions, the effects of the pedagogy on 15- and 16-year-old students' ability to perform historical contextualization were examined (N = 131). The results indicated that students in the experimental condition significantly improved their ability to perform historical contextualization compared to students in the control condition. These findings could be used to help teachers and other educational professionals design and implement historical contextualization tasks and instructions.

This chapter is based on: Huijgen, T. D., Van de Grift, W. J. C. M., Van Boxtel, C. A. M., & Holthuis, P. (2018). Promoting historical contextualization: The development and testing of a pedagogy. *Journal of Curriculum Studies*, *50*, 410-434. doi:10.1080/00220272.2018.1435724

6.1 Introduction

Scholars such as Seixas (2015), VanSledright (2011), and Wineburg (2001) emphasize that history education should not only focus on learning historical facts but also include promoting students' historical thinking and reasoning. Historical reasoning competencies have therefore become increasingly important in western history education (Erdmann & Hasberg, 2011). A key component of historical reasoning is the ability to perform historical contextualization (Lévesque, 2008; Seixas & Morton, 2013; Van Drie & Van Boxtel, 2008), which is the ability to situate phenomena and actions by people in the context of time, historical location, long-term developments, or particular events to give meaning to these phenomena and actions (Van Boxtel & Van Drie, 2012). Without this ability, for example, historical agents' actions cannot be explained and historical events cannot adequately be interpreted (Barton & Levstik, 2004; VanSledright, 2002).

Despite the importance of historical contextualization, research indicates that many students struggle when asked to perform historical contextualization tasks because they view the past from a present-oriented perspective (Foster, Ashby, & Lee, 2008; Hartmann & Hasselhorn, 2008; Huijgen, Van Boxtel, Van de Grift, & Holthuis, 2014; Shemilt, 2009). As Reisman and Wineburg (2008) noted: "Contextualized historical thinking runs counter to the narratives and frameworks that many students bring to class" (p. 203). Teachers should therefore explicitly teach students historical contextualization to help them overcome possible present-oriented perspectives.

Research on historical contextualization has focused on, for example, how students performed historical contextualization (e.g., Berti, Baldin, & Toneatti, 2009; Wooden, 2008) and how it can be observed (Huijgen, Van de Grift, Van Boxtel, & Holthuis, 2017), or promoted (e.g., Baron, 2016; Boerman-Cornell, 2015). However, experimental studies testing pedagogies on historical contextualization are scarce. This is unfortunate since teachers seem to struggle with developing instructional tools to engage students in historical reasoning processes (e.g., Achinstein & Fogo, 2015; Reisman, 2015; Saye & Social Studies Inquiry Research Collaborative, 2013). More examples of effective and practical instructional tools are therefore desired within the field of history education (e.g., Fogo, 2014; Grant & Gradwell, 2010; Reisman & Fogo, 2016).

The aim of the present study is therefore twofold: (1) to develop a pedagogy for promoting students' ability to perform historical contextualization and (2) to test this pedagogy for success in a pre- and post-test quasi-experimental design.

6.2 Theoretical framework

6.2.1 The concept of historical contextualization

Some studies define historical contextualization as a heuristic (in addition to sourcing and corroboration) to examine historical sources (e.g., Britt & Aglinskas, 2002; Rouet, Favart, Britt, & Perfetti, 1997; Wineburg, 1991). However, in history education, it is possible to contextualize historical agents' actions, historical events, and historical sources (Havekes, Coppen, Luttenberg, & Van Boxtel, 2012). Therefore, in this study, we conceptualize historical contextualization as the ability to situate phenomena and the actions of people in the context of time, historical location, long-term developments, or particular events to give meaning to these phenomena and actions (Van Boxtel & Van Drie, 2012).

A key component for performing historical contextualization successfully is students' understanding of the differences between the past and present (Seixas & Peck, 2004). Historical contextualization concerns:

> A temporal sense of difference that includes deep understanding of the social, political, and cultural norms of the time period under investigation as well as knowledge of the events leading up to the historical situation and other relevant events that are occurring concurrently. (Endacott & Brooks, 2013, p. 43)

Historical contextualization is therefore a complex skill because it not only requires historical factual knowledge and a sense of chronology but also the ability to identify gaps in this knowledge, the ability to formulate questions, and the ability to question information or conclusions (Wineburg, 1998). For example, to explain why Julius Caesar could not have had breakfast in Rome and dinner in the Gallic region of France on the same day, students have to contextualize the ancient Roman period, including the knowledge that the transportation necessary for such a day trip was not available in those times (Lévesque, 2008).

6.2.2 Teaching historical contextualization

Building on Wineburg's (1991) work, most intervention studies that provide insight into the teaching of historical contextualization consider contextualization to be one heuristic to be used (besides sourcing and corroboration) to examine historical documents. For example, Nokes, Dole, and Hacker (2007) tested the effect of heuristic instruction among 16- and 17-year-old students that explicitly taught sourcing, corroboration, and contextualization. Contextualization was taught by discussing the use and importance of contextualization, modeling contextualization, and asking students to create a historical context of a document to interpret the documents. In the pre- and post-test, the authors found that only 7% of the students used contextualization and therefore conducted no further analyses. Reisman (2012b) examined the effect of a curriculum intervention (focusing on sourcing, corroboration, close reading, and contextualization) in disciplinary reading among 11th-grade students. Contextualization was taught by cognitive modeling, guided practice, or independent practice. A historical reading strategy chart with guiding questions (e.g., What else was happening at the time this was written?) helped students perform contextualization. However, no significant intervention effect for contextualization was found, and Reisman (2012b) concluded that the question of how to teach contextualization remains unanswered. De La Paz et al. (2014) tested a curriculum intervention, including explicitly promoting contextualization, among eight-grade students to test their disciplinary writing skills. To promote contextualization, the students were provided a handout with questions focusing on the type of document (e.g., What type of document is this and where did it appear?) and the time period and setting of the document (e.g., What else was happening at the time?). The students' disciplinary writing skills improved, but no specific information is given on their improvement in contextualization.

In other studies, historical contextualization was the main dependent variable, and the focus was less on contextualization as a component of the critical examination of historical sources but more on the contextualization of particular events, situations, or the actions of people in the past. For example, Van Boxtel and Van Drie (2012) asked students aged 14–17 to interpret and date situations or events that are described in a historical document or shown in a historical image ("What is it about?"). They found that instruction focusing on the development of a rich associative network of historical knowledge and knowledge of landmarks helps students to interpret the historical situation described or depicted because they are better able to reconstruct a historical

context. Building upon the research literature on historical contextualization, Huijgen, Van de Grift, et al. (2017) suggested four teacher strategies that might improve students' ability to perform historical contextualization: (1) making students aware of the consequences of a present-oriented perspective when examining the past; (2) enhancing the reconstruction of a historical context; (3) enhancing the use of a historical context to explain historical phenomena, and (4) enhancing historical empathy.

These strategies can help students perform historical contextualization, not only when they have to contextualize historical sources but also when historical events and historical agents' actions are discussed in classrooms. In this study, these four teaching strategies were therefore used to develop and test a pedagogy for teaching historical contextualization. The following section describes a translation from the teachers' strategies into pedagogical design principles.

6.2.3 Pedagogical design principles of historical contextualization

6.2.3.1 Making students aware of present-oriented perspectives

Presentism, or viewing the past from a present-oriented perspective, is a bias in which people assume that the same values, intentions, attitudes, and beliefs existed in the past as they exist today (Barton & Levstik, 2004). We can never be perfectly non-presentist (e.g., Pendry & Husbands, 2000; VanSledright, 2001), but teachers should make students aware of their own values and beliefs and the consequences of this perspective when explaining the past (Seixas & Peck, 2004). Students will otherwise not succeed in explaining historical phenomena and historical agents' actions (e.g., Barton, 2008; Lee, 2005; Wineburg, 2001).

To make students in history classrooms aware of their presentism, Havekes et al. (2012) argued that creating cognitive incongruity that is aimed at testing students' assumptions or creating a conflict with their prior knowledge can promote historical contextualization. In previous research, we therefore explored the use of cognitive conflicts to trigger and prevent presentism among students (Huijgen & Holthuis, 2015). In this approach, possible present-oriented perspectives among students become "visible" by presenting a historical event that students find difficult to explain. When students display present-oriented perspectives when answering accompanying explanatory questions, the teacher would explain the consequences

(i.e., not being able to explain and understand the historical event under study) of viewing the past from this perspective. For example, students could be shown a 1932 election poster of Hitler's political party and be asked to explain whether a German person could have voted for this political party. This approach appears promising but has never been tested in an experimental study. In our pedagogy, we therefore aim to make students aware of the consequences of a present-oriented perspective when examining the past by creating cognitive incongruity.

6.2.3.2 Enhancing the reconstruction of a historical context

Different studies stress the importance of historical content knowledge (including chronological and spatial knowledge) to perform historical contextualization successfully (e.g., Van Boxtel & Van Drie, 2012; Wineburg, 2001). To reconstruct the historical context, students and teachers can use different frames of reference (De Keyser & Vandepitte, 1998): a chronological frame of reference, a spatial frame of reference, and a social frame of reference comprising socio-economic, socio-political, and socio-cultural knowledge. To examine the frames of reference and reconstruct a historical context, students can use different primary and secondary sources, such as movies (e.g., Metzger, 2012), visual images (e.g., Baron, 2016; Boerman-Cornell, 2015; Wilschut, 2012), and written documents (e.g., Fasulo, Girardet, & Pontecorvo, 1998).

In previous research, we found indicators that students who combine different frames of reference are more successful in reconstructing the historical context to explain historical agents' actions (Huijgen, Van Boxtel, Van de Grift, & Holthuis, 2017). To reconstruct a context successfully, it is important to provide good examples and scaffolds of contextualized thinking (Havekes et al., 2012; Huijgen & Holthuis, 2015; Reisman & Wineburg, 2008). For example, teachers could provide students with scaffolds that focus on examining the different frames of reference before students formulate arguments and present conclusions. In our pedagogy, we therefore use the different frames of reference to teach students how to reconstruct a historical context of the historical topic under study to answer and discuss historical questions.

6.2.3.3 Enhancing the use of a historical context to explain the past

Teachers should also create opportunities for students to reason using their historical context knowledge (Counsell, Burn, & Chapman, 2016; Halvorsen, Harris, Aponte Martinez, & Frasier, 2015). Historical context knowledge could, for example, be used to interpret a historical source (Reisman & Wineburg, 2008), formulate historical

questions (Logtenberg, Van Boxtel, & Van Hout-Wolters, 2011), or date and sequence historical events, documents, and images (Van Boxtel & Van Drie, 2012). Research indicates, however, that a strong focus in history classrooms on the transmission of historical content knowledge is preferred to creating opportunities for students to reason with their knowledge (e.g., Saye & Social Studies Inquiry Research Collaborative, 2013; VanSledright, 2011). Different studies distil the general image of a teacher who often uses the history textbook narrative and focuses on the transmission of historical content knowledge, such as memorizing (nationally) significant figures, events, and narratives (e.g., Achinstein & Fogo, 2015; Barton & Levstik, 2003). In our pedagogy, we therefore explicitly created opportunities for students to reason with their historical context knowledge to answer and discuss explanatory historical questions

6.2.3.4 Enhancing historical empathy

Historical empathy is "the ability to see and entertain, as conditionally appropriate, connections between intentions, circumstances and actions and to see how any particular perspective would actually have affected actions in particular circumstances" (Lee & Ashby, 2001, p. 25). Historical empathy is the ability to see and judge the past on its own terms by attempting to understand the historical agents' frames of reference and actions (Yilmaz, 2007). Despite some scholars claiming that historical empathy is idealistic and can never be fully achieved because many historical agents are absent (Metzger, 2012), most scholars agree that historical empathy and historical contextualization are closely related (e.g., Cunningham, 2009; Endacott & Brooks, 2013).

Historical empathy may serve as a "fall back rationale," i.e., when students are to contextualize historical events or actions but lack relevant historical knowledge (Berti et al., 2009). For example, students who did not possess adequate historical context knowledge regarding Germany in 1930 could successfully explain the actions of a historical agent based on affective connections and recognizable emotions, such as the fear of being unemployed (Huijgen, Van Boxtel, et al., 2017). In history classrooms, teachers could choose a historical agent relevant to the historical topic under study and instruct their students to examine the historical agents' lives to successfully perform historical contextualization. What was the social position of the historical agent in the society? Was the historical agent wealthy or poor? Did the historical agent belong to the elite? Answering these types of questions could result in a successful explanation of historical agents' decisions and an understanding of historical events. For example,

examining the life of a young man (Hannes) who lived in Germany in 1930 and must decide which political party he would vote for might result in a better understanding of the rise of Hitler (Hartmann & Hasselhorn, 2008; Huijgen, Van Boxtel, et al., 2017). Endacott and Pelekanos (2015) discussed introducing relevant historical agents and their situation to explain and understand social control in ancient Athens.

These studies suggested that when students use affective connections and focus on the role of a historical agent, they may be able to perform historical contextualization successfully. In our pedagogy, we therefore selected a relevant historical agent for each historical topic. Students were provided with a short description accompanied by two central questions that the students need to answer. To answer the questions successfully, the students needed to use affective connections and consider the role and (social) position of the historical agent

6.3 Research question

Since practical and effective instructional tools for teaching historical contextualization are lacking, this study focuses on identifying whether a developed pedagogy, based on the pedagogical design principles of historical contextualization, can improve students' ability to perform historical contextualization. For the present study, we formulated the following research question: What are the effects of a lesson unit based on the four design principles for teaching historical contextualization on 15- and 16-year-old students' ability to perform historical contextualization?

6.4 Method

6.4.1 Research design

We chose an empirical quasi-experimental pre-test–post-test design (Shadish, Cook, & Campbell, 2002) to test the pedagogy. Compared to the experimental designs, quasi-experimental designs lack the random assignment of participants to experimental or control groups. Random assignment was difficult because our research was conducted in an educational setting and we were dependent on the teachers' voluntary participation to implement an intervention. Within the quasi-experimental design, we established an experimental condition where the teachers

used the pedagogy and a control condition where the teachers used a more traditional lesson structure. The participating teachers in the experimental condition were asked to keep a diary (e.g., Bailey, 1990) during the intervention to describe examples of how students might improve in historical contextualization using the pedagogical framework. Post-intervention interviews with the teachers in the experimental condition were used to discuss the examples in the teachers' diaries. This additional qualitative method provided more insights on how the pedagogy was implemented and how students might have improved in historical contextualization.

6.4.2 Participants

Since we wanted as few differences as possible between the teachers, we used non-probability sampling to select teachers of a similar age, work experience as a history teacher, nationality, and educational degree from our professional network to participate in the intervention. All selected teachers had participated in a 1-day professionalization program at the institution of the first author but were not specifically trained in historical contextualization. All teachers participated voluntarily, held Dutch nationality, and had a masters-level educational degree. Their schools did not differ significantly from the total population regarding graduation and enrollment numbers (Statistics Netherlands, 2016). The participating teachers attended two training meetings (two hours per meeting) to understand the lesson structure and activities and how to administer the pre- and post-tests. Table 20 presents the teachers' characteristics. The average student class size was 20.2 students in the experimental condition and 14.0 students in the control condition. History is an elective in Dutch upper secondary education, and the classes can therefore differ in size.

Table 20. Teachers' characteristics

Teacher*	Class	Gender	Age	Years of work experience
Experimental groups				
Ben	1	Male	43	16
David	2	Male	41	14
Wendy	3	Female	50	15
Kim	4	Female	40	13
Lisa	5	Female	32	7
Control groups				
Ben	6	Male	43	16
Emily	7	Female	48	4

Note. *Names are pseudonyms.

A total of 101 secondary school students (44 male, 57 female) participated in the experimental condition. The mean students' age in this condition was 15.9 years and ranged from 15 to 18 years. The control condition yielded a total of 30 students (14 male, 16 female). The mean students' age in the control condition was 15.9 years, ranging from 15 to 19 years. All participating students were senior general secondary educational students (the second-highest secondary educational track in the Netherlands) and did not have extensive prior knowledge of the historical topic of the lesson unit. The historical topic for the experimental and control condition was the 17th and 18th century because this topic fits with the teachers' curriculum during the period in which we wanted to implement the intervention.

6.4.3 Historical contextualization instrument

To answer our research question, we developed and used a historical contextualization test. In two meetings with four experienced history teachers (all four teachers had more than 15 years of working experience each as history teachers), we constructed 30 items to test the students' ability to perform historical contextualization. All items consisted of a historical written source or image and an accompanying choice of two answers: one answer presented a present-oriented perspective, and the other offered a contextualized perspective on the historical source. For example, the students were provided with a source describing the arranged and forced marriage of an 11-year-old girl in the late Middle Ages. The students had to choose the statement that fit the source best: a present-oriented answer (i.e., an 11-year-old should not be forced to marry) or a contextualized answer (i.e., these marriages were based on profit for the families). The items in the test comprised historical topics from the ancient to the modern period. These 30 items were piloted among 158 secondary students from three different schools, with a mean age of 15.1 years old. The pilot results displayed a Cronbach's alpha (α) of .69.

Based on this test, the authors of this study constructed another eight items, yielding a total of 38 items. Next, we randomly assigned 19 items to the pre-test and 19 items to the post-test to reduce the carryover effect, i.e., the effect where students remember their answers from the pre-test and benefit from this retained information in the post-test (Bose & Dey, 2009). When analyzing the instruments' reliability, we found five items in the pre-test and five items in the post-test that threatened the internal consistency of the instruments ($\alpha < .60$). We therefore chose to delete these items. This resulted in a pre-test of 14 items ($\alpha = .70$) and a post-test of 14 items ($\alpha = .68$). There was a significant correlation between the pre-test and post-test ($r = .49$, $p < .01$).

Moreover, we asked two expert history teacher educators and two educational measurement experts to review the deleted items and the final version of the pre-test and the post-test to ensure face and content validity. The experts found no threats in deleting the 10 items and noted that the final pre- and post-tests measure the students' ability to perform historical contextualization and that the tests do not differ significantly in time needed to be completed by the students. The instruments' items were also piloted in four different history classes to test them for practical use. The four teachers who conducted the tests did not have any specific comments about the content or length of the items. Appendix E presents examples of the pre- and post-test items.

6.4.4 The historical contextualization pedagogy

To develop the pedagogy, we followed the guidelines of McKenney and Reeves (2012) for educational design research. We first explored, using focus group methodology, how history teachers might promote historical contextualization in classrooms without specific training or support. To develop an effective pedagogy, we were interested in what teachers might or might not do. Next, based on the exploration and pedagogical design principles of historical contextualization, we constructed the lesson activities from the historical contextualization pedagogy. Using focus group methodology, the pedagogy was reviewed and adjusted for practical use before being tested in a quasi-experimental design.

6.4.4.1 Exploring the teaching of historical contextualization

We used focus group methodology (Vaughn, Schumm, & Sinagub, 1996) to explore how history teachers might promote historical contextualization without specific training or support. The focus group consisted of 16 history teachers (ranging in work experience as history teachers from 1 to 42 years), and all teachers participated voluntarily. To structure the discussion, we first explained the four teachers' strategies of Huijgen, Van de Grift, et al. (2017) and asked which strategies the teachers employ in their lessons and how the strategies are implemented. Most attention was paid to the reconstruction of the historical context, and the least attention was paid to increase awareness among students of their possible present-oriented perspectives. Next, we provided the Framework for Analyzing the Teaching of Historical Contextualization (FAT-HC) of Huijgen, Van de Grift, et al. (2017) and a short explanation of the items and asked which indicators they frequently used in their lessons. The least attention was paid to items that focus on engaging students in historical contextualization

processes (e.g., the students place phenomena in long-term development). This is in line with previous research where we observed how history teachers promote historical contextualization in classrooms (Huijgen, Holthuis, Van Boxtel, & Van de Grift, 2018).

We ended the discussion by asking about the challenges teachers experienced when teaching historical contextualization. Most teachers acknowledged the importance of the indicators of the FAT-HC but noted that they did not have the time, expertise, or support to develop such lesson activities. Based on this exploration, we aimed to help teachers explicitly engage students in historical contextualization processes.

6.4.4.2 Lesson activities of the pedagogy

To construct the lesson activities, we used the four pedagogical design principles of historical contextualization as a starting point: (1) making students aware of the consequences of a present-oriented perspective when examining the past; (2) enhancing the reconstruction of a historical context; (3) enhancing the use of the historical context to explain a historical phenomenon, and (4) enhancing historical empathy.

The first lesson activity promotes awareness of students' possible present-oriented perspectives. For each lesson, we constructed a case centralizing a particular historical topic that students find difficult to explain without historical context knowledge (i.e., creating cognitive incongruity). Each case study was accompanied by an explanatory question that students had to answer and discuss in the classroom. During this classroom discussion, teachers explicitly explained the consequences of viewing the past from a present-oriented perspective. For example, we created a case centralizing the exchange of the colony of New Netherland, currently New York City, for Suriname in 1626. Most students generally find it difficult to explain why "the Dutch Republic exchanged a world-class city for a small country in South America." The central question of this case study was "Can you explain why the Dutch Republic exchanged New Netherland for Suriname in 1626?" In the following classroom discussion, the students were allowed to react and attempt to answer the question while the teacher corrected possible present-oriented perspectives and explicitly explained, by stressing the differences between past and present knowledge, beliefs, and values, that the case cannot be explained when using present-oriented perspectives.

The second lesson activity reconstructed the historical context. In each lesson, the students (in groups of four) had to reconstruct the historical context of the case using a chronological dimension (using a timeline), a spatial dimension (using geographical maps), and a socio-political, socio-economic, and socio-cultural dimension. To reconstruct the historical context, students were provided primary and secondary sources that addressed all frames of reference. Guiding questions were provided to help students examine the socio-political, socio-economic, and socio-cultural frames of reference (see Appendix F). The teachers in the experimental condition were provided the reconstructed historical context (i.e., the historical context knowledge of the different frames of reference), and each group had to present the reconstructed context to the teacher to check for correctness. For example, in the case of the exchange of New Netherland, the students received information to create a timeline of events. A geographical map of the Americas was displayed, and students were presented with historical sources that provide information on the Dutch political climate in the Dutch Republic and New Netherland around 1626, the economic importance of plantations, and the beliefs and values of different people in the 17th century. After the student groups reconstructed the context of the New Netherland exchange using the guiding questions, the teachers corrected mistakes and provided further explanation when needed.

The third lesson activity uses the historical context to explain historical phenomena. After the historical context of the case was reconstructed by the student groups, the teachers asked students in a classroom discussion again to answer the central question of the case but now while referring to their acquired historical context knowledge. Teachers explicitly stressed that considering the historical context could make students aware of their possible present-oriented perspectives while examining the past. For the case of the exchange of the colony of New Netherland, the students again had to answer the following question: "Can you explain why the Dutch Republic exchanged New Netherland for Suriname in 1626?" To answer this question, students had to, for example, compare the economic importance of Suriname (which had far more plantations and raw minerals) to the economic importance of New Netherland (which had far fewer plantations and raw minerals). At the end of this lesson activity, the teachers and students together evaluated any possible shift among the students from a present-oriented perspective towards a historically contextualized perspective.

The fourth lesson activity was a historical empathy task, where students had to study a historical agent related to the historical topic of the case. To design these

historical empathy tasks, we used the theoretical framework of Endacott and Brooks (2013), who argue that effective historical empathy tasks address three components: historical contextualization, affective connections, and perspective adoption. For the New Netherland case, the historical agent was Willem Bosman, a director of the Dutch West-India Company as well as a merchant and slave trader. The students were given a short description of the historical context and historical agent and had to answer a question similar to this: "If you were Willem Bosman, would you fear being prosecuted for crimes against humanity?" This question addresses the three components of the framework of Endacott and Brooks (2013) because the answer requires historical context knowledge (i.e., the economic and political circumstances of the Dutch Republic in the late 17th century), affective connections (i.e., seeking a connection between the life of Willem Bosman and the students' lives), and adopting the perspective of a historical agent (i.e., understanding Bosman's beliefs, position, and attitude).

6.4.4.3 Reviewing the pedagogy

Brown (1992) argues that educational interventions must be designed to inform practice. The intervention must therefore be easily translated from experimental classrooms to average classrooms and from experimental teachers to average teachers. Considering this important point and to further examine the ecological validity of the pedagogy, we established a focus group to review the developed pedagogy for its practical use. In total, 10 history teachers (all with more than 10 years of experience as a history teacher) participated. To structure the discussion, we presented the lesson activities of the pedagogy and asked the teachers to review each lesson activity for its practical use.

Most teachers found that the concept of the cases triggered presentism among the students, which was exciting and motivating for the students. However, three teachers had some feedback regarding two cases. Based on suggestions from these teachers, we developed two different cases. The teachers liked the structure of first presenting a case, reconstructing the context, and finally using historical context knowledge to explain the case. The teachers also approved of the historical empathy task but were concerned that it might be too strenuous for the students to cover in one lesson. We ended the discussion by asking for general remarks regarding the pedagogy. In general, the teachers noted that the students' ability to perform historical contextualization should be increased with the pedagogy. Despite the teachers' mild concern about the

length of the lesson unit, we chose to maintain the length of the intervention (eight lessons) because a shorter intervention may not result in a deeper understanding of the concept of historical contextualization (e.g., Reisman & Wineburg, 2008).

6.4.5 The control condition

To test the pedagogy, a control condition was designed using previous research in which we observed how teachers promote historical contextualization (Huijgen et al., 2018). In most of the observed lessons, the teachers first activated the students' prior knowledge by asking the students questions. Next, the teachers explained a historical event by reconstructing the historical context. Finally, the students had to finish the history textbook assignments, which were also evaluated after completion. We therefore used this lesson structure as the core for the control condition lessons. Dutch history textbooks do not contain assignments focusing explicitly on historical contextualization. Table 21 presents an overview of the different lessons in the experimental and control conditions. This first lesson of both conditions after the pre-test is described in more detail since the following lessons have the same lesson structure and activities but differ in historical topic.

Table 21. Overview of the lesson activities in the experimental and control condition

Lesson	Historical topic*	Experimental condition	Control condition
1 & 2	Pre-tests	Pre-test historical contextualization	Pre-test historical contextualization
3	Absolutism in the 17th century	Case to enhance awareness of present-oriented perspectives: In a classroom discussion, the teacher asks the students to explain why the palace of Versailles was so large and expensive while many French people suffered from a famine. In the discussion, the teacher uses the students' present-oriented answers to explain the consequences of viewing the past from a present-oriented perspective (i.e., not able to explain the case).	Prior knowledge activation: The teacher activates the students' prior knowledge on 17th-century absolutism by asking questions in a classroom discussion.
		Task to reconstruct the historical context: In groups of four, students reconstruct a historical context of 17th-century absolutism based on the different frames of reference (i.e., chronological, spatial, socio-economic, socio-political, and socio-cultural). The teacher checks the reconstructed context of the different groups for correctness and provides help when needed.	Teacher lecturing: The teacher explains the concept of 17th-century absolutism and the students take notes.
		Explanation of the case: The teacher asks the students to explain the case again but now to explicitly use the gained historical context knowledge. In this classroom discussion, the teacher explains the importance of historical contextualization by stressing the differences between the students' present-oriented answers (from the first lesson activity) and the contextualized answers.	Individual assignments: Students work individually to complete the history textbook assignments on absolutism, and the teacher helps the students when needed.
		Historical empathy task: The students have to answer two explanatory questions about Louis XIV. The students' answers are evaluated by the teacher for historical contextualization.	Whole-class discussion: Students' answers to the textbook assignments are discussed, and the teacher clarifies the answers when needed.
4	The Dutch Republic in the 17th century	Case to enhance awareness of present-oriented perspectives: Students have to explain why there was significant criticism of a former Dutch prime minister who said that we need to go back to the time of the Dutch East India Company.	Prior knowledge activation: The teacher activates the students' prior knowledge on the Dutch Republic in the 17th century.
		Task to reconstruct the historical context: Students reconstruct the historical context of the Dutch Republic.	Teacher lecturing: The teacher explains the origin and characteristics of the Dutch Republic.
		Explanation of the case: The teacher asks the students to explain the case again.	Individual assignments: Students work individually to complete the history textbook assignments.
		Historical empathy task: The task included two questions about the Dutch 17th-century politician Johan de Witt. For example, the students have to explain how De Witt viewed Louis XIV.	Whole-class discussion: Students' answers to the textbook assignments are discussed.

Table 21. continued

Lesson	Historical topic*	Experimental condition	Control condition
5	Worldwide trading in the 17th century	Case to enhance awareness of present-oriented perspectives: Students have to explain the exchange of the colony of New Netherland for Suriname.	Prior knowledge activation: The teacher activates the students' prior knowledge on trading in the 17th century.
		Task to reconstruct the historical context: Students reconstruct the historical context of trade in the 17th century.	Teacher lecturing: The teacher explains how people traded in the 17th century.
		Explanation of the case: The teacher asks the students to explain the case again.	Individual assignments: Students work individually to complete the history textbook assignments.
		Historical empathy task: The task focused on the Dutch slave trader Willem Bosman. For example, students have to explain why Bosman was not arrested by the government for conducting crimes.	Whole-class discussion: Students' answers to the textbook assignments are discussed.
6	The scientific revolution in the 17th century	Case to enhance awareness of present-oriented perspectives: Students have to explain why Copernicus' book on the Solar System was placed on a list of forbidden books.	Prior knowledge activation: The teacher activates the students' prior knowledge on the scientific revolution.
		Task to reconstruct the historical context: Students reconstruct the historical context of the scientific revolution.	Teacher lecturing: The teacher explains the origin and characteristics of the scientific revolution.
		Explanation of the case: The teacher asks the students to explain the case again.	Individual assignments: Students work individually to complete the history textbook assignments.
		Historical empathy task: The historical agent was Antoni van Leeuwenhoek, a Dutch inventor. Students, for example, have to examine how most people would have reacted when Van Leeuwenhoek said that he could see animalcules.	Whole-class discussion: Students' answers to the textbook assignments are discussed.
7	The Enlightenment in the 18th century	Case to enhance awareness of present-oriented perspectives: Students have to explain why Montesquieu's book on the Trias Politica was forbidden in many European countries and why Montesquieu even received death threats.	Prior knowledge activation: The teacher activates the students' prior knowledge on the Enlightenment.
		Task to reconstruct the historical context: Students reconstruct the historical context of the Enlightenment.	Teacher lecturing: The teacher explains the origin and characteristics of the Enlightenment.
		Explanation of the case: The teacher asks the students to explain the case again.	Individual assignments: Students work individually to complete the history textbook assignments.
		Historical empathy task: Voltaire was the historical agent of the task. For example, students have to explain why Voltaire fled to the Lorraine area after he had published Lettres anglaise in 1734.	Whole-class discussion: Students' answers to the textbook assignments are discussed.

Table 21. continued

Lesson	Historical topic*	Experimental condition	Control condition
8	Enlightened absolutism in the 18th century	Case to enhance awareness of present-oriented perspectives: Students have to explain why Catherine the Great, an enlightened absolutist monarch, became far stricter at the end of the 18th century.	Prior knowledge activation: The teacher activates the students' prior knowledge on Enlightened absolutism.
		Task to reconstruct the historical context: Students reconstruct the historical context of enlightened absolutism.	Teacher lecturing: The teacher explains the origin and characteristics of enlightened absolutism.
		Explanation of the case: The teacher asks the students to explain the case again.	Individual assignments: Students work individually to complete the history textbook assignments.
		Historical empathy task: The task focuses on Frederick the Great, an enlightened absolutist monarch and Prussian King. Students have to explain, for example, the extent to which Frederick the Great was an enlightened monarch.	Whole-class discussion: Students' answers to the textbook assignments are discussed.
9	Trans-Atlantic slave trade in the 18th century	Case to enhance awareness of present-oriented perspectives: Students have to explain why slavery in the Netherlands was abolished in 1863 while other European countries abolished slavery much earlier.	Prior knowledge activation: The teacher activates the students' prior knowledge on the Trans-Atlantic slave trade.
		Task to reconstruct the historical context: Students reconstruct the historical context of the Trans-Atlantic slave trade.	Teacher lecturing: The teacher explains the origin and characteristics of the Trans-Atlantic slave trade.
		Explanation of the case: The teacher asks the students to explain the case again.	Individual assignments: Students work individually to complete the history textbook assignments.
		Historical empathy task: The historical agent was Harriet Beecher Stowe, an American abolitionist and author. Students, for example, have to explain how Stowe's opinions were received in the southern United States.	Whole-class discussion: Students' answers to the textbook assignments are discussed.
10	Democratic revolutions in the 18th century	Case to enhance awareness of present-oriented perspectives: Students have to explain the joy of the French people when Marie-Antoinette, who was married to King Louis XVI, was executed in 1793.	Prior knowledge activation: The teacher activates the students' prior knowledge on the democratic revolutions.
		Task to reconstruct the historical context: Students reconstruct the historical context of the democratic revolutions.	Teacher lecturing: The teacher explains the origin and characteristics of the democratic revolutions.
		Explanation of the case: The teacher asks the students to explain the case again.	Individual assignments: Students work individually to complete the history textbook assignments.
		Historical empathy task: Joan Derk van der Capellen tot den Pol was the historical agent. He was a nobleman who wanted far more political influence for the Dutch people. Students, for example, have to explain how Van der Capellen tot den Pol's ideas were received by different sectors of the population.	Whole-class discussion: Students' answers to the textbook assignments are discussed.
11&12		Post-test historical contextualization	Post-test historical contextualization

Note. *The experimental and control lessons focus on the same historical topic.

6.4.6 Implementation fidelity

The implementation fidelity of the experimental and control condition was checked by post-intervention interviews (cf. Nelson, Cordray, Hulleman, Darrow, & Sommer, 2012). In the post-intervention interviews, we asked all the teachers to score how each lesson activity of the experimental and control conditions was implemented (0 = not implemented at all, 1 = partly implemented, and 2 = fully implemented). Table 22 presents the average implementation scores of the different lesson activities in both conditions on the 2-point scale.

Table 22. Implementation scores for the lesson activities (maximum score = 2.00)

Lesson activity	Implementation score
Experimental condition	
1. Case to enhance awareness of present-oriented perspectives	1.60
2. Task to reconstruct the historical context	1.60
3. Explanation of the case	1.48
4. Historical empathy task	1.12
Control condition	
1. Prior knowledge activation	1.90
2. Teacher lecturing	1.98
3. Individual assignments	1.66
4. Whole-class discussion	1.36

6.5 Results

6.5.1 Historical contextualization

Table 23 presents the students' mean historical contextualization pre- and post-test scores for the two conditions (experimental and control). The two conditions differ only slightly in their mean pre-test scores, but the mean post-test scores differ to a much greater extent. To assess the comparability of the conditions prior to the intervention, we evaluated the differences between the students' pre-test scores in the different conditions. This evaluation revealed no significant differences ($F(1,129) = 0.18$, $p = .89$, $\eta_p^2 = .00$). We did find a significant difference between the students' post-test scores in the different conditions ($F(1,129) = 10.70$, $p = .001$, $\eta_p^2 = .08$).

To examine the gains made by the experimental group, a paired sample test was conducted that revealed a significant difference between the students' pre-test and post-test scores in the experimental condition; $t(100) = -2.37$, $p = .02$. To further assess the gains of the experimental group, an effect size was calculated. Morris (2008) describes an effect size for the pre-test-post-test-control design where the standardized effect of the treatment is defined as the difference between groups

in the mean pre-post change divided by the standard deviation of the untreated population. In our case, this effect size is .72, which is an effect between intermediate and large. This standardized effect of the treatment is significant ($p = .001$).

Table 23. Students' mean scores on historical contextualization

Condition	*n*	Pre-test mean (SD)	Post-test mean (SD)
Experimental	101	11.00 (2.47)	11.53 (2.37)
Control	30	11.07 (1.98)	9.90 (2.43)
Total	131		

To examine the intervention effect, we first used a multilevel analysis to explore the extent to which the differences in student achievement on historical contextualization can be explained by the differences between classes. We specified classes as a random factor and the pre-test scores as a fixed factor (-2*LL* = 539.25). This model showed that the total variance of student achievement is 4.20 and that 22% of this variance (0.94) can be explained by the differences between classes. Next, we specified classes as a random factor and the pre-test scores and condition as fixed factors to examine the extent to which the differences in student achievement between the different classes can be explained by participating in the experimental condition (-*2LL* = 535.02, indicating a better fit). This model showed a total variance of 3.71, and 12% of this variance (0.46) can be explained by the differences between classes.

The comparison of the two models showed that the treatment only affected the variance explained by the differences between classes (which decreased from 0.94 to 0.46) and not the residual student variance, which remained the same. The result is that more than half (51%) of the differences between the different classes can be explained by participation in the experimental condition. The effect of the treatment on the differences between the classes was significant ($p < .05$). We calculated the effect size to examine the amount of variance within the experiment that is explained by the treatment. Our multilevel analyses showed that the treatment was responsible for 11% of the differences in student achievement between students in the experimental condition and those in the control condition, which is considered a medium effect (Cohen, 1988).

6.5.2 Students' improvement in historical contextualization

To further explore how students in the experimental condition might have improved in historical contextualization, we asked the teachers in the post-intervention interviews to evaluate the intervention based on their diary notes and experiences.

All teachers noted that the lesson structure of (1) present a historical case at the start of the lesson, (2) instruct students to reconstruct a historical context of this case, and (3) instruct students to evaluate the historical case again using their acquired historical context knowledge promoted historical contextualization. For example, Lisa described in the post-intervention interview that a student immediately reacted from a present-oriented perspective when she showed the painting of the enormous Palace of Versailles, the large building costs and the poor circumstances of many French people. This student noted that people in the past must be really stupid to accept that this palace could be built because the building cost could better be spent on preventing people from dying of starvation. After Lisa explained that one must consider the specific circumstances when explaining historical events and agents' actions and a historical context of was reconstructed (i.e., the political, economic, and cultural circumstances of 17th-century France) by the students, Lisa noticed that her students were more able to explain the building of the palace. For example, the student who displayed a present-oriented perspective at the beginning of the lesson now used historical context knowledge by considering that French kings in that time period saw themselves as substitutes for God and therefore ruled by absolutism. The student now understood that the French people did not have any political influence and that they could not protest such decisions. Moreover, Lisa noted that the student compared the historical context with the present political situation (i.e., elections to influence political decisions). When Lisa asked the student to explain why he had changed his answer from his answer in the first lesson activity, the student noted that he knew now that he had to consider the specific circumstances at that time to answer a question about the past.

Another example how students improved in historical contextualization using this lesson structure was provided by David. He experienced the same shift as Lisa among many of his students when he introduced the exchange of New Netherland for Suriname. Many students reacted with "That is insane" or "That is really not a good deal." These students viewed the historical event from a present-oriented perspective (i.e., exchanging a very economically important city for a nugatory country). After

the reconstruction of the historical context of this exchange (e.g., the Third Anglo-Dutch War, the plantations of Suriname, triangular trade), the students understood the historical event better because they considered chronological and economic historical context knowledge. For example, different students mentioned that people such as Stuyvesant could not have known that New Netherland would become New York City and that Suriname had far more plantations in the 17th century.

Moreover, all teachers noted that the historical empathy tasks promoted historical contextualization because by examining the life of historical agents their students learned how historical agents perceived historical events resulting in the consideration of the specific circumstances of a historical event. Wendy explicitly stressed the additional value of the historical empathy tasks besides the other three lesson activities. Wendy noted that her students found it very difficult to understand and explain the Enlightenment in the 18th century, even after the historical context of the Enlightenment was reconstructed and discussed. One of her students noted that it was not possible to understand the Enlightenment "because there is so much to understand." The historical empathy task consisted of a historical source that described the life of Voltaire and two accompanying questions focusing on how Voltaire saw the Church and why Voltaire risked arrestment. By examining the life of Voltaire, her students were able to understand the broader historical context of the Enlightenment because "the abstract became more concrete for them," as Wendy noted in the post-intervention interview. For example, one of her students noted that Voltaire criticized the absolute emperors and religious dogmas of his time. This student understood that Voltaire might have fled because these views were not common in that time period and could therefore triggered resistance among the rulers.

Despite these positive findings, the teachers noted three main issues than can be used to further improve the effectiveness of the intervention to promote historical contextualization. The first issue is that the different lesson activities took more time than estimated. Lisa and Wendy (who both hold an average implementation score of 1.00 out of a 2-point scale) noted that they did not complete a number of different lesson activities due to a lack of lesson time. They found eight lessons too long to implement an intervention because they had to prepare students for formal tests. The other teachers ranged in implementation scores between 1.59 and 1.88 and experienced this problem less but also acknowledge that the lesson activities took more time than expected. Because the lesson activities took longer than estimated,

the teachers skipped the historical empathy tasks the most because these tasks were scheduled at the end of each lesson. Each teacher, however, conducted at least four of the eight historical empathy tasks.

Secondly, all teachers noted that students became demotivated after three or four lessons due to the repetitive lesson structure. Instead of a repetitive structure, Ben suggested to use only four lessons and to present in the first lesson a historical case that might trigger present-oriented perspectives and an accompanying explanatory question. After the case has been discussed, the teacher could stress the danger of presentism, explain the importance of historical contextualization, and model historical contextualization (for example, by discussing the guidelines of Appendix F). This lesson is followed by two lessons where the students and teacher work together on reconstructing the historical context to answer the question of the historical case. In the fourth and final lesson, the teacher evaluates the answer to the question of the historical case with the students.

Finally, Lisa, Ben, and Wendy suggested to focus more on the differences between individual students because some of their students were already aware of the consequences of presentism while others viewed historical events from a dominant present-oriented perspective. Lisa suggested to use a different lesson structure to address student differences:

> Teachers might present a central historical case or problem and instruct students in groups to examine the historical case on their own rather than discussing the historical case directly in a classroom discussion. This provides the opportunity to evaluate how the different groups perform historical contextualization and then I can provide more customized instructions when students ask for help. For example, when groups keep viewing the past from present-oriented perspectives, I can explain the consequences of presentism to this group. When the students do not know how to reconstruct a historical context, I can provide a handout with the frames of reference as guiding questions.

6.6 Conclusions and discussion

The aim of this explorative study was to develop a pedagogy and to test it to assess its success in improving students' ability to perform historical contextualization using a quasi-experimental pre- and post-test design. In contrast to scholars who focused on contextualization as a heuristic to examine historical documents (e.g., Baron, 2016; Reisman, 2012b) or on students' knowledge and strategies to date historical sources and events (e.g., Van Boxtel & Van Drie, 2012; Wilschut, 2012), we explored whether the teaching strategies of Huijgen, Van de Grift, et al. (2017) could be used to develop a historical contextualization pedagogy. The results of a historical contextualization test showed that students in the experimental condition demonstrated more progress in their ability to perform historical contextualization compared to students in the control condition. A multilevel analysis indicated that the developed pedagogy had a medium effect on students' ability to perform historical contextualization.

The teachers' post-intervention interviews indicate that the structure: (1) presenting a historical case that triggers possible present-oriented perspectives, (2) instructing students to reconstruct a historical context, and (3) instructing students to use historical context knowledge to evaluate the historical case again, can promote historical contextualization. Similar approaches have been suggested by scholars such as Reisman (2012a) and Havekes et al. (2012), but positive indicators of this approach in promoting students' ability to perform historical contextualization were still missing. Moreover, in line with scholars such as Lee and Ashby (2001) and VanSledright (2001) who argue that historical empathy can promote historical contextualization, our findings seem to illustrate that the historical empathy tasks helped students perform historical contextualization. The historical empathy tasks might make historical events more concrete for students (cf. De Leur, Van Boxtel, & Wilschut, 2017) and let them grasp the "sense of a period," as Dawson (2009) calls it.

Despite the positive indicators, all teachers noted that the lesson activities took more lesson time than estimated. Especially the historical empathy tasks (which were scheduled at the end of each lesson) were therefore not always completed. Two teachers explicitly stressed that implementing all eight lessons would have left them little time to prepare their students for the formal test. To integrate the historical empathy tasks more within the other lesson activities a structure of Endacott and Pelekanos (2015) can be used where students are first introduced to historical agents (introduction phase), reconstruct a relevant historical context (investigation phase),

and finally demonstrate and reflect on their historical understanding (display and reflection phase). Following this structure, the historical empathy tasks of our study can be presented as historical cases which trigger possible present-oriented perspectives (introduction phase). For example, students can be provided with a description of a European slave trader who treats slaves badly and have to reason if this slave trader risked arrestment. Subsequently, students have to reconstruct a historical context in groups or dyads (investigation phase). Finally, the teacher and the students evaluate the historical case, for example, by reasoning if the slave trader got arrested (display and reflection phase). These lesson activities can be distributed across multiple (e.g., three or four) lessons resulting in more time and flexibility for teachers.

Spreading the lesson activities across different lessons might also motivate students more since there is no repetitive lesson structure. Teachers might also start with basic instructions (e.g., teachers create a historical context and explain the past) in the first lessons and progress to more complex instructions (e.g., students working with historical sources to create a historical context to explain the historical event) in following lessons to motivate students, (e.g., Merriënboer & Kirschner, 2007). One of the teachers suggested a similar approach to prevent a repetitive structure. Moreover, to motivate students it is also important to address differences between students (Ginsberg, 2005; Subban, 2006). Three teachers noted that the intervention does not address these differences. An improvement, for example, could be to provide the guiding questions only to the students who need help in reconstructing a historical context.

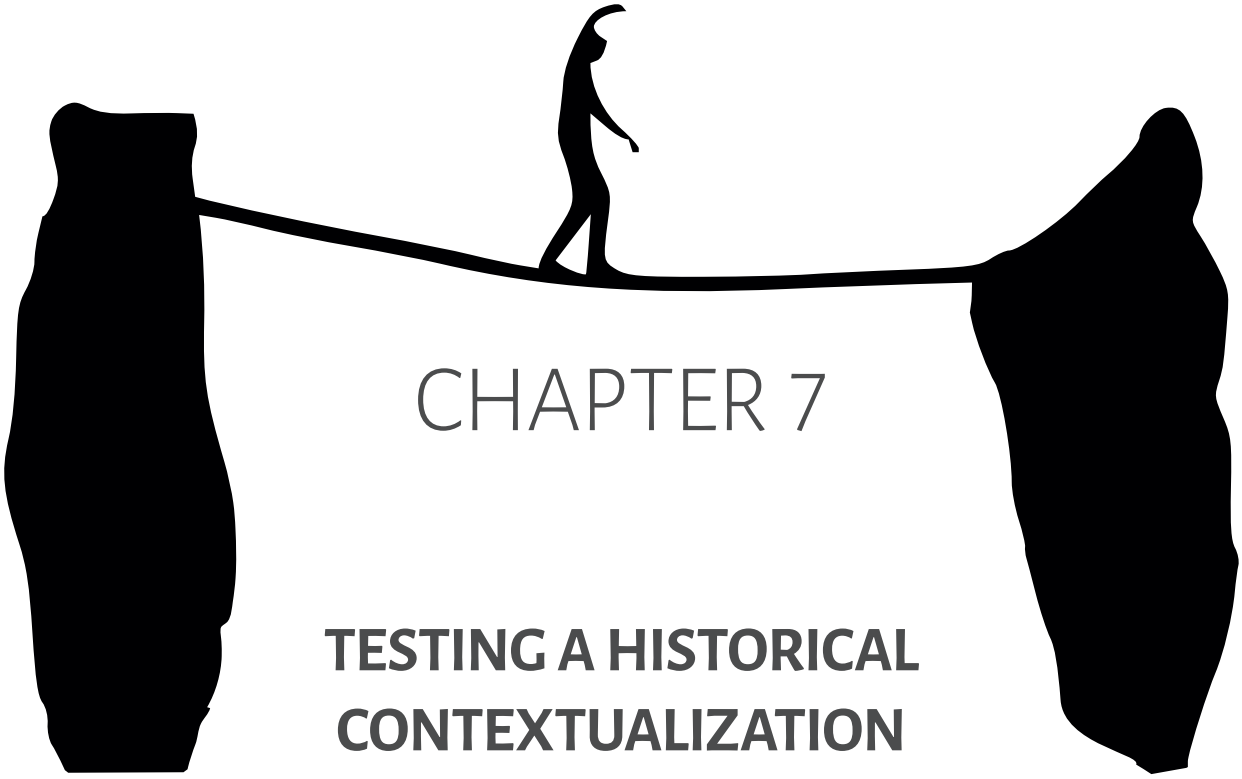
An important limitation of our explorative study is the small sample size, especially for the participants in the control condition (two teachers and 30 students). A design using more participants and random sampling would be preferred (cf. Raudenbush & Bryk, 2002). Experimental studies should also be repeated in different settings to confirm the findings (Collins, Joseph, & Bielaczyc, 2004). Another limitation is the tests used to measure the students' ability to perform historical contextualization. The internal consistency (Cronbach's alpha) of both instruments is on the lower end of what is considered acceptable. Refining the items by, for example, using thinking aloud protocols could provide insights into ways to increase the internal consistency. Moreover, the tests measure the ability to perform historical contextualization at a very basic level. Including History Assessments of Thinking in historical contextualization could provide other insights because these assessments also require student argumentation (Breakstone, Smith, & Wineburg, 2013). The implementation fidelity scores of the experimental condition might also be a limitation since not all lesson

activities were completed due to a lack of time. An approach where the lesson activities are more evenly distributed across different lessons is therefore preferred.

Future research on testing the pedagogy should also pay more attention to the use of mixed methods, as advocated by Shadish et al. (2002), because combining quantitative data with more qualitative data (e.g., thinking aloud protocols triggered by stimulated recall methods) provides insight into teachers' and students' motives and experiences during an intervention. In this study, teachers' diaries and post-intervention interviews were only used as a qualitative method to gain insights in how students improved in historical contextualization. A protocol analysis of a classroom discussion during the intervention and students' responses to contextualization tasks, as suggested by Reisman (2012b), could be more valuable to examine the students' progress in the ability to perform historical contextualization and their situational interest. Moreover, since research suggests that historical contextualization might also promote competencies such as learning about democratic citizenship, social perspective adoption, and the ability to adopt multiple perspectives (e.g., Barton, 2012; Gehlbach, 2004), it would be interesting to examine the effects of the pedagogical design principles for these competencies.

Finally, we discuss some practical implications for the teaching and learning of history. Since there might be a dichotomy between historical skills and knowledge in history education (Counsell, 2000) and teachers might experience problems when teaching historical reasoning competencies (e.g., Barton & Levstik, 2003; Hall & Scott, 2007), the pedagogy could help teachers combine the teaching of historical content knowledge and historical reasoning competencies in a practical manner. Teachers who want to explicitly teach historical contextualization could start with implementing the cases in their lessons to prevent presentism among their students.

To conclude, intervention studies are scarce within the field of history education research; however, more attention has been given recently to the use of this methodology to examine the learning and teaching of history (e.g., De La Paz et al., 2014; Reisman, 2012b; Stoel, Van Drie, & Van Boxtel, 2017). To contribute, we conducted an intervention study focusing on the learning and teaching of historical contextualization. The developed pedagogy may help teachers not only teach students historical facts but also actively engage them in the process of historical contextualization to understand and explain the differences and connections between the past and present.



CHAPTER 7

TESTING A HISTORICAL CONTEXTUALIZATION FRAMEWORK

This chapter elaborates on the findings of the previous chapter and presents an example of how the design principles of historical contextualization can be used to develop a three-stage framework. This framework was used to design a lesson unit on Cold War events. The effects of the lesson unit on students' ability to perform historical contextualization are explored in a quasi-experimental pre-test–post-test design with an experimental (n = 96) and control (n = 73) condition. The students' answers to a historical contextualization test were analyzed. The results indicate that students in the experimental condition increased their ability to perform historical contextualization and displayed less present-oriented perspectives in their answers compared to students in the control condition.

This chapter is based on: Huijgen, T. D., Holthuis, P., Van Boxtel, C. A. M., Van de Grift, W. J. C. M., & Suhre, C. J. M. (2018). Students' historical contextualization and the Cold War. *British Journal of Educational Studies.* Advance online publication. doi:10.1080/00071005.2018.1518512

7.1 Introduction

"I think it is just stupid when you lose your job when you married. And females should do all the household labor too, shouldn't they? Men are also grown-ups, right? Let them do the cooking and cleaning." This was said by Lisa, a 14-year-old secondary school student, when we asked her to explain why, until the late 1950s, Dutch female governmental officials lost their jobs when they married. Lisa reacted with disbelief and was not able to understand or explain this historical phenomenon. Lisa and many other students tend to view and judge the past from a present-oriented perspective instead of using historical context knowledge to explain and understand historical phenomena (Foster, Lee, & Ashby, 2008; Huijgen, Van Boxtel, Van de Grift, & Holthuis, 2014). To help students to view and judge the past in its own terms, it is necessary to increase their ability to perform historical contextualization (Wineburg, 2001). Historical contextualization is the ability to situate phenomena and actions in the context of long-term developments, their specific time, and the historical location to be able to give meaning to these phenomena and actions (Van Boxtel & Van Drie, 2012).

Previous research has indicated, however, that history teachers might demonstrate historical contextualization themselves in their lessons but do not explicitly engage students in historical contextualization processes (Huijgen, Holthuis, Van Boxtel, & Van de Grift, 2018). For example, the teachers included in the sample often reconstructed a historical context themselves instead of creating opportunities for students to use historical sources to create a historical context on their own. The students might therefore miss opportunities to practice their historical contextualization skills and keep viewing the past based on their own values and beliefs (Reisman & Wineburg, 2008).

Moreover, most studies focusing on the development and testing of contextualization pedagogies consider contextualization to be a heuristic along with, for example, sourcing, collaboration, and close-reading. Building upon the work of Wineburg (1991, 1998), Reisman (2012b), and De La Paz et al. (2014) examined the use of these heuristics in primary source instruction and in a disciplinary reading and writing curriculum intervention, respectively. In contrast to these studies, we focus solely on students' ability to perform historical contextualization and the role contextualization plays in preventing present-oriented perspectives among students.

Building upon previous work in which design principles of historical contextualization were operationalized (Huijgen, Van de Grift, Van Boxtel, & Holthuis, 2018), in this study, we present an example of how these design principles can be used as a three-stage framework to develop a lesson unit focusing on Cold War events. The effects of these lessons on students' ability to contextualize historical events are explored using a quasi-experimental pre-test–post-test design with an experimental and control condition.

7.2 Theoretical framework

7.2.1 Historical thinking and reasoning in classrooms

In history classrooms, students not only need to learn to memorize historical facts but should also be engaged in historical thinking and reasoning, such as working with historical sources, asking historical questions, determining change and continuity, and performing historical contextualization (Lévesque, 2008; Seixas & Morton, 2013; Van Drie & Van Boxtel, 2008). Influenced by the work of Peter Seixas, in Canadian states such as Ontario, historical thinking competencies are explicitly mentioned in the curriculum (Ontario Ministry of Education, 2015). In Australia, the history curriculum directs students to use different historical skills, such as understanding the different social, cultural, and intellectual contexts that shaped people's lives and actions in the past (National Curriculum Board, 2009). With the development of The College, Career, and Civic Life (C3) Framework for Social Studies State Standards (National Council for the Social Studies, 2013), more attention may also be given to implementing historical reasoning competencies in state curricula in the United States.

Similar to other West European countries (e.g., Erdmann & Hasberg, 2011), historical reasoning competencies are explicitly implemented in the formal Dutch history curriculum. Dutch students have to, for example, explain human behavior in the past (thinking and doing) based on the knowledge and values known and accepted at that specific time. Moreover, they should be able to recognize different value and belief frameworks when they are asked to provide a moral judgement about historical events and agents' actions (Board of Test and Examinations, 2017).

Despite the importance of historical reasoning in history curricula, most teachers seem not to engage students in historical reasoning. More than a decade ago, VanSledright

and Limón (2006) described an average history classroom where lecturing and storytelling by the teacher dominated. In such history classrooms, historical reasoning might not be encouraged since it requires active participation and input from the students (Van Boxtel & Van Drie, 2017). Recent research indicates that little has changed. For example, Reisman (2015), when analyzing videotaped history lessons, concluded that disciplinary discussions were surprisingly rare and that discussion that encouraged historical understanding was even more rare. Saye and Social Studies Inquiry Research Collaborative (2013) found that only 21% of the students in their sample attended classes that met the standards for moderately challenging teaching, such as engaging students in disciplined inquiry. A recent observation study of (Huijgen, Holthuis, et al., 2018) showed that the eight history teachers who were included in the sample rarely engaged students in historical contextualization processes. This study aims, therefore, to help teachers engage students in historical contextualization processes by examining the use of a historical contextualization framework.

7.2.2 Historical contextualization and presentism

Following Wineburg (1991, 1998), several scholars consider historical contextualization to be a heuristic that is used when reading historical texts, in addition to sourcing and corroboration. For example, De La Paz et al. (2014) viewed contextualization as the extent to which students identified and situated arguments and primary sources in the appropriate time, place, and setting. In this study, we use a broader definition of historical contextualization as the ability to situate phenomena and people's actions in the context of long-term developments, their specific times, and historical location to be able to give meaning to these phenomena and acts (Van Boxtel & Van Drie, 2012). Historical contextualization requires an understanding of the social, political, and cultural norms of the time period under investigation and knowledge of the events leading up to the historical situation and other relevant events that happened concurrently (Endacott & Brooks, 2013). However, historical contextualization should not lead to relativism among students, such as the justification of controversial people's actions in the past. Rather, it should help students to make reasoned ethical judgements and to understand and explain historical phenomena and people's actions based on a created historical context (Seixas & Morton, 2013).

Many scholars argue that historical contextualization could prevent *presentism* (e.g., Barton & Levstik, 2004; VanSledright, 2001). The term presentism is often used when

students examine the past with their own knowledge, values, and beliefs, which often results in misunderstanding historical phenomena and agents' actions (e.g., Hartmann & Hasselhorn, 2008; Seixas & Peck, 2004). Wineburg (2001) argues that many students naturally view the past from their own present-oriented perspectives and that historical thinking is therefore an "unnatural act" that needs to be learned in history classrooms. Teaching students historical contextualization could prevent viewing the past from a present-oriented perspective because an important component of historical contextualization is considering the specific circumstances of a historical period when examining the past (Endacott & Brooks, 2013; Reisman & Wineburg, 2008).

7.2.3 Engaging students in historical contextualization

To help teachers develop teaching and learning activities that engage students in historical contextualization processes, we developed four design principles of historical contextualization in previous research: (1) raising awareness of present-oriented perspectives, (2) reconstructing a historical context, (3) creating opportunities to practice historical contextualization to explain historical phenomena or agents' actions, and (4) enhancing historical empathy (Huijgen, Van de Grift, et al., 2018). For this study, we used the first three design principles to develop a three-stage framework in which the teacher (1) presents a historical case that triggers possible present-oriented perspectives, (2) instructs students to reconstruct a historical context for the historical case, and (3) instructs students to use historical context knowledge to interpret the historical case again.

We chose not to use the design principle of historical empathy because this principle can be incorporated in the different stages of the framework. For example, students can be asked to imagine the thoughts and feelings of individual historical agents using their own "similar" life experiences to reconstruct a historical context. The three-stage historical contextualization framework is visualized in Figure 4. First, awareness is raised concerning possible present-oriented perspectives by presenting a historical case that students find difficult to explain. Next, the historical context of the particular case is reconstructed. Finally, students and the teachers interpret the historical case again with their newly acquired historical context knowledge.

The first component of the framework is raising students' awareness of their possible present-oriented perspectives. Building upon work in the field of cognitive conflicts

(e.g., Johnson & Johnson, 2009), different scholars argue that historical tension might contribute to encouraging students' ability to perform historical contextualization (Havekes, Coppen, Luttenberg, & Van Boxtel, 2012; Huijgen & Holthuis, 2015). Historical tension is created when students are not able to explain a historical event or a historical agent's action because of their present-oriented perspectives. For example, the teachers might present a case about a 20-year-old man living in 1930 in Germany and ask the students if they can explain why this man might have voted for the Nazi Party (Hartmann & Hasselhorn, 2008). Students are often inclined to view and judge these types of historical events based on their own values, knowledge, and beliefs (Wineburg, 2001). For the lesson unit of this study, we therefore designed historical cases that encouraged historical tension to provide opportunities for teachers to discuss the consequences and limitations of viewing the past from present-oriented perspectives.

The second component of the framework is teaching students how to reconstruct a historical context successfully. The students therefore need explicit guidelines (Havekes et al., 2012; Reisman & Wineburg, 2008). For the lesson unit of this study, we used a chronological, spatial, political, economic, and cultural frame of reference as guidelines for students to reconstruct a historical context for a phenomenon or source (De Keyser & Vandepitte, 1998). These guidelines also function as a checklist since they provide students with the opportunity to review what they do and do not know about a historical event. For example, when students are asked to reconstruct the historical context of the Cuban Missile Crisis of 1962, they might forget the geographical context, which is essential to understanding and explaining this crisis. Considering all frames of reference reduces the chances of students missing important and relevant historical context knowledge. The guiding questions can be found in Appendix G.

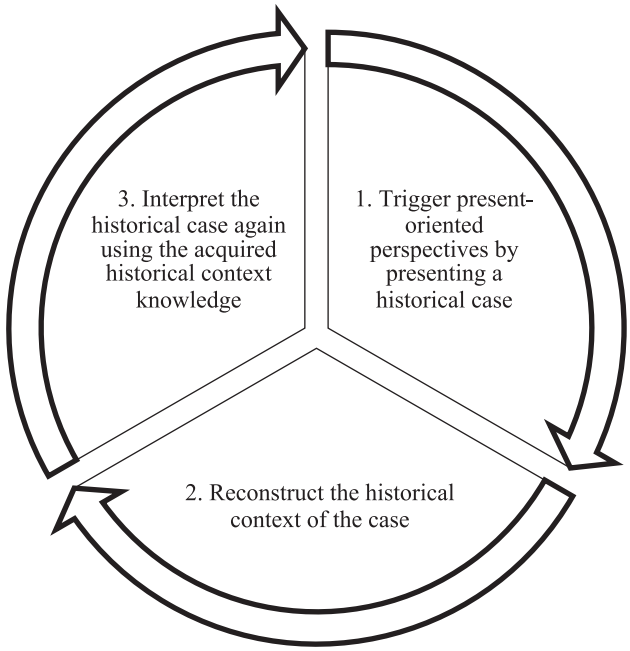


Figure 4. The three-stage historical contextualization framework

The third and final component of the framework is based on the idea that students in history classrooms often have to explain, compare, and interpret historical phenomena and sources (Haydn, Stephen, Arthur & Hunt, 2015; Lévesque, 2008; Seixas & Morton, 2013). Only becoming aware of a possible present-oriented perspective and knowing how to reconstruct a historical context is not enough to do this successfully. The students must also learn to use their abilities to perform historical contextualization to examine and interpret historical phenomena and sources. The third component aims, therefore, to create opportunities for students in the lessons to perform historical contextualization to explain, compare, and interpret historical phenomena and sources. In the lesson unit, we therefore created opportunities for students to use their newly acquired historical context knowledge to interpret the historical case again.

7.3 Research question

The central research question of this study is "What are the effects of a lesson unit, based on a three-stage historical contextualization framework, on 14- to 16-year-old students' ability to perform historical contextualization?" Since the intervention aimed at promoting historical contextualization, we expect that the intervention helps to increase the use of historical contextualization and to decrease the use of presentism.

7.4 Method

7.4.1 Research design

Based on the three-stage historical contextualization framework, a lesson unit (four lessons) focusing on Cold War events was developed. To explore the effects of the lesson unit on changes in students' ability to perform historical contextualization, we used a non-equivalent control group pre-test–post-test design (Shadish, Cook, & Campbell, 2002). A historical contextualization test of six open-ended questions was constructed to explore the students' progress in historical contextualization. This test was administered as a pre- and post-test. The students' answers to the pre- and post-tests were qualitatively analyzed using a coding scheme to explore possible progress in the students' ability to perform historical contextualization (Krippendorff, 2013).

7.4.2 Participants

We asked seven teachers from our professional network to participate in the study. In consultation with the teachers, we decided that four teachers would teach the lesson unit in one of their history classes (experimental condition) and that three teachers would teach the control condition in one of their history classes. Table 24 presents the teachers' characteristics. All teachers participated voluntarily, held a Master's degree in history education, and were Dutch nationals. The participating teachers in the experimental condition received a 2-hour training session given by one of the authors to teach the lesson activities in the lesson unit and to conduct the pre- and post-tests. The teachers participating in the control condition received instructions from one of the authors on how to apply the different lesson activities in the control condition and how they had to conduct the pre- and post-tests.

Table 24. Teachers' characteristics

Teacher	Gender	Age	Years' work experience	Condition	Educational track	Class size
T1	Female	32	6	Experimental	Senior general secondary education	26
T2	Male	50	5	Experimental	Senior general secondary education	26
T3	Male	34	9	Experimental	Pre-university education	20
T4	Female	30	6	Experimental	Pre-university education	24
Total						*96*
T5	Female	23	1	Control	Pre-university education	23
T6	Male	51	25	Control	Senior general secondary education	22
T7	Male	56	25	Control	Pre-university education	28
Total						*73*

In total, 169 secondary school students from the two highest Dutch educational tracks (senior general secondary education and pre-university education) participated in the study, ranging in age from 14 to 16 years old. The mean students' age in the experimental condition was 14.8 (SD = 0.56) years old compared to a mean age of 14.7 (SD = 0.53) years old for the control condition. The female and male distribution in the experimental condition was 48% and 52%, respectively. In the control condition, this distribution was 45% and 55%, respectively. Two students in the sample held non-Dutch nationalities while the other students held Dutch nationality.

7.4.3 The structure of the lesson unit

The three components of the historical contextualization framework were used to develop a 4-lesson unit focusing on Cold War events for secondary students aged 14 to 16 years old. This topic was chosen because it fitted the history teachers' curricula the best at the time of the intervention. The lesson topics (The start of the Cold War and the fear of the atomic bomb, the American fear of communism, and the Hungarian Revolt) are topics implemented in the formal history curricula for the two highest educational tracks in the Netherlands (Board of Tests and Examinations, 2017). In previous research, we used a repetitive lesson structure for eight lessons (Huijgen, Van de Grift, et al., 2018). However, the students and teachers became demotivated because they had to do the same work for each lesson.

In this study, we therefore used only four lessons and used Merrill's (2002) review study of instructional design theory to create a new and more motivating structure in the lesson unit. Five principles are elaborated by Merrill (2002): (1) problem-centered learning, (2) activation of existing knowledge, (3) demonstration of new knowledge, (4) application of new knowledge, and (5) integration of new knowledge. The first two

principles are implemented in the lesson structure because all lessons start with a problem (a historical case aiming to trigger a cognitive conflict), and prior knowledge is activated by asking students in a classroom discussion to examine this historical case by using their prior topic knowledge. Moreover, the first two lessons focused more on demonstrating how to perform historical contextualization successfully (*show me*) and the final two lessons focused more on the application (*let me*) and the integration (*watch me*) of the historical contextualization processes.

7.4.4 The experimental lessons

The historical topic of the first lesson was the start of the Cold War and the development and fear of the atomic bomb. First, the short movie *Duck and Cover* (Federal Civil Defense Administration, 1951) was shown to create historical tension and to trigger possible present-oriented perspectives among students. The film shows what to do in case of a nuclear explosion. The students had to discuss in dyads and in a classroom discussion whether they could imagine receiving similar atomic warfare training. Next, the students were provided with a handout presenting guiding questions for reconstructing a historical context (see Appendix G). The indicative questions were formulated to guide students' thinking. The teacher explained the different steps and the importance of reconstructing a historical context to explain historical events. Next, the teacher used the different frames of reference to reconstruct the context of the start of the Cold War. This context comprised a chronological context (timeline), a spatial context (geographical map), and an explanation of the following historical events: the Russian Revolution of 1917, the collaboration between the Soviet Union and the United States to defeat Nazi Germany, the Yalta and Potsdam conferences, the differences between the Soviet Union and the United States, and the development and fear of the atomic bomb. The lesson ended with the teacher asking the students to use their newly acquired historical context knowledge to review their answer from the first lesson activity. Could they now better explain why American secondary school students received atomic warfare training?

The second lesson focused on the American fear of communism during the Cold War. At the start of the lesson, the students were provided with a historical source about the execution of Julius and Ethel Rosenberg in 1953 to create historical tension. The historical source focused on the marginal evidence of the involvement of Ethel Rosenberg in espionage. In dyads and in a classroom discussion, the students were instructed to discuss whether they could explain the execution of the Rosenbergs.

Next, the teacher used the different frames of reference to reconstruct the context of the start of the Cold War. This context comprised a chronological context (timeline), a spatial context (geographical map), and the following historical events and developments: the enmity between the Soviet Union and the United States, the American fear of communism, and Senator McCarthy. At the end of the lesson, the teacher asked the students to use their newly acquired historical context knowledge to review their answer from the first lesson activity. Are they able to explain the execution of the Rosenbergs better?

In the first two lessons the focus was more on showing students how to perform historical contextualization successfully. Merrill (2002) argues that when information is presented through specific situations or cases, the students will remember and practice this information better. We therefore expected that using a specific historical case will result in a better application of historical contextualization processes. Furthermore, Merrill (2002) noted that learning is encouraged when procedures are demonstrated and behavior is modeled. The guiding questions of Appendix G had this goal. Moreover, procedures and processes must be visible (Merrill, 2002), and the final lesson activity (where the students had to use their newly acquired historical context knowledge to review the historical case again) therefore provided the opportunity for teachers to review and discuss successful and unsuccessful demonstrations of historical contextualization processes.

The third and fourth lesson focused on the Hungarian Revolution of 1956. At the start of the third lesson, the teacher organized students into dyads with two historical pictures (displaying a street name change) and asked them to discuss whether they could explain why a street in Amsterdam, called the *Stalin Lane,* was changed to *4 November Lane* in 1956. Next, the students were divided into groups of four and were provided with five written historical sources about the Hungarian Revolution. One historical source provided general information about the Hungarian Revolution. The second historical source addressed the demands of Hungarian students and the working class presented to the Hungarian government. The third historical source addressed the Soviet invasion from the perspective of a Hungarian journalist. The fourth source presented the perspective of a British journalist on the Hungarian Revolution, and the fifth source presented the perspective of a Russian tourist in Budapest on the Hungarian Revolution. The central task of the third and fourth lessons was to use the historical sources to reconstruct a historical context to explain why the Amsterdam

street name change happened. To reconstruct the historical context of the Hungarian Revolution, the students had to use the guidelines from Appendix G. At the end of the fourth lesson, the students had to present their answer to the other students and received feedback from the teachers. When presenting their answers, the students also had to explain if the task helped them to explain and understand the street name change and if they changed their initial answer from the start of the third lesson.

Compared to the first two lessons in which the focus was more on demonstrating historical contextualization (show me), the third and fourth lesson focused more on the application of new knowledge (let me) and the integration of this knowledge (watch me). In a review by Merrill (2002), it becomes clear that learning is encouraged more when students are required to use their new skills to solve problems. These problems should be real-world tasks instead of providing, for example, multiple-choice questions. We therefore developed an assignment for the third and fourth lessons to examine the Amsterdam street name change, where students have to apply their skills in historical contextualization to complete this assignment successfully. Moreover, in contrast to the first two lessons (where students received more support), the final two lessons involved less student guidance because the students (on their own) had to reconstruct the historical context of the Hungarian Revolt. This scaffolding is considered an effective way to apply new forms of knowledge (Merrill, 2002). Moreover, in effective instruction, there must be an opportunity for students to demonstrate their newly acquired skill of historical contextualization. Therefore, at the end of the fourth lesson, the students had the opportunity to demonstrate, reflect on, defend, and share what they had learned over the past four lessons (Merrill, 2002).

7.4.5 The control condition

Table 25 provides an overview of the lesson activities in the experimental and control condition. The lessons in the control condition comprised the same historical topics and contained different lesson activities, but the students did not receive explicit instruction in historical contextualization. In each of the control lessons, the students' prior knowledge was activated, historical phenomena were explained by the teacher, and the students completed assignments that were also discussed in a classroom discussion. Each lesson ended with a review of the most important historical phenomena of that particular lesson. The students' assignments (which students had to complete during the control lessons) were developed by the authors to prevent deviations between the different control classrooms. The assignments always related

to the lesson topic. For example, the lesson topic of the second lesson was the start of the Cold War (see Table 25), and the first assignment asked students to compare (in table form) the economic, political, and socio-cultural differences between the Soviet Union and the United States. The second question comprised a description of a country's economy. The students had to use this description to explain whether the country is communist or capitalist. The third question asked students to describe how the United States and the Soviet Union became involved in the Second World War. All assignments were examples of regular Dutch history textbook exercises.

Table 25. Overview of the lesson activities in the experimental and control condition

Lesson	Historical topic*	Experimental condition	Control condition
1	Pre-tests	Pre-test historical contextualization	Pre-test historical contextualization
2	Start of the Cold War	Case to enhance awareness of present-oriented perspectives: The teacher provides dyads with a handout with questions and shows the Duck and Cover movie. Central task for the dyads is to reason if they could imagine that they received such atomic warfare training. In a classroom discussion, the dyad's answers are discussed. In this discussion, the teacher uses the students' present-oriented answers to explain the consequences of viewing the past from a present-oriented perspective (i.e., not able to explain the case). Explanation of historical contextualization: The teacher provides the handout with the guiding questions (Appendix G) and explains the importance of historical contextualization. Reconstructing the historical context: The teacher explains the following historical phenomena: Russian Revolution, Collaboration between the USA and the Soviet-Union, Yalta and Potsdam conferences, differences between capitalism and communism, and the development and fear of the atomic bomb. Evaluating the case: The teacher asks the dyads to explain the case again but now to explicitly use the newly acquired historical context knowledge. In this classroom discussion, the teacher explains the importance of historical contextualization by stressing the differences between the students' present-oriented answers (from the first lesson activity) and the contextualized answers.	Prior knowledge activation: The teacher activates the students' prior knowledge by asking questions in a classroom discussion. Teacher lecturing: The teacher explains the following historical phenomena: Russian Revolution, Collaboration between the USA and the Soviet-Union, Yalta and Potsdam conferences, differences between capitalism and communism, and the development and the fear of the atomic bomb. Individual assignments: The teacher provides the students with 13 assignments. They work individually to complete questions 1 to 3. Whole-class discussion: Students' answers to the assignments are discussed. Recap: The teacher repeats the most important historical events explained in this lesson.

Table 25. continued

Lesson	Historical topic*	Experimental condition	Control condition
3	The Rosenbergs	Case to enhance awareness of present-oriented perspectives: The teacher provides dyads with a handout with questions on the execution of the Rosenbergs. Central task for the dyads is to explain why Ethel Rosenberg was executed despite the marginal evidence of espionage. In a classroom discussion, the dyad's answers are discussed. In this discussion, the teacher uses the students' present-oriented answers to explain the consequences of viewing the past from a present-oriented perspective (i.e., not able to explain the case). Reconstructing the historical context: The teacher explains the following historical phenomena: entities between the USA and the Soviet-Union, The rise and fall of McCarthy, and the American fear of communism. Evaluating the case: The teacher asks the dyads to explain the case again but now to explicitly use the newly acquired historical context knowledge. In this classroom discussion, the teacher explains the importance of historical contextualization by stressing the differences between the students' present-oriented answers (from the first lesson activity) and the contextualized answers.	Prior knowledge activation: The teacher activates the students' prior knowledge of the previous lesson. Teacher lecturing: The teacher explains the following historical phenomena: entities between the USA and the Soviet-Union, the rise and fall of McCarthy, and the American fear of communism. Individual assignments: Students work individually to complete questions 4 to 7. Whole-class discussion: Students' answers to the assignments are discussed. Recap: The teacher repeats the most important historical events explained in this lesson.
4	Hungarian Revolt	Case to enhance awareness of present-oriented perspectives: Students in groups of four are provided with two historical sources displaying a street name change in Amsterdam. The groups have to answer if they could explain why this street name change happened. They have to write down their answer on a handout. Reconstructing the historical context: In groups of four, the students are provided with five historical sources describing different perspectives on the Hungarian Revolt of 1956. Students are instructed to create a historical context based on the provided historical sources and to answer the central question (explaining the street name change in Amsterdam). Students can use the handout from the second lesson (guiding questions).	Prior knowledge activation: The teacher activates the students' prior knowledge of the previous lesson. Teacher lecturing: The teacher explains the following historical phenomena: origin of the Warsaw Pact, causes of the Hungarian revolt, and the beginning of the Hungarian Revolt. Individual assignments: Students work individually to complete questions 8 to 10. Whole-class discussion: Students' answers to the assignments are discussed. Recap: The teacher repeats the most important historical events explained in this lesson.

Table 25. continued

Lesson	Historical topic*	Experimental condition	Control condition
5	Hungarian Revolt	Reconstructing the historical context: In groups of four, the students continue working at the reconstructing of the historical context of the Hungarian Revolt of 1956. Evaluating the case: The student groups have to present their answers of the assignment for the other students at the end of the fifth lesson. What did they find? Furthermore, they have to discuss if their answer from the lesson start of fourth lesson has changed	Prior knowledge activation: The teacher activates the students' prior knowledge of the previous lesson. Teacher lecturing: The teacher explains the following historical phenomena: The end and consequences of the Hungarian Revolt. Individual assignments: Students work individually to complete questions 11 to 13. Whole-class discussion: Students' answers to the assignments are discussed. Recap: The teacher repeats the most important historical events explained in this lesson.
6	Post-tests	Post-test historical contextualization	Post-test historical contextualization

Note. *The experimental and control lessons focus on the same historical topic.

7.4.6 The historical contextualization test

To test the effects of the pedagogy on the students' ability to perform historical contextualization, we developed a historical contextualization test based on the History Assessments of Historical Thinking (HATs) developed by the Stanford History Education Group (e.g., Breakstone, Smith, & Wineburg, 2013) and instruments used in previous research on contextualization (Huijgen et al., 2014). These instruments offer more positive indicators for face and content validity compared to the construction of completely new instruments. Recently, we used a multiple-choice historical test (Huijgen, Van de Grift, et al., 2018), but this test did not provide the opportunity to examine the students' answers since it only provided quantitative results.

The test used in this study comprised six open-ended questions regarding different historical topics (see Appendix H) and was used as a pre- and post-test. Based on the work of Hartmann and Hasselhorn (2008) and Huijgen et al. (2014), the first question included a scenario of a young man living in Germany in 1930 who must decide which political party to vote for. The students had to explain whether this young man might have voted for the Nazi Party. The second question asks students to note what else they should know to answer this question successfully. This question aims to provide insights into students' consideration of what they do not know but should know to be able to answer the question. The third question uses a HAT format and displayed two statements about the German 1930s scenario. One statement displayed a present-oriented perspective and the other statement a contextualized perspective. The students had to choose and explain why they chose a particular statement. The fourth question was based on an instrument tested in a previous work (Huijgen et al., 2014) and focused on 19th-century slavery. The HAT format was again used to trigger a verdict about two statements. The fifth question concerns a young woman (Sophie) who reads in her history textbook that until the 1950s, women in the Netherlands lost their jobs when they married. Sophie reacts: "People were stupid in the past." Based on the HAT format, the task for students is to explain if they do or do not agree with Sophie. The sixth and final question used the same lay-out and HAT format of the fifth question but focuses on 16th-century witch hunts.

7.4.7 Data analysis

To check the implementation fidelity of the intervention, all teachers were asked to review each lesson during the intervention and to provide information on whether all lesson activities were successfully completed and whether any irregularities occurred.

No anomalies were noticed by the participating teachers. To examine the progress of the students in historical contextualization, a coding scheme was constructed that provided the opportunity to review the students' answers to the test questions. The coding scheme was based on literature on historical contextualization (e.g., Endacott & Brooks, 2013; Hartmann & Hasselhorn, 2008; Lee & Ashby, 2001). Since our research question focused on the display of presentism and historical contextualization before and after an intervention, we chose to work with these two coding categories (see Table 26). It is possible that an answer received a "presentism" and "contextualized" code when the student, for example, answered that slavery was a phenomenon in the 19th century (using chronological context knowledge) but that it was stupid to not bring people who committed atrocities to trial (present-oriented perspective).

The coding was first independently done by one of the authors, who holds a Master's degree in history education and taught history in a secondary school for 9 years. Next, the coding was reviewed by one of the other authors, who also holds a Master's degree in history education and taught history in a secondary school for over 40 years. Subsequently, all non-corresponding codes (approximately 15%) were discussed until a consensus was reached, resulting in the final coding. First, a frequency analysis (e.g., Krippendorff, 2013) was used to examine the possible gains in historical contextualization in both conditions. Next, a qualitative analysis of the students' answers was done to explore how the students might have improved their historical contextualization skills. The unit of analysis was the entire student answer.

Table 26. Coding scheme with the categories, category descriptions, and examples of students' answers

Category	Category description	Examples of students' answers
Displaying a present-oriented perspective when answering the question	Students display a present-oriented perspective in their answer. They view and judge the past based on their own beliefs, values, and/or knowledge. No consideration of the contextual circumstances at the time of the historical event takes place.	1. People living back then were just stupid. 2. Hannes should not vote for the Nazi Party considering what they have done to the world.
Using historical context knowledge to answer the question	Students use chronological, spatial, socio-political, socio-economic, and/or socio-cultural context knowledge in their answer. They consider the contextual circumstances at the time of the historical event.	1. In the 1950s, it was the law that women in governmental roles would lose their jobs when they married. 2. This source material is set before the beginning of the Second World War; thus, Hannes could not have known the outcome. 3. People in the 16th century thought that witches existed and were afraid of them.

7.5 Results

7.5.1 Presentism and contextualization in students' answers

Based on the coding of the students' answers, Table 27 presents the descriptive statistics (mean scores) of the presence of presentism and contextualization in the students' answers on the pre- and post-test in both conditions. The maximum score for each category is 6.00 when students use historical context knowledge or display presentism in all six questions. The effect sizes (Cohen's *d*) are for the dependent sample t-tests. The effect sizes indicate the standardized difference between the two means and displays this difference in standard deviation units. In both conditions, the students displayed less presentism when the intervention ended, but the students in the experimental condition used far less presentism in the post-test compared to the students in the control condition. The students in the experimental condition also used more contextualization in the post-test than in the pre-test, while students in the control condition showed slightly less use of contextualization in the post-test compared to the pre-test.

Table 27. Descriptive statistics of the presence of presentism and contextualization in the students' answers (maximum score = 6.00)

Condition	*n*	Presentism pre-test (SD)	Presentism post-test (SD)	Cohen's *d*	Sig. (*p*)	Contextualization pre-test (SD)	Contextualization post-test (SD)	Cohen's *d*	Sig. (*p*)
Experimental	96	1.26 (1.04)	.82 (0.87)	0.37	.001	3.96 (1.42)	4.35 (1.31)	0.26	.013
Control	73	1.52 (0.85)	1.41 (0.78)	0.14	.230	3.92 (1.24)	3.67 (1.29)	0.12	.066

A repeated measurement ANOVA was performed to test whether the students in the experimental condition displayed a significantly greater decline in the use of presentism than students in the control condition after the intervention ended. This repeated measurement analysis indicates that the decline of the use of presentism is significantly higher in the experimental condition than in the control condition ($F(1, 167) = 4.17$, $p = .04$, $\eta_p^2 = .02$, which is considered a small effect; Cohen, 1988). This means that the intervention contributed significantly to the decline in presentism in the post-test answers.

Another repeated measurement ANOVA was conducted to examine whether the intervention contributed significantly to the increased difference between the

experimental and control condition in the use of contextualization. This analysis indicates that the gain scores of the students in the experimental condition differ significantly from those of the students in the control group ($F(1,167) = 9.09$, $p = .003$, $\eta_p^2 = .05$, which is considered a small effect; Cohen, 1988). The students in the experimental group show improvement in the use of contextualization, whereas students in the control group show a slight decrease in the use of contextualization. The difference between both groups in the post-test is significant ($t = 3.37$, $p = .001$, $\eta_p^2 = .06$, which is considered a medium effect; Cohen, 1988).

7.5.2 Students' improvement in historical contextualization

The frequency analysis indicates that students in the experimental condition improved more in historical contextualization than students in the control condition. To further explore how they might have improved during the intervention, the students' answers were qualitatively analyzed. Overall, the students in the experimental condition displayed less presentism in the post-test questions compared to their answers in the pre-test questions. The framework used in the experimental condition seemed to teach students to set aside their present-oriented perspectives and to explicitly consider the differences in values, beliefs, and knowledge between the past and the present when answering the test questions. For example, David answered in the pre-test to the fifth question about the position of women in the 1950s: "I agree with Sophie. People were stupid back then. Women have the same rights as men. Some women could even be better than some men." In the post-test, David no longer viewed the past from a present-oriented perspective and provided a historical explanation in his answer: "I do not agree with Sophie. At that specific time [1950s], a woman was viewed as less than a man. When women married, they had to do all of the housekeeping, and the man was the breadwinner." Another example is Emma's answer to the fourth question of the pre-test about slavery: "I chose statement I because Simpson and his family were heavily abused, and this was not acceptable." In the post-test, she included more specific historical circumstances: "I chose statement II because at that time [18th century], it was more normal to keep and trade slaves. Nobody did anything because it was more common in society."

Another interesting finding is that although students such as David and Emma might possess historical context knowledge, this knowledge might be "blocked" by a dominant present-oriented perspective. David and Emma might already have had some knowledge of the historical context since they answered the questions

successfully in the post-test but never received topic knowledge of the test questions during the intervention. However, the collision between the students' current values and beliefs, and the values and beliefs in the past might result in a dominant present-oriented perspective when answering the pre-test question.

The framework might also teach students not to respond immediately but to consider and use historical context knowledge explicitly when answering historical questions and making ethical or moral judgements. For example, Nina answered the fifth question (about the firing of women) in the pre-test as follows:

> I agree with Sophie. Women were inferior to men. Why on earth would rape by the husband be not as bad as rape by somebody you do not know? And why were women allowed to work before they married and not when they married?

In the post-test, Nina explicitly noted that we cannot judge people's actions in the past without considering the historical context: "At that time [1950s], everything was different compared to our daily lives. Laws, opinions, thoughts; almost everything was different. We cannot judge people in the past right away without knowing the specific circumstances."

In the control condition, the decrease in the use of presentism was less recognizable in the post-test answers. For example, Thalia answered Question 5 in the pre-test: "I agree with Sophie because I can understand that women want to work less, but losing your job is really absurd. Rape was also allowed. This cannot be true and should not be allowed." In the post-test, Thalia still viewed the past from a present-oriented perspective when answering the same question: "I agree with Sophie because it is outrageous how people treated women". The answers to the question about 19th-century slavery (Question 4) also illustrated that many students in the control condition continued viewing the past from a present-oriented perspective. For example, Anna answered in the pre-test:

> I chose statement I because the plantation owner was involved in human trafficking, neglect, abuse and murder of people, and he should therefore be brought to trial. I did not choose statement II because you are not allowed to view a human being as a non-living creature.

In the post-test, she answered, "I chose statement I because he [the plantation owner] killed somebody. I did not choose statement II because one should live in freedom and slavery is illegal." The students in the control group, such as Thalia and Anna, had difficulty setting aside their personal emotions, values, and beliefs. Whereas students in the experimental condition (such as David and Emma) shifted from a present-oriented perspective towards a historical contextualized perspective, the students in the control condition (such as Thalia and Anna) continued viewing the past with their current values, beliefs and knowledge. This indicates that presentism might remain the dominant perspective when not teaching students explicitly to perform historical contextualization (e.g., by providing guiding questions and creating opportunities to practice historical contextualization).

A distinction can be made between chronological, spatial, socio-political, socio-economic, and socio-cultural context knowledge. The framework used in the experimental condition aimed to encourage the use of these frames of reference to reconstruct a historical context by providing guiding questions (see Appendix G). Spatial context knowledge was rarely used in the answers of the pre- and post-test questions (in total, 12 explicit references to spatial context knowledge). Moreover, we did not notice a large progression in the use of socio-political, socio-economic, and socio-cultural knowledge in the experimental condition. Most progression was found in the use of chronological knowledge. The students in the experimental condition used far more chronological knowledge in the post-test compared to the pre-test to answer the test questions. This knowledge mostly occurred in the form of sequencing historical events and mentioning a specific year or time period, often resulting in a shift from a present-oriented perspective towards a more historical and contextualized one. For example, Robert provided a present-oriented perspective when answering the fourth question in the pre-test: "I chose statement I because what the plantation owner did is not allowed, and therefore, he should be punished […]." In the post-test, Robert considered the chronological context of the 19th century more and proceeded to answer as follows:

> I chose statement II because at that specific time it was more normal to treat your slaves that way. I did not choose statement I because the law did not view slavery as something bad, so he [the plantation owner] could not be arrested for this.

The use of chronological knowledge was also more implicitly applied after the intervention ended. For example, Kathy displayed a present-oriented perspective (i.e., knowledge that was not yet available for people living in 1930) when she answered the third pre-test question about Hannes, who is inclined not to vote for the Nazi Party: "I chose statement I because nobody can justify what the Nazi Party has done to the world. I did not choose statement II because I do not think that Hannes would view Hitler as a strong leader." However, Kathy considered in her post-test answer that Hannes was living in 1930 and did not yet know the outcome of the Nazi Party's rise to power:

> I chose statement II because Hannes wants to keep his job and does not want the company to go bankrupt. I did not choose statement I because Hannes did not know then what the Nazi Party would do to the world. He is focused on helping out in his father's business.

In the control condition, no such increase was noticeable in the use of chronological knowledge. For example, whereas students in the experimental condition explicitly used chronological knowledge to answer Question 5 (e.g., recognizing that in the 1950s different laws, values, and beliefs applied), the students in the control condition used chronological indicators less often in their answers. The guiding questions focusing on reconstructing the chronological context (see Appendix G) might therefore teach students that historical events occurred in a different chronological context (compared to their contemporary context) and should therefore be examined in that particular context.

7.6 Conclusions and discussion

The aim of this study was to examine the effects of a lesson unit based on a three-stage historical contextualization framework on 14- to 16-year-old students' ability to perform historical contextualization. Repeated measures analyses of variance showed that students in the experimental condition displayed significantly less presentism and used more historical context knowledge in their answers compared to students in the control condition after the intervention ended. Further analysis of the students' answers indicated how students might have improved their historical contextualization skills during the intervention.

The framework seemed to teach students in the experimental condition to set aside their present-oriented perspectives and to explicitly consider the differences in values, beliefs, and knowledge between the past and the present when answering the test questions. Wineburg (2001) already noted that historical thinking is an unnatural act and that it should be taught to students in history classrooms. Our findings seem to illustrate this. In line with research into other historical reasoning competencies (e.g., Nokes, Dole, & Hacker, 2007; Reisman, 2012b; Stoel, Van Drie, & Van Boxtel, 2017), students need to explicitly be taught how to engage in historical reasoning. If students are not taught explicit historical contextualization, they might keep viewing the past from a present-oriented perspective, which often results in misunderstanding historical events and agents' actions (Barton & Levstik, 2004; Lévesque, 2008).

The framework also might "unblock" students' historical context knowledge since some students might already possess this knowledge, which they used in the post-test questions, but did not use this because of a dominant present-oriented perspective when answering the pre-test questions. Further research is needed to examine the unblocking function of the framework and which types of context knowledge students use and do not use when performing historical contextualization. Thinking aloud protocols (e.g., Van Someren, Barnard, & Sandberg, 1994) can be used to examine students' reasoning in more detail when answering historical questions. Moreover, the framework might have taught students to become aware of how they can approach a historical question or assignment. The students in the experimental condition not only displayed a less present-oriented perspective and applied more historical contextualization in their post-test answers but also explicitly mentioned, for example, that a moral judgement cannot be made without considering the historical context. Seixas and Morton (2013) consider this a "demonstration of powerful understanding" (p. 189) which contributes to thinking historically.

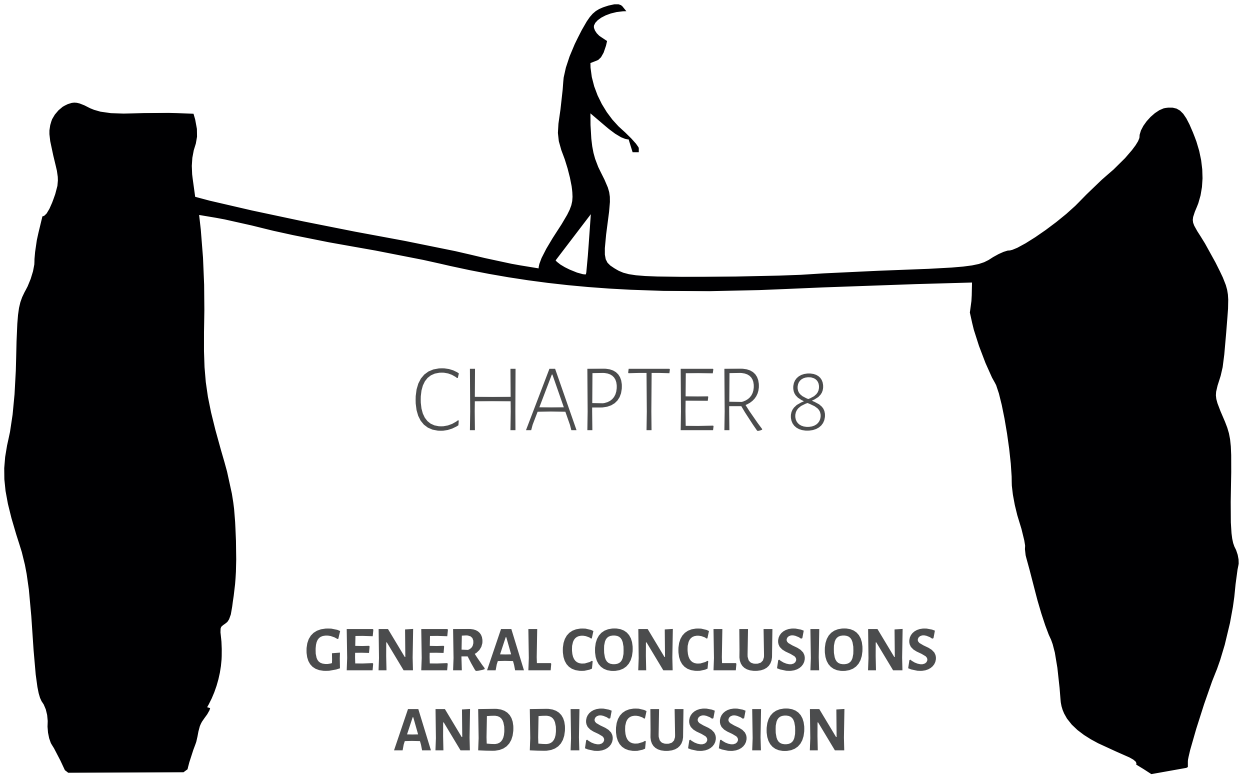
Students in the experimental condition also used more historical context knowledge after the intervention. In particular, they used more chronological knowledge, such as considering the specific time period or sequencing historical events to answer the post-test questions. This often resulted in a shift from a present-oriented perspective towards a historically contextualized perspective. Dawson (2009) argues that chronological knowledge could contribute to a "sense of a period." By explicitly considering knowledge of characteristics of a particular historical period, students might become aware that there are differences in values, beliefs, and knowledge in

different time periods. We noticed no extensive use of spatial, socio-political, socio-economic, or socio-cultural knowledge in the students' answers. When students used these types of knowledge, it was often less concrete (e.g., "they did things differently back then"). A possible reason might be that the students did not receive topic knowledge about the test questions. Future research should therefore also include test questions related to the historical topic of the lesson unit to provide more insight into the role of historical context knowledge in historical contextualization processes.

Beyond the scope of this article, but nonetheless important, is the role of the history teacher. How did they experience teaching with the intervention? What did they learn? The teachers in our sample participated voluntarily, but how do other teachers react to using the framework in their history lessons? Future research should therefore include the teachers' role, for example, by conducting teachers' beliefs interviews (e.g., Luft & Roehrig, 2007) before and after the intervention. Future research should also focus on the historical topics of the test questions. Do questions about recent topics (e.g., women's rights in the 1950s) trigger more present-oriented perspectives compared to more distant historical topics, such as the witch hunts in the 16th century? If this is the case, teachers should practice historical contextualization with distant historical topics and more recent historical topics.

Our study has several limitations. We used a quasi-experimental research design with non-probability sampling, while random assignment should be preferred (Cook & Payne, 2002). Moreover, the research was conducted with seven history teachers and 169 students. The seven history teachers also volunteered to participate in the study and they might be more motivated to improve their teaching compared to other history teachers. Follow-up studies using more teachers and students with different backgrounds and interventions focusing on different historical topics are needed to confirm the findings and to further examine possible gains in the students' ability to perform historical contextualization. Furthermore, only six items were used as a historical contextualization test, and despite the fact that the test format was derived from previous work, more items and information on the validity and reliability of these items are needed. The development of a parallel test could help overcome the carryover effect (Bose & Dey, 2009). The interesting Historical Thinking Competencies in History project of Trautwein et al. (2017), where items are developed that can measure historical reasoning competencies, might provide effective formats for the development of new test items.

Despite these limitations, this study shows promising indicators for using the three-stage framework to develop lesson activities that encourage historical contextualization. The study presents a concrete example for teachers who wish to teach students historical contextualization in their lessons. The framework can easily be implemented in one history lesson or can be taught in multiple lessons. Moreover, teachers can use one or more items from the historical contextualization test to gain insight into students' progress in historical contextualization and to teach them how to improve in historical contextualization as a form of formative assessment (e.g., Heritage, 2010). The question format for choosing a statement can easily be used to collect information about the students' ability to perform historical contextualization among a large and heterogenous group of students. The study offers positive indicators for the ways in which historical contextualization can be encouraged in classrooms. The success of the intervention is best illustrated by Lisa, the 14-year-old girl who displayed a present-oriented perspective on the firing of women when they married in the 1950s. She answered the same question in the post-test: "I can explain why it happened. I do not agree with it, but I know that in the 1950s, there were different beliefs and values compared to our contemporary society."



CHAPTER 8

GENERAL CONCLUSIONS AND DISCUSSION

This chapter first presents the general conclusions. Subsequently, limitations and suggestions for future research are discussed. Next, the scientific contribution of this thesis and the practical implications for teaching and learning historical contextualization are described.

8.1 Summary of the main findings

"Why on earth would you burn those women alive? It is obvious that witches do not exist, do they? That is just cruel. People in the past were indeed stupid." This statement is a reaction from a student when the witch hunts in the Early Modern Period were the topic under study. This student viewed the witch hunts from a present-oriented perspective in which the past is examined with present beliefs, values, and knowledge. This process often results in misunderstanding historical phenomena and historical agents' actions. Historical contextualization, the main focus of this thesis, can help prevent present-oriented perspectives among students when they examine the past (Wineburg, 2001). It is therefore considered a key component of historical thinking and reasoning (Lévesque, 2008; Van Drie & Van Boxtel, 2008).

Historical contextualization is the ability to situate phenomena and people's actions in a historical context in order to give meaning to these phenomena and actions (Van Boxtel & Van Drie, 2012). It is not the same as relativism or the justification of controversial historical events and agents' actions. Instead, historical contextualization should lead to reasoned ethical judgments about past events and help students to explain historical phenomena and agents' actions (Seixas & Morton, 2013). It is not wrong for students to think that, for example, slavery or child labor is awful. However, to explain and interpret such historical phenomena successfully, students need to become aware that the past differs from the present, that people in the past held different beliefs and values and that these people might not have possessed the same knowledge as the students themselves.

Teaching historical contextualization in this thesis was conceptualized as four interrelated components: (1) reconstructing a historical context, (2) raising awareness of present-oriented perspectives, (3) enhancing historical empathy, and (4) creating opportunities for students to practice historical contextualization to enable historical reasoning. All components of this theoretical framework for teaching historical contextualization should occur in interactions between teachers and students. The framework is visualized in Figure 1 in the first chapter of this thesis.

Despite the importance of historical contextualization in history education, there is a lack of (1) standardized instruments that measure students' ability to perform historical contextualization, (2) observation instruments that provide insight into

how teachers promote historical contextualization, and (3) classroom materials that promote historical contextualization. To examine these problems, we formulated the following main research question for this thesis: *How can students' ability to perform historical contextualization be promoted?* To answer the main research question, a total of five research questions were formulated:

1. How can we measure elementary and secondary school students' ability to contextualize historical agents' actions?
2. How successfully can secondary school students contextualize historical agents' actions?
3. What instrument can be used to observe how history teachers promote historical contextualization in classrooms?
4. How do history teachers promote historical contextualization in their classrooms?
5. What are the effects of a lesson unit designed to promote secondary school students' ability to perform historical contextualization?

The first two research questions focused on students' ability to contextualize historical agent's actions (historical perspective taking). The framework for teaching historical contextualization was used in these two studies to design and test instruments and to analyze how students performed historical contextualization. The third and fourth research questions examined teachers' instructions with regard to historical contextualization. In these two studies, the framework was used to develop an observation-instrument to explore how teachers promote historical contextualization in classrooms. To answer the final research question, the framework was used to design classroom materials. The effects of these materials on student's ability to perform historical contextualization were tested in two intervention studies.

8.1.1 Students' ability to perform historical contextualization

The first study (Chapter 2) focused on developing and testing two instruments that measure students' ability to perform historical contextualization and how students at different ages and educational levels perform historical contextualization. In the second study (Chapter 3), we used the most reliable instrument of the first study to examine students' reasoning when asked to perform the task of the instrument.

Study 1: How can we measure elementary and secondary school students' ability to contextualize historical agents' actions?

To answer this question, we examined two instruments. One instrument was adapted from Hartmann and Hasselhorn (2008) and focused on the rise of the Nazi Party in Germany in the 1930s. Using the same format, we developed a second instrument that focused on 19th-century slavery. Both instruments were tested for validity and reliability among 1,270 Dutch upper elementary and secondary school students, ranging in age from 10 to 17 years. Principal component analysis (PCA) and Cronbach's alpha coefficients showed that the Nazi Party instrument was the more valid and reliable instrument. Constructing a new instrument using the same format as the Nazi Party instrument proved to be a difficult task since we found low reliability scores for the slavery instrument. To explore how students of different ages and educational levels performed historical contextualization, we analyzed the data that we obtained from the Nazi Party instrument. The results indicated that upper elementary school students successfully performed some historical contextualization efforts starting at approximately the age of ten. However, older students achieved higher scores for historical contextualization than younger students. When examining differences between educational levels, we found that pre-university students achieved the highest historical contextualization scores compared to those of students in senior general secondary education and elementary education groups. The amount of prior topic knowledge might be an explanation for these differences since we found a correlation (.27) between students' topic knowledge and their contextualization scores.

Study 2: How successfully can secondary school students contextualize historical agents' actions?

In this study, we assessed with the Nazi Party instrument a sample of 15- and 16-year-old secondary school students ($n = 143$) to determine their ability to contextualize the actions of people in the past. Subsequently, we used the thinking-aloud methodology to explore the reasoning of 36 students to uncover their contextualization process. The results of this study showed that only seven of the 143 students (4.9%) had a mean historical contextualization score < 2.50 out of a maximum 4.00 score. Most students (57.3%) achieved a mean score ≥ 3.00 and < 3.50 and 22.4% obtained mean scores ≥ 3.50. The verbal protocols of the 36 students were analyzed for the use of present-oriented perspectives, historical empathy, and historical context knowledge. These analyses indicated that five students did not realize that people in the past did not

have the same knowledge as we have today. One student viewed the protagonist in the task as naïve. The other students were aware of the consequences of their present-oriented perspective when performing the historical contextualization task. Students used affective forms of historical empathy (e.g., love for your parents) to solve the task but they did not consider explicitly the role of the historical agent in society. Moreover, the students used chronological knowledge and knowledge of socio-economic, socio-political, and socio-cultural circumstances. Spatial knowledge (e.g., the geographical size of the Weimar Republic and the German Empire) was not explicitly considered. Students who obtained mean scores ≥ 3.00 seemed to use more historical context knowledge in their reasoning than students with mean scores < 3.00. These findings and the small correlation (.19) between students' topic knowledge and their contextualization scores indicate that historical content knowledge is necessary to successfully perform historical contextualization.

8.1.2 Teachers' instructions with regard to contextualization

To examine how history teachers might promote historical contextualization, an observation instrument was developed and tested in the third study (Chapter 4). This instrument was used in the fourth study (Chapter 5) to explore how eight history teachers promoted historical contextualization in their classrooms.

Study 3: What instrument can be used to observe how history teachers promote historical contextualization in classrooms?

This study presented the development of the Framework for Analyzing the Teaching of Historical Contextualization (FAT-HC). This is a high-inference observation instrument that focuses on history teachers' competency in promoting historical contextualization in classrooms. The instrument can be used to provide discipline-specific feedback to teachers. Using expert panels, positive indicators of the instrument's content validity were found. Furthermore, generalizability theory analysis (e.g., Brennan, 2001; Hill, Charalambous, & Kraft, 2012) provided indicators that the instrument is one-dimensional when used to evaluate how history teachers promote historical contextualization. Generalizability theory analysis also showed that a large proportion of the instrument's variance is explained by the differences between the observed teachers and a small proportion of the variance is explained by the differences in lessons and observers, which are conclusions that provide positive indicators for the instrument's reliability. To construct a reliable scoring design, a decision study (D-study) was conducted. A scoring design with one observer

evaluating two lessons seems to be the most effective method for research purposes. For formative teacher evaluations, a reliable scoring design in which two observers each evaluate two lessons or three observers each evaluate one lesson is most effective.

Study 4: How do history teachers promote historical contextualization in their classrooms?
The FAT-HC was used in this study by five trained raters to observe eight history teachers twice. The FAT-HC items were divided into eight categories while distinguishing between items that demonstrate historical contextualization and items that focus on engaging students in historical contextualization processes. The results of this study indicated that the teachers in the sample did not explicitly promote historical contextualization in the lessons. No teacher obtained a mean FAT-HC score > 2.00 on a 4-point scale. The teachers mainly demonstrated historical contextualization, while engaging students in historical contextualization processes was observed far less often. The highest scores could be found in the categories that focused on avoiding anachronisms and presenting the past as progress (mean score = 3.59) and the category that focused on reconstructing the historical context (mean score = 2.18). All other categories obtained mean scores < 2.00, with the category that focused on promoting students to use historical empathy (mean score = 1.25) and the category that focused on promoting students to use the historical context to explain the past (mean score = 1.24) displaying the lowest scores.

8.1.3 Promoting historical contextualization with classroom materials

The fourth study indicated that history teachers often did not engage students in historical contextualization. To help teachers, the final two studies (Chapters 6 and 7) explored the effectiveness of various practical tools to improve students' ability to perform historical contextualization.

Studies 5 and 6: What are the effects of a lesson unit designed to promote secondary school students' ability to perform historical contextualization?
In the fifth study the framework for teaching historical contextualization was used to formulate four design principles: (1) raising students' awareness of the consequences of a present-oriented perspective when examining the past, (2) enhancing the reconstruction of a historical context, (3) enhancing the use of a historical context to explain historical phenomena, and (4) enhancing historical empathy. These principles were used to develop a lesson unit focusing on the 17th and 18th centuries, which was

tested through a quasi-experimental pre-test–post-test design with an experimental and a control condition (N=131). The mean students' age was 16 years in both conditions. To raise awareness of possible present-oriented perspectives, a historical case was presented at the start of the lesson. This case relates to a historical phenomenon that students find difficult to explain due to present-oriented perspectives (e.g., the exchange of the colony of New Netherland for Suriname). In the next classroom activity, the historical context of the case was reconstructed using guiding questions. This exercise was followed by a classroom activity wherein the students were asked to use their newly acquired historical context knowledge to interpret the historical case again. Finally, the students conducted historical empathy tasks focusing on a relevant historical agent. A multiple-choice historical contextualization test was constructed to compare students' ability before and after the intervention. The analyses of pre- and post-test scores indicated that in the experimental condition, students' ability to perform historical contextualization significantly increased compared to that of students in the control condition. Moreover, the teachers who taught the intervention experienced that the practical implementation of the different design principles promoted historical contextualization.

In the sixth study, the framework for teaching historical contextualization was used to develop a three-stage historical contextualization framework: (1) raising awareness of possible present-oriented perspectives by presenting a historical case that students find difficult to explain, (2) reconstructing a historical context for the particular case, and (3) interpreting the historical case again with students' newly acquired historical context knowledge. This framework was used to develop a lesson unit on Cold War events. In contrast to the lesson unit of the fifth study, this lesson unit consisted of four lessons and contained a structure that varied from more basic instructions (e.g., teachers demonstrate historical contextualization) to more complex instructions (e.g., students perform historical contextualization on their own). The lesson unit on Cold War events was tested for its success in improving students' ability to perform historical contextualization through a quasi-experimental pre-test–post-test design with an experimental and a control condition (N = 169). The mean students' age was 15 years in both conditions. In contrast to the fifth study, this study measured students' ability to perform historical contextualization with six open-ended questions instead of multiple-choice items.

The analyses of pre- and post-test answers showed that students in the experimental condition used less *presentism* and more historical context knowledge in their post-test answers compared to the students in the control condition. Examples of students' answers suggests that the framework might teach students to set aside their present-oriented perspectives, unblock historical context knowledge, and become aware of the need to consider historical context knowledge to understand historical phenomena and agents' actions.

8.1.4 An answer to the main research question

The main research question of this thesis was: *How can students' ability to perform historical contextualization be promoted?* Students can increase their ability to perform historical contextualization. Teachers should therefore explicitly engage students in historical contextualization processes and should not only demonstrate historical contextualization themselves. The four design principles (awareness of a present-oriented perspective, the reconstruction of a historical context, the use of the historical context to explain historical phenomena, and historical empathy) can help teachers to develop classroom materials that engage students in historical contextualization processes.

The three-stage framework provides a promising lesson structure that can be used by teachers to promote historical contextualization among their students. Within this framework, the teacher first presents a case about a historical phenomenon that evokes a cognitive conflict among students. Next, the teacher instructs the students to reconstruct a historical context of the case and asks them afterwards to evaluate the case again. A structure that varies from basic instructions to more complex instructions is preferred. For example, the three stages can first be used to learn students how to perform historical contextualization (e.g., discussing step-by-step plans for reconstructing a historical context). Subsequently, the stages can be used to create opportunities for students to perform historical contextualization on their own to explain and interpret historical phenomena.

8.2 Limitations and suggestions for future research

This thesis contains a number of limitations. Three limitations that also provide suggestions for future research are described in more detail. The first important

limitation is that the first two studies focused on contextualizing historical agents' actions. In these two studies, we did not focus on other important historical reasoning competencies, such as the use of historical contextualization to interpret historical sources or to determine continuity and change in history. Future research should therefore develop instruments that test students' ability to use historical contextualization to perform these other historical reasoning competencies.

A second important limitation focuses on the instruments developed and used in this thesis. The Nazi Party instrument (used in the first two studies) showed positive indicators for reliability and validity. However, the instruments' scenario is fictional, focuses on one historical topic, and contains only a textual story. Developing instruments using different historical topics and sources is necessary to provide more insights into students' ability to perform historical contextualization when explaining historical events or agents' actions. Moreover, future research should examine how the internal consistency of the instrument can be increased since this was on the lower side of what is considered acceptable. The function of the role of the historical agent (ROA) items also remains unclear. As suggested in the second study, when researchers or teachers want to include ROA items in an instrument, they can split these items into two categories: (1) items that might trigger more affective processes of historical empathy (i.e., using recognizable or universal emotions) and (2) items that might trigger more cognitive processes (i.e., considering the role of the historical agent) of historical empathy. In the fifth study, two historical contextualization tasks were used, both containing multiple-choice items. The internal consistencies of both instruments were on the lower side of what is considered acceptable. Further, multiple-choice questions do not display students' reasoning as well as open-ended questions. In the sixth study, a historical contextualization test was therefore used based on the instruments tested in the first study and on the History Assessments of Thinking (HATs) of the Stanford History Education Group. However, this six-item test needs more validation by, for example, using expert panels and conducting thinking-aloud protocols with students of different ages, educational levels, and backgrounds.

The Framework for Analyzing the Teaching of Historical Contextualization (FAT-HC), which was developed in the third study and used in the fourth study, needs more examination to increase the reliability because nearly 35% of the variance (residual) could not be explained by teacher, observer, or lesson variance. Moreover, the number of items in the instrument (40 items) should be decreased because observers indicated

that 40 items are too many to score during one lesson. A larger generalizability study, including a decision study that focuses on how many items are necessary to achieve reliability, could provide these insights. Furthermore, the FAT-HC was only tested when observing videotaped lessons and not when observing live lessons. The Rasch methodology could also provide more information on the instrument's reliability and item difficulty as perceived by the history teachers (Fischer & Molenaar, 1995). Combining the use of the FAT-HC with other methods, such as teachers' self-reports and students' questionnaires, could also provide more insight into how historical contextualization is promoted in classrooms. When a reliable and valid observation system is established, research can focus on differences between teachers (e.g., novices versus experts; history teachers trained in contextualization versus non-trained history teachers). A reliable instrument also opens the way for cross-sectional research, longitudinal studies, and international comparisons.

The samples of teachers, students, and lessons used in the different studies are a third important limitation. Most studies were explorative in nature. Therefore, more teachers, students, and lessons should be included to confirm the findings of the different studies. What differences occur in reasoning between younger and older students when asked to create historical contexts? Are students less engaged in historical contextualization processes during history lessons? Which individual design principle best promotes historical contextualization?

8.3 Three directions for future research

Three possible directions for future research can be formulated based on the findings and limitations of this thesis. First, the field of history education research needs more valid and reliable instruments that can examine students' ability to perform historical reasoning competencies in more detail. As shown in this thesis, the development of such instruments is a complex process. It is a hopeful development that different studies have recently been published on this topic (Ercikan & Seixas, 2015; Reich, 2009; VanSledright, 2013). In addition, various initiatives have been launched such as Beyond the Bubble project of the Stanford History Education Group and the Historical Thinking Competencies in History project of Trautwein et al. (2017). International collaboration is needed to further develop and test such instruments in different contexts. Other important questions that also need to be answered are: how can

researchers and teacher educators help teachers to implement such instruments in their daily practice, and how can teachers help researchers and teacher educators to further improve these instruments?

Second, the observation instrument (FAT-HC) indicates that historical reasoning competencies can be observed in a practical manner. However, the instrument only focuses on historical contextualization. Developing instruments focusing on other historical reasoning competencies (such as using historical sources or determining causes and consequences) would provide more insights into how historical reasoning competencies are promoted in classrooms. The historical reasoning and thinking frameworks of Seixas and Morton (2013) and Van Drie and Van Boxtel (2008) and research on history teaching practices (e.g., Fogo, 2014) could provide a common core for constructing such domain-specific observation instruments. The PATH instrument of Van Hover, Hicks, and Cotton (2012) is a good example but needs further elaboration and more alternatives. The Teaching Historical Thinking and Reasoning instrument (Gestsdóttir, Van Boxtel, & Van Drie, submitted) is a promising effort.

In the third place, (quasi-)experimental studies are scarce within the field of history education research. More attention has recently been paid to conducting such studies (e.g., Reisman, 2012b; Stoel, Van Drie, & Van Boxtel, 2017) but this research method is still underexposed compared to the use of case studies and other qualitative methods. Intervention studies, where teachers and researchers work collaboratively to design pedagogies and practical tools, are highly needed to gain insight into what works within history education. These pedagogies should not be a "one solution for all strategy" but rather build upon each teacher's individual characteristics and work context (Korthagen, 2017).

8.4 Scientific contribution

There is a need to develop and test new assessment formats to make sense of how students learn history and how they improve in it (Ercikan & Seixas, 2015; Smith, 2018; VanSledright, 2013). The first contribution of this thesis is therefore that several instruments to assess students' ability to perform historical contextualization were constructed and tested. The first study showed, for example, that the Nazi Party instrument of Hartmann and Hasselhorn (2008) could be transferred and used in

a different country. However, the study also showed the complexity of constructing a new valid and reliable instrument when using the format of Hartmann and Hasselhorn (2008). Moreover, the second study, using thinking-aloud protocols, indicated that it is very important to formulate unambiguous items to construct valid and reliable instruments. The fifth study showed that it might be possible to measure the ability to perform historical contextualization with multiple-choice items. These items, however, measure historical contextualization at a very basic level; do students view the past from a present-oriented perspective or not? The sixth study tried to measure historical contextualization with open-ended questions. This approach seems promising, but more insight into the items' reliability and validity is needed.

Despite the fact that some empirical research has been conducted on how students contextualize historical phenomena and agents' actions (e.g., Berti, Baldin, Toneatti, 2009; Lee & Ashby, 2001), large-scale research studies on how students of different ages and educational levels perform historical contextualization are scarce in history education research. A second contribution of this thesis is that the first study provided more insight into these differences. Especially in the Dutch educational context, not much had been known about how upper elementary, lower secondary, and upper secondary students performed historical contextualization. This thesis also argues that we need to adopt a balanced attitude to statements that students view the past from a present-oriented perspective (cf. Lévesque, 2008; Seixas & Morton, 2013; Wineburg, 2001). Some young students in the first study performed elements of historical contextualization successfully, and most students included in the second study succeeded in contextualizing the actions of the historical agent.

A third contribution is that this thesis seems to confirm the views of scholars such as Reisman and Wineburg (2008) and Van Boxtel and Van Drie (2012) who argue that historical content knowledge is necessary to perform historical contextualization successfully. For example, the first and second studies displayed significant correlations between students' historical content knowledge (prior topic knowledge) and their instruments' scores. In addition to previous research, the results of the second study indicate that combining different types of knowledge (i.e., frames of reference) might improve the ability to perform historical contextualization and that students do not consider spatial knowledge and knowledge about the role of historical agents in particular. The thesis also argues (in the sixth study) that students might possess historical context knowledge on a particular topic but that this knowledge

might be "blocked" by their present-oriented perspectives. Students need therefore to be alerted explicitly to possible present-oriented perspectives when examining the past. Moreover, we agree with, for example, Endacott and Brooks (2013) that not only a cognitive form but also an affective form of historical empathy might promote historical contextualization since students in the second study used, for instance, recognizable emotions to explain historical agents' actions.

A fourth contribution of this thesis is the focus on teachers' instructions. Various scholars have argued that teachers might struggle with teaching historical reasoning competencies (e.g., Reisman, 2015; Saye & Social Studies Inquiry Research Collaborative, 2013), but observation studies on this topic have been scarce, resulting in a lack of insight into history teachers' instructions and specific professionalization needs (Van Hover, Hicks, & Cotton, 2012). This thesis included a domain-specific observation study whose results indicate that the teachers included in our sample did not engage students much in historical contextualization processes.

A final contribution is that a theoretical framework for teaching historical contextualization has been missing. By integrating previous research on historical contextualization, we developed a framework containing four interrelated components (see also Figure 1 in the first chapter), which can be used by researchers and teachers to design and test practical tools to promote students' ability to perform historical contextualization. The fifth and sixth studies illustrated that this framework could very well be used to develop such tools. This contribution is especially important since there might be a lack of practical and effective tools to promote students' historical reasoning competencies (e.g., Fogo, 2014; Grant & Gradwell, 2010; Reisman & Fogo, 2016).

8.5 Practical implications

What are the practical implications for the field of history education? First, the developed measure instruments, such as the Nazi Party instrument, could easily be implemented by teachers in classrooms to test students' ability to perform historical contextualization. Which students view the past from a present-oriented perspective, and who performs historical contextualization successfully? If teachers have this information, they can adapt their teaching to the specific needs of individual students.

The first study indicated, for example, that young students especially viewed the past from a present-oriented perspective. Historical contextualization should therefore also be explicitly taught in the first years of secondary education (lower secondary education) and not only in upper secondary education.

A second practical implication is the developed Framework for Analyzing the Teaching of Historical Contextualization (FAT-HC) that can be used as a tool to improve history instruction. History teachers and teacher educators could, for example, observe history lessons using the instrument and discuss their findings to further improve students' ability to perform historical contextualization. Moreover, the FAT-HC could function as a framework to collaboratively design new lessons that explicitly promote historical contextualization. This observation instrument was not designed to assess history teachers.

A third practical implication are the four design principles that can be used by teachers to design classroom materials that promote students' ability to perform historical contextualization. The fifth and sixth study transfer the design principles to practical examples for history teachers. The three-stage framework (presenting a historical case, reconstruct a historical context, and evaluate the case again) of the sixth study is very flexible because teachers can implement the framework at the start of the lesson or during the entire lesson or spread the stages over multiple lessons. To inspire teachers, we have included in Appendix I different examples of cases that can promote a cognitive conflict. Teachers can use these examples to develop lessons according to the three-stage framework.

The thesis showed, as a fourth practical implication, that students might profit when they are provided with explicit instructions to reconstruct a historical context. The different frames of reference (chronological, spatial, political, economic, and cultural) can be used by teachers as instructions and function as a checklist for students when creating a historical context. The guiding questions for reconstructing a historical context that are used in the fifth and sixth study can be used as examples.

A final practical implementation is related to the teaching of sensitive topics. A sensitive topic is, for example, the teaching of the Dutch colonial past (e.g., Savenije, Van Boxtel, & Grever, 2014). Currently, a public debate is ongoing about the removal of statues of historical agents, such as Jan Pieterszoon Coen and Michiel de Ruyter,

who both played important roles in the Dutch colonial past. Teachers might not know how to teach such topics to students. The three-stage framework might help them to design lessons. The framework, for example, can be used to develop a case about Jan Pieterszoon Coen. What should be done with his statue in the city of Hoorn? This question can be discussed after a short description of the acts of Jan Pieterszoon Coen. In the following lesson activity, the students can examine the different frames of reference to reconstruct a historical context. Finally, the students can provide a well-considered answer to the question based on a created historical context.

8.6 A call to conduct domain-specific research

The *Golden Coach* of the Dutch royal family shows a painting (in Dutch: *Hulde der Koloniën*) in which slaves offer some of their belongings to the royal family. In one of my lessons, I asked my eighth-grade students a provoking question: should this coach be banned? Look at the painting. Is it okay to glorify slavery in the past of the Netherlands? The majority of my students found that the coach should be banned and that the royal family should be ashamed of themselves for riding in such a coach. Next, I reconstructed the historical context of the coach with my students. When was it made? What was occurring in Europe and in the world at that time? Who made the painting? Afterwards, I asked my students the same question: should the coach be banned? Now, the students were far subtler in their reactions. They considered that the coach was built in the year 1898, when the possession of colonies was common among many European countries, and that there were beliefs and values different from present-day attitudes. One student noted that the coach should not be banned but rather be used to raise people's awareness of the differences between the past and the present. I had a great day.

How my teaching improved by conducting this PhD research is beyond the scope of this thesis, but I believe that my students are more engaged in history and that they are better equipped for participating in subsequent education and in our society. This research also taught me how to train student history teachers more effectively in teaching historical contextualization. This is therefore a call for teachers and teacher educators to conduct domain-specific research.

CHAPTER 9

SAMENVATTING (DUTCH)

9.1 Inleiding

Wanneer ik mijn leerlingen vroeg de winst van de NSDAP bij de verkiezingen in 1933 te verklaren, antwoordden veel van hen "dat die mensen wel gek moesten zijn om te stemmen op Hitler". De christenvervolgingen in het Romeinse Rijk vonden zij belachelijk, want "je bent toch vrij om je eigen geloof te kiezen?" Helemaal idioot vonden zij de ruil tussen Nieuw-Amsterdam (het huidige New York) en Suriname in de zeventiende eeuw. In plaats van het zoeken naar historische verklaringen voor deze gebeurtenis, ging het vooral over hoe "gaaf" het was wanneer Nederland eigenaar zou zijn van de New York Knicks en van Wall Street. Mijn leerlingen keken vanuit hun eigen waarden, normen en kennis naar het verleden. Dit noemen we *presentisme* en heeft vaak als gevolg dat leerlingen niet succesvol kunnen verklaren waardoor historische gebeurtenissen ontstonden (Wineburg, 2001).

De vaardigheid *historisch contextualiseren* helpt leerlingen bij het voorkomen van presentisme en bij het verklaren van historische gebeurtenissen (Barton & Levstik, 2004; Lévesque, 2008; Seixas & Morton, 2013). Contextualiseren is namelijk het onderzoeken van verschijnselen vanuit een gecreëerde en relevante historische context (Van Drie & Van Boxtel, 2008). Verschijnselen worden in dit proefschrift gedefinieerd als gebeurtenissen of handelingen van mensen in het verleden. Contextualiseren voorkomt dat leerlingen vanuit hun eigen perspectief naar verschijnselen kijken doordat zij rekening leren houden met de historische omstandigheden ten tijde van een verschijnsel en met de mogelijke verschillen tussen mensen in waarden, normen en kennis. De opkomst van de NSDAP kan bijvoorbeeld alleen worden verklaard door rekening te houden met de toenmalige slechte economische omstandigheden en dat mensen toen nog niet wisten tot welke gruwelen het naziregime allemaal in staat was.

Dit proefschrift richt zich op de vraag hoe bij leerlingen de historische vaardigheid contextualiseren kan worden verbeterd. Om deze vraag te beantwoorden, worden drie problemen onderzocht die tot nu toe onderbelicht zijn gebleven in het vakdidactisch geschiedenisonderzoek. Het eerste probleem is dat er weinig betrouwbare instrumenten zijn waarmee kan worden bepaald welk niveau van contextualiseren leerlingen hebben bereikt. Hierdoor is het lastig om de ontwikkeling van deze vaardigheid bij leerlingen te monitoren. Het tweede probleem is dat er weinig instrumenten zijn die inzicht geven in welke mate en op welke wijze leraren aandacht

besteden aan contextualiseren. Hierdoor is er niet veel bekend over de problemen die leraren ervaren bij het onderwijzen van contextualiseren. Het derde probleem is dat er weinig empirisch gefundeerd lesmateriaal is dat het niveau van contextualiseren van leerlingen bevordert. Dit komt doordat interventie-studies schaars zijn binnen het vakdidactisch geschiedenisonderzoek.

9.2 Theoretische achtergronden

In het proefschrift staat contextualiseren centraal. Daarom volgt eerst een conceptualisering van deze vaardigheid en wordt het belang van contextualiseren voor het geschiedenisonderwijs toegelicht. Deze sectie eindigt met de beschrijving van een didactisch model voor het onderwijzen van contextualiseren dat wordt gebruikt in de verschillende deelstudies van dit proefschrift. Een visuele weergave van dit model staat in Figuur 1 in het eerste hoofdstuk van dit proefschrift.

9.2.1 De historische vaardigheid contextualiseren

Wat wordt in dit proefschrift verstaan onder contextualiseren? Sommige onderzoekers zien contextualiseren alleen als een vaardigheid om historische bronnen te kunnen interpreteren (o.a. Baron, 2016; Reisman, 2012b; Wineburg, 1998). Contextualiseren draait dan om vragen zoals: wat gebeurde er op het moment dat de bron werd gemaakt? Hoe hebben deze omstandigheden de bron beïnvloed? Het is echter ook mogelijk om verschijnselen te contextualiseren (Havekes, Coppen, Luttenberg, & Van Boxtel, 2012). Hierbij gaat het dan om vragen zoals: waarom verbrandden ze vroeger mensen die verdacht werden van hekserij? Hoe kan het dat het begrip *handel* in de vroege middeleeuwen een hele andere betekenis heeft dan in de late middeleeuwen?

Doordat contextualiseren niet alleen een vaardigheid is om bronnen te kunnen interpreteren, wordt contextualiseren in dit proefschrift beschouwd als het creëren van een historische context bij gebeurtenissen en handelingen van mensen met als doel deze verschijnselen te verklaren en interpreteren. Wanneer het in dit proefschrift specifiek gaat om het contextualiseren van menselijke handelingen in het verleden, wordt de term *historisch perspectief nemen* gebruikt (vgl. Hartmann & Hasselhorn, 2008; Yeager & Foster, 2001).

Contextualiseren is niet het goedkeuren of relativeren van controversiële verschijnselen. Slavernij bijvoorbeeld kan niet worden goedgekeurd, maar leerlingen moeten wel in staat zijn dit verschijnsel te verklaren en interpreteren door rekening te houden met de historische context. Wanneer ontstond slavernij en hoe lang heeft het bestaan? Welke landen waren betrokken bij de slavenhandel? Wat waren de politieke en economische omstandigheden op dat moment? Welk wereldbeeld hadden mensen toen? Contextualiseren draait om het stellen van dit soort vragen.

9.2.2 Het belang van contextualiseren voor het geschiedenisonderwijs

Waarom focust het proefschrift op contextualiseren? Recentelijk zijn diverse publicaties verschenen over de kern van het schoolvak geschiedenis. Deze studies illustreren dat het schoolvak geschiedenis zich niet alleen moet richten op het aanleren van historische feiten, maar ook op het stimuleren van het historisch denken en redeneren van leerlingen (o.a. Carretero, Berger, & Grever, 2017; Chapman & Wilschut, 2015; Counsell, Burn, & Chapman, 2016; Metzger & McArthur Harris, 2018). Historisch denken en redeneren is het inzetten van historische vaardigheden om zo te kunnen redeneren over veranderingen, oorzaken, gevolgen, overeenkomsten en verschillen met betrekking tot verschijnselen en perioden (Van Boxtel & Van Drie, 2018). In het Nederlandse geschiedenisonderwijs moeten havo- en vwo-leerlingen bijvoorbeeld in staat zijn om historische bronnen te interpreteren, oorzaken en gevolgen te onderscheiden en historische onderzoeksvragen te formuleren (Board of Tests and Examinations, 2017).

Dit historisch denken is echter "onnatuurlijk", omdat mensen automatisch vanuit hun eigen waarden en normen naar het verleden kijken (Wineburg, 2001). Hierdoor slagen leerlingen er vaak niet in om verschijnselen accuraat te verklaren (Lee & Ashby, 2001; Reisman & Wineburg, 2008). Contextualiseren helpt leerlingen rekening te houden met de historische omstandigheden en geeft inzicht in de verschillen tussen hun eigen opvattingen en de waarden, normen en kennis die mensen in verschillende historische perioden hadden. Op deze manier kunnen ze verschijnselen beter verklaren en interpreteren.

Wanneer leerlingen bijvoorbeeld onderzoek doen naar de ontwikkeling van de handel in de oudheid en middeleeuwen, dient voor iedere periode rekening te worden gehouden met een historische context. Handel ten tijde van het Romeinse Rijk betekende namelijk heel iets anders dan handel in de vroege middeleeuwen of

handel in onze huidige samenleving. Ook de verandering van een pre-industriële samenleving naar een industriële samenleving in het achttiende-eeuwse Engeland kan alleen worden verklaard wanneer van beide perioden contexten worden gecreëerd en met elkaar worden vergeleken. Een ander voorbeeld is dat het schot uit het pistool van Gavrilo Princip in 1914, waardoor de Eerste Wereldoorlog uitbrak, alle betekenis verliest wanneer geen rekening wordt gehouden met het opkomende nationalisme, de diverse bondgenootschappen en het moderne imperialisme uit die tijd.

Contextualiseren speelt daarom een essentiële rol in het succesvol kunnen historisch denken en redeneren (Lévesque, 2008; Seixas & Morton, 2013; Van Drie & Van Boxtel, 2008). Leerlingen vinden contextualiseren echter lastig. Zo ondervinden ze problemen bij het beargumenteren of de Verenigde Staten altijd de belangrijkste speler zullen zijn op het wereldtoneel (Foster, Ashby, & Lee, 2008) en bij het verklaren waarom er kinderhuwelijken waren in de vijftiende eeuw (Angvik & Von Borries, 1997). Het proefschrift probeert daarom leerlingen te helpen zich te verbeteren in deze belangrijke vaardigheid.

9.2.3 Een didactisch model van contextualiseren

Op basis van een literatuurstudie is een didactisch model voor contextualiseren ontwikkeld dat wordt gebruikt in de verschillende studies van dit proefschrift. Dit model bevat vier aan elkaar gerelateerde onderdelen: (1) het creëren van een historische context, (2) het bewust maken van presentisme bij het bestuderen van verschijnselen, (3) het stimuleren van historische inleving en (4) het gebruiken van de vaardigheid contextualiseren voor het verklaren en interpreteren van verschijnselen. Figuur 1 in de algemene inleiding van het proefschrift presenteert een visuele weergave van dit model.

Het eerste onderdeel gaat om het creëren van een historische context. Leerlingen kunnen daarvoor gebruik maken van chronologische en ruimtelijke kennis en van kennis over de politieke, economische en sociaal-culturele omstandigheden ten tijde van een verschijnsel (De Keyser & Vandepitte, 1998). Chronologische kennis houdt in dat leerlingen verschijnselen chronologisch kunnen plaatsen in bijvoorbeeld een tijdvak of periode. Ruimtelijke kennis gaat over (geografische) locaties en schalen waarop een verschijnsel zich afspeelde. Voor het verklaren en interpreteren van de Cubacrisis in 1962, moeten leerlingen bijvoorbeeld weten dat Cuba relatief dicht bij de Verenigde Staten ligt en dat de Verenigde Staten zich daardoor erg bedreigd voelden.

De politieke, economische en sociaal-culturele omstandigheden gaan bijvoorbeeld over welk politiek bestuurssysteem, welke economische mate van welvaart en welke sociale verhoudingen er waren ten tijde van het bestudeerde verschijnsel.

Het tweede onderdeel is leerlingen bewust maken van presentisme wanneer zij verschijnselen onderzoeken. Presentisme is het kijken naar het verleden met hedendaagse waarden, normen en kennis. Een aanpak die presentisme kan voorkomen, is het stimuleren van *cognitieve conflicten* bij leerlingen. Deze ontstaan wanneer leerlingen geconfronteerd worden met tegenstrijdige ideeën of wanneer aangereikte informatie botst met hun voorkennis (Johnson & Johnson, 2009). In een geschiedenisles kan dit worden gedaan door het verleden als iets "vreemds" te presenteren (Havekes et al., 2012; Logtenberg, 2012). Geschiedenisleraren kunnen bijvoorbeeld aan leerlingen vragen of kinderarbeid in het negentiende-eeuwse Nederland verboden was. Op deze manier komt naar voren hoe leerlingen naar het verleden kijken. Houden zij rekening met een historische context of kijken zij alleen vanuit hun eigen perspectief naar het verleden?

Het derde onderdeel is het stimuleren van historische inleving. Door bijvoorbeeld een historisch persoon bij een verschijnsel centraal te stellen, kunnen leerlingen meer inzicht krijgen in de historische context van een verschijnsel (Doppen, 2000; Foster, 1999; Wooden, 2008). In dit proefschrift wordt een onderscheid gemaakt tussen twee vormen van historische inleving: het gebruik maken van affectieve connecties met een historisch persoon en het onderzoeken van de rol en positie van een historisch persoon in de maatschappij. Affectieve connecties gaan over herkenbare emoties voor leerlingen, bijvoorbeeld het kunnen begrijpen dat je jouw ouders wilt beschermen wanneer zij met ontslag worden bedreigd. Een meer cognitieve vorm van historische inleving is het onderzoeken van de rol en positie van een historisch persoon. Is de persoon afkomstig uit een elitaire familie? Is de persoon lid van een bepaalde politieke partij of stroming? Dit soort vragen is belangrijk voor het verklaren van hoe een persoon dacht en handelde in een specifieke situatie (Bermúdez & Jaramillo, 2001; Endacott & Brooks, 2013; Hartmann & Hasselhorn, 2008).

Het laatste onderdeel van het didactisch model betreft het inzetten van contextualiseren om verschijnselen te verklaren en interpreteren. Alleen leerlingen leren hoe zij een historische context moeten creëren, hen bewust maken van presentisme of laten inleven, is niet voldoende. Leerlingen moeten hiernaast ook

mogelijkheden krijgen om de vaardigheid contextualiseren te gebruiken voor het onderzoeken van verschijnselen. Leraren kunnen bijvoorbeeld aan leerlingen vragen om te onderzoeken waarom iemand in Duitsland in de jaren 1930 zou kunnen stemmen op de NSDAP of waarom Nieuw-Amsterdam werd geruild met Suriname in de zeventiende eeuw. Op deze manier wordt contextualiseren echt zinvol, omdat het dan expliciet wordt gebruikt voor het verklaren en interpreteren van verschijnselen (Van Boxtel & Van Drie, 2016).

9.3 Onderzoeksvragen

Hoewel contextualiseren een erg belangrijke vaardigheid is voor het hedendaagse geschiedenisonderwijs, ontbreekt het in vakdidactisch geschiedenisonderzoek aan kennis over (1) instrumenten waarmee we kunnen bepalen welk niveau van contextualiseren leerlingen hebben bereikt, (2) instrumenten waarmee we kunnen bepalen in welke mate en op welke wijze leraren aandacht geven aan contextualiseren en (3) empirisch gefundeerd lesmateriaal waarmee contextualiseren kan worden bevorderd. Deze problemen komen samen in de volgende hoofdvraag: *Hoe kan bij leerlingen de historische vaardigheid contextualiseren worden verbeterd?* Voor het beantwoorden van deze vraag zijn vijf deelvragen opgesteld:

1. Hoe kan bij leerlingen het niveau van contextualiseren worden gemeten?
2. Hoe succesvol zijn leerlingen in het contextualiseren van handelingen van personen in het verleden?
3. Hoe kan bij geschiedenisleraren worden geobserveerd in welke mate en op welke wijze zij aandacht besteden aan contextualiseren?
4. In welke mate en op welke wijze besteden geschiedenisleraren aandacht aan contextualiseren?
5. Wat is het effect van een lessenserie waarin expliciete aandacht is voor contextualiseren op het vermogen van leerlingen om te contextualiseren?

9.4 Samenvatting van de resultaten

De eerste twee deelvragen richtten zich op het niveau van leerlingen om handelingen van mensen in het verleden te contextualiseren. Het eerder beschreven didactisch model van contextualiseren werd in deze twee studies gebruikt voor de ontwikkeling en het testen van meetinstrumenten en als kader om te onderzoeken hoe leerlingen contextualiseren. Leraargedrag stond in de derde en vierde deelvraag centraal. In deze twee studies werd het model gebruikt voor de ontwikkeling van een observatie-instrument en als analysekader om te bepalen in welke mate en op welke wijze leraren aandacht besteden aan contextualiseren. De laatste deelvraag focuste op de ontwikkeling en het testen van lesmateriaal. Hiervoor werden twee interventie-studies uitgevoerd waarin het model werd gebruikt voor de ontwikkeling van ontwerpprincipes en om te analyseren welke vooruitgang leerlingen boekten in het niveau van contextualiseren. Hieronder volgen de belangrijkste bevindingen per deelvraag.

1. Hoe kan bij leerlingen het niveau van contextualiseren gemeten worden?

Aangezien er weinig betrouwbare instrumenten bestaan voor het meten van het niveau van leerlingen met betrekking tot contextualiseren, was het doel van de eerste studie (hoofdstuk 2) het ontwikkelen en testen van dit soort instrumenten. Twee instrumenten stonden centraal. Eén instrument was een vertaling van een instrument dat is ontwikkeld door Hartmann en Hasselhorn (2008). Dit instrument richt zich op de opkomst van de NSDAP in Duitsland in de jaren 1930. Hierin moeten leerlingen beredeneren of Hannes (de hoofdpersoon) kan stemmen op de politieke partij van Hitler. Het format van dit instrument is gebruikt om een tweede instrument te ontwikkelen dat was gericht op negentiende-eeuwse slavernij. Beide instrumenten zijn getest bij 1270 leerlingen variërend in leeftijd van 10 tot 17 jaar op validiteit en betrouwbaarheid. Door gebruik te maken van factoranalyses en het berekenen van de interne consistentie (Cronbach's alfa) bleek het NSDAP-instrument het meest betrouwbaar. De analyses lieten bovendien zien dat het lastig is om op basis van het format van Hartmann en Hasselhorn (2008) een nieuw instrument te ontwikkelen. Op basis van de studie van Hartmann en Hasselhorn (2008) moesten twee dimensies naar voren komen: een dimensie van contextualiseren en een dimensie van historische inleving. Het vertaalde en onderzochte NSDAP-instrument bevatte deze twee dimensies. Het slavernij-instrument bevatte echter drie dimensies en kan daardoor andere elementen toetsen dan het NSDAP-instrument. Verder scoorde het slavernij-instrument laag op de interne consistentie. Om deze reden is in een vervolgstudie

alleen gebruikt gemaakt van de scores op het NSDAP-instrument. Deze scores gaven aan dat leerlingen in hogere leerjaren hogere scores behaalden dan leerlingen in lagere leerjaren. Dit is mogelijk te verklaren doordat leerlingen in hogere leerjaren meer historische kennis hebben. Er was namelijk een positieve correlatie (.27) tussen hun voorkennis van het historische onderwerp en hun scores op het instrument.

2. Hoe succesvol zijn leerlingen in het contextualiseren van handelingen van personen in het verleden?

Het doel van de tweede studie (hoofdstuk 3) was te achterhalen hoe succesvol 15- en 16-jarige leerlingen zijn wanneer hen wordt gevraagd handelingen van een (fictief) historisch persoon te contextualiseren. Om dit te bepalen werd gebruik gemaakt van een kwantitatieve en kwalitatieve onderzoeksmethode. Voor 15- en 16-jarige leerlingen werd gekozen, omdat leerlingen van deze leeftijd kunnen contextualiseren (o.a. Berti, Baldin, & Toneatti, 2009; Hartmann & Hasselhorn, 2008). Hierdoor was het zeker dat redeneringen die leiden tot het succesvol voltooien van de opdracht konden worden onderzocht. Het NSDAP-instrument werd eerst gebruikt om te onderzoeken hoe succesvol (n = 143) leerlingen de taak oplosten. Ongeveer 5% van de leerlingen haalde een score < 2.50 op een vierpuntschaal. De meesten (57%) haalden een gemiddelde score ≥ 3.00 < 3.50; 22% bereikte een score > 3.50. Vervolgens werd de opdracht uit het NSDAP-instrument door leerlingen (n = 36) hardop-denkend uitgevoerd om te onderzoeken welke verschillende redeneringen leerlingen gebruikten. Hierbij werd gelet op presentisme, historische inleving en de soorten kennis die zij gebruikten. Uit de analyses van de hard-op-denk-protocollen bleek dat slechts vijf leerlingen zich lieten leiden door presentisme en dat één leerling de hoofdrolspeler in de opdracht (Hannes) naïef vond. Verder gebruikte het merendeel van de leerlingen historisch inleving om handelingen van personen te begrijpen. Dit gebeurde vooral doordat leerlingen sommige emoties van de hoofdpersoon herkenden (bijvoorbeeld liefde voor je ouders). Leerlingen hielden echter weinig rekening met de maatschappelijke positie van de hoofdpersoon. Ze gebruikten verder chronologische kennis en kennis over de politieke, economische en sociale omstandigheden om de taak op te lossen, maar geografische kennis (waar en op welke schaal het verschijnsel speelde) werd weinig gebruikt.

3. *Hoe kan bij geschiedenisleraren worden geobserveerd in welke mate en op welke wijze zij aandacht besteden aan contextualiseren?*

De derde studie (hoofdstuk 4) beschrijft de ontwikkeling van een observatie-instrument dat inzicht geeft in hoeverre geschiedenisleraren contextualiseren in hun lessen bevorderen. Dit instrument ontbrak, zodat wij weinig wisten over hoe leraren kunnen worden geholpen met het bevorderen van contextualiseren. Het ontwikkelde Framework for Analyzing the Teaching of Historical Contextualization (FAT-HC) bevat 40 items die gescoord kunnen worden op een schaal van 1 (zwak) tot en met 4 (sterk). Met behulp van generaliseerbaarheidstheorie (Brennan, 2001; Shavelson & Webb, 1991) werd de betrouwbaarheid van dit instrument onderzocht. Bij vijf leraren werden twee lessen gefilmd waarbij iedere les vervolgens door vijf getrainde observatoren werd gescoord. Uit de resultaten van deze 50 observaties kwam een positief beeld van de inhoudsvaliditeit en dimensionaliteit van het instrument naar voren. Een beslissingsstudie (d-studie) gaf informatie over de hoeveelheid lessen en observatoren die nodig zijn voor een betrouwbaar beeld van hoe leraren contextualiseren (Brennan & Kane, 1977). Voor onderzoeksdoeleinden is het voldoende betrouwbaar wanneer één observator met het FAT-HC twee lessen observeert van dezelfde leraar. Voor het geven van feedback aan leraren is het voldoende betrouwbaar wanneer twee observatoren ieder twee lessen van dezelfde leraar observeren of wanneer drie observatoren één les van dezelfde leraar observeren.

4. *In welke mate en op welke wijze besteden geschiedenisleraren aandacht aan contextualiseren?*

De vierde studie (hoofdstuk 5) had als doel om met het FAT-HC te onderzoeken in welke mate en op welke wijze geschiedenisleraren in hun lessen aandacht besteden aan contextualiseren. Deze studie beschrijft hoe vijf getrainde observatoren acht geschiedenisleraren tweemaal bekeken met behulp van het FAT-HC. De 40 items in het FAT-HC werden verdeeld in acht categorieën. Er waren aparte categorieën voor items waarbij de leraar contextualiseert (bijvoorbeeld de leraar schetst de politieke omstandigheden ten tijde van een verschijnsel) en categorieën voor items waarbij de leerlingen contextualiseren (bijvoorbeeld de leerlingen onderzoeken de politieke omstandigheden ten tijde van een verschijnsel). Per leraar werd voor iedere categorie een gemiddelde score berekend en een gemiddelde totaalscore (FAT-HC-score). Geen enkele leraar behaalde een gemiddelde FAT-HC-score > 2.00 op een vierpuntsschaal. Wanneer er wel gecontextualiseerd werd, creëerde de leraar meestal zelf een historische context bij een verschijnsel. De leerlingen werden door de leraren echter weinig gevraagd zelf te contextualiseren.

5. *Wat is het effect van een lessenserie waarin expliciete aandacht is voor contextualiseren op het vermogen van leerlingen om te contextualiseren?*

De vijfde en zesde studie zijn interventiestudies met als doel het ontwikkelen en testen van lesmateriaal dat leraren kunnen gebruiken om het niveau van hun leerlingen met betrekking tot contextualiseren te verhogen.

Op basis van het didactische model van contextualiseren zijn in de vijfde studie (hoofdstuk 6) vier ontwerpprincipes voor het bevorderen van contextualiseren opgesteld: (1) maak leerlingen bewust van hun hedendaagse perspectief, (2) leer leerlingen hoe zij een historische context kunnen creëren, (3) creëer mogelijkheden voor leerlingen om de vaardigheid contextualiseren in te zetten voor het verklaren en interpreteren van verschijnselen en (4) stimuleer historische inleving bij leerlingen. Tezamen vormden deze vier principes een didactiek waarbij eerst aan leerlingen werd gevraagd een casus te verklaren (bijvoorbeeld de ruil van Nieuw-Amsterdam voor Suriname). Op deze manier werd de aanwezigheid van presentisme bij leerlingen duidelijk. Daarna creëerden leerlingen een historische context bij de casus om vervolgens dezelfde casus nogmaals te verklaren. De bedoeling van dit terugblikken was dat leerlingen gingen inzien dat zij bij het verklaren van verschijnselen zich bewust moesten zijn van presentisme en de historische context moesten gebruiken om tot een goede interpretatie te komen. Het laatste onderdeel van de didactiek was het maken van een historische inlevingsopdracht. Deze didactiek werd gebruikt voor de ontwikkeling van een lessenserie van acht lessen voor havo- en vwo-leerlingen. Iedere les moesten leerlingen dezelfde lesactiviteiten uitvoeren met betrekking tot verschillende historische onderwerpen. Om te kijken naar het effect van de lessenserie op het niveau van contextualiseren werd een quasi-experimentele onderzoeksopzet gebruikt met een experimentele en controle conditie ($N = 131$). De gemiddelde leeftijd van de leerlingen was 16 jaar. Scores op een voor- en nameting (meerkeuzevragen) toonden aan dat leerlingen in de experimentele conditie na afloop significant hoger scoorden op de vaardigheid contextualiseren dan de leerlingen in de controle conditie. Uit interviews met de deelnemende leraren bleek verder dat zij de verschillende lesactiviteiten nuttig en effectief vonden voor het bevorderen van contextualiseren bij leerlingen.

In de zesde studie (hoofdstuk 7) werd het didactische ontwerp van de vijfde studie aangepast. De leraren die hadden deelgenomen aan het experiment vonden namelijk dat de lessen erg vol zaten en te veel uit herhaling bestonden. Dit demotiveerde leerlingen: iedere les moesten zij hetzelfde doen. In het nieuwe ontwerp vervielen

de historische inlevingstaken, werd het experiment ingekort naar vier lessen en werd een opbouwende moeilijkheidsgraad aangebracht. In de eerste twee lessen werden de drie stappen (presenteren casus, creëren historische context en terugblikken casus) gebruikt voor het oefenen en aanleren van contextualiseren en in de laatste twee lessen doorliepen leerlingen meer zelfstandig de drie stappen om een verschijnsel te onderzoeken. De effectiviteit van de lessenserie werd vervolgens onderzocht in een quasi-experimentele studie met een experimentele en controle conditie ($N = 169$). De gemiddelde leeftijd van deze havo- en vwo-leerlingen was 15 jaar. Een vergelijking tussen de twee condities op basis van een voor- en nameting (open vragen) gaf aan dat leerlingen in de experimentele conditie na afloop minder presentisme en meer kennis van de historische context gebruikten dan leerlingen in de controle conditie.

9.5 Algemene conclusie

De centrale vraag van dit proefschrift was: *Hoe kan bij leerlingen de historische vaardigheid contextualiseren worden verbeterd?* Leerlingen kunnen beter worden in contextualiseren als het expliciet onderwezen wordt. Hierbij is het van belang dat leerlingen zelf moeten contextualiseren. Dit kan worden bevorderd wanneer leraren de volgende ontwerpprincipes gebruiken: (1) maak leerlingen bewust van hun hedendaagse blik, (2) leer leerlingen hoe zij een context kunnen creëren, (3) creëer mogelijkheden voor leerlingen om de vaardigheid contextualiseren in te zetten voor het verklaren en interpreteren van verschijnselen en (4) stimuleer historische inleving bij leerlingen.

Een didactiek die effectief is gebleken heeft de volgende fasering: presenteer eerst een casus over een historisch onderwerp dat een cognitief conflict oproept bij leerlingen. Geef daarna leerlingen de instructie om een relevante historische context bij deze casus te creëren. Vraag ten slotte aan de leerlingen om nogmaals de casus te verklaren waarbij leerlingen gebruik moeten maken van de gecreëerde historische context. Een opbouwende moeilijkheidsgraad heeft hierbij de voorkeur. Een voorbeeld van een dergelijke opbouw is om de stappen eerst te gebruiken om leerlingen kennis te laten maken met contextualiseren en hen te leren contextualiseren (bijvoorbeeld door het bespreken van stappenplannen voor het creëren van een historische context). Daarna kunnen leerlingen meer zelfstandig de stappen doorlopen om een opdracht te voltooien waarin een verschijnsel moet worden verklaard (bijvoorbeeld de wijziging van een Amsterdamse straatnaam in 1956).

9.6 Discussie

Eerst worden de beperkingen van het onderzoek en suggesties voor toekomstig onderzoek beschreven. Daarna worden de wetenschappelijke en de praktische relevantie van dit proefschrift bediscussieerd. Een pleidooi voor meer ruimte voor historisch besef in het Nederlandse geschiedenisonderwijs vormt het slot.

9.6.1 Beperkingen en suggesties voor verder onderzoek

Dit proefschrift kent drie belangrijke beperkingen. De eerste beperking is dat de eerste twee studies zich richtten op het contextualiseren van handelingen van historische personen (historisch perspectief nemen). Contextualiseren is echter een breder concept. Ook ontwikkelingen en historische bronnen kunnen bijvoorbeeld worden gecontextualiseerd en zijn in deze twee studies niet meegenomen. Een tweede beperking bestaat uit de instrumenten die werden ingezet in de verschillende studies. Het NSDAP-instrument dat werd gebruikt in de eerste en tweede studie focust op maar één historisch onderwerp (de opkomst van Hitler in de jaren 1930) en betreft een fictief scenario. Meer onderzoek is daarom nodig naar hoe leerlingen contextualiseren bij andere historische onderwerpen en wanneer gebruik wordt gemaakt van niet-fictieve bronnen. In het zesde hoofdstuk werden meerkeuzevragen gebruikt voor het meten van contextualiseren. Net als de open vragen die gebruikt werden in het zevende hoofdstuk, vereisen deze vragen meer onderzoek naar de betrouwbaarheid en validiteit. Meer onderzoek is ook nodig naar de betrouwbaarheid en validiteit van het FAT-HC. Is het instrument inderdaad ééndimensionaal? Kunnen wij het aantal items verminderen zonder dat de betrouwbaarheid van het instrument in gevaar komt? Een derde beperking is het aantal participanten en lessen. De meeste studies in dit proefschrift waren exploratief van aard. Onderzoek met meer lessen, leerlingen en leraren is nodig om te kijken of de bevindingen worden ondersteund.

Het proefschrift geeft verder drie belangrijke richtingen voor vervolgonderzoek. Ten eerste zijn meer gestandaardiseerde instrumenten nodig die inzicht geven in hoe leerlingen historisch denken en redeneren. Dit proefschrift geeft een aanzet tot de ontwikkeling van instrumenten die het vermogen van leerlingen om te contextualiseren meten. Dit soort instrumenten is echter ook nodig voor andere historische vaardigheden, zoals het onderscheiden van oorzaak en gevolg en het analyseren van verandering en continuïteit. Daarnaast richt het FAT-HC zich alleen op contextualiseren. Een tweede richting voor vervolgonderzoek is daarom de

ontwikkeling van observatie-instrumenten die inzicht geven in hoeverre andere historische vaardigheden, zoals het omgaan met verandering en continuïteit, worden bevorderd in geschiedenislessen. Ten slotte dient vervolgonderzoek zich te richten op het uitvoeren van quasi-experimentele studies. Binnen het vakdidactisch geschiedenisonderzoek zijn deze studies schaars terwijl dit belangrijke inzichten kan opleveren in wat wel en niet effectief is in het geschiedenisonderwijs.

9.6.2 Wetenschappelijke relevantie

Wat heeft het onderzoek toegevoegd aan bestaande inzichten in de didactiek van het schoolvak geschiedenis? Hoewel verschillende studies zijn uitgevoerd naar het ontwikkelen en testen van instrumenten (o.a. Ercikan & Seixas, 2015; Smith, 2018; VanSledright, 2013), waren er niet veel instrumenten die zich richten op het meten van het niveau van contextualiseren bij leerlingen. Dit proefschrift presenteert verschillende instrumenten die het leerlingniveau van contextualiseren in kaart kunnen brengen. In de eerste studie wordt bovendien duidelijk hoe leerlingen van verschillende leeftijden en leerniveaus contextualiseren. Over deze verschillen was, zeker in de Nederlandse context, nog niet veel bekend. Daarnaast zijn er geen goede observatie-instrumenten die inzicht geven in het vraagstuk hoe geschiedenisleraren het historisch denken en redeneren van leerlingen bevorderen (Van Hover, Hicks, & Cotton, 2012). In de derde studie is daarom een observatie-instrument ontwikkeld dat in de vierde studie is gebruikt om te onderzoeken in welke mate leraren aandacht besteden aan contextualiseren. Ten slotte concluderen verschillende onderzoekers (o.a. Fogo, 2014; Grant & Gradwell, 2010; Reisman & Fogo, 2016) dat er weinig bekend is over de effecten van lesmateriaal op het historisch denken en redeneren van leerlingen. De laatste twee studies van dit proefschrift zijn interventiestudies en geven inzicht in hoe lesmateriaal ontwikkeld en getest kan worden op dit soort effecten.

9.6.3 Praktische relevantie

De relevantie van dit proefschrift voor de praktijk van het geschiedenisonderwijs kan worden samengevat in drie belangrijke punten. Ten eerste heeft het proefschrift verschillende instrumenten opgeleverd die leraren kunnen inzetten om te achterhalen hoe leerlingen contextualiseren. Voordat effectieve lessen kunnen worden ontworpen, moet eerst duidelijk zijn welke kennis, vaardigheden en concepties leerlingen hebben op het gebied van contextualiseren. Het NSDAP-instrument uit de eerste twee studies, de meerkeuzevragen uit de vijfde studie en de open vragen

uit de zesde studie kunnen hiervoor worden gebruikt. Het tweede punt is dat het proefschrift een observatie-instrument (het FAT-HC) heeft opgeleverd dat gebruikt kan worden om lessen van geschiedenisleraren te observeren. Dit instrument is geen beoordelingsinstrument, maar is bedoeld voor onderzoekers, lerarenopleiders en leraren (in opleiding) om elkaar te observeren en van elkaar te leren op het gebied van contextualiseren. Dit kan erg goed door het bespreken van de bevindingen na afloop van een observatie. Een derde punt is dat het onderzoek vier ontwerpprincipes heeft opgeleverd waarmee leraren onderwijsactiviteiten kunnen ontwerpen die het vermogen van leerlingen om te contextualiseren bevorderen. De ontwerpprincipes en voorbeelden van lesmateriaal staan beschreven in de vijfde en zesde studie. Om leraren te inspireren bevat Appendix I voorbeelden van casussen die een cognitief conflict bij leerlingen kunnen oproepen.

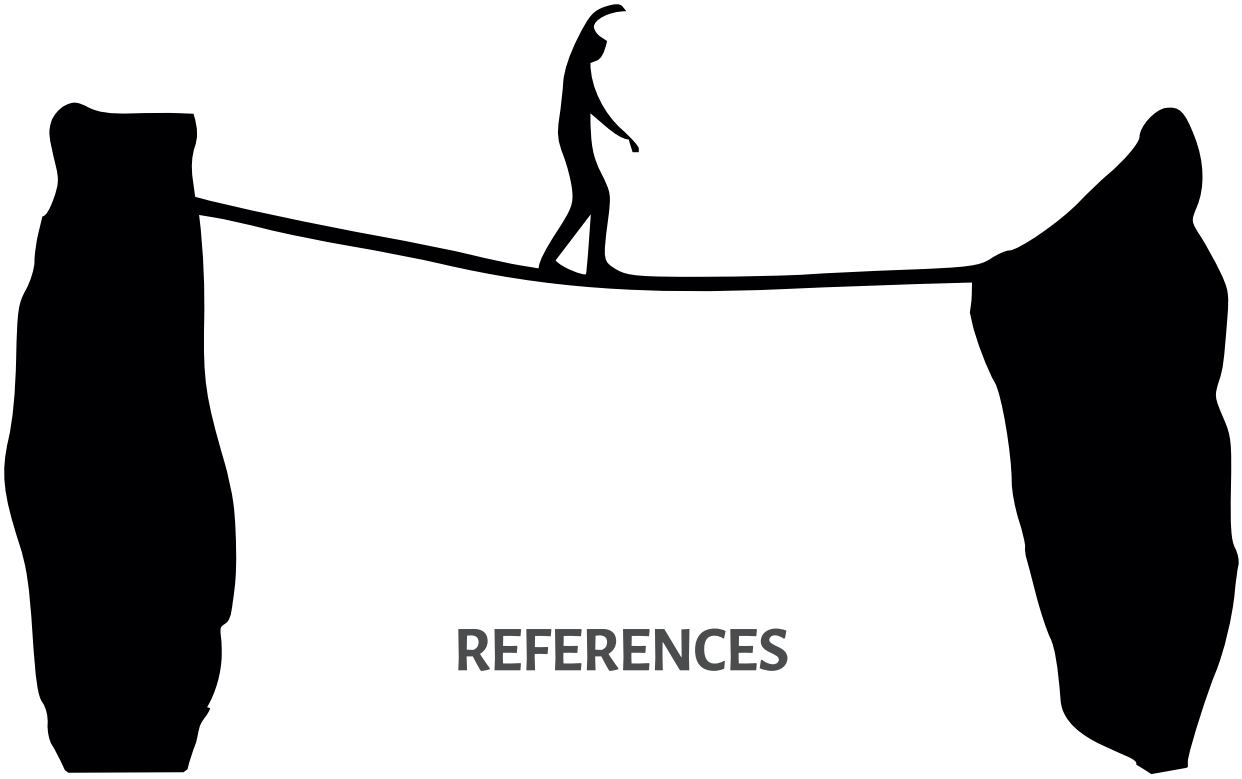
9.6.4 Maak meer ruimte voor historisch besef

In deze laatste paragraaf richt ik mij graag tot iedereen die werkzaam is in het onderwijs. Momenteel woedt in het kader van Curriculum.nu een maatschappelijke discussie over de doelen en inrichting van het Nederlandse onderwijs. Het prachtige schoolvak geschiedenis wordt door sommige mensen gezien als ouderwets, want historische feiten kun je toch gewoon online opzoeken? Dit proefschrift laat zien dat geschiedenisonderwijs meer is dan leerlingen leren om zoveel mogelijk feiten op te dreunen. Contextualiseren leidt niet alleen tot het beter kunnen verklaren en interpreteren van verschijnselen, maar ook tot het inzicht dat er verschillen bestaan tussen mensen in waarden, normen en kennis. Wanneer leerlingen zich hiervan bewust zijn, is de kans een stuk kleiner dat zij verzetsmonumenten vernielen of hun blote billen laten zien in een Cambodjaanse tempel. Wij willen leerlingen toch opvoeden tot burgers die zich ervan bewust zijn dat mensen en culturen verschillende opvattingen kunnen hebben? Dat is waar historisch besef voor mij om draait.

Er moet echter wel iets veranderen in onderwijsland om leerlingen dit historisch besef mee te kunnen geven. Het proefschrift laat namelijk zien dat het aanleren van contextualiseren alleen kan gebeuren wanneer er een dialoog is tussen leraren en leerlingen, wanneer leerlingen nieuwsgierig zijn naar het verleden en wanneer iedere leerling wordt uitgedaagd zichzelf verder te ontwikkelen. Voor het ontwerpen van onderwijsactiviteiten die hiervoor zorgen, hebben leraren veel meer tijd, ruimte en ondersteuning nodig. Om dit te realiseren is het noodzakelijk dat het aantal lesuren, dat een leraar per week moet geven, wordt gereduceerd. Een voorstel:

twintig procent van de aanstellingsomvang van een docent is voor scholing en het gezamenlijk ontwikkelen van onderwijs. Het is zonde wanneer leraren bijvoorbeeld de instrumenten die dit proefschrift heeft opgeleverd niet gaan gebruiken door een gebrek aan tijd.

Verder is het noodzakelijk dat het knellende eindexamenprogramma voor het schoolvak geschiedenis wordt losgelaten. Dit programma zorgt er namelijk voor dat projecten die historisch besef bij leerlingen stimuleren niet van de grond komen. Waarom wordt door leraren in de bovenbouw erg weinig tijd ingepland om samen met musea en andere maatschappelijke instellingen een tentoonstelling te organiseren waarvoor leerlingen mensen hebben geïnterviewd over hun leven tijdens de Koude Oorlog? Dit stimuleert het historisch besef van leerlingen toch veel meer dan het kunnen opdreunen van de kenmerkende aspecten? Dit zorgt toch voor meer motivatie bij leerlingen en leraren? Dit bereidt leerlingen toch veel beter voor op het vervolgonderwijs? Het antwoord op deze drie vragen is een volmondig "ja". De druk van het examen zit echter in de weg. Daarom moeten wij samen op zoek naar andere inrichtingen van het curriculum en naar andere toetsingsvormen. Maar dan graag zonder de molensteen van het examenprogramma om onze lerarennek.



REFERENCES

A

Achinstein, B., & Fogo, B. (2015). Mentoring novices' teaching of historical reasoning: Opportunities for pedagogical content knowledge development through mentor-facilitated practice. *Teaching and Teacher Education, 45*, 45-58. doi:10.1016/j.tate.2014.09.002

Adler, S. (2008). The education of social studies teachers. In L. S. Levstik and C. A. Tyson (Eds.) *Handbook of research in social studies education* (pp. 329–351). New York, NY: Routledge.

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

Angvik, M., & Von Borries, B. (1997). *A comparative European survey on historical consciousness and political attitudes among adolescents*. Hamburg, Germany: Körber Stiftung.

Ashby, R. (2004). Developing a concept of historical evidence: Students' ideas about testing singular factual claims. *International Journal of Historical Learning, Teaching and Research, 4*(2), 44–55.

Australian Curriculum, Assessment and Reporting Authority. (2014). *The Australian Curriculum*. Retrieved from www.australiancurriculum.edu.au

B

Bailey, K. M. (1990). The use of diaries in teacher education programs. In J. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 215-226). Cambridge, United Kingdom: Cambridge University Press.

Bain, R., & Mirel, J. (2006). Setting up camp at the great instructional divide educating beginning history teachers. *Journal of Teacher Education, 57*, 212-219. doi:10.1177/0022487105285892

Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education, 60*, 497-511. doi:10.1177/0022487109348479

Baron, C. (2016). Using embedded visual coding to support contextualization of historical texts. *American Educational Research Journal, 53*, 516-540. doi:10.3102/0002831216637347

Barton, K. C. (1996). Narrative simplifications in elementary children's historical understanding. In J. Brophy (Ed.), *Advances in research on teaching* (pp. 51-83). Greenwich, CT: JAI Press.

Barton, K. C. (1997). "Bossed around by the queen": Elementary students' understanding of individuals and institutions in history. *Journal of Curriculum and Supervision, 12*, 290-314.

Barton, K. C. (1997). "I just kinda know": Elementary students' ideas about historical evidence. *Theory & Research in Social Education, 25*, 407-430. doi:10.1080/00933104.1997.10505821

Barton, K. C. (2008). Research on students' ideas about history. In L. S. Levstik & C. A. Tyson (Eds.), *Handbook of research in social studies education* (pp. 239–258). New York, NY: Routledge.

Barton, K. C. (2012). Agency, choice and historical action: How history teaching can help students think about democratic decision making. *Citizenship Teaching & Learning, 7*, 131-142. doi:10.1386/ctl.7.2.131_1

Barton, K. C., & Levstik, L. S. (2003). Why don't more history teachers engage students in interpretation? *Social Education, 67*, 358-358.

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

Bermúdez, A., & Jaramillo, R. (2001). Development of historical explanation in children, adolescents and adults. In A. Dickinson, P. Gordon, & P. Lee (Eds.), *Raising standards in history education* (pp. 146–167). London, United Kingdom: Woburn Press.

Berti, A. E., Baldin, I., & Toneatti, L. (2009). Empathy in history. Understanding a past institution (ordeal) in children and young adults when description and rationale are provided. *Contemporary Educational Psychology, 34*, 278-288. doi:10.1016/j.cedpsych.2009.06.002

Bevir, M. (2002). How to be an intentionalist. *History and Theory, 41*, 209-217. doi:10.1111/0018-2656.0

Beyer, B. K. (2008). How to teach thinking skills in social studies and history. *The Social Studies, 99*, 196-201. doi:10.3200/TSSS.99.5.196-201

Birch, S. A., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18*, 382-386. doi:10.1111/j.1467-9280.2007.01909.x

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*, 25-31. doi:10.1016/S0010-0277(00)00096-2

Board of Tests and Examinations. (2015). *Examenprogramma geschiedenis vwo* [History exam program pre-university education]. Retrieved from https://www.examenblad.nl/vak/geschiedenis/2015/vwo

Board of Tests and Examinations. (2017). *Examenprogramma geschiedenis vwo* [History exam program pre-university education]. Retrieved from https://www.examenblad.nl/vak/geschiedenis/2018/vwo

Boerman-Cornell, W. (2015). Using historical graphic novels in high school history classes: Potential for contextualization, sourcing, and corroborating. *The History Teacher, 48*, 209-224.

Bose, M., & Dey, A. (2009). *Optimal crossover designs*. Singapore, Singapore: World Scientific Publishing.

Bransford, J., Brown, A., & Cocking, R. (2000). *How people learn. Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Breakstone, J. (2014). Try, try, try again: The process of designing new history assessments. *Theory & Research in Social Education, 42*, 453-485. doi:10.1080/00933104.2014.965860

Breakstone, J., Smith, M. D., & Wineburg, S. S. (2013). Beyond the bubble in history/social studies assessments. *Phi Delta Kappan, 94*(5), 53-57. doi:10.1177/003172171309400512

Brennan, R. L. (2001). *Generalizability theory*. New York, NY: Springer-Verlag.

Brennan, R. L., & Kane, M. T. (1977). An index of dependability for mastery tests. *Journal of Educational Measurement, 14*, 277–289. doi:10.1111/j.1745-3984.1977.tb00045.x

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

Brooks, S. (2008). Displaying historical empathy: What impact can a writing assignment have. *Social Studies Research and Practice, 3*(2), 130-146.

Brooks, S. (2011). Historical empathy as perspective recognition and care in one secondary social studies classroom. *Theory & Research in Social Education, 39*, 166-202. doi:10.1080/00933104.2011.10473452

Brophy, J., & VanSledright, B. A. (1997). *Teaching and learning history in elementary schools*. New York, NY: Teachers College Press.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences, 2*, 141-178. doi:10.1207/s15327809jls0202_2

C

Carretero, M., Asensio, M., & Rodriguez-Moneo, M. (Eds.). (2012). *History education and the construction of national identities*. London, United Kingdom: Information Age.

Carretero, M., Berger, S., & Grever, M. C. R. (Eds.). (2017). *Palgrave handbook of research in historical culture and education*. London, United Kingdom: Palgrave Macmillan.

Chapman, A., & Wilschut, A. H. J. (Eds.). (2015). *Joined-up history: New directions in history education research*. Charlotte, NC: Information Age.

Chávez, R. C. (1984). The use of high-inference measures to study classroom climates: A review. *Review of Educational Research, 54*, 237-261. doi:10.3102/00346543054002237

Ciardiello, A. V. (2012). Is Angel Island the Ellis Island of the west? Teaching multiple perspective-taking in American immigration history. *The Social Studies, 103*, 171-176. doi:10.1080/00377996.2011.606436

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). New York, NY: Academic Press.

Colby, S. (2008). Energizing the history classroom: Historical narrative inquiry and historical empathy. *Social Studies Research and Practice, 3*(3), 60-79.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences, 13*, 15-42. doi:10.1207/s15327809jls1301_2

Colton, D., & Covert, R. W. (2007). *Designing and constructing instruments for social research and evaluation*. San Francisco, CA: Wiley.

Cook, T. D., & Payne, M. R. (2002). Objecting to the objections to using random assignment in educational research. In F. Mosteller & R. Boruch (Eds.), *Evidence matters. Randomized trials in education research* (pp. 150-178). Washington, DC: Brookings.

Cooper, H., & Chapman, A. (2009). (Eds.) *Constructing history 11-19*. London, United Kingdom: Sage.

Counsell, C. (2000). Historical knowledge and historical skills. In J. Arthur & R. Phillips (Eds.), *Issues in history teaching* (pp. 54-70). London, United Kingdom: Routledge.

Counsell, C., Burn, K., & Chapman, A. (Eds.). (2016). *Masterclass in history education. Transforming teaching and learning*. London, United Kingdom: Bloomsbury.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, CA: Sage.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.

Cronbach, L., Gleser, G., Nanda, H., & Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements*. New York, NY: Wiley.

Cunningham, D. L. (2007). Understanding pedagogical reasoning in history teaching through the case of cultivating historical empathy. *Theory & Research in Social Education, 35*, 592-630. doi:10.1080/00933104.2007.10473352

Cunningham, D. L. (2009). An empirical framework for understanding how teachers conceptualize and cultivate historical empathy in students. *Journal of Curriculum Studies, 41*, 679-709. doi:10.1080/00220270902947376

D

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L. (2012). *Creating a comprehensive system for evaluating and supporting effective reaching*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan, 93*(6), 8-15. doi:10.1177/003172171209300603

Darling-Hammond, L., Wei, R., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession*. Washington, DC: National Staff Development Council.

Davies, I. (Ed.). (2011). *Debates in history teaching*. London, United Kingdom: Routledge.

Davis, O. L. (2001). In pursuit of historical empathy. In O. L. Davis, E. A. Yeager & S. J. Foster. (Eds.), *Historical empathy and perspective taking in the social studies* (pp. 1-12). Lanham, MD: Rowman & Littlefield.

Davis, O. L., Yeager, E. A., & Foster, S. J. (Eds.). (2001). *Historical empathy and perspective taking in the social studies*. Lanham, MD: Rowman & Littlefield.

Dawson, I. (2009). What time does that tune start? From thinking about "sense of period" to modelling history at key stage 3. *Teaching History, 135*, 50–57.

De Graaff, H. J. C., Koopman, G. J., Anikina, Y., & Westhoff, G. (2007). An observation tool for effective L2 pedagogy in content and language integrated learning (CLIL). *International Journal of Bilingual Education and Bilingualism, 10*, 603-624. doi:10.2167/beb462.0

De Keyser, R., & Vandepitte, P. (1998). *Historical formation. Design of vision*. Brussels, Belgium: Flemish Board for Catholic Secondary Education.

De La Paz, S., Felton, M., Monte-Sano, C., Croninger, R., Jackson, C., Deogracias, J. S., & Hoffman, B. P. (2014). Developing historical reading and writing with adolescent readers: Effects on student learning. *Theory & Research in Social Education, 42*, 228-274. doi:10.1080/00933104.2014.908754

De Leur, T. L., Van Boxtel, C. A. M., & Wilschut, A. H. J. (2017). 'I saw angry people and broken statues': Historical empathy in secondary history education. *British Journal of Educational Studies, 65*, 331-352. doi:10.1080/00071005.2017.1291902

Den Heyer, K. (2003). Between every "now" and "then": A role for the study of historical agency in history and citizenship education. *Theory & Research in Social Education, 31*, 411-434. doi:10.1080/00933104.2003.10473232

Department for Education. (2013). *National curriculum in England: History programmes of study*. Retrieved from https://www.gov.uk/government/publications/national-curriculum-in-england-history-programmes-of-study

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*, 181-199. doi:10.3102/0013189X08331140

Desimone, L. M., Smith, T. M., & Ueno, K. (2006). Are teachers who need sustained, content-focused professional development getting it? An administrator's dilemma. *Educational Administration Quarterly, 42*, 179-215. doi:10.1177/0013161X04273848

Doppen, F. (2000). Teaching and learning multiple perspectives: The atomic bomb. *The Social Studies, 91*, 159-169. doi: 10.1080/00377990009602461

Dutch National Institute for Educational Measurement. (2013). Terugblik en resultaten 2013 Eindtoets Basisonderwijs [Review and results final test elementary education 2013]. Retrieved from http://www.cito.nl/onderzoek%20en%20wetenschap/achtergrondinformatie/primair_speciaal_onderwijs/eindtoets_onderzoek_achtergrond

Dutch Inspectorate of Education (2012). De Staat van het onderwijs: onderwijsverslag 2012/2013 [The state of education in the Netherlands 2012/2013]. Retrieved from https://www.onderwijsinspectie.nl/documenten/publicaties/2014/04/16/de-staat-van-het-onderwijs-onderwijsverslag-2012-2013

Dutch Ministry of Education, Culture, and Science. (2011). *Personeel voortgezet onderwijs* [Secondary school teachers]. Retrieved from https://duo.nl/open_onderwijsdata/databestanden/vo/onderwijspersoneel

E

Eisenberg, N. (2000). Empathy. In K. E. Alan (Ed.), *Encyclopedia of psychology* (Vol. 3, pp. 179-182). New York, NY: Oxford University Press.

Endacott, J. L. (2010). Reconsidering affective engagement in historical empathy. *Theory & Research in Social Education, 38*, 6-47. doi:10.1080/00933104.2010.10473415

Endacott, J. L., & Brooks, S. (2013). An updated theoretical and practical model for promoting historical empathy. *Social Studies Research and Practice, 8*(1), 41–58.

Endacott, J. L., & Pelekanos, C. (2015). Slaves, women, and war! Engaging middle school students in historical empathy for enduring understanding. *The Social Studies, 106*, 1-7. doi:10.1080/00377996.2014.957378

Endacott, J. L., & Sturtz, J. (2014). Historical empathy and pedagogical reasoning. *The Journal of Social Studies Research, 39*, 1-16. doi:10.1016/j.jssr.2014.05.003

Ercikan, K., & Seixas, P. (Eds.). (2015). *New directions in assessing historical thinking*. New York, NY: Routledge.

Erdmann, E., & Hasberg, W. (Eds.). (2011). *Facing–mapping–bridging diversity Part 1 & 2. Foundation of a European discourse on history education*. Schwalbach, Germany: Wochenschau-Verlag.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Evans, R. W. (2011). *The hope for American school reform: The cold war pursuit of inquiry learning in social studies*. New York, NY: Palgrave Macmillan.

F

Fasulo, A., Girardet, H., & Pontecorvo, C. (1998). Seeing the past: Learning history through group discussion and iconographic sources. In J. F. Voss, & M. Carretero (Eds.) *Learning and reasoning in history. International review of history education* (Vol. 2, pp. 132–153). London, United Kingdom: Woburn.

Federal Civil Defense Administration (Producer) & Rizzo, A (Director). (1951). *Duck and Cover* [Motion picture]. United States of America: Archer Productions. Retrieved from https://archive.org/details/DuckandC1951

Field, S. (2001). Perspectives and elementary social studies: Practice and promise. In O. L. Davis, E. A. Yeager & S. J. Foster. (Eds.), *Historical empathy and perspective taking in the social studies* (pp. 115-138). Lanham, MD: Rowman & Littlefield.

Fischer, G. H., & Molenaar, I. W. (Eds.). (1995). *Rasch models: Foundations, recent developments, and applications*. New York, NY: Springer.

Fogo, B. (2014). Core practices for teaching history: The results of a Delphi panel survey. *Theory & Research in Social Education, 42*, 151-196. doi:10.1080/00933104.2014.902781

Fordham, M. (2013). O brave new world, without those levels in't: Where now for key stage 3 assessment in history? *Teaching History. The Historical Association Curriculum Supplement, 153*, 16-23.

Forsten, C., Grant, J., & Hollas, B. (2002). *Differentiated instruction. Different strategies for different learners*. Peterborough, NH: Crystal Springs Books.

Foster, S. J. (1999). Using historical empathy to excite students about the study of history: Can you empathize with Neville Chamberlain? *The Social Studies, 90*, 18-24. doi:10.1080/00377999909602386

Foster, S. J. (2001). Historical empathy in theory and practice: Some final thoughts. In O. L. Davis, E. A. Yeager, & S. J. Foster (Eds.), *Historical empathy and perspective taking in the social studies* (pp. 167-181). Lanham, MD: Rowman & Littlefield.

Foster, S. J., Ashby, R., & Lee, P. (2008). *Usable historical pasts: A study of students' frameworks of the past: Full research report ESRC end of award report. RES-000-22-1676*. Swindon, United Kingdom: ESRC.

Foster S. J., & Yeager, E. A. (1998). The role of empathy in the development of historical understanding. *International Journal of Social Education*, *13*(1), 1-7.

Foster, S. J., & Yeager, E. A. (1999). "You've got to put together the pieces": English 12-year-olds encounter and learn from historical evidence. *Journal of Curriculum and Supervision*, *14*, 286-317.

G

Gaddis, J. L. (2002). *The landscape of history: How historians map the past*. Oxford, United Kingdom: Oxford University Press.

Gehlbach, H. (2004). Social perspective taking: A facilitating aptitude for conflict resolution, historical empathy, and social studies achievement. *Theory & Research in Social Education, 32*, 39-55. doi:10.1080/00933104.2004.10473242

George, A. L., & Bennett, A. (2004). *Case studies and theory development in the social sciences*. London, United Kingdom: MIT Press.

Gertsen, R., Baker, S. K., Haager, D., & Graves, A. W. (2005). Exploring the role of teacher quality in predicting the reading outcomes for first-grade English learners: An observational study. *Remedial and Special Education*, *26*, 197-206. doi:10.1177/07419325050260040201

Gerwin, D., & Visone, F. (2006). The freedom to teach: Contrasting history teaching in elective and state-tested courses. *Theory and Research in Social Education*, *34*, 259–282. doi:10.1080/00933104.2006.10473307

Gestsdóttir, S. M., Van Boxtel, C. A. M., & Van Drie, J. P. (2017). *An observation instrument of the teaching of historical thinking and reasoning*. Manuscript submitted for publication.

Ginsberg, M. B. (2005). Cultural diversity, motivation, and differentiation. *Theory Into Practice*, *44*, 218-225. doi:10.1207/s15430421tip4403_6

Grant, S. G., & Gradwell, J. M. (2009). The road to ambitious teaching: Creating big idea units in history classes. *Journal of Inquiry and Action in Education*, *2*(1), 1-26.

Grant, S. G., & Gradwell, J. M. (2010). *Teaching with big ideas: Cases of ambitious teaching*. Lanham, MD: Rowman & Littlefield.

Grossman, P. L., Loeb, S., Cohen, J., Hammerness, K., Wyckoff, J. H., Boyd, D. J., & Lankford, H. (2010). *Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores*. Cambridge, MA: National Bureau of Economic Research.

Grossman, P. L., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal*, *45*, 184-205. doi:10.3102/0002831207312906

H

Hall, T. D., & Scott, R. (2007). Closing the gap between professors and teachers: "Uncoverage" as a model of professional development for history teachers. *The History Teacher*, *40*, 257-263.

Hallam, R. (1970). Piaget and thinking in history. In M. Ballard (Ed.), *New movements in the study and teaching of history* (pp. 162-178). Bloomington: Indiana University Press.

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

Halvorsen, A. L., Harris, L. M., Aponte Martinez, G., & Frasier, A. S. (2015). Does students' heritage matter in their performance on and perceptions of historical reasoning tasks? *Journal of Curriculum Studies*, *48*, 457-478. doi:10.1080/00220272.2015.1092585

Harris, R., & Foreman-Peck, L. (2004). 'Stepping into other people's shoes': Teaching and assessing empathy in the secondary history curriculum. *International Journal of Historical Learning, Teaching and Research*, *4*(2), 1-14.

Hartmann, U., & Hasselhorn, M. (2008). Historical perspective taking. A standardized measure for an aspect of students' historical thinking. *Learning and Individual Differences*, *18*, 264–270. doi:10.1016/j.lindif.2007.10.002

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

Havekes, H. G. F., Coppen, P. A. J. M., Luttenberg, J. M., & Van Boxtel, C. A. M. (2012). Knowing and doing history: A conceptual framework and pedagogy for teaching historical contextualisation. *International Journal of Historical Learning, Teaching and Research*, *11*(2), 72–93.

Haydn, T. (2011). Secondary history. Current themes. In I. Davies (Ed.), *Debates in history teaching* (pp. 30-45). New York, NY: Routledge.

Haydn, T., Stephen, A., Arthur, J., & Hunt, M. (1997). *Learning to teach history in the secondary school: A companion to school experience* (1st ed.). New York, NY: Routledge.

Haydn, T., Stephen, A., Arthur, J., & Hunt, M. (2015). *Learning to teach history in the secondary school: A companion to school experience* (4th ed.). New York, NY: Routledge.

Haydn, T., & Counsell, C. (2002). *History, ICT and learning*. London, United Kingdom: Routledge-Falmer.

Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Thousand Oaks, CA: Corwin.

Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Observational systems and a case for the G-study. *Educational Researcher, 41*, 56–64. doi:10.3102/0013189X12437203

Huijgen, T. D., & Holthuis, P. (2015). 'Why am I accused of being a heretic?' A pedagogical framework for stimulating historical contextualisation. *Teaching History, 158*, 56-61.

Huijgen, T. D., Holthuis, P., Van Boxtel, C. A. M., & Van de Grift, W. J. C. M. (2018). Promoting historical contextualization in classrooms: An observational study. *Educational Studies.* Advance online publication. doi:10.1080/03055698.2018.1509771

Huijgen, T. D., Holthuis, P., Van Boxtel, C. A. M., Van de Grift, W. J. C. M., & Suhre, C. J. M. (2018). Students' historical contextualization and the Cold War. *British Journal of Educational Studies*. Advance online publication. doi:10.1080/00071005.2018.1518512

Huijgen, T. D., Van Boxtel, C. A. M., Van de Grift, W. J. C. M., & Holthuis, P. (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. doi:10.1007/s10212-014-0219-4

Huijgen, T. D., Van Boxtel, C. A. M., Van de Grift, W. J. C. M., & Holthuis, P. (2017). Toward historical perspective taking: Students' reasoning when contextualizing the actions of people in the past. *Theory & Research in Social Education, 45*, 110-144. doi:10.1080/00933104.2016.1208597

Huijgen, T. D., Van de Grift, W. J. C. M., Van Boxtel, C. A. M., & Holthuis, P. (2017). Teaching historical contextualization: The construction of a reliable observation instrument. *European Journal of Psychology of Education, 32*, 159-181. doi:10.1007/s10212-016-0295-8

Huijgen, T. D., Van de Grift, W. J. C. M., Van Boxtel, C. A. M., & Holthuis, P. (2018). Promoting historical contextualization: The development and testing of a pedagogy. *Journal of Curriculum Studies, 50*, 410-434. doi:10.1080/00220272.2018.1435724

Husbands, C. T. (1996). *What is history teaching? Language, ideas and meaning in learning about the past*. Buckingham, United Kingdom: Open University Press.

J

Johnson, D. W., & Johnson, R. T. (2009). Energizing learning: The instructional power of conflict. *Educational Researcher, 38*, 37–51. doi:10.3102/0013189X08330540

Jonassen, D., & Grabowski, B. (1993). *Handbook of individual differences, learning and instruction*. Hillsdale, NJ: Erlbaum.

K

Kennedy, J. (1983). Assessing the relationship between information processing capacity and historical understanding. *Theory & Research in Social Education, 11*, 1-22. doi:10.1080/00933104.1983.10505446

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*, 945-980. doi:10.3102/0034654315626800

Kitson, A., Husbands, C. T., & Steward, S. (2011). *Teaching and learning history 11-18: Understanding the past*. Maidenhead, United Kingdom: Open University Press.

Kohlmeier, J. (2006). "Couldn't she just leave?": The relationship between consistently using class discussions and the development of historical empathy in a 9th grade world history course. *Theory & Research in Social Education, 34*, 34-57. doi:10.1080/00933104.2006.10473297

Korthagen, F. (2017). Inconvenient truths about teacher learning: Towards professional development 3.0. *Teachers and Teaching, 23*, 387-405. doi:10.1080/13540602.2016.1211523

Kosso, P. (2009). Philosophy of historiography. In A. Tucker (Ed.), *A companion to the philosophy of history and historiography* (pp. 9-25). Malden, MA: Blackwell.

Krippendorff, K. (2013). *Content analysis: An introduction to its methodology* (3rd ed.). Thousand Oaks, CA: Sage.

Kyriakides, L. (2008). Testing the validity of the comprehensive model of educational effectiveness: A step towards the development of a dynamic model of effectiveness. *School Effectiveness and School Improvement, 19*, 429-446. doi:10.1080/09243450802535208

Kyriakides, L., Campbell, R. J., & Christofidou, E. (2002). Generating criteria for measuring teacher effectiveness through a self-evaluation approach: A complementary way of measuring teacher effectiveness. *School Effectiveness and School Improvement, 13*, 291-325. doi:10.1076/sesi.13.3.291.3426

Kyriakides, L., Christoforou, C., & Charalambous, C.Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education, 36*, 143-152. doi:10.1016/j.tate.2013.07.010

L

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174. doi:10.2307/2529310

Lavigne, A., & Good, T. (2015). *Improving teaching through observation and feedback: Beyond state and federal mandates*. New York, NY: Routledge.

Lee, P. (2005). Putting principles into practice. In M. S. Donovan & J. D. Bransford (Eds.), *How students learn: History, mathematics, and science in the classroom* (pp. 31–78). Washington, DC: The National Academies Press.

Lee, P., & Ashby, R. (2001). Empathy, perspective taking, and rational understanding. In O. L. Davis, E. A. Yeager, & S. J. Foster (Eds.), *Historical empathy and perspective taking in the social studies* (pp. 21–50). Lanham, MD: Rowman & Littlefield.

Lee, P., Dickinson, A., & Ashby, R. (1997). "Just another emperor": Understanding action in the past. *International Journal of Educational Research, 27*, 233–244. doi:10.1016/S0883-0355(97)89731-5

Lee, P., & Howson, J. (2009). "Two out of five did not know that Henry VIII had six wives." Historical literacy and historical consciousness. In L. Symcox & A. H. J. Wilschut (Eds.), *National history standards. The problem of the canon and the future of teaching history* (pp. 211–261). Charlotte, NC: Information Age.

Lei, P. W., Smith, M., & Suen, H. K. (2007). The use of generalizability theory to estimate data reliability in single-subject observational research. *Psychology in the Schools, 44*, 433-439. doi:10.1002/pits.20235

Leinhardt, G., Beck, I. L., & Stainton, C. (1994). *Teaching and learning in history*. Hillsdale, NJ: Erlbaum.

Leinhardt, G., & McCarthy Young, K. (1996). Two texts, three readers: Distance and expertise in reading history. *Cognition and Instruction, 14*, 441–486. doi:10.1207/s1532690xci1404_2

Lévesque, S. (2008). *Thinking historically. Educating students for the twenty-first century*. Toronto, Canada: Toronto University Press.

Levi, P. M. (1988). *The Drowned and the Saved*. London, United Kingdom: Abascus.

Levstik, L. S. (1997). "Any history is someone's history:" Listening to multiple voices from the past. *Social Education, 61*(1), 48-51.

Levstik, L. S. (2001). Crossing the empty spaces: New Zealand adolescents' understanding of perspective-taking and historical significance. In O.L. Davis, E. A. Yeager, & S. J. Foster (Eds.), *Historical empathy and perspective taking in the social studies* (pp. 69–96). Lanham, MD: Rowman & Littlefield.

Levstik, L. S. (2011). Learning history. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 108 -126). New York, NY: Routledge.

Levstik, L. S., & Barton, K. C. (2011). *Doing history: Investigating with children in elementary and middle schools*. New York, NY: Routledge.

Levstik, L. S., & Groth, J. (2002). "Scary thing, being an eighth grader": Exploring gender and sexuality in a middle school US history unit. *Theory & Research in Social Education, 30*, 233-254. doi:10.1080/00933104.2002.10473193

Levstik, L. S., & Smith, D. (1996). "I've never done this before": building a community of historical inquiry in a third-grade classroom. In J. Brophy (Ed.), *Advances in research on teaching* (pp. 85-114). Greenwich, CT: JAI Press.

Levstik, L. S., & Tyson, C. A. (Eds.). (2008). *Handbook of research on social studies education*. New York, NY: Routledge.

Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational Researcher, 35*, 3-14. doi:10.3102/0013189X035003003

Liang, J. (2015). Live video classroom observation: An effective approach to reducing reactivity in collecting observational information for teacher professional development. *Journal of Education for Teaching, 41*, 235-253. doi:10.1080/02607476.2015.1045314

Logtenberg, A. (2012). *Questioning the past. Student questioning and historical reasoning*. (Unpublished doctoral dissertation). University of Amsterdam, Amsterdam.

Logtenberg, A., Van Boxtel, C. A. M., & Van Hout-Wolters, B. A. H. M. (2011). Stimulating situational interest and student questioning through three types of historical introductory texts. *European Journal of Psychology of Education*, *26*, 179-198. doi:10.1007/s10212-010-0041-6

Luft, J. A., & Roehrig, G. H. (2007). Capturing science teachers' epistemological beliefs: The development of the teacher beliefs interview. *Electronic Journal of Science Education*, *11*(2), 38-63.

Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. *British Educational Research Journal*, *29*, 861-878. doi:10.1080/0141192032000137349

M

Macpherson, I., Brooker, R., & Ainsworth, P. (2000). Case study in the contemporary world of research: Using notions of purpose, place, process and products to develop some principles for practice. *International Journal of Social Research Methodology*, *3*, 49-61. doi:10.1080/136455700294923

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

Marcus, A. S. (2005). "It is as it was": Feature film in the history classroom. *The Social Studies*, *96*, 61-67. doi:10.3200/TSSS.96.2.61-67

Marriott, G. (2001). *Observing teachers at work*. Oxford, United Kingdom: Heinemann.

Martell, C. C. (2013). Learning to teach history as interpretation: A longitudinal study of beginning teachers. *The Journal of Social Studies Research*, *37*, 17-31. doi:10.1016/j.jssr.2012.12.001

Martin, D., Maldonado, S.I., Schneider, J., & Smith, M. (2011). *A report on the state of history education. State policies and national programs*. Retrieved from the National History Education Clearinghouse website: http://teachinghistory.org/files/teachinghistory_special_report_2011.pdf

Matsumura, L. C., Garnier, H. E., Slater, S. C., & Boston, M. D. (2008). Toward measuring instructional interactions "at-scale". *Educational Assessment*, *13*, 267-300. doi:10.1080/10627190802602541

Maulana, R., Helms-Lorenz, M., & Van de Grift, W. J. C. M. (2014). Development and evaluation of a questionnaire measuring pre-service teachers' teaching behaviour: A Rasch modelling approach. *School Effectiveness and School Improvement*, *26*, 169-194. doi:10.1080/09243453.2014.939198

Maulana, R., Helms-Lorenz, M., & Van de Grift, W. J. C. M. (2015). Pupils' perceptions of teaching behaviour: Evaluation of an instrument and importance for academic motivation in Indonesian secondary education. *International Journal of Educational Research*, *69*, 98-112. doi:10.1016/j.ijer.2014.11.002

McBride, B. (2004). Data-driven instructional methods: "One-strategy-fits-all" doesn't work in real classrooms. *T.H.E Journal*, *31*(11), 38-40.

McCoy, J., & Ketterlin-Geller, L. (2004). Rethinking instructional delivery for diverse student populations. *Intervention in School and Clinic*, *40*, 88-95. doi:10.1177/10534512040400020401

McCully, A. (2012). History teaching, conflict and the legacy of the past. *Education, Citizenship and Social Justice*, *7*, 145-159. doi:10.1177/1746197912440854

McKenney, S., & Reeves, T. C. (2012). *Conducting educational design research*. London, United Kingdom: Routledge.

Mercer, N. (2010). The analysis of classroom talk: Methods and methodologies. *British Journal of Educational Psychology*, *80*, 1-14. doi:10.1348/000709909X479853

Merriënboer, J. V., & Kirschner, P. A. (2007). *Ten steps to complex learning*. Mahwah, NJ: Erlbaum.

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

Metzger, S. A. (2012). The borders of historical empathy: Students encounter the Holocaust through film. *Journal of Social Studies Research*, *36*, 387-410.

Metzger, S. A., & MacArthur Harris, L. (Eds.), *The Wiley international handbook of history teaching and learning* (pp. 149–176). Hoboken, NJ: Wiley-Blackwell.

Meuwissen, K. W. (2017). "Happy professional development at an unhappy time": Learning to teach for historical thinking in a high-pressure accountability context. *Theory & Research in Social Education*, *45*, 248-285. doi:10.1080/00933104.2016.1232208

Monte-Sano, C. (2011). Learning to open up history for students: Preservice teachers' emerging pedagogical content knowledge. *Journal of Teacher Education*, *62*, 260-272. doi:10.1177/0022487110397842

Monte-Sano, C., & Cochran, M. (2009). Attention to learners, subject, or teaching: What takes precedence as preservice candidates learn to teach historical thinking and reading? *Theory & Research in Social Education*, *37*, 101-135. doi:10.1080/00933104.2009.10473389

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*, 364-386. doi:10.1177/1094428106291059

Muhr, T. (1991). Atlas/ti — A prototype for the support of text interpretation. *Qualitative sociology*, *14*, 349-371. doi:10.1007/BF00989645

Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation*, *12*, 53-74. doi:10.1080/13803610500392236

N

National Center for History in the Schools. (1996). *National standards for history. Basic edition*. Retrieved from https://phi.history.ucla.edu/nchs/history-standards

National Council for the Social Studies. (2013). *The College, Career, and Civic Life (C3) Framework for social studies state standards: Guidance for enhancing the rigor of K–12 civics, economics, geography, and history*. Retrieved from https://media.nationalgeographic.org/assets/file/C3-Framework-for-Social-Studies.pdf

National Curriculum Board. (2009). *National history curriculum: Framing paper*. Retrieved from the Australian Curriculum, Assessment and Reporting Authority website: http://docs.acara.edu.au/resources/National_History_Curriculum_-_Framing_Paper.pdf

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research*, *39*, 374-396. doi:10.1007/s11414-012-9295-x

Netherlands Institute for Curriculum Development. (2016). *Geschiedenis. Vakspecifieke trendanalyse 2016* [History. Subject-specific analysis 2016]. Retrieved from http://downloads.slo.nl/Repository/geschiedenis-vakspecifieke-trendanalyse-2016.pdf

Nikitina, S. (2006). Three strategies for interdisciplinary teaching: Contextualizing, conceptualizing, and problem-centring. *Journal of Curriculum Studies*, *38*, 251-271, doi:10.1080/00220270500422632

Nokes, J. D. (2010). Observing literacy practices in history classrooms. *Theory & Research in Social Education*, *38*, 515–544. doi:10.1080/00933104.2010.10473438

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

O

O'Leary, M. (2014). *Classroom observation: A guide to the effective observation of teaching and learning*. London, United Kingdom: Routledge.

Ontario Ministry of Education. (2015). *The Ontario curriculum grades 11 and 12. Canadian and world studies*. Retrieved from www.edu.gov.on.ca/eng/curriculum/secondary/2015cws11and12.pdf

O'Reilly, K. (1991). Informal reasoning in high school history. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.) *Informal reasoning and education* (pp. 363–379). Hillsdale, NJ: Erlbaum.

Osborne, K. (2006). 'To the past': Why we need to teach and study history. In R. Sandwell (Ed.), *To the past: History education, public memory and citizenship in Canada* (pp. 103-131). Toronto, Canada: University of Toronto Press.

P

Pauly, N. (2003). Interpreting visual culture as cultural narratives in teacher education. *Studies in Art Education*, *44*, 264-284.

Peck, C., & Seixas, P. (2008). Benchmarks of historical thinking: First steps. *Canadian Journal of Education*, *31*, 1015-1038.

Pendry, A., & Husbands, C. T. (2000). Research and practice in history teacher education. *Cambridge Journal of Education*, *30*, 321-334. doi:10.1080/03797720600625275

Perfetti, C. A., Britt, M. A., & Georgi, M. C. (1995). *Text-based learning and reasoning: Studies in history*. Hillsdale, NJ: Erlbaum.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*, 109-119. doi:10.3102/0013189X09332374

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS)*. Baltimore, MD: Brookes.

Pontecorvo, C., & Girardet, H. (1993). Arguing and reasoning in understanding historical topics. *Cognition and Instruction, 11*, 365-395. doi:10.1080/07370008.1993.9649030

R

Rantala, J. (2011). Assessing historical empathy through simulation – How do Finnish teacher students achieve contextual historical empathy? *Nordidactica: Journal of Humanities and Social Science Education, 1*, 58-76.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Reich, G. A. (2009). Testing historical knowledge: Standards, multiple-choice questions and student reasoning. *Theory & Research in Social Education, 37*, 325-360. doi:10.1080/00933104.2009.10473401

Reisman, A. (2012a). The 'Document-Based Lesson': Bringing disciplinary inquiry into high school history classrooms with adolescent struggling readers. *Journal of Curriculum Studies, 44*, 233-264. doi:10.1080/00220272.2011.591436

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

Reisman, A. (2015). Entering the historical problem space: Whole-class text-based discussion in history class. *Teachers College Record, 117*(2), 1-44.

Reisman, A., & Fogo, B. (2016). Contributions of educative document-based curricular materials to quality of historical instruction. *Teaching and Teacher Education, 59*, 191-202. doi:10.1016/j.tate.2016.05.018

Reisman, A., Kavanagh, S. S., Monte-Sano, C., Fogo, B., McGrew, S. C., Cipparone, P., & Simmons, E. (2018). Facilitating whole-class discussions in history: A framework for preparing teacher candidates. *Journal of Teacher Education, 69*, 278-293. doi:10.1177/0022487117707463

Reisman, A., & Wineburg, S. S. (2008). Teaching the skill of contextualizing in history. *The Social Studies, 99*, 202–207. doi:10.3200/TSSS.99.5.202-207

Richardson, V., Anders, P., Tidwell, D., & Lloyd, C. (1991). The relationship between teachers' beliefs and practices in reading comprehension instruction. *American Educational Research Journal, 28*, 559-586. doi:10.3102/00028312028003559

Riley, K. (1998). Historical empathy and the Holocaust: Theory into practice. *International Journal of Social Education, 13(1)*, 32-42.

Ritter, J. K. (2012). Modeling powerful social studies: Bridging theory and practice with preservice elementary teachers. *The Social Studies, 103*, 117-124. doi:10.1080/00377996.2011.596857

Rothstein, R. (2004). We are not ready to assess history performance. *Journal of American History, 90*, 1381-1391. doi:10.2307/3660358

Rouet, J. F., Favart, M., Britt, M. A., & Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction, 15*, 85–106. doi:10.1207/s1532690xci1501_3

Rüsen, J. (2004). How to overcome ethnocentrism: Approaches to a culture of recognition by history in the twenty-first century. *History and Theory, 43*, 118-129. doi:10.1111/j.1468-2303.2004.00301.x

S

Savenije, G. M., Van Boxtel, C. A. M., & Grever, M. C. R. (2014). Sensitive 'heritage' of slavery in a multicultural classroom: Pupils' ideas regarding significance. *British Journal of Educational Studies, 62*, 127-148. doi:10.1080/00071005.2014.910292

Sawada, D., Piburn, M., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics, 102*, 245-253. doi:10.1111/j.1949-8594.2002.tb17883.x

Saye, J., & Brush, T. (2004). Promoting civic competence through problem-based history learning experiments. In G.E. Hamot, J.J. Patrick, & R.S. Leming (Eds)., *Civic Learning in Teacher Education* (Vol. 3, pp. 123-145). Bloomington, IN: The Social Studies Development Center.

Saye, J., & Social Studies Inquiry Research Collaborative. (2013). Authentic pedagogy: Its presence in social studies classrooms and relationships to student performance on state-mandated tests. *Theory & Research in Social Education*, *41*, 89–132. doi:10.1080/00933104.2013.756785

Schoenfeld, A. H. (2013). Classroom observations in theory and practice. *ZDM*, *45*, 607-621. doi:10.1007/s11858-012-0483-1

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*, 454-499. doi:10.3102/0034654307310317

Seixas, P. (1996). Conceptualizing the growth of historical understanding. In D. Olsen & N. Torrance (Eds.), *Handbook of education and human development: New models of learning, teaching, and schooling* (pp. 765–783). Oxford: Blackwell.

Seixas, P. (1998). Student teachers thinking historically. *Theory & Research in Social Education*, *26*, 310–41. doi:10.1080/00933104.1998.10505854

Seixas, P. (2006). *Benchmarks of historical thinking: A framework for assessment in* Canada. Retrieved from The Historical Thinking Project website: http://historicalthinking.ca/sites/default/files/files/docs/Framework_EN.pdf

Seixas, P. (2015). Translation and its discontents: Key concepts in English and German history education. *Journal of Curriculum Studies*, *48*, 427-439. doi:10.1080/00220272.2015.1101618

Seixas, P., & Colyer, J. (2012). *Assessing historical thinking. A report on the national meeting of the historical thinking project*. Retrieved from The Historical Thinking Project website: http://historybenchmarks.ca/sites/default/files/files/docs/2012Report_EN.pdf

Seixas, P., & Morton, T. (2013). *The big six historical thinking concepts*. Toronto, Canada: Nelson.

Seixas, P., & Peck, C. (2004). Teaching historical thinking. In A. Sears & I. Wright (Eds.), *Challenges and prospects for Canadian social studies* (pp. 109-117). Vancouver, Canada: Pacific Educational Press.

SERVE. (2006). *How to assess student performance in history: Going beyond multiple-choice tests*. Retrieved from http://www.serve.org/uploads/publications/AssessHistory.pdf

Sewell Jr., W. H. (2005). *Logics of history: Social theory and social transformation*. Chicago, IL: University of Chicago Press.

Shadish, W.R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

Shavelson, R. J., Webb, N. M., & Rowley, G. L. (1989). Generalizability theory. *American Psychologist*, *44*, 922-932. doi:10.1037/0003-066X.44.6.922

Shemilt, D. (1983). The devil's locomotive. *History and Theory*, *22*(4), 1–18. doi:10.2307/2505213

Shemilt, D. (1987). Adolescent ideas about evidence and methodology in history. In C. Portal (Ed.), *The history curriculum for teachers* (pp. 39–61). London, United Kingdom: The Falmer Press.

Shemilt, D. (2009). Drinking an ocean and pissing a cupful: How adolescents make sense of history. In L. Symcox & A. H. J. Wilschut (Eds.), *National history standards: The problem of the canon and the future of teaching history* (pp. 141–209). Charlotte, NC: Information Age.

Skolnick, J., Dulberg, N., & Maestre, T. (2004). *Through other eyes: Developing empathy and multicultural perspectives in the social studies*. Toronto, Canada: Pippin.

Slavin, R. (1996). *Education for all*. Lisse, The Netherlands: Swets & Zeitlinger.

Smit, N., Van de Grift, W. J. C. M., De Bot, C. L. J., & Jansen, E. P. W. A. (2017). A classroom observation tool for scaffolding reading comprehension. *System*, *65*, 117-129. doi:10.1016/j.system.2016.12.014

Smith, M. D. (2018). New multiple-choice measures of historical thinking: An investigation of cognitive validity. *Theory & Research in Social Education*, *46*, 1-34. doi:10.1080/00933104.2017.1351412

Spoehr, K. T., & Spoehr, L. W. (1994). Learning to think historically. *Educational Psychologist*, *29*, 71–77. doi:10.1207/s15326985ep2902_2

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*, 430–456.

Stallings, J. A., & Kaskowitz, D. H. (1974). *Follow through classroom observation evaluation, 1972-1973*. Menlo Park, CA: Stanford Research Institute.

Statistics Netherlands. (2012). Statline database VO [Statline database secondary education]. Retrieved from statline.cbs.nl/Statweb

Statistics Netherlands. (2014). Statline database VO [Statline database file secondary education]. Retrieved from statline.cbs.nl/Statweb

Statistics Netherlands. (2016). Statline database VO [Statline database secondary education]. Retrieved from statline.cbs.nl/Statweb

Stearns, L.M., Morgan, J., Capraro, M. M., & Capraro, R. M. (2012). A teacher observation instrument for PBL classroom instruction. *Journal of STEM Education: Innovations and Research*, *13*(3), 7-16.

Stern, B. S. (1998). Addressing the concept of historical empathy: Wit frauen: German women recall the third reich. *International Journal of Social Education*, *13*(1), 43-48.

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education*, *17*, 213-226. doi:10.1016/S0742-051X(00)00052-4

Stoel, G. L., Van Drie, J. P., & Van Boxtel, C. A. M. (2015). Teaching towards historical expertise. Developing a pedagogy for fostering causal reasoning in history. *Journal of Curriculum Studies*, *47*, 49-76. doi: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*, 321-337. doi:10.1037/edu0000143

Stradling, R. (2003). *Multiperspectivity in history teaching: A guide for teachers*. Strasbourg, France: Council of Europe.

Subban, P. (2006). Differentiated instruction: A research basis. *International Education Journal*, *7*, 935–947.

Suen, H. K., & Ary, D. (2014). *Analyzing quantitative behavioral observation data*. New York, NY: Psychology Press.

Symcox, L., & Wilschut, A. H. J. (2009). *National history standards: The problem of the canon and the future of teaching history*. Charlotte, NC: Information Age.

T

Tashakkori, A., & Teddlie, C. (Eds.). (2010). *Sage handbook of mixed methods in social & behavioral research*. Thousand Oaks, CA: Sage.

Teddlie, C., Creemers, B., Kyriakides, L., Muijs, D., & Yu, F. (2006). The International System for Teacher Observation and Feedback: An evolution of an international study of teacher effectiveness constructs. *Educational Research and Evaluation*, *12*, 561-582. doi:10.1080/13803610600874067

Tomlinson, C. A. (2002). Different learners different lessons. *Instructor, 112*(2), 21-25.

Tomlinson, C. A., & Kalbfleisch, M. (1998). Teach me, teach my brain: A call for differentiated classrooms. *Educational Leadership, 56*(3), 52-55.

Trautwein, U., Bertram, C., Von Borries, B., Brauch, N., Hirsch, M., Klausmeier, K., … Zuckowski, A. *Kompetenzen historischen denkens erfassen. Konzeption, operationalisierung und befunde des projekts "Historical thinking - competencies in history" (HiTCH)* [Capturing historical thinking competencies. Conception, operationalization and findings of the project "Historical thinking - competencies in history" (HiTCH)]. Münster, Germany: Waxmann.

Tuithof, J. I. G. M. (2017). *The characteristics of Dutch experienced history teachers' PCK in the context of a curriculum innovation* (Unpublished doctoral dissertation). Utrecht University, Utrecht.

Tully, J. (Ed.). (1988). *Meaning and context: Quentin Skinner and his critics*. Princeton, NJ: Princeton University Press.

V

Van Boxtel, C. A. M., & Grever, M. C. R. (2011). Between disenchantment and high expectations. History education in the Netherlands, ca. 1968–2008. In E. Erdmann & W. Hasberg (Eds.), *Facing—Mapping—Bridging diversity. Foundation of a European discourse on history education* (pp. 83–116). Schwalbach, Germany: Wochenschau-Verlag.

Van Boxtel, C. A. M., & Van Drie, J. P. (2004). Historical reasoning: A comparison of how experts and novices contextualize historical sources. *International Journal of Historical Learning, Teaching and Research*, *4*(2), 89–97.

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

Van Boxtel, C. A. M., & Van Drie, J. P. (2016). Redesigning history education to improve pupils' understanding: Implications for theory and research. In C. Counsell, K. Burn, & A. Chapman (Eds.), *Masterclass in history education: Transforming teaching and learning* (pp. 201-208). London, United Kingdom: Bloomsbury.

Van Boxtel, C. A. M., & Van Drie, J. P. (2017). Engaging students in historical reasoning: The need for dialogic history education. In M. Carretero, S. Berger, & M. C. R. Grever (Eds.), *Palgrave handbook of research in historical culture and education* (pp. 573-589). London, United Kingdom: Palgrave Macmillan.

Van Boxtel, C. A. M., & Van Drie, J. P. (2018). Historical reasoning: Conceptualizations and educational applications. In S. A. Metzger & L. MacArthur Harris (Eds.), *The Wiley international handbook of history teaching and learning* (pp. 149–176). Hoboken, NJ: Wiley-Blackwell.

Van de Grift, W. J. C. M. (2007). Quality of teaching in four European countries: A review of the literature and an application of an assessment instrument. *Educational Research*, *49*, 127-152. doi:10.1080/00131880701369651

Van de Grift, W. J. C. M. (2009). Reliability and validity in measuring the value added of schools. *School Effectiveness and School Improvement, 20*, 269-285. doi:10.1080/09243450902883946

Van de Grift, W. J. C. M., Helms-Lorenz, M., & Maulana, R. (2014). Teaching skills of student teachers: Calibration of an evaluation instrument and its value in predicting student academic engagement. *Studies in Educational Evaluation*, *43*, 150-159. doi:10.1016/j.stueduc.2014.09.003

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

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

Van Hover, S., Hicks, D., & Cotton, S. (2012). "Can you make 'historiography' sound more friendly?": Towards the construction of a reliable and validated history teaching observation instrument. *History Teacher, 45*, 603-612.

Van Hover, S., & Yeager, E. A. (2004). Challenges facing beginning history teachers: An exploratory study. *International Journal of Social Education*, *19*(1), 8-21.

VanSledright, B. A. (2001). From empathetic regard to self-understanding: Im/positioning, empathy, and historical contextualization. In O. L. Davis, E. A. Yeager, & S. J. Foster (Eds.), *Historical empathy and perspective taking in the social studies* (pp. 51-68). Lanham, MD: Rowman & Littlefield.

VanSledright, B. A. (2002). Fifth graders investigating history in the classroom: Results from a researcher-practitioner design experiment. *The Elementary School Journal*, *103*, 131-160. doi:10.1086/499720

VanSledright, B. A. (2008). Narratives of nation-state, historical knowledge, and school history education. *Review of Research in Education*, *32*, 109–146. doi:10.3102/0091732X07311065

VanSledright, B. A. (2010). What does it mean to think historically…and how do you teach it? In W. C. Parker (Ed.), *Social studies today: Research and practice* (pp. 112–120). New York, NY: Routledge.

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

VanSledright, B. A. (2013). *Assessing historical thinking and understanding. Innovative designs for new standards*. New York, NY: Taylor & Francis.

VanSledright, B. A., & Afflerbach, P. (2000). Reconstructing Andrew Jackson: Prospective elementary teachers' readings of revisionist history texts. *Theory & Research in Social Education*, *28*, 411–444. doi:10.1080/00933104.2000.10505915

VanSledright, B. A., Kelly, T., & Meuwissen, K. (2006). Oh, the trouble we've seen: Researching historical thinking and understanding. In K. C. Barton (Ed.), *Research methods in social studies education: Contemporary issues and perspectives* (pp. 207-233). Greenwich, CT: Information Age.

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). Mahwah, NJ: Erlbaum.

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. London, United Kingdom: Academic Press.

Vaughn, S., Schumm, J. S., & Sinagub, J. M. (1996). *Focus group interviews in education and psychology*. Thousand Oaks, CA: Sage.

Virja, A., & Kouki, E. (2014). Dimensions of historical empathy in upper secondary students' essays. *Nordidactica: Journal of Humanities and Social Science Education*, *2*, 137-160.

Virta, A. (2002). Becoming a history teacher: Observations on the beliefs and growth of student teachers. *Teaching and Teacher Education*, *18*, 687-698.

Virta, A. (2007). Historical literacy: Thinking, reading and understanding history. *Journal of Research in Teacher Education*, *14*(4), 11-25.

Von Borries, B. (1994). (Re-)Constructing history and moral judgment: On relationships between interpretations of the past and perceptions of the present. In M. Carretero & J. F. Voss (Eds.), *Cognitive and instructional processes in history and the social sciences* (pp. 285–308). Hillsdale, NJ: Erlbaum.

W

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. doi:10.1016/j.tate.2016.02.005

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655-684. doi:10.1111/1467-8624.00304

Wils, K., & Verschaffel, T. (2012). Longing for the present in the history of history education. *Paedagogica Historica*, *48*, 793-799. doi:10.1080/00309230.2012.739182

Wilschut, A. H. J. (2012). *Images of time: The role of a historical consciousness of time in learning history*. Charlotte, NC: Information Age.

Wineburg, S. 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*, 73-87. doi:10.1037/0022-0663.83.1.73

Wineburg, S. S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science*, *22*, 319–346. doi:10.1207/s15516709cog2203_3

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

Wineburg, S. S., & Fournier, J. (1994). Contextualized thinking in history. In M. Carretero & J. F. Voss (Eds.), *Cognitive and instructional processes in history and the social sciences* (pp. 285–308). Hillsdale, NJ: Erlbaum.

Witte, T. C. H. (2008). *Het oog van de meester. Een onderzoek naar de literaire ontwikkeling van havo- en vwo-leerlingen in de tweede fase van het voortgezet onderwijs* [The eye of the master: A study of students' literary development in upper secondary education]. Delft, The Netherlands: Eburon.

Wooden, J. A. (2008). 'I had always heard Lincoln was a good person, but..' A study of sixth graders' reading of Lincoln's views on black-white relations. *The Social Studies*, *99*, 23-31. doi:10.3200/TSSS.99.1.23-32

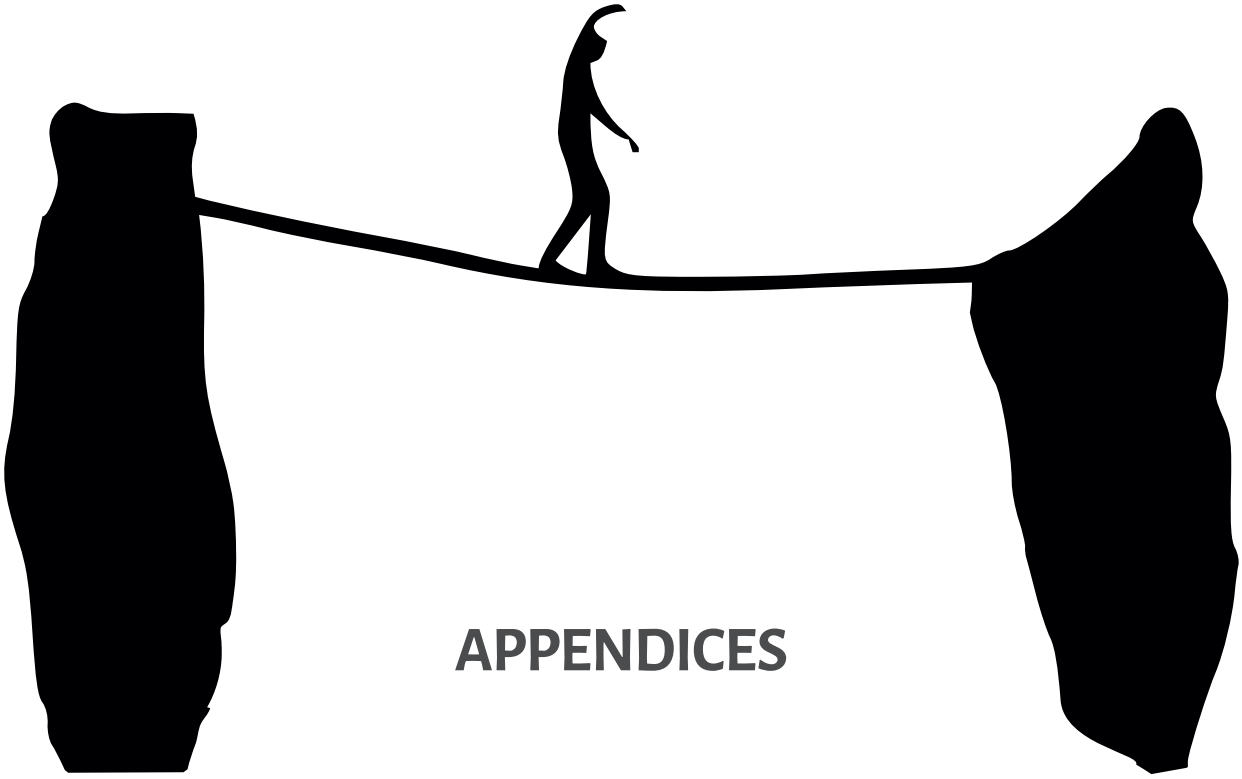
Wragg, T. (1994). *An introduction to classroom observation*. New York, NY: Routledge.

Y

Yeager, E. A., & Foster, S. J. (2001). The role of empathy in the development of historical understanding. In O. L. Davis, E. A. Yeager, & S. J. Foster (Eds.), *Historical empathy and perspective taking in the social studies* (pp. 21–50). Lanham, MD: Rowman & Littlefield.

Yilmaz, K. (2007). Historical empathy and its implications for classroom practices in schools. *The History Teacher, 40*, 331-337.

Yoder, P. J., & Symons, F. J. (2010). *Observational Measurement of Behavior*. New York, NY: Springer.



APPENDICES

Appendix A: The Nazi Party instrument

Dusseldorf, Germany in 1930. Hannes (20 years old) is the son of a man who owns a small factory that makes handmade shoes. One day Hannes meets with his friend Gerd. They talk about the situation in Germany and the upcoming elections. Hannes says: "My father's company might close down. Since the war ended, everything is getting worse and worse. After the economic crisis of 1923, we began to feel some hope again. But now it is worse than ever. I don't know how this is going to end. Right now, I still have a job in my father's business. But when he closes down, I have no idea where to get a job. We have always been wealthy people – and look at us now!". 'Gerd replies. "You are right. What has happened to our country? Look at what is going on today. No one has work. Hannes replies: "My father always says that we were better off during the time of the German Empire. What can we do if our country is suffering from a crisis and the winners of the war are hurting us wherever they can? Our politicians are not decisive and do us no good. It's time that Germany is ruled by someone who knows what he is doing and who really takes the lead. During the last election, I supported the German Democratic Party, but I do not know if they have got the right people to save our country."

Below you will find some statements. Read through all the statements first. Then, try to take Hannes' perspective and mark for every statement how well it fits his situation. Could Hannes vote for an anti-democratic party like the NSDAP?

Item		Doesn't fit his situation at all	Doesn't fit his situation too much	Fits his situation somewhat	Fits his situation very well
POP 1	He will definitely not vote for the NSDAP. No one can approve what this party has done to the world.	☐	☐	☐	☐
POP 2	He will see that only in a democracy can people take part in decision-making. That is why he will decide wisely and that does not mean NSDAP.	☐	☐	☐	☐
POP 3	He will not vote for the NSDAP. Their ideas are easy to see through. It is clear that this party wants a war.	☐	☐	☐	☐
ROA 1	As a member of a wealthy family, he would like to return to the German Empire, where his family was better off. He can vote on an anti-democratic party.	☐	☐	☐	☐
ROA 2	As son of a businessman, he would likely vote for a party that strives to keep things as they are, but not necessarily for the NSDAP.	☐	☐	☐	☐
ROA 3	Because his father's business is almost bankrupt, he could vote for a party that protects small business owners.	☐	☐	☐	☐
CONT 1	Hannes has little experience with democracy. He probably does not know the risk that the NSDAP entails and thus will probably vote for the NSDAP.	☐	☐	☐	☐
CONT 2	To him, Hitler probably represents a strong leader. He probably would not think too much about the threats connected with the NSDAP.	☐	☐	☐	☐
CONT 3	In his situation, he only sees the disadvantages of democracy. Therefore, he might fall for the ideas of the NSDAP.	☐	☐	☐	☐

Note. POP = present-oriented perspective, ROA = role of the historical agent, and CONT = historical contextualization.

Appendix B: The slavery instrument

Harry Knox, a journalist working for the respected American newspaper *Austin Press*, interviewed in 1891 the 70-year-old Ben Simpson. The enslaved Simpson worked for over 20 years at an American plantation in Texas and told Knox the following story about his life:

> "The plantation-owner was in charge of a large plantation. When he pulled me and the others of the boat, he chained us around our necks. The chains were fixed to the horses. With the chains we—my mother, my sister Emma, I and the other slaves had to walk all the way to his plantation in Texas. Somewhere along the way it started to snow, but the plantation-owner did not care about our bare feet. We had to sleep in the snow on the ground. The plantation-owner had a long whip, made of leather. And if one of us fell behind, then he would hit him with it. We had no tents. When the night came, he fixed our chains to a tree. The ground was our bed. A little raw meat and corn were the only things we ate. Often I ate weeds and I was very hungry. He let us never eat during the day and forced us to walk the whole day without any breaks. He branded me, my mother and my sister. At the border of Texas my mother couldn't go any further. Her feet were broken and bleeding, her legs were swollen. The master took his gun and shot her. He didn't burry her, he left her lying where he had shot her."

Which statements can be useful in order to describe the relationship between slaves and plantation owners in 19th-century America?

Item		Not useful	Not too useful	Somewhat useful	Very useful
POP 1	The plantation owner committed a severe crime. He should be arrested and brought to trial.	☐	☐	☐	☐
POP 2	Slaves are also humans. That is why the plantation owner should treat them right.	☐	☐	☐	☐
POP 3	Food is a primary vital need. This is why the plantation owner should provide the slaves with enough food.	☐	☐	☐	☐
ROA 1	Simpson was an eye-witness. He told his story to the *Austin Daily*, because he thought it was a reliable newspaper.	☐	☐	☐	☐
ROA 2	Simpson was the son of a woman who was killed by a plantation owner. That's why he might not be objective in his story.	☐	☐	☐	☐
ROA 3	Because Simpson was an enslaved man who suffered terribly, he probably only remembered the bad things.	☐	☐	☐	☐
CONT 1	The plantation owners were strict, because they wanted the slaves to work harder in order to increase profit.	☐	☐	☐	☐
CONT 2	The plantation owners were strict, because they feared rebellion among the slaves.	☐	☐	☐	☐
CONT 3	Plantation owners saw slaves as economic products and not as people. They thought: if a product is broken, I am buying a new one.	☐	☐	☐	☐

Note. POP = present-oriented perspective, ROA = role of the historical agent, and CONT = historical contextualization.

Appendix C: Prior knowledge test (Germany between 1900 – 1950)

Answer these four questions about the history of Germany during the period 1900 to 1950. Circle the letter that corresponds to the correct answer.

1. What were the years of the First World War?

A. 1910–1914

B. 1912–1916

C. 1914–1918

D. 1916–1920

2. In what year did Hitler become the political leader of Germany?

A. 1931

B. 1933

C. 1935

D. 1939

3. In what year did the New York Wall Street crisis occur?

A. 1923

B. 1925

C. 1927

D. 1929

4. In what year was the Treaty of Versailles signed?

A. 1914

B. 1918

C. 1919

D. 1923

Appendix D: The Framework for Analyzing the Teaching of Historical Contextualization (FAT-HC)

Explanatory notes 1: weak, 2: more weak than strong, 3: more strong than weak, and 4: strong

	The teacher…	**1**	**2**	**3**	**4**
1.	Activates relevant prior knowledge				
2.	Shows visual material				
3.	Uses historical sources				
4.	Gives time indicators regarding phenomena				
5.	Gives the duration of phenomena				
6.	Shows phenomena on a timeline				
7.	Gives geographical/spatial indicators regarding phenomena				
8.	Shows phenomena on a geographical map				
9.	Appoints political/governance characteristics at the time of phenomena				
10.	Appoints economic characteristics at the time of phenomena				
11.	Appoints socio-cultural characteristics at the time of phenomena				
12.	Appoints causes and consequences of phenomena				
13.	Appoints change and continuity regarding phenomena				
	The students…	**1**	**2**	**3**	**4**
14.	Give time indicators regarding phenomena				
15.	Give the duration of phenomena				
16.	Give geographical/spatial indicators regarding phenomena				
17.	Appoint political/governance characteristics at the time of phenomena				
18.	Appoint economic characteristics at the time of phenomena				
19.	Appoint socio-cultural characteristics at the time of phenomena				
20.	Appoint socio-cultural characteristics at the time of phenomena				
21.	Appoint change and continuity regarding phenomena				
	The teacher…	**1**	**2**	**3**	**4**
22.	Centralizes a historical agent				
23.	Moves self into the past to explain phenomena (if I..)				
24.	Outlines a recognizable role for students to foster empathy (as a businessman / like a father)				
	The students…	**1**	**2**	**3**	**4**
25.	Make affective / emotional connections with historical agents				
26.	Consider the role of the historical agent to explain historical decisions				
27.	State what they would have decided regarding historical decisions				
	The teacher…	**1**	**2**	**3**	**4**
28.	Compares phenomena with other times				
29.	Compares phenomena with other places				
30.	Places phenomena in long-term developments				
31.	Outlines phenomena from different perspectives				

	The students…	**1**	**2**	**3**	**4**
32.	Compare phenomena with other times				
33.	Compare phenomena with other places				
34.	Place phenomena in long-term developments				
35.	Outline phenomena from different perspectives				
	The teacher…	**1**	**2**	**3**	**4**
36.	Does *not* use anachronisms				
37.	Does *not* present the past as progress				
38.	Creates historical tension (the past as different)				
39.	Presents conflicting historical sources				
40.	Presents learning strategies for historical contextualization				

Appendix E: Pre- and post-test example items (historical contextualization)

Item (pre-test)

Instruction: Read the following historical source describing a day program of the games in Ancient Rome.

"The gladiator fighting constituted the highlight of the day. First, the hunters demonstrated their expertise with different weapons. In the afternoon, prisoners were thrown to the wild animals. After that, adventurous gladiators began fighting. If there was no decisive victor in the fight, the people who witnessed the fight could decide which gladiator might live."

Choose the statement that best matches this source:

O People should not have the power to decide on life and death.

O Gladiator games were common entertainment for the Roman people.

Item (pre-test)

Instruction: Read the following historical source about marriage in the Middle Ages.

"In the Middle Ages, girls were sometimes married at the age of eleven. The family arranged the marriage. After her marriage, her inheritance was automatically transferred to her husband. Therefore, knights often sought a rich heir."

Choose the statement that best matches this source:

O Women have the right to choose their own husbands.

O These were marriages of convenience that often did not involve much love.

Item (post-test)

Instruction: Read the following historical source about Roman Emperor Nero and the fire of Rome.

"To suppress the rumor that the fire was lit on [imperial] command, Emperor Nero blamed a group of Christians and subjected them to the most ingenious punishments. A huge mass of people was sentenced—not because of the crime of arson but because of hatred towards humanity. In addition, their dying was coupled with scorn: they were, for example, covered with wild animals hides and torn apart by dogs or nailed to crosses."

Choose the statement that best matches this source:

O The Romans were afraid of the Christians and tried to suppress them.

O Everyone is entitled to religious freedom, and therefore, Nero violated the law.

Item (post-test)

Instruction: Read the following historical source about punishments on ships in the late Middle Ages.

"Keelhauling is a punishment that could be imposed by a ship captain on crew members. With a rope, the person was dragged under the ship. Because the ship's hull was always covered with shells, keelhauling caused severe injuries to the victim."

Choose the statement that best matches this source:

O The captain's authority on a ship is holy and the law.

O A court must pronounce the punishment instead of the captain.

Appendix F: Guiding questions for reconstructing a historical context

Socio-political context

1. Was there a government?
2. What kind of governance was present (democracy / dictatorship / monarchy / aristocracy /oligarchy)?
3. Which political parties existed (liberalism / socialism / confessionals)?
4. Who had political power?
5. Did the country have colonies?
6. Who could participate in the political process?
7. Was there a central authority?
8. Was there any military / political conflict?
9. Was there separation of political powers (executive, judicial, and legislative)?
10. Was there separation between church and state?

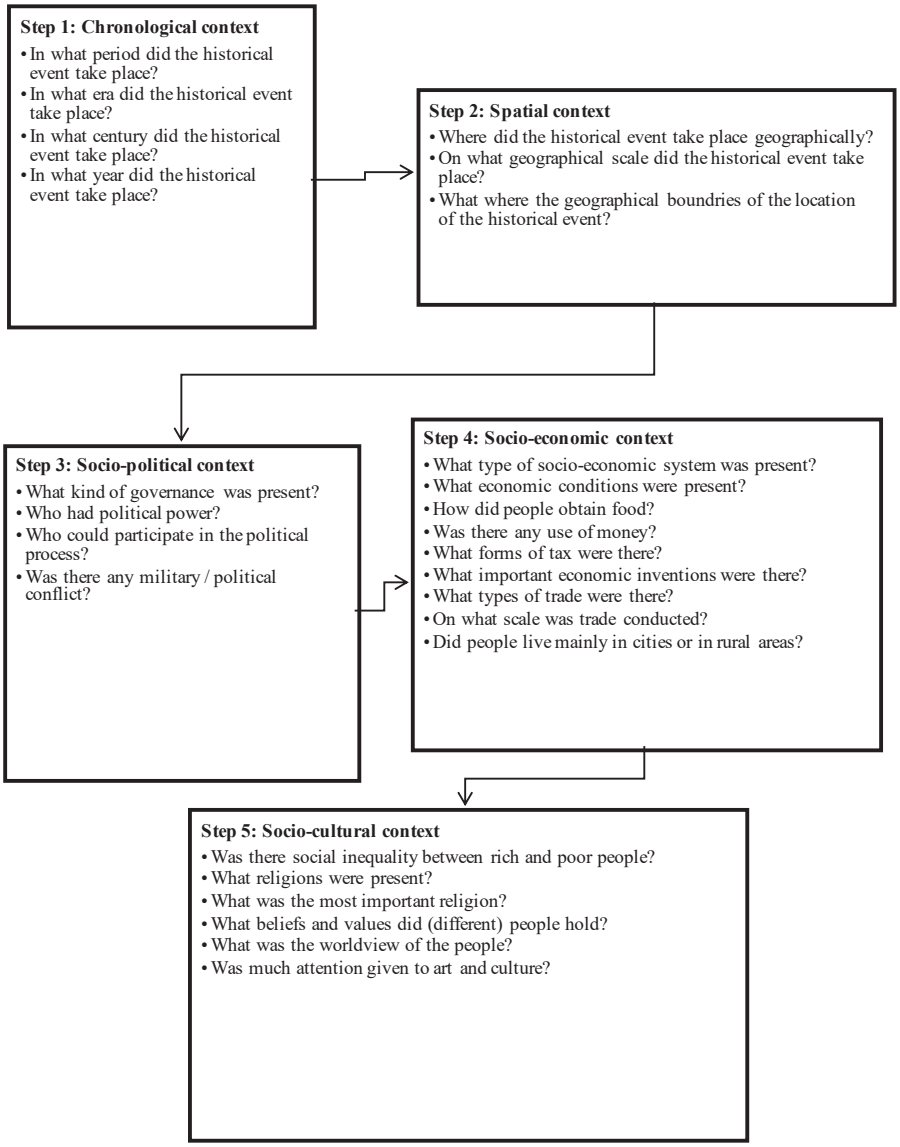
Socio-economic context

1. What type of socio-economic system was present (agricultural / agricultural-urban / industrial)?
2. Which economic conditions were present (prosperity / crises / famine)?
3. What kind of economy was present (self-sufficient, free trade / protection)?
4. Were there factories?
5. What forms of tax were there?
6. Who had to pay taxes?
7. Which economic inventions were there?
8. What types of trade were there, and on what scale was trade driven?
9. Which economic sectors existed (agriculture / industry / services)?
10. Did people live mainly in cities or in the countryside (urbanization / suburbanization)?

Socio-cultural context

1. Was there social inequality between people (grades / positions / wealth / poverty)?
2. Which religions were allowed / suppressed?
3. Was there censorship / freedom of expression?
4. Which freedoms did people have?
5. What role did faith play?
6. Was the society multicultural?
7. What did people believe in?
8. What was the worldview of the people?
9. Were there many scientific discoveries?
10. Was there much attention given to art and culture?

Appendix G: Guiding questions to reconstruct a historical context



Appendix H: Historical contextualization test

Item 1

Dusseldorf, Germany in 1930. Hannes (20 years old) is the son of a man who owns a small factory that makes handmade shoes. One day Hannes meets with his friend Gerd. They talk about the situation in Germany and the upcoming elections. Hannes says: 'My father's company might close down. Since the war ended, everything is getting worse and worse. After the economic crisis of 1923, we began to feel some hope again. But now, it is worse than ever. I don't know how this is going to end. Right now, I still have a job in my father's business. But when he closes down, I have no idea where to get a job. We have always been wealthy people—and look at us now!' Gerd replies. 'You are right. What has happened to our country? Look at what is going on today. No one has work. Hannes replies: 'My father always says that we were better off during the time of the German Empire. What can we do if our country is suffering from a crisis and the winners of the war are hurting us wherever they can? Our politicians are not decisive and do us no good. It's time that Germany is ruled by someone who knows what he is doing and who really takes the lead.

Based on the information in this text, explain how likely it is that Hannes will vote for Hitler's political party: the Nazi Party.

Item 2

Explain what else you should know to answer correctly as to how likely it is that Hannes will vote for the Nazi Party.

Item 3

Read the text about Hannes again. The Nazi Party is Hitler's political party.

Two statements:

I. Hannes will not vote on the NSDAP. Nobody can approve what this party has done to the world.

II. Hannes will vote on the NSDAP. Hitler is in his eyes a strong leader.

Try to take Hannes' perspective and choose the statement that suits his situation best.

I do choose statement number: ..... because ...

I do not choose statement number: ..... because ...

Item 4

Harry Knox, a journalist working for the respected American newspaper Austin Press, interviewed in 1891 the 70-year-old Ben Simpson. The enslaved Simpson worked for over 20 years at an American plantation in Texas and told Knox the following story about his life:

"The plantation-owner was in charge of a large plantation. When he pulled me and the others of the boat, he chained us around our necks. The chains were fixed to the horses. With the chains we—my mother, my sister Emma, I and the other slaves had to walk all the way to his plantation in Texas. We had to sleep in the snow on the ground. The plantation-owner had a long whip, made of leather. And if one of us fell behind, then he would hit him with it. When the night came, he fixed our chains to a tree. The ground was our bed. At the border of Texas my mother couldn't go any further. Her feet were broken and bleeding, her legs were swollen. The master took his gun and shot her. He didn't bury her, he left her lying where he had shot her."

Two statements accompanying the text:

I. The plantation owner committed a crime. He should have been arrested by the police and brought to trial.
II. Slaveholders saw slaves as products. They thought: If a product brakes, you just buy a new product.

Which statement suits the events described by Simpson the best?

I do choose statement number: ..... because ...

I do not choose statement number: ..... because ...

Item 5

Sophie reads in her history textbook:

"Until the 1950s, women in the Netherlands automatically lost their jobs as soon as they married. For female officials this was legally laid down. Also, rape was not punishable when this happened by the spouse."

Sophie responds to this text by saying that the people were then retarded.

Do you agree with Sophie's statement?

Circle your choice and explain your answer:

I *agree* / *disagree* with this statement made by Sophie because ...

Item 6

David reads in his history textbook:

"In the 16th century witch hunting took place in northern Netherlands. The women who were suspected of witchcraft were immediately sentenced to death without trial, for example by drowning or the pillar of fire."

David responds to this text by saying that these witch hunts were not so strange.

Do you agree with David's statement?

Circle your choice and explain your answer:

I *agree* / *disagree* with this statement made by David because ...

Appendix I: Example cases to promote cognitive conflicts

Case #1: Bloodletting Look at the ancient Greek painting on a vase. The physician is bleeding a patient. Bloodletting is the withdrawal of blood from a person to prevent or cure illness and disease. How could you cure people with these manners? Link: tinyurl.com/ycfx887l
Case #2: Medieval Marriages If you were a young female living in Europe in the Middle Ages, it was very normal that the parents arranged your marriage. It was all about gaining economic profit. You did not have any choice. When you reached the age of 12 your marriage could be arranged. Link: tinyurl.com/y9gwrchm
Case #3: Drinking a Beer Nowadays you have to be 18 years old to drink a beer in the Netherlands. However, in the Middle Ages even very young Dutch children drank a beer regularly, even at breakfast. Moreover, the average consumption of beer was around 300 litres of beer a year. Did these people not know any better? Link: tinyurl.com/ya6y3b62
Case #4: The Dutch Trio Read the descriptions of the lives of Jan Pieterszoon Coen, Joannes van Heutsz, and Michiel de Ruyter. Coen and Van Heutsz murdered many people in the Dutch East Indies and De Ruyter recaptured Fort Elmina in Africa to continue the Dutch slave trade. Do these people earn statues in the Netherlands? Link: tinyurl.com/y965eujk
Case #5: Bib Hill No1 Look at the picture taken in the beginning of the 20th century. The two children are very young and working barefoot(!) on the spinning machine. These children often had to work all day long and could easily get hurt. Should the owner of the factory not be arrested for putting such young children at work? Link: tinyurl.com/yaz6ln6j
Case #6: The Swimwear Police View the picture. The beach patrol was measuring bathing suit length in 1922 in the USA. If the bathing suit was too short, a woman was asked to leave the beach. Is that not ridiculous? Should women not decide for themselves what to wear? Link: tinyurl.com/y9k6um97
Case #7: Warner Bros Cartoons View the cartoon (7.09 min) made by Warner Bros in the 1930s. Is this not very racist? Why would somebody make such video's? Should these videos not be banned? Link: tinyurl.com/ybkerghu
Case #8: The Gas This picture was taken at the Empire Pool in London on August 21, 1938. You can see a father, mother, and two children wearing bathing suits but also a gas mask. Are these people crazy? Why would you go swimming wearing a gas mask? Link: tinyurl.com/y7vjq2tl
Case #9: The Guillotine Study the picture. You can see the last public execution by the guillotine in 1939 in the French Republic. Do you see the large crowd watching this execution? Can you imagine yourself as one of them? Link: tinyurl.com/y82zahxg

Case #10: The Tendeloo Resolution

The Dutch law stated until 1956 that females working for the government lost their jobs when they got married. Moreover, all women had to ask their husbands for permission to travel and to buy new clothes. As a Dutch magazine noted in 1949: "For a husband it is very difficult to stay home all day to pay the milkman and baker. He has something better to do." Were these people stupid?

Link: tinyurl.com/ybh6wjc6

Case #11: The Comic Book Ban

Read the different strips. That is racism, is it not? The black people are depicted as stupid and European people as civilized. Should these comic books not be banned? Is it normal that you can buy these books in shops? What will young readers think when they read these comic books?

Link: tinyurl.com/yat4gjeb

CURRICULUM VITAE

Tim Huijgen (1985) was raised in a family of educators. His mother worked as an elementary school teacher and as coach for special educational need students. His father worked as a geography teacher before becoming an educational advisor in primary and secondary education. Between 1998 and 2004, Tim attended pre-university education at the Drachtster Lyceum. In 2004, he started with his Bachelor of Arts in History at the University of Groningen which he completed in 2007. Between 2007 and 2009, he finished two masters at the University of Groningen: A Master of Arts in History and a Master of Arts in Education in the field of history and civics. He graduated his Master of Arts in Education with honor.

In 2008, Tim started to work as a history teacher at the Stadslyceum (a secondary school for senior general secondary and pre-university education) in Groningen. In 2010, he also began to work as an educational textbook author and as a course development officer at the University of Groningen. Between 2011 and 2012, he was the coordinator of the Educational Minor program. Besides this function, he also started to work as a history teacher educator as a colleague of dr. Paul Holthuis. In 2011, Tim became a certified teacher educator of the Dutch Association for Teacher Educators and and he obtained his University Teaching Qualification in 2012.

In august 2012, his PhD research proposal on historical contextualization was awarded a grant from the Netherlands Organisation for Scientific Research (NWO) and he became a PhD candidate at the Department for Teacher Education of the University of Groningen, while he continued to work as a history teacher educator and as a secondary school history teacher.

Tim has published articles on the teaching and learning of history in international scientific and professional journals and presented his work in international and national context. He was also a member of different review and audit panels (e.g., Netherlands Institute for Curriculum Development, Netherlands Organisation for Scientific Research, Accreditation Organisation of the Netherlands and Flanders). Moreover, he is a reviewer for journals such as *Education Inquiry*, *Historical Encounters*, and *Theory & Research in Social Education*.

Currently, Tim is working as a history teacher educator and researcher at the Department for Teacher Education of the University of Groningen. His research interests focus on teaching and learning history. He is also still working as a secondary school history teacher at the Stadslyceum in Groningen. In all of his work, Tim tries to combine educational experience and theory.

Links:

- rug.nl/staff/t.d.huijgen
- researchgate.net/profile/Tim_Huijgen
- linkedin.com/in/timhuijgen

PHD RESEARCH OUTPUT

Scientific publications

Huijgen, T. D., Holthuis, P., Van Boxtel, C. A. M., & Van de Grift, W. J. C. M. (2018). Promoting historical contextualization in classrooms: An observational study. *Educational Studies*. Advance online publication. doi:10.1080/03055698.2018.1509771

Huijgen, T. D., Holthuis, P., Van Boxtel, C. A. M., Van de Grift, W. J. C. M., & Suhre, C. J. M. (2018). Students' historical contextualization and the Cold War. *British Journal of Educational Studies*. Advance online publication. doi:10.1080/00071005.2018.1518512.

Huijgen, T. D., Van Boxtel, C. A. M., Van de Grift, W. J. C. M., & Holthuis, P. (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. doi:10.1007/s10212-014-0219-4

Huijgen, T. D., Van Boxtel, C. A. M., Van de Grift, W. J. C. M., & Holthuis, P. (2017). Toward historical perspective taking: Students' reasoning when contextualizing the actions of people in the past. *Theory & Research in Social Education*, *45*, 110-144. doi:10.1080/00933104.2016.1208597

Huijgen, T. D., Van de Grift, W. J. C. M., Van Boxtel, C. A. M., & Holthuis, P. (2017). Teaching historical contextualization: The construction of a reliable observation instrument. *European Journal of Psychology of Education*, *32*, 159-181. doi:10.1007/s10212-016-0295-8

Huijgen, T. D., Van de Grift, W. J. C. M., Van Boxtel, C. A. M., & Holthuis, P. (2018). Promoting historical contextualization: the development and testing of a pedagogy. *Journal of Curriculum Studies*. Advance online publication. doi:10.1080/00220272.2018.1435724

Professional publications

Huijgen, T. D., & Holthuis, P. (2013). Het verleden beter begrijpen [A better understanding of the past]. *KLEIO, 6*, 50-57.

Huijgen, T. D., & Holthuis, P. (2015). 'Why am I accused of being a heretic?' A pedagogical framework for stimulating historical contextualisation. *Teaching History, 158*, 56-61.

Huijgen, T. D., & Holthuis, P. (2018). 'Man, people in the past were indeed stupid.' Using a three-stage framework to promote historical contextualisation. *Teaching History, 172*, 8-16.

Huijgen, T. D., Holthuis, P., Knoops, C., Hofstede, J., & Eeftink, E. (2012). Droegen de Romeinen dan geen spijkerbroeken? [Did the Romans not wear jeans?]. *Didactief*, 46 - 48.

Staats, B., Nederhoed, A., Huijgen, T. D., & Holthuis, P. (2012). Vaardig in contextualiseren: Doorlopende leerlijnen van PO naar bovenbouw VO [Performing historical contextualization successfully in elementary and secondary education]. *KLEIO*, *2*, 28-33.

Visions of Education (Producer). (2017, May 3). *Episode 54: Students' historical perspective taking with Tim Huijgen* [Audio podcast]. Retrieved from https://visionsofed.com

Posters and presentations

Huijgen, T. D. (2013, March). *Naar excellent meesterschap: Het meten van de vakdidactische kwaliteit van leraren* [Towards mastery: Measuring the quality of teachers' domain-specific instructions]. Round table session at the VELON 2013 meeting, Groningen, The Netherlands.

Huijgen, T. D., Holthuis, P., Van Boxtel, C. A. M., & Van de Grift, W. J. C. M. (2016, September). *Testing a pedagogy for promoting historical contextualization in classrooms*. Paper presented at the HEIRNET 2016 meeting, Murcia, Spain.

Huijgen, T. D., Van Boxtel, C. A. M., Holthuis, P., & Van de Grift, W. J. C. M. (2014, June). *Naar excellent historisch redeneren* [Towards excellent historical reasoning]. Poster session presented at the Onderwijs Researchdagen 2014, Groningen, The Netherlands.

Huijgen, T. D., Van Boxtel, C. A. M., Holthuis, P., & Van de Grift, W. J. C. M. (2014, September). *Towards excellent Historical Reasoning: The competency of historical perspective taking (HPT).* Paper presented at the HEIRNET 2014 meeting, Ljubljana, Slovenia.

Huijgen, T. D., Van Boxtel, C. A. M., Van de Grift, W. J. C. M., & Holthuis, P. (2015, September). *Developing a pedagogy for promoting historical contextualization in classrooms*. Paper presented at the HEIRNET 2015 meeting, London, United Kingdom.

Huijgen, T. D., Van de Grift, W. J. C. M., Van Boxtel, C. A. M., & Holthuis, P. (2015, April). *Teaching historical contextualization: The construction of an observational instrument*. Paper presented at the AERA 2015 Annual Meeting, Chicago, IL, United States.

DANKWOORD (DUTCH)

"The mountains are calling & I must go & I will work on while I can,
studying incessantly."

John Muir, 1873

Het schrijven van een proefschrift is vergelijkbaar met een uitdagende trektocht in de bergen. Je plant een route, markeringspaaltjes wijzen de weg en obstakels zorgen soms voor een omweg. Onderweg bedwing je hoge bergen, loopt door diepe dalen en ontmoet interessante mensen. Deze trektocht was echter niet mogelijk zonder hulp en steun van een aantal ervaren reizigers en gidsen.

Lieve Carla, je bent een geweldig mens met een enorme passie voor het (geschiedenis) onderwijs. Jij haalt het beste in mensen naar boven en bent een voorbeeld voor velen. Dank voor alle steun en goede begeleiding. Ik hoop dat je mij in de toekomst weer wilt vergezellen tijdens diverse bergtochten. Beste Paul, we kennen elkaar al erg lang. In chronologische volgorde was jij mijn geschiedenisleraar, mijn vakdidacticus, mijn collega en mijn copromotor. Je bent echter vooral mijn mentor en vriend. Dank voor de fijne en inspirerende gesprekken en ik hoop deze nog lang te kunnen voeren met jou. Beste Wim, in 2012 stamelde ik iets over onderzoek willen doen naar geschiedenisleraren. Gelukkig was je direct enthousiast. Dank voor het vertrouwen en de discussies over onderwijs en onderzoek waar ik veel van heb geleerd. Carla, Paul en Wim, zonder jullie had ik het nooit gered. Jullie zorgden ervoor dat ik niet te ver afdwaalde van de geplande route, maakten een kampvuurtje wanneer het kouder werd en zorgden ervoor dat ik steeds beter in mijn eentje de tent kon opzetten.

Dank ook aan alle (oud-)collega's van de Lerarenopleiding. Wat een prachtige club mensen! Jullie zorgden ervoor dat ik mij nooit eenzaam heb gevoeld tijdens mijn reis. Velen liepen stukjes met mij mee. Cor en Rikkert, dank voor het meedenken en jullie hulp met de statistische analyses. Ook vond ik het erg fijn dat ik altijd met vragen terecht kon bij andere (oud-)onderzoekers van onze afdeling. Mijn dank voor deze collegialiteit is zeer groot! Anna, Deniz, Gerrit, Joyce, Marjon en Marieke, jullie waren (of zijn) mijn examencommissiegenoten. Dank voor het meeleven en het fijne gezelschap! Fenna, Janneke, Jasmijn, Joyce en Nienke (The Ladies) dank voor alle blikken op mijn gebruik van het Engels en het meedenken. Theo, dank voor jouw hulp met het aanscherpen van de

Nederlandse samenvatting. Alle fouten zijn voor mijn rekening. Ruth, heel erg bedankt voor al onze leuke koffiedates en gesprekken. Roomie Carien en opvolger Zwanie, met jullie is een werkkamer stukken gezelliger! Heel veel dank ben ik verschuldigd aan de dames van het secretariaat die altijd klaar staan voor anderen: Annemarie, Sindy, Suus en alle ECABO-secretaressen van de afgelopen jaren. Dik, Ellen, Klaas, Michelle en Siebrich, dank voor het vertrouwen en de ruimte die ik kreeg voor het voltooien van deze tocht! Ook wil ik mijn collega's van de faculteit Letteren bedanken. Tijdens mijn tochten in het Academiegebouw en Harmoniegebouw werd vaak belangstellend gevraagd hoe het mij en mijn onderzoek ging. Dat heb ik zeer gewaardeerd.

Ook (oud-)collega's van het Stadslyceum liepen stukken van de route met mij mee. Wanneer ik een ansichtkaart stuurde met een verhaal hoe de reis verliep, kreeg ik hartverwarmende reacties terug. Dit deed mij erg goed en daarvoor wil ik jullie allemaal heel erg bedanken! Caren, Co en Wieteke, dank dat ik deze reis kon beginnen en voltooien. Ook alle teamleiders wil ik danken voor de steun. Theo, dank voor de prettige gesprekken over onderzoek en onderwijs. Als “grote baas” van het openbare onderwijs in de stad Groningen, hoop ik dat je onderzoek binnen de scholen een warm hart blijft toedragen. Allert, Bert, Ellen, Jan, Lieke, Ilona, Roel, Siger, Tessa, Wibe en Wim, jullie zijn mijn (oud-)sectiegenoten van het Stadslyceum. In liefdevolle herinnering horen Esther en Menke hier ook bij. Dank jullie allemaal voor de belangstelling, het meedenken en de ruimte die ik kreeg. Het rooster- en formatieteam (de heren Bas, Michel en Remo): dank voor alle ingewilligde verzoekjes. Ook voor het andere onderwijsondersteunende personeel van het Stadslyceum: mijn dank is groot. Speciale dank aan de financiële administratie van Openbaar Onderwijs Groningen voor het bijhouden van de subsidie. Bert en zijn hulptroepen, dank voor de heerlijke broodjes, de koffie en de gesprekken! Oud-collega Hedzer: dank voor de hulp met het Engels! Natuurlijk wil ik ook al mijn (oud-)leerlingen danken. Met jullie is mijn reis eigenlijk begonnen. Het werken met jullie geeft mij enorm veel plezier en energie. Ik hoop dit nog lang te kunnen blijven doen.

Ik ben alle scholen, leraren, studenten, student-assistenten en leerlingen die hebben meegewerkt aan dit onderzoek ook ontzettend dankbaar. Dit zijn er gewoon te veel om op te noemen, maar jullie allemaal hebben het onmogelijke mogelijk gemaakt. Samen onderzoek doen met leraren vind ik het leukste dat er is. Ik heb genoten van de discussies, de informele etentjes, de borrels en het samen ontwerpen van lesmateriaal. Dank voor jullie tomeloze inzet en enthousiasme! Ik hoop dat ik in de toekomst weer een beroep op jullie mag doen.

Onderweg kwam ik veel ervaren reizigers tegen. Albert, Bjorn en Gerhard, dank voor alle hulp, inspiratie, samenwerking en ontspanning. Ook Arie, Dick, Geerte, Hanneke, Harry, Huub, Jannet, Karel, Kaat, Marc, Marjan, Mark, Michiel, Susan, Tessa, Uddhava en alle andere leden van de vakdidactische onderzoeksgroep, heel erg bedankt voor alle feedback en het meeleven! Ik hoop dat wij in de toekomst gezamenlijk mooie projecten kunnen uitvoeren. Ook dank aan alle andere collega-vakdidactici van andere universiteiten en hogescholen voor de warme belangstelling en het meedenken! A special thanks to the members of the History Educators International Research Network. The conferences are always like a warm bath after a long hike in the mountains. Ik wil graag NWO bedanken voor de promotiebeurs en het vertrouwen. Hierdoor werd ik gedeeltelijk vrijgesteld van mijn onderwijstaken en kon het onderzoek worden uitgevoerd. Ferdi, dank voor de hulp bij het ontwerpen van het eerste figuur. Ferdinand, dank voor de opmaak van het complete proefschrift en het fijne contact.

Durk-Rein, Fenneke, Henri, Marije en Ronald, jullie vormden vanaf het begin mijn trouwste klankbord. Dank voor al het werk dat jullie voor mij hebben gedaan. Wat zijn jullie gepassioneerde en goede leraren! Marije, jij bent tevens mijn collega bij de Lerarenopleiding en mijn paranimf. Dank voor alle gesprekken, jouw goede kritische blik en jouw vriendschap. Ik hoop in de toekomst samen veel te kunnen betekenen voor het geschiedenisonderwijs. Annemiek, jij bent mijn andere paranimf. Samen naar conferenties in het buitenland, samen (met natuurlijk koffie) in de trein naar Amsterdam, ik geniet van onze gesprekken en van jouw betrokkenheid bij het vakdidactisch geschiedenisonderzoek.

Lieve familie en vrienden. Ik prijs mijzelf erg gelukkig met jullie. Samen delen we lief en leed. Jullie zorgden vaak voor de broodnodige ontspanning en steun. Dank daarvoor! Beste Rick en Martijn, jullie zijn mijn broers en tegelijkertijd mijn vrienden. Betere broers kun je niet wensen. Dank voor het meeleven, de discussies en het slap ouwehoeren. Lieve Ries en Marian, zonder jullie was dit onderzoek niet mogelijk geweest. Door de vele reizen en de boeken die altijd beschikbaar waren, kreeg ik belangstelling voor het verleden en voor andere culturen. Jullie hebben mij gevormd en ik ben erg dankbaar voor jullie liefdevolle opvoeding. Altijd ondersteunend, inspirerend en belangstellend. Jullie zorgen ervoor dat Siegerswoude een fijne en warme plek is. Een plek waar ik altijd welkom ben en een plek met erg goede herinneringen.

Lieve Antje, ik ben iedere dag trots en dankbaar dat ik jou mijn vrouw mag noemen. Jij leefde de afgelopen jaren intens mee en stond altijd voor mij klaar. Ik houd mij vooral bezig met het verleden, maar de toekomst ligt voor ons en daar kijk ik heel erg naar uit!