Kasvatustieteellisiä tutkimuksia, numero 178
Helsinki Studies in Education, number 178

Amna Khawaja

Historical literacy in Finnish primary classrooms
Teaching history or things about the past?

Doctoral dissertation, to be presented for public discussion with the permission of the Faculty of Educational Sciences of the University of Helsinki, in the Small Festive Hall, University Main Building, Fabianinkatu 33, on the 19 January 2024 at 13 o'clock.

Helsinki 2024

Pre-examiners
Associate Professor Karel Van Nieuwenhuyse, KU Leuven
Associate Professor David Rosenlund, Malmö University

Custos
Professor Jukka Rantala, University of Helsinki

Supervisors
Professor Jukka Rantala, University of Helsinki
Docent Najat Ouakrim-Soivio, University of Helsinki

Opponent
Professor Emerita Arja Virta, University of Turku

ISBN 978-951-51-9998-0 (pehmeäkantinen)
ISBN 978-951-51-9999-7 (PDF)
ISSN 1798-8322
ISSN 2489–2297 verkkojulkaisu

Punamusta, Joensuu
2024

Do not merely practise your art, but endeavour also to fathom its inner meaning; it deserves this effort. For only art and science can raise man to the divine.

–Ludwig van Beethoven, letter to Emilie, 1812

What can be said at all can be said clearly; and whereof one cannot speak thereof one must be silent.

–Ludwig Wittgenstein, Tractatus Logico-Philosophicus, 1921

University of Helsinki, Faculty of Educational Sciences
Helsinki Studies in Education, number 178

Amna Khawaja

Historical literacy in Finnish primary classrooms
Teaching history or things about the past?

Abstract

The present study is an attempt to identify possible impediments to, and potential for, teaching historical literacy in the context of primary history education in Finland. The question is approached through assessment, classroom instruction and teachers´ thinking. A theoretical framework of recontextualization and a modified version of the pedagogic device (Bernstein, 1990, 2000) are applied in order to place the impediments and potential in a wider educational context. The three research articles that complement this dissertation utilize two separate datasets and two wider approaches, namely design-based research (Article I) and observational case study (Articles II and III).

In the first article, the aim was to produce assessment material that could help teachers implement historically literate assessment. The design process focused on testing the validity of an assessment tool designed to assess pupils´ historical literacy rather than general literacy or the capacity to memorize isolated historical facts. The design process took place in three phases. The assessment tool was tested with 67 participants, seven of whom took part in think-aloud interviews.

The data for Articles II and III were gathered in nine Finnish primary school classrooms. A total of 74 lessons were observed over the course of one year, and the data sources include numerical data (i.e., frequency of using text types), field notes and materials used in the classrooms. All of the teachers were interviewed at the end of the observation period. The interviews focused on teachers´ decisions concerning the planning and implementation of lessons. In addition, teachers were asked to reflect on the importance of a set of teaching practices and the meaning of disciplinary concepts. Article II examines the text types and the manner in which they were used in all nine classrooms. Article III focuses on two of the nine teachers and investigates the forms of knowledge present in classrooms by using the concepts of horizontal and vertical discourses. Additionally, Article III seeks to uncover features in primary teachers´ recontextualization process.

The findings identify a number of impediments to teaching historical literacy. First, the design process in Article I revealed that pupils have difficulties in detecting contradictions between texts, and that they tend to read texts for

information rather than as evidence. As a result, materials used for assessing historical literacy need to provide sufficient scaffolding for pupils by guiding them towards disciplinary literacy strategies rather than general strategies. The impediments reported in Article II include the predominant role played by the body text in textbooks, and consequently the scarcity of primary sources. None of the nine teachers gave explicit instructions on using history-specific literacy strategies. Instead, content-area literacy was applied. Texts other than those in textbooks were typically used only briefly and hence could not be analysed or interpreted. The findings in Article III suggest that because substantive knowledge took precedence over procedural knowledge and second-order concepts, a space for historical literacy was limited. Moreover, everyday knowledge was not addressed through disciplinary means, thus leaving the horizontal and vertical discourses unconnected. Finally, the recontextualization process of the two teachers was either limited and resembled the reproduction of knowledge, or was aimed at objectives other than historical literacy.

Although the study reveals more impediments than potential, it nevertheless suggests that primary teachers have a tentative understanding of the nature of historical knowledge, and some primary teachers even express in-depth insights into disciplinary aspects of history. Moreover, most teachers viewed the use of primary sources not only as possible but also as desirable if they were made available. The first article also revealed that primary pupils were well-equipped for reading multiple sources and familiarizing themselves with the historical context within a one-hour timeframe. Therefore, there seem to be no insurmountable obstacles to using several textual documents in an assessment tool for historical literacy.

The study suggests that if historical literacy is to be promoted in primary history classrooms, several measures need to be introduced in order to overcome the identified impediments. All the agents of the pedagogic device, including teacher educators, textbook authors, curriculum designers and teachers themselves, can contribute towards a more historically literate instruction. The results call for teacher educators to equip primary teachers with more procedural knowledge and ways of teaching it to pupils. Moreover, teachers should be provided with the means to deconstruct historical culture and students' everyday experiences through disciplinary lenses. The study also suggests that agents responsible for curricular documents should produce materials for primary teachers´ use. This would clarify the aims stated in curricular documents and potentially reduce the influence of textbooks in the recontextualization process.

Keywords: historical literacy, primary education, recontextualization, assessment, vertical and horizontal discourses

Helsingin yliopisto, Kasvatustieteellinen tiedekunta
Kasvatustieteellisiä tutkimuksia, numero 178

Amna Khawaja

Historian tekstitaidot suomalaisissa alakouluissa
Opetetaanko historiaa vai asioita menneisyydestä?

Tiivistelmä

Tämän väitöskirjatutkimuksen tarkoitus on selvittää historian tekstitaitojen opetuksen esteitä ja mahdollisuuksia suomalaisessa alakoulun historianopetuksessa. Aihetta lähestytään arvioinnin, luokkahuoneopetuksen sekä opettajien ajattelun näkökulmista. Tutkimuksessa hyödynnetään rekontekstualisoinnin teoreettista viitekehystä (Bernstein, 1990, 2000) ja sovelletaan pedagogista muunninta (device). Niiden avulla tunnistetut esteet ja mahdollisuudet asetetaan laajempaan koulutuksen viitekehykseen. Väitöskirjan osatutkimukset rakentuvat kahdesta aineistokokonaisuudesta ja kahdesta menetelmällisestä lähestymistavasta: osatutkimuksessa I käytetään menetelmänä kehittämistutkimusta, osatutkimuksissa II ja III havainnoivaa tapaustutkimusta.

Ensimmäisen osatutkimuksen tavoite oli kehittää arviointimateriaali, jota voidaan käyttää historian tekstitaitojen arvioimiseen alakoulussa. Kehittämisprosessissa keskityttiin validiteetin tarkasteluun, jotta arvioinnin kohteena olisivat historian tekstitaidot yleisten tekstitaitojen sijaan. Kehittämistutkimus sisälsi kolme vaihetta, joiden aikana arviointimateriaalin testaamiseen osallistui yhteensä 67 oppilasta. Heistä seitsemän osallistui ääneenajatteluhaastatteluihin.

Toisen ja kolmannen osatutkimuksen aineisto kerättiin havainnoimalla yhdeksää suomalaista luokanopettajaa yhden lukuvuoden aikana. Havainnoituja oppitunteja oli yhteensä 74, ja aineisto muodostui havaintomuistiinpanoista, numeerisesta aineistosta (esim. erilaisten tekstityyppien esiintyminen oppitunneilla) sekä luokkahuoneissa käytetyistä materiaaleista. Lisäksi opettajia haastateltiin havaintojakson jälkeen, jolloin opettajilta kysyttiin heidän suunnittelu- ja opetusprosesseistaan. Opettajia pyydettiin pohtimaan tiedonalakohtaisten käsitteiden merkityksiä sekä suhdettaan historianopetuksen erilaisiin opetuskäytänteisiin. Toisessa osatutkimuksessa perehdyttiin erilaisten tekstien käyttöön yhdeksässä luokkahuoneessa. Kolmannessa osatutkimuksessa näkökulma tarkennettiin kahden opettajan opetuksen tiedonmuotoihin analysoimalla opetusta horisontaalisen ja vertikaalisen diskurssin käsittein. Lisäksi kolmannessa osatutkimuksessa eriteltiin luokanopettajien käyttämiä rekontekstualisoinnin tapoja.

Tutkimuksessa tunnistettiin useita tekijöitä, jotka rajoittavat historian tekstitaitojen opettamista. Arviointimateriaalia kehittäessä kävi ilmi, että oppilailla oli vaikeuksia havaita ristiriitaisuuksia tekstien välillä. Lisäksi oppilaat suhtautuivat teksteihin informaation lähteinä eikä todistusaineistoina. Toisessa osatutkimuksessa esille tulleista esteistä keskeisimpiä ovat oppikirjojen leipätekstin hallitseva asema suhteessa esimerkiksi alkuperäislähteisiin sekä tiedonalakohtaisiin tekstitaitoihin ohjeistamisen puute. Kukaan opettajista ei mallintanut oppilaille historian tekstitaitojen strategioita. Oppikirjaa ja sen leipätekstiä lukuun ottamatta tekstien parissa vietettiin vain vähän aikaa. Tämä ei mahdollistanut tekstien tutkimista ja tulkitsemista. Kolmas osatutkimus osoitti, että proseduraalisen tiedon osuus suhteessa muihin tiedon muotoihin oli vähäinen, mikä rajoitti historian tekstitaitoihin vaadittavaa tiedollista tilaa. Arkitietoa ei myöskään purettu tiedonalakohtaisen käsitteistön avulla, jolloin horisontaalinen ja vertikaalinen diskurssi jäivät toisistaan irrallisiksi. Viimeisenä esteenä voidaan pitää havaittua rekontekstualisoinnin tapaa, joka oli joko rajoittunutta tai suuntautui tavoitteisiin, jotka eivät olleet yhteensopivia historian tekstitaitojen kanssa.

Lukuisista esteistä huolimatta tutkimuksessa tunnistettiin myös mahdollisuuksia suhteessa historian tekstitaitoihin. Kaikilla opettajilla oli vähintäänkin alustava ja joillakin opettajilla varsin syväkin ymmärrys historiallisen tiedon luonteesta ja tiedonalakohtaisista käsitteistä. Opettajat myös suhtautuivat myönteisesti alkuperäislähteiden käyttöön opetuksessa. Kehittämistutkimus puolestaan osoitti, että alakouluikäiset oppilaat kykenivät lukemaan useita lähteitä ja ottamaan haltuun uutta kontekstitietoa oppitunnin mittaiseksi tarkoitetun tehtävän puitteissa. Siten alakoulussa ei näytä olevan esteitä historian tekstitaitoihin kohdistuvien tekstipohjaisten arviointimateriaalien käytölle.

Mikäli historian tekstitaitojen opettamista halutaan tukea, tarvitaan ratkaisuja tunnistetuille esteille. Kaikki pedagogisen muuntimen toimijat, kuten opettajankouluttajat, oppikirjojen tuottajat, opetussuunnitelmien laatijat ja opettajat voivat vaikuttaa myönteisesti historian tekstitaitojen tukemiseen tutkimuksessa ehdotetuilla tavoilla. Opettajankoulutuksessa tulisi painottaa proseduraalisen tiedon merkitystä ja tarjota välineitä sen opettamiseen. Arkitiedon ja historiakulttuurin käsittelyä historian tiedonalan kautta tulisi myös vahvistaa opettajankoulutuksessa. Opetussuunnitelmista vastaavien toimijoiden tulisi puolestaan tarjota materiaaleja, jotka konkretisoisivat opetussuunnitelman tavoitteita ja mahdollisesti vähentäisivät oppikirjojen hallitsevaa roolia rekontekstualisointiprosessissa.

Avainsanat: historian tekstitaidot, alakoulu, rekontekstualisointi, arviointi, horisontaalinen ja vertikaalinen diskurssi

Acknowledgements

Completing a dissertation, as anything worthwhile in life, takes a village. This process would not have been possible without the help of a number of people to whom I will always be indebted. While writing the thesis itself has been anything but a straightforward endeavour, the difficulty in writing these few words at the end of such a long process caught me by surprise, as the words do not seem to do justice to the gratitude I feel.

First and foremost, I would like to thank my supervisors, Jukka Rantala and Najat Ouakrim-Soivio, who have provided indescribable support throughout the years. Being surrounded by your expertise, selflessness and kindness has been a privilege awarded to just a lucky few. I thank you both for the artistry of knowing when to give advice and when to refrain from doing so. Another vital person in the thesis process has been Mikko Puustinen, whom I wish to thank not only for a fruitful but also for a fun collaboration. You are a source of brilliant ideas while simultaneously having the ability to listen to the ideas of others, which makes discussions with you particularly rewarding.

I would also like to thank my pre-examiners, Karel Van Nieuwenhuyse and David Rosenlund, for their insightful comments on the manuscript. Your suggestions have been invaluable and have broadened my thinking. Although I cannot name the pupils and teachers who participated in this study, without your contribution, this study would not even exist. I have the deepest admiration for the nine teachers who opened their classroom doors to me.

While some communities are imagined ones, I have been fortunate enough to be a part of some that are not. I would like to extend my thanks to our subject-specific group members, Jari Salminen, Janne Säntti and Tanja Taivalantti, in addition to Jukka and Mikko, for stimulating debates where everyone means what they say and says what they mean. The sense of humour shared within this group and the mutual respect shown for one another is something quite extraordinary. Another essential community is that of the ‘ruokaryhmä’, where colleagues, many working along the same corridor, make day-to-day life at the office meaningful and have therefore contributed immensely to the progress of this dissertation. Leea Lakka and Penni Pietilä, a special thanks to both of you for providing peer support and friendship without the expectation of keeping up appearances. I am also most grateful to Pia Mikander for help at the beginning of the process and to Eeva Haataja and Jenni Marjokorpi for technical support during the final steps. Maiju Wuokko, thank you too for showing excitement in my study from the history department perspective.

While working on the dissertation over the years, I have been part of several research projects and networks, such as HisLit, HistoryLab, PARVI and KOSS. I

extend my heartfelt thanks to all of you on these projects. When starting out as a PhD student, I could not have imagined that I would have the opportunity to interact with so many esteemed researchers. A special thank-you to Jonna Pulkkinen, Laura Ketonen, Toni Mäkipää, and Taina Tammelin-Laine for close collaboration during the last couple of years. In addition to everything I have learnt about assessment, your respective subjects and about conducting large studies, I would also like to thank you for being so flexible with schedules when this dissertation required it. Within the KOSS network, I wish to thank Arthur Chapman and Kenneth Nordgren for the opportunity to co-write and hence to deepen my understanding of different aspects of history education. Not forgetting the entire Strand 2 in the KOSS network – thanks for the lively discussions that have introduced me to thoughts and literature essential for this thesis.

I have never taken the idea of having friends for granted and therefore feel particularly lucky to have had so many of you around me, especially during the last few years, which have been challenging in many ways. Thanks for sticking by me and for accepting me for who I am, warts and all.

Last but by no means least, I want to thank my family and especially my father, Jahangir Khawaja. Abbaji, your love for academic research, the English language as well as your ability to encourage and help me towards expressing my thoughts more precisely have been a constant companion during the years spent on the dissertation. Special thanks are also due to my nephew Benjamin, who was born when I began the thesis and with whom I have shared many unforgettable reflections on the limits of the possible and the impossible. As for all the love and help I have always received from my entire family, there are no words. Äiti, Abbaji, Tamim, Saara, Benjamin and Saima, with you I always feel like I am home.

At the office in Siltavuorenpenger, on the 12th of December, 2023

Amna Khawaja

List of original publications

I Khawaja, A. (2018). Designing an assessment tool for historical literacy: the case of Copernicus. *Nordidactica – Journal of Humanities and Social Science Education 8*(3*)*, 1–26.

II Khawaja, A. & Puustinen, M. (2022). The Nature of Literacy Instruction in Elementary School History Lessons. *Language and Literacy 24*(3), 20–44.

III Khawaja, A. & Puustinen, M. (2022). Recontextualising history in primary school: discourses in the classroom. *History Education Research Journal, 19*(1), 1–16.

Contents

1 Introduction

1.1 The intellectual powers of history

"One thing seems clear: if all students are helped to the full utilization of their intellectual powers, we will have a better chance of surviving as a democracy in an age of enormous technological and social complexity."

These words were not inspired by the emergence of the internet, the rise of globalization or the threat of climate change. Instead, they were written in 1960 by psychologist and educator Jerome Bruner. Thus, educational aims have addressed technological and societal changes long before our own time and, in a sense, not much has changed. On the other hand, it is hard to ignore the rapid change brought about by the internet in the quantity as well as the quality of information. Even though Lowenthal (2015, p. 11) warns us about the illusion that "every quarter-century seems especially earth-shaking to eyewitnesses", the 21st century has undeniably seen a shift in the information space. Misinformation and disinformation and the polarization of societies have introduced new challenges for maintaining stable democracies.

In addition to older forms of texts such as (text)books, news, documentaries and films, the textual space of students in the 21st century has been expanded for example by YouTube videos, websites, and social media platforms. Young people encounter claims about the past and its connection to the present through all these channels. For the students themselves as well as for democratic societies, it matters whether students have the means to question the validity of these claims or whether the claims are taken at face value. To that end, interpreting texts as evidence, understanding the motives and biases of those behind the texts, and distinguishing facts from interpretations all serve a purpose. All the aforementioned processes can be developed through historical literacy, and hence one way to utilize what Bruner calls the "intellectual powers" of students and to strive for an in-depth understanding of the world is to acquire historical literacy. However, this requires counter-intuitive thinking, which goes beyond common sense (Wineburg, 2001).

Historical literacy refers to the ability to interpret texts using strategies and ways of thinking specific to history (Downey & Long, 2016). A historically literate person has an understanding of the discipline of history (Lee, 2007) and knows how historical knowledge is constructed (Nokes, 2010). Moreover, students are able to use their knowledge about the discipline for constructing their own evidence-based historical interpretations. An integral part of historical literacy entails taking into consideration the motives and biases behind the

sources, comparing sources and examining them in their historical context (Wineburg, 1991). These goals are not merely aimed at understanding the past but also at understanding how the past is related to the present (Downey & Long, 2016). Without a historical gaze (Bertram, 2008) produced by historical literacy, students may struggle to take part in contemporary debates, which inevitably have their roots in the past.

In a globally linked world, history-related discussions are not confined to national borders. For instance, debates about removing controversial statues are global and reach people through their physical surroundings. The traditional ways of approaching history teaching, referred to as memory-history (Lévesque, 2008) and heritage education (VanSledright, 2016), focus on memorizing specific events from the past as well as learning and celebrating fixed narratives. This encourages object veneration rather than evaluation of claims (VanSledright, 2016). Historical literacy, on the other hand, aims at providing pupils with disciplinary tools to evaluate artefacts such as statues, and thereby to understand what history is about. Teaching history is profoundly different from teaching things about the past.

Most of the research on teaching historical literacy has focused on the secondary level (e.g., Cuban, 2016; Hartzler-Miller, 2001; Nieuwenhuyse et al., 2017; Nokes, 2010, 2012; Rosenlund, 2016). Fewer studies have addressed the primary level[1] (e.g., Hughes, 2021; Nokes, 2014; Stolare, 2017). There may be multiple reasons for the lower interest in the primary level, one being the persistence of the Piagetian age-related framework, according to which young learners are not considered capable of the complex and abstract thinking (Piaget, 1988) required to understand history and think historically.

Booth´s (e.g., 1980) work was among the first to question the narrowness of Piaget´s model and its applicability to learning history. Since then, a considerable body of work has provided us with evidence that understanding history is not a simple age-related question and that given the opportunity, pupils aged 11–12 are able to engage with history in a disciplinary way (e.g., Barton, 1997; Booth & Husbands, 1993; Fillpot, 2012; Foster & Yeager, 1999; Lee & Ashby, 2000; Nokes, 2014; VanSledright, 2002, Wissinger, et al., 2018). As Booth (1994) has concluded, "the limits of children´s ability to understand history, to 'think historically', seem to be set not so much by cognitive factors but on a wide range of issues such as the teaching context, the use of accessible and problematic historical materials or the teaching styles and subject knowledge of the teacher" (Booth, 1994, p. 65). The present study is interested in some aspects of this "wide range of issues", such as teaching, teachers and teaching materials, rather than in children's abilities because, as mentioned, there is already a compelling body of

[1] By this, I mean the Finnish primary level which covers grade levels 1–6 and is typically attended by pupils aged 7–12. Moreover, as explained in Chapter 2, history is taught in grades 4–6, when pupils are aged 10–12.

work suggesting that given the right context, guidance, and support, even 11–12-year-olds are able to think about history in a disciplinary way.

Another explanation for the relatively greater emphasis on the secondary level may not stem from pupils´ but rather from teachers´ abilities: subject teachers have a closer connection to the discipline of history and presumably a higher level of expertise in teaching history. As primary school teachers have been described as having a broad but shallow knowledge base (Stolare, 2017), their approach to teaching may not be considered equally relevant or interesting compared with that of history subject teachers. Yet it is in primary school that pupils are first introduced to the subject of history. Hence the manner in which history is taught and assessed during these years is likely to influence pupils´ perceptions of what history is about. Both the importance of history teaching in primary schools as well as the scant attention it has received in previous research provided the motivation for embarking on this study, the results of which are detailed in this dissertation.

1.2 Aims of the study

At the curricular level, many countries have shifted away from memory-history. The emphasis is now on the interpretation of history, with curricula structured around concepts such as historical literacy, historical thinking, and historical reasoning (e.g., Bertram, 2019; Ni Cassaithe et al., 2022; Ormond, 2017; Rosenlund, 2016). Although the movement advocating interpretational history began in the United States as early as the 1960s, and in the United Kingdom a decade later, it seems that the change has not reached many classrooms, where traditional ways of teaching and assessing history still prevail (Cuban, 2016; Nokes, 2010).

In Finland, historical literacy is stated as one of the main instructional objectives in teaching history, starting at primary level (Finnish National Board of Education, 2014). However, little is known about how history is actually taught at primary level, and whether or not the ideals of historical literacy are realized in classrooms. To date, research on the implementation of the history curriculum in Finnish primary classrooms has been restricted to a study on integrating history with other subjects (Mård, 2020). Further, although the existing assessment materials in teachers´ guides have been criticized for lacking a disciplinary approach (Lyhty, 2020), there is little information on how best to assess pupils´ historical literacy at the Finnish primary level. To that end, the present study aims to contribute to the slowly increasing body of knowledge on Finnish primary history education by observing nine primary teachers and developing new assessment material.

On the whole, this study is an effort to find out how historical literacy as an instructional aim could best be incorporated into Finnish primary classrooms. Thus, my main research question is:

What kind of impediments to and potential for teaching historical literacy can be identified in the context of Finnish primary history education?

In order to answer the main question, the three articles included in the study ask:

1. What kind of assessment tool is appropriate for assessing historical literacy among 11–12-year-old primary school pupils? (Article I)
2. To what extent the observed literacy practices are compatible with the ideals of historical literacy? (Article II)
3. What kind of discourses exist in primary history lessons? (Article III)
4. How does teachers´ thinking explain the observed discourses, and the ways in which teachers recontextualize history? (Article III)

By impediments, I refer to factors that may inhibit, and by potential to I refer to factors that may have a positive impact on the teaching of historical literacy. This division into impediments and potential is based upon the theoretical work described in Chapters 3 and 4, where I first define historical literacy (Chapter 3) and then go on to elaborate the ideals for teaching it (Chapter 4). The three research articles included in the dissertation utilize two separate datasets and two wider approaches: design-based research (Article I) and observational case study (Articles II and III).

The first article is rooted in the realization that without proper means for assessment, teachers will be unlikely to adopt new learning objectives (Seixas et al., 2015), such as historical literacy. Although textbooks should serve the purpose of implementing the curriculum, the majority of history textbooks available in Finland provide assessment materials that focus on memory-history (Lyhty, 2020). Thus, my aim in the first article has been to produce an alternative that could help teachers to implement historically literate assessment. The article seeks not only to identify impediments to and potential for assessing historical literacy, but also to find solutions to the identified constraints through an iterative design process. The design process focuses on testing the validity of the assessment tool to ensure that it assesses pupils´ historical literacy, rather than only general literacy or their capacity to memorize.

Articles II and III both utilize the data gathered in nine Finnish primary classrooms. I gathered this observational and interview data to investigate how teaching practices and the thinking of primary school teachers relate to the ideals of teaching historical literacy. In view of previous research on secondary history

instruction and teachers´ thinking (e.g., Nokes, 2010; Rosenlund, 2016), a number of constraints on historical literacy were expected in the Finnish context as well. The second article focuses solely on the literacy practices of the nine teachers, using only observational data. The aim of this article is to determine the extent to which the literacy practices are specific to history and what possible role do other types of literacies, for example content-area and cultural literacy (see Khawaja & Puustinen, 2022) play in the observed classrooms.

The third article employs both observational and interview data, but in order to gain an in-depth understanding of the entire teaching process from planning to implementation, the article focuses on two of the nine teachers. I use Bernstein´s (1990, 2000) work on recontextualization as the theoretical framework to analyse teachers´ role in defining the kind of knowledge that is present in history lessons. By recontextualization, I refer to a process whereby knowledge is transformed when it is relocated and refocused from its disciplinary origins for educational purposes (Bernstein, 1990; Hordern, 2021). The article seeks to explore how the power struggles inherent in recontextualization (see Chapter 2) affect the role of teachers in this process.

It should be noted that in the present study, teaching is considered to consist of a wide range of processes, one of which is assessment. As assessment is meant to support learning and inform teaching, it is implicitly or explicitly present in all phases of teaching (Koppinen et al., 1994). Thus, assessment is not explicated as a concept in the main research question but is included in the concept of teaching. This decision can be further justified by the fact that in the Finnish (primary) context, individual teachers have both the possibility and responsibility to carry out assessment as part of their teaching: Finnish basic education does not include any large-scale or national tests.

Although historical literacy is a key concept in the current history syllabus (National Board of Education, 2014), my interest in investigating historical literacy is not based solely on curricular relevance but on a broader theoretical framework. The theoretical basis of this study lies in a disciplinary approach, where structures and characteristics of different disciplines are at the core of education and learning (see Bruner, 1960; Gardner, 1999; Schwab, 1978). Bruner´s curricular hypothesis stating that "any subject can be taught effectively in some intellectually honest form to any child at any stage of development" (1960, p. 33) has been considered central to the curricular reform, which emphasized the role of disciplinary thinking (e.g., Deng, 2004; Lévesque, 2008, Tanner & Tanner, 1980). Thus, while my research question specifically addresses historical literacy, the study also aims to engage in a wider discussion on a disciplinary approach to teaching history, and ultimately to a discussion on the relationship between disciplines, school subjects and teachers (see Bernstein, 2000).

1.3 Structure

The study follows a traditional structure. The theoretical framework is divided into three chapters, the first of which – Chapter 2 – utilizes Bernstein´s (1990) pedagogic device to introduce the Finnish educational context. With the help of both the pedagogic device and the concept of recontextualization, I discuss some existing tensions that affect history teaching at the primary level. These tensions can be identified, for example, between different agents in the educational field as well as between the aims and influence of different school subjects such as history and Finnish language and literature. Chapter 3 deals with this latter tension as I approach the definition of historical literacy from the perspectives of both literacy education and history education. Chapter 3 also situates historical literacy in the context of other related concepts and thereby aims to decrease confusion around terminology. In the last chapter of the theoretical framework – Chapter 4 – I describe the ideals of historical literacy at the primary level. These ideals derive from the theoretical basis of the concept as well as the empirical work carried out in previous research.

Chapter 5 introduces the two methodological approaches of the study as well as the procedures and analyses of the three research articles. The most relevant findings in relation to the research question, namely the impediments to and the potential for teaching historical literacy, are included in Chapter 6. I conclude the study in Chapter 7 with a discussion of the results as well as a reflection on the wider implications of the present work.

2 Recontextualizing history in the Finnish context

In this chapter, I use Bernstein´s pedagogic device (1990) to introduce the context in which Finnish primary teachers teach history. I describe the process and the related power struggles through which the school subject of history is recontextualized in Finland, and reflect on primary teachers´ role in the equation. The aim is to shed light on the aspects that frame the space in which primary history is recontextualized. These aspects vary from the education of teachers to societal pressures and traditions, and from state-controlled curricular decisions to the role of textbooks. The chapter begins with an introduction to the pedagogic device and continues with the device´s application in the Finnish context.

2.1 A pedagogic relay

Knowledge undergoes a transformative process when changing its context from a discipline to a school subject. In other words, history as a discipline always differs from the school subject of history in some way. Bernstein (1990, 2000) has conceptualized this process through the idea of recontextualization, which involves appropriating, relocating, refocusing, and relating knowledge, or discourses, from one context to another (2000, p. 33). In order to understand how and by whom knowledge is transformed, Bernstein developed a pedagogic device depicting the process through which the knowledge from the disciplines recontextualizes into something else for the educational context. In his theorization, Bernstein identified the various stages and agents that are involved in the recontextualizing process. As I will elaborate later, recontextualization is not free from tensions. In every educational context, different agents have differing interests and aspirations about the form and shape that a school subject should take (Hordern, 2021). For example, state departments, textbook authors, subject teachers, and primary teachers may have dissimilar views on what history as a school subject should include and emphasize. Finnish teachers are often viewed as having considerable autonomy over their instruction (Kansanen, 2014). Thus, they have the opportunity to influence how subjects are taught. However, Finnish primary teachers´ relationship with and influence over school subjects has been explored less. In what follows, I examine the kind of space that the Finnish educational system and societal conditions provide for teachers´ recontextualization. The pedagogic device is a useful instrument for this purpose.

Bernstein´s pedagogic device (1990, 2000) is a complex model, covering multiple aspects of the sociology of knowledge. In the present work, however, I will introduce and explore those aspects of the device which are useful in understanding the circumstances in which history is taught in Finland at the primary level. A modified version of Bernstein´s (1990) original device is shown in Figure 1, where I take into account the Finnish context.

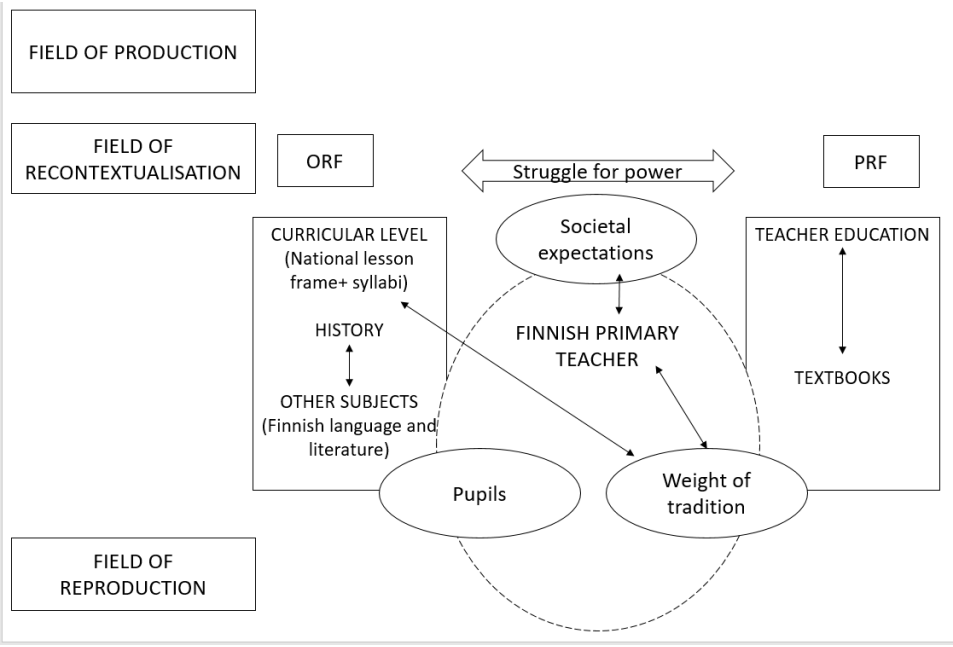

Figure 1. A modified version of Bernstein´s (1990) pedagogic device. Different power struggles within the device are depicted with various arrows.

The "pedagogic device" portrays the pedagogization of knowledge as a relay, starting from the field of production, continuing to the field of recontextualization and finally reaching the field of reproduction. The primary focus of the device is not on what is being relayed per se, but on "the constitution of the relay itself", including the rules and agents involved in the process (Bernstein, 2000, p. 25). The field of production is responsible for generating new knowledge, typically by researchers and experts in various disciplines. The second field, the field of recontextualization is where, according to Bernstein (1990, 2000), disciplinary knowledge is deliberately recontextualized so that it would serve educational goals. In other words, school subjects are created based on their parent disciplines within the field of recontextualization. Two different kinds of agents can be identified, who attempt to determine the aims and substance of the school subject.

The agents in the official recontextualization field (ORF) are typically state departments, which are responsible for creating official documents such as national curricula. The pedagogic recontextualization field (PRF), on the other hand, comprises teacher educators and different types of organizationa, for example. In my view, although not mentioned by Bernstein, textbook authors could also be included in the PRF. The views of the agents in the two fields (ORF and PRF) might differ considerably and thus create a power struggle. For example, the textbook authors may leave out aspects of the discipline that are considered essential by those designing national curricula, and vice versa.

Surprisingly, teachers are not part of the field of recontextualization in Bernstein´s pedagogic device (1990, 2000). Instead, teachers enter the relay only in its final stage, in the field of reproduction, where their role is to convey what has already been recontextualized by agents in both of the recontextualizing fields (ORF and PRF). Bernstein does not assign a significant role to teachers in defining the nature of school subjects as they are placed at the end of the pedagogization relay. Instead of transforming and recontextualizing knowledge, teachers are thought to acquire and transmit knowledge (2000, p. 37).

Bernstein (1990) intended the pedagogic device to be a grammatical tool regulating the relations between and within the levels of production, recontextualization and reproduction. He underscored that "the outer boundaries and inner possibilities of each level is a matter of the historical and ideological context of the device" (1990, p. 189). Hence, the device should not be understood as a "one-size-fits-all solution", but as an instrument through which power and structures related to knowledge could be identified in educational contexts differing in time and space. As Hordern (2021, p. 2021) phrases the issue: "The policy context and the politics of the recontextualization field shape the extent to which teachers can exercise autonomy and have a role in formulating the curriculum".

The device has been influenced by Bernstein´s familiarity with the British educational system, where the process of 'curriculum making' largely determines the content to be taught (Kitson, 2020; Deng, 2022). However, Kitson´s work (2020), also in the British context, emphasizes the role of teachers as recontextualization agents. A similar view on the teacher´s role in recontextualization has been expressed by Nordgren (2021) in the Nordic context. As mentioned, Finnish teachers are typically described as having considerable autonomy in their work (Kansanen, 2014; Puustinen, 2022), and placing them in the field of reproduction may seem unwarranted. Therefore, the Finnish primary teachers' space in Figure 1 covers the fields of recontextualization and reproduction, but because of the differing views, it is represented by a dotted circle.

The view that school subjects are transformed from pre-existent disciplinary knowledge has also been criticized for the demands that it places on teachers.

Stengel (1997) argues that, unlike disciplinary scholars, teachers usually do not have a sophisticated and in-depth understanding of disciplinary knowledge, and their inclusion as members of disciplines has also been debated (see Fordham, 2016). However, in the Finnish context, these reservations have little relevance, especially concerning subject teachers. While subject teachers may not take part in producing new knowledge in the disciplines (Fordham, 2016), they are nevertheless experts in their respective disciplines, which they study for five years for a master´s degree. Thus, as subject teachers have an in-depth understanding of their respective disciplines (Puustinen, 2022), they can be assumed to use their access to the discipline in their recontextualization process. Primary teachers, on the other hand, have no direct link or exposure to the field of production, in this case the discipline of history, and therefore expectations about their role in the recontextualization process or their membership in a disciplinary community are less obvious (see Fordham, 2016).

Before moving on to the tensions and power struggles inherent in the device, it should be noted that the focus here is on structures and entities through which knowledge is recontextualized. Thus, pupils´ influence on teaching, although recognized as significant, is not discussed in this chapter.

2.2 A struggle for power

As noted, the recontextualization process involves several agents who struggle for the power to determine the type of knowledge that is taught in a given school subject, and for the power to "regulate consciousness" (Bernstein, 2000, p. 37). However, the pedagogic device is less explicit about whose consciousness it regulates and where the "arena of struggle" exists (p. 38). Recontextualization is a problematic concept if it is not linked to the role of teachers (Hordern, 2021). Hence, I take the view that the rivalry between various agents in the recontextualization process takes place in relation to teachers, who are the last link in the relay and who determine what is actually taught in classrooms (see also Vesterinen, 2022). I have attempted to convey this centrality of teachers in Figure 1. The knowledge advocated by curricular documents (ORF) or textbooks (PRF), for example, holds little power by its mere existence if it is not used by teachers. Hence, multiple agents advocating a certain approach to knowledge aim to influence teachers, who may or may not be convinced. The central position of teachers is emphasized in the Finnish context, where no school inspections or national tests exist, and where teachers are free to choose their materials and methods in the classroom. I begin by describing the power struggle between the official and pedagogic recontextualization fields in the Finnish context in general, before moving on to some more specific sources of tension.

The wider struggle for power in recontextualization is between the official and the pedagogic recontextualization fields, represented by the horizontal arrow in Figure 1. The official recontextualization field (ORF) is largely associated with drawing up curricula (e.g., Bernstein, 2000; Hordern, 2021), but its power could be viewed to extend beyond the content of curricula. For example, the ORF also uses its power to determine how much each subject is taught and at what age. In Finland, pupils start studying history at the earliest in the fourth grade, and at the latest in the fifth grade. The national distribution of lessons assigns 1–2 hours of history per week for two years. However, schooling providers are free to decide how to distribute the lessons during grades four to six. Schools can, for example, introduce history in the fourth grade with one lesson per week and continue with two lessons per week in the fifth grade. In that case, only civics/social studies is taught in the sixth grade, the final year of primary school. Alternatively, it is possible to study civics/social studies in the fourth grade and to start studying history from the fifth grade (Finnish National Agency for Education, 2020).

The ORF's core area is the national curriculum, a binding document for teachers in Finland that is revised every ten years or so. While in some countries, politicians have a considerable influence on the ORF (see e.g., Smith, 2017), the Finnish curricular process, although it may be relatively less politicized, is still recognized as political and as a means of political influence (Rantala & Ouakrim-Soivio, 2020). Hence, it follows that even in Finland the process is initiated and governed by political actors, who may have a say in the composition of the group selected to design the national curriculum.

The two most recent Finnish national core curricula (FNAE, 2004, 2014) have recontextualized primary history in a manner that emphasizes disciplinary aims at the primary level. The 2004 National Core Curriculum (Finnish National Agency for Education, 2004) utilized historical thinking concepts such as change and continuity and cause and consequence (see e.g., Lévesque, 2008, Peck & Seixas, 2008), and also pointed towards historical empathy for the first time (Rantala, 2011). The same document also identified the nature of historical knowledge as one of the main aims of learning history. The ratio between aims and content to be taught was tilted towards content; there were five aims (related to historical thinking and the nature of historical knowledge) and eight core content areas, starting with pre-history and ending with the French Revolution. Although disciplinary in nature, it was suggested that the future curriculum should reduce the amount of content in order to highlight the role of disciplinary history (Rantala, 2011).

Indeed, the following National Core Curriculum of 2014, which has been in use throughout the duration of this study, did take an even more disciplinary turn. The aims of the current curriculum are structured under three titles: 1) acquiring information about the past, 2) understanding historical phenomena, and 3) applying historical knowledge. Historical thinking concepts are used under these

titles to frame the eleven individual learning objectives, all related to some aspect of disciplinary history. The way that the eleven learning objectives (Table 1) are constructed implies that they are meant to develop both historical thinking and historical literacy at the primary level. The curriculum emphasizes multi-perspectivity and an interpretational process in relation to texts. It does not promote an empiricist epistemological stance (cf. Wilke et al., 2022), which assumes that interrogation of sources would lead to revealing the truth about the past (see Rosenlund, 2015).

The current National Core Curriculum (2014) defines only five core content areas, as opposed to eight in the previous 2004 curriculum. In addition, in the current curriculum, the core content areas are defined in broader terms and therefore give teachers the freedom to choose individual topics.

Table 1. The 11 objectives for history teaching at the primary level, stated in the 2014 National Core Curriculum (Finnish Agency for Education).

	Objective of instruction
O1	to guide the pupil to become interested in history as a field of knowledge and a subject that builds his or her identity
O2	to guide the pupil to recognise different sources of history
O3	to guide the pupil to notice that historical information can be interpreted in different ways
O4	to help the pupil to understand different ways of dividing history into eras and using the related historical concepts
O5	to guide the pupil to understand the motives of human activity
O6	to help the pupil to perceive different reasons for historical events and phenomena and their consequences
O7	to help the pupil to identify changes in the history of his or her family or community and to understand how the same changes may have meant different things to different people
O8	to teach the pupil to perceive the continuity in history
O9	to guide the pupil to propose reasons for changes
O10	to instruct the pupil to explain how interpretations may change as a consequence of new sources or new ways of examining them
O11	to guide the pupil to explain human activity

As the aims of both the 2004 and 2014 National Core Curriculum suggest, the ORF has consistently emphasized a disciplinary approach to history since the turn of the millennium, but the two curricula also differ in some respects. In the most recent curriculum, the reduced role of substantive knowledge in content areas

underscores that the recontextualization process by the ORF has aimed at a close relationship between the discipline of history and the school subject of history.

The curricular process in Finland starts with the National Core Curriculum and continues with cities and schools drawing up their own local curricula based on the national one. The local curricula interpret the national one in terms of the aims and content and define more specifically what should be taught at each grade level (Ouakrim-Soivio, 2015). At this local stage, it is difficult to establish to what extent the process remains part of the official recontextualization field because it is mostly teachers who are involved. Although cities and schools make their own curricular interpretations, these local documents should not contravene the national curriculum, which is a normative document. To date, however, there is no research on the concordance between national, city and school-level curricula in the Finnish context. In other words, it is only assumed that the curricula of individual schools are in compliance with the National Core Curriculum (2014) and, in the case of history, promote a disciplinary approach; it may be that the emphasis in the local curricula is different.

Assessment criteria defined in the National Core Curriculum are not intended to be modified locally but expected to be applied as such across the board. The assessment criteria in national curricula have been binding since the 2004 curriculum and the aim has been to create a system where pupils throughout the country would be assessed in a fair and equitable way (Ouakrim-Soivio, 2015). Unlike many other countries (see e.g., Alvén, 2021), there are no national tests at the basic educational level (grades 1–9).

Recently, the Finnish Agency for Education revised the assessment criteria for the final assessment conducted at the end of basic education (year 9). This reform was initiated in response to the growing concern that assessment across Finland was not unified, and that grading varied significantly between students with similar competence. In history, for example, students who had similar learning outcomes differed by up to two grades (on a scale of 4–10) in their final assessment (Ouakrim-Soivio, 2012). In order to unify the assessment, the existing criteria for "good performance" (grade 8) were complemented with a set of criteria for grades 5, 7 and 9. By setting more specific criteria, the assessment would be less contingent upon the teachers and the assessment culture in schools. While the reform for the final assessment came into effect in autumn 2021, the Finnish Agency for Education also initiated a reform concerning the assessment criteria used at the end of the primary level (year 6). This reform has come into effect in autumn 2023. In other words, in recent years, the ORF has increased its share of influence mainly through assessment reforms.

One of the main agents in the pedagogical recontextualization field (PRF) is teacher education. Unlike in many other countries, in Finland many have aspired to become primary teachers, although the trend has changed in recent years. However, in 2017, only 11% of applicants were accepted into the primary teacher

programmes in universities (Finnish National Agency for Education, 2017). Hence, at least until lately, prospective primary teachers have had to perform well in their upper secondary studies in order to enter teacher training programmes in university (see Tallavaara & Rautiainen, 2020). Due to the popularity of primary teacher education, up to 96% of primary teachers working in schools are qualified, which means that they have a master´s degree in education (Finnish National Agency for Education, 2017).

In Finland, the curriculum for the five-year master´s programme varies from university to university, but typically class teachers major in educational sciences and complete multidisciplinary studies (60 ECTS) comprising short courses in all school subjects listed in the National Core Curriculum (2014). Every university determines the scope and content of the history didactics course or courses designed for prospective elementary teachers, but typically only 3–5 ECTS are required (see Tallavaara & Rautiainen, 2020). At the University of Helsinki, the single, compulsory history didactics course consists of 6 lectures and 6 group sessions extending over a period of 8 weeks, which qualifies students to teach history in primary schools. Many specialize in a school subject by completing a one-year extra course (60 ECTS).

As described above, prospective primary teachers spend five years in teacher education, but their encounter with individual school subjects during their studies is relatively short. Therefore, initial teacher education as an agent of the PRF has an extremely narrow time window to convey its views on the nature of school subjects. The importance of initial teacher education is further emphasized by the fragmented and non-systematic in-service teacher education (Heikkinen et al., 2015). Recent research also suggests that despite the efforts of teacher education departments to produce primary teachers who can draw on research in their work, only a minority have any connection to educational research (Koski et al., 2023). Thus, in-service teachers most likely rely on their initial teacher education when teaching school subjects. The brief history didactics course during the initial teacher education raises questions about the extent to which primary teachers are expected to familiarize themselves with the school subject and act as agents of recontextualization in their future work. On the other hand, through the disciplinary aims of the National Core Curriculum in history (see Table 1), the ORF expects primary teachers to be acquainted with the nature of historical knowledge, thereby conveying a different kind of signal.

For Bernstein (2000, p. 33), autonomy of education means that the PRF can have an impact on pedagogic discourse, independently of the ORF. This requires that the state does not interfere with the possible impact of the PRF on the recontextualization process or does not attempt to weaken the PRF, something that Bernstein feared was happening in the UK context. In Finland, the PRF can be considered relatively independent of the ORF as, for example, textbooks have not been subject to state control for several decades. Although the absence of

school inspection and other forms of control by the Finnish state is more concerned with the field of reproduction than the field of recontextualization in Bernstein´s model, it nevertheless implies that the ORF has limited power in the Finnish education system (see Simola et al., 2009). Recent developments suggest that the state wishes to increase its role in schools, however. The new government elected in 2023 has expressed an interest in using legislative power to define the role and use of digital devices in classrooms, for example (Finnish Government, 2023).

2.2.1 Tensions between subjects

Apart from the wider struggle for power between the ORF and PRF, there are also power struggles within the fields. These are represented by vertical arrows in Figure 1. In the case of the ORF, tensions arise from the different status afforded to different school subjects. Whilst the national framework for the distribution of lesson hours is intended as a legal and administrative document, inferences can be made about what it conveys about the relative standing of school subjects. One aspect of the tension concerns the age at which pupils should start learning a subject. This also has implications for the total number of lessons taught during primary school; the earlier a subject is introduced, the more hours it gains. For example, religious studies and ethics begin as early as the first grade and continue throughout the basic education. History is taught at the earliest in year 4, but usually in year 5. As a consequence, relatively more time is devoted to religious studies and ethics during the six years of primary school.

Finnish language and literature is the most taught subject at the primary level. In grades 5 and 6, it is taught for 4–5 hours per week, and the number of hours is even higher during the first years of primary school. History, on the other hand, is taught 1–2 hours per week for two years. For primary teachers, who teach all subjects, the national framework for the distribution of lesson hours may establish a certain hierarchy between school subjects. When it comes to teaching literacy, this means that it is mainly viewed through the subject of Finnish language and literature, where general literacy is taught. The predominant standing of Finnish language and literature gives rise to speculation as to whether other subjects, and especially their literacies, are viewed as auxiliary platforms on which to develop general literacy. Do the objectives of Finnish language and literature affect the recontextualization of other subjects as well? Be that as it may, the decisions made by the ORF on the distribution of lesson hours between subjects are far from tension-free.

In addition to the number of hours to be taught, the age at which different school subjects are introduced also tells us something about how the ORF views the knowledge of different school subjects and its accessibility to pupils. Apart from the B1 language (Swedish or Finnish depending on the pupil´s mother

tongue), history is the last subject to be introduced at the primary level. This suggests that at the state level, historical knowledge is viewed as unattainable for younger pupils, unlike in many other countries where history teaching begins much earlier. Through these types of documents, the ORF (the state) communicates what kind of knowledge pupils need at different stages and when they are considered capable of constructing that knowledge.

These preferences by the ORF can be assumed to have an effect on teachers, starting from teacher education: prospective teachers know which subjects they will be expected to teach the most and may orientate their studies accordingly. For example, those specializing in history are relatively few: in the whole of Finland, only 22 prospective primary teachers completed 60 ECTS in history in 2020 and thus became qualified to teach history at the lower secondary level (Vipunen, 2022).

2.2.2 Teacher education and textbooks

I approach the inner tensions of the PRF by examining the way in which teacher education and textbooks recontextualize history (see Figure 1). The starting point of history didactics has been studied relatively actively in recent decades (see e.g., Rantala & Khawaja, 2021; Tallavaara & Rautiainen, 2020; Virta, 2001). In all these studies, prospective primary school teachers were asked at the beginning of the history didactics course what they considered important in teaching history. Although the results show some variation, in all three studies, the most important reason for teaching history was to increase understanding about the past and thus to understand the present. In addition, acquiring general knowledge about the past was a high priority (Rantala & Khawaja, 2021; Tallavaara & Rautiainen, 2020). No systematic research exists on how history didactics courses in various teacher education units proceed from this starting point and in which direction they recontextualize history.

However, some tentative inferences can be made based on the course description and general aims detailed in research articles. It seems that at least some history didactics courses in Finland introduce the nature of historical knowledge and the interpretation of history and could therefore be described as promoting a disciplinary approach (Rantala & Khawaja, 2021; Virta, 2001). As Ahonen´s review (2020) of Finnish history education points out, history educators have chosen differing orientations and different approaches concerning international trends. Some, like Ahonen herself, identify with the German tradition of historical consciousness, while others have drawn upon Anglo-American concepts of historical thinking and literacy. It can be assumed that these orientations and approaches identified in history educators´ research may correspond to the history didactics courses conducted by them. Despite varying

orientations, there are no indications that history didactics courses in universities promote a purely memory-history approach to history, which is associated more with history textbooks.

Apart from teacher education, textbooks and their authors are important agents in the PRF, particularly because textbooks have been shown to play a significant role in the Finnish education system (Heinonen, 2005). There are different interpretations about the way that textbooks relate to the ORF, the surrounding historical culture and the discipline. For example, according to Williams (2014, p. 1), "formally, a medium for the transmission of educational content, the school textbook is also an instrument of the state, a national project, as it were". On the other hand, textbooks as well as the school subject itself have been interpreted as having a close connection to popular historical culture (van Niuwenhuyse, 2018). While acknowledging the influence that both national states and historical culture have on textbooks, in the present work I perceive textbooks as the product of the PRF, mainly because of their authorship. In Finland, textbook authors are researchers, teacher educators and experienced teachers (see Löfström, 2019), all of whom are considered part of the disciplinary community (Fordham, 2015) and who draw considerably on historical research (Ahonen, 2017). This proximity of textbook authors to the discipline makes it difficult to perceive history textbooks mainly as a product of historical culture. However, as mentioned earlier, the recontextualization process of textbooks involves much more than a link to the discipline. Historical culture, traditions related to history textbooks as a genre (see Klerides, 2010) and the ORF in the form of the national curriculum are likely to influence the process of writing history textbooks. Although Finnish textbooks are produced by commercial publishers and have not been inspected since 1992 (Löfström, 2019), there is an underlying expectation that they should not be disconnected from the national curriculum. As teachers are expected to follow the national curriculum and textbooks are designed to be a teaching resource, it may be assumed that textbooks help teachers to implement the curriculum. Textbook authors also refer to this link between the textbook and the national curriculum in the preface to their textbooks.

The connection between curriculum and textbooks in the Finnish context has been studied by Aalto and Kemppinen (2020), who argue that "textbooks are interpretations of a national curriculum valid at the time" (p. 101) and that the "national curriculum produces textbooks" (p. 99). However, several other studies suggest that textbooks do not always mirror national curricula, even though from the teacher´s perspective, it is assumed that they do. Rantala´s (2017) review of Finnish history textbooks suggests that although there have been considerable changes in the way textbooks recontextualize history, many of them continue to build national narratives. After the curricular shift towards disciplinary history in 2004, it was expected that there would be a similar shift in textbooks. Instead, only one textbook series clearly abandoned a coherent narrative and focused on

multiperspectivity and the interpretation of sources (Rantala, 2017). Vesterinen´s (2022) analysis of Finnish history textbooks also underscores that while some textbooks have become part of a disciplinary shift, the majority of them promote a collective memory approach.

Lyhty (2020) concluded that teacher´s guides to primary history textbooks tend to contain limited disciplinary aspects and that the materials have a weak connection to the National Core Curriculum (National Board of Education, 2014). In only one textbook series were the curricular aims present and to some extent even identified. A similar discordance between the National Core Curriculum and textbooks was found in Norppa´s (2019) work, where she analysed lower and upper secondary history textbooks. Moreover, as Finnish history textbooks often emphasize a Western perspective on history (Mikander, 2016), this poses challenges to the aim of multiperspectivity. Based on these studies, it could be suggested that some textbook authors´ recontextualization process may not be aligned with the recent trends in teacher education and curricular reforms. As the content and approach of textbooks are not dictated by any agent in the ORF, textbook authors can recontextualize freely and textbooks may even become a powerful counterforce to national curricula.

2.2.3 Tradition and societal expectations

The most intangible factors defining primary teachers´ space for recontextualization are tradition and societal expectations. By tradition, I refer to the tradition of teaching history in a manner that teachers are familiar with from their own school days. Moreover, teaching communities themselves sustain traditions (VanSledright, 1996). These traditions have long upheld a certain way of teaching history, often described as the heritage approach (Seixas, 2016b; VanSledright, 2016), as well as memory-history (Lévesque, 2008), both of which are suggested as opposites of disciplinary history (Lowenthal, 1998; Nora, 1996). Heritage and memory are often selective, collective (and national) as well as celebratory in their approach to the past (Nora, 1996; VanSledright), while "history, on the other hand, is the reconstruction, always problematic and incomplete, of what is no longer" (Nora, 1996, p. 3). Thus, memory-history and heritage could be described as being invested in preserving and celebrating the past, while history aims at distancing itself from it by investigating the past through analytical and sceptical inquiry (VanSledright, 2016). In the present work, the term memory-history rather than heritage is used as a counterpoint to disciplinary history because of the multiple meanings given to heritage (see e.g., van Boxtel et al., 2016).

Memory-history in classrooms typically translates into lists of content to be memorized through narratives presented by the teacher and/or textbooks. It is about information transmission and factual recall (Vesterinen, 2022), typically

practised through close-ended questions. Studies suggest that memory-history is prevalent in a number of countries; memorizing facts and narratives continues to make up a considerable part of history lessons (e.g., Cuban, 2016; Nokes, 2010; Rosenlund, 2016).

There are several reasons for the success of memory-history. First, some teachers consider that the aim of teaching history is to raise "an all-round cultivated person", and that teaching a maximum amount of information about historical events is a means to that end (Puustinen & Khawaja, 2022, p. 15). Thus, memory-history is rooted in the belief that history is a collection of facts students are entitled to acquire in order to build the big picture of history (see Veijola, 2013). The second reason for the persistence of the traditional way can be traced back to the practical aspect of teaching. Even during their studies, prospective history teachers identify as either "idealists" or "realists", where the former represent a disciplinary and the latter a traditional, "old" way of teaching (Veijola, 2013, p. 237). This division implies that a disciplinary approach to history is viewed as unrealistic in terms of materials and time constraints, for example. Hence, memory-history is viewed as more compatible with the structures of school, making it a strong contender in the minds of teachers. A study on prospective primary teachers´ views on disciplinary thinking showed that while a vast majority of the participants considered a disciplinary approach to be a sound way of teaching history, almost half had doubts about whether they would be able to implement such an approach in practice (Rantala & Khawaja, 2021).

The third reason for the power of memory-history is the tradition itself. A tradition becomes a self-maintaining system, the dismantling of which requires a substantial amount of energy from all agents in the system, but particularly from teachers. Hence, teachers may understandably continue with the traditional way of teaching history until they are sufficiently convinced that a different kind of approach is both required and possible.

Societal expectations in the pedagogic device are the result of a public discourse taking place throughout society and especially in the media, including social media. While only a small minority of primary teachers may actively take part in the discussion about history education, all teachers are nevertheless subjected to the public discourse. Politicians' statements, news stories and social media debates convey what is at any given time considered important in terms of what is learnt about the past, and through which means. For example, were societal expectations targeted at building and maintaining national unity, it would undoubtedly resonate with teachers one way or another. Thus, it is important to note that one way to react to societal expectations is to go against them. Whether or not teachers comply with or resist societal expectations, they nevertheless have a crucial role in the pedagogic device.

3 A disciplinary approach to historical literacy

Historical literacy can be approached through two research fields (see Figure 2). On the one hand, it can be considered a part of history education, representing one of several segments of disciplinary history (historical thinking, historical reasoning, historical consciousness, etc.). The main focus in this dissertation is on the approach that history education takes towards historical literacy.

On the other hand, as Figure 2 shows, in language and literacy education, historical literacy can be perceived as one of many disciplinary literacies (mathematical literacy, musical literacy, etc.). As learning and teaching historical literacy cannot be fully understood without its connection to literacy education, the latter part of this chapter discusses the debates on disciplinary and content-area literacy and their relevance at the primary level.

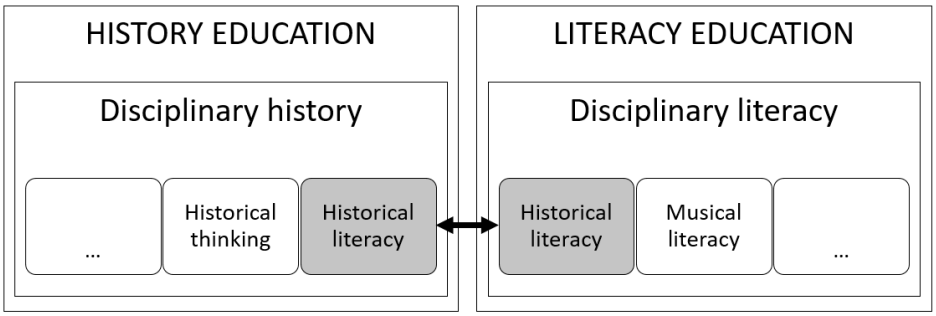

Figure 2. Historical literacy approached through two research fields.

3.1 Disciplinary history

I begin by considering the structure of historical knowledge and addressing three forms of knowledge, the understanding of which is then used in defining historical literacy. To demarcate the space of historical literacy and avoid confusion with other similar concepts, I set it within the context of other concepts representing disciplinary history.

3.1.1 The structure of historical knowledge

Whilst in everyday life historical knowledge may often be equated with factual information about events and dates, it is a far more complex construct, comprising

several forms of knowledge. In an attempt to conceptualize historical knowledge, researchers have utilized concepts such as content knowledge, substantive, factual, syntactic, procedural, epistemic, second-order, and disciplinary knowledge. I use three forms of knowledge, namely substantive, second-order and procedural, to describe the elements essential for building historical knowledge.

Substantive knowledge in history, also referred to as content knowledge (e.g., Chapman, 2021; Nokes, 2012; Nordgren, 2021; Vesterinen, 2021) or factual knowledge (Downey & Long, 2016), refers to knowledge about the people, events, dates and places of the past (see e.g., Bertram, 2019; Lee, 2005, 2011; Lévesque, 2008; Ormond, 2017). Unlike Downey and Long (2016, p. 7), for whom substantive knowledge is the "raw material" of historical knowledge, I perceive historical sources as the "raw material" of history and substantive knowledge as one form of knowledge needed to interpret that "raw material". A source´s content, its meaning and significance may remain vague without interaction with prior knowledge about the era, places, people, and events relevant to the source. This applies to things that the source contains, but also to what it does not; without prior substantive knowledge, it is difficult to detect things that have been left out of the source.

Second-order knowledge is constructed by employing second-order concepts. These concepts are also called disciplinary and historical thinking concepts (see e.g., Downey & Long, 2016; Seixas, 2006) and their role in constructing historical knowledge is essential. The choice to use the term second-order concept in the present study should not be interpreted as viewing knowledge acquired through second-order concepts as subsidiary to substantive knowledge (see the concerns raised by Downey & Long, 2016 p. 18). On the contrary, the role of second-order concepts is to organize ideas (Chapman, 2021, p. 8) and thus to give meaning to the when, what and who questions arising from substantive knowledge. For example, by employing the concepts of change and continuity, different and often apparently isolated events are given a historical meaning.

The third form of knowledge is procedural knowledge, which is typically seen – as the concept implies – as procedures and strategies involved in constructing historical knowledge. Reading strategies specific to history, such as sourcing, contextualizing and corroboration identified by Wineburg (1991), or Questioning the Author strategies (QTA) by Beck and McKeown (2002), are often included in these strategies (e.g., Downey & Long; 2016; Nokes, 2012). VanSledright (2002) has introduced strategies specifically applicable to the primary context. These strategies include asking a set of questions depending on the source type (the use of strategies is discussed further in Chapter 4).

In addition to reading strategies, procedural knowledge also includes writing strategies, which typically emphasize argumentation (e.g., Nokes, 2017; Reisman & McGrew, 2018). Reading and writing strategies that can be applied to any domain, such as summarizing, note-taking, and concept-mapping, represent

content-area literacy (Fang, 2012; Shanahan & Shanahan, 2012) and are therefore not crucial to the process of building historical knowledge.

While epistemic knowledge is often presented as a dimension of its own (e.g., Nokes, 2012), it has also been paired with procedural knowledge (see Chapman, 2021). As procedures for investigating history are derived from the epistemic fundamentals of the discipline, I suggest that procedural and epistemic knowledge are intertwined and thus belong to the same dimension of knowledge. Hence, apart from specific strategies and procedures, in this study, the concept of procedural knowledge also includes wider epistemic issues related to constructing historical knowledge. This epistemic aspect of procedural knowledge may include discussions on the nature of historical knowledge and historians´ way of thinking and working – in other words, habits of mind (Nokes, 2010, 2012; Moje, 2015) as well as the criteria for evaluating historical knowledge (Hughes, 2021).

Forms of knowledge have also been discussed by Bernstein (1999) sought to go beyond such distinctions as school and common sense knowledge or official and local knowledge because he considered them to be ideologically positioned. Instead, he introduced two types of discourse, vertical and horizontal, the characteristics of which could be described at a more general level. Vertical discourse comprises symbolic structures of explicit, specialized knowledge, which is not context-dependent. It has a set of rules that define how knowledge is accessed, transmitted and evaluated. Vertical discourse is typically set in official and institutional contexts; for example, Nordgren (2021) equates vertical discourse with an academic discipline. Thus, the aforementioned substantive, procedural and second-order knowledge are part of the vertical discourse – to use Bernstein´s (1999) terms – because they represent specialized knowledge that is not context-dependent. These forms of knowledge interact with each other and contribute to the process of producing historical knowledge. Therefore, I consider all these forms of knowledge to fall under the term disciplinary knowledge in history.

Horizontal discourse, on the other hand, is both context-dependent and specific, and is segmentally organized. In other words, different segments of horizontal discourse exist independently of each other, and contradictions can be found across but not within these segments. When defining horizontal discourse, Bernstein uses terms such as competences, strategies, and skills rather than knowledge, except when in quotation marks (1999, pp. 159–161). This further emphasizes the different roles assigned to horizontal and vertical discourses in Bernstein´s thinking. Although Bernstein´s concept of horizontal discourse seems to refer for the most part to learnable skills and strategies (tying one´s shoes) rather than experiences, later researchers have included experiences in horizontal discourse (Dempster, 2021; Nordgren, 2021). Horizontal discourse is often associated with common-sense knowledge (Bernstein, 1999, p. 159), everyday knowledge and experience-based knowledge (Nordgren, 2021). Arising from the

common problems of living, horizontal discourse is typically transferred orally within communities such as homes and workplaces.

Partly because the term "skills" is closely associated with horizontal discourse (see Bernstein, 1999), I have avoided using it when referring to any of the three forms of knowledge described above. Although many others have used the term "skills" in connection with second-order and procedural knowledge (e.g., Seixas, 2016a), I agree with Downey and Long (2016, p. 43), who suggest that skills point more to generic than domain-specific processes. The term has been widely associated with approaches that emphasize the generalizability of thinking processes. For example, 21st century skills (see Kouki & Virta, 2017) as well as the 'Future 2' in Young´s (2008) work apply the term in reference to general thinking capabilities. Hence, the Finnish term *historian tekstitaidot* (literally 'history´s textual skills') used for historical literacy (e.g., Rantala & Khawaja, 2016; Rantala et al., 2020; Rantala & van den Berg, 2013; Veijola, 2016; Wuokko & Räsänen, 2023) is particularly challenging because skills are incorporated into the concept itself and thus the role of knowledge in historical literacy is de-emphasized. However, as discussed later, historical literacy presupposes abilities that could be characterized as skills, such as general reading practice (VanSledright, 2002). Thus, it could be said that skills are not entirely excluded from historical literacy.

It has also been suggested that the use of "skills" prompts a counterproductive dichotomy between different elements of history teaching (e.g., Lee, 2011, p. 64; Lévesque, 2008, pp. 30–31). Lee (2011, p. 64) warns that "once we slip into 'skills' language, we are in in trouble" because the use of skills implies that repeated practice is sufficient for learning (e.g., writing neatly). Lee argues that because history requires reflection and judgement, it cannot be learnt only through practice. The steps involved in historical enquiry and source work will not lead to understanding without a grasp of the concept of evidence (Lee, 2011).

If looking into the biases of sources and reflecting on change over time are described as skills rather than procedural and second-order knowledge, they are by definition excluded from the process of constructing historical knowledge. This further underscores their role as something additional, extra and potentially unnecessary. For example, in Vesterinen´s study (2022), the observed teacher was sceptical of a disciplinary approach because it did not focus on the essence of history, "the command over content knowledge". Conceptualizing disciplinary knowledge as skills in the Finnish National Curriculum (2014) misled the teacher into thinking that knowledge was under threat in a disciplinary approach (Vesterinen, 2022).

3.1.2 Historical literacy

At the core of historical literacy, and in fact literacy in general, is the interaction between a text and its interpreter. In this work, the definition of "text" includes all kinds of "physical representational resources" (Cope & Kalantis, 2000), duly committing to a broad idea of what constitutes text. In addition to traditional written texts, this definition includes resources such as music, videos, visual texts (e.g., photographs, paintings) websites and artefacts (e.g., statues, household objects). The difference between the concepts of literacy and reading can also be traced back to how texts are perceived. The use of "reading" typically refers to constructing meaning through words, sentences and paragraphs, although some history educators have used "historical reading" in a broader sense (see Reisman, 2012; Wineburg, 1991). "Literacy", on the other hand, includes but extends beyond reading traditional texts: a literate person is expected to construct meaning with all types of textual resources, not only with traditional ones (Cope & Kalantis, 2000; Nokes, 2012). While some scholars use literacy for describing competencies and familiarity with something, such as cultural (Hirsch, 1988) and assessment literacy (Atjonen et al., 2022), this meaning of literacy is not relevant to the present study.

Many of the traits associated with historical literacy can be traced backed to the American contribution to "New History", which advocated a disciplinary approach to teaching history and took its first steps as early as the 1890s. However, at that time, "New History" did not spread widely and remained in the background to some extent. It was revived in the 1960s and 70s when reforms, particularly in Germany, the United States and Britain changed the ideals of history education to a large extent (Lévesque, 2008; Seixas, 2017). These reforms have largely been responsible for a myriad of concepts used for disciplinary history. This terminological challenge is discussed in the next section.

The development in the United States led to a disciplinary approach which first and foremost placed emphasis on the importance of interacting with texts in a disciplinary way. The Amhurst Project played a significant part in determining the direction that disciplinary history took in the United States.

In the Amhurst Project, history was approached through the processes and procedures essential for historians. Students were guided to act as inquirers, answering open-ended questions, which would ideally be linked to their own lives. Students´ agency was emphasized as learning was understood as "something the student did, rather than something done to him or her by the teacher". Realizing that the existing textbooks were unable to support this way of learning, extensive materials consisting of primary sources were developed (Brown, 1996).

The epistemological issues of "how do we know about the past" and "what are the limits of knowledge in history" are central to the American tradition and are

approached mainly through textual means. The emphasis on texts and reading strategies has its roots in the Amhurst Project, but the research initiated by Wineburg´s (1991) seminal work on the reading heuristics of historians has further strengthened the connection between history and literacy. In addition, as the Common Core State Standards launched in the 2010s have focused on improving literacy across subjects, literacy and history teaching continue to have a close relationship in the United States (Masuda, 2014; Seixas, 2017). Despite its American origins, the concept of historical literacy is not exclusive to the United States as it is widely used in other countries as well (e.g., Evans & Midford, 2021; Lee, 2007, 2011; Maposa & Waaserman, 2009; Rantala & van den Berg, 2013; Veijola & Rantala, 2018). As a result, several definitions of historical literacy exist.

The definitions of historical literacy vary in regard to the range of processes that the definition covers. In their review article, Maposa & Wassermanin (2009) arrive at the conclusion that historical literacy has been defined in so many different ways that nearly all aspects of school history need to be included in the definition. Nokes´s (2010, 2012, 2014) definition, on the other hand, represents one of the narrower definitions in the research literature, and is one that I lean on considerably in my work. The connection to texts is at the core of Nokes´s definition as all the elements he includes in historical literacy involve either reading or producing texts. He views historical literacy as a student´s ability to gather and weigh evidence from multiple sources, as well as use historical accounts to construct and defend an interpretation (Nokes, 2012). Historical literacy can thus be defined as "the ability to appropriately negotiate and create the texts and resources that are valued within the discipline of history using the methods approved by the community of historians" (Nokes, 2012, p. 13). One essential aspect is engaging with the historical process, that is to say building historical knowledge (Nokes, 2010), which in turn requires multiple forms of knowledge (Downey & Long, 2016).

In terms of different forms of knowledge, historical literacy calls for substantive, procedural and second-order knowledge, and the interaction between them. Downey and Long (2016, p. 7) underscore that students should "become literate about the discipline as well as the subject matter of history". This suggests that the interplay between procedural knowledge – including epistemological considerations – and substantive knowledge is viewed as particularly crucial. Although Downey and Long´s book also introduces second-order concepts, their role in forming historical literacy remains vague. Similarly, others (Nokes, 2010, 2012) have emphasized that there is a cyclic relationship between substantive and procedural knowledge because in order "to make sense of what a source may say", substantive knowledge is required (Kitson, 2021, p. 36).

In addition to the three forms of knowledge, all part of vertical discourse, historical literacy presupposes a connection with horizontal discourse, namely

everyday and personal experiences as well as texts produced by historical culture (Downey & Long, 2016; Hughes, 2021; Nokes, 2012). This contradicts Bernstein´s (1999) original idea because, for him, recontextualization involves selecting, appropriating, relocating and refocusing elements of vertical discourse, in other words the knowledge produced by the disciplines. He does not recommend that the pedagogic discourse in schools should include horizontal discourse due to its context dependency, not even for the purpose of accessing vertical discourse. On the other hand, classics (Bruner, 1960, p. 11) as well as later research have underscored the importance of including horizontal discourse when the aim is to produce meaningful knowledge (Dempster, 2021; Nordgren, 2021).

Horizontal discourse does not form the basis of historical literacy as vertical discourse does. Nevertheless, developing pupils´ historical literacy cannot take place by excluding the experiences and texts they face outside the classroom. Historical literacy should prepare pupils to face horizontal discourse, such as historical fiction, through disciplinary means (Nokes, 2012). Participation in today's society demands interaction with it. Were horizontal discourse excluded from the definition of historical literacy, it would run the risk of pupils forming parallel and separate ways of looking at history; one concerning everyday life and one school history.

One of the most widely agreed-upon features of historical literacy is that reading texts involves analysing and interrogating them (e.g., Downey & Long, 2016; Nokes, 2010, 2012; Rantala & van den Berg, 2013). Sources may be used as evidence in answering historical questions (Ashby, 2011). In other words, conceptual understanding of evidence is needed because not all sources are automatically evidence (Downey & Long, 2016; Lee, 2005, 2007) although the concept of evidence is sometimes equated with primary sources (see Nokes, 2012). A historically literate person understands the limitations of different source types. For example, primary sources as fragments of the past require interpretative processes in order to reconstruct the past. Secondary and tertiary sources, on the other hand, already have an interpretative layer built into them (Downey & Long, Nokes, 2012). The definitions of different source types are described in Chapter 4.

Empirical studies suggest that historians have a distinct way of approaching texts (Shanahan & Shanahan, 2008; Wineburg, 1991). As the aim of any disciplinary literacy is to adopt the practices and habits of mind, these history-specific literacy strategies are at the core of historical literacy. Wineburg´s (1991) heuristics of sourcing, contextualization and corroboration are processes that are perhaps not restricted to historians, but typical of them. Instead of starting with the content of the text, historians tend to pay attention to the background of the source, addressing its author, date and general circumstances. Moreover, historians take into account the possible bias of the author, which will influence the interpretational process. Historians also aim to recognize their own bias

regarding the author´s bias (e.g., opposing values) and accept that both author and reader are positioned (Shanahan & Shanahan, 2008). Unlike in some other disciplines, historians do not embark upon their reading by looking at the content of the text (Wineburg, 1991; Shanahan & Shanahan, 2008).

Historians also cross-check and compare sources because the validity of claims is often difficult to establish using just one source. However, merely possessing multiple sources does not result in corroboration; the texts should be compared for a specific purpose (Wineburg, 1991). The third heuristic suggested by Wineburg is contextualization, whereby a source is placed in its historical context. In order to make meaning with a text, historians attach it to the circumstances in which the source was created. This includes anything from wider political and societal structures to considering the physical environment where the source was created. Contextualization also requires constructing a chronological timeline in which the text is situated. Contextualization can be viewed as a part of historical empathy (Lévesque, 2008), one of the most relevant second-order concepts for historical literacy.

Historical literacy requires empathetic reading, which develops through historical empathy; in other words, the reader attempts to put herself in the position of those who have lived in the past (Lévesque, 2008). Imagining, interpreting and experiencing the world through the belief systems of past societies is the key to historical empathy. This is by no means a simple exercise as a different set of values and morals existing in the past have to be taken into account, duly requiring one to distance oneself from the present (Lee, 2005; Rantala & van den Berg, 2013). As contemporary actors, we must ask how, with what reservations and under what conditions we can presume to understand the views and ideas of others (Seixas 1996, in Lévesque, 2008). One of these conditions is the extent to which a person in the present is assumed to be able to relive the feelings and thoughts of historical actors (Lévesque, 2008). According to Shemilt (1984), re-imagining the past may result in an attempt to mentally become the subject in question and lose one´s own identity in the process. This approach makes the assumption that no boundaries exist between now and then and dismisses the fact that "there are limits to what the mind can imagine, feel or accurately relive" (Lévesque, 2008, p. 145). The opposite pitfall in the imaginative process is to frame past action in a way that makes it plausible and recognizable for contemporary audiences. Empathetic and historical literacy needs a third way where the interpreter does not aim to become one with historical actors but to "mentally relive events from the situation" (White, 1978).

For Lee (2007), historical literacy includes both knowledge of the discipline and orientating in time. He states that the aim of historical literacy is to understand "the nature of historical claims so that students can arbitrate between rival stories on historical grounds", as well as to enable students to "make sense of their world in time" (Lee, 2007, p. 60). This dual approach to historical literacy therefore

bears similarities to historical consciousness. Despite including a time orientation perspective, Lee´s (2005, p. 210) conception of historical literacy is firmly rooted in the concept of evidence and its interpretation: "historical literacy demands that students learn to evaluate arguments and decide which positions, given the evidence, are more or less plausible, better or worse".

Although reading heuristics have gained considerable attention in the literature concerning historical literacy, writing is an inseparable part of literacy. Reading and writing are to an increasing degree seen as involving similar cognitive processes and as supporting one another (see Graham & Hebert, 2010). They could be described as two sides of the same coin, as both are part of a historian´s process. However, it should be noted that historically literate writing is based on evidence, which in turn requires reading. Historically literate reading, on the other hand, may benefit from writing, but can exist independently of writing. Historically literate writing is most often described as argumentative, as the aim is to use evidence to justify an interpretation (Nokes, 2012; Nokes & De La Paz, 2018) and to utilize the conventions of historians (De La Paz et al., 2016).

Contrary to what Ahonen (2020a) suggests, the aim of historical literacy, or "New History" in general, is not to prepare pupils to produce "objective explanations" (p. 133) about the past. Instead, historical literacy and historical thinking require pupils to understand the interpretive nature of history, which by definition means acknowledging the existence of several competing interpretations, and that reconstructing history is not an objective endeavour (e.g., Fillpot, 2012; Lee, 2005; Lévesque & Clark, 2018; Nokes, 2012; Reisman & McGrew, 2018; VanSledright, 2002).

In conclusion, in this study, historical literacy refers to specific and rather narrowly defined processes related to understanding and producing texts in ways that are characteristic of the discipline of history. A historically literate person is able to interpret a text by considering by whom, why and in what kind of context it was produced, and to reflect on the motives and biases of those behind the text. The process includes an active interaction between different forms of knowledge and an intent to build one´s own evidence-based and justified interpretation. Rather than acquiring things about history, historical literacy aims at understanding what history is about.

Some have argued that risks are involved in defining the concepts used for history teaching in too concise a manner because narrow definitions may result in decreasing the diversity inherent in teaching history (Veijola, 2016). However, I propose that the advantages of using narrow rather than broad concepts outweigh the risks in two ways. First, the use of broad concepts has pedagogical implications. If research is to inform and benefit the classroom level, the concepts should be easily understandable, not only to researchers but to teachers as well. Using several but more narrowly defined concepts may help in this quest. Second, the quest for an all-encompassing concept (see van Drie & van Boxtel, 2008;

Maposa & Wasserman, 2009), which would describe all the main aims of history teaching, makes discussion among researchers more difficult. Furthermore, using different concepts for similar phenomena also affects the accumulation of empirical knowledge because comparing the results becomes more challenging. This terminological complexity is addressed next.

3.1.3 A web of concepts

Terms such as historical thinking (e.g., Husbands, 1996; Seixas, 2006, 2017; VanSledright & Frankes, 2000; Wineburg, 2001), historical reasoning (e.g., Luis & Rapanta, 2020; van Drie & van Boxtel, 2008), historical consciousness (e.g., Nordgren & Johansson, 2015; Köbl & Konrad, 2015), historical understanding (e.g., Monte-Sano & Reisman, 2016) and historical literacy (e.g., Downey & Long, 2016; Nokes, 2010; Lee, 2007) have been used for conceptualizing the aims of history teaching. The myriad of concepts in history education has led to a lack of a common vocabulary (Seixas, 2016b), as even definitions of a single term can differ significantly. In addition to the quantity of similar terms, their quality poses a challenge: while some terms refer to an activity or a process (thinking, reasoning), others point to an ability (literacy) or to a state of being (consciousness, understanding). In what follows my aim is to further illustrate historical literacy through these related concepts.

The review article by Luis and Rapanta (2020) shows that while historical thinking, reasoning and consciousness emphasize different aspects of history, the terms also partly overlap. For example, the Dutch model for historical reasoning (van Drie & van Boxtel, 2008) shares many elements with the Canadian model for historical thinking (Seixas, 2006), which in turn has been influenced by German, English and American traditions (Seixas, 2017). As a result of the terminological controversy, no consensus has been reached on the hierarchy of the terms. Depending on the definitions, some view historical reasoning as a broader term than historical thinking (Luis & Rapanta, 2020), while others point out that historical thinking may go "beyond the borders of historical reasoning" (Munoz & Balmaceda, 2022, p. 3). Both historical literacy (Maposa & Wasserman, 2009) and historical thinking (Parkes & Donnelly, 2014) have been suggested as umbrella terms encompassing all the others. Lee (2011, p. 63), on the other hand, views historical literacy as nesting "in the wider idea of historical consciousness". Thus, although research on teaching history has resulted in general agreement about the elements that should be included in disciplinary history, differing views on the terminology and the relationship between the terms used still exist.

All of these models and approaches are the result of a recontextualization process conducted by researchers, who have differing views on 1) the overall purpose of teaching history, 2) what aspects of history as a discipline are most

relevant for learning history, and 3) the extent to which things external to history (e.g., the ideal of an active citizen) should play a part in school history. Hence, I argue that the diversity of these terms and theoretical models makes it difficult to pinpoint "the disciplinary" way of teaching history (cf. Vesterinen, 2022). Instead, historical thinking, reasoning, consciousness, understanding and literacy all have their distinct ways of representing disciplinary history.

Historical thinking is perhaps the most used and conceptualized among all the related concepts. Its origins go back to the 1970s when the Schools Council History Project (SCHP) in England identified a conceptual structure for history. The aim of the SCHP was to introduce a more disciplinary approach to history, which would, for example, give pupils a more active role in learning, and would assess historical understanding rather than the memorization of facts (Dawson, 1989). As Seixas (2017) points out, the significance of the British contribution to developing disciplinary history teaching from the 1980s onwards lies in the fact that theoretical, conceptual groundwork (e.g., Booth, 1987; Portal, 1987; Sansom, 1987) was complemented with a body of empirical work on pupils´ progression in historical thinking (e.g., Ashby & Lee, 1987; Shemilt, 1987). As the SCHP succeeded in replacing the traditional chronological curriculum with one focusing on history as a distinct form of knowledge, it has been described as one of the most successful reform projects in the field (Stearns et al., 2000).

The various frameworks for historical thinking that emerged in the decades that followed are largely based on the SCHP. For example, history educators in Canada created a model consisting of six historical thinking concepts: 1) significance, 2) evidence, 3) continuity and change, 4) cause and consequence, 5) historical perspective, and 6) moral dimension (Seixas, 2006). The framework by Lévesque (2008), on the other hand, does not include cause, consequence and the moral dimension, and instead of historical perspective, Lévesque uses the concept of historical empathy. Most of the frameworks for historical thinking include five or six concepts with some variation. In addition, Downey and Long (2016) include the concepts of time and chronology as well as agency in their framework. The concepts of historical thinking are also referred to as disciplinary concepts (Downey & Long, 2016), procedural concepts (Lévesque, 2008) and perhaps most often second-order concepts (e.g., Chapman, 2021; Lee, 2005; Nordgren, 2021; VanSledright, 2011).

Despite apparent similarities between historical thinking and historical literacy, their exact relationship has seldom been defined. The concepts have often been used by the same authors, even in the same publication (e.g., Downey & Long, 2016), without defining their relationship. Nokes (2014, p. 377) does define historical literacy "as a vital element of historical thinking" but uses the concepts interchangeably because of "the significant overlap" between them. Seixas (2006), on the other hand, has defined these two concepts in relation to each other.

In the Canadian Benchmarks of historical thinking, Seixas (2006, p. 2) states that "historical thinking concepts, taken together, form historical literacy".

I view the two concepts as related but not interchangeable. Historical thinking is based on second-order knowledge, which aims to organize, structure and explain the who, when, where and what questions in history (see Chapman, 2021, p. 13). Historical thinking includes a broad range of processes that aim to capture all the aspects considered relevant for achieving historical understanding. I agree with Reisman (2015, p. 6), for whom "historical thinking depends upon but extends beyond historical reading". Historical literacy always requires an interaction with a text of some type. Some aspects of historical thinking, on the other hand, are not text-dependent. For example, reflecting on the changes and continuities does not necessarily require interaction with texts. Moreover, while evidence is only one of the elements of historical thinking (Downey & Long, 2016; Lee, 2005; Lévesque, 2008; Seixas, 2006), for the purposes of historical literacy, evidence and accounts play a more central role than some of the other second-order concepts.

3.2 The purpose of literacy

The tensions between different school subjects, described in Chapter 2, have implications for the way in which the nature and purpose of literacy is perceived. My aim below is to describe the ways in which disciplinary and content-area literacy differ in their approach to the kind of literacy that is needed to learn history. Moreover, the section touches upon literacy development in order to find out the prerequisites for teaching historical literacy at the primary level.

3.2.1 Content-area and disciplinary literacy

Content-area literacy means that literacy is viewed as a general ability that is not dependent on or specific to various disciplines (Faggella-Luby & Deshler, 2008). Thus, content-area, also referred to as general (e.g., Nokes, 2010) or generic (Heller, 2010) literacy, assumes that the same processes and ways of thinking about texts can and should be applied to texts in any context (see Shanahan & Shanahan, 2012). It is suggested that strategies such as summarizing, paraphrasing and reading comprehension exercises are sufficient for learning in any subject areas (see Fang, 2012).

One suggested benefit of content-area literacy is that it may help in comprehending and memorizing the content of a text (Friend, 2001; Shanahan & Shanahan, 2012). The aim is then to use literacy to extract the content provided by different disciplines and school subjects. Instead of constructing knowledge, the focus is – as the term content implies – on adopting content (McKenna & Robinson, 2014). As noted, in content-area literacy, the logic of "one literacy fits

all" is used for covering content in different subject areas. However, the reverse may also apply: subject areas such as history can be used for improving general literacy if they are perceived as platforms from which to develop basic reading and writing skills.

Content-area literacy in history classrooms would entail a focus on the acquisition of information regarding content such as dates, places and events of the past. This would lead students to summarize textbook chapters, write bullet points on the most essential facts, and find out the meaning of given concepts, to name but a few. As for visual texts such as paintings, the emphasis would be on recognizing the object depicted in the painting, such as a king or a battle, and memorizing the facts. As Moje (2008) points out, these general literacy strategies run the risk of creating passive information consumers rather than preparing learners to create and critique.

The premise of disciplinary literacy is different and to some extent the opposite of that of content-area literacy. A disciplinary literacy approach underscores that because disciplines and disciplinary experts have their own unique ways of both producing and processing texts, literacy instruction should not ignore these differences (e.g., Fang, 2012; Moje 2008, 2015; Shanahan & Shanahan 2008 2012). Thus, disciplinary literacy can be defined as "specialized literacy practices of a given disciplinary domain" (Moje, 2015, p. 256). In addition to literacy practices, disciplinary literacy may refer to domain-specific cognitive processes, "habits of mind" (Fang, 2012, p. 20).

Differences between disciplines in their manner of approaching texts are not only based on theoretical assumptions but also supported by some comparative empirical studies (e.g., Shanahan & Shanahan, 2008) as well as studies focused on specific disciplines. For example, Shanahan and Shanahan (2008) compared the way experts in mathematics, chemistry and history read disciplinary texts and used reading strategies. They found that while the focus of the mathematics expert was on close reading (e.g., differentiating between "a" and "the"), the chemistry expert compared alternative representations (e.g., graphs, tables). The history experts, on the other hand, began their reading process by identifying the author and their possible bias.

As described above, disciplinary literacy has been defined through discipline-specific literacy practices and ways of thinking. However, language and literacy experts also look at disciplinary literacy in relation to vocabulary, grammar, language structure and functionality, and further suggest that disciplinary learning is primarily a linguistic process (see Fang, 2012). For example, Shanahan and Shanahan (2012) describe how words of Latin and Greek origin can be used for supporting disciplinary understanding in science classrooms. Instead of aiming to memorize and organize words, and repeatedly match words with their meaning (content-area literacy), concepts such as annual, biennial, and perennial should be analysed. This entails encouraging students to think about the purpose of this

scientific terminology and the way words with Latin and Greek roots are constructed.

Another linguistic approach to disciplinary literacy is that of functional linguistics, which emphasizes the importance of recognizing and understanding the discourse grammar of disciplinary texts, which refer to texts produced by disciplinary experts (Fang, 2012). These language patterns, once identified and analysed, can be useful in understanding both the nature of the discipline as well as the content at hand (Schleppegrell, 2004, 2016). For example, Fang and Schleppegrell (2008) found that when describing causation, texts within natural science shift focus from social agents to natural agents. History, on the other hand, makes its meaning more through the use of verbs, and when addressing causation, human agency has a central role. Unpacking these structures may decrease the level of abstraction in disciplinary texts and help students to understand the differences between disciplines.

The debate over the advantages and disadvantages of disciplinary and content-area literacy approaches is ongoing, and at times, a lively process (see Moje 2008; Heller, 2010; Moje, 2011). In reality, there is a realization on both sides that both types of literacies are needed and that the debate is mainly about when and how these literacies should be presented in classrooms. Critics of disciplinary literacy have been concerned about introducing disciplinary literacy too early (Heller, 2010), and to learners who struggle with general literacy (Faggella-Luby et al., 2012). Heller´s (2010) concerns stem from the perception that disciplinary literacy is synonymous with "postsecondary literacy instruction", that is, literacy taught in the disciplines (Heller, 2010, p. 270). Thus, unlike others (e.g., Fang, 2012; Shanahan & Shanahan, 2008), he does not recognize the possibility of adjusting disciplinary literacy according to the educational level.

Irrespective of the stand one takes in the debates described above, the underlying assumption seems to be that literacy is a separate entity from disciplines or school subjects. In the research concerning language and literature, literacy is often coupled with verbs such as "infuse" (Hinchman & O´Brien, 2019), "incorporate" (Howard et al., 2021) or "integrate" (Howell et al., 2021; Orr et al., 2014). Thus, literacy is brought into the space of school subjects, and depending on the stand taken, the literacy introduced in subject-area classrooms would emphasize either content-area or disciplinary literacy.

In Finland, the discussion on literacy has been less dichotomous. The influence of disciplinary differences on literacy seems to have been widely recognized among language experts (e.g., Juvonen & Routarinne, 2018; Sulkunen & Luukka, 2014; Sulkunen et al., 2019). However, the Finnish literacy discussion involves terms which could be described as general rather than discipline-specific. For example, critical literacy is often referred to as a general ability applied to a number of school subjects (see e.g., Kouki & Virta, 2017; Luukka, 2014). While critical literacy shares many features with historical literacy (Kouki & Virta,

2017), it suggests, however, that the act of being critical in relation to texts can be applied across subjects. Further, literacy experts often use the term "multiliteracies" instead of or alongside disciplinary literacy (e.g., Juvonen & Routarinne, 2018). Nevertheless, the role of disciplines in literary instruction has not been questioned and one of the reasons for the strong position of disciplinary literacy in the Finnish literacy discussion may be traced back to the obligatory nature of national curricula: disciplinary literacy is stated as an instructional aim (Finnish National Agency of Education, 2014). This may have contributed towards a less controversial approach than in the United States, for example, where national curricula do not exist. However, the lack of active Finnish discussion on the balance between disciplinary and content-area literacy cannot be taken as confirmation that disciplinary literacy has a strong position in classrooms.

3.2.2 Literacy at the primary level

Literacy can be perceived as a continuum, where different ways of using language enter pupils´ lives at different stages. Several models have been built on accumulating language and literacy abilities, where new ways of understanding and using the language do not replace the existing ones (e.g., Halliday, 2004; Shanahan & Shanahan, 2008; VanSledright, 2002). According to Halliday (2004), children go through three essential stages in developing their language and literacy. After learning the spoken language during the early years and using it in everyday life, children are introduced to written language quite universally between the ages of four and six. This, according to Halliday, does not result only from practical issues related to schooling but because children at that particular age begin to process abstract meaning and use abstract symbols. Through this basic literacy, pupils in school are able to move from commonsense knowledge to educational knowledge, which includes extending abstraction to categories as well, thereby thinking in more general terms. The third stage introduces "the language of specialized disciplines", knowledge that is technical and grounded in a theory of some sort (Halliday, 2004, p. 19). This stage starts at the age of nine and continues throughout the educational path. However, it should be noted that Halliday´s model (2004) is mainly concerned with the grammatical side of literacy and less with ways of thinking in the disciplines.

A literacy progression model by Shanahan and Shanahan (2008) also suggests that pupils first develop basic, then intermediate and finally disciplinary literacy. This model is presented as a pyramid, where the basic literacy level forms the broadest part at the base, narrowing towards the disciplinary level at the apex. According to Shanahan and Shanahan, most students in the United States achieve the intermediate level by the end of middle school (aged 11–13), although some struggle with general literacy. The disciplinary literacy level is suggested to begin

in middle and high school in the American context. The fact that the model is presented as a pyramid conveys that not everyone reaches the disciplinary level. However, the Shanahans point out that while the difficulty of disciplinary literacy is a contributing factor, the narrowing tip is mainly a result of not teaching these higher-level processes.

Both progression models (Halliday, 2004; Shanahan & Shanahan, 2008) emphasize that disciplinary literacy is progressively built on basic literacies, which continue to develop alongside disciplinary literacy. Further, these models suggest that pupils as young as nine can be introduced to some features of disciplinary literacy. Empirical work on primary pupils´ progression concerning different types of literacies is scarce. However, in their intervention study, Swan and colleagues (2008), for example, investigated whether fifth graders used more general or history-specific literacy. Using VanSledright´s (2002) literacy continuum, the researchers found that approximately half of the participants´ written answers fitted into the category of "historical reading and interpretation", while half had "global reading and interpretation" features. Prior to the intervention, the 71 participants in the study had not received explicit instruction in interpreting conflicting primary sources. Thus, only limited conclusions can be drawn from the study about the literacy progression of primary pupils.

In the Finnish context, the primary level continues until pupils are typically 12 years old and can be expected to interact with texts in a disciplinary way. The circumstances for introducing disciplinary literacy in Finnish primary classrooms are favourable. The results of the 2016 Progress in International Reading Literacy Study (PIRLS) showed that Finnish fourth-graders have a good command of basic literacy. Of the 50 countries taking part in the study, Finnish pupils scored the fifth highest, and although there was some decline compared to the 2011 PIRLS, there was no significant change in the level of performance (Leino et al., 2017). The results also reveal that Finnish fourth-graders were much more proficient in retrieving information from texts than interpreting them, and this difference remained considerable compared to countries with similar education systems, such as Norway, Sweden and Denmark.

From the perspective of primary teachers, the FNCC (National Board of Education, 2014) requires a three-pronged approach to literacy. First, primary teachers teach the subject of mother tongue and literature, which sets its own objectives for literacy. Basic reading and writing competencies are one of the main aims at the primary level. Second, the syllabi of other school subjects, such as history, have their own disciplinary literacy aims. A third form of literacy stated in the FNCC (Finnish National Board of Education, 2014, pp. 168–169) is that of multiliteracy, which is defined as a "transversal competency" and is therefore intended to be addressed across school subjects and throughout instruction.

Although the concept of multiliteracy cannot be discussed in detail here, its definition in the FNCC reveals a detachment from disciplinary literacy. Neither

the general description (pp. 22–23) nor the more specific description of multiliteracy in grades 3–6 (pp. 169–170) includes the concept of "discipline", as the following excerpt shows:

> The pupils' multiliteracy is developed in all school subjects, progressing from everyday language to mastering the language and presentational modes of different ways of knowing" (FNCC, p. 22).

Thus, while the lengthy definitions of multiliteracy in the FNCC acknowledge the existence of several literacies, they do not specify disciplines as their origins. In addition, disciplinary ways of thinking and practices are not included in the concept of multiliteracy. Consequently, multiliteracy as defined in the FNCC has little in common with disciplinary literacy and, as a result, with historical literacy (cf. Rantala & van den Berg, 2013, p. 395).

As Kela (2012) notes, multiliteracy as defined in the FNCC through a cluster of attributes (but excluding disciplines) poses a challenge for teachers; the national curriculum as a normative and therefore quite pragmatic document should provide teachers with concreteness, which multiliteracy in its current form lacks. I argue that if multiliteracies were to have a more explicit connection to disciplines, there would be an increase in the level of concreteness. This, in turn, would help primary teachers to navigate between different literacies.

4 The ideals of teaching historical literacy

This chapter discusses the expectations and ideals set for historical literacy in the primary school context. The ideals introduced here are founded upon the definition of historical literacy described in the previous chapter, and address texts, teaching practices and assessment. The focus is on the primary level, but since, as mentioned earlier, empirical studies have focused on the secondary rather than the primary level, I also utilize research concerning the secondary level.

4.1 Texts

Teaching historical literacy requires not only exposure to and engagement with a variety of text types, such as written, visual and artefactual, but also the ability to differentiate between the various functions that texts have. In what follows texts are discussed as primary, secondary and tertiary sources according to their function in investigating history. Fictional texts are also included in the discussion. All of these serve a different purpose in developing historical literacy.

4.1.1 Primary sources

"Primary sources provide a foundation upon which the discipline of history is built", (Nokes, 2012, p. 81). Since one of the main aims of historical literacy is to engage in cognitive processes similar to those of historians and to help pupils become knowledge constructors (Hughes, 2021a), it follows that pursuing historical literacy involves the use of primary sources.

In the present study, the term primary source covers varying text types, such as written accounts, photographs, paintings and artefacts (see Downey & Long, 2016). Rather than the text type itself, primary sources are defined here by their relationship to the historical question for which they intend to provide evidence. Therefore, any text´s function can be a primary source depending on the topic to be investigated. For example, even a school history textbook can be a primary source when examining how textbooks at a given time have portrayed a certain event or topic. Before elaborating on the advantages of using different types of primary sources, it should be emphasized, as most scholars in the field have done (e.g., Booth 1994; Nokes, 2012; Wineburg, 1991), that the mere existence of primary sources in the classroom does not result in historical thinking or literacy; pupils need to be taught how to interact with the texts. This aspect is discussed later in this chapter.

One way to understand the interpretative process involved in constructing historical knowledge is to interact with multiple, possibly conflicting primary sources as evidence. Although primary sources are influenced by the motivations and points of view of those who created them, they are not distorted by other people´s interpretations (Nokes, 2012; VanSledright, 2002). Hence, primary sources provide the opportunity to engage in the process of building knowledge, and thus to better understand the limitations of secondary and tertiary sources. The essential role of primary sources is further emphasized at the primary level, where pupils may not yet have the knowledge needed for evaluating interpretations presented in secondary (e.g., historical research) and tertiary (e.g., textbooks, documentaries) sources. The findings in several studies suggest that interacting with primary sources develops different aspects of historical literacy, such as the ability to evaluate and cite texts (Rouet et al., 1996), to read for historical meaning (Hynd-Shanahan et al., 2004), and to use discipline-specific strategies (Fillpot, 2012).

Despite constant efforts at the curricular level to introduce primary sources into history classrooms, the prevalence and manner of using primary sources has not been studied extensively, especially in the form of large-scale studies. One exception is from Belgium, where Van Nieuwenhuyse and colleagues (2017) observed a total of 88 lessons in 55 classrooms to investigate the use of primary sources. No primary sources were used in 22 lessons, and in 17 lessons one primary source was used. On average, 3.65 primary sources were present per lesson. The majority (71%) of primary sources used were visual texts. Case studies carried out at the secondary level have also increased our knowledge of the use of primary sources. The teacher in Monte-Sano´s case study (2011) used primary sources extensively, as 82% of all texts were primary sources, used for 92% of the time. On the other hand, the nine teachers in Nokes´s study (2010) used tertiary sources four times more than primary sources on average, although there were differences among teachers. In his case study, Cuban (2015) reported four teachers out of seven who were relatively active in employing primary sources, although he did not report any quantitative data.

In the Nordic context, teachers have been reported to use primary sources at the secondary level, but in Iceland, for example, teachers used primary sources mainly for setting the context or for close reading (Gestsdóttir et al., 2019). Rosenlund´s study (2016) did not include classroom observations, and nor did it examine the type of texts used, but an analysis of teacher-made assessment tasks (n=940) revealed that more than 93% of the assessment tasks merely focused on historical content and were thus unlikely to utilize primary sources. Only 1.2% of the tasks were concerned with a category named “historical methods”, where primary sources may have been used. Results from our observations at the secondary level (Puustinen & Khawaja, 2022) point to a similar trend; the role played by primary sources in classrooms was a minor one.

Similar studies at the primary level and on the types of texts used in classrooms are few and far between. While it has been shown that researchers teaching primary pupils employ primary sources effectively (Nokes, 2014; VanSledright, 2002), only a few case studies have focused on observing the kind of texts that are used by primary teachers in non-intervention studies. One of the rare examples is Hughes´s (2021a) study on a third-grade teacher´s literacy practices. The teacher used versatile texts, both primary and secondary sources as well as fiction. The primary sources included photographs, engravings, lithographs, maps, political cartoons, and documents, and the use of evidence was frequent. However, as Hughes (2021b) notes, the teacher was an expert teacher. Although Hughes (2021b) offers no exact definition of what differentiates an expert from a non-expert, the teacher considered an expert in his study had participated in professional development projects, and had focused on disciplinary literacy throughout her career. Thus, a non-expert teacher in a primary history classroom could be someone who does not actively and consistently interact with matters related to teaching history.

Bruner´s idea that "Intellectual activity anywhere is the same, whether at the frontier of knowledge or in third-grade classrooms" (1960, p. 22) draws a parallel between what takes place in the disciplines and what could take place in primary classrooms. Applying this idea to texts underscores the significance of primary sources, which are the basis of the "intellectual activity" of the work conducted by historians. However, since historical literacy is a work in progress at the primary level, some essential differences have to be considered. One of these is the process of acquiring primary sources. Work conducted in archives is a key part of the historian´s work "at the frontier of knowledge", that is to say in the disciplines where knowledge is produced. Young pupils, on the other hand, lack both the means and the expertise to engage in similar work and most of the primary sources used by young learners are provided by the teacher. In other words, pupils cannot be expected to find versatile and relevant primary sources on their own. Thus, the texts used as evidence in primary classrooms are pre-selected by the teacher or by textbook authors.

Primary sources have to be made accessible to pupils who, in addition to learning disciplinary literacies, are simultaneously strengthening their general literacies (Duhaylongsod et al., 2015; Shanahan & Shanahan, 2008). This means that the length, vocabulary and structure of the text may have to be modified in order to offer accessibility to primary pupils. The process may also include translating sources from their original language. While historians would not compromise the authenticity of primary sources, modifying sources is considered not only acceptable but also necessary by researchers, even at the secondary level (e.g., Fillpot, 2012; Reisman, 2012).

One way to work with primary sources without subjecting them to modification is to increase the role of visual texts and artefacts in the primary

context (see e.g., Hughes, 2021a; Johansson, 2019; Levstik et al., 2014). Interpreting artefacts can be a complex process because it requires knowledge about the historical context and about the visual language of a certain era. However, it does not involve decoding of written text, something that may be time- consuming for primary pupils. Johansson (2019) investigated pupils´ historical enquiry and interpretation about the Viking age. Instead of reading what Johansson describes as ethnocentric narratives about the Vikings in textbooks, the pupils in the study were provided with artefactual texts. The results show that pupils as young as ten are able to interpret primary sources with the right kind of scaffolding. In fact, Johansson suggests that primary pupils should begin their journey with history by investigating artefacts rather than through textbook narratives, which have their own shortcomings.

4.1.2 Interpretative and fictional texts

Stories and narratives are a common way of looking at history because they are able to capture how experiences are naturally ordered and structured (see Coffin, 1997). A heritage approach to history in particular relies on narratives because they help to convey the desired version of history (VanSledright, 2016). On the other hand, a narrative, "a coherent representation of past human actions as the story of these events" (Lévesque, 2008, p. 24), has been claimed to force order onto a chaotic, disorganized and messy past: in order to produce a coherent narrative, historians have to reconstruct and recreate the past, not only to describe or explain it (Tosh, 2015).

Narratives come in many shapes and forms, and understanding the nature and limitations of different kinds of narratives is a prerequisite for historical literacy. I make a distinction here between fictional and non-fictional narratives and refer to the latter with the term interpretative texts. These include both secondary and tertiary sources. The former are constructed by historians´ work on primary sources and the latter include text types such as history textbooks, which may not be based on primary sources (Nokes, 2012). Nevertheless, both secondary and tertiary sources are produced through an interpretative process, which is why the common umbrella term is used for both functions.

The challenges related to interpretative texts are mentioned in Coffin´s (1997) work, where she identifies different genres found in texts about history. The genres in textbooks are of special interest because studies show that a textbook is one of the main textual resources in history classrooms (Cuban, 2015; Nokes, 2010). Textbooks have been criticized for several reasons, many of them related to their narrative and authoritative nature (Bain, 2006; Paxton, 2002; Schleppegrell et al., 2004).

In simple terms, Coffin´s work on systemic functional linguistics (SFL) reveals two main genres, namely narrative and argumentative, which in turn are further

divided into more specific genres. Narrative genres, such as recounts and accounts, are typically used for describing a chronologically organized past, where the author´s voice is impersonalized and hidden. As the timespan of recounts covers several generations, it follows that a significant amount of selection, editing and linearizing is involved. This in turn results in a portrayal of history as a straightforward, value-free and objective record of the past, where events merely follow one another. Accounts as a genre share many features with recounts, the difference being that in addition to using only time as a principle in sequencing, accounts also contain some causal links.

Tertiary sources such as textbooks are mostly written in these narrative genres, as shown by a study in which eighth-grade history textbooks were analysed by Schleppegrell and colleagues (2004). Although no systematic analyses of the genres of Finnish primary history textbooks have been conducted, recounts and accounts seem to be the predominant genres here as well (see Rantala et al., 2020). Herein lies the pitfall of using mostly textbooks in the classroom: pupils adopt a version of history as a set of simplified stories told by an invisible narrator. The narrative genres do not make visible the varying interpretations, biases and choices unavoidably included in interpretative texts. Argumentative genres (Coffin, 1997, p. 198), on the other hand, not only explicitly "advocate a particular interpretation of the past" but also negotiate with differing interpretations. Moreover, the varying aspects and uncertainty of knowledge are not hidden from the reader.

Young children are generally exposed to narrative rather than argumentative genres (Duke, 2000), which are typically found in secondary sources. Nevertheless, one of the aims of history education in many countries is to teach students to produce argumentative answers or essays (Coffin, 1997). This predicament is aptly summarized by Nokes and De La Paz: "students who rarely read argumentative texts are unlikely to produce argumentative texts" (2018, p. 554). Historically literate instruction acquaints pupils, and subsequently students, with these argumentative genres they are expected to write in. Hence, interacting only with the narrative genres found in textbooks does not suffice.

Despite the importance of primary, secondary and tertiary sources in learning history, history is present in people´s lives largely through their everyday environments (Rantala & Ahonen, 2015). This includes what Nokes (2012) calls public history and what is also referred to as historical culture (Rantala & Ahonen, 2015). Texts such as films, TV series, games, books and comics, to name but a few, utilize history, and while serving the purpose of entertaining people, they also exert a significant influence people´s perception of history (Butler et al., 2009), perhaps more than academic and school history. Finns in particular are active consumers of history-related fiction (Torsti, 2012). At the Finnish secondary level, Vesterinen observed the importance of historical culture for students (2022). Moreover, Rantala (forthcoming) found that Finnish student

teachers construct their perception of the events of the Second World War in Finland mainly on the basis of different versions of the film based on Väinö Linna's novel *The Unknown Soldier*.

Teachers use fictional texts such as films in classrooms mostly for their content, and in some cases for inducing historical empathy. Films are not employed for building procedural knowledge or for reasoning about the films (van Niuwenhuyse, 2016). From the perspective of historical literacy, the narratives produced by historical culture, family histories and pupils´ own experiences pose a challenge if their role is not defined and explained to pupils. Yet the exclusion of historical culture and students´ experiences from history classrooms is considered equally disadvantageous by many researchers (e.g., Hughes, 2021a; van Niuwenhuyse, 2016; Nordgren, 2021). Bernstein´s (2000) concepts of horizontal and vertical discourses are helpful in understanding the role of different types of texts in the classroom.

Vertical discourse, which is concerned with disciplinary knowledge includes texts used (primary sources) and produced by the discipline (secondary sources). Furthermore, tertiary sources such as textbooks can also be included in vertical discourse because they lean heavily on historical research (Ahonen, 2017). Horizontal discourse, on the other hand, is based on common sense and everyday knowledge and is therefore represented by fictional and experiential texts, as mentioned above. Historically literate instruction requires recognizing and engaging with both vertical and horizontal discourses and the texts produced by them. The interaction between the discourses does not, however, mean that the texts in different discourses have an equal standing or a similar role in understanding history (Nordgren, 2021). Ideally, teachers´ pedagogical recontextualization would bring horizontal discourse into history classrooms. Vertical discourse would then provide the disciplinary tools through which the initial conception of the horizontal discourse might be transformed. Thus, pupils´ initial take on historical narratives suggested by films and games, for instance, would be viewed through disciplinary lenses. Excluding horizontal texts from classrooms would leave pupils without any tools to process the texts they face in their daily lives. However, the recontextualization of horizontal and vertical discourses is challenging because teachers have been shown to struggle with applying the same level of disciplinary scrutiny to horizontal discourse as they do to vertical discourse (Hughes, 2021; Johansson, 2019).

The variety of different text types and the role they play in historical literacy is depicted in Figure 3. Texts functioning as primary sources comprise the core layer because of their vital role in knowledge construction, the understanding of which is the basis of historical literacy. The knowledge needed in the core layer is a prerequisite for evaluating texts in the subsequent layers as well: in order to evaluate the plausibility, soundness and credibility of the interpretations presented in secondary and tertiary sources as well as fictional texts, access to and

knowledge about primary sources is necessary (Nokes, 2012, p. 21). While the layer with secondary sources is closely attached to the core, the layer with tertiary sources may or may not utilize primary sources (i.e., in the form of Wikipedia, see Nokes, 2012). The outermost layer with texts deriving from historical culture and pupils´ own experiences surrounds the entire figure.

Teaching historical literacy involves all of the aforementioned layers with versatile text types and, as mentioned, interaction between horizontal and vertical discourses. However, primary sources have special significance for historical literacy as the epistemological foundation for historical knowledge is closely connected to primary sources (Nokes, 2012; Rosenlund, 2015). Remaining at the outer levels only would, from a disciplinary point of view, skip a fundamental epistemological step. As noted by Lévesque (2008, p. 26), making sense of or critically employing historical interpretations calls for an understanding of the way they were created in the first place. In other words, if students do not engage with the process of producing historical knowledge, which means using primary sources as evidence in the core layer, they may find it hard to locate the basis on which to evaluate existing narratives in the outer layers. On the other hand, historical literacy is not achievable by looking at primary sources only (see Reisman & McGrew, 2018).

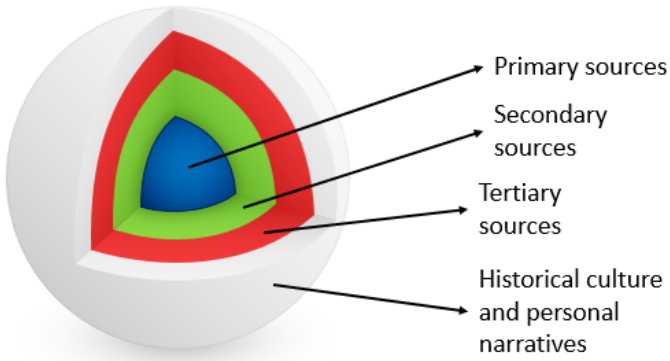

Figure 3. Different textual layers involved in teaching historical literacy.

The challenges of narrative genres in both factual and fictional texts are a reason for their inclusion, not for their exclusion. Secondary sources and the argumentative genres they convey are an important way to show the characteristics of history. Narratives in textbooks offer pupils, especially at primary level, much-needed textual support as they are not yet able to read through

a vast number of texts. Hence, the aim at primary level is not to avoid textbooks but to understand their genre and convey the characteristics and limitations of that genre to pupils.

4.2 Teaching practices

As described in the previous chapter, historical literacy requires different forms of knowledge and especially the interplay between them. In this vein, I will now explore the relevant teaching practices in relation to the forms of knowledge they aim to promote. This distinction between procedural and conceptual knowledge is made for the sake of clarity, while acknowledging that multiple aims and forms of knowledge exist simultaneously in teaching practices.

4.2.1 Towards procedural knowledge

Some researchers have expressed doubts about emphasizing procedural knowledge in classrooms (Bain, 2000, 2017). In Bain`s (2000) view, it is ill-advised to teach only procedures as this may lead to a mechanical approach. Students should also understand the thinking behind the procedures, otherwise they run the risk of developing a naïve epistemological stance, believing that applying certain procedures to texts would result in obtaining absolute knowledge about the past (Wilke et al., 2022). Also typical of a naïve epistemology is the belief that knowledge is replicable from one context to another (Grecic, 2022). In this sense, it can also be described as a copier stance, which is considered disadvantageous for disciplinary history (e.g., Rosenlund, 2016; Wilke et al., 2022). Another critical perspective raised by Ahonen (2020a, p. 137) is that in the pursuit of historical literacy, "history lessons may be reduced to intellectual handicrafts instead of preparing citizens to justify their future choices".

In order to avoid a mechanical approach to literacy, in his intervention study with fifth-graders, VanSledright (2002) connected literacy strategies to a wider epistemological discussion about the origin and nature of historical knowledge. In order to help the pupils, VanSledright introduced history-specific reading strategies and incorporated them into classroom routines. He employed a chart, where definitions as well as the interpretational issues related to primary and secondary sources were outlined. The chart shows the commonalities in the interpretational processes of primary and secondary sources. For example, limitations applying to both to primary and secondary sources include: "author may not remember all the details, author may see/hear/understand things differently (because of his or her point of view) and author´s ideas may be influenced by someone else who was there" (2002, pp. 43–44). In addition to commonalities, the chart also advises pupils on the differences between primary

and secondary sources. Another chart titled "Questions historical detectives ask to solve the mysteries of the past" listed the steps to be taken and questions to be asked during the inquiry. Pupils were asked to copy this list into their notebooks so that it could be used in future inquiries as well. The definitions and investigative steps were short and simplified.

Similar teaching practices related to explicit literacy strategies can also be found in other studies. The expert teacher in Hughes´s (2021) study also created routines for her third-grade pupils and taught them, for example, to reply in unison to the question of what one should do when encountering a source. The teacher also modelled many of the essential steps involved in reading sources. The teaching practices served an inquiry project in which pupils had to answer the inquiry question "How do people overcome hardship?" in the context of African American history. Explicit strategies were not limited to reading but extended to writing evidence-based claims. In one instance, the teacher modelled the initial steps involved in building a claim that would rely on the given evidence, consisting of 12 visual texts. The teacher underscored that it would not suffice to build claims based on personal opinions. The modelling was interactive in the sense that the pupils expressed their ideas on how and why to proceed with their historical inquiry.

In addition to providing explicit reading strategies, Hughes (2021) identified a teaching practice resonating with Bain´s (2000) concern about a mechanical approach. A significant amount of time was spent on understanding the nature of historical knowledge. This ensured that pupils would not merely execute the given literacy instructions without simultaneously developing an epistemological understanding. Whole class discussions played a central role in these epistemological reflections, as in VanSledright´s (2002) classroom. In both instances, teachers asked questions to lead the discussion, which revealed pupils´ epistemological stances and possible misconceptions.

A third expert/researcher study on successful literacy practices in the primary context is an intervention study conducted by Nokes (2014). He used a document-based lesson structure (see Reisman, 2012a), which is essentially one form of historical inquiry. Nokes taught two fifth-grade classes one history lesson per week over the course of a year. The structure of his lessons largely followed a pattern whereby he gave a 5- to 15-minute lecture, culminating in a historical question or controversy to be addressed using the evidence provided. Although much of the work was done in small groups, Nokes noted that a more in-depth understanding of the sources was gained if read through and discussed at the beginning of the lesson with the whole class. On some occasions, pupils used a graphic organizer which was designed to point towards relevant questions about the sources.

Taken together, the results of VanSledright´s (2002), Hughes´s (2021) and Nokes´s (2014) studies show that providing pupils with procedural knowledge in

the form of explicit strategies and epistemological questions translates into improved historical literacy (see also Wissinger et al., 2018). Although Hughes did not present data on pupils´ performance, he infers that the teacher´s practices provided the pupils with the opportunity to build historical literacy. Both VanSledright and Nokes assessed pupils´ knowledge and proficiency in historical literacy or related issues such as their epistemological stances. VanSledright reported that there was an improvement in pupils´ intertextual reading, their ability to understand the subtext, and to identify the author´s perspective. Using Wineburg´s (1991) terminology, sourcing and corroboration were enhanced during the intervention and pupils resorted less to general literacy strategies. Among the changes recorded by Nokes (2014) were those affecting the role of pupils, texts, ideas, teachers and peers. For instance, before the intervention, pupils described their role to include listening to lectures, and recalling and reporting information. After the intervention, pupils viewed their roles to include weighing evidence, using background knowledge, and developing their own interpretations. In addition, there was a shift from an objectivist stance to a criterialist stance, that is, from viewing history as an objective account of the past to recognizing that multiple interpretations exist and that evidence should be employed to evaluate those interpretations.

Despite these results, the effectiveness of teaching procedural knowledge for interpreting texts has been questioned (Stolare, 2017). The study that Stolare refers to is one conducted by van Boxtel and van Drie (2012) on contextualization, defined by them as situating "phenomena and acts of people in the context of time, historical location, long-term developments […]" (p. 114). The study shows that students who received "knowledge training" (50 minutes of substantive knowledge coaching) performed better than those who had "strategic support" (four directive questions). The relevance of these results for historical literacy seems low for three reasons. First, contextualization is only one of the processes involved in historical literacy (see Wineburg, 1991). Thus, the study may not warrant drawing conclusions on the ineffectiveness of procedural knowledge on the interpretational process as a whole. Second, the "strategic support" was not taught at all, only introduced in written form in the post-intervention test. Third, the "strategic support" mainly included general strategic steps, not historical literacy-related strategies.

The available data on the introduction of procedural knowledge in classrooms, not taught by expert teachers or researchers, concerns the secondary level. It seems that explicit instruction for the purpose of procedural knowledge is rare, as was the case in Iceland, where only 7% of the observed lessons included explicit strategy instruction (Gestsdóttir et al., 2019). Nokes (2010) reported that no procedural knowledge for interpreting texts in a disciplinary way was provided in the eight secondary classrooms he observed. In contrast, in Belgium, primary sources were used more frequently to foster strategic knowledge (24%) than for

illustrative purposes, although for the majority of time (55%), primary sources were used for fostering substantive knowledge (van Nieuwenhuyse et al., 2017).

Owing to the recontextualization process (Bernstein, 1990), the procedures and strategies taught in schools are transformed into something other than those that exist in the parent discipline. Hence, procedural knowledge at the primary level differs from the knowledge used by historians. For example, the strategies introduced by VanSledright (2002) for interpreting different types of sources have been recontextualized for pupils´ needs. Similarly, the teacher in Hughes´s (2021a) study had simplified the procedures and epistemic questions for third-graders. Although pupils are guided to mimic some of the procedures used by historians, such as defining the source type and looking into possible biases, pupils rarely encounter all of the epistemic questions faced by historians. Yet familiarity with the existence of the procedures used by historians is vital for understanding the epistemic foundation of the discipline. Just as procedural knowledge forms the basis for the initial years of schooling in general literacy (e.g., de-coding, summarizing, making mind-maps, choosing key words) and mathematics (e.g., bridging through ten, subtraction), it may do the same in history lessons.

4.2.2 Towards conceptual knowledge

My critical reflections on van Boxtel and van Drie´s (2012) results in relation to historical literacy should not be taken to undermine the essential role of substantive knowledge in developing historical literacy. On the contrary, as mentioned in Chapter 3, historical literacy requires knowledge about the historical context. Substantive knowledge informs us how to read texts and, conversely, the texts we read contribute to our knowledge of the context. This cycle is an inseparable part of historical literacy (Rantala et al., 2020). Hence, Wineburg and Wilson´s (1991) suggestion that lack of substantive knowledge can be offset by a strong understanding of disciplinary knowledge seems unwarranted.

Particularly at the primary level, where pupils´ conceptual understanding is still an emergent one, first-order conceptual knowledge cannot merely be "fed in" but should be taught. When studying 12-year-old pupils´ understanding of history, Pilli (1988) found that pupils struggled with many of the central first-order concepts found in textbooks. Concepts that had both concrete and abstract meanings in Finnish were particularly confusing for pupils. Many of these concepts could be described as having a different meaning in vertical and horizontal discourses. For example, the word *kauppa* in Finnish refers to trade as an abstract transaction, as well as to a shop in more concrete terms. The former is used in vertical discourse in a disciplinary context, and the latter for the most part in horizontal discourse in pupils´ everyday lives. Thus, it is understandable that pupils tend to go for the meaning provided by horizontal discourse.

According to Pilli (1988), these conceptual challenges should not be taken as confirmation that primary history should not introduce abstract concepts. Instead, Pilli calls for textbook authors to clarify the meaning of those central concepts that are used repeatedly and are only assumed to be familiar to pupils. The challenges at the conceptual level also place demands on teaching practices. How can teachers ensure that the substantive knowledge learnt contributes towards historical literacy and not towards merely memorizing substantive knowledge for the purpose of adopting a given historical narrative?

One possibility is to draw on Reisman´s (2012a) document-based lesson structure for primary-level needs. In each lesson, substantive knowledge relevant to the inquiry at hand is provided through a short lecture or by reading, usually lasting 5–15 minutes and utilizing tertiary sources. Taking Pilli´s (1988) findings into account, even short lectures at the primary level may benefit from ensuring at regular intervals that pupils comprehend the concepts used and content covered. Following the structure of the document-based lesson, the acquired knowledge is then put to use when interpreting a set of sources. Reisman´s (2012b) study shows that secondary students master substantive knowledge better if it is built as part of an interpretational process, as opposed to when the aim is to acquire substantive knowledge per se. There is no reason to assume that primary pupils´ conceptual understanding would not benefit from a similar interpretational approach. Hence, even if the acquisition of substantive knowledge were the primary goal of history education, connecting substantive knowledge with the process of investigating and interpreting history seems to be effective.

The medium through which substantive knowledge is acquired is also relevant. Studies on the comprehension of texts show that reading texts with a visible author, whose point of view and authorial presence is known, elicit active engagement with the content because students seem to ask questions from the texts. Expository texts, such as those in textbooks with an anonymous author, raise fewer comparative questions and arouse less interest in the topic (Logtenberg et al., 2011.) This tendency to engage actively with secondary sources has been shown even at the primary level (Wolfe & Goldman, 2005). In his intervention with fifth-graders, VanSledright (2002) opted to use a secondary source for getting the pupils acquainted with relevant substantive knowledge for an inquiry project. He chose a chapter from a book and conducted a “guided reading” (p. 83), which included an introduction explaining why the chapter needed to be read. He then used lessons on two days for proceeding with the chapter in steps, duly adopting a different approach to substantive knowledge compared to Reisman´s (2012a) document-based lesson.

Whatever the chosen approach, understanding the concepts and making meaning with substantive knowledge requires time. Therefore, attempting to cover all the substantive knowledge provided in textbooks or other equivalent

materials runs the risk of treating the topics merely on a superficial level. That, in turn, does not contribute towards becoming historically literate.

The tendency to cover as many chapters or topics as possible may not arise from the needs of history altogether. As discussed in earlier chapters, the multiple literacy aims (the aims of content-area literacy and multiliteracy) may influence the decisions that primary teachers make in history lessons. If the aim is to strengthen pupils´ general literacy, going through substantive knowledge in history lessons can be a platform for practising reading comprehension and other general literacy aims.

Apart from substantive knowledge, another type of conceptual knowledge also plays a part in historical literacy. Second-order concepts and knowledge are particularly useful when pupils construct and write their own accounts of history. Ideally, when constructing their own interpretations, primary pupils use second-order concepts to build claims about changes, continuities, causes and consequences (Nokes, 2014). Thus, second-order concepts help to structure the thinking that is produced through reading multiple sources and by the possible use of texts as evidence. Since primary textbooks are built on accounts and recounts, which seldom contain second-order concepts (Coffin, 1997), their introduction to primary classrooms therefore relies on the teacher's initiative.

4.3 Assessment

Assessment is often viewed in terms of its various aims. First, assessment can aim at improving teaching and learning (Webb, 1992). This type of assessment is seen as formative "when the information gathered in the assessment is actually used to adapt the teaching to meet student needs" (Black and Wiliam, 1998b, p. 140). Second, assessment can be used for student accountability. The focus of this kind of summative assessment is on what has already been learnt, not on what should be learnt in the future, and how it should be learnt (Boud, 2000; Webb, 1992). Unlike the first two aims, the third purpose of institutional accountability (Webb, 1992) has little relevance in the Finnish context for basic education, where schools are not publicly rated based on student learning outcomes.

In the Finnish context, classroom assessment refers to both summative and formative assessment. Thus, it differs from educational contexts where, in contrast to large-scale assessment, classroom assessment is equated with formative assessment (e.g., Black & Wiliam, 1998a). Although the distinction between formative and summative assessment is commonly acknowledged (Bennet, 2011; Boud & Falchikov, 2006; Taras, 2005), the distinction may not be self-evident in a classroom setting. Any assessment, including assessment conducted at the end of a teaching period, can be considered formative if it informs future teaching. In the Finnish educational setting, where teachers are solely responsible for assessment, the information gathered on pupils´ learning through assessment can

affect teachers´ future decisions. Therefore, when introducing some of the relevant literature on assessment materials, I do not make a distinction between tasks used for summative and formative assessment.

The significance that assessment has in both teaching and learning processes is described as the backwash effect (e.g., Brown, 2011; Vesalainen & Huhtala, 2018). The backwash effect refers to a phenomenon whereby what is being assessed defines what is being learnt and taught. Therefore, it has been suggested that changes in teaching and learning require change in assessment (Brown et al., 1997; Watkins et al., 2005). What teachers choose to assess also conveys to learners what is valued and considered important (Vesalainen & Huhtala, 2018). Thus, historical literacy cannot be successfully taught by conducting assessment that is incompatible with its aims, even if all other teaching practices are. As assessment has a unique power to either undermine or support a given aim, it has to be aligned with other aspects of teaching.

4.3.1 Issues of validity

Wineburg´s (2018) description of a century of history assessment in the United States has tragicomic features. The tragedy lies in the fact that large-scale testing was originally designed to rank students, not to assess their actual knowledge. Yet Wineburg shows that the results have, for one hundred years, been interpreted in such a way as to reveal students´ poor understanding of history. Striving for a symmetrical bell curve has ensured that "winners create losers" (p. 17). As a consequence, even if teachers were to succeed in raising an exceptionally bright and capable cohort of students, the bell curve would not allow for the majority of students to score at the top. According to Wineburg (2018), even the so-called standard-based national tests used today in the States are still embedded in the ideal of creating spread among students and conforming to a normal curve. Hence, "no national test can allow students to show themselves historically literate" (p. 20). The comical feature is that each generation repeats the same fallacy of interpreting the next generation´s test scores as exemplifying ignorance and a decline in knowledge.

In history education, there are two related but to a certain extent separate discussions on assessment. On the one hand, the discussion on high-stakes and other types of large-scale assessment addresses, or at least should address, questions on validity but cannot avoid issues related to costs, efficiency, and feasibility. Apart from the States, this kind of discussion is also taking place in Sweden, for example, where national tests exist as well (e.g., VanSledright, 2014; Eliasson et al., 2015; Wineburg, 2018). On the other hand, assessment is taking place in classrooms where individual teachers are wholly responsible for it, from planning and implementation to scoring and grading. Since Finland has no large-

scale testing during the first nine years of schooling, I focus in the following on the assessment of historical literacy in classrooms only.

As issues of scalability and cost do not have to be considered in classroom assessment, the main concern is validity. To be more precise, assessment theorists such as Messick (1980) and Cronbach (1971) have underscored that it is not a test or a task in itself which is valid or invalid. Instead, "validity is the overall degree of justification for test interpretation and use" (Messick, 1980, p. 1014). In his later work, Messick (1989, p. 13) also included "other modes of assessment" alongside tests in his validity definition. A fitting example of invalid inferences on test scores is the aforementioned case of one century of history assessment tradition in the US (see Wineburg, 2018), where test scores have been falsely used to make claims about young people's ignorance. Hence, when assessing historical literacy, the main question is whether a task, project or any form of work used for assessment purposes produces valid inferences about pupils´ ability to engage with texts in a historically literate way.

In order to understand the challenges of assessing historical literacy, I look into some central aspects related to validity. Theorization on validity has been dealt with more extensively, for example in Ouakrim-Soivio´s (2013) work, where modern validity theory, its developments and different aspects of validity are discussed in depth. Here, I focus mainly on construct validity, which is seen as the basis for unified validity (e.g., Messick, 1980; Kane, 2001; Ouakrim-Soivio, 2013) and described as "interpretive meaningfulness" (Messick, 1980, p. 1015). Construct validity demands that a construct such as historical literacy has been conceptualized and linked to a theoretical background in a satisfactory manner (Kane, 2006; Rantanen, 2003). Construct validity goes beyond the consideration of content (or content relevance, see Messick, 1980), which in the case of history would entail ensuring that the era, events and topics under assessment are relevant.

Interpretative meaningfulness arises from the connection with the theoretical background of historical literacy. This means that assessment should provide pupils with the opportunity to demonstrate their proficiency in interpreting a variety of texts, building historical knowledge and justifying their claims. This in turn requires interaction with different forms of knowledge. If a task claiming to assess historical literacy focused only on substantive knowledge, the construct validity would be compromised. Similarly, assessment that assumes only general literacy cannot provide valid interpretations of pupils´ abilities. Assessment that is valid in terms of historical thinking is not necessarily valid for historical literacy, although the two concepts are interrelated. For example, if pupils were asked to reflect on the causes and consequences of a conflict without using any textual material, it may prove to be valid for historical thinking but not for historical literacy.

When looking at construct validity, it is also important to consider the way in which different elements of the same construct are dealt with. Reisman (2015b)

suggests that when teachers assess students' historical reading, they should assess 1) background knowledge, 2) the ability to evaluate the reliability and relevance of textual evidence, and 3) the ability to produce their own accounts. Reisman (2015b) continues that "these three assessments need not be administered together" (p. 37). However, if assessed separately, the interaction between these elements is not assessed, which may compromise the construct validity.

One way to produce evidence for construct validity is to look at cognitive validity, in other words the cognitive processes which an assessment task or item elicits (Breakstone, 2014; Ercikan et al., 2015; Kaliski, et al., 2015; Smith, 2018). Smith (2018) investigated whether a Historical Thinking Test (HTT) created by Reisman (2012a) elicited the intended cognitive historical thinking processes. Smith (2018) as well as Ercikan and colleagues (2015) used think-aloud protocols to uncover the thinking that took place while completing the tasks. For the most part, both studies were able to provide a solid argument for validity through think-aloud protocols.

4.3.2 Means of assessment

The very definition of historical literacy stipulates that any material attempting to assess historical literacy should involve interaction with texts. The questions and assignments that this interaction should consist of are still a matter of debate. Using multiple-choice questions as the basis for assessment has been widely criticized by history educators (e.g., Breakstone, 2014; Reich, 2015; Reisman, 2015; VanSledright, 2014). It has been pointed out that multiple-choice questions are unlikely to reveal students´ reasoning or induce complex thinking, and that construct validity is sacrificed for test reliability (VanSledright, 2014, 2015). Traditionally, multiple-choice questions have at least been viewed as a valid means to test students´ factual, content-related knowledge. However, even this has been questioned by suggesting that answering correctly is not entirely dependent on factual knowledge. Reich´s (2009) think-aloud interviews with high-school students show that almost a fifth of the responses to multiple-choice questions were correct despite the fact that the participants did not know the key information. The participants often arrived at the correct answer through their general literacy skills or by using what Reich refers to as test-wiseness, by utilizing the information given in other items, for example.

On the other hand, there have been attempts to develop multiple-choice questions which are able to capture disciplinary aspects of history. Seixas and colleagues (2015) included some multiple-choice items in their one-hour historical thinking assessment, which also provided a follow-up opportunity where students´ thinking would be visible. Smith (2018) tested whether multiple-choice questions intended to elicit sourcing, corroboration and contextualization

were able to do so. The results showed that the six items generally elicited the intended cognitive processes, although construct-irrelevant reasoning was also found. When students produced false positive answers, it was due to 1) guessing, 2) irrelevant item feature (e.g., an arbitrary date), 3) factual recall, and 4) decontextualized reasoning (e.g., viewing people´s thinking in the past as similar to today). In spite of the limitations observed, Smith (2018) tentatively regards the results as encouraging for using the Historical Thinking Test's multiple-choice questions as a means of assessing history in a disciplinary way. Another approach to multiple-choice questions was introduced by VanSledright (2014) in the form of weighted options. Instead of three incorrect and one correct option, weighted multiple choice questions provide several correct options but to a varying degree. The efficacy of this approach is tested in the present study (Article I).

Whether or not multiple-choice items are able to elicit cognitive processes relevant to historical literacy, they pose a challenge for assessment. In a classroom setting, teachers seldom conduct think-aloud interviews and therefore the only traces of pupils´ thinking in multiple-choice questions are the encircled options. Constructed-response items generate more information on pupils´ thinking, albeit introducing a new set of challenges for validity. On the other hand, the scoring of multiple-choice items is fast and does not require the teacher to interpret pupils´ answers, as in constructed responses.

Ideally, constructed responses to open-ended questions should provide pupils with a means of expressing their thinking and a channel through which they can produce evidence-based interpretations. However, written answers also presuppose general literacy and, more specifically, writing. In order to make valid interpretations about pupils´ historical literacy through assessment, teachers need to identify the extent to which general literacy contributes to a successful answer. As noted earlier, writing in general involves cognitive processes such as planning, translating (i.e., the production of text), as well as reviewing and revising (Hayes & Flower, 1980; Kellogg, 1999). These processes are present in most writing irrespective of the subject, and thus are part of content-area literacy. By using content-area literacy strategies, it is possible to structure a coherent answer irrespective of the subject or topic. Historical writing (Nokes & De La Paz, 2018), however, is considered to focus in particular on argumentative writing, in which data and claims are used to build a sound argument. Further, a historical argument has to take into account the nature of historical evidence and the contexts, duly differentiating a historical argument from other arguments.

Any validity argument of an open-ended item should take into account the kind of writing that is assumed. If a pupil can provide an exemplary answer to an open-ended question solely by proficiently executing the planning, translation and revision phases (Hayes & Flower, 1980), the item generates little information about pupils´ historical literacy. Consequently, open-ended questions should necessitate building a historical argument, which in the primary context can be a

short one. Short answers to open-ended questions have been developed by Breakstone (2014) in the form of History Assessments of Thinking materials. While these assessment material designs, like most others, are targeted at the secondary level, they can also be effectively used at the primary level because they utilize visual or short written texts.

The assessment materials available for Finnish primary teachers mainly derive from teacher's guides published as part of textbook series. Lyhty´s (2020) analysis of these assessment materials reveals that hardly any of the materials in the guides link the items of assessment tasks to the disciplinary aims stated in the national curriculum, making it difficult for teachers to recognize what is being assessed. Moreover, while some items were suitable for assessing disciplinary aims, the vast majority of tasks and items were concerned with assessing the recall of factual information. Thus, although the Finnish context of classroom assessment would allow for versatile and disciplinary assessment materials, they are not easily available for primary teachers. It follows that if teachers wanted to use assessment materials aligned with curricular aims, they would have to produce these materials themselves. Given the number of subjects that primary teachers have in their schedule, creating new assessment materials may seem a less viable option than using ready-made ones.

5 Methodological approaches

I have employed two research approaches in this study as I believe that this methodological versatility has the potential to unravel different aspects related to teaching historical literacy. The first is a design-based approach, in which I have sought to develop the existing practices (Article I), whereas through the second, namely the observational multiple case study, I have aimed at understanding teachers´ actions and thinking (Articles II and III). While the former is research for education, the latter is research about education (Juuti & Lavonen, 2006).

The direction of the research process is generally thought to proceed from larger entities to smaller ones: methodologies define procedural rules, which in turn result in methods and ultimately in the production of knowledge (Brewer, 2000). However, research processes are often messy and may not proceed linearly. Rather than starting from fundamental methodological assumptions, smaller entities, such as the choice of methods, may serve as the initial spark for research (Brewer, 2000). In my research, I was motivated by a desire to conduct think-aloud interviews and observations. However, my interest in specific data collection methods does not mean that the theoretical dependence of these methods on the broader methodological issues would be dismissed. Thus, I reflect below on both the ontological and epistemological assumptions underlying this study.

In this research, an assumption of an observable reality is made, reflected by my interest in observing classroom instruction. A realist ontology holds that there is a reality that is independent of people´s perceptions of it (Brewer, 2000). As a result, this independent nature of reality sets a limit on what we can know (Maton & Moore, 2009). However, unlike naïve realists, I do not claim that it would be possible to represent an objective reality accurately. A more moderate realist stance takes into account the possibility that there are restrictions concerning the representation of a reality. Thus, while the objects of the study are dependent on social constructs such as language and culture, moderate realist stances emphasize that reality is not constructed by the researcher´s beliefs and theories (Raatikainen, 2004). In short, I acknowledge that while observing an observer-independent reality, a representation of that reality is always presented through lenses and thus is liable to become distorted in some way.

In the theoretical framework, I have discussed the epistemological foundations of knowledge in relation to history and, in doing so, have drawn on social realism. The same assumptions apply in relation to knowledge throughout this study. As a school of thought, social realism has sought to move away from the assumption that positivist absolutism and constructivist relativism are the only choices available. Knowledge in this dichotomy is conceptualized either as

"decontextualized, value-free, detached and 'objective'" or as only socially constructed and concerned with questions of interest and power (Maton & Moore, 2009, pp. 1–2). The critique towards postmodern relativism has addressed the attempt to reduce knowledge to knowers and their experiences and to equate it with knowledge (Moore & Muller, 1999; Young, 2008). On the other hand, it is equally misleading to deny the social nature of knowledge, which in Young´s (2008) words leads to "a distinctly impoverished view of the sociology of education". Hence, rather than promoting an either/or position, social realism promotes a "both/and" approach. Epistemologically, this means that "rational objectivity in knowledge is acknowledged as itself a fact (we do actually have knowledge), but it is also recognized as a social phenomenon (it is something that people do in a socio-historical context) and it is fallible rather than absolute or merely relative" (Maton & Moore, 2009, p. 2).

5.1 Creating new: design-based research

Logically, research about history teaching would precede research aimed at improving history teaching: knowledge about the former would inform the latter. In this dissertation, however, the first of three articles deals with designing an artefact that would have the potential to improve existing practices. As described earlier, the existing assessment materials in Finland have largely focused on recalling substantive knowledge (Lyhty, 2020) and therefore would not help teachers in assessing the historical literacy-related construct objectives stated in the National Core Curriculum (2014). This incompatibility between curricular expectations for assessing pupils´ historical literacy and available assessment materials led me to embark upon a design process aimed at developing an assessment tool for historical literacy at the primary level. Below, I use the terms test, tool and task interchangeably when describing the artefact, while acknowledging that these terms may carry different meanings. However, with all these terms, I refer only to classroom assessment, not to large-scale testing.

My design process has been significantly influenced by the research conducted in North America on developing assessment materials. Canadians have built on their Historical Thinking Project (see Peck & Seixas, 2008) to design a one-hour test targeting a selected number of historical thinking concepts (e.g., Seixas et al., 2015; Ercikan et al., 2015). In the United States, the Stanford Group in particular has been active in designing (Reisman, 2012) and testing the validity of assessment materials (Smith, 2018). Breakstone (2014) describes his design process in detail, from selecting historical sources to re-designing the assessment tasks after testing. As Breakstone (ibid.) notes, research on designing new assessment materials is scarce, given the widely acknowledged problems with existing assessment practices in history classrooms (e.g., Wineburg, 2018). To date, designing and testing assessment materials for history teaching at the

primary level has been done only rarely, and there are no such materials in the Finnish context.

Design-based research (DBR) is often described as a change- and future-oriented approach (Barab, 2014; Campanella & Penuel, 2021). The relevance in DBR lies in what education could or even should be (Campanella & Penuel, 2021). In this respect, DBR is prescriptive rather than descriptive as it suggests new – and presumably better – practices. The pragmatic nature of DBR (Barab & Squire, 2004; Juuti & Lavonen, 2006) can be perceived in two ways. On the one hand, the needs as well as the opportunities and constraints of practice constitute the foundation of DBR (The DBR Collective, 2003). On the other hand, the design process should result in usability of the design product and thus contribute to the learning and teaching praxis (Juuti & Lavonen, 2006). This design product is commonly referred to as an artefact.

In addition to being pragmatic and creating usable products, DBR is also expected to contribute to theoretical understanding of the phenomenon at a general level. In other words, DBR has a "dual commitment" (Sandoval, 2014, p. 20), whereby it simultaneously aims to increase "fundamental understanding" and serve "considerations of use" (Stokes, 1997, p. 73). The need to include practice in knowledge building is also advocated by social realists. For example, Young (2008) dismisses the assumption that "theory" should be developed independently of the needs of practice.

One of the most essential characteristics of DBR is its iterative nature (Juuti & Lavonen, 2006). The design, testing, evaluation and reflection phases are repeated over several cycles. The initial design of an artefact should be based on theoretical assumptions about how learning happens or how learning is supported, rather than on the designer´s personal experiences and beliefs (Juuti & Lavonen, 2006; Sandoval, 2014). In this study, the artefact was initially constructed based on the literature concerning primary pupils´ ability to think historically and to engage with texts in a historically literate way.

During the testing phase, design decisions based on theoretical assumptions are exposed to the "messiness of real-world practice" (Barab & Squire, 2004, p. 3), which in the educational context is often the classroom. The requirement to carry out work in a real-life setting is also considered one of the strengths of DBR. If and when DBR recognizes and integrates real-world practice with theoretical claims, they may have real explanatory value (Barab, 2014). My encounter with the "messiness" of the "real world" has perhaps been more limited than in many other DBR studies due to the nature of my artefact. A tool for summative assessment, which could also be characterized as a test or a task, is normally handed out to pupils and collected after completion. Compared with testing curricular interventions or new learning environments, for example, there are fewer opportunities for the complexity of the classroom reality to interact with the artefact. Instead, in this study, the real-world context entailed testing the artefact

for the type of thinking it elicits in pupils, both through pen-and-paper tests and think-aloud protocols.

Teachers were not included in the design process, which makes this a somewhat atypical DBR study. Typically, the design involves three elements: a) a designer, b) a practitioner, and c) an artefact (Juuti & Lavonen, 2006). In my research, the participants involved in the study were not practitioners (e.g., teachers) but pupils for whom the artefact was designed. The object of study in DBR is the artefact, not those participating in the study. As a consequence, participants are perceived as collaborators or co-participants who contribute to the development of the artefact (Barab & Squire, 2014). The role of the participants can extend to decision-making in the design process (see Severance et al., 2016) and even to the analysis (Barab & Squire, 2004). Even if participants are not included in the actual design and analysis phases, they are acknowledged as contributors. Particularly with regard to children and research, this approach may respond to the call for giving participants stronger agency in the research process and thereby working towards more democratic research practice (e.g., Waldron, 2006).

A key method used in my design process is think-aloud interviews, also referred to as verbal report protocols (VanSledright et al., 2006) and think(ing)-aloud protocols (e.g., Ericsson & Simon, 1993; Smith et al., 2018). Think-aloud protocols (hereafter TAPs) are part of a wider spectrum of verbalizing procedures (retrospective and introspective reports) developed in the field of psychology (Ericsson & Simon, 1993). Whereas retrospective verbal reports may be recorded hours or days after the completion of a task, TAPs are concurrent reports that seek to gain insights into the actual thinking process of participants while engaging with a text or an activity (Ericsson & Simon, 1993).

Thinking aloud requires the participant to attempt to verbalize their thinking while attending to information (Ericsson & Simon, 1993). Usually the focus is on the content of the participant's thoughts rather than on explaining their reasoning process (see Martin & Wineburg, 2008). On the other hand, TAPs can elicit descriptions or explication of thought content as well as explanations of thought processes (Ericsson & Simon, 1993). However, forming descriptions and explanations requires access to additional information and "as a result, the sequence of thoughts is changed, because the subjects must attend to information not normally needed to perform a task" (Ericsson & Simon, 1993, xiii). Ideally, the think-aloud method therefore aims to capture the thinking process in as authentic a form as possible, which means that TAPs would also include such rawness of thinking as hesitation, discontinuity and illogicality of thought (Martin & Wineburg, 2008).

Some limitations duly need to be considered when conducting think-aloud interviews. First, people vary in their ability to verbalize their thoughts. Adults in general are more fluent than those still maturing cognitively when it comes to

turning ideas into verbal articulation (VanSledright et al., 2006), but some of the differences may also be attributed to personal traits and cultural factors. Second, although Ericsson and Simon (1993) conclude that there is a close relationship between thought and the corresponding vocalization, they do not view thought as internalized speech (Ericsson & Simon, 1993). It follows that no matter how strictly participants are instructed to only verbalize their thoughts as such, the verbal reports may never fully correspond with actual thinking. Third, even though thinking aloud does not increase the measurable time taken to complete the task, participants may feel that thinking aloud slows down and even interferes with their thinking process (Dansereau & Gregg, 1966).

In design-based research, TAPs have been used for identifying whether or not an artefact elicits the type of thinking that was intended by the researchers, also in the context of historical thinking (e.g., Shanahan & Shanahan, 2008; Wineburg, 1991) and its assessment (Ercikan et al., 2015; Kaliski, Smith & Huff, 2015; Smith 2018). As in the present study, TAPs have also been used for addressing the validity of assessment (see Ercikan et al., 2015; Kaliski, Smith & Huff, 2015; Reich, 2015).

5.1.1 An artefact for historical literacy

The artefact created in this study is an assessment task. While its testing and development phases are reported in the results, I describe its initial structure and content here in brief. The questions in the assessment task are designed to tap into four curricular aims (see Table 1), namely:

- to guide the pupil to recognize different sources of history (O2)
- to guide the pupil to notice that historical information can be interpreted in different ways (O3)
- to guide the pupil to understand the motives behind people´s actions (O5)
- to instruct the pupil to explain how interpretations may change as a consequence of the new sources or new ways of examining them (O10)

The task is intended to be completed within one lesson (see Seixas et al., 2015) and is therefore restricted to two A4 sheets of paper. The task is entitled "Copernicus – friend or foe of the Catholic Church?", a title that is meant to serve simultaneously as a research question. Three written sources are utilized, all shortened to less than 100 words, with the language modified in order to make them accessible to pupils. The first written source (source 3 in the artefact) is an excerpt from a preface written by Copernicus. It was published as part of his work presenting the heliocentric model. The tone of the preface is conciliatory and respectful towards the Catholic Church, which is also reflected in the dedication of the book to the Pope. The second written source is an excerpt from a secondary history textbook, published in 1982. Contrary to the preface, this source suggests that Copernicus and the Catholic Church had a mutually hostile relationship. The third source is an excerpt from a letter written by Cardinal Schönberg to

Copernicus in 1536. This letter is also published as part of Copernicus's book *On the Revolutions of the Heavenly Spheres* (1543). In the letter, the Cardinal shows his respect for and appreciation of Copernicus´s work. Thus, the task is built on a contradiction between the secondary and primary sources.

The tool begins with a set of questions designed to determine pupils' knowledge of source types as well as overall contextual knowledge, such as the extent of the power that the Catholic Church had over medieval society, as well as medieval conceptions of the universe. For this purpose, two visual sources are used, one drawn by the Portuguese cartographer and cosmographer Bartolomeu Velho in 1568. The other is the heliocentric model by Copernicus. The aim of the background questions is not to highlight differences among pupils, but to provide the teacher with information on whether pupils have the required knowledge, both procedural and substantive, to complete the test.

All written sources used in the task are excerpts, meaning that none of them are presented in their original form. While the authenticity of sources is thus compromised, Reisman (2012a) thinks that it is necessary to modify sources in such a way that they are understandable to pupils, both visually and cognitively. The modification process was similar to that of Reisman´s (2012a): focusing, simplification and presentation. I took only the most essential parts from each source and simplified the vocabulary, while trying to maintain a tone typical of the era. I shortened the preface from 2,086 words to 83, the letter from 303 words to 97, and chose a 77-word excerpt from the textbook. While Reisman (2012a), working with high school students, aimed at keeping the sources below 250 words, sources used in the present study were each less than 100 words. In addition to the three written documents, two drawings were used as sources, one drawn by Copernicus himself and one by the Portuguese cartographer and cosmographer Bartolomeu Velho in 1568. These drawings were used in the background questions, where pupils were asked to identify the heliocentric and geocentric models.

5.1.2 Procedures for testing the artefact

As noted, DBR is characterized by its iterative and cyclical nature, making it difficult to separate methods from results, as they become intertwined in the development process. Thus, the design and redesign procedures are described as part of the results although they are undeniably also method-related. Below, I describe the participants as well as some of the procedures involved in setting up and conducting the study.

Table 2. Study participants by data collection method.

Phase	Year	Place	Pen-and-paper participants	Think-aloud participants	Total
Phase 1	2015	Helsinki	18	2	20
Phase 2	2016	Helsinki	21	-	21
Phase 3	2017	Uppsala	21	5	26
Total			60	7	67

The artefact of the study, the assessment task, was tested three times between November 2015 and October 2017. As Table 2 shows, the first two phases were conducted in Finland and the last phase in Sweden, to which end, the artefact was translated into Swedish. All of the participants were primary pupils aged 11–12, with most of them participating by completing a pen-and-paper task. Seven pupils took part in think-aloud interviews. All pupils in both countries were studying in municipal schools of average size, also including those pupils with learning difficulties. None of the teachers helping with the study had seen the task in advance. The teachers told me that the participants were familiar with the historical context of the assessment task and had worked with historical sources before.

The selection of participants for the think-aloud interviews was left to the teachers, who knew their pupils better than I did. The pupils´ overall knowledge of history or their grades were not a defining factor in the selection process. As thinking aloud required the pupils to engage in an unfamiliar activity and to work with an unfamiliar researcher, I asked the teachers to recommend pupils for whom this setting would cause as little stress as possible. Despite taking these precautions, out of the six pupils originally participating in the interviews in 2017, one found the situation extremely stressful. I consequently terminated the interview, resulting in the final number of five think-aloud participants in phase three (see Table 2).

I started each think-aloud interview with an informal interaction and went on to describe the participant´s role as someone helping me to develop the artefact, not as someone to be tested or evaluated. As thinking aloud is not a common activity, we practised on some Donald Duck comics to begin with, which many Finnish children read in their free time. I asked the pupil to read the text in the comics and to tell me what they were thinking at regular intervals. During this preparatory phase, I noticed that many struggled to voice their thoughts fluently and there were long pauses. This led me to take a less strict stance on the

recommended TAP procedures, which normally avoid questions and interactions with participants during the interviews.

Typically, only neutral suggestions, such as “carry on” or “continue”, are recommended during think-aloud interviews (see Ericsson & Simon, 1993). Due to the difficulties experienced by many pupils in the warm-up exercises, I decided to take a more active role in the interviews if needed. Thus, in addition to encouraging participants to “carry on”, I also asked questions about the text if the participants paused for extended periods of time.

Those participating in the pen-and-paper test were given up to one hour to complete the task. All participants and their parents were informed beforehand that taking part in the study would not have an effect on their overall assessment in history. Thus, participants were aware that they did not have a personal stake in completing the task, either through TAPs or pen-and-paper tests. This may have influenced the level of participants´ answers in two ways. On the one hand, some participants may not have been as fully engaged as they would have been had the test influenced their assessment. On the other hand, some participants may have produced unusually high-level answers due to the lack of pressure to succeed. Each phase needed new participants because a session with one version of the artefact would leave a mental residue (VanSledright et al., 2006). In other words, a participant would not be able to forget what had been asked in the previous session and what they were thinking at the time.

I carried out a content analysis of the data generated by the open-ended questions in the pen-and-paper test. The analysis covered the scope, content and quality of participants´ answers. In other words, I looked into whether the questions prompted answers that would correspond with my design expectations in terms of issues addressed and concepts used. The aim was to judge the overall level of difficulty of each question based on the answers it generated, not to rate the performance of individual participants. However, I drew inferences about the level of sophistication concerning answers to the open-ended questions.

In the case of weighted multiple-choice questions, the analysis of pen-and-paper data focused on whether the participants had understood the logic of answering this new question type. For example, choosing multiple options would suggest the contrary. In addition, the distribution of chosen options was analysed. However, as the pen-and-paper test could not reveal the basis on which participants had chosen an option, only limited conclusions could be drawn from the data.

I transcribed both the Finnish and the Swedish TAPs. A Swedish-speaking researcher read through the transcriptions of the Swedish data and clarified some words. Although the TAPs were analysed for their content just like the pen-and-paper data, the analysis was more detailed as the transcribed interviews contained not only participants´ thinking, but also pauses and hesitations. In addition to the

questions, the TAPs included data about the way in which participants read historical sources and progressed through the task.

5.2 Approaching classroom reality through a case study

In articles II and III, I set out to investigate the standing that historical literacy has in Finnish primary education. This was conducted as part of a historical literacy project (HisLit: Engaging in disciplinary thinking: historical literacy practices in Finnish general upper secondary schools) in conjunction with another researcher, Mikko Puustinen, whose focus was on similar issues but at the secondary level. As described earlier, our interests included understanding teachers´ practices and choices in relation to texts, for example. Additionally, we focused on the nature of the knowledge that was being promoted in history lessons. Although it was evident that this work should be carried out in a setting where teaching takes place, namely in classrooms, it was less evident what kind of study design would best serve this purpose. We collaborated considerably when developing the study design. However, I was solely responsible for the decisions and implementation of the study to be carried out at the primary level. Hence, in what follows I refer to myself rather than to both of us, while acknowledging the role of my colleague in the process.

When investigating teachers´ practices and thinking, and especially the potential discrepancy between the two, case study is one of the most widely used approaches in history education (e.g., Hartzler-Miller, 2001; Howard et al., 2021; Hughes, 2021; Kitson, 2020; Monte-Sano, 2011; Nokes, 2010; Stolare, 2017). This may be because it allows researchers to explore the complexity of a chosen phenomenon in an in-depth manner (Simons, 2009) and in a real-world context (Yin, 2014). As Yin (2014) points out, case study definitions differ on whether they emphasize the nature of the object (i.e., individuals, organizations, programmes) or the procedures related to the inquiry (i.e., data collection techniques, data analysis). However, what most definitions have in common is a demand for particularization and acknowledgment of the circumstances (e.g., Simons, 2009; Stake, 1995). Stake´s (1995, p. xi) definition aptly describes the approach taken in the present study: “Case study is the study of the particularity and complexity of a single case, coming to understand its activity within important circumstances”.

Case study cannot be defined as a methodology as it can serve differing epistemological standpoints, such as realist and relativist positions (Yin, 2014, p. 17). Further, similarly to Simons (2009), I refer to case study as an approach rather than a method (cf. Yin, 2014, p.16) because case studies can employ a variety of methods, both qualitative and quantitative (Brewer, 2000; Simons, 2009).

As noted, the aim of case study is to gain an understanding of a specific case. A single case study may be designed to test an existing theory, capture common

or unusual cases, or follow a case longitudinally (Yin, 2014). Although these single case studies hold explanatory power on their own, I opted for a multiple case study (see e.g., Kitson, 2020; Nokes, 2010) as this would allow for taking into account the potential diversity that exists in teaching history at the primary level. Multiple case studies aim to "form a collective understanding of the issue or question" (Stake, 1995, pp. 3–4) and evidence from multiple cases has been described as more compelling (Herriot & Firestone, 1983), partly because cross-case comparisons can be made (Yin, 2014).

Although case studies neither seek nor are able to make generalizations about populations, the question of generalizability cannot be avoided when conducting a case study, especially since the inability to generalize is often suggested as a failing of this approach. Despite common misconceptions, generalization and case studies are not altogether mutually exclusive (Simons, 2009; Yin, 2014). When generalizability is conceived as usability of findings or as ways to transfer findings to other contexts, case studies – and especially multiple case studies – may contribute to generalization (Simons, 2009). The cross-case analysis conducted in this study aims at identifying both the differences and interconnecting themes present in the multiple cases. This provides the opportunity to make abstractions and general propositions, although the meaning is based on these specific cases (Simons, 2009) and is not extended to an entire population of teachers. In other words, I do not claim to describe the typical or average way in which Finnish primary teachers teach history. Instead, my aim has been to describe some of the characteristics and themes that exist in relation to teaching historical literacy at the primary level.

The primary data- gathering method used in this case study is classroom observation, although teacher interviews were conducted alongside observations. Therefore, an observational multiple case study (see Borg & Gall, 1989) is perhaps the most accurate description of my approach. When making decisions about the ways in which to conduct the observations, the nature and aims of a multiple case study guided the research design. A compromise was reached between an in-depth understanding of the phenomenon (depth) and gathering sufficient data for cross-case comparison (breadth).

The first issue regarding depth and breadth is the number of participants. Multiple case studies focusing on history teachers have typically included two to five participants (e.g., Kitson, 2020; Monte-Sano, 2013), but others such as Nokes (2010) have included eight teachers when investigating teaching historical literacy at the secondary level. With fixed resources (i.e., the number of researchers), an increase in the number of participants would reduce the time spent in one classroom (Herriot & Firestone, 1983). Hence, observing dozens of teachers would have entailed observing each teacher´s instruction only once or twice, possibly resulting in what Borg and Gall (1989, p. 402) refer to as a "shallow case study". Despite offering breadth and enhancing generalization, a limited number

of visits to each classroom would have incurred severe risks in relation to the aims of this study: to observe typical lessons conducted by each teacher. If each teacher´s instruction is observed for just one or two lessons (see Gestsdóttir et al., 2019), the risk of observing atypical lessons might increase as teachers may be tempted to prepare and design a lesson especially for the researchers´ benefit. It is impossible to define an exact number of observed lessons that would guarantee a researcher access to only typical lessons. Nevertheless, an assumption was made that frequent visits would decrease the risk of atypical lessons. As a result, eight to ten visits in each of the nine participants´ classrooms afforded the study some breadth while still enabling an in-depth approach to each of the cases.

Another point of view regarding balancing depth and breadth concerns data-gathering procedures. As Herriot and Firestone (1983) point out, cross-case analysis and comparison benefit from using shared definitions of concepts and data-collection procedures. Thus, researchers conducting multiple case studies need to resolve the extent to which they will use "structured" data-gathering procedures (Firestone & Herriot, 1982). Again, an increase in structure may increase the breadth as more data can be gathered and compared (see e.g., Huijgen et al., 2019 for entirely structured observation), but on the other hand detailed, less-structured descriptions of each case may increase the depth of the data. I arrived at a compromise similar to that of Nokes (2010), who conducted observations both through a structured instrument and field notes. The data-gathering process is described in more detail later.

Reposted methodological approaches often overlap and draw on similar traditions. For example, case study and ethnography are connected through their interest in the particular. In addition, ethnographic and case study research often employ similar methods, with observation being the most essential. However, while all ethnographies involve case study, case study is not synonymous with ethnography as the former may serve both qualitative and quantitative research traditions (Brewer, 2000). Below, I elaborate on what distinguishes the present study from ethnography methodologically and thereby seek to further clarify the nature of the study.

Although I have attempted to gain an in-depth understanding of the observed classrooms, three interconnected features in this study set it apart from ethnography. First, unlike ethnographers, I was not concerned with "the full range of social behaviour within location, event or setting" (Pole & Morrison, 2003, p. 3). My interest did not extend to all types of interactional contexts in the classrooms, which would have also included more informal situations. Instead, I focused only on specific lessons and particularly on teachers' actions. It is also worth noting the primary school context, where primary teachers teach several subjects. My observations included lessons on only one subject. Thus, my observational approach does not warrant interpretations about these primary teachers in general or about their classroom culture. Second, my observations

were organized around the specific concept of historical literacy, and teachers´ practice and thinking regarding that concept. This may be too narrow a focus for conducting ethnographic research, which strives to gain a more holistic understanding of a phenomenon (similar to Kitson, 2020). For example, in her ethnography, Vesterinen (2022) studied the enacted history curriculum as a whole. Third, rather than taking on the role of a participant observer, as assumed in ethnographies, my role as a researcher was less immersive (Adler & Adler, 1987).

There are several ways to characterize the role or position of the researcher. A widely used characterization is that of a participant (e.g., DeWalt & DeWalt, 2010; Spradley, 1980) or a non-participant observer. This dichotomous division is problematic as it oversimplifies the researcher´s role and does not provide scope for more subtle role descriptions. The idea of a non-participant observer exists hypothetically for the most part as anyone present in an observation situation does participate in some way, irrespective of whether or not that is his or her intention. Rather than becoming a participant observer, I aimed for a more detached role in the classroom (cf. Kitson, 2020). This was not done in "denial of my own presence" (Simons, 2009, p. 163), or to claim that observations would as a result produce neutral interpretations. Rather, interaction with pupils and the teacher during the lessons was minimal for two reasons. First, as I will describe in more detail, the observation instrument required coding every five minutes. This would have made participating in classroom activities difficult. Second, active participation in activities and communicating with pupils during lessons may have revealed my own ideals for teaching history, which in turn may have influenced the teacher's instruction. While recognizing that observing in the classroom would entail membership in that space, I consider that membership to be a distant one. Adler and Adler (1987, p. 1) refer to this type of role as peripheral membership, which assumes that researchers refrain from "participating in activities that stand at the core of group membership and identification".

5.2.1 Gathering the observation and interview data

The nine teachers participating in the study taught in eight different schools in the Helsinki metropolitan region. Each teacher was selected through recommendations by administrative staff or other teachers, not through a randomized process. The recommendation did not imply that the teachers were known for a particularly advanced approach to teaching history, but that they were willing to let researchers into their classrooms. The main criterion for enrolling in the study was several years´ working experience as primary teachers. In addition, the participants were diverse in terms of specialization in history as a school subject: five of them had no extra credits in history, and four had credits to a

varying degree.[2] As history is typically taught in grades 5 and 6, it was also considered advantageous to include teachers from these two grade levels.

The observations were carried out between autumn 2018 and spring 2019, with each teacher observed for 7–10 consecutive lessons, although the original plan included making eight visits to each classroom. The variation in the number of classroom visits was largely due to the difficulty in coordinating the schedules of several teachers. In addition, some lessons were cancelled due to teachers´ absence or school trips. In total, I observed 74 lessons, which amounted to 3,325 minutes. While I was present in every lesson, a colleague visited eight teachers´ classrooms once so that the reliability of the observations could be ensured.

During the observations, I used an instrument (Appendix A) and made field notes. A starting point for the observation instrument was one described by Nokes (2010). His instrument was further developed and tested five times in different settings to ensure its compatibility with the Finnish context. This testing process necessitated modifications, for example in the number of categories and in the duration of the coding period.

The final instrument consists of three sections: text types, classroom activities and historical literacy heuristics. The first section includes 16 text types, such as "textbook", which was coded whenever the textbook was used in the classroom. When the use of the textbook included text types other than the body text, for example maps and visual texts, they were coded in their own categories but marked as originating from the textbook. This was done in order to capture the use of the textbook in as much detail as possible. Each text type had a written description to enable the most precise coding possible. The coding period was five minutes, during which time each text used in the lesson was marked in the instrument. Primary sources were not coded during observations but subsequently identified by analysing field notes. This was done to ensure that determining which texts were primary sources would not be compromised due to the hectic pace of the observations.

As for classroom activities, the ten categories in the instrument aim to capture what was being done and by whom. For example, three of the categories concern teacher-pupil interaction: direct instruction, discussion, and Initiation-Response-Feedback interaction (hereafter IRF). Direct instruction refers to a teacher-centred approach, where the teacher conveys information to students (Wells, 1998). The other teacher-centred activity included in the instrument concerns asking close-ended questions using the IRF structure where, despite student participation, they lack opportunities to influence the direction of the dialogue or to take initiative (Lemke, 1990). Unlike texts, only the most predominant activity was coded.

[2] For a more detailed description of the teachers, their work experience, the grade levels taught, and materials used, see Table 1 in Article II.

Although Wineburg´s (1991) heuristics are included in the instrument, they do not have a discernible role later in the analysis. I coded the heuristics during the observations, but this section was not able to yield reliable data. This can be traced back to the development and testing phase. While testing the instrument, very few examples of these strategies were found, making it difficult to draw conclusions on the usability of the heuristics. As the data gathering progressed, it became evident that identifying sourcing, corroboration or contextualization in fast-paced classroom situations would have required a more specialized instrument.

Each lesson generated 1–2 pages of field notes on average, which included observations not only about teachers´ actions and utterances, but also teacher-pupil and pupil-pupil interaction. Similarly to the coding period of five minutes in the observation instrument, I structured the field notes in 5-minute intervals so that the data from the instrument could easily be combined with the data from the field notes. Without meaning to exclude any aspects of the lessons, all notes were written through a specific lens – that of historical literacy. For example, the notes seldom include descriptions of historical facts (or lack thereof) in teachers´ talk, but instead contain inferences about text-related instructions, comments, and the way history as a domain is addressed. In addition, the field notes include the use of historical thinking concepts, such as cause and consequence. Most lesson descriptions include notes about the overall atmosphere in the classroom, as well as non-history-related debates and discussions on school rules and practices, for example.

I interviewed each teacher after the observation period. The interview included questions about a) the observation period itself, b) the teacher's planning and teaching process in general, and c) the teacher's conceptions of history as a discipline and school subject. Concerning the observation period, I asked about specific instances observed in the lessons and asked the teacher to reflect on them in retrospect. I was interested in each teacher's intended outcomes and their views on what had taken place. In terms of planning, I asked each teacher to describe their typical planning process both in terms of the whole year and for individual lessons. I used Fogo´s (2014) Delphi panel results to ask what teachers considered essential teaching practices and whether they themselves used those practices. Finally, I asked teachers to define some central disciplinary concepts as well as to reflect on the nature of historical knowledge. My aim was to identify what the concepts meant as far as the teachers were concerned, and whether or not they were familiar in the first place. I appreciated the fact that terminology and language used by researchers and practitioners may differ. However, as the concepts discussed in the interviews are part of the national curriculum (National Board of Education, 2014), history didactics courses as well as textbooks, they should not be considered only of academic interest and beyond teachers´ grasp. Primary teachers´ knowledge base concerning individual subjects has been described as shallow (Stolare, 2017), although there is little research on the topic.

Therefore, I considered it essential to gain a better understanding of what the teachers knew about the disciplinary aspects of history. Each interview lasted an hour on average. For article III, I interviewed the two teachers, Amy and Eve, a second time as I needed more detailed data from them than from the other seven teachers. These interviews lasted for 20 and 17 minutes for Amy and Eve respectively.

5.2.2 Analyses of articles II and III

The analysis of article II utilized three types of data: 1) numerical data gathered through the observation instrument, 2) field notes, and 3) classroom materials. The main research question was approached through auxiliary questions, which concerned the types of texts used, how the texts were used, and what kind of activities were favoured by the teachers. As to what kind of texts were used, numerical data were employed for the most part, although some coding required analysing the classroom materials more thoroughly in order to define the right category. The manner in which teachers used a text was analysed to begin with using numerical data to define the length of use. Cases where a text was used for more than five minutes were then traced back to the field notes, which provided a detailed description of the incident. This phase of the analysis focused on identifying whether or not literacy strategies specific to historical literacy were used and by whom (the teacher or the pupils). The types of activities in the observed classrooms were analysed, again with the help of the numerical data first. The analysis continued by investigating the qualitative traits of some of the categories, such as direct instruction and discussion. For example, an examination of the field notes revealed incidences where procedural knowledge was part of the teachers´ direct instruction. These incidences typically included remarks on the work of historians.

The results answering the auxiliary questions helped to answer the main research question in article II and, as a result, teacher profiles were built based on five criteria: (1) instances where texts were present for more than five minutes, (2) instances of procedural knowledge, (3) use of textbook, (4) teacher-centred activities, concerning direct instruction in particular, and (5) high student participation activities, concerning discussion in particular. High numbers in the first two criteria increased the disciplinarity of a profile. First, engaging with texts in a historically meaningful way requires time. Second, students need procedural knowledge to understand how historical knowledge is constructed. High instances of textbook use contributed to a less disciplinary profile because, similar to Nokes (2010), we consider the dominant role of the textbook to be counterproductive for learning historical literacy and consistent with furthering cultural literacy.

For reliability purposes, I compared observation sheets from lessons where both authors of the article were present. In terms of texts, a point-by-point

comparison was made for each row (i.e., text type, see Appendix A), where at least one observer had coded texts. Rows left blank by both observers were not taken into account. The comparison revealed an 87% agreement on texts. As for activities, instead of comparing rows, each column was compared (i.e., each 5-minute period, see Appendix A) as only the most predominant activity was coded. This comparison revealed an 84% agreement on activities.

Article III had two aims, first to explore what kind of discourses existed in the lessons, and second, to understand how teachers´ thinking explains the observed discourses and the way teachers recontextualize history. To answer these questions, field notes and teacher interviews were analysed. For the discourses, theory-based content analysis (Gläser-Zikuda et al., 2020) was conducted on the field notes. The discourses and their subtypes were predetermined and these categories functioned as the coding scheme for the field notes. I did, however, take into account that not all data would fall under the predetermined categories.

The field notes were written in five-minute periods, which thus formed a coding unit. For each of these units, it was determined whether vertical (V) or horizontal (H) discourses could be detected (see Table 3). An appropriate subtype was duly determined. While all the subtypes of the vertical discourse, namely substantive, second-order or procedural knowledge, were formed based on existing research literature, some of the subcategories of horizontal discourse (contemporary phenomena, historical culture) were formed based on the data.

Table 3. Coding scheme for the discourses.

code	explanation
Hpe	Horizontal discourse, personal experience
Hhc	Horizontal discourse, historical culture
Hcp	Horizontal discourse, contemporary phenomenon
Vsk	Vertical discourse, substantive knowledge
V2oc	Vertical discourse, 2nd order concepts addressed specifically
Vpk	Vertical discourse, procedural knowledge
V→H	Vertical discourse used for understanding a horizontal discourse
H→V	Horizontal discourse used for understanding a vertical discourse
H+V	Two discourses existing in close proximity to each other but lacking a clear direction

In order to examine the interaction between discourses, I also identified instances where horizontal discourse led to vertical discourse or vice versa (see Dempster, 2021). However, as the analysis progressed, another category (H+V) was created to describe situations where a discourse consisted of vertical and horizontal elements, but where the direction between the two discourses could not be detected.

Qualitative content analysis was also applied to the interview data in order to investigate the teachers' role in recontextualization. Again, because the themes and subthemes were identified based on the theoretical background of the study, the analysis was theory-driven. Three themes structured the analysis of the interviews. The first theme addressed teachers' familiarity with vertical discourse (for example, substantive knowledge, nature of historical knowledge, central concepts), namely the potential for conscious recontextualizing. The second theme dealt with the teachers' planning process and the role of curricula and teaching materials in it. Moreover, the second theme addressed teachers´ choices for pacing and sequencing their teaching. The third interview theme concerned the way teachers thought about teaching practices.

A close examination of the analyses of articles II and III indicates an inconsistency concerning the procedural knowledge found in Eve´s and Amy´s classrooms. For example, in article II, Eve has seven instances of procedural knowledge, and in article III the equivalent number is eight. This difference is explained by the different rationales in the coding processes. In article II, the duration of an instance of imparting procedural knowledge was not taken into account; an instance was coded whether it lasted for one minute or ten. In article III, on the other hand, there was a coding unit of five minutes according to which everything was coded. In addition, as described earlier, in article III some of the procedural knowledge was coded as part of an interaction between the discourses, not as an independent entity.

6 Results

The aim of this study is to discover both the impediments and the potential related to teaching historical literacy at the Finnish primary level. The results of the three articles are described vis-à-vis the impediments and the potential found in the data. First, I present some of the constraining factors by describing them article by article, before providing an outline of those aspects considered to have potential for incorporating historical literacy into Finnish classrooms.

6.1 Impediments

6.1.1 Article I

The first article set out to develop a valid assessment tool for assessing primary pupils´ historical literacy. I was interested in the validity of the task, namely whether or not it would be able to elicit thinking relevant for historical literacy. The constraints found during the design process are the material´s: either in relation to the overall clarity of the instructions and individual items, or in relation to participants´ reading and reasoning processes. Pupils´ written answers, as well as those given during the think-aloud protocols (TAPs), revealed not only unclear instructions but also whether or not the prompt elicited a different kind of thinking than I had originally expected. In what follows I introduce some of the factors that seem to demarcate the space in which historical literacy can be assessed at the primary level. In keeping with the nature of design-based research, I have attempted to resolve some of the detected impediments while further developing the assessment tool.

Difficulties in detecting contradictions

When originally creating the assessment tool, one of the underlying assumptions was that pupils would notice the intentionally constructed contradiction between the three written sources. In order to further emphasize the existence of differing views about the relationship between Copernicus and the Catholic Church, I gave the tool a dichotomous title: “Copernicus – friend or foe of the Catholic Church?”.

In the first version of the artefact, no specific instructions were given for reading the three written sources as my assumption was that pupils would detect the contradiction between them. The two TAPs in phase 1 (see Table 2) did not elicit any changes to the instructions: one of the TAP participants was not able to

verbalize any thoughts while reading the sources, but the other TAP participant expressed some understanding of the different points of view of the sources, while largely focusing on their astronomical content. The five TAPs in phase 3 underscored the pupils´ tendency to focus on the astronomical content of the sources rather than on the conflicting views about the relationship between the Catholic Church and Copernicus. In the process, they formed a single, consistent narrative where the contradictory features of the sources were dismissed. The fact that pupils were instructed to read all the sources consecutively may have contributed to merging the sources into a coherent narrative.

Owing to the difficulties pupils experienced in detecting the contradiction between sources, I added specific instructions for reading the sources to the final version of the tool. The instructions advised pupils to ponder "What kind of relationship did Copernicus and the Catholic Church have according to the source?". Further, the instructions advised pupils to "pause after each source" and "circle an appropriate image" reflecting the source´s stance on the issue. These images were three emoticons expressing good/friendly, neutral, and bad/hostile relationship between the two parties involved. Circling the emoticons had two aims: first, simply to stop pupils from reading all the sources one after another at a fast pace. Second, identifying the sources´ position on Copernicus´s work would emphasize the contradiction.

One of the reasons for the difficulties in detecting the contradictions in the sources may be the non-existent attention received by the title of the entire assessment tool, "Copernicus – friend or foe of the Catholic Church?". Phase 3 showed that none of the TAP participants read the aforementioned title of the task aloud. This would suggest that many of the participants taking the pen-and-paper test were also unaware of the title´s significance to the task as a whole. The title was designed as an inquiry question that would ideally guide pupils´ reading process as it insinuates the conflicting views on the subject. Thus, as the title went unnoticed by the pupils, they read the sources for overall information rather than for evidence.

Tendency to read sources for information rather than for evidence

The first of the two open-ended questions was designed to assess pupils´ ability to evaluate the reliability of different source types by asking "Is the information given in the textbook (source 4) as reliable as sources 3 and 5 regarding the relationship between Copernicus and the Catholic Church?".[3] Hence, pupils were

[3] Source 3: an excerpt from the preface written by Copernicus in 1543.
Source 4: an excerpt from a Finnish history textbook written in 1982.
Source 5: an excerpt from a letter written by Cardinal Schönberg to Copernicus in 1536.

expected to engage in sourcing by considering why, when and by whom the sources were produced. In addition, the phrasing of the question also directed pupils to evaluate the limitations of primary and secondary sources in general. The first testing phase suggested that the participants had not fully grasped the idea of reliability. Half of the pen-and-paper answers (9/18) focused on the content of sources and did not consider by whom and why the source was produced. Although four participants did refer to the properties of source types in general and three participants viewed authorship to be relevant for reliability, the item did not meet its objectives.

In consequence, I modified the first open-ended question. Its focus shifted from assessing each source´s reliability to deciding which source pupils would use if asked to write a "credible account". In an attempt to clarify the item, I also divided the question into two parts. Instead of revealing in the prompt which sources were primary, pupils were asked to identify them in the first part of the question. This was intended to increase pupils´ awareness of different source types, which in turn would lead to using more disciplinary argumentation in the second part of the question ("Imagine that you were asked to write a credible account of the relationship between Copernicus and the Catholic Church. You could use only one source to help you. Which source would you use? Justify your choice".)

Despite the modifications to the item, a third of the participants in the second testing phase chose a source based on the amount of information it provided. These participants typically paraphrased the content of the sources without giving an actual argument for their choice. On the other hand, twelve participants did justify their choice of source by referring to authorship, which points to a more disciplinary way of reading the sources. Of these twelve participants, only two compared two sources with one another. However, the item should have elicited the act of corroboration among more than two participants, neither of whom compared all three sources. Interestingly, all twelve who suggested authorship as a reason for choosing a source opted for Copernicus´s preface. Despite being a primary source, Cardinal Schönberg´s letter was not chosen by anyone.

In order to encourage pupils to compare the sources and take into account the nature of primary and secondary sources, the question was slightly modified after phase two. Instead of having to choose only one source, the pupils were given the option to choose either one or two sources. This, I believed, would increase the comparison between sources instead of merely choosing one.

The written answers in testing phase 3 did show a change in the type of answer that the item generated. Twelve participants, comprising the majority of the class, based their answers on the properties, not the content, of the source. Similarly to the previous phase, there were some (4) who pointed directly to the author when considering the credibility of sources. However, instead of leaning only on authorship as an argument, answers included other aspects as well. Of these twelve

answers, five referred to the age of the source and three used the concept of primary source to reflect on the properties of different source types in general.

While the written answers suggested some change to the previous phase, the TAPs in the third testing phase pointed to reading strategies similar to those in the earlier phases: most of the participants associated credibility with the amount of information that a source provided and did not spontaneously compare all three sources. Only one participant mentioned all three sources when reasoning:

> I would choose sources 3 and 5 because they are primary sources and one can trust them a lot. Because with source 4 one can never know if everything important has been used for interpretation. (Think-aloud, pupil 65, phase 3)

Although only one TAP drew a spontaneous comparison between all three sources, the other four produced worthwhile reflections when asked to explain why they would not choose a certain source:

> Partly because it´s not a primary source, because it feels, like, somehow more credible when one knows what didn´t happen and there are people who tell what actually happened and their view of it. This [the secondary source] on the other hand is something where someone has found facts from different places and we don´t really know if it´s true. (Think-aloud, pupil 64, phase 3)

Since the think-aloud participants were able to produce sound argumentation using negation, that is, explaining why they were not doing something, the item was modified accordingly. In the final version of the task, pupils are asked to justify in a graphic organizer both why they would and would not choose certain sources when constructing a credible account of the relationship between Copernicus and the Catholic Church.

Overall, the testing phases of the first open-ended question showed that pupils relied on general comprehension strategies and focused on the content of sources; the more detailed information a source was perceived to provide, the more credible it was considered to be. However, as the item transformed through the testing phases, there was some increase in the use of source types in pupils´ argumentation.

Limited validity of weighted multiple-choice items

Weighted multiple-choice questions (WMC) offer several options that are correct to a varying degree (scoring e.g., 0, 1, 2 and 4 points). Ideally, contemplating

different but very similar options would necessitate pupils to engage in in-depth thinking, while simultaneously minimizing the need to write lengthy answers. In the case of both WMC questions, the TAPs revealed thinking which was partly in keeping with the target constructs, namely objectives 2, 3, 5 and 10 in the national curriculum (see Table 1). On the other hand, both WMC questions also elicited thinking which was questionable in terms of historical literacy.

The first WMC item was designed to assess pupils´ ability to recognize the motives and intentions of those who have produced historical sources. The item asked students to choose an alternative that would best explain why Copernicus dedicated the preface of his work to the Pope. However, of the seven TAPs, only two included reflections on those options that resulted in scoring points. The other five participants decided on one option, discarding other viable options relatively quickly. Thus, the intended aim of producing in-depth, complex thinking was not realized. One reason for this may lie in the fact that the scoring was not made visible to the pupils and therefore they presumably treated the item as a traditional multiple-choice item: once the participants found a suitable option, they assumed that the other options would be incorrect. In order to underscore that the item did not include just one entirely correct option, the scoring was added to the prompt after the third testing phase. This would, ideally, lead to a more thorough weighing process between those options that are correct to a varying degree.

The other shortcoming of the first WMC item was that the TAPs in both phases suggested that while participants did consider Copernicus´s motives, the historical context did not play a significant role. For example, most participants did not take into account the role and power of the Catholic Church in the Middle Ages. Instead, some participants merely referred to Copernicus´s feelings and explained his actions through general conceptions of how people think and feel.

The second WMC question, which dealt with the construction of historical knowledge, was relatively more successful in eliciting the intended thinking: all seven TAPs in both testing phases suggested that this item made participants reflect on the processes related to constructing historical knowledge. Two of the participants in the third testing phase did not compare all the options and quickly decided on one. In addition, there was considerable variation in the level of sophistication of participants´ reflections. However, all seven think-aloud participants did reflect on how sources are used by textbook authors in producing narratives. One of the think-aloud participants pondered the issue for eleven minutes, considering aspects such as the expertise of textbook authors and the standing of individual sources within a larger historical context. During the think-aloud interview, this participant also pointed out the difficulty caused by the three rather similar options in the question.

The TAPs did not explicitly reveal false positive answers, namely correct answers originating from entirely unintended or incorrect thinking processes (see Shemilt, 2015; Smith, 2018), as the participants did not, for example, justify their

answers by citing the length of the given option. However, this may have occurred, especially when taking into account the pen-and-paper tests. In the last phase, where 21 participants produced a written answer, 14 chose the highest scoring option for the first WMC item and as many as 19 for the second WMC item. In both cases, the highest scoring option was the lengthiest one. Based on both the TAPs and the written answers, the WMC items seem to have compromised validity, although in the case of some participants, the items were able to elicit historically literate thinking.

6.1.2 Article II

The second article utilized only observational data and focused on text types, activities and literacy strategies employed in primary history lessons.

Predominance of textbooks and scarcity of primary sources

The observations revealed that textbooks dominated the textual space in the classrooms (see Table 4). Although the frequency of textbook use varied among teachers (30–56%), it was the most used text type in all but one classroom. On average, textbooks were used for 45% of the time in history lessons. It should be noted that the data presented on the use of textbooks include all text types, with the body text being one of them. However, reading the body text was the most time-consuming way of using the textbook; other text types (paintings, drawings, maps) were seldom subjected to lengthy inspection or discussion. In addition to the textbook, another dominant explanatory text type was teacher´s notes, typically in the form of PowerPoint slides but occasionally also in the form of handwritten notes on the blackboard. Again, there was considerable variation among teachers regarding the use of notes (2–39%).

Table 4. Texts and activities observed in the classrooms in relation to the total time.

		Teacher								
		Amy	Brian	Chris	Daphne	Eve	Fiona	George	Henry	Ida
Texts (%)										
	textbook	43.9	35.5	49.3	50.0	55.7	30.1	47.5	55.9	33.8
	teacher´s notes	20.0	14.4	13.9	1.6	7.0	38.3	39.3	22.1	19.0
	non-fiction text	1.2	10.5	8.8	3.1	2.3	6.8	36.0	0	22.0
	newspaper	0	0	1.3	0	0	0	0	0	5.4
	administrative document	0	1.3	0	0	0	0	0	0	0
	photograph	3.6	0	0	0	3.5	12.3	13.1	1.5	2.7
	painting or drawing	4.8	15.7	12.7	9.3	14.0	21.9	37.8	13.2	20.0
	lyrics	4.9	2.6	0	9.3	0	0	26.2	0	0
	maps	6.0	10.4	1.3	1.6	16.0	9.6	13.1	10.3	2.7
	statistics	0	0	0	3.1	0	0	1.6	0	5.4
	video	3.6	0	13.9	4.7	14.0	11.0	11.5	0	5.4
	buildings and artefacts	4.8	0	6.3	4.7	8.1	9.6	27.9	11.8	8.1
	other	0	0	0	6.3	11.0	0	6.6	0	14.0
	No texts	30.4	30.2	17.7	26.6	26.1	24.6	18.0	25.0	20.3
Activities (%)										
	direct instruction	7.3	22.4	17.7	3.1	33.0	35.6	13.1	13.2	13.5
	IRF	13.4	14.5	12.7	28.1	10.2	11.0	0	11.8	5.4
	giving instructions	8.5	15.8	11.4	17.2	12.5	15.1	16.4	19.1	13.5
	reading	14.6	9.2	11.4	15.6	4.5	6.8	0	11.8	1.4
	discussion	15.9	14.5	5.1	0	4.5	2.7	0	11.8	2.7
	group work	8.5	10.5	17.7	7.8	6.8	8.2	0	8.8	9.5
	individual work	19.5	2.6	6.3	9.4	12.5	9.6	54.1	8.8	33.8
	video	3.7	0	10.1	3.1	4.5	5.5	11.5	0	5.4
	other	0	2.6	3.8	6.3	2.3	4.1	1.6	4.4	2.7
	no activity	8.5	7.9	3.8	9.4	9.1	1.4	3.3	10.3	12.2
	activities in total	100	100	100	100	100	100	100	100	100

Note. We coded all texts used in the classroom during the 5-minute coding period. Because there were coding periods with multiple texts, the sum total of percentages does not add up to 100. The corresponding figures for the activities do add up to 100 because only the most predominant activity (at least three minutes) was coded.

During the lessons, textbooks were mainly approached as neutral repositories of information. For example, teachers did not comment on the selection of various topics covered in the book and how the topics were sequenced. Moreover, the perspectives through which the textbook narratives were built were not commented on. Some historically controversial expressions such as "voyages of discovery" were addressed occasionally by Brian, Chris and Eve, but the relevance or the choice of verbs for the narrative was not discussed.

Although the textbook and teacher´s notes were by far the most used resources as a whole, there was nevertheless versatility in the textual space in the classrooms. As Table 4 shows, more than eleven different text types were present in the lessons. The category "other" included text types such as music excerpts, pupils´ presentations and artwork.

One text type was particularly under-represented in the observational data. News items (either newspaper articles or online news) were used only by Chris and Ida, and in both cases, only once. Ida´s pupils read a news article on new archaeological discoveries made in Egypt. Chris, on the other hand, began his lesson on European exploration in the Americas by presenting and discussing news on the dispute between Spain and Mexico, where the latter demanded an apology from the former. In both of these cases, the teachers were able to use the news for demonstrating the relevance and consequences of the past in relation to the present. To this end, as news items were used infrequently, their potential in most classrooms was not exploited.

Written primary sources were scarce in the lessons, as only Ida and Brian utilized them in their instruction in a manner in which pupils knew they were dealing with primary sources. In both of these cases, the written primary sources were provided by the textbook materials in the series *Mennyt* ('The Past'). In addition to using written primary sources from the textbook, Brian also visited a digitized archive (Diplomatarium Fennicum) on one occasion. He did this when asking the pupils what and how they had come to know about the Middle Ages. Brian had pre-selected an administrative letter from the archive, and in addition to reading a few words from the letter he showed a photo of the original handwritten letter to the pupils. This was the only time during the 3,325 minutes of observation that archives either as a resource for primary sources or as a general concept were introduced to pupils.

Three of the nine teachers, Eve, Fiona and George, used the textbook *Ritari* ('Knight'), which includes spreads under the heading "Investigate". While these investigative spreads give the expression of providing primary sources, for the most part the text excerpts are not primary sources but explanatory texts, the origin of which is not stated. Even in one or two cases where *Ritari* does provide written sources that could be considered primary, the textbook does not convey this to the pupils, and in Eve´s, Fiona´s and George´s classrooms this was not explained by the teachers either. Therefore, although George´s pupils, for example, did interact

once with a written primary source on people´s reactions to the Plague, they were unaware that it was a primary source and did not approach the text any differently compared to explanatory texts or body texts in the textbook. Hence, this encounter was not counted as an interaction with a written primary source.

Visual texts were presented frequently in all of the classrooms, with most of them originating from the textbooks. However, the visual texts in Fiona´s and George´s (84% and 70% respectively) lessons were mainly retrieved from the internet and were included in their notes, which, as stated, formed the backbone of their lessons. For all teachers, visual texts were more common than written ones as primary sources. In particular, pictures of artefacts such as pottery, weapons and buildings were used. Visual texts, such as paintings and drawings, formed a significant part of the texts but their role as primary sources was less significant as they were often created in a different time period from the event they depicted. For example, one of the textbooks presented a 19th century painting depicting the events of the Club War (1596–1597).[4] In addition to the text type of texts, the study revealed the way in which texts were used in the lessons.

Hectic pace and lack of explicit disciplinary literacy strategies

One of the most important findings that emerged through the observations was that both visual and written texts, other than the textbook or teacher´s notes, tended to be used only for a short period of time. As shown in Table 5, it was relatively rare for texts to be addressed for more than five minutes. Again, there was considerable variation among teachers, as Chris and Henry had only two instances of addressing texts for more than five minutes, while I found eleven corresponding instances in Ida´s classroom. Teachers also differed in whether they provided pupils with the opportunity to engage with a text for a longer period of time or whether they themselves were the ones engaging with texts. In four classrooms, pupils never worked with texts (other than the textbook or teacher´s notes) for longer than five minutes.

This tendency to address texts only briefly contributed to a hectic textual atmosphere: texts (again, other than the textbook and teacher´s notes) were looked at in passing before moving on to the next one. This hastiness was even commented on and objected to by pupils in Amy´s classroom.

In terms of the function of texts, the fast pace allowed mainly two approaches. First, texts served as supporting material for narratives presented either by the teacher or the textbook chapter. Second, texts were used as visual stimuli and for illustrative purposes, which was the case particularly when they were included in

[4] A peasant uprising which was part of the power struggle for the throne in Sweden. The war was also connected to the tensions between Catholics and Protestants.

PowerPoint presentations, where their origin, purpose and context was often omitted. In neither of these cases were texts introduced as sources to be analysed and interpreted. My field-note excerpt dealing with the European invasion of America aptly illustrates the fast pace with which texts were introduced to pupils in Fiona´s classroom:

> Photo of a present-day Inca as well as some kind of drawing. A photo of Machu Pichu.
> All texts so far part of a PowerPoint presentation made by the teacher. Some texts might be primary sources, but students won´t be aware of it because teacher doesn´t tell them anything about the texts. The slides don´t have anything written in small print, which would help me to trace the sources.
> New drawing about the Incas, origin not given. Visible for about 10 seconds
> New drawing about Atahualpa and Pizarro, origin not given. Visible for about 15 seconds.
> New drawing about gold. Visible for about 5 seconds.
> Back to Atahualpa.
>
> (Field notes, 14 February, 2019)

Even when texts were present for more than five minutes, it was rare for any historical literacy strategy to be mentioned, let alone applied (see Table 5). The category "instances with any historical literacy strategy" includes cases where a reference was made to any process relevant for interpreting texts in a historically literate way. Typically, this meant that the teacher mentioned the age, author or reliability of a text, or a textbook assignment expected the pupils to do so. However, none of the nine teachers gave explicit instructions nor modelled how to interpret texts by employing historical literacy strategies. This, together with the fact that texts were only arbitrarily identified as primary or secondary sources, could be described as one of the main impediments to teaching historical literacy in the classrooms observed. Instead of disciplinary literacy strategies, teachers gave specific instructions on general reading and writing strategies. These included instructions on constructing mind-maps, constructing essay-type answers and drawing inferences about the importance of a passage based on whether the text was highlighted in bold. Thus, general rather than disciplinary literacy was promoted.

Table 5. Instances where texts* were used for longer than five minutes and instances of introducing procedural knowledge.

	Amy	Brian	Chris	Daphne	Eve	Fiona	George	Henry	Ida
text addressed for more than 5 minutes									
by the teacher	3	8	2	2	6	4	2	2	4
by the students	3	0	0	1	2	0	5	0	7
In total	6	8	2	3	8	4	7	2	11
Out of which instances with any historical literacy strategy	3	3	1	0	2	0	2	1	4
Instances of procedural knowledge	7	12	2	1	7	2	2	3	11

*other than the textbook or teacher's notes

A typical occurrence in classrooms, also indicating an emphasis on teaching general literacy, was when teachers directed pupils to find specific information in the textbooks. This required pupils to locate the information in the textbook and transfer it into their notebooks, thus resembling a copy-paste phenomenon.

IRF interaction over discussion

Teachers used IRF interaction (Initiation-Response-Feedback) twice as often as discussion, for 12% of the time on average. In keeping with the nature of IRF interaction in general, the IRF sequences in the observed classrooms focused on closed-ended questions. The initiations by the teachers dealt with the recollection of dates and names as well as comprehension of first-order concepts for the most part. The feedback phase of the interaction was largely conducted by stating that pupils´ responses were either correct or incorrect, duly lacking a qualitative component to the feedback. IRF was used either right after reading a textbook chapter (Amy, Daphne, Henry) or for revising the previous lesson and the related textbook chapter (Eve). Although IRF could, in theory, also be used for asking closed-ended questions about procedural knowledge, in all of the observed classrooms, IRF – when used – dealt only with ensuring the comprehension and memorization of substantive knowledge.

Discussion was employed for 6% of the time on average. While in George´s and Daphne´s classrooms there was no discussion at all, Amy´s pupils engaged in discussion for almost 16% of the time. During the analysis, I formed three categories describing the quality of discussions: lacking historical relevance (Amy and Henry), some historical relevance (Chris and Fiona), and high historical relevance (Brian, Eve, and Ida). This categorization was based on the extent to which the discussions were able to prompt justified interpretations rather than personal opinions. While, as noted, some of the discussions were historically relevant and not merely stating opinions, none of the teachers engaged the pupils in structured discussions guided by texts and related questions. Perhaps due to the spontaneity of the discussions, they were also brief, rarely lasting for more than five minutes.

Profiles

Profiles were created for each teacher, based on the criteria described in Chapter 5. Depending on what kind of literacy teachers´ classroom practices supported more, they were placed along a continuum ranging from content-area to historical literacy (Figure 4).

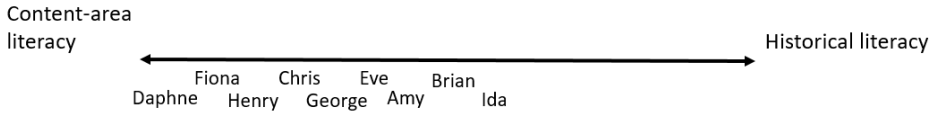

Figure 4. The nine participants placed on a literacy continuum.

None of the teachers´ classroom practices supported an approach that could be considered historically literate. However, there was considerable variation between teachers. Moreover, the composition of each teacher´s profile differed in terms of the five criteria.

Daphne and Ida represent the two extremes of the continuum. In Daphne´s lesson, the textbook was used about 50% of the time. Students were asked to read a chapter and answer questions stated at the end of it. These questions required only reading comprehension as the answers could be found in the text as such. At 28%, IRF interaction was the highest among all teachers. No instances of procedural knowledge could be identified in Daphne´s lessons. Ida, on the other hand, allocated more time to texts than any other teacher. This applied to all kinds of texts, including visual ones. Pupils themselves, not only the teacher, interacted with texts, and the textbook was used moderately (34%). On four occasions, Ida applied history-specific literacy strategies and there were 11 instances of

procedural knowledge. Although all of these aspects increased the disciplinarity of her instruction, most assignments required only text comprehension and collecting information. In addition, discussion was extremely rare (3%) despite the high number of participatory activities (47%).

6.1.3 Article III

In the final article, I set out to investigate primary teachers´ recontextualization process and the type of knowledge that exists in primary history. The article utilized both observation and interview data. Out of nine teachers, this article focuses on Amy and Eve.

A disconnect between vertical and horizontal discourses

The composition of the vertical discourse in both classrooms was similar: substantive knowledge was heavily emphasized over second-order and procedural knowledge. In Amy´s instruction, I coded 32 instances of substantive knowledge, zero instances of second-order knowledge and 4 instances of procedural knowledge. The corresponding numbers for Eve´s instruction were 40 (substantive), 2 (second-order) and 8 (procedural). In Amy´s case, most of the substantive knowledge originated from the textbook narratives as pupils read textbook chapters, completed related assignments and copied ready-made notes provided by the teacher´s guide. In addition to the aforementioned sources, substantive knowledge in Eve´s classrooms also originated from lecturing (see Table 4). Procedural knowledge was detected when Amy and Eve mentioned aspects of historians´ work, made remarks about properties of sources, or when textbook assignments required procedural knowledge. However, as described earlier, procedural knowledge did not take the form of explicit descriptions of the processes involved (either in textbooks or in teachers´ talk) in historical inquiry, and teachers did not model these steps to pupils.

Horizontal discourse in Eve´s classroom mainly comprised historical culture, either on its own or in close proximity to substantive knowledge. However, on the whole, horizontal discourse was seldom introduced in Eve´s classroom. Amy was significantly more active in including horizontal discourse through pupils´ personal experiences, contemporary phenomena and historical culture.

In addition to identifying the discourses individually, the analysis in the article focused on whether and how the discourses were connected to each other. One of the main findings of the article was that vertical discourse was used for accessing horizontal discourse only once. This took place in Amy´s classroom. No equivalent instance could be detected in Eve´s instruction. The lack of movement from the vertical to the horizontal resulted in instruction where personal

experiences, historical culture and contemporary phenomena were not deconstructed and reconceptualized through disciplinary tools and concepts. A connection from horizontal to vertical was somewhat more common, although only in Amy´s classroom.

There were several instances where a clear direction between the discourses could not be defined. Here, I considered a coexistence between the discourses to be the most apt description for their relationship. Where the discourses coexisted, elements of both were present in the same sequence of teaching, but a connection between the two was never established by the teacher. I identified one of these sequences in Eve´s classroom, where as part of a project on the Club War (1596–1597), pupils were asked to draw a cartoon depicting the events of the war. Pupils used textbook chapters, Wikipedia as well as a YouTube video for the project. The following excerpt from my field notes describes the activity of watching the YouTube video with the whole class:

> Eve: Now we'll watch a YouTube video that will last approximately ten minutes. You already have some kind of idea about the Club War.
> Student: Why is the music like this? (*soon after the start of the video*)
> Student: Yeah, why does the music have to be so epic? (*teacher does not react to students' comments*)
> (Video continues for another 5 minutes)
> (*The video includes lots of direct quotes. Where are they from? There are references at the end, but they are not introduced or mentioned to the students. After the video has ended, Eve, sitting next to me [A.K., at the back of the classroom, watching the video with the teacher], whispers that the video portrayed quite a romanticized version of the war, but this is not shared with the students*.)
>
> (Field notes, 15 February, 2019)

Here, the video represents horizontal discourse and, more specifically, historical culture, where history is depicted in an entertaining way. In the video, the events and characters of the war are introduced mainly through 19th and 20th century interpretations, which generally promote a national romantic perspective. The "epic" tone of the pompous music, noticed by the pupils, is also part of the same phenomena. Although the creator of the video listed these 19th and 20th century sources at the end of the video, Eve did not draw pupils´ attention to them. The national romantic tone of the video was, however, noticed by Eve as she pointed it out to me.

The video is a telling example of the disconnect between discourses. Vertical discourse in the form of substantive knowledge both precedes and follows horizontal discourse, but the two remain separate, disconnected entities. The

horizontal discourse in the form of a YouTube video is not unpacked, analysed or explained through disciplinary concepts and procedures.

Limited or misdirected recontextualization process

As defined in this study, recontextualization requires active, intentional and conscious decision-making (Alvunger, 2018) that takes into account the existence of different kinds of opportunities for action, and the exercise of choice (Priestly et al., 2017). This entails selecting, appropriating and refocusing knowledge as part of the recontextualization process (Hordern, 2021).

Amy´s way of planning and implementing her instruction relied heavily on the textbook and teacher´s guide (part of the textbook material). The interviews revealed that her planning process involved only the textbook, not the national or local curricula. Amy did not mention the curriculum at all in relation to planning, and when explicitly asked about the curricula (both national and local), she was unsure about the aims stated in them in terms of history teaching.

Amy explained that her planning process began by going through textbook chapters and by estimating the time needed for them. Her decision-making, in terms of planning, was restricted to deciding how many chapters to cover within one lesson. Amy emphasized that she would not be able to omit chapters from the textbook, and if pressed for time, she would rather combine chapters in order to cover all of them. Thus, the act of choosing something over something else did not take place. She did not question the choices made in the textbooks regarding the topics or the way they were sequenced.

Amy´s views on the importance of different teaching practices revealed that she had not previously thought about conveying procedural knowledge in the form of modelling how to read texts. When I asked the teachers to rate the importance of different teaching practices using a scale of 1–5 (five being extremely important), Amy did not, at first, grasp what was meant by modelling historical reading. After hearing the definition, she valued this teaching practice while simultaneously acknowledging that she had not employed it: “I gave it a four but I don´t think I´ve done this”. This explains the relatively low incidence of procedural knowledge in Amy´s lessons. When asked about the practice of “selecting and adapting historical documents”, Amy gave three reasons why she had given it a value of two. First, she thought that there were already sufficient historical documents available, and second, that the task of finding and modifying materials was too laborious given the number of other subjects to be taught. Third, Amy pointed out that using historical sources was not the only way to teach history successfully. This line of reasoning suggests that, for Amy, using only textbook materials stemmed from both time constraints and from the belief that selecting and adapting historical texts is not a necessary component of high-quality history

teaching. Instead, Amy relied heavily on passing on the already once recontextualized narratives in textbooks: recontextualization during the planning phase was limited at best, and could even be characterized as reproduction in Bernstein´s (1990) terms. Thus, whether or not historical literacy was promoted depended on the textbook´s approach to it.

While it seems that little recontextualization took place during the planning phase, Amy did introduce both vertical and horizontal discourses in the lessons and there was even some interaction between the discourses. However, as noted, this potential was not fully realized as the interaction and connection between the discourses was not explicated.

Eve´s planning differed considerably from Amy´s: her starting point was an overall plan, where textbooks were only one component. Eve´s process began by considering whether the study unit had the potential for a project. She then actively referred to curricula, both local and national. Her dissatisfaction with the most recent textbook series had led her to use two older ones. Although Eve used textbooks just as Amy did, their approach differed. For Eve, textbooks served as a useful resource to help her strive towards the goals she had set for her instruction:

> I mainly rely on ready-made materials, simply because there is not enough time. I just don´t have the time to create my own things. First, I always try to find good stuff in the textbooks because it makes my work a lot easier.

The excerpt shows that, as with Amy, searching for and selecting historical documents was not a realistic endeavour due to time constraints. In Eve's case, however, the use of textbooks seems to have been dictated by necessity and was not her preferred choice.

Similar to Amy, Eve found the idea of omitting topics altogether difficult, and hence rather than leave things out, she would cover the content by assigning additional homework or by using extra time for history lessons. The inability to skip topics was a limiting factor in Eve´s recontextualization. On the other hand, Eve reflected on how her way of sequencing the topics differed from the local curriculum. Thus, using multiple resources, sequencing topics and considering the curriculum in her planning point to an intentional and active recontextualization process.

As described above, in Eve's case, many of the criteria for recontextualization were met, and thus she cannot be described as only reproducing knowledge. The fact that Eve was consistent in her thinking and actions also suggests intentionality in her teaching. For instance, she regarded the practice of explaining content as extremely important, which is consistent with what she did during the lessons.

However, while Eve did recontextualize, the process took a direction that is incompatible with the aims of historical literacy. One of the contributing factors

for this was Eve´s thoughts about the ability of, or necessity for, primary pupils to examine sources in a disciplinary way. An excerpt in which Eve comments on her decision not to unpack the YouTube video with disciplinary tools reveals her reservations concerning the age of pupils in relation to historical literacy:

> Sure, the way the music and everything had been constructed, it had a sense of drama – things that get pupils interested in the topic, I guess. Of course, one can explain and teach a lot to pupils, but maybe at their age one doesn´t need to reduce everything to atoms. Maybe that´s what I thought.

Eve´s interviews revealed three interconnected reasons for a recontextualization outcome differing from the ideals of historical literacy. First, Eve chose not to emphasize procedural knowledge in her lessons, despite having a solid understanding of it. As she mentioned when reflecting on the qualities of different types of tasks:

> In this task, one deals with the nature of historical knowledge. And it involves comparing different interpretations, which is key in learning that history is always based on a historian´s interpretation of a set of sources. And it´s good that it states that the pupil should construct an argument.

Second, she perceived her own role as someone whose main responsibility is to explain the substantive knowledge encountered by pupils. Finally, in terms of creating a connection between horizontal and vertical discourses, Eve viewed the connection with pupils´ lives as a matter of sequencing topics rather than as an interaction with texts in a way that would be meaningful to pupils.

6.2 Potential

No unsurmountable impediments to using document-based tasks

Apart from the constraining factors described earlier, document-based tasks may entail different types of problems if pupils are not able to understand the content and the context of the task, or if the task proves to be too laborious. Here, I present my findings on pupils´ reasonable command over the historical context, key concepts and general literacy.

When testing the assessment task, the answers to the background questions showed that pupils were relatively well acquainted with the historical context. Of

the 67 participants, only one pupil thought that the Catholic Church did not have any societal power in the 15th century. The rest of the participants viewed the Catholic Church as extremely (N=54) or somewhat (N=12) powerful. The participants had no difficulties in identifying geocentric and heliocentric models either, as the vast majority of participants were able to identify which model was closer to the views of the Catholic Church in the 15th century.

Their understanding of the meaning of primary and secondary sources was also relatively good after the first testing phase, where instead of only two correct options, there were also two incorrect options. In the first phase, ten out of 18 pupils chose an incorrect option. To mitigate confusion, four options were reduced to two. In addition, instead of using only one Finnish term for the primary source (*ensikädenlähde*), another possibly more familiar one was added (*alkuperäislähde*) because one of the think-aloud participants stated that the terms used in the task were unfamiliar. As phases two and three generated few (4/42) incorrect answers, no further modifications were made to this item. To this end, it can be said that most participants understood the meaning of central concepts such as primary and secondary sources.

When working with written sources, basic reading comprehension is needed to access historical literacy because difficulties encountered in the former would restrict access to the latter. Therefore, testing the task provided an opportunity to reflect on whether the quantity and difficulty of written excerpts was compatible with pupil´s abilities to decode and understand the text within a certain timeframe (lessons typically last 45 to 60 minutes). The first testing phase revealed that 9 out of 18 participants in the pen-and-paper test used 23 minutes to complete the task. When 27 minutes had passed, 16 participants had completed the task. The last participant was ready after 42 minutes. The maximum time reserved for the task was 60 minutes. These results would suggest that the task required a level of general literacy that was achievable by the participants. This was also confirmed by the TAPs, where pupils did not struggle with reading the sources and the information box in which concepts were clarified and context provided.

As the participants did not struggle with the historical context, concepts and general literacy, it could be suggested that this type of task type utilizing historical sources is a sound option for primary school pupils. Previous studies in the US have already shown that primary pupils can successfully work with sources (e.g., Hughes, 2021; Nokes, 2014; VanSledright, 2002). The present study suggests similar findings in the Finnish context.

Teachers´ understanding of disciplinary concepts and practices

Despite multiple impediments identified in the nine classrooms observed, the interviews carried out in the case study revealed aspects that I regard as enabling factors for teaching historical literacy. During these interviews, I specifically asked the teachers about their knowledge base, beliefs and practices. Below, I provide some examples of the teachers´ thinking, which underscores the potential that primary teachers have for teaching historical literacy. I focus on three aspects, namely teachers´ understanding of the nature of historical knowledge, their conceptions of historical thinking, and their views on the use of primary sources.

Given that one of the main aims of historical literacy is to construct historical knowledge (Nokes, 2010), understanding its nature is essential for teaching it. Similarly, views on historical thinking are relevant as historical literacy, as defined in this study, is part of the larger entity of historical thinking.

I asked the teachers to reflect upon the nature of historical knowledge and possibly to compare it with knowledge produced by other disciplines. Most of the nine teachers engaged in lengthy reflections and everyone was able to produce an answer. The theme mentioned most often was interpretation in one form or another (the interpretative nature of knowledge, the process of interpreting, etc.). For example, in Amy´s view, "even if something [in history] has been researched and it is known that this is what took place, it is nevertheless… can be an interpretation". Connected to the interpretative nature of history, a common view was that historical knowledge is "less absolute" (Amy) or "less exact" (Chris, Eve, George) compared to knowledge produced by the natural sciences or mathematics. Fiona, on the other hand, also pointed out that historical knowledge is fundamentally no different from other knowledge, as "all knowledge is valid only until it´s proven invalid".

Only three of the nine teachers (Chris, Eve and Ida) used the concepts of evidence or sources in their reflections on historical knowledge. Ida, whose literacy profile I considered the most promising in terms of historical literacy, emphasized the constant change that historical knowledge undergoes due to the emergence of new evidence, and therefore, according to Ida, "in history the past changes". Similarly, Eve described history as "source-based detective work" where there is always the possibility that new sources will "shake it [existing knowledge]".

Amy, Brian and Henry emphasized the meaning of differing perspectives in knowledge production. As far as they were concerned, history unfolded through multiple competing perspectives. This multiperspectivity was seen as a challenge for knowledge construction. Ida, on the other hand, while underscoring that knowledge changed with new evidence, had a more cumulative view of historical knowledge. She described history as a patchwork quilt, where an understanding of history is created by finding new pieces.

Historical thinking was characterized both in broad terms and through specific elements. None of the teachers included all of the elements most commonly

referred to in the literature (e.g., Lévesque, 2008), but some teachers provided a relatively comprehensive characterization. The ability to put oneself in the position of those living in the past – historical empathy – was the most frequently mentioned part of historical thinking. Understanding cause and consequence as well as detecting change were mentioned by several teachers. Perspective-taking and developing a sense of chronology each received a mention. Perhaps the most significant finding was that all the answers were related to second-order concepts; none of the teachers identified historical thinking with the memorization of substantive knowledge.

Although teachers expressed some reservations, in principle they were all in favour of using primary sources in primary history education. An epistemological aspect was raised by many teachers. For example, Fiona articulated that "through primary sources, pupils can see where the textbook interpretations originate from". Another theme was that of authenticity, albeit largely for motivational purposes. According to Brian, the authenticity of primary sources would enhance pupils´ experience of "doing something real". However, George stressed that the authenticity of primary sources was not a must, as long as they appear to be authentic.

Interestingly, teachers emphasized that artefacts, such as the Rosetta Stone, were the most important primary sources at the primary level. For some, this was a practical question, as they felt that written sources were not easily available, while for others, the question related to pupils´ ability to read primary sources, particularly if they were not shortened and modified. Other reservations concerned lack of time and the extent to which primary sources should be used: Eve considered primary sources essential but expressed her concern about the idea "that every single thing that pupils should learn should come from primary sources".

As the results in the previous section show, the underlying potential described here was rarely visible in the lessons observed. Nevertheless, the results indicate that primary teachers are acquainted with the nature of historical knowledge. They also understand at least some aspects of historical thinking, and most view the use of primary sources as not only possible but also desirable. Had the interviews shown the opposite, the aim of incorporating historical literacy into primary instruction would ostensibly prove much more difficult.

7 Discussion

My point of departure for this study was to examine the impediments to and potential for teaching historical literacy that exist at the primary level, particularly in the Finnish context. As shown in the previous section, the impediments identified in the three articles seem to overshadow the potential. Hence, if historical literacy is considered a worthwhile instructional aim, there is a need to find solutions to the identified impediments. I begin the chapter by discussing the findings of this study in relation to previous research. Based on the results, I then suggest some measures for tackling the impediments. Before an overall reflection on the need for a disciplinary shift in history education, I consider some of the limitations of the study.

7.1 Conclusions

Comparing the results of the present study with those of earlier research is challenging because, as described in the theoretical framework, comparable studies carried out at the primary level are few and far between. Although most of the available studies are intervention studies or related to history instruction by expert teachers, they nevertheless offer a point of reflection. The results of Article II show that teachers rarely employed and never modelled history-specific literacy strategies. Instead, literacy instruction in history lessons focused more on text comprehension as well as on summarizing, locating information and constructing coherent answers, thus contributing more to content-area than historical literacy. Further, if epistemological issues related to history were addressed at all, they were addressed in an inconsistent way. These results have little in common with expert studies conducted at the primary level (Hughes, 2021; Nokes, 2014; VanSledright, 2002): the experts modelled explicit history-specific literacy strategies, established a set of procedures to be used when interpreting different types of texts, and engaged pupils in discussing the constructed nature of historical knowledge.

However, when looking at how non-expert teachers have been reported to teach the interpretation of sources, the results more closely resemble those presented here. Stolare (2017) reported that a primary teacher had difficulty maintaining a disciplinary approach when teaching about sources and their use. An emphasis on content and textbook narrative gradually replaced the more disciplinary approach taken at the beginning of the intervention. Neugebauer and Blair (2020) also noted a lack of disciplinary literacy strategies when studying the nature of literacy instruction across subjects. Thus, it seems that introducing

procedural knowledge is one of the differences between expert and non-expert teachers. However, the lack of disciplinary literacy is not unique to the primary level. Several studies show that subject teachers also rarely include history-specific literacy strategies (e.g., Gestsdottir et al., 2019; Hartzler-Miller, 2001; Nokes, 2010; Puustinen & Khawaja, 2022). Thus, subject-specific expertise does not automatically translate into practices that promote historical literacy or other disciplinary aims. This was also evident in the present study (Article II), where teachers with the most extra credits in history did not promote historical literacy more than those with little or no specialization in history.

In the Nordic social study context, studies have suggested that subject teachers do connect vertical and horizontal discourses frequently (Aashamer & Klette, 2023; Alvunger, 2018). However, neither of these studies discuss the interaction between the discourses in terms of historical literacy. Instead, Aashamer and Klette (ibid.) applied the framework for Authentic Intellectual Work (see Saye et al., 2018) and Alvunger (ibid.) looked at teacher agency. As a result, both studies seem to have a different view from that of the present study on what is required for the discourses to be viewed as connected, and further, to support knowledge construction. In order to contribute to historical literacy, the interaction between the discourses has to unpack the meanings and contexts of concepts and ideas, not only mention or introduce first-order concepts, as the example in Alvunger´s study (2018) suggests. The unpacking of meaning is at the core of recontextualization since concepts and ideas often have different meanings depending on whether they are used in horizontal or vertical discourse, in an everyday or disciplinary setting (see Pilli, 1988.) To that end, I found little interaction between the discourses that would support historical literacy (Article III).

Both articles II and III indicate a scarcity of procedural knowledge in relation to other forms of knowledge. While historical literacy requires procedural knowledge (Nokes, 2010, 2012; Downey & Long, 2016), it would be counterproductive to determine an ideal ratio between different forms of knowledge; there is no exact formula for historical literacy. However, the results presented in this study do encourage a discussion on the infrequent use of procedural knowledge and second-order concepts in classrooms.

Although discussion was rare and Initiation-Response-Feedback (IRF) interaction relatively frequent in the observed classrooms, the findings concerning activities also point to an encouraging tendency for historical literacy. The nine teachers in the present study used direct instruction considerably less than secondary teachers (Nokes, 2010; Puustinen & Khawaja, 2022) and middle school teachers in the US (Neugebauer & Blair, 2020). The willingness and ability of teachers to let pupils do more than listen to lectures has the potential to engage pupils in knowledge construction. However, the essential question is what takes place when pupils work independently or in groups – whether they are required to

analyse and interpret texts or locate and re-locate information. My findings are more consistent with the latter approach.

The predominant role of textbooks was similar to what has been reported about primary social studies instruction (Kon, 1995; Neugebauer & Blair, 2020; Sunal & Sunal, 2008; Zhao & Hoge, 2005). As the choices made by textbook authors and the genres they are written in (see Coffin, 1997) were not addressed, the potential that textbook narratives offer for learning historical literacy was not utilized. Other text types, especially their use as primary sources, were considerably fewer. Apart from a few examples (Van Nieuwenhuyse et al., 2017; Nokes, 2010; Puustinen & Khawaja, 2022), previous history education research has not examined the amount of time spent with texts. Although Nokes´s (2010) study shows the frequency of using different text types, it does not reveal the length of time spent with an individual text. My findings show that the majority of teachers demonstrated only a few instances where a text (other than the textbook or teacher´s notes) was used for more than five minutes. This means that pupils do not have even a theoretical chance to make meaning with any text type other than the textbook. Van Nieuwenhuyse and colleagues (2017) reported similar results at the secondary level, where on average teachers spent 2 minutes and 27 seconds per primary source. The role of visual texts and artefacts was typically to support a given narrative and to make it more vivid. The same tendency is reflected in Stolare and colleagues' study (2021, p. 272), where they found that a significant motivation for primary teachers to visit heritage sites was to make history "more vivid". These results, similarly to my own, reveal something about how primary teachers tend to perceive texts other than the textbook narrative. Whether written, visual or artefactual (includinge.g., excavation sites), primary teachers seem to use texts other than the textbook narrative as an illustrative support for these narratives.

The central role of the textbook offers an interesting perspective on the recontextualization process. I found textbooks to be either the key or sole component of teachers´ planning process, and thus to wield considerable influence in determining whether or not recontextualization took place and, if so, in which direction. While some teachers reported using curricular documents to guide their work, they often interpreted the documents from a less disciplinary perspective. Neither of the two teachers in Article III demonstrated recontextualization that would lead to historical literacy. In one case, the teacher largely engaged in the reproduction of knowledge stated in the textbook. In the other case, the teacher´s choices in relocating and sequencing knowledge, while pointing to recontextualization (Bernstein, 1990; Nordgren et al., 2021) were not compatible with the aims of historical literacy. Moreover, a historically literate interaction between vertical and horizontal discourses was limited in both classrooms.

Although it would be tempting to explain these results only by the fact that primary teachers lack direct access to the discipline of history and are not part of

the disciplinary community (see Fordham, 2016), the results require a more comprehensive explanation. Subject teachers have been reported to rely on their disciplinary knowledge while planning (Hartzler-Miller, 2001; Monte-Sano, 2013), especially those identified as agents of recontextualization (Kitson, 2020). However, as noted earlier, the existence of a disciplinary connection does not guarantee its application in the classroom (Puustinen & Khawaja, 2021, 2022; Vesterinen, 2022; Wilke et al., 2022), but seems nevertheless to make secondary teachers less dependent on textbooks and the way textbook authors recontextualize history (Nokes, 2010).

While primary teachers´ knowledge of the discipline of history cannot match that of subject teachers (see Puustinen, 2022), my results suggest that all nine teachers had at least a tentative understanding of the nature of historical knowledge; some even expressed an in-depth insight. Thus, while the results reveal that for some primary teachers a modest knowledge base is a significant reason for not recontextualizing – especially towards historical literacy – a knowledge base in itself is not the only explanation. Instead, the reasons for the recontextualization outcomes identified in this study are manifold and best understood through the pedagogic device (see Figure 1).

First, teachers' recontextualization in this study was affected by doubts concerning pupils´ ability to understand interpretational aspects of history. These doubts are not exclusive to the primary level (see Hartzler-Miller, 2001; Wilke et al., 2022; Wuokko & Räsänen, 2023). One of the core functions of teacher education is to provide research-based knowledge to prospective teachers about pupils´ abilities. If, however, teachers are unaware of or uncertain about what pupils are able to do given the right circumstances, it can be taken as a reflection of the relatively weak power of teacher education in the pedagogic recontextualization field (PRF). Thus, within the PRF, the power struggle between teacher education and textbooks seems tilted towards the latter.

Second, I found that primary teachers were not free from the influence of tradition (see Figure 1). In Stolare´s (2017, p. 47) words, "it becomes clear how strong school and the subject tradition are". Apart from other factors, recontextualization was limited to, or directed towards, aims other than historical literacy because teachers considered the traditional way to be sufficient or even preferable. The third reason points to the inner tensions of the official recontextualization field (ORF), where different school subjects compete for teachers´ time and attention. In this study, teachers related the lack of time specifically to teaching a variety of school subjects. Recontextualizing requires time resources, and as primary teachers are responsible for up to ten subjects, they may prioritize some subjects over others in their planning. The findings of Zhao and Hoge (2005, p. 220), while not identical to those of the present study, illustrate how the tension within the ORF may affect primary teachers:

> We can report that the textbook-centred approach to social studies was driven by teachers' desires to fulfil the minimum requirements of state and local curriculum guides while saving their best instructional efforts for more highly valued subjects, such as reading and math.

The results of articles II and III suggest that historical literacy was not taught actively or systematically in the nine observed classrooms. One reason for the results may be the lack of disciplinary assessment materials available for primary teachers. A more disciplinary approach to history requires assessment materials aligned with that objective (Brown et al., 1997). To date, the ORF has not provided any assessment materials for the primary level. As a result, teachers rely on the textbook materials, which seldom offer tasks compatible with disciplinary aims (Lyhty, 2020). Moreover, the time constraints for planning reported by teachers in this study suggest that they may be unlikely to develop new assessment materials for historical literacy themselves (Article III).

As the Article I design process shows, designing historically literate assessment material needs to take into account the balance between pupils´ emerging general literacy and the aims of disciplinary literacy. The balancing act between these two is recognized when discussing assessment at the secondary level (e.g., Reisman, 2015b; Seixas et al., 2015), but is even more apparent at the primary level. Pupils in the present study tended to use reading comprehension strategies for the most part, and to read the texts for information rather than for evidence. When scaffolding in the prompts and instructions was increased, the number of pupils who utilized disciplinary literacy also increased. However, the design process raises questions about the limits of scaffolding in an assessment task. At which point does scaffolding defeat the purpose of assessment if pupils are guided too specifically towards the way of thinking that is expected? Testing the final version in the future may shed light on this question.

My attempt to avoid the pitfalls of traditional multiple-choice questions (see Eliasson et al., 2015; VanSledright, 2014) led me to explore weighted options. Although the TAPs showed that the weighted options induced relevant cognitive processes and reflections among some pupils, the validity of the questions was limited. The challenges involved in weighted multiple choice (WMC) questions should not discourage new multiple-choice experiments as the problems identified in the traditional ones still exist. In addition, while secondary school assessment tasks can utilize constructed-response items exclusively (see Eliasson et al., 2015; Seixas et al., 2015), in the primary context there are limits as to how much pupils can be expected to write, especially during a one-hour test. Therefore, new ways of assessing complex cognitive processes in history are needed. Although Smith´s (2018) results on the cognitive validity of new multiple-choice items are inconclusive, they nevertheless show promise.

7.2 The way forward

If teaching history at the primary level is to promote historical literacy, the impediments identified need solutions. Whether or not historical literacy is a desired objective is up for debate and discussed in the final chapter. Here however, I focus on how different agents of the pedagogic device could contribute to more historically literate instruction if such an aim is deemed important.

The PRF has two strong agents: teacher education and textbooks. Their influence on teachers´ thinking differs considerably, both in timing and duration. Teacher educators´ interaction with prospective teachers is brief, whereas textbooks have a continuous relationship with in-service teachers. Despite the time constraints of teacher education, its contribution can be helpful for prospective teachers in two ways. As university-level instruction, it can provide research-based knowledge about children´s ability to learn history. Further, although brief history didactics courses cannot provide an in-depth understanding of history as a discipline, they can nevertheless provide some disciplinary understanding. This understanding is crucial if primary teachers are to recontextualize history in any manner. In view of the present results, procedural knowledge in particular should be included in history didactics courses, not only for the purpose of acquiring procedural knowledge but also for introducing ways to teach that knowledge, for example through learning how to model history-specific literacy strategies.

The second way in which teacher education may address the identified impediments concerns teacher education programmes in general, not history education courses in particular. If courses across subjects were to address the structure, transformation and meaning of knowledge, it would help prospective teachers to view themselves as active agents of the recontextualization process. An emphasis on knowledge may help them apply didactical trends in a critical manner. For example, instead of viewing participatory teaching as an end in itself, prospective teachers would learn to ask what that participatory activity is able to offer in terms of knowledge. Moreover, teacher education is the only agent of the device that has the possibility of making teachers aware of their role in the power struggles. Whether or not it is Bernstein´s (1990) framework that is used for this purpose is beside the point. What is important is that teachers are given the opportunity to see their role from a wider perspective than the classroom context. Teaching should be set in a wider political and societal framework in order to avoid the decontextualization of education (Säntti et al., 2023). Awareness of the different agents and their power struggles over knowledge is a prerequisite for teachers to find their place in the equation.

In addition to the initial teacher training, it may be useful to consider the potential that in-service teacher education has for promoting historical literacy. However, the current Finnish in-service education has been described as fragmented and non-systematic (Heikkinen et al., 2015). The advantage of in-

service teacher education is its ability to reach teachers who have already developed basic classroom routines and can direct their attention to deepening their understanding of individual school subjects. Therefore, compared with new graduates, in-service teachers may be better equipped to utilize new knowledge about ways to use texts in a disciplinary manner, for example. Moreover, in initial teacher training, the timespan between a history didactics course and working in a classroom may be several years. In-service education courses, on the other hand, provide teachers with the opportunity to apply new knowledge in their classrooms soon after the course.

As noted, textbooks seem to shift their emphases slowly (Lyhty, 2020; Norppa 2019; Rantala, 2017). This may be explained in part by publishers´ desire to tap into what teachers are assumed to want from a textbook – a memory-history approach to history. This assumption may prompt publishers to hesitate when it comes to publishing textbooks with a different kind of approach. In recent years, however, two textbook series have taken a more interpretational approach, providing primary sources and guidance on the interpretation of history at the primary level. One of these series was used in the present study by two teachers who were also the ones placed closest to historically literate instruction on the literacy continuum (Figure 4). Of course, no causality can be shown here between the textbook used and teachers´ literacy practices, but the finding provides food for thought. In any case, if greater change is desired when it comes to the approach adopted by textbooks, teachers are in a pivotal position to exert an influence, more so than researchers or ORF officials. Publishers are only likely to produce new kinds of material if they are convinced that there is a demand for such materials.

If the agents working in the ORF wish the curricular documents to have an impact on teachers´ thinking and practices, I suggest that it needs to produce materials aligned with curricular aims. There is a particular need for assessment materials because of the backwash effect; assessment defines much of what is taught and possibly learnt (see Brown, 2011). In the absence of assessment materials provided by the ORF, the renewed assessment criteria for year 6, applied for the first time in spring 2024, may function as a lever for change. This would, however, require the criteria to be viewed by teachers as a comprehensible, adoptable and useful tool. If not, the renewed criteria are unlikely to create a backwash effect. The aims of the National Core Curriculum need to be translated into materials by the same agents who create the official documents. This would help teachers to understand what is required in practice. In addition, materials from the ORF would at least to some extent provide an alternative to textbooks. The results of this study suggest that 1) textbooks are used extensively (Article II), 2) teachers rely on textbooks due to lack of time (Article III), and 3) that developing assessment materials for historical literacy requires time (Article I). Hence, it seems that teachers would benefit from materials provided by the ORF, if not comprehensively for all teaching, at least as a point of reference and to

generate discussion among colleagues, as reported in Sweden regarding their national tests (Eliasson, 2013). Otherwise, the parallel existence of curricula and classroom reality may persist (see Vesterinen, 2022).

In view of the evidence presented, I suggest including primary teachers in both recontextualization and reproduction fields. I further suggest that primary teachers´ recontextualization should be understood as the process of negotiating between official and pedagogic recontextualization fields and other factors, while determining their own position and power in the pedagogic device. Without access to each school subject´s parent discipline, primary teachers´ recontextualization cannot be identical to that of subject teachers (see Kitson, 2020). Nevertheless, given the right circumstances, primary teachers are able to recontextualize knowledge. Therefore, as agents of the device, some of the solutions for overcoming impediments are in teachers´ hands.

As decathletes of teaching, primary teachers cannot be expected to focus merely on one subject or to increase their workload considerably in the face of new approaches, for example to history. What can be expected is being open-minded to new ways of maximizing the resources at hand, both material and otherwise. By this, I mean a readiness to consider alternative ways of planning and implementing instruction. For example, deciding not to cover all the content in textbooks, and guiding pupils to analyse and interpret the available texts for longer periods of time, does not require hours of planning but rather a different frame of mind in relation to history. Thus, historically literate instruction does not necessarily require teachers to provide pupils with primary sources from various archives each week. Instead, teachers can use multiple ways to help pupils build a disciplinary relationship with texts in history classrooms. As the results show, most teachers did understand the epistemological foundations of history, but this was not reflected in their instruction.

Finally, the present study offers a perspective on the pros and cons of the Finnish education system, which has often been presented through a celebratory narrative (Thrupp et al., 2023). Teacher autonomy in particular has been viewed as the cornerstone of a system which was, until recently, able to produce excellent PISA results. However, autonomy also has its unintended consequences, such as those seen in the nine classrooms observed here. When teachers do not receive any external feedback, either through student assessment or someone observing their teaching, the aims set by teachers can differ considerably from those stated in official documents. While I am not suggesting the introduction of national tests or the revival of school inspections, I call for an open and wide-ranging discussion on some of the implications of the current education system.

7.3 Limitations

Despite my best efforts to conduct this study with rigour from data-gathering to analysis and to the inferences drawn, limitations are an inevitable part of any research. I have already discussed the limitations concerning the chosen methodological approaches in Chapter 5, and here I consider the topic further in the context of the present work. The primary goal of the study has been to identify what kind of impediments to and potential for teaching historical literacy exist. In terms of the construct validity of a test or research, Messick (1980) points out that it should be relevant, comprehensive, and technically sound in relation to the theoretical framework. In this vein, I touch on three aspects that should be taken into account when making validity claims about this study. The first two concern the comprehensiveness and the third the decisions concerning the execution of the research.

First, although the three articles contribute towards answering the research question, they offer only a limited perspective. The role of pupils in the study is not in keeping with the significance they have in the instructional process. In Article I, the pupils´ contribution to the design process was vital; I based my design decisions on the knowledge I gathered in relation to pupils´ thinking. However, in articles II and III, the focus is solely on teachers´ actions and thinking, and pupils are included only through their reactions to teachers´ actions in the lessons. Pupils´ learning, their views and perceptions are not included, and hence, this study is not able to produce knowledge on the way pupils perceive history teaching at large and, more specifically, the way they experience the interaction with different kinds of texts. To that end, primary pupils´ perceptions and reflections about history teaching and literacy are an exciting avenue for future research, particularly because very few studies have examined the learner's perspective in history classrooms (e.g., Neugebauer & Blair, 2020).

Secondly, despite the central role of textbooks in the classrooms and in teachers´ recontextualization process, an analysis of the textbooks is not part of this study. Textbooks series differ considerably in their approach to history. Therefore, the validity claims about this study would have been stronger had I examined each textbook series and the decisions made by textbook authors in relation to topics, forms of knowledge and materials for teachers, to name but a few. The lack of these two aspects alone – pupils and textbooks – limits the possibility to claim that the study is able to produce a comprehensive answer to the research question.

The third aspect is whether valid inferences can be drawn about the results considering the way the work was executed. Both datasets have limitations that should be acknowledged. Article I would have benefitted from more think-aloud participants throughout the three design phases. This would have provided a stronger foundation for the modifications made to the artefact. In addition, the usability and reliability of the assessment criteria drawn up for the items were not

tested with the help of teachers. In particular, the criteria for constructed responses require a close examination of how uniformly teachers are able to interpret the scoring criteria.

All of these limitations arose from lack of time and resources. As mentioned earlier, the observational data for articles II and III relied solely on an observation sheet and field notes. The decision not to gather video data was made at the time with my colleague. While realizing that video recordings would have allowed us to re-visit and re-interpret the observed lessons, we also had reservations concerning the effect that recording would have had on the teachers. In the end, we opted for what we considered to be a less invasive data-gathering approach. The first two validity limitations are ones that can be addressed in a future study. Systematic research on Finnish primary school history textbooks, as well as the way pupils experience history lessons, are both interesting avenues for further research.

I have sought to be candid about the decisions I have made throughout the study, thereby allowing the reader to critically review this work. In addition to methodological decisions, I have tried to be transparent about the theoretical premise – the usefulness of historical literacy as an instructional aim. To this end, the study is not a descriptive one, but normative in the sense that by identifying what enables and disables historically literate instruction, it implies what should and should not be done if aiming for historical literacy. The distinction between impediments and potential is based on previous theoretical and empirical work, not on my personal opinions. It should be noted, however, that the study does not take a stand on what makes history teaching "good" or "bad" in general, and that the results are only concerned with one aspect of history teaching, namely historical literacy.

7.4 Why insist on historical literacy?

The results of the present study suggest that teaching things about the past takes precedence over teaching what history is about. It thus corroborates the view expressed in many previous studies that introducing disciplinary history into instruction is an ongoing but somewhat uphill task. Teachers´ practices and thinking have not yet fully shifted towards disciplinary history (see e.g., Cuban, 2016; Hartzler-Miller, 2001; James, 2008; Nokes, 2010; Stolare, 2017; Vesterinen, 2022) and valid assessment materials are only slowly being developed for classroom use (Breakstone, 2014; Seixas et al., 2015; Smith, 2018). The difficulties in implementing disciplinary approaches such as historical literacy, and the lasting appeal of memory-history, compel one to ask the question: If teachers seem inclined to teach ready-made narratives, why force the issue of historical literacy, for example? If history classrooms seem resistant to change, is advocating change necessary, reasonable, or even advisable?

A recent article provides an interesting perspective when it comes to answering the aforementioned questions. Wuokko and Räsänen (2023) studied the views of those teaching introductory courses in history departments in several Finnish universities. Their results show that university teachers regarded epistemological understanding and historical thinking as the most central components of expertise for prospective historians. The teachers felt that the secondary level focused too much on substantive knowledge at the expense of epistemological issues. Hence, the university teachers reported that when students began to study history at university level, it was their job to "destroy" or "blow up" students´ ideas of history, which relied heavily on the everyday conception of history as being synonymous with the past (p. 206). Yet the informants underlined that because of time constraints and how difficult it was to grasp the nature of history as a discipline, the first years of studying academic history could touch upon epistemological issues only tentatively. According to one informant, the aim was that "the pieces should start to fall into place during their master´s degree studies" (p. 204). Wuokko and Räsänen (2023) conclude their paper with a suggestion that because the epistemological aspect of history is difficult and unfamiliar to university students, studies may and should initially focus on other things, such as self-reflection, creating motivation, and practical experiences.

Wuokko and Räsänen's (2023) results and conclusions can be situated at the end of a kind of relay, where each educational level postpones disciplinary aspects and, in a manner of speaking, passes the buck to the next level. Many prospective primary teachers consider it premature to reflect on the construction of historical knowledge (James, 2008; Rantala & Khawaja, 2021), as do secondary teachers (Hartzler-Miller, 2001; VanSledright & Limón, 2006; Vesterinen, 2022). The teacher in Vesterinen´s (2022) study considered the disciplinary approach "too university-like" (p. 153). And as described, when the relay reaches the university level, there is hesitation concerning when and how to introduce epistemological aspects of history because students´ prior knowledge about the way historical knowledge is constructed is tentative at best.

This very relay of avoidance, or the mere possibility of postponement, makes the shift towards historical literacy or other disciplinary approaches challenging. However, if the pursuit of change is abandoned, the consequences extend far beyond whether or not history departments are provided with students who are historically literate. My argument for the redundancy of the relay of disciplinarity avoidance and for a change towards historical literacy is threefold. First, historical literacy helps learners to understand what history is about, which has value in and of itself without being instrumental to other goals. The disciplines and the knowledge they produce are inseparable (Schwab, 1978). If school subjects exclude the way knowledge is built in their respective disciplines, students will not be provided with the cognitive tools to evaluate and even question the knowledge they encounter both in and outside the classroom. Through historical

literacy, it is easier to justify why the school subject is called history rather than heritage studies or collective memory studies.

The second argument for promoting historical literacy despite apparent impediments concerns the individual level. The results presented in this study (Article II) show that pupils are not given the time to engage with texts other than those in the textbook, the content of which is to be decoded and adopted, not analysed. Even when primary or secondary sources were present in lessons for more than five minutes, it was typically the teacher, not the pupils, who engaged with the texts. This reduces pupils´ chances of connecting with texts, making meaning with them, and building agency (see Wineburg & Martin, 2005). While Ahonen (2020a) claims that interpreting texts runs the risk of losing what is meaningful in history, I ask the reverse: How can history become meaningful without the possibility of participating in the process of knowledge construction? Article I in this study, and especially its think-aloud interviews, suggest that with the right kind of scaffolding, text-related tasks encourage pupils to interact with the texts as well as ponder and reflect on their views and interpretations. Thus, I suggest that meaningfulness can be perceived through participation, agency in relation to texts, and the development of one´s thinking. Moreover, learning to interpret different types of sources does not exclude questions about connecting history to pupils at a personal or family level (cf. Ahonen, 2020b).

At the individual level, an emphasis on historical literacy also relates to accessibility and the right to knowledge (see Puustinen & Khawaja, 2021). As reported in article III, teachers may consider it irrelevant to share knowledge about the interpretational layers underlying historical narratives. This results in knowledge becoming a commodity of the few, knowledge of the powerful (Young, 1971). Awareness of the existence and nature of primary sources and everyone´s right to interpret them makes knowledge "everyone´s business". If university teachers have to "blow up" the common-sense way of perceiving history (Wuokko & Räsänen, 2023), it suggests that the majority of pupils do not acquire a disciplinary way of thinking about history during their basic education (see also Ouakrim-Soivio & Kuusela, 2011; van den Berg, 2011). This in turn narrows an individual´s role as an active citizen. Impediments to accessing knowledge and its construction have implications not only for individuals but for society at large.

I began this work with Bruner´s (1960) idea of young people´s "intellectual powers", the use of which he considered essential for maintaining democratic societies. Without citizens who are able to evaluate knowledge and find reliable information, democracies face multiple challenges (Breakstone et al., 2014; Hobbes, 2010). Although historical literacy is only one way to utilize these intellectual powers, its contribution at the societal level is essential. It provides an insight into how conflicting interpretations of the past are formed and used in the present. It makes explicit the multiperspectivity of history and thereby also the

multiperspectivity of the present, which is essential for maintaining democracies. Moreover, through historical literacy, detecting biases and the motives behind texts become a default way of approaching historical and societal texts. The current way of teaching in the United States has resulted in a majority not being able to differentiate between a news story and an advertisement, and not being able to detect the connection between the bias of a website and its content (Breakstone et al., 2021). If the aim is to raise students who build and renew democracy, these results are a cause for concern. People with limited resources for evaluating the trustworthiness of information are ill-equipped in the face of false claims, such as those used for instigating the attack on Capitol Hill in the US in 2021.

To conclude, my argument for the need to change history teaching is based on the needs of the school subject itself, as well as individual and societal needs. If such change is to take place, the primary level has a unique position; it has the possibility to ensure that the "relay of avoidance" is not put into motion in the first place, and that disciplinary history is not postponed, perhaps indefinitely. To expect this from primary teachers may seem unreasonable since they are less familiar with the disciplinary aspects of history than subject teachers are. However, I think a more disciplinary approach to history is feasible at the primary level for two reasons. First, the impediments identified in this study are not set in stone. They can be tackled and overcome. Second, I found teachers who introduced disciplinary aspects into their teaching, even if they had not yet achieved the ideals. In addition, I found potential for historical literacy in teachers´ reflections on the nature of historical knowledge and historical thinking. Thus, if the agents involved in the recontextualization process support teachers by providing materials and by increasing their existing knowledge, history teaching can be historically literate already at the primary school level.

References

Aashamar, P., & Klette, K. (2023). Powerful knowledge in the social studies classroom and beyond. *Journal of Curriculum Studies, 55*(4), 388–408, DOI:10.1080/00220272.2023.2234427

Aalto, I., & Kemppinen, L. (2020). Arkeologinen näkökulma alakoulun oppikirjojen kuvauksiin Suomen esihistoriasta. [An archaeological perspective on primary school textbooks´ descriptions of Finnish pre-history] *Kasvatus & Aika, 14*(2), 99–128.

Ahonen, S. (2017). Suomalaisuuden monet myytit: Kansallinen katse historian kirjoissa. [The many myths of being Finnish: The national gaze in history books.] Gaudeamus.

Ahonen, S. (2020a). Taaksepäin katsoen: puoli vuosisataa historianopetussotia. [Looking backwards: half a century of history education wars.] *Kasvatus & Aika, 14*(2), 129–140. https://doi.org/10.33350/ka.91260

Ahonen, S. (2020b). Form or substance? Weighing critical skills against identity narratives in history education. In C. Berg, & T. Christou (Eds.), *The Palgrave Handbook of History and Social Studies Education* (pp. 145–163). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-37210-1_7

Alvén, F. (2021). Pojkars och flickors resultat på det nationella provet i historia: Bestämmer uppgifters historiska innehåll och form pojkars respective flickors resultat? [Boys' and girls' performance in the national history test: Do the historical content and form of tasks determine boys' and girls' performance?] In D. Ludvigsson & L. Andersson Hult (Eds.), *Att bedöma i historieämnet.: Perspektiv på historisk kunskap, bedömning och prov* (pp. 205–226). Stockholm: Historielärarnas Förening.

Alvunger, D. (2018). Teachers' curriculum agency in teaching standards-based curriculum. *The Curriculum Journal, 29*(4), 479–98.

Ashby, R. (2011). Understanding historical evidence: Teaching and learning challenges. In I. Davies (Ed.), *Debates in history teaching* (pp. 137–47). London: Routledge.

Ashby, R. & Lee, P. (1987). Children´s concepts of empathy and understanding in history. In C. Portal (Ed.), *The History Curriculum for Teachers* (pp. 62–88). London: The Falmer Press.

Atjonen, P., Pöntinen, S., Kontkanen, S., & Ruotsalainen, P. (2022). In *Enhancing Preservice Teachers' Assessment Literacy: Focus on Knowledge Base, Conceptions of Assessment, and Teacher Learning*. *Frontiers in Education, 7*. doi: 10.3389/feduc.2022.891391

Bain, B. (2006). Rounding up unusual suspects: Facing the authority hidden in the history classroom. *Teachers College Record, 108*(10), 2080–2114. https://doi.org/10.1111/j.1467-9620.2006.0077.

Barab, S. (2014). Design-based research: A methodological toolkit for engineering change. In *The Cambridge handbook of the learning sciences* (2nd ed.), (pp. 151–170). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.011

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences*, *13*(1), 1–14.

Barton, K. (1997). "I just kinda know": Elementary students´ ideas about historical evidence. *Theory and Research in Social Education, 25*(4), 407-430.

Beck, I. & McKeown, M. (2002). Questioning the Author: Making sense of social studies. *Educational Leadership, 60*(4), 44–47.

Bennett, R. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25, DOI: 10.1080/0969594X.2010.513678

van den Berg, M. (2012). Vastaukset historian taitojen kuvastajina: laadullinen tutkimus historian tehtävistä. [Answers as a reflection of skills in history: a qualitative analysis]. In N. Ouakrim-Soivio & J. Kuusela. Historian ja yhteiskuntaopin oppimistulokset perusopetuksen päättövaiheessa 2011. Opetushallitus.

Bernstein, B. (1990). *The Structuring of pedagogical discourse*. London: Routledge.

Bernstein, B. (1999). Vertical and horizontal discourse: an essay. *British Journal of Sociology of Education, 20*(2), 157–173.

Bernstein, B. (2000). *Pedagogy, symbolic control and identity*. Maryland: Rowman and Littlefield.

Bertram, C. (2008). Exploring the concept of historical gaze. *Yesterday & Today, 3*, 1-9.

Bertram, C. (2019). What is powerful knowledge in school history? Learning from the South African and Rwandan school curriculum documents, *The Curriculum Journal*, 30(2). DOI: 10.1080/09585176.2018.1557536

Black, P. & Wiliam, D. (1998a). Assessment and classroom learning, Assessment in Education. *Principles, policy and practice, 5*(1), 7–74. DOI:10.1080/0969595980050102

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–44. https://doi.org/10.1177/003172171009200119

Booth, M. (1980). A modern world history course and the thinking of adolescent pupils. *Educational Review, 32*, 245–257.

Booth, M. (1987). Ages and concepts: A critique of the Piagetian approach to history teaching. In C. Portal (Ed.), *The history curriculum for teachers* (pp. 22–38). London: The Falmer Press.

Booth, M. (1994). Cognition in history: A British perspective. *Educational Psychologist, 29*(2), 61–69.

Booth, M., & Husbands, C. (1993). The history national curriculum in England and Wales: assessment at Key Stage 3. *The Curriculum Journal, 4*, 21–36.

Borg, W., & Gall, M. (1989). *Educational Research: An Introduction*. Fifth Edition. New York: Longman.

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151–167.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning, *Assessment & Evaluation in Higher Education, 31*(4), 399-413. DOI: 10.1080/02602930600679050

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

Breakstone, J. (2014). The process of designing new history assessments. *Theory and Research in Social Education, 42*(4), 453–458.

Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., & Saavedra, A. (2021). Students' Civic Online Reasoning: A national portrait. *Educational Researcher*. https://purl.stanford.edu/cz440cm8408

Brown, G. (2011). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters, 3*, 45–70.

Brown, R. (1996). Learning How to Learn: The Amherst project and history education in the schools, *The Social Studies, 87*(6), 267–273. DOI: 10.1080/00377996.1996.10114499

Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing student learning in higher education*. London: Routledge.

Brewer, J. (2000). *Ethnography*. Open University Press.

Bruner, J. (1960). *The Process of Education*. Cambridge: Harvard University Press, 1960.

Campanella, M., & Penuel, W. (2021). Design-based research in educational settings: motivations, crosscutting features and considerations for design. In Z. Philippakos, E. Howell, & A. Pellegrino *Design-based research in education: theory and applications.* The Guilford Press.

Ní Cassaithe, C., Waldron, F., & Dooley, T. (2022). "We can't really know cos we weren't really there": Identifying Irish primary children's bottleneck

beliefs about history. *Historical Encounters, 9*(1), 78-100. https://doi.org/10.52289/hej9.105

Chapman, A. (2021). Historical knowledge and the knowledge turn. In A. Chapman (Ed.), *Knowing History in Schools: Powerful knowledge and the power of knowledge (*pp. 1–31*)*. London: UCL Press.

Coffin, C. (1997). Constructing and giving value to the past. In J. R. Martin & F. Christie (Eds.), *Genres and Institutions: Social processes in the workplace and school* (pp. 196–230). Pinter.

Cope, B., & Kalantzis, M. (2000). *Multiliteracies: Literacy learning and the design of social futures*. Routledge.

Cronbach, L., (1971). Test validation. In R. Thorndike (Ed.), *Educational measurement (2nd ed.) (pp. 443-507)*. Washington, DC: American Council on Education.

Cuban, L. (2016). *Teaching history then and now: A story of stability and change in schools*. Cambridge, Massachusetts: Harvard Education Press.

Dansereau, D., & Gregg, L. (1966). An information processing analysis of mental multiplication. *Psychonomic Science, 6*(2), 71–72.

Dawson, I. (1989). The schools history project: A study in curriculum development. *The History Teacher, 22*(3), 221–238.

De La Paz, S., Monte-Sano, C., Felton, M., Croninger, R., Jackson, C. & Pientadosi Worland, K. (2016). A Historical writing apprenticeship for adolescents: Integrating disciplinary learning with cognitive strategies. *Reading Research Quarterly, 52*(1), 31-52.

Dempster, E. (2021). Power and control in science: A case study of a syllabus for science and technology. *British Journal of Sociology of Education, 42*(1), 113–26.

Deng, Z. (2004). The Fallacies of Jerome Bruner´s hypothesis in 'The Process of Education': A Deweyan perspective. *The Journal of Educational Thought, 38*(2), 151–170.

Deng, Z. (2022). Powerful knowledge, educational potential and knowledge-rich curriculum: Pushing the boundaries. *Journal of Curriculum Studies, 54*(5), 599-617. DOI: 10.1080/00220272.2022.2089538

Design-Based Research Collective (2003). Design-based research: An emerging paradigm for educational enquiry. *Educational Researcher 32*(1), 5–8. Retrieved from: https://search.proquest.com/docview/216901476?accountid=11365

DeWalt, K., & DeWalt, B. (2011). *Participant observation: a guide for fieldworkers* (2nd ed.). Rowman & Littlefield.

Downey, M., & Long, K. (2016). *Teaching for historical literacy: Building knowledge in the history classroom*. Routledge.

Duhaylongsod, L., Snow, C., Selman, R., & Donovan, M. (2015). Toward disciplinary literacy: Dilemmas and challenges in designing history

curriculum to support middle school students. *Harvard Educational Review, 85*(4), 587–608. doi:10.17763/0017- 8055.85.4.587

Duke, N. (2000). 3.6 minutes per day: the scarcity of informational text in first grade. *Reading Research Quaterly, 35*(2), 202–224.

van Drie, J. & van Boxtel, C. (2008). Historical reasoning: Towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review, 20(*2), 87–110. DOI 10.1007/s10648-007-9056-1

Eliasson, P. (2013). Historiemedvetande pÍ prov – nationella ämnesprov i historia i Sverige [Historical awareness put to the test – national subject tests in history in Sweden]. *Kasvatus & Aika, 7*(3). Retrieved from https://journal.fi/kasvatusjaaika/article/view/68651

Eliasson, P., Alvén, F., Axelsson Yngvéus, C., & Rosenlund, D. (2015). Historical consciousness and historical thinking reflected in large-scale assessment in Sweden. In K. Ercikan and P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 171–182). New York: Routledge.

Ercikan, K., Seixas, P., Lyons-Thomas, J., & Gibson, L. (2015). Cognitive validity evidence for validating assessments of historical thinking. In K. Ercikan and P. Seixas (Eds.) *New directions in assessing historical thinking* (pp. 206–220). New York: Routledge.

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

Evans, R., & Midford, S. (2021). Teaching historical literacies to digital learners via popular culture. *Arts and Humanities in Higher Education, 21*(3), 285–301.

Faggella-Luby, M., & Deshler, D. (2008). Reading comprehension in adolescents with LD: What we know; what we need to learn. *Learning Disabilities Research & Practice, 23*(2), 70–78.

Faggella-Luby, M., Graner, P., Deshler, D., & Drew, S. (2012). Building a house on sand: Why disciplinary literacy is not sufficient to replace general strategies for adolescent learners who struggle. *Topics in Language Disorders, 32*(1), 69–84. https://doi.org/10.1097/TLD.0b013e318245618e

Fang, Z. (2012). Language correlates of disciplinary literacy. *Topics in Language Disorders, 32*(1), 19–30. https://doi.org/10.1097/TLD.0b013e31824501de

Fang, Z., & Schleppegrell, M. (2008). *Reading in secondary content areas*. Ann Arbor: University of Michigan Press.

Fillpot, E. (2012). Historical thinking in third grade. *The Social Studies, 103*(5), 206–217.

Finnish Government (2023). A strong and committed Finland: Programme of Prime Minister Petteri Orpo's government 20 June 2023. Accessed 1

August 2023. https://valtioneuvosto.fi/en/governments/government-programme#/

Finnish National Board of Education (2004). National Core Curriculum for Basic Education. Accessed February 3, 2022. https://www.oph.fi/sites/default/files/documents/perusopetuksen-opetussuunnitelman-perusteet_2004.pdf

Finnish National Board of Education (2014). National Core Curriculum for Basic Education 2014. Helsinki: Finnish National Agency for Education. Accessed 15 March 2022. https://www.oph.fi/en/statistics-and-publications/publications/new-national-core-curriculum-basic-education-focus-school

Finnish National Board of Education (2017). Opettajat ja rehtorit Suomessa 2016. *Raportit ja sevitykset 2017*(2) [Teachers and principals in Finland in 2016].

Fogo, B. (2014). Core practices for teaching history: The results of a Delphi panel survey. *Theory & Research in Social Education, 42*(2), 151–196.

Fordham, M. (2016). Teachers and the academic disciplines. *Journal of Philosophy of Education 50*(3), 419–431.

Foster, S., & Yeager, E. (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*(4), 286–317. Retrieved from https://search.proquest.com/docview/196371949?accountid=11365

Friend, R. (2001). Teaching and summarization as a content area reading strategy. *Journal of Adolescent & Adult Literacy, 44*(4), 320–329.

Gardner, H. (1999). *The Disciplined mind.* Penguin Books.

Gestsdóttir, S., van Drie, J., van Boxtel, C. (2019). Teaching historical thinking and reasoning in upper secondary schools in Iceland: results of an observation study. *Nordidactica – Journal of Humanities and Social Science Education, 2,* 90–113.

Gläser-Zikuda, M., Hagenauer, G., & Stephan, M. (2020). The potential of qualitative content analysis for empirical educational research. Forum: *Qualitative Social Research, 21*(1), 17.

Graham, S., & Hebert, M.A. (2010*). Writing to read: Evidence for how writing can improve reading*. A Carnegie Corporation Time to Act Report. Washington, DC: Alliance for Excellent Education.

Grecic, D. (2022). The epistemological chain: A tool to guide TNE development. *Journal of Studies in International Education* https://doi.org/10.1177/10283153221145078

Halliday, M. (2004). *The language of science* (Vol. 5 in The collected works of M. A. K. Halliday, J. Webster Ed.). London: Continuum

Hammond, L. (2022). Recognizing and exploring children´s geographies in school geography. *Children´s Geography, 20*(1), 64–78.

Hartzler-Miller, C. (2001). Making sense of "best practice" in teaching history. *Theory and Research in Social Education, 29*(4), 672–695. https://doi.org/10.1080/00933104.2001.10505961

Hayes, J., & Flowel, L. (1980). Identifying the organization of writing processes. In L. Gregg & E. Steinberg (Eds.). *Cognitive processes in writing* (pp. 3–30). Hillsdale, NJ: Lawrence Erlbaum.

Heinonen, J.-P. (2005). Opetussuunnitelmat vai oppimateriaalit. Peruskoulun opettajien käsityksiä opetussuunnitelmien ja oppimateriaalien merkityksestä opetuksessa [Curricula or learning materials. Primary school teachers' perceptions of the role of curricula and learning materials in teaching]. Soveltavan kasvatustieteen laitos. Tutkimuksia 257. Helsingin yliopisto.

Heikkinen, H., Aho, J., & Korhonen, H. (2015). Ope ei saa oppia. Opettajankoulutuksen jatkumon kehittäminen. [The teacher should not learn. Developing continuity in teacher education]. Jyväskylä: Koulutuksen tutkimuslaitos.

Heller, R. (2010). In praise of amateurism: A friendly critique of Moje's "Call for change" in secondary literacy. *Journal of Adolescent & Adult Literacy, 54*(4), 267–273. https://doi.org/10.1598/JAAL.54.4.4

Herriot, R., & Firestone, W. (1983). Multisite qualitative policy research: Optimizing description and generalizability. *Educational Researcher, 12*(2), 14–19.

Hinchman, K., & O´Brien, D. (2019). Disciplinary literacy: From infusion to hybridity. *Journal of Literacy Research, 51*(4) 525–536.

Hirsch, E. (1988). *Cultural literacy: What every American needs to know*. Vintage Books.

Hobbs, R. (2010). Digital and media literacy: A plan of action. The Aspen Institute. https://assets.aspeninstitute.org/content/uploads/2010/11/Digital_and_Media_Literacy.pdf

Hordern, J. (2021). Recontextualisation and the teaching of subjects. *The Curriculum Journal, 32(*4), 592–606.

Howard, C., Adams- Budde, M., Lambert, C. & Myers, J. (2021). Engaging literacy experiences in history classrooms: A multiple case study of novice teachers´ beliefs and practices. *Literacy Research Instruction, 60*(1), 36–55.

Howell, E., Barlow, W., & Dyches, J. (2021). Disciplinary literacy: success and challenges of professional development. *Journal of Language and Literacy Education, 17*(1), 1–26.

Howard, C., Adams-Budde, M., Lambert, C., & Myers, J. (2021). Engaging literacy experiences in history classrooms: A multiple case study of novice teachers' beliefs and practices. *Literacy Research and*

Instruction, 60(1), 36-55. https://doi.org/10.1080/19388071.2020.1779878

Huijgen, T., Holthuis, P., van Boxtel, C., & van de Grift, W. (2019). Promoting historical contextualisation in classrooms: an observational study. *Educational Studies, 45*(4), 456–479. DOI: 10.1080/03055698.2018.1509771

Hughes, R. (2021a). Apprenticing third graders in disciplinary literacy in history. *Literacy Research and Instruction, 60*(2), 127–151. https://doi.org/10.1080/19388071.2020.1789249

Hughes, R. (2021b). "What is slavery?": Third-grade students' sensemaking about enslavement through historical inquiry, *Theory & Research in Social Education, 50*(1),1–45. DOI: 10.1080/00933104.2021.1927921

Husbands, C: (1996). *What Is History Teaching?* Open University Press.

Hynd, C., Holschuh, J. P., & Hubbard, B. P. (2004). Thinking like a historian: College students' reading of multiple historical documents. *Journal of Literacy Research, 36*(2), 141–176. https://doi.org/10.1207/s15548430jlr3602_2

James, J. (2008). Teachers as protectors: Making sense of preservice teachers' resistance to interpretation in primary history teaching. *Theory and Research in Social Education, 36*(3), 172–205.

Johansson, P. (2019). Historical enquiry with archaeological artefacts in primary school. *Nordidactica – Journal of Humanities and Social Science Education, 2019*(1), 77–104.

Juuti, K., & Lavonen, J. (2006). Design-based research in science education: one step towards methodology. *Nordic Studies in Science Education, 4*, 54–68. doi.org/10.5617/nordina.424

Juvonen, R. & Routarinne, S. (2018). Toimintaa lukemisen ja kirjoittamisen rajapinnalla: neljäsluokkalaisten luetun ymmärtämisen tehtävä ympäristöopin tunnilla. [Activities at the interface between reading and writing: a reading comprehension task for fourth-graders in an environmental studies lesson]. In L. Lehti, P. Peltonen, R. Routarinne, V. Vaakanainen, & V. Virsu (Eds.), *Uusia lukutaitoja rakentamassa – Building New Literacies* (pp. 134–158). AnFiLan vuosikirja. Jyväskylä.

Kaliski, P., Smith, K. & Huff, H. (2015). The importance of construct validity evidence in history assessment: what is often overlooked or misunderstood? In K. Ercikan and P. Seixas (Eds.), *New directions in assessing historical thinking*. (pp. 195–205). New York: Routledge.

Kane, M. (2001). Current concerns in validity theory. *Journal of Educational measurement,* 38, 319–342.

Kane, M. T. (2006). Validation. In R. Brennan (Ed.). *Educational Measurement (4th edition)* (pp. 17–64). Westport, CT: American Council on Education and Praeger Publishers.

Kansanen, P. (2014). Teaching as master's level profession in Finland: Theoretical reflections and practical solutions. In O. McNamara, J. Murray, M. Jones, M. (Eds.), *Workplace learning in teacher education: International practice and policy*. (pp. 279–292). Springer.

Kela, M. (2012). Suunvuoro. [A comment]. *Virittäjä, 4*. Retrieved from https://journal.fi/virittaja/issue/view/1036

Kellogg, R. (1999). Components of working memory in text production. In M. Torrance & G. Jeffery (Eds.), *The cognitive demands of writing: Processing capacity and working memory in text production* (pp. 42–61). Amsterdam: Amsterdam University Press.

Khawaja, A. & Puustinen, M (2022). The nature of literacy instruction in elementary school history lessons. *Language and Literacy, 24*(3), 20–44.

Kitson, A. (2020). *Teachers as recontextualization agents: A study of expert teachers' knowledge and their role in recontextualization process across different subjects*. PhD thesis, UCL, London, UK.

Kitson, A. (2021). How useful is the theory of powerful knowledge to history educators?'. In A. Chapman (Ed.), *Knowing History in Schools: Powerful knowledge and the power of knowledge* (pp. 31–52*)*. London: UCL Press.

Klerides, E. (2010). Imagining the Textbook: Textbooks as discourse and genre. *Journal of Educational Media, Memory, and Society, 2*(1), 31–54.

Kon, J. H. (1995). Teachers' curricular decision making in response to a new social studies textbook. *Theory and Research in Social Education, 23*(2), 121–146. https://doi.org/10.1080/00933104.1995.10505748

Koppinen, M-L., Korpinen, E. & Pollari, J. (1994). *Mitä arviointi on?* [What is assessment?] Juva: WSOY.

Koski, A., Puustinen M., Säntti J. & Salminen, J. (2023). Exploring the research-based approach of academic classroom teachers in Finland. *Teachers and Teaching*. DOI: 10.1080/13540602.2023.2208034

Kouki, E. & Virta, A. (2017). Towards critical literacy: Students´ reading skills and source evaluation. In N. Pyyry, L. Tainio, K. Juuti, R. Vasquez, & M. Paananen (Eds.), *Changing subjects, changing pedagogies: Diversities in school and education* (pp. 34–51). Suomen ainedidaktinen tutkimusseura.

Kölbl, C., & Konrad, L. (2015). Historical consciousness in Germany: Concept, implementation, assessment. In K. Ercikan and P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 17–28). New York: Routledge.

Lee, P. (2005). Putting principles into practice: understanding history. In S. Donovan & J. Bransford (Eds.*), How Students Learn: History in the*

Classroom (pp. 31–78). Washington, DC: The National Academies Press. https://doi.org/10.17226/11100.

Lee, P. (2007). From national canon to historical literacy. In M. Grever & S. Stuurman (Eds.), *Beyond the canon: History for the Twenty-First Century* (pp. 48–62). Basingstoke: Palgrave Macmillan.

Lee, P. (2011). History education and historical literacy. In I. Davies (Ed.), *Debates in History Teaching* (pp. 63–72*)*. London: Routledge.

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

Levstik, L., Henderson, G., & Lee, Y. (2014). The beauty of other lives: Material culture as evidence of human ingenuity and agency. *The Social Studies, 105*(4), 184-192. https://doi.org/10.1080/00377996.2014.886987

Logtenberg, A., van Boxtel, C., & van Hout-Wolters, B. (2011). Stimulating situational interest and student questioning through three types of historical introductory texts. *European Journal of Psychology of Education, 26*(2), 179–198.

Lowenthal, D. (2015). *The Past Is a Foreign Country - Revisited*. Revised and updated edition. Cambridge: Cambridge University Press.

Luis, R., & Rapanta, C. (2020). Towards (re-)defining historical reasoning competence: A review of theoretical and empirical research. *Educational Research Review, 31*. https://doi.org/10.1016/j.edurev.2020.100336

Lyhty, J. (2020). Kuinka arvioida asetettuja tavoitteita: alakoulun historian opetussuunnitelma ja opettajanoppaat. [How to assess the stated aims? The primary level curriculum and teacher´s guides] *Kasvatus & aika, 14*(4), 21-38. https://doi.org/10.33350/ka.91153

Löfström, J. (2019). Yhteiskuntaoppi. Social studies in Finland. A country report. *Journal of Social Science Education 18*(4). 103–116. DOI 10.4119/jsse-1583

Maposa, M., & Wasserman, J. (2009). Conceptualising historical literacy: Review of the literature. *Yesterday & Today, 4*, 41–66.

Martin, D., & Wineburg, S. (2008). Seeing thinking on the web. *The History Teacher, 41*(3), 305–319.

Masuda, A. (2014). Exploring preservice teachers' emerging understandings of disciplinary literacy. *Literacy Research and Instruction, 53*(4), 326-346. DOI:10.1080/19388071.2014.927940

Maton, K., & Moore, R. (2009). Introduction: coalitions of the mind. In K. Maton & R. Moore (Eds.), *Social realism, knowledge and the sociology of education* (pp. 1–13). London: Continuum.

McKenna, M., & Robinson, R. (2014). *Teaching through text: Reading and writing in the content areas* (2nd ed.). Boston, MA: Pearson.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35*(11), 1012–1027.

Messick, S. (1989). Validity. In R. Linn (Ed.), Educational measurement (3rd ed.) (pp. 13-103). New York: American Council on Education and Macmillan.

Mikander, P. (2016). *Westerners and others in Finnish school textbooks*. Helsinki: University of Helsinki.

Moje, E. (2008). Foregrounding the disciplines in secondary literacy teaching and learning: A call for change. *Journal of Adolescent & Adult Literacy, 52*(2), 96– 107. https://doi.org/10.1598/JAAL.52.2.1

Moje, E. (2011). Response: Heller's "In praise of amateurism: A friendly critique of Moje's 'Call for change' in secondary literacy". *Journal of Adolescent & Adult Literacy, 54*(4), 275–278. https://doi.org/10.1598/JAAL.54.4.5

Moje, E. (2015). Doing and teaching disciplinary literacy with adolescent learners: A social and cultural enterprise. *Harvard Educational Review, 85*(2), 254–278. https://doi.org/10.17763/0017-8055.85.2.254

Monte-Sano, C. (2011). Beyond reading comprehension and summary: learning to read and write in history by focusing on evidence, perspective, and interpretation. *Curriculum Inquiry, 41*(2), 212–249. DOI: 10.1111/j.1467-873X.2011.00547.x

Monte-Sano, C., & Budano. C. (2013). Developing and enacting pedagogical content knowledge for teaching history: An exploration of two novice teachers' growth over three years. *Journal of the Learning Sciences, 22*(2), 171–211. DOI: 10.1080/10508406.2012.742016

Monte-Sano, C., & Reisman, A. (2016). Studying historical understanding. In L. Corno & E. Anderman (Eds.), *The Handbook of Educational Psychology, 3rd edition*. Washington, D.C.: American Psychological Association.

Moore, R., & Muller, J. (1999). The Discourse of 'Voice' and the Problem of Knowledge and Identity in the Sociology of Education. *British Journal of Sociology of Education, 20*(2), 189–206. DOI: 10.1080/01425699995407

Mård, N. (2020). History in multidisciplinary education: A case study in a Finnish primary school. *Education 3-13*, 1–16. https://doi.org/10.1080/03004279.2020.1737172

Muñoz, N., & Balmaceda, C. (2022). Between school and ethical–political everyday action: a comprehensive framework of the development of historical thinking, *Journal of Curriculum Studies 54*(4), 1–21 DOI: 10.1080/00220272.2021.2018502

Neugebauer, S., & Blair, E. (2020). "I know how to read and all…but". Disciplinary reading construction of middle school students of color. *Journal of Literacy Research, 52*(3), 316–340.

Van Nieuwenhuyse, K. (2016). Is seeing believing? On the educational use of mainstream historical films in the history classroom. *International*

Journal of Research on History Didactics, History Education, and Historical Culture, 37 (1), 191–212.

Van Nieuwnhuyse, K. (2018). Towards a postcolonial mindset in a postcolonial world? Evolving representations of modern imperialism in Belgian history textbooks (1945–2017). In K. Van Nieuwnhuyse and J. Pires Valentim (Eds.), *The colonial past in history textbooks: historical and social psychological perspectives* (pp. 155–176). Information Age Publishing, Inc.

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

Nokes, J. (2010). Observing literacy practices in history classrooms. *Theory and Research in Social Education, 38*(4), 515–544. https://doi.org/10.1080/00933104.2010.10473438

Nokes, J. (2012). *Building students´ historical literacies. Learning to read and reason with historical texts and evidence*. New York: Routledge.

Nokes, J. (2014). Elementary students' roles and epistemic stances during document-based history lessons. *Theory and Research in Social Education, 42*(3), 375–413. https://doi.org/10.1080/00933104.2014.937546

Nokes, J. (2017). Exploring patterns of historical thinking through eighth-grade students' argumentative writing. *Journal of Writing Research, 8*(3), 437–467.

Nokes, J., & De La Paz, S. (2018). Writing and argumentation in history education. In S. Metzger & L. McArthur Harris (Eds.), *The Wiley International Handbook of History Teaching and Learning*. Wiley.

Nordgren, K. (2021). Powerful knowledge for what? History education and 45-degree discourse. In A. Chapman (Ed.), *Knowing History in Schools: Powerful knowledge and the power of knowledge*. London: UCL Press.

Nordgren, K., & Johansson, M. (2015). Intercultural historical learning: A conceptual framework. *Journal of Curriculum Studies, 47*(1), 1–25.

Norppa, J. (2019). Historian oppikirjojen ja opetussuunnitelmien tavoitteet ristiriidassa? [Are the aims of history textbooks and curricula in contradiction with one another?]. *Koulu ja menneisyys, 57*, 33–53.

Ormond, B. (2017). Curriculum decisions: The challenges of teacher autonomy over knowledge selection for history. *Journal of Curriculum Studies, 49*(5), 599–619, DOI: 10.1080/00220272.2016.1149225

Ouakrim-Soivio, N. (2013). *Toimivatko päättöarvioinnin kriteerit? Oppilaiden saamat arvosanat ja Opetushallituksen oppimistulosten seuranta-arviointi koulujen välisten osaamiserojen mittareina*. [Do the

assessment criteria work as they should?] Raportit ja selvitykset 2013(9). Opetushallitus.

Ouakrim-Soivio, N. (2015). Oppimisen ja osaamisen arviointi. [Formative and summative assessment]. Helsinki: Otava.

Ouakrim-Soivio, N., & Kuusela, J. (2012). *Historian ja yhteiskuntaopin oppimistulokset perusopetuksen päättövaiheessa 2011*. [The learning outcomes of history and social studies in the final stage of basic education 2011]. Helsinki: Opetushallitus.

Owings Swan, K., Hofer, M., & Locascio, D. (2008). Historical scene investigation (HSI) project: Examining the use of case based historical instruction in the fifth grade social studies classroom. *International Journal of Social Education, 22*(2), 70–100.

Parkes, R., & Donnelly, D. (2014). Changing conceptions of historical thinking in history education: An Australian case study. *Revista Tempo e Argumento, Florian polis, 6*(11), 113–136.

Paxton, R. (2002). The influence of author visibility on high school students solving a historical problem. *Cognition and Instruction, 20*(2), 197–248.

Peck, C., & Seixas, P. (2008). Benchmarks of historical thinking: First steps. *Canadian Journal of Education*, 31(4), 1015–1038.

Piaget, J. (1988). Lapsi maailmansa rakentajana. Kuusi esseetä lapsen kehityksestä [The child as a builder of his or her world. Six essays on child development]. Porvoo: WSOY

Pilli, A. (1988). *Historiatiedon ja historian ymmärtämisen luonne ja lapsen tapa ymmärtää historiaa*. [Understanding historical knowledge and a child´s way of understanding history]. Turku: Turun yliopisto.

Pole, C. & Morrison, M. (2003). *Ethnography for Education*. Buckingham: Open University Press. *Canadian Journal of Education, 31*(4), 1015–1038.

Portal, C. (1987). Empathy as an objective for history teaching. In C. Portal (Ed.), The History Curriculum for Teachers (pp. 89–102). London: The Falmer Press.

Priestley, M., Biesta, G., & Robinson, S. (2017). *Teacher agency: An ecological approach*. London: Bloomsbury Academic.

Puustinen, M. (2022). Teacher education and history teachers' powerful professional knowledge. In B. Hudson, N. Gericke, C. Olin-Scheller, & M. Stolare (Eds.), *International perspectives on knowledge and quality: Implications for Innovation in Teacher Education Policy and Practice*. Bloomsbury Academic.

Puustinen, M., & Khawaja, A. (2021). Envisaging the alternatives: From knowledge of the powerful to powerful knowledge in history classrooms. *Journal of Curriculum Studies, 53*(1), 16-31. https://doi.org/10.1080/00220272.2019.1703273

Puustinen, M., & Khawaja, A. (2022). Ambitious aims, traditional reality? Observing historical literacy in Finnish upper secondary schools. *Nordidactica– Journal of Humanities and Social Science Education*, 3, 53–76. https://journals.lub.lu.se/nordidactica/article/view/23810

Raatikainen, P. (2004). *Ihmistieteet ja filosofia*. [Humanities and philosophy]. Helsinki: Gaudeamus.

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*(1), 58-76. http://kau.diva-portal.org/smash/record.jsf?pid=diva2:416331 >

Rantala, J. (2017). Oppikirjoja sadan vuoden ajalta: luonteenkasvatuksesta kriittisyyteen. [Half a century of textbooks: from character building to a critical stance] In P. Hiidenmaa, M. Löytönen, & H. Ruuska (Eds.), *Oppikirja Suomea rakentamassa* (pp. 247–281). Suomen tietokirjailijat.

Rantala, J. (forthcoming). Tuntematon sotilas -elokuvat kollektiivisen muistin vahvistajana ja proteettisen muistin tuottajana. [The Unknown Soldier films as a building block for a collective memory].

Rantala, J., & Ahonen, S. (2015). *Ajan merkit. Historian käyttö ja opetus*. [Signs of time. The use and teaching of history.] Helsinki: Gaudeamus.

Rantala, J., & van den Berg, M. (2013). Lukiolaisten historian tekstitaidot arvioitavina. [Assessing historical literacy among Finnish high-school students] *Kasvatus, 44*(4), 394–407.

Rantala, J., & Khawaja, A. (2016). Historian tekstitaitojen arvioiminen peruskoulun kuudesluokkalaisilla [Assessing historical literacy among sixth-graders] (pp. 77–93). In H-M. Pakula, E. Kouki, H. Silfverberg, & E. Yli-Panula (Eds.), *Uudistuva ja uusiutuva ainedidaktiikka*..

Rantala, J., & Khawaja, A. (2021). Prospective primary school teachers' confidence in teaching disciplinary history. *Teaching and Teacher Education, 107*. https://doi.org/10.1016/j.tate.2021.103492

Rantala, J., Oukrim-Soivio, N. (2020). Historianopetuksen muutoksesta ja pysyvyydestä [On the change and continuity of history teaching]. *Kasvatus & Aika, 14*(2), 5–8. https://doi.org / 10.33350/ka 91311

Rantala, J., Puustinen, M., Khawaja, A., van den Berg, M. & Ouakrim-Soivio, N. (2020). *Näinkö historiaa opitaan*? [Is this how history is learnt?]. Helsinki: Gaudeamus.

Rantanen, P. (2003). *Enemmän vähemmällä. Monivalintatehtävien mittaustarkkuuden nostaminen: Kasvatusalan tutkimuksia*. [More with less: increasing the reliability of multiple-choice items]. Turku: Suomen kasvatustieteellinen seura.

Reich, G. (2009). Testing historical knowledge: Standards, multiple choice questions and student reasoning, *Theory & Research in Social Education, 37(*3), 325–360. DOI: 10.1080/00933104.2009.10473401

Reich, G. (2015). Measuring up? Multiple-choice questions. In K. Ercikan and P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 221–232). New York: Routledge.

Reisman, A. (2012a). The 'Document-Based Lesson': Bringing disciplinary inquiry into high school history classrooms with adolescent struggling readers. *Journal of Curriculum Studies, 44*(2), 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. (2015a). Entering the historical problem space: Whole-class text-based discussion in history class. *Teachers College Record, 117*(2), 1–44. https://doi.org/10.1177/016146811511700206

Reisman, A. (2015b). The difficulty of assessing disciplinary historical reading. In K. Ercikan & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 29–39). New York: Routledge.

Reisman, A., & McGrew, S. (2018). Reading in history education: text, sources and evidence. In S. Metzger & L. McArthur Harris (Eds.), *The Wiley International Handbook of History Teaching and Learning, 1st edition*. John Wiley & Sons, Inc.

Rosenlund, D. (2015). Source criticism in the classroom: An empiricist straitjacket on pupils´ historical thinking? *Historical Encounters: A journal of historical consciousness, historical cultures, and history education, 2*(1), 47–57. https://doi.org/10.52289/hej2.100

Rosenlund, D. (2016). *History education as content, methods or orientation? A study of curriculum prescriptions, teacher-made tasks and student strategies*. Frankfurt am Main: Routledge.

Rouet, J.-F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology, 88*(3), 478–493. https://doi.org/10.1037/0022-0663.88.3.478

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences, 23(*1), 18-36. DOI: 10.1080/10508406.2013.778204

Saye, J., Stoddard, J., Gerwin, D., Libresco, A., & Maddox, L. (2018). Authentic pedagogy: examining intellectual challenge in social studies classrooms. *Journal of Curriculum Studies, 50*(6), 865–884.DOI: 10.1080/00220272.2018.1473496

Schleppegrell, M. (2004). *The language of schooling: A functional linguistics perspective*. Lawrence Erlbaum Associates.

Schleppegrell, M. (2016). Content-based language teaching with functional grammar in the elementary school. *Language Teaching, 49*(1), 116–128. doi:10.1017/S0261444814000093.

Schleppegrell, M., Achugar, M., & Oteiza, T. (2004). The grammar of history: Enhancing content-based instruction through a functional focus on language. *Tesol Quarterly, 38*(1), 67–93.

Schwab, J. (1978). Education and the structure of disciplines. In I. Westbury & N.J. Wilkof (Eds.), *Science, curriculum and liberal education*. University of Chicago Press.

Seixas, P. (1996). Conceptualizing the growth of historical understanding. In D. Olson and N. Torrance (Eds.), *Handbook of education and human development* (pp. 765–83). Oxford: Blackwell.

Seixas, P. (2006). *Benchmarks of historical thinking: A framework for assessment in Canada*. Centre for the study of historical consciousness. http://archive.historybenchmarks.ca/documents/benchmarkshistoricalthinkingframework-assessment-canada

Seixas, P. (2016a). Translation and its discontents: Key concepts in English and German history education. *Journal of Curriculum Studies, 48*(4), 427–439.

Seixas, P. (2016b). Are heritage education and critical historical thinking compatible? Reflections on historical consciousness from Canada. In C. Boxtel, M. Grever & S. Klein (Eds.), *Sensitive pasts: Questioning heritage in education* (pp. 40–58). Berghahn.

Seixas, P. (2017). A model of historical thinking. *Educational Philosophy and Theory, 49*(6), 593–605.

Seixas, P., Gibson, L., & Ercikan, K. (2015). A design process for assessing historical thinking: the case of a one-hour test. In K. Ercikan and P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 102–116). New York: Routledge.

Seixas, P., Stearns, P. & Wineburg, S. (2000). *Knowing, teaching and learning history: National and international perspectives*. New York: New York University Press.

Severance S., Penuel W., Sumner, T. & Leary H. (2016). Organizing for teacher agency in curricular co-design. *Journal of the Learning Sciences, 25*(4), 531–564. DOI: 10.1080/10508406.2016.1207541

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review, 78*(1), 40–59. https://doi.org/10.17763/haer.78.1.v62444321p602101

Shanahan, T., & Shanahan, C. (2012). What is disciplinary literacy and why does it matter? *Topics in Language Disorders, 32*(1), 7–18. https://doi.org/10.1097/TLD.0b013e318244557a

Shemilt, D. (1984). Beauty and the philosopher: Empathy in the classroom. In A. Dickinson, P. Lee, P. Rogers (Eds.), *Learning history*. London: Heinemann.

Shemilt, D. (1987). Adolescent ideas about evidence and methodology in history. In C. Portal (Ed.), *The History Curriculum for Teachers* (pp.39–61). London: The Falmer Press.

Simola, H., Rinne, R., Varjo, J., Kauko, J., & Pitkänen, H. (2009). Quality Assurance and Evaluation (QAE) in Finnish comprehensive schooling – a national model or just unintended effects of radical decentralisation? *Journal of Education Policy, 24*(2), 163–178. https://doi.org/10.1080/02680930902733139

Simons, H. (2009). *Case study research in practice*. Los Angeles: Sage.

Smith, J. (2017). Discursive dancing: Traditionalism and social realism in the 2013 English history curriculum wars. *British Journal of Educational Studies, 65*(3), 307–329. DOI: 10.1080/00071005.2017.1279274

Smith, M. (2018). New multiple-choice measures of historical thinking: An investigation of cognitive validity. *Theory & Research in Social Education, 46*(1), 1–34. DOI: 10.1080/00933104.2017.1351412

Smith, M., Breakstone, J. & Wineburg, S. (2019). History assessments of thinking: A validity study. *Cognition and Instruction*,37(1), 118–144. DOI: 10.1080/07370008.2018.1499646

Spradley, J. (1980). *Participant observation*. Holt, Rinehart and Winston.

Stearns, P., Seixas, P., & Wineburg, S. (2000). *Knowing, teaching, and learning history: national and international perspectives*. New York: New York University Press.

Stengel, B. (1997). 'Academic discipline' and 'school subject': Contestable curricular concepts. *Journal of Curriculum Studies, 29*(5), 585-602. DOI: 10.1080/002202797183928

Stokes. (1997). *Pasteur's quadrant: basic science and technological innovation*. Brookings Institution.

Stolare, M. (2017). Did the Vikings really have helmets with horns? Sources and narrative content in Swedish upper primary school history teaching. *Education 3–13, 45*(1), 36–50. https://doi.org/10.1080/03004279.2015.1033439

Stolare, M., Ludvigsson, D., & Trenter, C. (2021). The educational power of heritage sites. *History Education Research Journal, 18*(2), 264–79. https://doi.org/10.14324/

Sulkunen, S., Luukka, M-R. (2014). Monilukutaito ja tiedonalakohtaiset tekstitaidot. [Multiliteracy and disciplinary literacies]. *Kielikukko 4*/2014, 2–7.

Sulkunen, S., Luukka, M-R., Saario J., & Veistämö, T. (2019). Monilukutaito lukion historian opetuksessa. [Multiliteracy in secondary history teaching]. *Ainedidaktiikka, 3*(2), 2–23.

Sunal, C. & Sunal, D. (2008). Reports from the field: Elementary teacher candidates describe the teaching of social studies. *International Journal of Social Education, 22*(2), 29–48.

Säntti, J., Puustinen, M., & Hansen, P. (2023). Unmentioned challenges of Finnish teacher education: Decontextualisation, scientification and the rhetoric of the research-based agenda. In M. Thrupp, P. Seppänen, J. Kauko & S. Kosunen (Eds.), *Finland's Famous Education System: Unvarnished Insights into Finnish Schooling* (pp. 105–119*)*. Springer. https://link.springer.com/chapter/10.1007/978-981-19-8241-5_7

Tallavaara, R., & Rautiainen, M. (2020). What is important in history teaching? Student class teachers' conceptions. *History Education Research Journal, 17* (2), 229–42. https://doi.org/10.14324/HERJ.17.2.07

Tanner, D., & Tanner, L. (1980). *Curriculum development: theory into practice (2nd Ed.)*. New York: Macmillan Publishing Co., Inc.

Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466–478.

Thrupp, M., Seppänen, P., Kauko, J., & Kosunen, S. (2023). Introduction. In M. Thrupp, P. Seppänen, J. Kauko & S. Kosunen (Eds.), *Finland´s famous educational system. Unvarnished Insights into Finnish Schooling*. Springer. https://doi.org/10.1007/97

Torsti, P. (2012). *Suomalaiset ja historia.* [Finns and history.] Helsinki: Gaudeamus.

Tosh, J. (2015). *The pursuit of history aims, methods and new directions in the study of history (Sixth edition.*). Routledge/Taylor & Francis Group.

VanSledright, B. (1996). Closing the gap between school and disciplinary history? Historian as high school history teacher. In J. Brophy (Ed.), *Advances in research on teaching: Teaching and learning history,* (pp. 257–289). Greenwich, Connecticut: Jai Press.

VanSledright, B. (2002). *In search of Americas past*. New York: Teachers College Press.

VanSledright, B. (2014). *Assessing historical thinking and understanding*. New York & London: Routledge.

VanSledright, B. (2015). Assessing for learning in the history classroom. In K. Ercikan and P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 75–88). New York: Routledge.

VanSledright, B. (2016). The continuous threat of excess? A cautionary tale about heritage celebration and object veneration in the U.S.A. In C. Boxtel, M. Grever & S. Klein (Eds.), *Sensitive pasts: Questioning heritage in education* (pp. 40–58). Berghahn.

VanSledright, B. & Frankes, L. (2000). Concept- and strategic-knowledge development in historical study: A comparative exploration in two fourth-grade classrooms. *Cognition and Instruction, 18*(2), 239–283. DOI: 10.1207/S1532690XCI1802_04

VanSledright, B., Kelly, T., & Meuwissen, K. (2006). Oh, the trouble we´ve seen. Researching historical thinking and understanding. In K. Barton (Ed.), *Research Methods in Social Studies Education: contemporary issues and perspectives*. Greenwich: Information Age Publishing.

VanSledright, B., & Limón, M. (2006). Learning and teaching social studies: A review of cognitive research in history and geography. *Handbook of Educational Psychology*. Routledge.

Vesalainen, M., & Huhtala, A. (2018). Arviointi muutoksessa myös kieltenopetuksessa? Opettajien ja opettajaksi opiskelevien käsityksiä arvioinnista. [Is assessment also changing in language education? Conceptions of teachers and trainee teachers concerning assessment] *Yliopistopedagogiikka 25*(1), 5–29.

Veijola, A. (2013). *Pedagogisen ajattelun kehittyminen aineenopettajakoulutuksessa. Tutkimus suoravalituista historian opettajaopiskelijoista*. [Development of pedagogical thinking during subject teachers´ education. A study of directly selected teacher trainees.] University of Jyväskylä. Dissertation.

Veijola, A. (2016). Historiatietoisuus, historiallinen ajattelu ja historian tekstitaidot: uuden opetussuunnitelman moninaiset lähtökohdat. [Historical consciousness, historical thinking and historical literacy: a diverse basis for the new core curriculum]. *Kasvatus & Aika, 10*(2), 6–18.

Vesterinen, I. (2022). *Orders of history: An ethnographic study on history education and the enacted curriculum*. Jyväskylä: University of Jyväskylä.

Virta, A. (2001). Student teachers´ conception of history. *History Education Research Journal 2*(1), 1–11.

Waldron, F. (2006). Children as co-researchers: Developing a democratic research practice with children. In K. Barton (Ed.*), Research methods in social studies education. Contemporary issues and perspectives*. Greenwich: Information Age Publishing.

Watkins, D., Dahlin, B. & Ekholm, M. (2005). Awareness of the backwash effect of assessment: A phenomenographic study of the views of Hong Kong and Swedish lecturers. *Instructional Science, 33*(4), 283–309.

Webb, N. (1992). Assessment of students' knowledge of mathematics: Steps toward a theory. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 661–683). New York: Macmillan.

Wells, G. (1998). Some questions about direct instruction: Why? To whom? How? And when? *Language Arts, 76*(1), 27–35.

White, H. (1978). Historical text as literary artifact. In R. Canary & H. Kozicki (Eds.), *The writing of history: Literary form and historical understanding* (pp. 81–100). Madison: University of Wisconsin Press.

Wilke, M., Depaepe, F., & Van Nieuwenhuyse, K. (2022). The interplay between historical thinking and epistemological beliefs: A case study with history teachers in Flanders. *Historical Encounters, 9*(1), 198–219.

Williams, J. H. (2014). Nation, state, school, textbook. In J. H. Williams (Ed.), *(Re)Constructing Memory: School Textbooks and the Imagination of the Nation* (pp. 1–9). Sense Publishers.

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

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

Wineburg, S. (2018). *Why learn history (When it's already on your phone)*. The University of Chicago Press.

Wineburg, S., & Martin, D. (2005). Tampering with history: adapting primary sources to struggling readers. *Social Education, 73*(5), 212–216.

Wineburg, S., and S. Wilson. (1991). Models of wisdom in the teaching of history. *The History Teacher, 24* (4), 395–412.

Wissinger, D., Ciullo, S., & Shiring, E. (2018). Historical literacy instruction for all learners: evidence from a design experiment. *Reading & Writing Quarterly, 34*(6), 568–586, DOI: 10.1080/10573569.2018.1510795

Wolfe, M., & Goldman, S. (2005). Relations between adolescents´ text processing and reasoning. *Cognition and Instruction, 23*(4), 467–502.

Wuokko, M. & Räsänen, M. (2023). Tieto, taito, into ja yhteisö: yliopisto-opettajien näkemykset historian asiantuntijuudesta ja sen kehittymisen tukemisesta opintojen alkuvaiheessa. [Knowledge, skills, enthusiasm and community: University teachers´ views on the nature and development of expertise during the first years of university]. *Historiallinen aikakauskirja, 121*(2), 200–211.

Yin, R. (2014). *Case Study Research Design and Methods*. Thousand Oaks, CA: Sage.

Young, M. (1971). *Knowledge and control: New directions for the sociology of education*. Collier-Macmillan.

Young, M. (2008). *Bringing knowledge back in: From social constructivism to social realism in the sociology of education*. Routledge: London.

Zhao, Y., & Hoge, J. (2005). What elementary students and teachers say about social studies. *The Social Studies, 96*(5), 216–221. DOI: 10.3200/TSSS.96.5.216-221

Appendix A

Observation sheet

Teacher _____
Lesson _____
Date _____
Observer _____

Circle if the text originates from a textbook.

Time used (5 minutes)	In total

Texts

textbook									
teacher´s notes									
other non-fiction text									
administrative document									
biography									
caricature									
newspaper									
photograph									
painting or drawing									
fiction									
poetry/lyrics									
map									
statistics									
video									
music									
other									

Classroom activities

direct instruction									
classroom discussion									
IRF interaction									
pair or group work									
reading									
instructions from the teacher									
individual work									
video									
other									
no activity									

Teacher provides specific instructions and examples on how to read and interpret sources, or the assignment orients students towards practising historical literacy strategies.

sourcing									
corroboration									
contextualization									
close reading									

Addition information about the materials

Textbook pages, individual assignments and exercises:

A digital platform:

Additional materials: