Supporting reading comprehension in history education

The use and usefulness of a digital learning environment

Marlies ter Beek

ico
Interuniversity Center for Educational Sciences

ISBN: 978-94-034-2505-3 (printed version)
ISBN: 978-94-034-2501-6 (electronic version)

Cover design and lay-out by Marlies ter Beek
Coloured icons from Flaticon.com
Black icons by Gan Khoon Lay via thenounproject.com
Printed by Ipskamp | proefschriften.net

© Marlies ter Beek/Rijksuniversiteit Groningen, 2020

NRO NATIONAAL REGIEORGAAN ONDERWIJSONDERZOEK

GazeLLe

Most of the research in this dissertation was financially supported by the Netherlands Organisation for Scientific Research (NWO), as part of the research program 'Innovation in education through research' (NRO-PPO: project number 405-15-551).

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, or by photocopying, recording or otherwise, without prior written permission of the author.

rijksuniversiteit
groningen

Supporting reading comprehension in history education

The use and usefulness of a digital learning environment

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

maandag 4 mei 2020 om 16:15 uur

door

Marlies ter Beek

geboren op 30 juni 1990
te Enschede

Promotor
Prof. dr. J.W. Strijbos

Copromotores
Dr. M.C.J.L. Opdenakker
Dr. M.I. Deunk

Beoordelingscommissie
Prof. dr. C.A.M. van Boxtel
Prof. dr. P.C.J. Segers
Prof. dr. A.J. Visscher

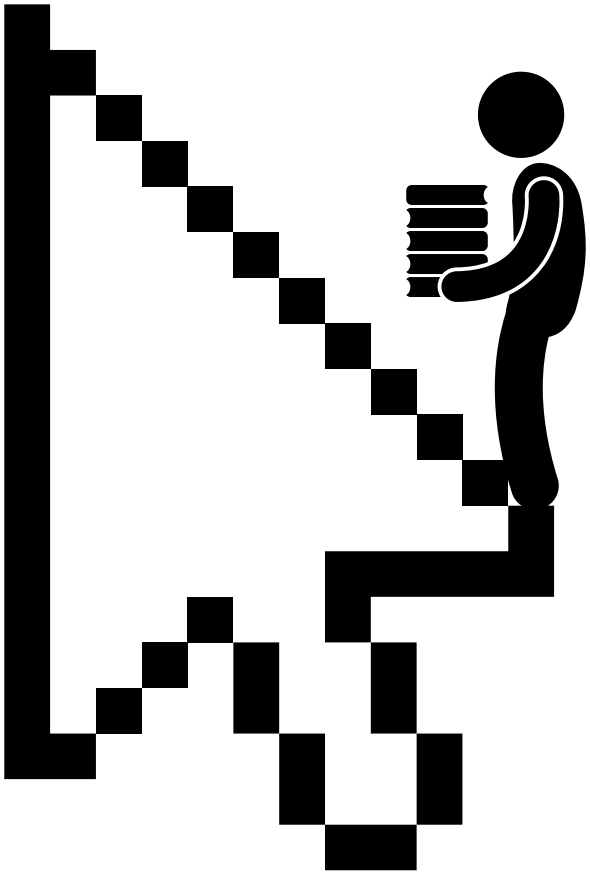
Table of contents

Waarom zou men in de geschiedenisles computers gebruiken? En als men dat dan doet, is het dan alleen maar voor de aardigheid, of is een computer een nuttig instrument dat een belangrijke verrijking voor het geschiedenisonderwijs zou kunnen betekenen?

Why would anyone use computers during history lessons? And if one does, will it be just for fun, or will the computer appear to be a useful instrument and a valuable enrichment to history education in general?

(Wilschut & Bitter, 1992, p. 359)

1



Chapter 1

General introduction

General introduction

Reading comprehension is an essential skill for processing textual information and acquiring knowledge. Therefore, being able to read and comprehend informational texts is a crucial prerequisite for academic success in every subject. Its importance also transcends the academic context since the abundant information provided in our modern, digital 21st-century society calls for critical, informed, and skilful readers (Alexander & The Disciplined Reading and Learning Research Laboratory, 2012; Kamil, Afflerbach, Pearson, & Moje, 2011). Even though the transfer of information has become increasingly visual with the rise of new media such as YouTube, text comprehension skills are still essential for general reading tasks in daily life, such as correctly following instructions in medication leaflets, understanding legal terms and conditions, or distinguishing fake from real news (Raad voor Cultuur & Onderwijsraad, 2019). Reading experts advocate that instruction and practice in reading comprehension should be an essential part of every academic subject's curriculum (Pereira & Nicolaas, 2019).

This finding is of specific relevance for history education because students are required to read a vast amount of fact-dense, expository texts for this subject. However, research has shown that secondary school students in particular often consider these texts to be difficult (Guthrie, Wigfield, & Klauda, 2012). Over the last decade, technology-enhanced learning environments have been used to support students' reading and learning processes, which has proven to be effective for reading comprehension in general (Cheung & Slavin, 2012; Lan, Lo, & Hsu, 2014; Moran, Ferdig, Pearson, Wardrop, & Blomeyer, 2008) and for history education in particular (O'Neill & Weiler, 2006; Poitras, Lajoie, & Hong, 2012). Nevertheless, more research is needed to investigate the role of reading comprehension in history education, especially in the Dutch context. The following paragraphs describe the role of reading comprehension in history education, the concerns in relation to Dutch adolescent students' reading performance and motivation levels, the factors that affect students' reading process, and how digital technology supports students' comprehension of expository history texts as well as teachers' instructional practice.

Reading Texts in History Education

Reading comprehension involves the construction of a mental representation or model of what is written in texts (Kintsch & Rawson, 2005; Perfetti, Landi, & Oakhill,

2005). For the subject of history in lower secondary education, students are required to read an abundance of informative texts, and the expository format of these texts commonly found in textbooks often contains difficult or new vocabulary (Mastropieri, Scruggs, & Graetz, 2003; Ramsay, Sperling, & Dornisch, 2010; Swanson et al., 2016). While reading, students have to infer relationships between sentences and connect information from paragraphs and textbook chapters to update their mental model. In addition, it is often considered essential that students are able to reason historically about the topics and the authors' perspectives they encounter in texts; this reflects the way in which expert historians interact with primary or secondary source material (Shanahan, Shanahan, & Misischia, 2011; van Boxtel & van Drie, 2018; van Drie & van Boxtel, 2008; Wineburg, 1991, 1998).

In the educational literature about the reading of texts for the subject of history, many authors have advocated for a disciplinary approach to literacy instruction (Moje, 2015; Monte-Sano, 2011; Reisman, 2012; Shanahan & Shanahan, 2008; Wineburg & Reisman, 2015). The ability to apply adequate, relevant strategies when reading texts for a specific school subject is commonly known as content-area literacy or disciplinary literacy. For the subject of history, disciplinary literacy practices entail the application of strategies such as sourcing (where does this information come from?), contextualising (what is the historical context in which this was written?), and corroborating (do other sources show similar or different information?). These strategies have shown to be beneficial for students' comprehension of historical topics (Girard & McArthur Harris, 2012; Learned, 2018, Monte-Sano, 2011; Wineburg & Reisman, 2015). However, several studies have emphasised the barriers to this disciplinary approach for students in (lower) secondary history classes: For students who have not fully mastered generic reading comprehension skills, the practice of reading like an expert historian is not yet attainable at this stage of academic development (Duhaylongsod, Snow, Selman, & Donovan, 2015; Nokes, 2011; Perfetti, Britt, & Georgi, 1995). Faggella-Luby, Graner, Deshler, and Drew (2012) use the striking metaphor of "building a house on sand" with regard to a primary focus on disciplinary literacy and conclude that generic reading strategy instruction is pivotal to adolescent students' text comprehension, especially for struggling students.

Concerns about Dutch Adolescents' Reading Comprehension and Motivation Levels

Students' reading performance has often been an area of concern for teachers,

researchers, and policymakers. Already fifteen years ago, a study by Hacquebord, Linthorst, Stellingwerf, and De Zeeuw (2004) indicated that the reading comprehension levels of about 25% of Dutch seventh-graders were insufficient for reading and understanding their textbooks. International reading assessments, such as PIRLS (Gubbels, Netten, & Verhoeven, 2017; Mullis, Martin, Foy, & Hooper, 2017) or PISA (Cito, 2012; Feskens, Kuhlemeier, & Limpens, 2016; Kordes, Bolsinova, Limpens, & Stolwijk, 2013; Organization for Economic Cooperation and Development [OECD], 2016a) showed that the average reading performance of Dutch ten- to fifteen-year-old students has gradually declined over the past two decades, even though the average national score was still among the top-performing countries. However, the latest PISA results, published in December 2019, showed an alarming and 'increasingly negative' trend with regard to Dutch students' reading performance; the average score was the lowest in 15 years and indicated a significant decrease. Moreover, the average Dutch performance was lower than the average performance of all countries participating in PISA (OECD, 2019). An in-depth analysis of the PISA-2015 results showed that only 8% of the fifteen-year-old students read at an advanced comprehension level, while almost one in five students could be classified as low literate, which means that they have difficulties in developing and functioning in today's literate society (Feskens et al., 2016). In 2018, the percentage of students classified as low literate had risen to 24% (Gubbels, van Langen, Maassen, & Meelissen, 2019).

Furthermore, international assessments of reading motivation show that Dutch students' motivation to read is considered weak. Almost half of the fifteen-year-olds do not read for their enjoyment at all, and even the students that do read, only read for a short amount of time per day (OECD, 2016b). For younger students, the numbers are even more alarming: almost a third of all Dutch ten-year-old students in PIRLS-2016 indicated that they did not like to read, leaving the Netherlands at the bottom part of the list of all participating countries with regard to reading enjoyment (Mullis et al., 2017). The same accounts for the Dutch fifteen-year-olds in PISA 2018; more than 40% of the students viewed reading for pleasure as a waste of time, and only 60% only reads texts when necessary (Gubbels et al., 2019). Students' lack of reading motivation may lead to even lower future academic performance since research has shown that the two concepts are related, especially for lower secondary students (Guthrie, Klauda, & Ho, 2013).

These developments did not pass unnoticed. The Dutch Language Union recently published a worrying report concerning the national levels of students'

reading comprehension and reading motivation, accompanied by a call for change. The report highlights five key bottlenecks that likely contributed to the decline in reading comprehension and motivation levels (Pereira & Nicolaas, 2019). First, the authors mention that there is insufficient attention to students' reading motivation in the current educational system. Although research has shown a strong link between reading motivation and academic performance (cf. Guthrie et al., 2013), many teachers view reading motivation as something problematic, yet separate from their instructional practice. Second, it is stated that some students experience reading as boring, because the texts commonly found in educational textbooks in primary education are often simplified and, thus, offer little challenge. A third bottleneck is teachers' reading instruction, which is often ineffective because it lacks evidence-based practices—it seems that the gap between research and practice is too wide in this respect. Fourth, there are differences in how schools monitor their students' reading progress, and commonly used standardised assessments offer limited insight into students' reading skills. Moreover, teachers often do not set specific reading goals nor adapt or personalise their instruction based on students' reading performance. The fifth and final bottleneck concerns the fact that practising reading comprehension shows little coherence with other school subjects. Reading comprehension is often taught as a separate subject, even though it is important to embed this skill in the entire curriculum. Secondary school teachers may express the desire to incorporate reading instruction within their core subject, but often do not feel competent enough to do so (cf. Hall, 2005).

The above-mentioned report focuses mainly on primary education, but it applies just as well to lower secondary education. It is important to stimulate students' reading comprehension and motivation throughout their academic career, but it might be especially crucial after the transition from primary to secondary education. During this transition, the process of 'learning to read' (e.g., decoding, word recognition, or fluency) shifts towards a process of 'reading to learn', for which students' reading skills and knowledge about reading strategies are constantly evolving (Alexander, 2005). However, compared to primary education, relatively little is known about students' reading processes in secondary education (Barnes, 2015). The following paragraphs explain what is known about students' reading process in terms of individual student factors, the influence of teachers' instructional behaviour and the supportive role of digital technology.

Factors that Affect Students' Reading Process

According to the Model of Domain Learning by Alexander (1998, 2005), students of all ages progress from acclimation through competence to proficiency (or expertise) in a certain domain. During this developmental process, there is a strong interaction between students' knowledge, strategy use, and interest; these concepts can also be referred to as students' cognition, metacognition, and motivation (cf. Donker, de Boer, Kostons, Dignath-van Ewijk, & van der Werf, 2014). Many scholars have attempted to unravel the process of students' reading comprehension, resulting in a plethora of scientific articles in this field. These studies focus, among other things, on students' knowledge, awareness and self-regulated use of reading strategies, and motivation to read. Moreover, the educational context in which reading activities take place, including teachers' reading strategy instruction, also affects students' reading comprehension performance.

Cognition: Constructing a mental representation of written texts.

According to the situation model of Kintsch (1998), reading comprehension involves the complex practice of combining textual units into meaningful and coherent mental representations (Kintsch & Rawson, 2005; Perfetti et al., 2005; van den Broek, 2010). This is also the case for students' comprehension of history texts. While reading, students have to connect new information with their prior knowledge to gain a deeper understanding of what is written in paragraphs, chapters, or multiple texts. For example, when students read that both boys and girls were trained in wrestling and javelin throwing in Spartan society, they need to connect this new information to their prior knowledge about Spartan warfare and the fact that both men and women participated in the army. This practice, beyond the level of isolated word or sentence comprehension, is referred to as global text comprehension (Kintsch & Rawson, 2005), and is especially diverse and important after the transition from primary to secondary education; in secondary education, students have to study—on their own and with little support—multiple texts in various subjects (Jetton & Lee, 2012).

However, adolescent students of the same age differ in their cognitive reading skills. For example, vocabulary knowledge, inference skill, and fluency all contribute to students' expository text comprehension but are found to vary between students (Welie, 2017). This might lead to student differences in same-age classrooms; some students may encounter more reading difficulties and are often described as struggling readers (Kendeou, van den Broek, Helder, & Karlsson, 2014). These students may benefit from practising text reading with instructional support, for example with

regard to knowledge of vocabulary and connectives (Welie, Schoonen, Kuiken, & van den Bergh, 2017) or the application of specific reading strategies (Swanson et al., 2016).

Metacognition: Awareness and self-regulated use of strategies before, while, and after reading. Metacognition, as the term suggests, comprises the awareness, knowledge, and control of cognition (Pintrich, Smith, García, & McKeachie, 1991; Zimmerman & Moylan, 2009). In the context of reading comprehension, metacognition is an important factor of influence, since students need to be aware of the cognitive activities and strategies they engage in while reading (McKeown & Beck, 2009; Mokhtari & Reichard, 2002). Reading strategies are deliberate actions, which enable a student "to control and modify the reader's efforts to decode text, understand words, and construct meanings of texts" (Afflerbach, Pearson, & Paris, 2008, p. 368). Students may know that they can consult a dictionary to search for the definition of unknown words (i.e., cognitive knowledge of a reading strategy), but it is equally important to monitor if or when the application of this strategy is necessary (i.e., metacognitive knowledge; Flavell, 1979). Metacognitive awareness is positively related to students' reading performance (Duffy et al., 1987; Mokhtari & Reichard, 2002; Paris, Cross, & Lipson, 1984; Schraw & Dennison, 1994). In a comprehensive meta-analysis on the effectiveness of strategy instruction on academic performance, it was found that metacognitive strategy knowledge significantly improved students' reading comprehension performance (Donker et al., 2014).

According to McKeown and Beck (2009), metacognition is, among other things, partly rooted in self-regulation theory. The metacognitive process of reading—in which students plan, monitor, and regulate their activities before, during, and after reading—is comparable to the cyclical model of self-regulated learning by Zimmerman (2000) and Zimmerman and Moylan (2009), which includes a forethought, performance, and self-reflection phase. Self-regulated learners are metacognitively, motivationally and behaviourally active participants in their own learning process, with the ability to use learning strategies and adapt their behaviour when they encounter problems before, during, of after learning (Boekaerts, 1997; Zimmerman, 2008). Therefore, self-regulated learning (SRL) and metacognition are important elements in the process of reading and comprehending texts (Artelt, Schiefele, & Schneider, 2001; Zimmerman, 2008).

Metacognition can be measured in various ways. In their Motivated Strategies for Learning Questionnaire (MSLQ), Pintrich et al. (1991) focus on students' active and

metacognitive self-regulation with regard to planning, monitoring, and regulating learning activities. In the specific context of reading texts, the Metacognitive Awareness of Reading Strategies Inventory (MARSI; Mokhtari & Reichard, 2002) includes a construct of metacognition based on an extensive inquiry into the reading strategies that are known and used in a self-regulated manner by skilled readers. This resulted in a categorisation of students' awareness of problem-solving, global, and supporting reading skills that students can perform to support their text comprehension.

Motivation: Students' intrinsic motivation, task value, and self-efficacy beliefs. In addition to students' cognitive and metacognitive skills, their motivation to read as well as their motivation for the subject of history in general are also considered crucial factors in the complex process of adolescent students' text comprehension (Guthrie, Klauda, & Ho, 2013; Pintrich & De Groot, 1990; Winne & Hadwin, 2008; Zimmerman, 2008). In general, students who struggle with texts have lower reading motivation than their more competent peers do. This is mainly because when students are motivated, they are more likely to engage in the reading activity and to use cognitive or metacognitive strategies to adapt their reading process (Guthrie & Klauda, 2016; Guthrie, Klauda, & Ho, 2013; Guthrie & Wigfield, 2017). In a study with ninth-grade students, motivational beliefs and attitudes towards school reading were found to relate with students' reading comprehension performance (Wolters, Barnes, Kulesz, York, & Francis, 2017).

However, motivation is a broad and complex concept, which can be approached and measured in different ways. In the context of reading comprehension, motivation can be defined as "the individual's personal goals, values and beliefs with regard to the topics, processes and outcomes of reading" (Guthrie & Wigfield, 2000, p. 406). Students' motivation comprises several aspects, such as intrinsic motivation, task value, and self-efficacy beliefs. Each of these aspects is known to contribute to students' reading performance (Anmarkrud & Bråten, 2009; Retelsdorf, Köller, & Möller, 2011; Schiefele, Schaffner, Möller, & Wigfield, 2012; Taboada, Tonks, Wigfield, & Guthrie, 2009; Unrau & Schlackman, 2006). Intrinsic motivation encompasses students' interest for a certain subject, such as history, or the enjoyment of a certain task, such as reading texts. In contrast, students' extrinsic motivation refers to behaviour that is driven by external rewards, such as grades—which is typical for the Dutch educational system. In the context of reading comprehension, task value refers to students' perceived usefulness of a reading task (Guthrie & Wigfield, 2017), and self-efficacy entails students' perceived ability to be successful in future reading

tasks (Bandura, 1982). It is important to note here that these aspects of motivation and their relation to academic performance may vary between students (Guthrie & Klauda, 2016).

Given the concerns about adolescent students' motivation outlined earlier in this introduction, the stimulation of intrinsic motivation, task value, and self-efficacy during educational reading activities seems indispensable. An important stepping stone for the research at hand is the experimental study by Souvignier and Mokhlesgerami (2006), which showed that self-regulated, strategy-oriented reading instruction programs were effective for fifth-graders' reading comprehension, strategy knowledge, and self-efficacy. Most importantly, a combination of (meta)cognitive and motivational strategy support had the strongest effects on a long-term retention test, compared to solely cognitive or a combination of cognitive and motivational support. This finding shows the importance of combining all three types of support when stimulating students' reading comprehension and motivation. In addition, a more recent study by Zepeda, Richey, Ronevich, and Nokes-Malach (2015) revealed that students who received metacognitive instruction and training during reading showed significantly higher levels of task value and self-efficacy, indicating that metacognitive support can enhance students' motivation as well.

Reading environment: Teachers' instructional practice. An additional and more comprehensive factor that affects students' reading process is the environment in which the reading activity is carried out. Research in this field has shown that home literacy environment (e.g., parental involvement in literacy tutoring, or socio-economic status; Senechal, 2006) and print exposure (e.g., amount of books or reading at home; Mol & Bus, 2011) are of influence on reading comprehension skills, but most reading activities for the subject of history are carried out in a classroom context. In secondary education, teachers provide instruction on separate subjects, such as history and geography, and reading comprehension skills are often taught exclusively in Dutch language lessons. However, it is known that in general, teacher instruction on reading comprehension strategies is effective for students' academic performance (de Jager, Reezigt, & Creemers, 2002; Palincsar & Brown, 1984; National Reading Panel, 2000; Okkinga et al., 2018). Teachers should not merely transmit information, but actively guide and monitor their students' learning processes (de Jager et al., 2002), since a lack of guided instruction has shown to be ineffective for novice learners such as lower secondary students (Kirschner, Sweller, & Clark, 2006). Using an instructional model based on direct or guided instruction to provide

students with reading strategy instruction has proven to enhance students' use of these strategies, their text comprehension, and their engagement (Guthrie & Davis, 2003; Kirschner et al., 2006; Smale-Jacobse, 2013).

However, research on reading strategy instruction in subjects such as history has shown that this type of instruction rarely occurs in daily practice (Linthorst & de Glopper, 2015; Ness, 2016). Teachers occasionally explain word meanings or ask students if they understand what has been read, but this instruction does not reach an explicit level wherein the usefulness or application of certain reading strategies is explained or discussed (Moje, 2008). Possible explanations suggest that teachers often do not feel responsible or sufficiently qualified to provide reading instruction within their own subject, or that they encounter difficulties in determining suitable reading instruction for each individual student (Greenleaf, Schoenbach, Cziko, & Mueller, 2001; Hall, 2005; Ness, 2016; O'Brien, Stewart, & Moje, 1995).

Digital Technology Supporting Students' Reading Process and Teachers' Instruction

Nowadays, technological applications are apparent in every classroom. With the rise of the concept of personalised learning, blended learning, and bring-your-own-device policies in schools, educators increasingly use computers, tablets, and laptops to provide students with educational materials and assignments. In return, the monitoring systems in many of these digital learning environments allow teachers to support, monitor, and evaluate individual students' learning processes (Azevedo & Gašević, 2019). Meta-analyses and other studies have shown that digital technology, such as tools and learning environments, has a positive effect on students' reading performance in both primary and secondary education (Cheung & Slavin, 2012; Lan et al., 2014; Lynch, Fawcett & Nicholson, 2000; Moran et al., 2008).

While educational technology develops rapidly, the research on the effectiveness of these technological developments on students' reading performance progresses at a slightly slower pace. Reported effect sizes are often small, few studies are aimed at secondary grade levels, and outcome measures seldom include elements of students' metacognition or motivation (Moran et al, 2008). Moreover, it remains somewhat unclear which support characteristics in computer-supported learning environments contribute to secondary students' reading process because studies often lack a detailed description of the actual content or focus of the provided support (Devolder, van Braak, & Tondeur, 2012; ter Beek, Brummer, Donker, & Opdenakker, 2018).

Therefore, it is essential to carefully analyse if and how digital support systems with supportive elements influence students' reading progress in an authentic classroom setting in secondary education.

Supporting students' reading process. Digital technology can support both students' text comprehension and students' self-regulated reading process. This support is often provided in the form of scaffolds such as hints, which can be defined as "tools, strategies and guides to support students in regulating their learning" (Lajoie, 2005, p. 547; Pea, 2004). Hints function as strategy activators, providing information about how to complete a specific learning task without disclosing the correct answer (Aleven & Koedinger, 2002; Azevedo, 2007; Devolder et al., 2012). In the educational literature, a distinction is made between cognitive, metacognitive, and motivational support (Donker et al., 2014), and several studies have focused on the effectiveness of either one or a combination of multiple types of support (Azevedo, 2005). Cognitive support is meant to help the student solve a problem on his or her own by providing information regarding the content of the learning material (Lajoie, 2005). Metacognitive support aims at improving students' regulation of learning (e.g., by planning, monitoring, or evaluating) which is an effective strategy in the context of reading (Donker et al., 2014). Motivational support is meant to enhance student interest, learner control, or affect (Lajoie, 2005).

Providing students with cognitive support, or a mixture of cognitive and metacognitive support, has been shown to have a positive effect on students' learning outcomes in general as well as the use of cognitive and metacognitive strategies (Berthold, Nückles, & Renkl, 2007; Souvignier & Mokhlesgerami, 2006). A systematic review by Devolder et al. (2012) addressed the effectiveness of computer-based scaffolds focused on self-regulated learning in the domain of science. The authors found that most effective scaffolds, such as prompts, focused on students' cognition. Metacognitive scaffolds were offered less often, and no clear conclusions about motivational scaffolds could be drawn due to the small number of scaffolds aimed at increasing or sustaining students' motivation. In addition, the authors note that most studies paid little attention to student characteristics, such as prior knowledge, self-beliefs, or motivation, or task characteristics, such as the frequency of scaffolds provided, which could also be of influence on students' performance.

A meta-analysis on the effects of computerised reading contexts by Lan et al. (2014) showed that instruction on metacognitive (self-) regulation had positive effects on seventh-grade students' reading performance. It must be noted, however,

that this finding was based on two studies (cf. Johnson-Glenberg, 2005; Puntambekar & Stylianou, 2005), since only a third of the participants in the studies included in this meta-analysis were from secondary education. Moreover, one of these two studies focused solely on students with poor reading comprehension. Nevertheless, the aforementioned findings indicate that including metacognitive support in computerised reading contexts, such as a Digital Learning Environment (DLE), may provide an effective form of improving students' reading comprehension performance.

Supporting teachers' instructional practice. Although scaffolded DLEs enable students to learn in a predominantly self-regulatory manner, this does not imply that teachers can go into standby mode. Cheung and Slavin (2012) found that the effects on students' reading achievement were larger when teachers were actively involved in using computer environments, for example by tailoring their instruction to complement the digital information provided. The technology behind online DLEs enables teachers to draw from a large source of data, such as log files, to monitor their students' reading activities and progress, which they can subsequently use to adapt their instruction. The process of using student data to inform instructional practice is also known as Data-Based Decision Making (DBDM; Schildkamp, Lai, & Earl, 2013), and its use has been associated with increased student performance (Campbell & Levin, 2009). A study by Lai, Wilson, McNaughton, and Hsiao (2014) showed that a DBDM intervention, in which teachers collaboratively practised profiling based on student assessment data, had positive effects on students' reading comprehension performance in secondary education. This finding indicates that students benefit when teachers apply educational technology and digital data to substantiate their instruction.

Similar to the provision of reading strategy instruction, the use of digital data about students' performance to inform teachers' instructional practice is not always evident in daily educational practice, especially in secondary education (Deunk, Smale-Jacobse, De Boer, Doolaard, & Bosker, 2018; Kippers, Wolterinck, Schildkamp, Poortman, & Visscher, 2018). In addition, it is known that digital data output is often quite extensive and is only easily interpretable for expert, skilled teachers (Vanhoof, Verhaeghe, Van Petegem, & Valcke, 2013). There seems to be a need for professional development training for teachers with regard to DBDM (Fisher, Frey, & Lapp, 2011; Kippers et al., 2018; Mandinach & Jimerson, 2016; Poortman, Schildkamp, & Lai, 2016). More specifically, teachers need to be able to integrate data skills with subject

matter content knowledge in an authentic context (Staman, Visscher, & Luyten, 2014). Studies have shown that professional development training in using data to substantiate teachers' instructional practice is beneficial for both teachers and students (Schildkamp, Lai, & Earl, 2013).

The Overarching Research Project: Gazelle

The present dissertation is grounded in a practice-oriented research project carried out between September 2015 and January 2019, officially known as "Cognitive, metacognitive and motivational hints to support self-regulated learning in secondary education: Research into an effective supportive learning environment".[1] This overarching research project stemmed from the urgent call coming from both school leaders and teachers to support students' self-regulation skills in the context of reading comprehension. The research project focused on text comprehension in lower secondary education, especially for subjects in which these students have to read many expository texts, such as history and geography. The current dissertation focuses specifically on the implementation and results for the subject of history.

The overarching research project has been approved and financed by the Netherlands Organisation for Scientific Research (NWO) and has been carried out according to the ethical guidelines of the department of pedagogy and educational sciences of the University of Groningen (April 2012). The official project report (ter Beek, Spijkerboer, Brummer, & Opdenakker, 2018) provides, among other things, an extensive description of the whole research project (in Dutch); in the following subparagraphs, relevant elements from this report have been adapted and translated to describe the research design and context.

The digital learning environment ("Gazelle"). This research project aimed at stimulating students' cognition, metacognition, and motivation with the help of a digital learning environment (DLE) and in the context of the self-regulated reading of expository texts for the subjects of history and geography, since it is known from the research literature that these three factors all influence students' reading comprehension. Cognitive, metacognitive, and motivational scaffolds, called 'hints', were incorporated in a DLE to support students' comprehension of expository texts. This DLE was given a name that was easily recognizable for both teachers and students, while also referring to relevant elements under study: "Gazelle", a Dutch acronym

1 The original project title is: "Cognitieve, metacognitieve en motivationele hints ter bevordering van zelfgestuurd leren in het secundair onderwijs: onderzoek naar een effectieve ondersteunende leeromgeving".

for "Gemotiveerd, actief en zelfstandig lezen" (motivated, active, and independent reading).

The DLE contained expository texts written by the project researchers in cooperation with the participating teachers. For the subject of history, these texts covered the topic of ancient Greece, which was in line with the historical period that was treated in the regular curriculum. Each text had a length of approximately 550 words and was written in an expository format. Teachers integrated the use of the DLE with their regular lessons, which each lasted about 50 minutes. After students logged into the DLE, they had to read the text and summarise it directly after reading. Consequently, each student had to answer ten text-related multiple-choice questions. At the end of each lesson, students could reflect on their summary and rate their own work on a scale of one to ten. While working in the DLE, students could consult cognitive hints alongside the multiple-choice questions that presented strategic information about the literal contents of the text, and metacognitive hints that presented strategic information to guide students' regulation of their learning process before, during, and after reading. Motivational hints pointed out the value of the reading task (i.e., the 'why' of the task) and what students might learn by reading the texts. Since students deliberately had to click on a light bulb-shaped button to open the contents of the hints (each type of hint was represented by a different icon), the embedded scaffolding depended on students' self-regulated, help-seeking initiative (Narciss, Proske, & Koerndle, 2007). The DLE automatically logged students' behaviour, such as time on task, given answers on open-ended and multiple-choice questions, and the consultation of hints.

Design and timeline of the Gazelle project. After the design and contents of the DLE were created and piloted in cooperation with both teachers and students (ter Beek, Spijkerboer et al., 2018), the first intervention year started in October 2016. Three comparable secondary schools from the northern part of the Netherlands participated, resulting in a total of 228 seventh-grade students who worked with the DLE. During this first year (2016–2017), the research project focused on the effectiveness of the availability and use of either a combination of cognitive and metacognitive hints (i.e., Phase 1) or a combination of cognitive, metacognitive, and motivational hints (i.e., Phase 2) for students' text comprehension, self-regulated learning, and motivation. After a general initial reading comprehension test (Aarnoutse, 1987), students from Experimental condition A were provided with hints focusing on reading strategies for the subject of history, while students from Experimental condition B were provided

with similar hints for the subject of geography (but not for history). Students from the control condition were not provided with hints for either subject. Questionnaires, based on the MSLQ (Pintrich et al., 1991) and MARSI (Mokhtari & Reichard, 2002), were administered prior to and after each series of six DLE lessons to measure students' self-regulated learning and motivation. Subsequently, questionnaire data and log files from the DLE were used to compare the two experimental conditions and the control condition.

In the second year of the research project (2017–2018), the two experimental conditions from year 1 continued working with the DLE for the subject of history, albeit with a new cohort of seventh-grade students. Two other schools were added to the project and only used the DLE for the subject of history. In total, 328 students from thirteen classrooms and nine different history teachers participated. During the second year, all students were provided with the combination of cognitive, metacognitive, and motivational hints in the DLE while reading expository history texts, similar to Phase 2 of year 1. Just as in year 1, the effects of students' hint use could be analysed based on data from the DLE.

Additionally, in year 2 the project also focused on the teachers; they received digital data output based on their students' performance in the DLE, to enable them to provide informed or personalised instruction. During Phase 1 of year 2, teachers from Experimental group A were able to consult basic and extended visualised data output, whereas teachers from Experimental group B and the control group were only able to consult basic data output. In Phase 2, four teachers in Experimental group A additionally received a professional development training and a guiding manual halfway through the school year to support the translation of students' extended performance data into effective reading strategy instruction. Two teachers in Experimental group B received both basic and extended data output, but no training. Teachers from the control group again only received the basic visualised data output. Lesson observations, interviews, focus group meetings, and teacher questionnaires were conducted to compare teachers from both experimental groups and the control group.

Figure 1.1 shows the timeline of the overarching research project as well as the individual studies included in this dissertation. It is important to note that the structure and contents of the questionnaires and DLE lessons in both intervention years 1 and 2 were highly similar, with the exception of the added motivational hints in Phase 2 of year 1 and the historical content knowledge (HICK) test halfway through

year 2. The T1, T2, T3, and T4 questionnaires each measured students' intrinsic motivation, task value, self-efficacy, metacognitive self-regulation, and awareness of reading strategies in a similar way. The HICK test was designed specifically for this dissertation; therefore, it was only added for those students who used the DLE for the subject of history.

Dissertation Overview

The general aim of this dissertation is to investigate the use and usefulness of a scaffolded DLE (i.e., 'Gazelle'), implemented in regular lessons to support students' comprehension of expository history texts. It is assumed that using the DLE affects both students and teachers. As a result, two main research questions will be explored simultaneously:

- How do seventh-grade students and their history teachers use and experience a DLE enriched with strategic hints and visualised student data (i.e., how do they use it)?
- What are the effects of using the DLE on students' reading process and learning outcomes, and on history teachers' instructional practice (i.e., how useful is it)?

Each individual study or chapter is part of the overarching research context, integrating the technology-enhanced, self-regulated reading of texts within a regular seventh-grade history curriculum. Figure 1.2 provides an overview of the concepts incorporated in this dissertation, as well as the studies and chapters in which they are included.

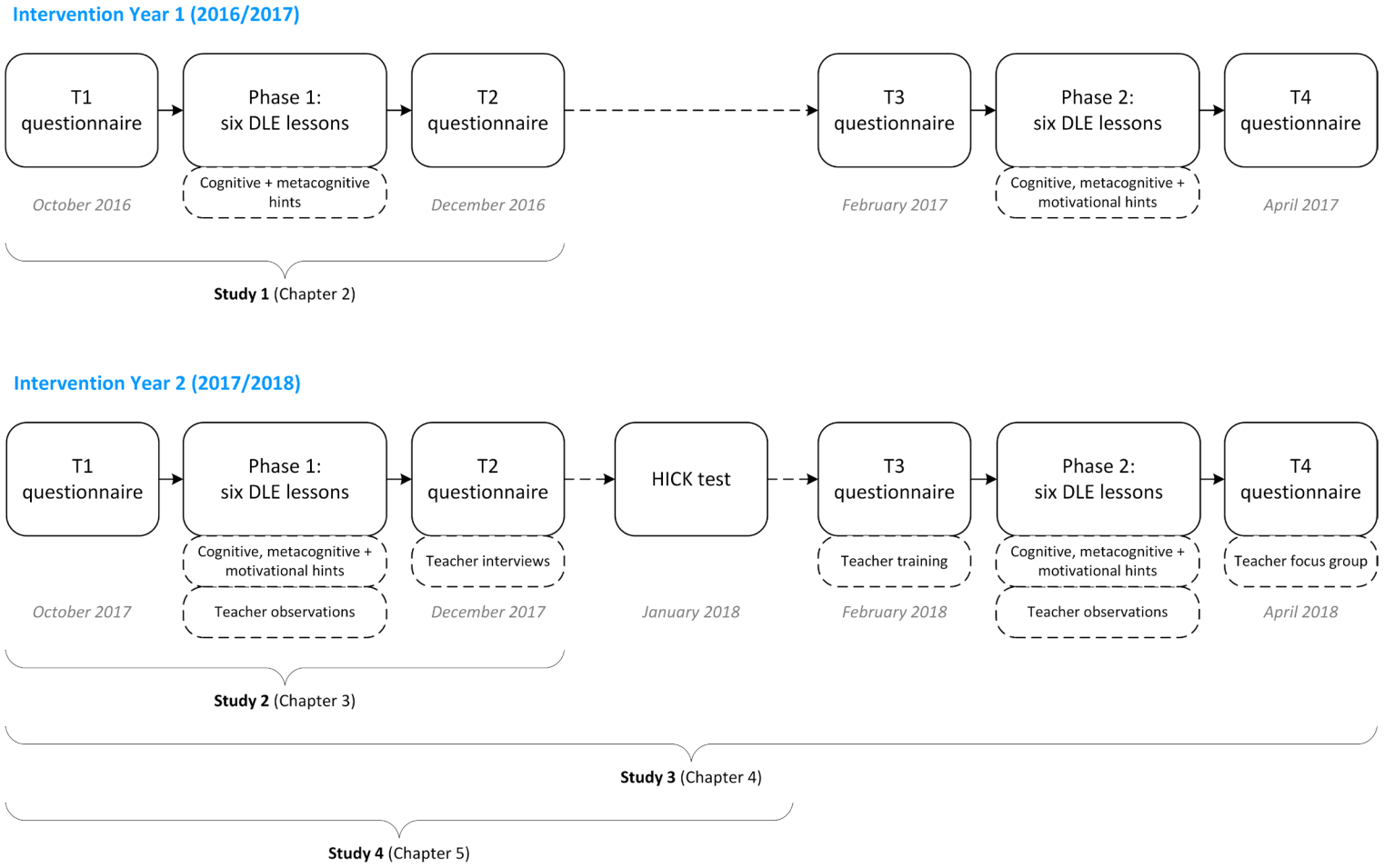


Figure 1.1 Timeline of studies included in this dissertation. DLE = digital learning environment; HICK = historical content knowledge.

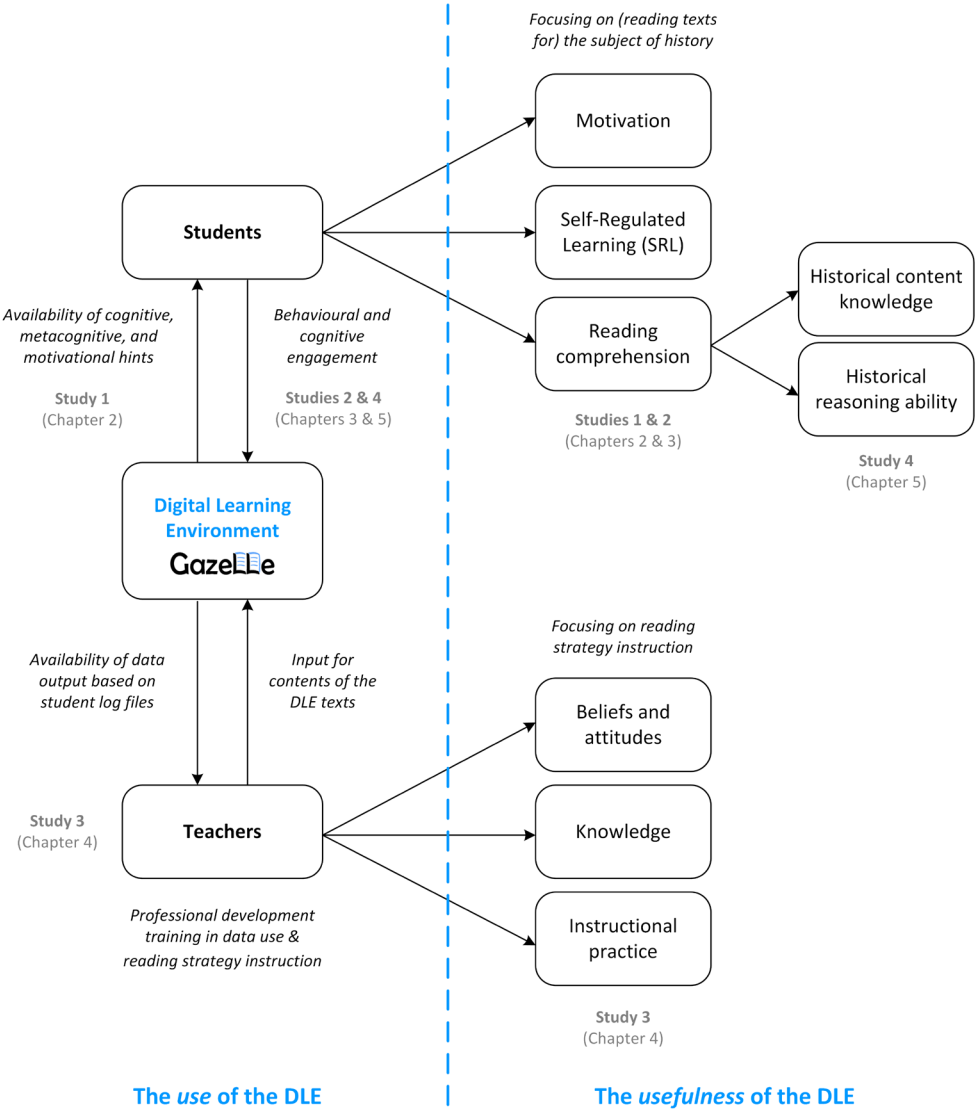


Figure 1.2 Overview of concepts included in the studies and chapters of this dissertation.

The first two studies answer research questions with a specific focus on the students in the first year (Chapter 2) and second year (Chapter 3) of the intervention. The third study focuses on teachers' use of the DLE in the second year of the intervention (Chapter 4). Chapter 5 describes a fourth and follow-up study on the effects of practising expository text reading on students' historical content knowledge and historical reasoning ability. All studies incorporate findings on the use and usefulness of the DLE. Lastly, Chapter 6 summarises and discusses the main findings

of the studies included in Chapters 2 through 5, and provides recommendations for future research and practice.

There may be some overlap between the chapters in this dissertation since they were written as independent journal articles (i.e., it is possible to read every chapter separately). Therefore, it was occasionally necessary to repeat theoretical background elements or descriptions of research methods and instruments. The following subparagraphs describe the individual studies in more detail; however, extensive descriptions and visualisations of the complex study designs are reported in the individual chapters.

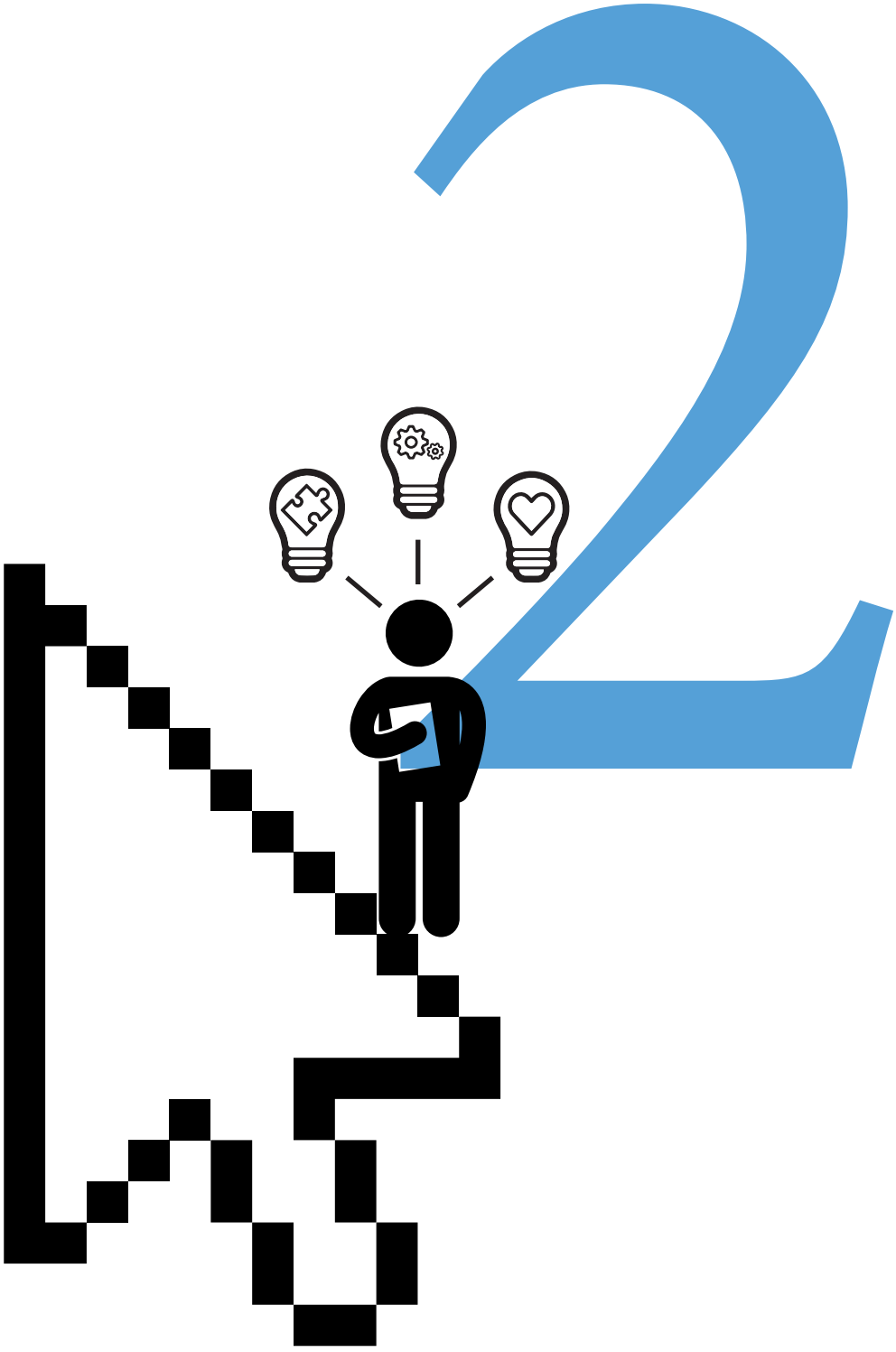
Effects of scaffolds on text comprehension, self-regulation, and motivation (Study 1, Chapter 2). The first study describes an experimental study that compares students who worked in the DLE and were provided with scaffolds (hints) for the subject of history (Experimental condition A), students who were provided with hints for the subject of geography, but not for history (Experimental condition B), and students with no hints provided for either subject. It uses data from the first half of the first year of the Gazelle research project and compares students from Experimental conditions A and B and the control condition in terms of text comprehension, self-regulated learning, and motivation. Additionally, the study focuses on the effects of the actual use of these hints. Lastly, this study also explores the effects of working in the DLE and using supportive hints for students with below-average, average, and above-average initial reading levels.

The relations between motivation, engagement, and comprehension (Study 2, Chapter 3). Whereas the first study focuses on differences between experimental groups, hint users, and groups of students with different initial reading comprehension levels, the second study uses data from the second year to identify subgroups of students by using latent profile analysis (LPA) based on their engagement within the DLE. Focusing on students' cognitive and behavioural engagement profiles enables the adoption of a person-centred approach, which complements the variable-centred approach used in the first study (Hickendorff, Edelsbrunner, McMullen, Schneider, & Trezise, 2018; Marsh, Lüdtke, Trautwein, & Morin, 2009). Subsequently, the study explores the relations between these engagement profiles and students' motivation as well as their text comprehension.

Supporting history teachers' reading strategy instruction (Study 3, Chapter 4). The third study focuses on how teachers use visualisations of student

performance in the DLE and how this affects their knowledge, attitudes, beliefs, and instructional behaviour with regard to embedding reading strategies in their history lessons. It uses data from the second year and provides a micro-level analysis of several types of observed reading strategy instruction during history lessons. In addition, it uses qualitative data, such as interview and focus group meeting transcripts, to address teachers' experiences with the use of a DLE and a professional development training on the provision of reading strategy instruction and the use of log file data. Following the content analyses of the teacher interviews, the study additionally highlights contextual barriers, such as time pressure and logistic problems, which can induce implementation challenges.

Effects on historical content knowledge and historical reasoning ability (Study 4, Chapter 5). Throughout the overarching research project, many history teachers asked the question: "But what did my students actually learn while reading those DLE texts?" The fourth study focuses on this question by analysing the relations between subject-specific reading skills (e.g., identifying causal relations) included in the DLE and students' delayed historical content knowledge about the topics of the DLE texts. Since the ability to reason historically is crucial for students' understanding of history in general (van Boxtel & van Drie, 2008), it also focuses on the relations between the subject-specific reading skills and students' historical reasoning ability. Additionally, this study integrates the five profiles identified in the second study (Chapter 3) to explore the influence of students' behavioural and cognitive engagement in a DLE on their historical content knowledge and historical reasoning ability. By doing so, it explores whether and how history lessons that include the use of a DLE to focus on reading comprehension are beneficial for seventh-grade students.



Chapter 2

Scaffolding expository history text reading: Effects on adolescents' comprehension, self-regulation, and motivation

This chapter is published as:

ter Beek, M., Opdenakker, M.-C., Spijkerboer, A. W., Brummer, L., Ozinga, H. W., & Strijbos, J. W. (2019). Scaffolding expository history text reading: Effects on adolescents' comprehension, self-regulation, and motivation. *Learning and Individual Differences*, *74*(101749), 1–12. https://doi.org/10.1016/j.lindif.2019.06.003

Abstract

Reading comprehension is an important predictor for academic success, yet many adolescents in secondary education face difficulties when reading their textbooks. In this quasi-experimental study, we developed a digital learning environment to scaffold students' expository text reading in seventh-grade history classrooms. Students in the experimental condition could use hints comprised of cognitive and metacognitive reading strategy instruction, whereas students in the control condition received no additional support. A comparison of posttest comprehension between conditions showed no significant differences. However, students in the experimental condition who accessed hints during history text reading performed significantly better on the posttest than students who did not use hints at all. We found no differences between conditions regarding students' self-regulated learning or motivation, but students' awareness of problem-solving reading strategies significantly increased in the experimental condition. Finally, a comparison of students with different reading levels showed that below-average readers benefitted most from digital reading practice.

3

schools/conditions

174

students

4

research questions

Highlights

- Hint users outperformed students who did not use hints significantly on measures of text comprehension.
- Providing (meta)cognitive hints did not lead to differences between conditions.
- Providing hints increased students' problem-solving reading strategy awareness.
- Digital practising had no significant negative effects on below-average students' reading comprehension.

Introduction

Reading comprehension is an important prerequisite for learning, particularly in history classes given the abundant use of broad expository texts (Mastropieri, Scruggs, & Graetz, 2003). Reading requires the application of both topic knowledge and domain knowledge (i.e., knowledge about reading strategies), which continually develop after the transition from primary to secondary education (Alexander, 2005). Despite extensive research on reading comprehension in primary education, relatively little is known about reading in secondary education (Barnes, 2015). Most studies focus on reading challenges for struggling adolescent students (Faggella-Luby, Graner, Deshler, & Drew, 2012; Mastropieri et al., 2003; Ness, 2016; Ramsay, Sperling, & Dornisch, 2010). However, all students need to learn how to correctly apply reading comprehension strategies, preferably using relevant and domain-specific content (Lan, Lo, & Hsu, 2014; McKeown, Beck, & Blake, 2009; Shanahan & Shanahan, 2008).

Comprehending Expository History Texts

Most history textbooks in secondary education contain fact-dense texts written in expository prose, with difficult vocabulary and obscure internal references (Mastropieri et al., 2003; Ramsay et al., 2010; Swanson et al., 2016). This expository format contrasts the narrative texts that are more common in primary education or language art classes, which makes it difficult for young secondary students to adapt to it (Fry & Gosky, 2007). Therefore, it is important to provide adolescent readers with adequate generic and domain-specific reading strategy instruction.

A case in point is research by Vaughn et al. (2013), which showed that eighth-grade students performed significantly better on content acquisition and reading comprehension when they were provided with specific reading strategy instruction during expository text reading, such as guiding questions for the text. A replication study yielded similar results (Vaughn et al., 2015). Other types of reading strategy instruction based on text content, such as elaborative interrogation or identifying and generating main ideas, have also proved effective for expository text comprehension in history classrooms (McKeown et al., 2009; Ramsay et al., 2010). To comprehend history texts, students need to know how and when to apply relevant reading strategies.

Reading and Self-Regulated Learning (SRL)

In general, learners are self-regulated to the degree that they are metacognitively, motivationally and behaviourally active participants in their own learning process. Self-regulated learners can apply learning strategies and adapt their learning behaviour when confronted with problems (Zimmerman, 2008). In line with this definition, self-regulated learning (SRL) is an important skill in the process of reading and comprehending texts (Artelt, Schiefele, & Schneider, 2001; Zimmerman, 2008). When students study their textbooks, they have to regulate their own learning, which includes that they decide which reading strategies they apply from the set of strategies they have at their disposal. According to Mokhtari and Reichard (2002), "awareness *and* monitoring of one's comprehension processes are critically important aspects of skilled reading" (p. 249, italics in the original). In fact, Mason (2013) showed that explicit reading strategy instruction combined with students' self-regulated learning before, while, and after reading has positive effects on students' performance. This three-step approach relates to the SRL model by Zimmerman (2000) and Zimmerman and Moylan (2009).

The cyclical model of SRL by Zimmerman (2000) and Zimmerman and Moylan (2009) is widely used in educational research (Panadero, 2017). It distinguishes three phases in student learning: the forethought phase, the performance phase, and the self-reflection phase. In line with this model, students can self-regulate their reading process by applying reading strategies before, during, and after reading. During the forethought phase, students might set goals for reading, determine the value of the reading task, or indicate the perceived difficulty of the task. During the performance phase, students might monitor their own reading, apply reading strategies, or seek help. During the self-reflection phase, students can evaluate their own reading process in various ways and decide to proceed to a new forethought phase. Each phase encompasses both metacognitive and motivational processes. Recent research on SRL and reading often includes motivational and affective aspects of learning to explore the complex learning processes of adolescent students (Guthrie, Klauda, & Ho, 2013; van Steensel, van der Sande, Bramer, & Arends, 2016).

Reading and Student Motivation

Since motivation is related to both performance and SRL, it is also essential to consider students' motivation in reading research (Guthrie et al., 2013; Schiefele, Schaffner, Möller, & Wigfield, 2012; Schunk & Zimmerman, 2008; Winne & Hadwin, 2008;

Zimmerman, 2011). For example, students' intrinsic goal orientation is an important element of self-regulation: without a clear goal, it is difficult to apply adequate learning strategies (Pintrich, 2000). In addition, students need to recognise the value of a reading task or decrease the perceived difficulty of a reading task (i.e., increase their self-efficacy beliefs) to be motivated to read texts (Guthrie et al., 2013; Pajares, 2008). Students' intrinsic motivation can be increased by stimulating feelings of competence, relatedness, and autonomy (Guthrie et al., 2013; Ryan & Deci, 2000). For example, students' feelings of autonomy are stimulated when they are able to decide which tasks to perform with regard to reading texts. Instruction on SRL strategies can also enhance students' motivation. A study by Zepeda, Richey, Ronevich, and Nokes-Malach (2015) revealed that students who received metacognitive instruction and training showed significantly higher levels of task value, self-efficacy, and mastery-approach goals.

Struggling Readers

Since SRL and motivation contribute to text comprehension, it can be argued that struggling readers—who have difficulties applying relevant strategies when reading expository texts—will benefit most from practice in reading combined with instructional support. In fact, Swanson et al. (2016) showed that struggling readers who daily received specific reading strategy instruction significantly improved on measures of knowledge acquisition, content reading comprehension, and vocabulary recall when compared with struggling students in a business-as-usual condition. Welie, Schoonen, Kuiken, and Van den Bergh (2017) discovered that eighth-grade students' knowledge of connectives (i.e., words that signal coherence in a text, like 'because' or 'therefore') was associated with expository text comprehension and metacognitive knowledge. More specifically, students with more metacognitive knowledge showed a stronger relationship between knowledge of connectives and text comprehension, indicating that students with less knowledge of connectives might benefit from metacognitive instruction to better comprehend expository texts.

Cognitive and Metacognitive Scaffolding

There are various ways to support students' text comprehension, SRL, and motivation. Strategy instruction is often used to enhance students' knowledge about which actions might improve their reading. A recent meta-analysis on the effectiveness of reading strategy interventions in whole classrooms showed a small but significant effect of reading strategy interventions on researcher developed-comprehension tests

(Cohen's d = .43); the effect sizes were largest for students in grades 6–8 (Okkinga et al., 2018). With regard to strategy knowledge and strategy use, the authors found small effects (Cohen's d = .37 and .36, respectively); in terms of strategic ability, larger effect sizes were obtained for low-achieving students. The authors conclude that "both knowledge about the different strategies and students' awareness of the type of strategies that are taught can be increased by the reading strategy interventions" (Okkinga et al., 2018, p. 1230).

In their meta-analysis of the effectiveness of learning strategy instruction on academic performance, Donker, de Boer, Kostons, Dignath-van Ewijk, and Van der Werf (2014) make a distinction between cognitive and metacognitive strategies. Cognitive strategies refer to domain or task-specific information; metacognitive strategies are higher-order strategies that regulate students' cognition, such as planning, monitoring, and evaluating. Although a small effect was found for reading comprehension (Hedges' g = .36), metacognitive knowledge significantly improved student performance. Moreover, Askell-Williams, Lawson, and Skrzypiec (2012) concluded from an extensive inventory with 1388 students that there was room for improvement in early adolescent students' cognitive and metacognitive strategy use. Scaffolding strategy use with learning protocols raised students' levels of strategy knowledge, although this finding was slightly limited. Lastly, an experimental study by Souvignier and Mokhlesgerami (2006) showed that a combination of cognitive, metacognitive, and motivational support is most effective for stimulating students' (long-term) reading comprehension.

Strategy instruction or support can be provided in the form of scaffolds, which contain strategy instruction or guidelines for answering questions. Scaffolds can be defined as "tools, strategies and guides to support students in regulating their learning" (Lajoie, 2005, p. 547), and can include cognitive, metacognitive or motivational processes. Often these scaffolds provide information about how to complete a specific learning task, without disclosing the correct answer (Aleven & Koedinger, 2002; Devolder, van Braak, & Tondeur, 2012; Lysenko & Abrami, 2014; McNamara, 2007). Scaffolds can differ in terms of their function, type of delivery, and the tool or mechanism by which they are presented. A typical scaffold is the prompt or 'hint', which is viewed as a strategy activator (Berthold, Nückles, & Renkl, 2007).

Berthold et al. (2007) provided undergraduate students with either cognitive prompts, metacognitive prompts, a mixture of cognitive and metacognitive prompts, or no prompts at all while writing a learning protocol. They found that participants

who received cognitive or mixed prompts performed significantly better on learning outcomes, and showed significantly more cognitive learning strategies than students who received metacognitive or no prompts. Additionally, students who received prompts (in either way) showed significantly more metacognitive strategy use compared to students who received no prompts. Therefore, they argued that the provision of strategy prompts leads to more cognitive and metacognitive strategy use in students' learning activities.

Reading in a Digital Environment

Over the past decades, the possibilities of instructional technology expanded research on the effects of digital learning environments (DLEs) on students' academic performance (Zheng, 2016). With regard to reading comprehension, it has been shown that both instruction and support in DLEs positively affects students' reading comprehension (Cheung & Slavin, 2012; Lan et al., 2014; Lysenko & Abrami, 2014; Moran, Ferdig, Pearson, Wardrop, & Blomeyer, 2008). DLEs enhance students' autonomy and provide individual flexibility and support. Devolder et al. (2012) concluded from their systematic review on scaffolding in computer-based learning environments that digital hints appear to be effective scaffolds, especially as support to stimulate the use of learning strategies. For example, hints can improve students' effort regulation by suggesting what actions to perform when confronted with difficulties while reading texts.

Strategy instruction and SRL supports are established predictors of reading performance. However, many existing studies that use digital or computer-supported environments mainly investigate the effects of support in primary or higher education, even though reading comprehension is equally essential for secondary education (Cheung & Slavin, 2012; Zheng, 2016). Moreover, many studies do not investigate the combined effects of cognitive and metacognitive instruction or scaffolds (Devolder et al., 2012; Lan et al., 2014). Finally, a recent systematic review by Ter Beek, Brummer, Donker, and Opdenakker (2018) showed that hardly any research has been conducted in the field of computer-supported expository text reading in secondary education. Therefore, the current study focuses on the combined use of cognitive and metacognitive support in a digital setting in secondary education, in the specific context of expository history text reading. To our knowledge, this combination of subject-specific, computer-supported research in secondary education has not been conducted in the field of reading comprehension or SRL research so far. By doing

this, the current study provides helpful insights for researchers and teachers who wish to integrate supportive educational technology in their lessons.

Research Aims and Expectations

The purpose of this study is to determine the effect of cognitive and metacognitive support (i.e., scaffolding through hints) in a digital learning environment on secondary students' expository history text comprehension, SRL, reading strategy awareness, and motivation. Since the use of hints was optional and relied on students' autonomous decisions to use them, the sub-question for each research question focuses on differences between students who accessed hints and students who did not. Additionally, we will analyse if there are different effects for secondary students with below-average, average, and above-average reading levels. We will address the following research questions:

1. What is the effect of the provision and use of cognitive and metacognitive hints on students' history text comprehension?
2. What is the effect of the provision and use of cognitive and metacognitive hints on students' SRL and reading strategy awareness?
3. What is the effect of the provision and use of cognitive and metacognitive hints on students' motivation for history in terms of task value and self-efficacy?
4. What are the effects of the provision and use of cognitive and metacognitive hints on text comprehension, SRL, reading strategy awareness, and motivation for students with different reading levels?

We expect that students who actually use the provided cognitive and metacognitive hints, compared to students who do not use them, will show higher or better (a) text comprehension (cf. Donker et al., 2014), (b) SRL and reading strategy awareness (cf. Berthold et al., 2007; Okkinga et al., 2018), and (c) motivation (cf. Souvignier & Mokhlesgerami, 2006) at posttest. With regard to the students with different reading levels, we expect that students with below-average reading levels will benefit most from this intervention (cf. Okkinga et al., 2018; Swanson et al., 2016), resulting in a larger increase in reading comprehension performance compared to average and above-average readers.

Method

Participants

In the school year of 2016–2017, six seventh-grade classrooms from three Dutch secondary schools participated by using a digital learning environment (DLE) to read expository texts. Initially, the sample consisted of 174 students. There was an equal distribution of boys (n = 88) and girls (n = 86). The average age at the start of the intervention was 12.5 years (SD = 0.42). All classrooms in each school consisted of a mixed educational level of general secondary and pre-university education[1]. The current study did not require submission for ethical approval at the local institutional review board, since it already obtained approval from a governmental review board involved in assessing the grant application. Nevertheless, parents or caretakers of all participating students were informed about the research project via a personal letter and were able to refuse the use of their child's data. We did not receive any such statements.

Design

This study investigates the effects of a DLE called 'Gazelle'[2] on students' reading of expository texts (ter Beek, Spijkerboer, Brummer, & Opdenakker, 2018). We developed the DLE, to be used in history and geography courses in secondary education, in collaboration with teachers and (non-participating) students. Three secondary schools volunteered to participate in the intervention. All schools were comparable in terms of gender distribution, educational level, denomination, and average final exam results. We randomly assigned two seventh-grade classrooms of each school to a research condition to ensure that all students within a school would be treated equally. This resulted in a quasi-experimental design with two experimental groups (A and B) and one control group (see Table 2.1).

Students in Experimental group A could consult hints while reading history texts in Gazelle but did not use the program to read geography texts. In Experimental group B, students used Gazelle to read both history and geography texts but were

1 In Dutch secondary education, many schools mix the educational levels of higher general secondary education (havo) and pre-university education (vwo) in seventh and eighth grade to determine the final educational level of a student at a later stage, based on his or her performance during the early secondary years. The higher general secondary education level grants access to higher vocational education, while pre-university education also grants access to university education.

2 Gazelle is a Dutch acronym for 'Gemotiveerd en Actief Zelfstandig Lezen', which roughly translates into 'Motivated and Active Independent Reading'.

Table 2.1 Overview of experimental and control groups

School	*N* students at school level	*M* final result (history)	Group	Hints	Subject(s)	*N* classrooms	*N* students	% girls
A	1364	6.7/10.0	Experimental A	History	History	2	57	52.6
B	1376	6.8/10.0	Experimental B	Geography	History, geography	2	61	52.5
C	1087	6.8/10.0	Control	None	History, geography	2	56	42.9

Note. The average final exam results are based on pre-university students' scores.

only able to consult hints while reading the geography texts. Originally, we intended to apply a cross-subjects design to test the transfer effects of the available support in one subject on the outcomes of the other subject (i.e., geography or history); unfortunately, the geography teachers in Experimental group A decided not to participate in this intervention shortly after the start of the school year. Students in the control group used Gazelle in both subjects but did not receive the opportunity to access hints in either subject (see Table 2.1). The present study focuses on the effects of using Gazelle on the outcomes for the subject of history. In doing so, Experimental group B functions as a separate condition to test the transfer effects of the provision of hints for geography on the outcomes for history.

Procedure

Before the intervention started, students completed a reading assessment to determine their initial reading comprehension level. Additionally, they completed a questionnaire to determine students' initial SRL, focusing on metacognitive strategy use and awareness of reading strategies, as well as a questionnaire on students' motivation (i.e., T1; see Figure 2.1).

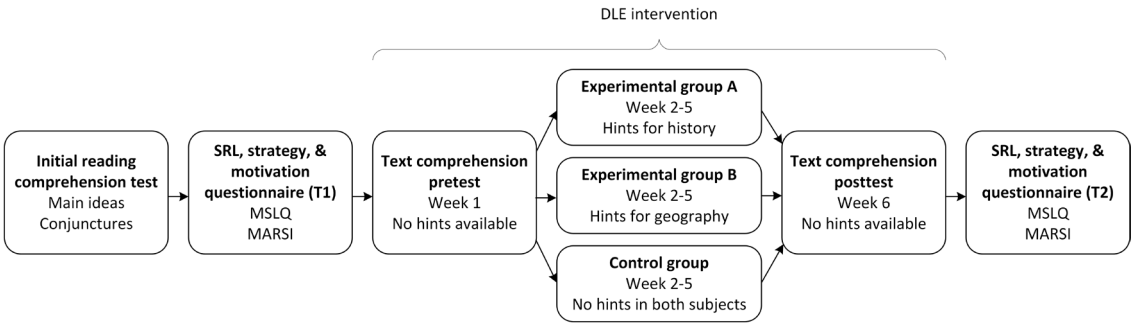


Figure 2.1 Timeline for the study and data collection. SRL = self-regulated learning; MSLQ = Motivated Strategies for Learning Questionnaire; MARSI = Metacognitive Awareness of Reading Strategies Inventory.

During a six-week intervention, carefully scheduled between two school holidays, students weekly read one expository text for each subject. Teachers integrated the use of Gazelle with their conventional lessons, which lasted about 50 minutes. For history, the topic of the texts in Gazelle was Ancient Greece, which was in line with the seventh-grade curriculum. We ensured that the texts in Gazelle presented new information to supplement the regular textbooks, whilst taking into account the

comparability of the contents and difficulty levels. Each text contained approximately 550 words. Students had to summarise the text directly after reading. Consequently, each student had to answer ten text-related multiple-choice questions. Students could continuously view the text on-screen to rule out the potential influence of memorisation. At the end of each lesson, students assessed their work on a scale of 1 to 10 and reflected on their summary.

Students in the experimental groups were able to consult both cognitive and metacognitive hints while reading texts and answering questions. Cognitive hints appeared alongside the multiple-choice questions and presented strategic information about the literal contents of the text (e.g., "A reason can be found after the appearance of words like *because* or *therefore*"), but they did not provide correct answers. Metacognitive hints presented strategic information about students' regulation of their learning process before, during, and after reading (e.g., "Evaluate your own work by focusing on your progress or concentration, instead of focusing on results"). These hints appeared during reading, summarising, and reflecting. The textual contents of the hints only appeared on screen when students deliberately clicked on a lightbulb-shaped button (see Figure 2.2). The hints were static; their contents were equal for all students (Puntambekar & Hubscher, 2005).

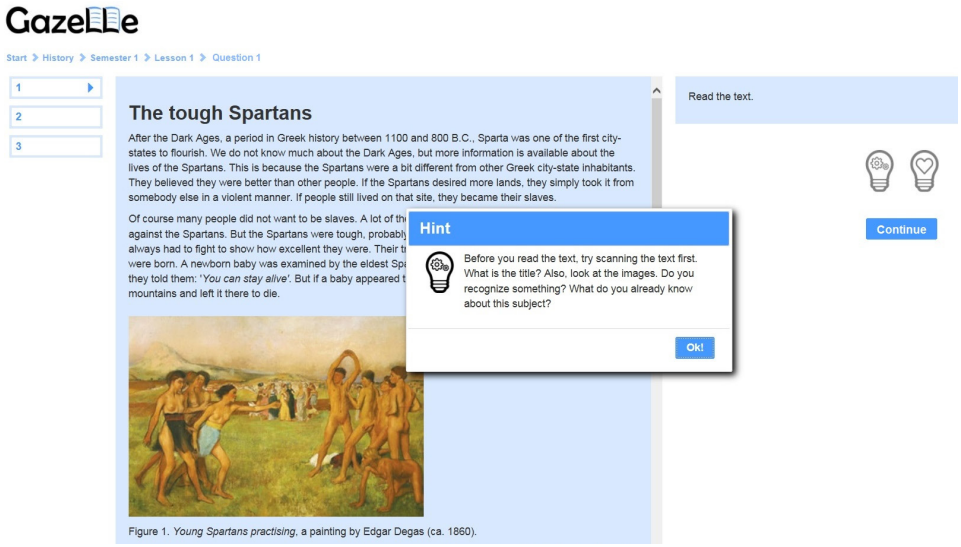


Figure 2.2 Screenshot of the Gazelle-program showing the contents of a metacognitive hint.

Using hints was optional in weeks 2, 3, 4, and 5. During the first and last week of the intervention, none of the students in the different conditions had access to

hints. Therefore, the results of week 1 and week 6 allow us to compare students' text comprehension before and after the intervention. We again administered the SRL and motivation questionnaires after the last week of the intervention (i.e., T2).

Instruments

We adopted three commonly used instruments to measure students' initial reading comprehension, SRL, strategy awareness, and motivation. However, we slightly modified them by translating items from English to Dutch and by adding specific subjects (i.e., '*in my history class*' or '*while reading history texts*'; ter Beek et al., 2018). Prior to the intervention, we discussed the items in two focus groups with seventh-grade students who did not participate in this study to ensure that the items were understandable for this age group.

Initial reading comprehension. Before the start of the intervention, we assessed students' initial reading comprehension levels by using a recognised Dutch reading instrument (Aarnoutse, 1987). The original instrument consists of four subtests: 'main ideas', 'conjunctures', 'synonyms', and 'antonyms'. According to Aarnoutse, the subtests for 'main ideas' and 'conjunctures' relate to higher levels of reading comprehension, such as recognising relationships between parts of the text, whereas 'synonyms' and 'antonyms' relate to vocabulary knowledge (1987). Therefore, we decided to administer only the 'conjunctures' and 'main ideas' subtests in this study (see Figure 2.3 for examples of subtest questions). We updated the old-fashioned language used in the original instrument and shortened the original 'main ideas' subtest from 21 to 8 items due to time constraints and possible overlap with the topics of texts in Gazelle. The original 'conjunctions' subtest consisted of 23 items; we excluded two items that substantially lowered the internal consistency. The final 29 items yielded a Cronbach's α of .63 and a Guttman's λ2 of .65.

Self-regulated learning (SRL). We measured students' SRL using two components of the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, García, & McKeachie, 1991). This instrument is widely used to measure students' metacognitive use of learning strategies across different content areas and student populations (García Duncan & McKeachie, 2005). Metacognitive Self-Regulation (MSR; 10 items) focuses on students' metacognitive processes such as planning, monitoring, and regulating (e.g., "I ask myself questions while reading history texts to check whether I understand the information"). Effort Regulation (ER; 3 items) concerns students' control of effort and attention when faced with difficult

Conjunctures

1. The old man fell down ... broke his leg.

Choose the right option:

a) after
b) and
c) but
d) none of the above

Main ideas

1. A female python can lay almost 100 eggs. A ring snake lays only one egg. Many species of snakes leave the eggs that they have laid, but the cobra remains curled up on the eggs until they hatch. Some snakes give birth to their young alive. They carry the unborn young in their body. When they are born, the mother leaves them immediately and lets them take care of themselves. The first thing these young snakes do is shake their skin off.

Which sentence best describes what the whole story is about?

a) Many snakes lay eggs.
b) Hatching snake eggs.
c) How snakes reproduce.
d) Let young snakes take care of themselves.

Figure 2.3 Examples of 'conjunctures' and 'main ideas' questions from the initial reading comprehension test (Aarnoutse, 1987; translated from Dutch to English). The correct answers are b and c.

or tedious tasks (e.g., "I work hard in history class, even if I don't like what I'm doing"), and is related to students' use of learning strategies (Pintrich et al., 1991). In line with De Boer, Hagenbeek, De Waal, Weening, & Admiraal (2013), we reduced the original seven-point Likert-type scale to increase the comprehensibility for the seventh-graders as well as the comparability with other instruments used in this study. Hence, all items were measured on a five-point Likert scale ranging from 1 (*not true at all for me*) to 5 (*absolutely true for me*).

Because this research focuses on reading comprehension strategies, we also administered the Metacognitive Awareness of Reading Strategies Inventory (MARSI; Mokhtari & Reichard, 2002). This inventory provides insight into students' global, problem-solving, and supportive reading strategy awareness. Global reading strategies (GLOB) are related to a global analysis of text (e.g., "I think about what I already know to help me understand what I read for history"). Problem-solving strategies (PROB) aim at what to do when the text becomes too difficult (e.g., "I try to guess the meaning of unknown words or phrases in history texts"). Support reading strategies (SUP) encompass strategies students use to actively support their own reading process (e.g., "I write summaries to reflect on key ideas in the history

text"). All items were measured on a five-point Likert scale ranging from 1 (*not true at all for me*) to 5 (*absolutely true for me*). Table 2.2 shows the reliability for all SRL components.

Motivation. We measured students' motivation with three components of the MSLQ. Task Value (TV; 6 items) refers to the student's evaluation of how interesting or useful a task or course is (e.g., "I am very interested in the contents of my history course"). Self-Efficacy for learning and performance (SE; 8 items) measures the perceived ability to master a task such as reading textbooks (e.g., "I am confident I can understand the basic concepts taught in my history course"). Intrinsic Goal Orientation (IGO; 4 items) provides an indication of student's involvement for reasons such as challenge, curiosity, or mastery (e.g., "For history I prefer texts that really challenge me so I can learn new things"). Similar to the SRL components, all

Table 2.2 Reliability indicators for SRL and motivation subscales (N_{T1} = 172; N_{T2} = 162)

Scale	*N* items	Cronbach's α (T1)	Guttman's λ2 (T1)	Cronbach's α (T2)	Guttman's λ2 (T2)
SRL					
MSR	10	.84	.84	.85	.85
ER	3	.71	.72	.74	.74
GLOB	13	.80	.81	.89	.89
PROB	8	.75	.76	.82	.82
SUP	9	.77	.78	.81	.82
Motivation					
TV	6	.78	.81	.78	.80
SE	8	.87	.87	.88	.88
IGO	4	.59	.60	.59	.61

Note. SRL = self-regulated learning ; MSR = metacognitive self-regulation; ER = effort regulation; GLOB = global reading strategies; PROB = problem-solving strategies; SUP = support reading strategies; TV = task value; SE = self-efficacy; IGO = intrinsic goal orientation.

items were measured on a five-point Likert scale ranging from 1 (*not true at all for me*) to 5 (*absolutely true for me*). Table 2.2 shows the reliability for the motivation

components. We report both Cronbach's α and Guttman's λ2; although Cronbach's α is commonly used in educational research, Guttman's λ2 is a better reliability estimator according to Drenth and Sijtsma (2005).

The reliability indicators in Table 2.2 are comparable with those of the original instruments, although the scores for TV and IGO are lower than the original alpha values of .90 and .74 (Pintrich et al., 1991). Just like Zepeda et al. (2015), we decided not to further analyse the results for IGO due to the low reliability.

Text comprehension. During the six-week intervention, students weekly answered ten text-related multiple-choice questions. These questions covered relevant reading skills, such as recognising causal relationships (e.g., "How did the Spartans become such good soldiers?") or explaining historical events (e.g., "Explain why the 300 Spartan soldiers went into battle against 10,000 Persians"). All multiple-choice questions of weeks 1 and 6 were comparable in terms of addressing different skills and covering text contents. Students received one point per correct answer, which led to a maximum score of 10 points. Consequently, we used the results on the multiple-choice questions of weeks 1 and 6 as pretest and posttest measures of students' text comprehension.

Hint use. Log-files in Gazelle registered whether students accessed cognitive or metacognitive hints. In weeks 2 through 5, students could access 16 metacognitive hints before, during, and after reading the text and 80 cognitive hints while answering the multiple-choice questions.

Analyses

To calculate initial reading comprehension, we computed overall mean scores for the combined 'main ideas' and 'conjunctures' subtests. For SRL and motivation, we computed a mean score for each subscale (i.e., MSR, ER, GLOB, PROB, SUP, TV, and SE) if a student answered at least 80% of the scale's items. We calculated sum scores for the multiple-choice questions in week 1 (pretest) and week 6 (posttest). Hint use was determined for each hint separately as a dichotomous variable (no use = 0, use = 1) and subsequently aggregated.

We used variance analysis with General Linear Models (GLM), paired samples *t*-tests, and post hoc Bonferroni tests to answer the research questions. All tests were performed as two-sided tests. The use of the terms 'ANOVA' and 'ANCOVA' in the results section refer to the variance analyses with GLM. We report effect sizes using

partial eta squared, or partial η^2, since this is a commonly used measure of effect sizes in the educational research literature (Richardson, 2011). Partial eta squared refers to the magnitude of the effect of the intervention controlled for the covariates, which gives a more realistic impression of the effect of the intervention. We consider effect sizes as small when partial $\eta^2 < 0.06$, medium when $0.06 <$ partial $\eta^2 < 0.14$, and large when partial $\eta^2 > 0.14$ (cf. Cohen, 1988; Zepeda et al., 2015). When comparing two groups, we also report Cohen's *d* as an effect size, for which a value of 0.2 can be considered a small effect, 0.5 a medium effect, and 0.8 a large effect.

Missing values. For the initial reading comprehension test, we excluded the results of six students because they did not execute the test seriously (e.g., their time spent on the test was two standard deviations below average or severe negative outliers). Data were missing for two students who were sick on the day of administration. Therefore, the final sample for the initial reading comprehension test was 166 students. For the SRL and motivation questionnaires on T1, data for two students were missing. Therefore, the final sample on T1 is 172. On T2, data were missing for 12 students (7% of the total sample) due to sickness or classroom migrations. In total, 160 students completed both questionnaires (i.e., T1 and T2). With regard to text comprehension, all students completed the multiple-choice questions for week 1, but data of three students were missing for week 6. Therefore, 171 students completed both the multiple-choice pre and posttest in Gazelle. The number of students who completed all measurements was 155 (89% of the initial sample).

Independent groups and subgroups. To answer RQs 1, 2, and 3, we compared students between and within the three different research conditions: Experimental group A, Experimental group B, and the control group. To analyse of the use of hints, we focused solely on the students in Experimental group A, who were provided with hints while reading history texts ($N = 57$). To examine whether the intervention affected students within the conditions differently, we made a distinction between students based on their results on the initial comprehension test ($N = 166$; $M = 22.56$, $SD = 3.46$). We categorised students who scored below one standard deviation (i.e., 19 points or lower) as 'below-average readers' ($n = 31$); students who scored 20 up to 25 points were categorised as 'average readers' ($n = 102$); and students who scored above one standard deviation (i.e., 26 points or higher) as 'above-average readers' ($n = 33$). We use this distinction to answer RQ4.

Results

Preliminary Analyses

Initial reading comprehension. An ANOVA revealed a significant difference between the three research conditions in initial reading comprehension performance, $F(2, 163) = 11.66$, $p < .001$, partial $\eta^2 = .13$, with the initial comprehension in Experimental group B ($M = 20.90$, $SD = 3.57$) being lower than Experimental group A ($M = 23.54$, $SD = 3.06$) and the control group ($M = 23.37$, $SD = 3.09$). The initial reading comprehension test was only used to distinguish below-average, average, and above-average readers ($N = 166$).

Text comprehension at pretest. An ANOVA revealed a significant difference between the three research conditions for comprehension performance at the pretest, $F(2, 171) = 3.13$, $p = .046$, partial $\eta^2 = .04$, with a lower score for the control group ($M = 6.77$, $SD = 1.84$) compared to Experimental groups A and B ($M = 7.61$, $SD = 1.70$; $M = 7.16$, $SD = 1.85$, respectively). Bonferroni post hoc tests showed that the control group performed significantly lower than Experimental group A, $p = .040$. Because the pretest was more similar to the posttest than the initial reading comprehension test, and the three research conditions significantly differed from each other, we decided to include the pretest performance as a covariate in further analyses ($N = 171$).

Hint use. Out of the 57 students in Experimental group A, 30 students used a cognitive or metacognitive hint at least once (i.e., the 'hint users'); 27 students did not use any hints (i.e., the 'non-hint users'). These 30 students used a total of 156 cognitive hints and 30 metacognitive hints; an average of 3.26 hints per student (see Table 2.3). Out of these 30 students, nine students only used a single hint. The average number of hints used decreased throughout the intervention. Since students' decision to use hints (or not) can be affected by a variety of factors, we approached the analysis in two ways. The first approach stresses students' deliberate decision and, thus, if students opened at least one of the available hints during the entire intervention they were categorised as 'hint users' ($n = 30$) and compared to non-hint users ($n = 27$). However, since accessing only a single hint could also have been caused by curiosity rather than an actual need for help, the second approach also stresses the use of more than one hint, and thus, we categorised the students as those who accessed multiple hints ($n = 21$), a single hint ($n = 9$), and no hints ($n = 27$). Students who accessed

metacognitive hints always accessed cognitive hints as well; therefore, we were not able to compare different types of hint users with regard to the contents of the hints.

Table 2.3 Average number of hints used per student in Experimental group A (*N* = 57)

Hints	Week 2 *M* (*SD*)	Week 3 *M* (*SD*)	Week 4 *M* (*SD*)	Week 5 *M* (*SD*)	Total *M* (*SD*)
Cognitive	1.72 (2.72)	0.53 (1.30)	0.26 (0.96)	0.23 (0.73)	2.74 (4.39)
Metacognitive	0.56 (0.82)	0.30 (0.65)	0.14 (0.44)	0.09 (0.29)	0.53 (1.18)
Total	2.28 (3.31)	0.82 (1.75)	0.40 (1.24)	0.32 (0.81)	3.26 (5.21)

Difficulty of multiple-choice questions in weeks 1 and 6. Analysis of the mean scores on each of the ten multiple-choice questions showed that week 6 included relatively more difficult questions than week 1, leading to an decline in performance ($M = 3.35$, $SD = 1.62$). We corrected the scores of week 6 using an equation procedure similar to a method frequently used in Dutch national final exams (College voor Toetsen en Examens & Cito, 2011). Based on the cumulative frequencies of the scores on the pre and posttest, we concluded that we had to add 3.2 points to the posttest sum scores to provide a more representative impression of students' performance. The corrected posttest sum scores were used in all analyses.

Effects on Text Comprehension (RQ1)

Experimental vs. control conditions. An ANCOVA with pretest comprehension performance as a covariate showed no significant differences in posttest comprehension performance between the three research conditions, $F(2, 167) = 1.39$, $p = .252$, partial $\eta^2 = .02$. Paired samples *t*-tests showed that the posttest comprehension performance of all groups significantly declined (see Table 2.4).

Hint use. With 'hint use' operationalised as a deliberate decision to use at least one hint during the intervention, a comparison of hint users and non-hint users showed no significant difference in comprehension performance on the pretest, $F(1, 55) < .01$, $p = .948$, partial $\eta^2 < .01$, indicating that the hint users were not mainly low or high performers. However, there was a significant difference in posttest performance in favour of the hint users ($M = 6.68$, $SD = 1.39$) versus the non-hint users ($M = 5.90$,

Table 2.4 Mean pre and posttest scores on reading comprehension by group/subgroup (*N* = 171)

Group/*subgroup*	*n*	Pretest *M* (*SD*)	Posttest *M* (*SD*)
Experimental A	57	7.61 (1.70)	6.31 (1.43) ***
Experimental B	60	7.17 (1.87)	6.38 (1.41) **
Control	54	6.78 (1.87)	5.88 (1.55) **
Within Experimental A:			
hint users (1 or more)	30	7.60 (1.81)	6.68 (1.39) *
non-hint users (0 hints)	27	7.63 (1.60)	5.90 (1.39) ***
multiple-hint users (2>)	21	7.86 (1.59)	6.71 (1.45) *
single-hint users (1 hint)	9	7.00 (2.24)	6.60 (1.33)
non-hint users (0 hints)	27	7.63 (1.60)	5.90 (1.39) ***

Note. The significance represents within-group comparisons, * $p < .05$. ** $p < .01$. *** $p < .001$. The italicised subgroups refer to comparisons between two different operationalisations of hint users.

SD = 1.39), $F(1, 55) = 4.46$, $p = .039$, partial $\eta^2 = .08$, $d = .56$. Additionally, hint users descriptively performed better on the posttest than students in Experimental group B and the control group (see Table 2.4).

Similar analyses with the operationalisation of 'hint use' including the use of multiple hints (i.e., comparing multiple hints, a single hint, or no hints) also showed no significant differences in pretest performance. There was a small difference in posttest performance in favour of the multiple-hint users ($M = 6.71$, $SD = 1.45$) and the single-hint users ($M = 6.60$, $SD = 1.33$) versus the non-hint users ($M = 5.90$, $SD = 1.39$); however, these differences were not significant, $F(1, 54) = 2.22$, $p = .119$, partial $\eta^2 = .08$. The decline in performance was significant for all students with 'hint use' operationalised as a deliberate decision, comparing 'hint users' and 'non-hint users'. However, when operationalising 'hint use' as also including the use of more than one hint—comparing multiple hints, a single hint, or no hints—the decline was not significant for the single-hint users (see Table 2.4).

Effects on SRL (RQ2)

Experimental vs. control conditions. A comparison between the three research

conditions yielded no significant differences on all SRL scales (i.e., MSR, ER, GLOB, PROB, and SUP) at T1 and T2. However, when comparing T1 with T2, the PROB scale increased for all conditions, but this increase was only significant for Experimental groups A and B (see Figure 2.4 and Table 2.5). GLOB and SUP significantly decreased in the control group ($p < .001$ and $p = .003$, respectively); SUP also significantly decreased in Experimental group B ($p = .012$).

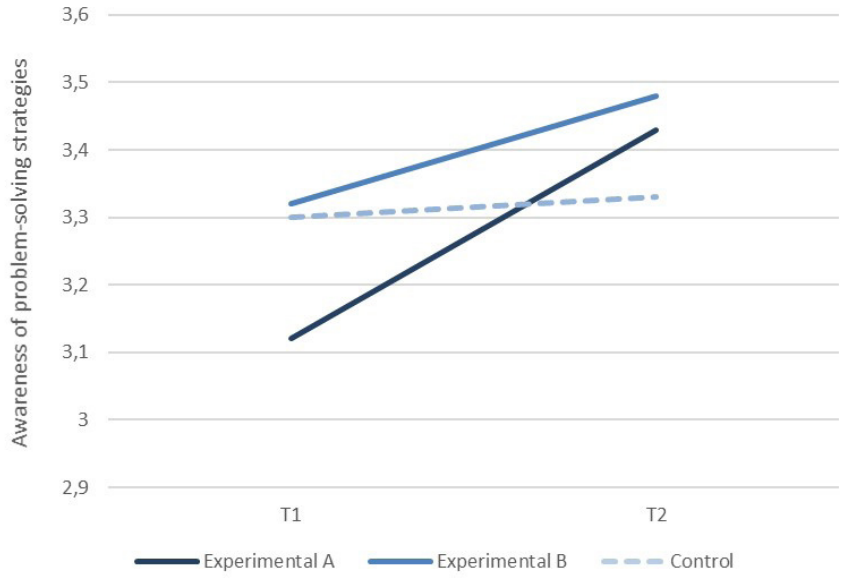


Figure 2.4 Mean problem-solving strategies at T1 and T2 by group for the subject of history (cf. ter Beek et al., 2018).

Hint use. With 'hint use' operationalised as a deliberate decision to use at least one hint during the intervention, ANOVA analyses comparing the SRL scales of hint users and non-hint users yielded no significant differences at T1 and T2. However, paired samples *t*-tests comparing the SRL scales at T1 and T2 showed a significant increase in PROB, $p < .001$, for both hint users and non-hint users. With the operationalisation of 'hint use' also including the use of multiple hints, the analyses also did not yield significant differences between the three groups at T1 and T2. However, both the multiple-hint users and the non-hint users showed a significant increase in PROB, $p = .002$ for multiple-hint users and $p < .001$ for non-hint users. Additionally, single-hint users showed a significant increase in MSR, $p = .012$.

Effects on Motivation (RQ3)

Experimental vs. control conditions. The mean score on TV at T1 was significantly higher for the control group compared to Experimental group A, $F(2, 170) = 4.45$, $p = .013$, partial $\eta^2 = .05$. In addition, the mean score on SE at T1 was significantly higher for the control group compared to Experimental group A and B, $F(2, 169) = 9.95$, $p < .001$, partial $\eta^2 = .11$. An ANCOVA with the mean scores at T1 as a covariate yielded no significant results between conditions on both motivation scales at T2. When comparing T2 with T1, TV and SE decreased for all groups, but not significantly (see Table 2.5). Therefore, the provision of hints did not result in significant changes in motivation of all groups.

Hint use. With 'hint use' operationalised as a deliberate decision to use at least one hint during the intervention, ANOVA analyses comparing hint users and non-hint users on TV and SE showed no significant differences at T1 and T2. This also indicates that students' motivation did not influence their hint use. Additionally, *t*-tests comparing T1 with T2 for hint users and non-hint users showed no significant differences. With the operationalisation of 'hint use' also including the use of multiple hints, we also found no significant differences for TV and SE.

Students with Different Reading Levels (RQ4)

Text comprehension. There was a significant difference between the pretest comprehension performance of below-average readers, average readers, and above-average readers, $F(2, 163) = 4.58$, $p = .012$, partial $\eta^2 = .05$. Bonferroni post hoc tests showed that above-average readers performed significantly better than below-average and average readers, $p = .013$ and $p = .043$, respectively. An ANCOVA with pretest comprehension performance as a covariate and Bonferroni post hoc testing showed no significant difference in posttest comprehension performance for the three reader types. Paired samples *t*-tests showed that the performance of average and above-average readers significantly declined, $p < .001$. However, the decrease was not significant for below-average readers, $p = .112$ (see Table 2.6). Below-average readers even outperformed average readers on the posttest.

Table 2.5 Mean T1 and T2 scores on SRL and motivation scales by condition ($N = 160$)

	Experimental group A ($n = 56$)		Experimental group B ($n = 58$)		Control group ($n = 46$)	
Scale	*M* (*SD*) T1	*M* (*SD*) T2	*M* (*SD*) T1	*M* (*SD*) T2	*M* (*SD*) T1	*M* (*SD*) T2
SRL						
MSR	3.06 (0.50)	3.20 (0.51)	3.09 (0.51)	3.16 (0.51)	3.17 (0.80)	3.12 (0.72)
ER	3.58 (0.62)	3.58 (0.61)	3.55 (0.68)	3.61 (0.61)	3.91 (0.64)	3.76 (0.74)
GLOB	3.27 (0.48)	3.25 (0.47)	3.23 (0.46)	3.32 (0.50)	3.46 (0.66)	3.14 (0.73) ***
PROB	3.12 (0.51)	3.43 (0.49) ***	3.32 (0.47)	3.48 (0.52) **	3.30 (0.74)	3.33 (0.72)
SUP	3.12 (0.52)	3.02 (0.54)	3.06 (0.49)	3.24 (0.54) *	3.30 (0.71)	3.04 (0.76) **
Motivation						
TV	3.26 (0.69)	3.15 (0.57)	3.34 (0.54)	3.29 (0.55)	3.67 (0.57)	3.60 (0.71)
SE	3.53 (0.46)	3.52 (0.53)	3.40 (0.59)	3.37 (0.46)	3.84 (0.43)	3.81 (0.48)

Note. SRL = self-regulated learning; MSR = metacognitive self-regulation; ER = effort regulation; GLOB = global reading strategies; PROB = problem-solving strategies; SUP = support reading strategies; TV = task value; SE = self-efficacy. The significance represents within-group comparisons of T1 and T2, * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 2.6 Mean pre and posttest scores on reading comprehension by reader type ($N = 164$)

Type	*n*	Pretest *M* (*SD*)	Posttest *M* (*SD*)
Below-average readers	30	6.87 (1.81)	6.23 (1.50)
Average readers	101	7.25 (1.65)	6.07 (1.43) ***
Above-average readers	33	8.06 (1.52)	6.61 (1.44) ***

Note. The significance represents within-group comparisons, * $p < .05$. ** $p < .01$. *** $p < .001$.

Self-regulated learning (SRL). There were no significant differences between the three reader types at T1 for all SRL scales, indicating that although their initial comprehension performance varied, their SRL did not. Furthermore, there were no significant differences between the reader types at T2. However, paired samples *t*-tests showed that there was a significant increase on MSR and PROB for average readers, $p = .002$ and $p < .001$, respectively. There were no changes in any of the SRL scales for below-average readers. There was a significant decrease in SUP for above-average readers, $p = .037$ (see Table 2.7).

Motivation. A comparison of the three reader types indicated no significant differences in motivation at T1 and T2. Paired samples *t*-tests showed that SE significantly decreased for below-average readers, $p = .043$. There were no significant differences in motivation for average readers. However, TV decreased significantly for above-average readers, $p = .011$ (see Table 2.7).

Interaction effects. To test whether the provision of hints during expository history text reading had differential effects on comprehension performance for the three reader types, we tested for possible interaction effects. Levene's Test indicated there was no violation of the assumption of equal error variances, $F(8, 155) = 0.82$, $p = .588$. An ANCOVA with reader type and experimental condition as predictors, experimental condition×reader type as the interaction term, and pretest performance as a covariate yielded no significant difference in posttest performance between the conditions for any of the three reader types, $F(4, 154) = 0.44$, $p = .777$, partial $\eta^2 = .01$. Similar analyses for the SRL and motivation scales also yielded no significant results.

Hint use. To test whether the deliberate use of hints (i.e., using at least one hint) during expository history text reading had differential effects on comprehension

Table 2.7 Mean T1 and T2 scores on SRL and motivation scales by reader type ($N = 155$)

	Below-average readers ($n = 30$)		Average readers ($n = 93$)		Above-average readers ($n = 32$)	
Scale	*M* (*SD*) T1	*M* (*SD*) T2	*M* (*SD*) T1	*M* (*SD*) T2	*M* (*SD*) T1	*M* (*SD*) T2
SRL						
MSR	3.20 (0.60)	3.13 (0.53)	3.01 (0.57)	3.19 (0.59) **	3.20 (0.64)	3.09 (0.59)
ER	3.80 (0.61)	3.67 (0.60)	3.59 (0.71)	3.61 (0.69)	3.80 (0.59)	3.70 (0.62)
GLOB	3.29 (0.59)	3.29 (0.55)	3.29 (0.48)	3.23 90.55)	3.35 (0.64)	3.18 (0.64)
PROB	3.28 (0.56)	3.37 (0.53)	3.18 (0.56)	3.44 (0.59) ***	3.30 (0.65)	3.39 (0.61)
SUP	3.07 (0.65)	3.19 (0.59)	3.13 (0.50)	3.11 (0.61)	3.21 (0.65)	2.98 (0.67) *
Motivation						
TV	3.51 (0.58)	3.40 (0.53)	3.33 (0.62)	3.32 (0.60)	3.50 (0.69)	3.26 (0.82) *
SE	3.56 (0.53)	3.39 (0.59) *	3.53 (0.56)	3.58 (0.50)	3.72 (0.45)	3.63 (0.53)

Note. SRL = self-regulated learning; MSR = metacognitive self-regulation; ER = effort regulation; GLOB = global reading strategies; PROB = problem-solving strategies; SUP = support reading strategies; TV = task value; SE = self-efficacy. The significance represents within-group comparisons of T1 and T2, * $p < .05$. ** $p < .01$. *** $p < .001$.

performance for the three reader types, the interaction effects between reader type and the hint users vs. non-hint users within Experimental group A were investigated. Levene's Test indicated there was no violation of the assumption of equal error variances, $F(5, 50) = 1.22$, $p = .312$. An ANCOVA with reader type and hint use as predictors, hint use×reader type as the interaction term, and pretest performance as a covariate yielded no significant difference in posttest performance between hint users and non-hint users for any of the three reader types, $F(2, 49) = 0.51$, $p = .606$, partial $\eta^2 = .02$. Similar analyses for the SRL and motivation scales also yielded no significant results. Finally, similar analyses with 'hint use' operationalised as also including the use of multiple hints did not yield significant differences in posttest performance, SRL, or motivation for any of the three reader types.

Discussion

Research has shown that instructional support in DLEs can have a positive effect on students' reading comprehension and academic performance (Cheung & Slavin, 2012; Lysenko & Abrami, 2014; Moran et al., 2008). This study added elements of autonomy and self-regulation to a DLE: students were able to decide whether and when to use cognitive and metacognitive support during history text reading. As such, the present study also addressed the possible effects of hint use on students' SRL and motivation.

Summary of Findings

Regarding text comprehension (i.e., RQ1), results showed no significant differences in posttest comprehension performance between the three conditions. A possible explanation might be that, in general, students in Experimental group A hardly used hints. In essence, students who did not use hints were identical to students in the control group, making it hard to compare them. However, we did find a significant difference in posttest comprehension performance in favour of the operationalisation of 'hint users' as students who deliberately used one or more hints during the intervention compared to students who did not use hints. Analyses in which 'hint users' were operationalised otherwise—as students who accessed *single* or *multiple* hints—did not yield any significant results, but effect sizes (partial $\eta^2 = .08$) were similar for both operationalisations of hint users. Thus, our expectation regarding hint users outperforming non-hint users was partially confirmed. We will discuss and reflect on the findings with regard to 'hint users' in the following section using

the operationalisation of students who used one or more hints unless we explicitly state this otherwise.

The finding that hint users outperformed non-hint users (albeit only with a specific operationalisation of these groups) is in line with earlier research studies, in which students in the experimental conditions who were provided with strategy instruction outperformed students who did not receive such instruction (Berthold et al., 2007; Mason, 2013; McKeown et al., 2009; Ramsay et al., 2010; Souvignier & Mokhlesgerami, 2006; Vaughn et al., 2013, 2015; Zepeda et al., 2015). However, implications of our findings with regard to performance must be considered carefully, given the fact that performance scores declined significantly in all groups throughout the intervention. This probably has more to do with students' decline in motivation than with the provision and use of hints.

There were no significant differences between conditions or between hint users with regard to SRL and strategy awareness (i.e., RQ2); this expectation was not met. Nevertheless, students in Experimental group A and B, who were provided with hints during expository history or geography text reading, significantly increased their levels of problem-solving strategy awareness, indicating that this practice might be useful for students' SRL—irrespective of whether students actually use the hints. While reading expository texts, students in the experimental groups could decide for each question whether they wanted to use a supportive hint. This option may have stimulated students' problem-solving strategies, since some students preferred to figure out the problem without using the hint; some students even expressed that they considered using hints as 'cheating'. This corresponds with the work of Roll, Baker, Aleven, and Koedinger (2014), who state that avoiding help is sometimes "associated with *better* performance than seeking help on steps for which students have low prior knowledge" (pp. 537–538, italics in the original). In contrast, awareness of global and support reading strategies significantly decreased for the control group, indicating that the lack of support might have had a negative influence on students' overall reading strategy awareness.

Regarding students' motivation (i.e., RQ3), there were no significant differences for the different conditions and hint users; thus, our expectations were not confirmed. Task value and self-efficacy decreased throughout the intervention in all groups, indicating that student motivation for the history course in general diminished over time, albeit not significantly. Teachers also stressed the low motivation levels of their students, since the average time spent on reading texts in the DLE declined

in all conditions. The decrease in students' motivation probably influenced students' posttest reading comprehension performance, which was unexpectedly lower than their pretest performance. Unfortunately, empirical research regarding effective digital support on intrinsic reading motivation is scarce, especially in secondary education (Moran et al., 2008; van Steensel et al., 2016). Many existing reading interventions focus on motivation in terms of competence, social aspects, and rewards, whereas only a few address the value of reading (van Steensel et al., 2016). Although research has shown that relatively short, two to four-week interventions yield larger effect sizes in this type of research (cf. Moran et al., 2008), it seems that the repetitive character of our six-week intervention fostered reluctance to work with the DLE, and, consequently, had a negative effect on students' general motivation for history. This relates to the findings of Azevedo, Cromley and Seibert (2004), who found less stated interest in students when they were scaffolded with domain-specific guiding or hints.

With regard to students with different reading levels (i.e., RQ4), it seems that practising in the DLE had diverse effects. First, there is a discrepancy between reading comprehension and self-efficacy of below-average readers. In contrast to the average and above-average readers, below-average readers' performance did not significantly decline during the intervention; thus, our expectation was partially confirmed. This corresponds to earlier research studies focusing on struggling students or students with learning disabilities in social studies text interventions (Swanson et al., 2014, 2016). Below-average readers even descriptively outperformed average readers on the posttest. However, their self-efficacy beliefs significantly decreased over time. This might be due to the direct feedback on each multiple-choice question: Multiple indications of incorrect answers confronted below-average readers with their lack of reading comprehension, which in turn might have lowered their self-efficacy beliefs. Second, above-average readers' task value and reading comprehension performance declined significantly. Above-average readers possibly felt no need to practise their reading, provoking decreased motivation, metacognitive self-regulation, and effort regulation and in turn leading to lower performance on the posttest.

Limitations

The main limitation of the present study is students' use of hints in general. We must treat findings for 'hint users' versus 'non-hint users' with caution: out of the 57 students who could use hints during history text reading, only 30 used a hint

at least once, and nine students only opened a single hint during the intervention. Given the low number of hints used, we dichotomised the hint use variable, but this precluded the opportunity to analyse hint use more extensively. Moreover, a different operationalisation of hint users (i.e., with three groups) did not yield any significant differences; this is probably related to a power problem due to the low number of students in the single-hint user group. Nevertheless, an interesting finding was the fact that single-hint users' MSR increased significantly throughout the intervention. Paradoxically, the increased self-regulation of these students is probably invoked by not using more than a single hint. Therefore, it might be useful to uncover why individual students did or did not use the cognitive and metacognitive hints in more detail.

Although hint users initially did not differ from non-hint users in terms of task value and self-efficacy, it seems likely that intrinsic motivation is related to hint use. Additionally, the students in this study might not have had sufficient metacognitive knowledge or metacognitive skills to decide whether they needed a hint or not. Even if they did use hints, it should not be assumed that they were able to use the strategic information offered by the hint effectively (Aleven & Koedinger, 2000; Azevedo, Moos, Greene, Winters, & Cromley, 2008; de Kock, 2016). Furthermore, it is possible that the provision and use of hints in the DLE increased students' cognitive load, since the hints contained even more text to read (cf. Berthold, Röder, Knörzer, Kessler, & Renkl, 2011; Kirschner, 2002). Future research should also include qualitative research data, such as student interviews or trace data, to be able to explain the findings with regard to students' hint use and strategy awareness in more detail.

Another limitation of this study is the fact that students in Experimental group A only used the DLE for reading history texts, whereas both other groups used it for both history and geography lessons. Unfortunately, the geography teachers of Experimental group A unexpectedly decided not to participate in the intervention. Our current design presupposed the use of the DLE in an ecologically valid context; however, it was also prone to challenges in the case of classroom or teacher attrition. Moreover, students from Experimental group B and the control group, who read texts for both subjects, complained about the density and repetitive character of the DLE. Satiation or boredom with the initially new programme might have resulted in lower motivation, effort regulation, or strategy use.

In the present study, students worked by design independently in the DLE without any help or instruction from their teacher. However, Azevedo et al. (2008)

discovered that externally facilitated learning, in which teachers have an active role in guiding students' SRL, leads to higher knowledge gains and more effective metacognitive strategy use by students. This lack of guidance might have resulted in lower and ineffective hint use, which could explain the decline in comprehension performance for hint users. It is probably helpful to train teachers in using the DLE in their classroom context to stimulate the text comprehension, strategy use, and motivation of their students.

A final limitation of this study is that we did not include students' vocabulary and background knowledge in our analyses; two factors that contribute to text comprehension in adolescent readers (Cromley & Azevedo, 2007). By focusing on students' ability to generate correct main ideas and conjunctions, we might have overlooked basic vocabulary skills and knowledge as important prerequisites for comprehending expository history texts. Therefore, the results regarding the reader types (below-average, average, and above-average readers) should be interpreted with caution. Future research on expository text reading should address multiple components of reading comprehension to obtain a more comprehensive picture of what influences students' comprehension performance during reading interventions.

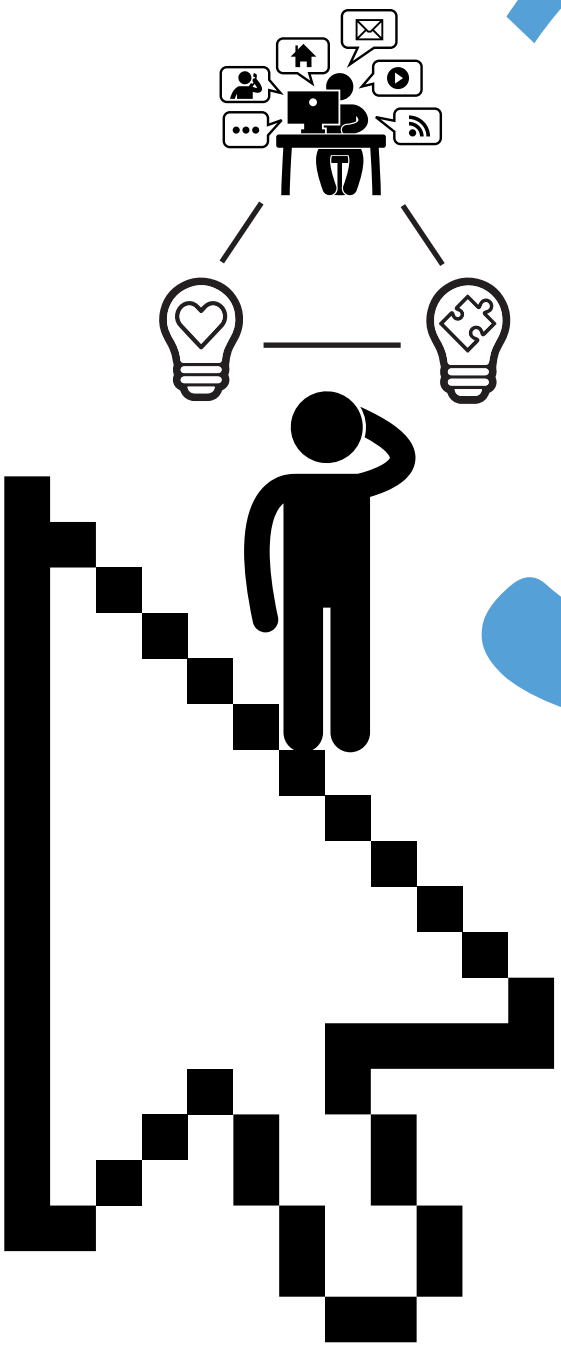
Practical and Scientific Implications

Practising expository text comprehension in a DLE, with or without reading strategy support, can improve below-average readers' text comprehension. However, for above-average readers, this practice might be detrimental for their motivation for the subject of history. Therefore, we suggest that teachers carefully consider which students can benefit from digitally supported reading practice. Moreover, not all students in this study who had the option to use hints also made use of the available support. It should not be expected that using an apparently rich and complex learning environment automatically results in more self-regulated use of hints. Therefore, teachers need to ensure that students consider help seeking as 'normal' and stimulate their students to use the support offered—regardless of the specific DLE used in this research. Teachers should also not assume that all seventh-grade students already possess the necessary self-regulated learning skills for reading expository texts, or that students are fully aware of when and how to apply relevant reading strategies.

This research shows that the use of a DLE with integrated strategy instruction can be beneficial to provide differentiated practice for students who struggle with reading their textbooks. On a scientific level, this research complements the

existing knowledge about the use of computer-supported or digital learning in educational practice, whilst also highlighting the possible challenges posed by this type of practical research. In addition, it provides a good example of the impact of methodological decisions on the outcomes, such as the operationalisation of subgroups. Notwithstanding these challenges, it is of continued importance to keep up with the rapid technological innovations of the 21st century by analysing 'what works' in education, and thereby to ensure that the use of technology in the classroom contributes to the development of individual students.

3



Chapter 3

Using learning analytics and latent profile analysis to explore the relations between reading engagement, motivation, and comprehension

This chapter has been submitted for publication as:

ter Beek, M., Opdenakker, M.-C., Deunk, M. I., & Strijbos, J. W. (2019). *Using learning analytics and latent profile analysis to explore the relations between reading engagement, motivation, and comprehension.*

Abstract

This study investigates how real-time reading engagement in a Digital Learning Environment (DLE), motivational aspects of reading, and expository text comprehension are related. Seventh-grade students read six history texts in a DLE, which recorded log file data related to their behavioural and cognitive engagement. Consequently, these log file data were used to identify engagement profiles using latent profile analysis. Five identified profiles were compared in terms of students' task value, self-efficacy, intrinsic motivation, and text comprehension. Results from this learning analytics approach show that highly engaged students initially have significantly higher task value and intrinsic motivation compared to students who show little engagement. Likewise, highly engaged students show better text comprehension. Although these results seem promising, it is important to note that the majority of students scored relatively low on all engagement, indicating that there is room for improvement in (fostering) students' engagement when using digital technology to read texts.

4

schools

325

students

3

research questions

Highlights

- Combining learning analytics and LPA can provide useful insights in students' real-time engagement when using technology for reading texts.
- Students who are highly engaged also show high levels of task value and intrinsic motivation.
- The more engaged a student works in a DLE, the better his or her reading performance is expected to be.

Introduction

Academic success in secondary education is, among others, influenced by the interplay between students' reading motivation, engagement, and comprehension, since reading texts is essential for almost every subject (Anmarkrud & Bråten, 2009; Guthrie, Wigfield, Metsala, & Cox, 1999; Morgan & Fuchs, 2007; Retelsdorf, Köller, & Möller, 2011; Taboada, Tonks, Wigfield, & Guthrie, 2009). To study and understand the information provided in their textbooks, students have to be motivated to read and have to be actively engaged in their reading process. This is especially the case for subjects like history, for which students often have to read broad, fact-dense expository texts (Mastropieri, Scruggs, & Graetz, 2003). Recently, the concept of student engagement has been studied extensively in educational research, for example in the field of reading research (Guthrie & Wigfield, 2017) and in research on the use of educational technology (Rashid & Asghar, 2016).

Digital Learning Environments (DLEs) provide a powerful, yet challenging way to examine students' cognition, metacognition, motivation, and engagement (Azevedo & Gašević, 2019; Azevedo et al., 2013). Over the past few years, DLEs have been improved with possibilities to collect and translate data to detect, analyse, and foster students' learning (Bouchet, Harley, Trevors, & Azevedo, 2013; Azevedo & Gašević, 2019). Methods such as educational data mining and learning analytics provide the opportunity to determine and examine students' learning processes through log file data and, subsequently, to adapt the instructional support to suit students' individual needs. However, there is an ongoing debate about the academic benefits of students' engagement with technology in education, and the research literature on this subject includes studies reporting positive effects as well as studies reporting negative or no effects (Rashid & Asghar, 2016). To contribute to this research field, the current study explores the relations between adolescent students' engagement in a DLE and their motivation and performance in the context of reading comprehension.

Motivation, Engagement, and Reading Comprehension

There is scientific consensus about the existence of a relationship between reading motivation, engagement, and reading performance (Guthrie & Klauda, 2016; Guthrie, Klauda, & Ho, 2013; Guthrie & Wigfield, 2017). A recent study by Wolters, Barnes, Kulesz, York, and Francis (2017) specifically examined the relation between reading motivation and reading comprehension performance among ninth-grade students.

The authors argue that "adolescents' engagement and performance at reading tasks are tied to the motivational beliefs and attitudes they have about reading for school" (p. 99). Schiefele et al. (2012) extensively reviewed several dimensions of reading motivation and their relation to reading behaviour and reading competence. They found that students' intrinsic motivation to read positively contributes to reading skills and comprehension. However, the causal role of reading motivation and the mediating role of reading behaviour in students' reading competence remained unclear. Guthrie and Wigfield (2017) recently presented an updated version of their conceptual engagement model of reading development. Based on this model, it is expected that classroom instruction influences students' reading motivation and cognition, which then leads to individual differences in students' engagement and, consequently, in their reading achievement.

Motivation. In the educational research literature, motivation is often regarded as an essential aspect of students' learning (Pintrich & De Groot, 1990; Winne & Hadwin, 2008). Students' motivation can refer to motivation for a subject in general as well as for a specific task within that subject, such as reading. Following this line of thought, a student who enjoys the subject of history is more likely to invest time and effort in a reading task for history than a student who thinks history is boring, regardless of the contents of the history texts. Guthrie and Wigfield (2000) extensively studied motivation in the context of reading, and define reading motivation as "the individual's personal goals, values and beliefs with regard to the topics, processes and outcomes of reading" (p. 406). Students' motivation comprises several distinct but related aspects, such as value, self-efficacy, and intrinsic motivation (Schiefele, Schaffner, Möller, & Wigfield, 2012; Guthrie & Wigfield, 2017).

Task value, self-efficacy beliefs, and intrinsic motivation are known to contribute to students' reading motivation and performance (Retelsdorf et al., 2011; Taboada et al., 2009; Unrau & Schlackman, 2006). Task value refers to students' perceived usefulness of a task or subject, or the belief that a (reading) task is useful and beneficial (Guthrie & Wigfield, 2017). The concept of self-efficacy entails students' perceived ability to be successful in future tasks (Bandura, 1982), for example, confidence of one's ability to read and understand texts (Guthrie & Wigfield, 2017). Lastly, intrinsic motivation encompasses students' perceived interest and enjoyment, for example when reading texts. In the context of reading comprehension, these aspects of motivation and their relation to academic performance may vary between students (Guthrie & Klauda, 2016). However, in general, research has shown a decline in students' intrinsic

motivation for content area reading around the time when students transition from primary to secondary education (Guthrie & Davis, 2003), which is also apparent in the Dutch educational context (Gubbels, Netten, & Verhoeven, 2017).

Engagement. According to Guthrie and Wigfield (2017), intrinsic motivation, self-efficacy, and value "are motivations that drive the engagement that flows out of them" (p. 58). However, educational engagement seems to be a difficult concept to grasp (Azevedo & Gašević, 2019; Fredricks, Blumenfeld, & Paris, 2004). Previous research on student engagement included measurements of (among others) students' effort, involvement, active participation, commitment, affect, enthusiasm, or persistence, resulting in a fuzzy construct. Fredricks et al. (2004) distinguished three main aspects of engagement: behavioural, cognitive, and emotional engagement. Behavioural engagement focuses on elements like time spent on a task, whereas cognitive engagement is related to the quality of processing learning content, like the use of strategy support. Students' emotional engagement, which encompasses positive and negative reactions to teachers, classmates, and school itself, is beyond the scope of the current study.

Reading comprehension. The goal of reading a text is to comprehend its contents, and in order to comprehend a text, a reader must be able to construct a mental representation of what has been written, also known as a situation model of the text (Kintsch & Rawson, 2005). In the reading engagement model by Guthrie and Wigfield (2017), reading comprehension is one of the aspects of the general concepts of reading achievement, together with reasoning, fluency, decoding, and phonemic awareness. Research has shown that these elements of reading achievement continually develop throughout a student's academic career (Alexander, 2005). For students who transition from primary to secondary education, the ability to comprehend lengthy expository texts, for example by distinguishing main ideas from irrelevant details, becomes increasingly important.

Adopting a Person-Centred Approach

The aforementioned consensus about the relationship between reading motivation, engagement, and performance is based on studies that typically adopt a variable-centred approach, using (group) mean scale or item scores as part of structural equation modelling or regression analyses. Many studies report positive correlations between measures of reading motivation and reading amount or comprehension, and the relationship can be mediated by behavioural engagement (De Naeghel, Van Keer,

Vansteenkiste, & Rosseel, 2012; Guthrie et al., 1999; Taboada et al., 2009; Wigfield et al., 2008). To complement results from the variable-centred approach and to identify individual student differences or different groups of individuals, the person-centred approach received more attention over the last years (Marsh, Lüdtke, Trautwein, & Morin, 2009). With the person-centred approach, it is possible to distinguish different learner profiles and to classify students as distinct learning types, (Flunger et al., 2015, 2017), providing teachers with the opportunity to differentiate their instruction according to various student needs.

An increasingly common way to adopt a person-centred approach is to cluster continuous data using latent profile analysis (LPA). LPA, which is a model-based type of cluster analysis, enables researchers to cluster homogeneous subgroups of individual students from a heterogeneous sample, such as students with similar patterns of characteristics. Its application is relevant for the educational research field, because it recovers hidden groups from observed data, and, thus, provides researchers and teachers with the opportunity to take into account individual or group differences in students' characteristics and learning processes (Hickendorff, Edelsbrunner, McMullen, Schneider, & Trezise, 2018). Compared to more traditional clustering methods, LPA is advantageous in the sense that the number of clusters can be determined based on statistical tests and goodness-of-fit indices, which leads to a better model fit.

Schiefele and Löweke (2018) adapted a person-centred approach using LPA with regard to motivation for recreational reading of elementary students in grades 3 and 4. Results showed that the profile with high levels of intrinsic motivation outperformed the low-intrinsic motivation profile on measures of reading comprehension. The authors mention that the use of LPA in reading motivation research remains scarce, especially in secondary education. With regard to secondary students' engagement, LPA was applied in studies concerning homework time and effort (Flunger et al., 2015, 2017) and engagement (van Rooij, Jansen, & van de Grift, 2017). After identifying four to five student profiles, results showed that higher levels of homework time and effort or academic engagement were positively related to students' academic performance (Flunger et al., 2015; van Rooij et al., 2017). Both studies used self-report measurements to establish the predictor variables that formed the basis of the LPA.

More recently, LPA also has been used in studies with regard to digital or online learning environments, such as the study by Tze, Daniels, Buhr, and Le (2017) on

the relationship between students' affective profiles and online engagement. By using LPA, they identified different affective profiles and found that positive measures of affect were associated with increased student engagement. However, they did not include actual usage data in their analyses (e.g., frequency of use or time spent on course materials). In the discussion section, the authors stress the importance of including this type of objective engagement data in future work. The same applies to the study by Vanslambrouck et al. (2019), who used LPA to study students' online self-regulation in blended learning environments and suggest that online measures should extend the commonly used self-reports.

Digital Measures of Students' Behavioural and Cognitive Engagement

Reading behaviour is often conceptualised in terms of reading frequency, reading pleasure, or reading environment (e.g., amount of books held at home), which all have a strong focus on recreational reading for pleasure instead of reading for school (Schiefele & Löweke, 2018). Moreover, these measurements often rely on self-reports, sometimes even assessed with a single item (Flunger et al., 2015; Tze et al., 2017). Although these studies led to interesting results, to define and measure student engagement remains a complex and challenging task. According to Azevedo (2015), it is important to triangulate process, product (e.g., performance), and self-reports to capture the complex nature and role of engagement in student learning.

The use of web-based log files or trace data is a common way to explore students' interactions with DLEs, which is also known as the concepts of educational data mining and learning analytics (Azevedo et al., 2013; Sheard, 2010; Siemens & Baker, 2012). Whereas educational data mining allows for extracting relevant information from large-scale datasets to process it for analytical purposes, learning analytics "seeks to interpret the collected data and draw conclusions from it … to optimize the individual learning process by exploiting the provided raw data" (Jülicher, 2018, p. 49). For example, existing learning analytics research focuses on the use of log files or trace data to distinguish students' navigational patterns in open-ended web environments or online courses (Lee, Kirschner, & Kester, 2016), and to cluster students according to their behaviour in these environments (Tze et al., 2017).

Log files, such as navigational data derived from digital systems, have been used in previous clustering research (cf. Barab, Bowdish, & Lawless, 1997; Bouchet et al., 2013; Sheard, 2010). However, to our knowledge, there are currently no studies using

clustering methods like LPA based on log files of real-time digital reading behaviour of students in secondary education to explain student differences with regard to reading motivation and reading comprehension.

Aims of the Current Study

Based on the aforementioned literature, we expect that students' behavioural and cognitive engagement is an important predictor of their reading comprehension performance, while at the same time this behavioural and cognitive engagement is influenced by students' motivation. Moreover, we suggest that this process is more cyclical than linear in nature; for example, students' motivation can influence their behavioural and cognitive engagement, but their engagement can also contribute to their motivation. Inspired by the model of Guthrie and Wigfield (2017), we designed a conceptual framework for the current study (see Figure 3.1).

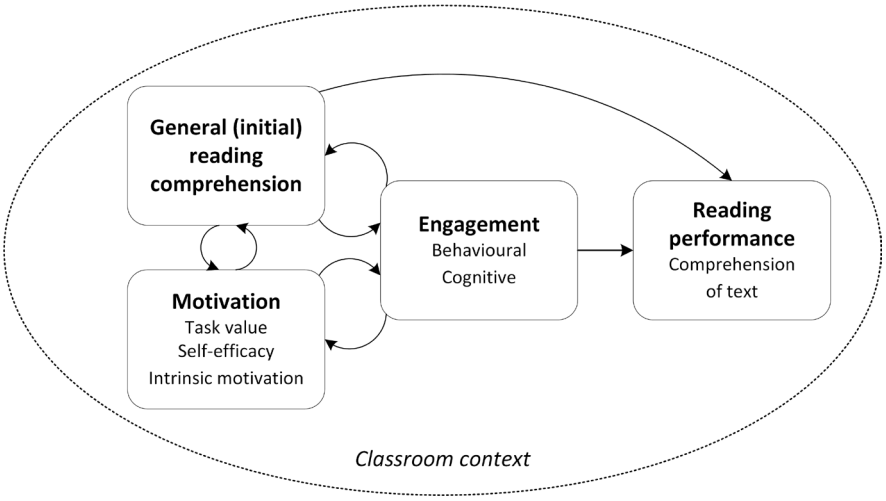


Figure 3.1 Conceptual framework for studying the relations between motivation, engagement, and reading comprehension.

Using a person-centred approach, including learning analytics based on digital log files, may offer unique and useful insights for this topic. Therefore, the purpose of the present exploratory study was threefold. First, to distinguish profiles based on students' real-time behavioural and cognitive engagement in a DLE while reading expository history texts. Second, to evaluate how these engagement profiles relate to three aspects of (reading) motivation: task value, self-efficacy, and intrinsic motivation. Third, to investigate how the profiles relate to students' posttest text

comprehension. In this study, we will address the following research questions:

1. Which meaningful profiles can be identified based on log files about students' behavioural and cognitive engagement in a DLE and what are their characteristics?
2. To what extent are these engagement profiles related to the motivational aspects of task value, self-efficacy, and intrinsic motivation?
3. To what extent are these engagement profiles related to students' expository history text comprehension?

Method

Participants

At first, 327 seventh-grade students from four secondary schools and thirteen classrooms participated in this study. The current study did not require submission for ethical approval at our local institutional review board, since it already obtained approval from a governmental review board involved in assessing the grant application. Nevertheless, parents or caretakers of all participating students were informed about the research project via a personal letter and could refuse the use of their child's data. This was the case for two students; their data were removed from all datasets. Therefore, the initial sample consisted of 325 students, of which 47.7% was female (n = 155) and 52.3% was male (n = 170). Students' average age was 12.5 years (SD = 0.45). Ten classrooms consisted of a mixed educational level of general secondary and pre-university education; three classrooms had a predominantly prevocational educational level[1]. Due to exclusion of students with missing data, the final sample consisted of 311 students (see 'Attrition and missing data' for a detailed description).

Design and Context

We designed a Digital Learning Environment (DLE) called 'Gazelle'[2] in cooperation

1 In Dutch secondary education, many schools mix the educational levels of prevocational (vmbo), higher general secondary (havo), and pre-university education (vwo) in seventh and eighth grade to determine the final educational level of a student at a later stage, based on his or her performance during the early secondary years. Pre-vocational education grants access to vocational education. Higher general secondary education grants access to higher vocational education, whilst pre-university education also grants access to university education.

2 Gazelle is a Dutch acronym for 'Gemotiveerd en Actief Zelfstandig Lezen', which roughly translates into 'Motivated and Active Independent Reading'.

with teachers and non-participating seventh-grade students (ter Beek, Spijkerboer, Brummer, & Opdenakker, 2018). The DLE contained expository texts for the subject of history. In line with the regular seventh-grade curriculum, the main theme of all texts was 'The time of Greeks and Romans'. We carefully analysed the contents of different regular textbooks to prevent overlap or duplicate information, since the lessons in which students used the DLE replaced six regular history lessons. Students worked in the DLE during six consecutive weeks. During this intervention, all students from each school read six expository texts about the ancient Greeks. Each text consisted of approximately 550 words and a lesson lasted approximately 50 minutes. Figure 3.2 provides an impression of the DLE contents.

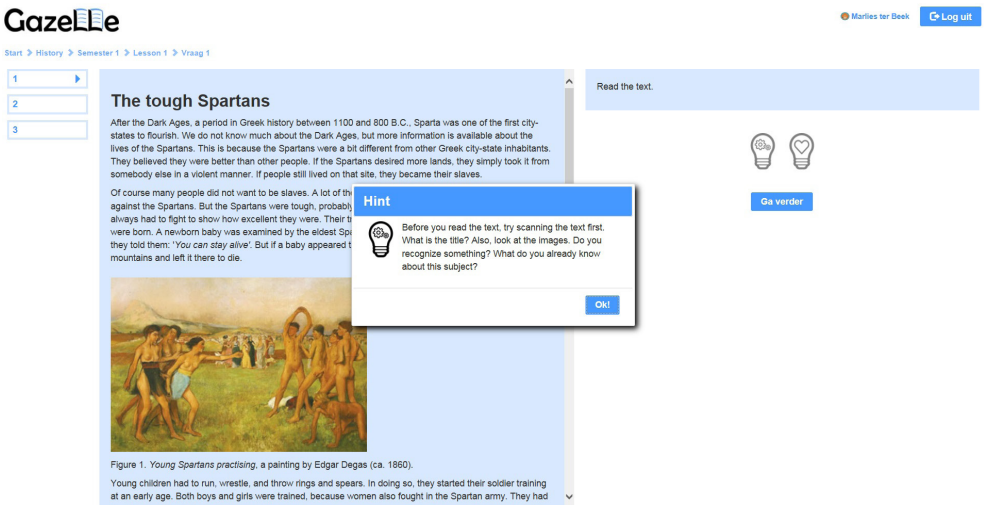


Figure 3.2 Screenshot of the Gazelle-program showing the contents of a supportive hint (translated from Dutch to English).

Lessons in the DLE. At the start of each lesson in the DLE, students had to answer an open-ended question about the value of the reading task ahead (not used in this study). After this, students received a prompt to read the text. Next, students had to summarise this text using a maximum of 150 words, after which ten text-related multiple-choice questions followed. The text remained visible during all assignments to minimise potential impediments caused by memorisation problems. After each multiple-choice question, students had to indicate their confidence in the correctness of their answer on a scale of 1 (*low*) to 5 (*high*) stars, which functions as an indicator of their judgment of learning (JOL). The lesson ended with two open-ended

questions in which students had to reflect on their summary and had to write down a piece of advice for themselves for the following lesson (not used in the current study).

Strategic hints. The DLE offered additional support during reading in the form of hints, which students could deliberately decide to access when they thought they needed them. There were three types of hints: cognitive, metacognitive, and motivational. Cognitive hints consisted of strategy instruction or explanations about the content of the text (e.g., "A reason can be found after the appearance of words like *because* or *since*"). Metacognitive hints aimed at students' regulation of their own learning process (e.g., "Evaluate your own work by focusing not only on your results, but also on your progress or your emotions"). Motivational hints pointed out the value of the reading task (i.e., the 'why' of the task) and what students might learn by reading the text (i.e., the usefulness of the task: "If you write down why reading this text is useful to you, you will look at this task in a more positive way"). Throughout the six-week intervention, students could access a maximum of 80 cognitive hints concurrently with the multiple-choice questions, and a further 24 metacognitive and 28 motivational hints during the summary assignment and the open-ended questions at the start and end of each lesson.

Procedure

Prior to the intervention, students completed two questionnaires: one to determine their initial (general) reading comprehension level, and another to determine their initial motivation for the subject of history in terms of task value, self-efficacy, and intrinsic motivation (i.e., T1; see Figure 3.3). Two weeks after completing the questionnaire, all students started working in the DLE in the same week. During weeks 1 and 6, none of the students had access to hints to ensure the comparability of all students. In addition, students only had one opportunity to answer the multiple-choice questions in weeks 1 and 6. During weeks 2–5, students were given the opportunity to access hints and to correct an incorrect multiple-choice answer after their first try. Cognitive hints were accessible for the multiple-choice questions, whereas metacognitive and motivational hints were accessible during the summary assignment and the open-ended questions. If a student's answer was incorrect, an on-screen pop-up provided the following feedback message: "Unfortunately, this answer is incorrect. Please try again. Maybe using a hint can help you?" The DLE recorded the actions of all students throughout the entire intervention. After the six-week intervention, we administered the motivational questionnaire again (i.e., T2).

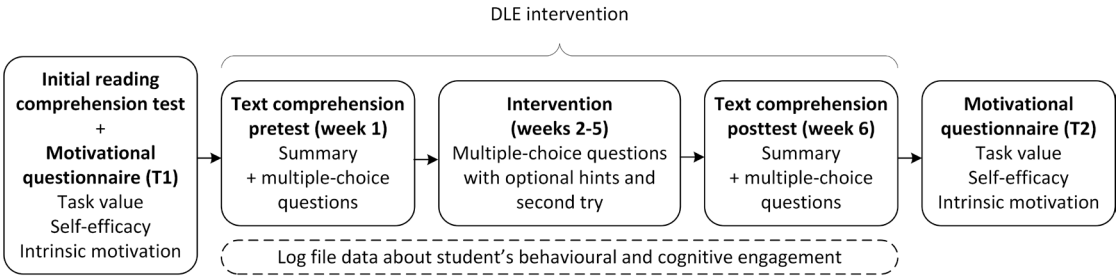


Figure 3.3 Timeline for the study procedure and data collection.

Measures

Background variables. Since we wanted to identify detailed characteristics of students assigned to the latent profiles, we included measures of gender (1 = male, 2 = female), educational level of the classroom the students are in (0 = predominantly vocational, 1 = mixed general secondary and pre-university), and initial reading comprehension level. Information about students' gender and educational level was provided by the participating schools; however, because the participating students all recently transitioned from primary to secondary education, we were not able to include equal estimates of prior performance in the specific domain of history (e.g., grades or test scores). Therefore, students' initial reading comprehension level was determined with a validated Dutch instrument by Aarnoutse (1987). The original instrument consists of four subtests: 'main ideas', 'conjunctures', 'synonyms', and 'antonyms'. The contents of these subtests are generic in nature and not related to a specific subject such as history. According to Aarnoutse, the subtests for 'main ideas' and 'conjunctures' relate to higher levels of reading comprehension, such as recognising relationships between parts of the text, whereas 'synonyms' and 'antonyms' relate to vocabulary knowledge.

Although it is a widely recognised and reliable instrument to measure students' reading comprehension (Aarnoutse, 1987), we updated the old-fashioned language of the original instrument. Due to time constraints with regard to testing the students, we shortened the original 'main ideas' subtest from 21 to 10 items, the 'conjunctions' subtest from 23 to 20 items, the 'synonyms' subtest from 30 to 20 items, and the 'antonyms' subtest from 39 to 20 items. Since one of our previous studies showed that only administering two subtests appeared to be restrictive to obtain a comprehensive overview of students' reading comprehension skills (ter Beek, Opdenakker, Deunk,

& Strijbos, 2019; see Chapter 2), we decided to use all four subtests and include a composite score as a background variable in this study. The scale scores based on the final 70 items yielded a Cronbach's α of .87. These values are similar to the reported reliability values referring to the subtest scores in the original instrument, which ranged from α = .80 to α = .87 (Aarnoutse, 1987).

Task value and self-efficacy. To measure students' motivation, we adopted existing scales from commonly used instruments. The original items were translated from English to Dutch, and we added the specific subject to the items to ensure domain specificity (i.e., '*in my history class*' or '*while reading history texts*'; ter Beek et al., 2018). We measured students' perceived task value and self-efficacy beliefs with subscales from the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, García, & McKeachie, 1991). We conceptualise task value (TV) as students' interests in and beliefs about the specific subject of history. The TV subscale refers to students' perception of how interesting, important, or useful a task or course is in general (e.g., "I am very interested in the contents of my history course"). We assessed students' TV with six items measured on a five-point Likert scale from 1 (*not true at all for me*) to 5 (*absolutely true for me*). Cronbach's α for the scale scores of this scale was .81 on T1 and .82 on T2. The alpha value as reported by Pintrich et al. (1991) was .90.

We define self-efficacy (SE) as students' beliefs about their ability to comprehend or execute domain-specific history tasks. The SE subscale measures students' perceived ability to master a task (e.g., "I am confident I can understand the basic concepts taught in my history course"). We assessed students' SE with eight items measured on a five-point Likert scale with identical anchors as for the TV subscale. The reliability estimates for this scale were good (Cronbach's α = .87 on T1 and .91 on T2); the Cronbach's alpha reported by Pintrich et al. (1991) was .93.

Intrinsic motivation. We define intrinsic motivation (IM) as student enjoyment of or interest in reading texts for the subject of history. We administered an eight-item composite scale to measure students' IM using six items from the Adolescent Motivations for School Reading questionnaire (AMSR; Coddington, 2009) and two items from the Motivations for Reading Information Books School questionnaire (MRIB-S; Guthrie et al., 2009). We obtained written permission to adapt and use these items, provided that we would clarify the alterations made to the original instrument. We made the items history-specific by changing the term 'language arts/reading' into 'history'. All items were measured on a five-point Likert scale ranging

from 1 (*not true at all for me*) to 5 (*absolutely true for me*). Appendix B contains the original and adapted items for the IM scale. We deliberately included two negatively worded items to prevent students from selecting the same answers for every item. The reliability estimates for the scale scores of this scale were good; Cronbach's α was .89 on T1 and .91 on T2, which is comparable with the original reported alpha values of .92 (Coddington, 2009) and .85 (Guthrie, Wigfield, & Klauda, 2012).

Expository history text comprehension. Students' expository history text comprehension was measured within the DLE, using results from the texts from weeks 1 and 6 about ancient Greece. We operationalise students' text comprehension in terms of their answers on multiple-choice questions and which main ideas they included in summaries.

Multiple-choice questions. Each expository text was accompanied by ten multiple-choice questions. These questions resembled regular textbook questions and focused on text features relevant for the subject of history, such as causal relations (e.g., "How did the Spartans become such good soldiers?"), or explaining historical events (e.g., "Explain why the 300 Spartan soldiers went into battle against 10,000 Persians"). The multiple-choice questions of weeks 1 and 6 addressed similar text features. During weeks 1 and 6, students did not have the opportunity to correct their answer. They received one point per correct answer, which led to a maximum score of 10 points. We used sum scores of the ten multiple-choice questions of week 1 (pretest) and week 6 (posttest) as indicators for students' text comprehension performance.

Summaries. In weeks 1 and 6, students had to write a summary in the DLE, reproducing the main ideas of the text with a maximum of 150 words. Presence of main ideas in summaries can be considered a measure of text comprehension, since reproducing main ideas from texts is an indicator of students' comprehension (Kintsch & van Dijk, 1978). The first author and three research assistants jointly trained the rating of students' summaries with a fixed scoring protocol that included the five main ideas from each text (e.g., "The summary mentions that Spartan society was characterised by warfare, fighting, or the training of soldiers"). The maximum score for each summary was 5 points, one for each main idea. After a 2-hour training, all raters scored six randomly selected summaries; three from week 1 and three from week 6. Since multiple researchers rated the summaries and the five items in the protocol were scored nominally (present = 1, absent = 0), we used Krippendorff's alpha to determine interrater reliability (Krippendorff, 2004) and obtained a sufficient reliability estimate of .70.

Predictive engagement variables. We extracted raw log file data about students' actual behaviour from the DLE and transformed them into output files with continuous and dichotomous variables for each open-ended and multiple-choice question in the DLE. Subsequently, we computed mean scores for either weeks 1–6 or weeks 2–5 of the intervention (see Table 3.1). We selected five variables, based on log file data from the DLE, as indicators of students' behavioural and cognitive engagement and predictors in our latent profile analysis. Together these predictor variables provide a comprehensive and interpretable overview of students' engagement while working in the DLE.

Table 3.1 Overview of weekly data used for average or total scores on predictor variables

Predictor variable	Week 1	Weeks 2–5	Week 6	Score
Time on task	x	x	x	Average
Cognitive hints		x		Total no.
Metacognitive + motivational hints		x		Total no.
MCQ score at first try		x		Average
JOL accuracy	x	x	x	Average

Note. MCQ = multiple-choice questions; JOL = judgment of learning.

Time on task. Time spent on learning tasks can regarded as an indicator of behavioural engagement (Fredricks et al., 2004). The DLE measured students' time on task from the moment they started a lesson. However, it tracked time as long as the DLE was active in the browser. Hence, if a student did not close the DLE properly after finishing a lesson, the value for time on task was very high. Two students were severe outliers with regard to their average time on task. Close examination revealed that they spent approximately four hours on one of the six lessons—a highly unrealistic value, and very different from their time on task for the other five lessons. We therefore changed all values above 50 minutes (i.e., higher than the regular lesson time) to missing values. For 19 students, this meant that one or two values for time on task were left out when computing their average time on task. We did not exclude very low values for time on task, since this could be a realistic indicator of students' behaviour. The average time on task was included as a continuous variable. Since students' time on task declined throughout the weeks, we used data from weeks

1–6 to be able to represent the average time on task for the entire intervention as accurately as possible.

Hint use. Strategy use can be considered as a form of cognitive engagement (Fredricks et al., 2004; van Rooij et al., 2017). We have no measurements of students' actual strategy use while working in the DLE; however, we do know whether students accessed supportive hints containing cognitive, metacognitive, or motivational strategy information. We used data from weeks 2–5, since these were the only weeks in which students could use the hints. To distinguish between cognitive engagement during multiple-choice questions and during open-ended questions, we included (a) the total amount of accessed cognitive hints and (b) the total amount of accessed metacognitive and motivational hints combined as count variables. Since metacognitive and motivational hints were both accessible during open-ended questions and students used these hints very little in general, we decided to combine these two types of hints into one variable.

Average score at first try on multiple-choice questions in weeks 2–5. The DLE functions as a means of practising reading expository texts through answering multiple-choice questions. If a student aims to answer the questions correctly at the first try (and succeeds), this can be seen as an indicator of students' mental effort in completing learning tasks, and, thus, as cognitive engagement (Fredricks et al., 2004). We did not include students' scores at second try, since some students did not need a second attempt and because these scores are possibly influenced by the result from the first attempt. Therefore, we included the average score on students' first try of answering the multiple-choice questions of weeks 2–5. We first calculated sum scores for all four weeks separately, followed by a mean score across the four weeks; the latter was included as a continuous variable. We only used data from weeks 2–5 because students' score at first try in weeks 1 and 6 was already used as a measure of pretest and posttest reading comprehension.

JOL accuracy. Students had to indicate their confidence in the correctness of their multiple-choice answers at their first try, which we here operationalise as a form of cognitive engagement. Students' JOL accuracy, that is, the correspondence between students' certainty of a selected answer and the actual result, was calculated separately for weeks 1 through 6 using the following formula by Schraw (2009) for the Absolute Accuracy Index:

$$1/n \sum_{i=1}^{n} (c_i - p_i)^2$$

where n = number of items (= 10 multiple-choice questions per week), ci = confidence rating per question (i.e., 1 star = 0.0; 2 stars = 0.25; 3 stars = 0.5; 4 stars = 0.75; and 5 stars = 1.0), and p_i = performance score for the corresponding question on the first try (i.e., 0 = incorrect; 1 = correct). The absolute accuracy index ranges from 0.0–1.0, for which scores close to zero correspond to high accuracy, while scores toward the maximum correspond to low accuracy. After calculating the index for each week separately, we computed the mean accuracy across the six weeks and included it as a continuous variable. We transformed the absolute accuracy index scores by subtracting the initial value from 1 to create a variable where a higher score is associated with better JOL accuracy. By doing so, the correlations between JOL accuracy and the other predictor variables are easier to interpret.

Statistical Analyses

Attrition and missing data. After completion of the initial reading comprehension test, but prior to the start of the six-week intervention, two students changed schools. Furthermore, 12 students did not complete all six lessons in the DLE. We could not determine reliable engagement profiles for these 14 students (4.3% of the total sample), because they missed several lessons—including the last lesson, which functions as the reading comprehension posttest—and, thus, their predictor variables with regard to engagement were incomplete. Since the number of excluded students did not exceed 5% of the total sample, and these students did not significantly differ from the included students in terms of gender, educational level, initial reading comprehension, and motivation, we found it acceptable to apply listwise deletion (Graham, 2009; cf. Schiefele & Löweke, 2018). The final sample consisted of $N = 311$ students.

With regard to the T2 questionnaire on students' motivation, data for an additional 24 students were missing.[3] A Missing Value Analysis using Little's test of Missing Completely At Random (MCAR), including all three motivation subscales at T1 and T2, was not significant, $\chi^2 = 3.479$, $df = 3$, $p = .323$, indicating that these data were missing at random. Because we were able to determine engagement profile membership for these 24 students as well as their reading comprehension performance

3 This was probably caused by the fact that the T2 questionnaire was administered in the week before Christmas, a week in which many students missed lessons due to other activities.

at posttest, and since the self-report measurement of motivation was similar at T1 and T2, we imputed their missing data at T2 using expectation maximisation instead of excluding these students from the dataset.

Identifying engagement profiles. We used five predictor variables to identify profiles by conducting LPA using Latent GOLD 5.0 (Vermunt & Magidson, 2013). We adapted a three-step approach (Hickendorff et al., 2018). First, we included the predictor variables in our analysis and fitted solutions with 1–8 profiles; expecting more than eight profiles was considered practically and theoretically unreasonable. Second, we determined the best profile solution to fit our data and assigned all students to the profile for which their membership probability was highest. Third, we used these profiles to analyse the associations between profile membership and students' motivation and text comprehension performance.

In the second step, we assessed each profile solution based on a combination of three criteria often used in LPA research: statistical model fit, parsimony, and interpretability (Hickendorff et al., 2018; Marsh et al., 2009). We used several statistical indicators to determine model fit: Akaike information criterion (AIC), Bayesian information criterion (BIC), and the entropy statistic. Lower values for log likelihood, AIC, and BIC indicate a better fit; higher entropy values (ranging from 0–1) indicate less classification error (Collins & Lanza, 2010), and entropy values above .75 indicate good classification accuracy. However, consistent with the findings of Nylund, Asparouhov, and Muthén (2007) for latent profile models, we favoured the BIC over other fit indices for selecting the number of profiles; BIC is stronger in selecting the correct number of profiles compared to the AIC and entropy values (Tein, Coxe, & Cham, 2013). Therefore, we mostly focused on the BIC values when determining the best profile solution fit. In addition, we took into account the interpretability and practical value of the final profile solutions; similar to Van Rooij et al. (2017), the percentage of students assigned to the smallest profile should be no less than five to ensure its practical value.

Associations between profile membership and external variables. We investigated differences between the latent profiles on motivation and expository text comprehension using variance analysis with General Linear Models (GLM) and post hoc comparisons using Bonferroni adjustment. We report effect sizes using partial eta squared, or partial η^2. We consider effect sizes as small when partial $\eta^2 < 0.06$, medium when $0.06 <$ partial $\eta^2 < 0.14$, and large when partial $\eta^2 > 0.14$ (cf. Cohen, 1988).

Results

Descriptive Statistics and Variable Correlations

Table 3.2 shows the descriptive statistics and the correlations between the variables used. All significant correlations were positive in their direction. Student's initial reading comprehension correlated significantly with T1 task value (TV), T1 intrinsic motivation (IM), all measures of comprehension, and all predictor variables except for cognitive hint use. Correlations between TV, self-efficacy (SE), and IM were significant at both T1 and T2. Task value at T1 also correlated significantly with pretest and posttest reading comprehension scores for multiple-choice questions (MCQ) and main ideas in summaries (SUM), average time on task in the DLE, and average score at first try. Students' SE at T1 correlated significantly with students' pretest MCQ performance, while SE at T2 correlated significantly with posttest MCQ performance. IM correlated significantly with all measures of reading comprehension performance and time on task in the DLE. Measures of reading comprehension also significantly correlated with each other, except for posttest MCQ and posttest SUM, and with students' average score at first try and time on task in the DLE. Cognitive, metacognitive and motivational hint use correlated significantly with MCQ posttest scores and average time on task in the DLE.

Identifying Engagement Profiles (RQ1)

Determining the number of latent profiles. Table 3.3 shows the model fit values for one to eight profiles. We carefully analysed the BIC values using a scree plot and concluded that the marginal gains in model fit dropped at the five-profile solution level (i.e., the "elbow criterion"; Masyn, 2013). Although the BIC indicated that the six to eight-profile solutions suggested a better fit compared to the five-profile solution, these solutions yielded small profiles including only a few students. Following Flunger et al. (2015) and Van Rooij et al. (2017), we therefore also considered the percentage of students assigned to the smallest profile as well as the interpretability and practical value of the profile solutions. We preferred the solution with fewer profiles if a solution with more profiles only included minor variations of profiles already identified. Compared to the four-profile solution, which included a profile with high time on task and high amount of hints used, the five-profile solution yielded an additional profile with high time on task but *low* amount of hints used. The six-profile solution did not yield a new distinctive profile compared to the five-profile solution. Since the percentage of

Table 3.2 Descriptive statistics and bivariate correlations ($N = 311$)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. IRC	-															
2. T1 TV	.21**	-														
3. T2 TV	.09	.69**	-													
4. T1 SE	.02	.57**	.39**	-												
5. T2 SE	-.03	.43**	.58**	.63**	-											
6. T1 IM	.16**	.76**	.56**	.48**	.33**	-										
7. T2 IM	.03	.61**	.75**	.42**	.50**	.70**	-									
8. T1 MCQ	.36**	.18**	.11	.16**	.08	.14*	.10	-								
9. T2 MCQ	.19**	.22**	.11*	.09	.13*	.23**	.15**	.23**	-							
10. T1 SUM	.33**	.17**	.08	.05	-.05	.13*	.02	.30**	.18**	-						
11. T2 SUM	.14*	.25**	.14*	.06	-.02	.16**	.02	.19**	.11	.37**	-					
12. Time o/t	.14*	.28**	.22**	.11	.10	.19**	.08	.29**	.28**	.34**	.36**	-				
13. Hints C	.03	.08	.05	-.02	.00	.06	.00	.03	.14*	.08	.11	.22**	-			
14. Hints M	.12*	.07	.05	.00	-.01	.06	-.01	.04	.20**	.10	.09	.19**	.48**	-		
15. 1st score	.34**	.16**	.10	.03	.03	.11	-.01	.36**	.22**	.26**	.24**	.41**	.10	.13*	-	
16. JOL acc.	.12*	.10	.05	.08	.00	-.02	-.03	.09	.04	.08	.10	.13*	-.08	-.04	.24**	-
M	51.73	3.18	3.11	3.24	3.24	2.64	2.68	6.68	6.19	1.37	1.27	17.14	6.27	2.37	5.14	0.70
SD	9.53	0.67	0.67	0.58	0.64	0.80	0.83	1.98	1.39	1.20	1.33	4.95	7.73	3.41	1.34	0.07

Note. IRC = initial reading comprehension; TV= task value; SE = self-efficacy; IM= intrinsic motivation; MCQ = multiple-choice questions; SUM = summary; Time o/t = time on task; Hints C = total number of accessed cognitive hints; Hints M = total number of accessed metacognitive and motivational hints; 1st score = average score for multiple-choice answers at first try; JOL acc. = judgment of learning accuracy. $N = 301$ for IRC. *$p < .05$, two-tailed. **$p < .01$, two-tailed.

students in the smallest profile was 4.8% for the five-profile solution and 1.9% for the six-profile solution, we opted for the five-profile solution as the best fit for our data.

Table 3.3 Model fit for estimated models

Model	Npar.	LL	AIC	BIC	Entropy
1-profile	8	-3488.7476	6993.4953	7023.4136	1.00
2-profile	17	-2855.6336	5745.2673	5808.8437	0.90
3-profile	26	-2699.0390	5450.0781	5547.3127	0.86
4-profile	35	-2633.3349	5336.6698	5437.5626	0.80
5-profile	**44**	**-2595.6885**	**5279.3770**	**5443.9279**	**0.79**
6-profile	53	-2550.5239	5207.0478	5405.2568	0.79
7-profile	62	-2519.2037	5162.4073	5394.2745	0.78
8-profile	71	-2495.0524	5132.1047	5397.6300	0.81

Note. Npar. = number of free parameters; LL = Log Likelihood.

Latent profile characteristics. We labelled the five latent profiles to distinguish the differences in students' reading engagement they represent. Table 3.4 shows the background characteristics of each latent profile. The average score on the initial reading comprehension test differed significantly between the identified profiles, $F(4, 296) = 2.43$, $p = .048$, partial $\eta^2 = .03$.

Naïve readers. The largest profile (n = 110; 35.4%) scored relatively low on all indicators of engagement. This means that these students spent little time in the DLE, accessed few cognitive, metacognitive, and motivational hints, had low scores at first try, and had lower JOL accuracy. Thus, these students had low performance, but did not appear to be (fully) aware of this and did not change their behaviour accordingly. Therefore, we decided to name this profile the 'naïve readers'. Students in this profile had the lowest average score on the initial reading comprehension test; post hoc analysis with Bonferroni adjustment showed that the naïve readers differed significantly from the independent readers, $p = .025$.

Stubborn readers. The second largest profile (n = 73; 23.5%) showed some similarities to the naïve readers: students in this profile also had relatively low scores on time on task, used almost no hints at all, and had lower scores at first try. However,

Table 3.4 Background characteristics of the latent profiles

Characteristic	Total sample	Naïve readers	Stubborn readers	Help-seeking readers	Independent readers	Uncertain readers
% of students (number)	100 (311)	35.4 (110)	23.5 (73)	22.5 (70)	13.8 (43)	4.8 (15)
% female (vs. male)	47.9	44.5	41.1	48.6	58.1	73.3
% students in prevocational education (vs. general secondary and/or pre-university education	23.2	23.6	24.7	27.1	14.0	20.0
Average score on initial reading comprehension test (*SD*)	51.73 (9.53)	$50.25\ (10.00)_{a}$	$51.22\ (9.60)_{a,b}$	$52.23\ (9.46)_{a,b}$	$55.47\ (8.09)_{b}$	$51.33\ (7.88)_{a,b}$

Note. For the initial reading comprehension test, the maximum score was 70, and $N = 301$. Within-row means with different subscripts differ significantly at $p < .05$.

their JOL accuracy was significantly higher than that of the naïve readers, indicating that these students were more aware of their low performance. Since these students did not show the type of cognitive engagement that could possibly improve their scores (i.e., by using hints), we named this profile the 'stubborn readers'. There was a relatively high amount of male students in this profile. The average score on the initial reading comprehension test of the stubborn readers was 51.22, which is comparable with the total sample average.

Help-seeking readers. Students in the third profile (n = 70; 22.5%) scored around the mean sample average for most of the indicator variables. Compared to the first two profiles, these students used significantly more cognitive, metacognitive and motivational hints. We therefore indicate this profile as the 'help-seeking readers'. Compared to the total sample, there was a relatively high amount of prevocational students in this profile (27.1%). The average score on the initial reading comprehension test of the help-seeking readers was 52.23, which is slightly higher than the total sample average.

Independent readers. The fourth profile (n = 43; 13.8%) scored relatively high on time on task, and the highest on scores at first try and JOL accuracy. In contrast, their supportive hint use was relatively low compared to all other profiles. Apparently, students in this profile were able to perform well at first try without accessing the additional support. Therefore, we decided to name this profile the 'independent readers'. Compared to the total sample, there was a relatively high amount of female students in this profile (58.1%), and a relatively low amount of prevocational students (14.0%). Students in this profile had the highest average score on the initial reading comprehension test, and differed significantly from the naïve readers, p = .025.

Uncertain readers. The fifth and last profile consisted of a small number of students (n = 15; 4.8%), whose scores were relatively high on almost all engagement indicators, especially time on task and hint use. However, their JOL accuracy was relatively low, indicating that they often misjudged their correct answers. We named this profile the 'uncertain readers'. Female students were overrepresented in this profile (73.3%). The average score on the initial reading comprehension test of the uncertain readers was 51.33, which is comparable with the total sample average and the stubborn readers.

Profiles and predictor variables. Post hoc comparisons with Bonferroni adjustment showed that the five profiles differed significantly in various ways on

the predictor variables. Table 3.5 shows the mean scores and standard deviations on predictor variables per latent profile and post hoc comparisons. The largest effects of profile membership appeared in the measures of cognitive hint use (R^2 = 0.83), metacognitive and motivational hint use (R^2 = 0.35), and time on task (R^2 = 0.32). All profiles, except for the naïve and independent readers, differed significantly from each other on measures of cognitive hint use ($p < .001$) and metacognitive and motivational hint use ($p < .05$).

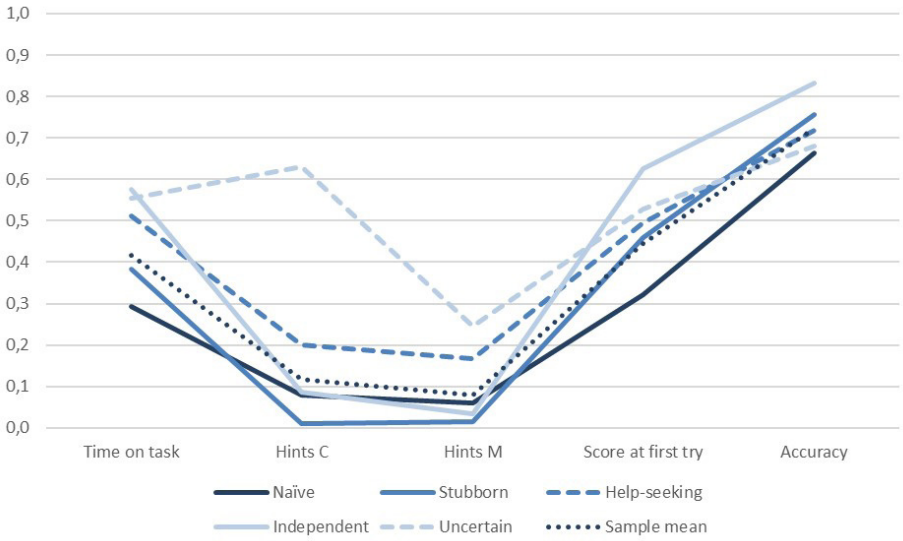


Figure 3.4 Normalised means [0–1] plot for the five latent profiles and the sample mean.

With regard to time on task, the naïve and stubborn readers differed significantly from each other ($p = .002$) and from the other three profiles ($p < .01$). Figure 3.4 shows the five profiles and the sample average on a 0–1 means plot, which depicts the profile-specific means rescaled into a 0–1 range.

Relations between Student Motivation and Engagement Profiles (RQ2)

Since student motivation and engagement are closely related (cf. Wolters et al., 2017), we analysed how profile membership relates to students' motivation prior to and after the intervention. Items for TV and SE focused on the central subject of history, whilst IM items aimed specifically at reading texts for history. Table 3.6 shows the average motivation per subscale for each profile at T1 and T2.

Task value. The naïve readers had the lowest average score on TV at T1; the

Table 3.5 Mean scores and standard deviations on predictor variables per latent profile and post hoc comparisons

Variable	Total sample mean (*SD*)	Naïve readers	Stubborn readers	Help-seeking readers	Independent readers	Uncertain readers	*p*	*F*	R^2
Time on task (in minutes)	17.14 (4.95)	$14.03\ (3.34)_a$	$16.34\ (4.22)_b$	$19.61\ (4.79)_c$	$21.21\ (4.35)_c$	$20.69\ (4.58)_c$	< .001	36.12	0.32
Cognitive hints (overall)	6.27 (7.73)	$4.20\ (2.20)_a$	$0.56\ (0.67)_b$	$10.67\ (4.12)_c$	$4.63\ (2.41)_a$	$33.40\ (9.25)_d$	< .001	370.66	0.83
Metacognitive and motivational hints (overall)	2.37 (3.41)	$1.81\ (1.43)_a$	$0.47\ (0.78)_b$	$5.00\ (3.72)_c$	$1.00\ (1.29)_{a,b}$	$7.40\ (8.70)_d$	< .001	40.44	0.35
Average score at first try (0–10)	5.14 (1.34)	$4.35\ (1.07)_a$	$5.23\ (1.32)_b$	$5.46\ (1.15)_b$	$6.31\ (1.06)_c$	$5.68\ (1.38)_{b,c}$	< .001	25.99	0.25
JOL accuracy (0–1)	0.70 (0.07)	$0.68\ (0.08)_a$	$0.72\ (0.05)_b$	$0.70\ (0.07)_{a,b}$	$0.76\ (0.03)_c$	$0.68\ (0.05)_{a,b}$	< .001	14.11	0.16

Note. JOL = Judgment of learning. Within-row means with different subscripts differ significantly at $p < .05$.

independent and uncertain readers the highest. There was a significant difference between the five latent profiles on subject-specific TV at T1, $F(4, 306) = 3.60$, $p = .007$, partial $\eta^2 = .05$. Post hoc comparisons showed a significant difference between the naïve readers ($M = 2.99$, $SD = 0.67$) and the independent readers ($M = 3.35$, $SD = 0.63$), $p = .031$. When controlling for TV at T1, there was no significant difference between the profiles on T2 TV, $F(4, 305) = 0.26$, $p = .902$, partial $\eta^2 = .00$.

Self-efficacy. Uncertain and naïve readers had the lowest average scores on SE at T1; the help-seeking readers displayed the highest average SE at T1. There were no significant differences between the latent profiles on SE at T1, $F(4, 306) = 0.61$, $p = .660$, partial $\eta^2 = .01$. The same accounts for T2, $F(4, 306) = 0.68$, $p = .609$, partial $\eta^2 = .01$.

Intrinsic motivation. IM at T1 was highest for the help-seeking readers and lowest for the naïve readers. There was a significant difference between the profiles on IM at T1, $F(4, 306) = 3.42$, $p = .009$, partial $\eta^2 = .04$. Post hoc comparisons showed a significant difference between the naïve readers ($M = 2.48$, $SD = 0.83$) and the help-seeking readers ($M = 2.90$, $SD = 0.74$), $p = .005$. When controlling for T1 IM, there was still a significant difference between the profiles on T2 IM, $F(4, 305) = 4.83$, $p = .001$, partial $\eta^2 = .06$. This time, post hoc comparisons showed a significant difference between the stubborn readers ($M = 2.83$, $SD = 0.87$) and the help-seeking readers ($M = 2.65$, $SD = 0.72$), $p < .001$.

Relations between Engagement Profiles and Text Comprehension (RQ3)

To determine whether and how the profiles related to students' text comprehension, we compared the profiles with regard to their performance on the multiple-choice questions (MCQ) and summaries (SUM). Table 3.6 shows the text comprehension performance per latent profile.

Multiple-choice questions. The total sample mean of students' pretest MCQ performance was 6.68. There was a significant difference between the profiles on the MCQ pretest, $F(4, 306) = 2.98$, $p = .020$, partial $\eta^2 = .04$. Post hoc comparisons showed a significant difference between the naïve readers ($M = 6.26$, $SD = 2.15$) and the independent readers ($M = 7.44$, $SD = 1.75$), $p = .009$. When controlling for pretest MCQ, there was also a significant difference between the profiles at posttest MCQ, $F(4, 305) = 4.22$, $p = .002$, partial $\eta^2 = .05$. This time, post hoc comparisons showed

Table 3.6 Motivation and comprehension performance per latent profile

Predictors or outcomes	Total sample mean (*SD*)	Naïve readers (*n* = 110)	Stubborn readers (*n* = 73)	Help-seeking readers (*n* = 70)	Independent readers (*n* = 43)	Uncertain readers (*n* = 15)	*p*	*F*	Partial η^2
Motivation									
Task value (T1)	3.18 (0.67)	$2.99\ (0.67)_{a}$	$3.23\ (0.73)_{a,b}$	$3.27\ (0.64)_{a,b}$	$3.35\ (0.63)_{b}$	$3.36\ (0.40)_{a,b}$	.007	3.60	.04
Task value (T2)	3.11 (0.67)	2.95 (0.67)	3.13 (0.75)	3.19 (0.58)	3.28 (0.63)	3.23 (0.63)	.902	0.26	.00
Self-efficacy (T1)	3.24 (0.64)	3.19 (0.60)	3.25 (0.58)	3.31 (0.60)	3.25 (0.51)	3.13 (0.50)	.660	0.61	.01
Self-efficacy (T2)	3.24 (0.64)	3.18 (0.65)	3.27 (0.68)	3.28 (0.65)	3.34 (0.57)	3.15 (0.49)	.609	0.68	.01
Intrinsic mot. (T1)	2.64 (0.80)	$2.48\ (0.83)_{a}$	$2.60\ (0.86)_{a,b}$	$2.90\ (0.74)_{b}$	$2.75\ (0.68)_{a,b}$	$2.59\ (0.67)_{a,b}$	.009	3.42	.04
Intrinsic mot. (T2)	2.66 (0.82)	$2.51\ (0.83)_{a,b}$	$2.83\ (0.87)_{a}$	$2.65\ (0.72)_{b}$	$2.75\ (0.78)_{a,b}$	$2.81\ (0.90)_{a,b}$	.001	4.83	.06
Comprehension									
Pretest MCQ	6.68 (1.98)	$6.26\ (2.15)_{a}$	$6.74\ (1.84)_{a,b}$	$6.74\ (1.86)_{a,b}$	$7.44\ (1.75)_{b}$	$6.93\ (1.98)_{a,b}$	.020	2.98	.04
Posttest MCQ	6.19 (1.39)	$5.88\ (1.24)_{a}$	$5.97\ (1.31)_{a,b}$	$6.43\ (1.47)_{a,b}$	$6.67\ (1.41)_{a,b}$	$7.04\ (1.58)_{b}$	.002	4.22	.05
Pretest SUM	1.37 (1.20)	$1.06\ (1.09)_{a}$	$1.34\ (1.15)_{a}$	$1.41\ (1.26)_{a,b}$	$1.98\ (1.12)_{b}$	$1.80\ (1.37)_{a,b}$	< .001	5.39	.07
Posttest SUM	1.27 (1.33)	0.96 (1.17)	1.04 (1.34)	1.63 (1.45)	1.72 (1.28)	1.67 (1.40)	.019	2.99	.04

Note. Intrinsic mot. = intrinsic motivation; MCQ = multiple-choice questions; SUM = summary. Within-row means with different subscripts differ significantly at $p < .05$. TV at T2 was controlled for TV at T1; IM at T2 was controlled for IM at T1; posttest MCQ was controlled for pretest MCQ; posttest SUM was controlled for pretest SUM.

a significant difference between the naïve readers (M = 5.88, SD = 1.24) and the uncertain readers (M = 7.04, SD = 1.58), p = .039.

Summaries. The total sample mean of students' pretest SUM performance was 1.37. Similar to the multiple-choice questions, there was a significant difference between the profiles on the SUM pretest, $F(4, 306) = 5.39$, $p < .001$, partial $\eta^2 = .07$. Post hoc comparisons revealed a significant difference between the naïve readers (M = 1.06, SD = 1.09) and the independent readers (M = 1.98, SD = 1.12), $p < .001$, and between the stubborn readers (M = 1.34, SD = 1.15) and the independent readers, p = .048. When controlling for pretest SUM, there was also a significant difference between profiles at posttest SUM, $F(4, 305) = 2.99$, $p = .019$, partial $\eta^2 = .04$. However, post hoc comparisons revealed no significant differences between the profiles.

Discussion

The purpose of the present exploratory study was to distinguish profiles based on students' real-time behavioural and cognitive engagement in a DLE while reading expository history texts. Consequently, we explored the relationships and differences between these engagement profiles and students' motivation and text comprehension.

Summary of Findings

In line with previous research (cf. Retelsdorf et al., 2011), measures of students' perceived task value and intrinsic motivation correlated positively with text comprehension performance. In addition, engagement in terms of average scores at first try, supportive hint use, and time on task in the DLE all correlated positively with students' text comprehension, supporting the idea that behavioural and cognitive engagement and students' understanding of texts are related. The person-centred approach used in this study provided a detailed overview of students' digital reading engagement and the relations between engagement profile membership, motivation, and text comprehension.

We distinguished five different engagement profiles based on the log files from the DLE. Supportive hint use was an important predictor of profile membership. However, hint use is not necessarily good or bad in terms of engagement (Roll, Baker, Aleven, & Koedinger, 2014), so it is valuable to present a holistic overview of student engagement using multiple predictor variables. In doing so, we were able to characterise the five different profiles based on their behavioural and cognitive

engagement. More than half of the students in our sample belonged to the profile we classified as 'naïve readers', a profile with relatively low scores on all indicators of engagement. This result is in line with findings from Vanslambrouck et al. (2019), who report a high amount of students in their lowest self-regulated learning profile. Naïve readers have the lowest average score on the initial reading comprehension test, indicating that their reading comprehension skills and lack of engagement with the DLE are possibly related. In contrast to the naïve readers, the students in the profiles we conceptualised as the independent and uncertain readers—profiles with relatively high scores on indicators of engagement—were predominantly female.

Our conceptual model assumes that there is a bidirectional relationship between student motivation and engagement. Our results showed that the engagement profiles differed significantly in terms of task value and intrinsic motivation prior to the intervention (T1). Independent readers showed the highest initial task value, which seems reasonable; students who perform well probably know the value of educational tasks such as reading. The lowest task value and intrinsic motivation were found for the naïve readers. Nevertheless, task value decreased in all profiles after the intervention. This is not an exceptional finding: Students' academic motivation in general as well as their motivation to read school-related texts are known to decline throughout the first years of secondary school (Guthrie & Davis, 2003; Opdenakker, Maulana, & den Brok, 2012; Unrau & Schlackman, 2006). There were no significant differences between the profiles in terms of self-efficacy; these values remained rather stable throughout the intervention.

An interesting finding was the fact that help-seeking readers showed the highest intrinsic motivation. Moreover, there was a significant difference in the intrinsic motivation of naïve readers and help-seeking readers, in favour of the latter. A possible explanation for this finding could be that help-seeking readers are more mastery-oriented, or motivated to solve problems on their own, even if this requires the use of additional hints. Help-seeking readers and uncertain readers use relatively many hints, indicating that these profiles probably consist of students who are able to estimate when they need support and who do not hesitate to access it when needed.

With regard to students' text comprehension, there were already significant differences between the profiles on the pretest: independent readers performed highest on both multiple-choice questions and summaries, whereas naïve readers performed lowest; these profiles differed significantly from each other. The various

profiles also differed significantly from each other at posttest, when controlling the differences at pretest. However, effect sizes of profile membership for posttest text comprehension were small. Although stubborn and help-seeking readers had similar scores on both the multiple-choice and summary pretest, the decrease at posttest was larger for the stubborn readers, indicating that the help-seeking readers (i.e., students who accessed more supportive hints) might have benefitted more from using the hints.

Limitations and Suggestions for Future Research

The role of extrinsic motivation. We did not include a grading system in our DLE to ensure that it would be a safe practice environment, minimising the possible impact of students' fear of failure. However, according to the participating teachers, students were less motivated to read the texts in the DLE because there was a lack of reward if the form of, for example, an extra grade or bonus points. Earlier research has shown that the effects of academic reading motivation are only significant for reading frequency, but not for reading engagement and reading comprehension (De Naeghel et al., 2012). Moreover, there is a shift towards a primarily extrinsic reading motivation for students in secondary education, which undermines the positive effects of students' intrinsic motivation on performance (Schiefele & Löweke, 2018). This indicates that when secondary students have to read texts for school, they are probably extrinsically motivated to do so. Students will engage more in reading when they expect to receive a grade on a test based on the contents of the text. Since we did not measure extrinsic reading motivation, we cannot explore the relations between extrinsic motivation and our behavioural engagement profiles. Future research should also include measures of extrinsic motivation to test the effects of extrinsic factors, such as grading systems, on students' behaviour when reading expository texts in DLEs.

Classroom context. Classroom context, which includes the classroom environment and the (instructional) behaviours of teachers and students, can either support or hinder both students' motivation and engagement. In their model of reading motivation and engagement, Guthrie and Wigfield (2017) stress the influence of classroom instruction on students' motivation to read, engagement in reading, and reading achievement. In this study, we did not include measures of classroom context, but the instructional choices made by teachers might have influenced the ways in which students interacted with the DLE. Future research should include

measures of classroom practices, such as observational data or teacher and student interviews, to determine whether and how the classroom context relates to students' motivation, comprehension, and digital reading engagement.

Determining engagement based on log file data. The current study provides a unique contribution to the field of reading research by its use of digital log file data to analyse students' behavioural and cognitive engagement while working in a DLE. However, although digital technologies provide the opportunity to register students' reading activities through log files, this method only collects these activities at a surface level (e.g., clicks or navigational patterns; Veenman, 2016). By doing so, the researcher constructs meaning from data without being fully able to explain the findings from a students' perspective. Therefore, it is important to evaluate the operationalisation of engagement through digital log file data critically.

For example, in the current study, we considered hint use to be a form of cognitive engagement and included this as a predictor variable in our LPA. However, for independent readers, not using the supportive hints was not necessarily a sign of little engagement; these students apparently performed well without using the available support. Therefore, it is suggested that the use of log file measures to determine engagement should be triangulated with other real-time measures of students' strategic learning behaviour and motivation, such as concurrent think-aloud or eye-tracking methods (Azevedo & Gašević, 2019; Veenman, 2016), to provide a more in-depth analysis of student engagement.

Measures of motivation. We measured task value and self-efficacy on a subject-specific level (i.e., history in general) without explicitly including the domain of reading (i.e., reading texts for history). Although we found some differences between profiles with regard to students' task value, adding a domain-specific element to items from existing questionnaires in the field of reading research might contribute to even more detailed and valid measures of students' motivation *to read* for a specific school subject.

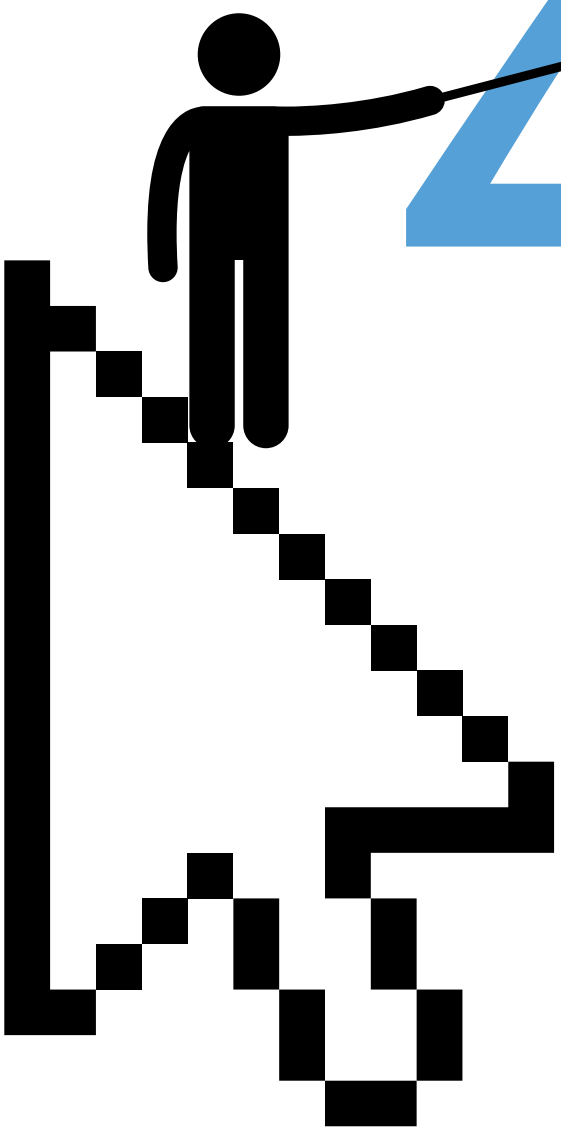
Practical and Scientific Implications

This study has shown that the majority of students who worked in the DLE scored relatively low on all measures of engagement, indicating that either there is room for improvement in students' digital reading behaviour, or that working in a DLE is less suitable for this group of students in terms of reading expository history

texts. However, students who did invest relatively more time in working with the DLE and showed higher levels of cognitive engagement consequently performed better on both measures of text comprehension. Simply stated, the more a student engages with working in a DLE, the better his or her comprehension and academic performance is expected to be. Therefore, in line with Van Rooij et al. (2017), we stress the importance of students' behavioural and cognitive engagement while reading in secondary education, especially when working with digital learning environments. Highly engaged students also show high levels of task value and intrinsic motivation. By stimulating these two aspects of motivation, teachers can indirectly foster students' engagement as well.

Using the engagement model of reading development by Guthrie and Wigfield (2017) as a conceptual model, our study adds to the scientific consensus that motivation, engagement, and reading performance are related, especially in the context of reading texts in a DLE. Although there are many ways to operationalise and measure students' engagement, this explorative study has shown that learning analytics, such as the use of digital log file data, and clustering these data through LPA can provide useful insights in students' real-time engagement when using technology for reading expository texts.

4



Chapter 4

Teaching reading strategies in history lessons: A micro-level analysis of professional development training and its practical challenges

This chapter is published as:

ter Beek, M., Opdenakker, M.-C., Deunk, M. I., & Strijbos, J. W. (2019). Teaching reading strategies in history lessons: A micro-level analysis of professional development training and its practical challenges. *Studies in Educational Evaluation*, *63*, 26–40. https://doi.org/10.1016/j.stueduc.2019.07.003

Abstract

Reading comprehension is an important skill in secondary education, yet many history teachers find it difficult to provide adequate reading strategy instruction. In this study, we designed a digital learning environment to support teachers' instruction of reading strategies based on student data. We provided history teachers in the experimental conditions with a visualisation of student performance data. Additionally, these teachers received professional development (PD) training and a guiding manual on how to translate these data into structured, explicit reading strategy instruction. Teachers in the control condition were only provided with basic data. We investigated teachers' personal experiences through micro-level analysis of qualitative interview data. Our results show that teachers in the experimental condition improved the variation of their strategy instruction and used modelling behaviour more often after the PD training. However, we also identified multiple contextual implementation barriers that provided us with important suggestions for future practice-oriented educational research.

3

conditions

9

teachers

3

research questions

Highlights

- History teachers acknowledge the importance of reading comprehension instruction.
- PD training in reading strategy instruction leads to higher instructional variety.
- PD training in data use does not automatically lead to actual extensive data use.
- Assessment of PD training feasibility requires small-scale, explorative research.
- Micro-level analysis of teachers' experiences is essential in educational research.

Introduction

Comprehending expository texts is an essential skill for students in secondary education, especially for the subject of history. While reading their history textbooks, students are required to infer difficult word meanings, identify main ideas, explain causal relationships, or question the author's assumptions. Therefore, it is essential for a student to know how and when to apply generic and subject-specific reading strategies (Hall, 2005; Heller & Greenleaf, 2007; Nokes, Dole, & Hacker, 2007; Shanahan & Shanahan, 2008). Reading strategies are deliberate actions that a student can attempt "to control and modify the reader's efforts to decode text, understand words, and construct meanings of texts" (Afflerbach, Pearson, & Paris, 2008, p. 368). Research has shown that explicitly teaching students to use strategies while reading can improve students' comprehension of text, which has led to an increased focus on literacy instruction in content areas (e.g., science or social studies) in the last two decades. Providing reading strategy instruction is no longer exclusive to language art class, but attainable in subjects like math, science, or history; reading strategies can be taught and applied across subjects (Hall, 2005; National Reading Panel, 2000).

Although the importance of integrating reading strategies in the history curriculum is widely acknowledged, observational studies in secondary education have shown that teachers do not always fully or correctly instruct them during their lessons. When observing lessons in Dutch secondary education, Linthorst and De Glopper (2015) calculated that 11.9% of the average lesson time was spent on reading instruction. When providing reading instruction, social studies teachers, including history teachers, mostly focused on vocabulary instruction and monitoring of students' comprehension. A similar study by Ness (2016), in which 600 minutes of lessons in middle school social studies classrooms were observed, showed that 10% of the overall lesson time was dedicated to reading strategy instruction, focusing on text structure, question answering, and summarisation. Finally, research suggests that content area teachers often do not feel responsible or qualified to provide reading strategy instruction, indicating a need for professional training in this area (Greenleaf, Schoenbach, Cziko, & Mueller, 2001; Hall, 2005; Ness, 2016).

Teaching Reading Strategies

Research has shown that the instruction of comprehension-fostering reading strategies in secondary school classrooms can elicit positive effects on students' academic performance (de Jager, Reezigt, & Creemers, 2002; Fisher & Frey, 2008;

Palincsar & Brown, 1984; National Reading Panel, 2000). Teaching reading strategies is especially effective for adolescent students of 12–13 years old; in the Dutch educational system, this is around the time when they experience the transition from primary to secondary education. A meta-analysis on reading strategy interventions in whole classrooms showed the largest effect sizes for intervention studies conducted with researcher-developed tests in grades 6–8, compared to other grades (Okkinga, van Steensel, van Gelderen, van Schooten, et al., 2018).

Unfortunately, there seems to be a decrease in content area teachers' actual literacy instruction around eighth grade, such as encouraging self-questioning, summarising, or monitoring comprehension (Guthrie & Davis, 2003). Research has shown that content area teachers encounter difficulties when incorporating literacy strategies in their lessons (Hall, 2005; O'Brien, Stewart, & Moje, 1995). Teachers and textbook methods focus mostly on asking questions about the content of the text (i.e., cognitive knowledge) and provide little explicit strategy instruction on how students can monitor or improve their reading comprehension (i.e., metacognitive knowledge; de Jager et al., 2002; Fisher & Frey, 2008). Moreover, disciplinary literacy instruction—instruction of reading strategies with a subject-specific approach, which is often regarded as even more effective—rarely occurs in content area classrooms (Moje, 2008).

Effective reading strategies. Students can apply several reading strategies before, during, and after reading to increase their comprehension of text. A meta-analysis by Donker, Kostons, Dignath-van Ewijk, and Van der Werf (2014) on effective learning strategies showed that a combination of cognitive and metacognitive strategies, such as orienting, planning, structuring, reflecting, and evaluating, led to positive results on measures of reading comprehension. More specifically, the practices of activating prior knowledge, defining difficult words, identifying main ideas, summarising, and reflecting on the contents of the text are all helpful strategies to support students' text comprehension (Afflerbach et al., 2008; Okkinga, van Steensel, van Gelderen, van Schooten, et al., 2018; Palincsar & Brown, 1984). Other strategies linked to text comprehension are expectation strategies (e.g., predicting the subject of the text), adjustment strategies (e.g., adapting reading behaviour according to one's comprehension of the text), and motivating strategies (e.g., focusing on the usefulness or pleasure of reading a text).

Instructional methods. There are multiple strategies to instruct, and multiple methods to instruct strategy use. For example, a teacher can explain strategies in

front of the classroom by providing verbal information. Furthermore, a teacher can also engage the students during the instruction by asking them questions about the instructional material. A possible way to further classify instructional modes is to make a distinction between implicit and explicit instruction, as done by Ellis (2009) and Dignath-van Ewijk, Dickhäuser, and Büttner (2013). It is important to note here that the distinction between implicit and explicit instruction is not straightforward, and that the terms have been operationalised in different ways in educational research (Ellis, 2009). A detailed description of the debate around implicit and explicit instruction is beyond the scope of this paper. We will follow the interpretation of Ellis (2009) and Dignath-van Ewijk et al. (2013).

First, teachers can implicitly provide instruction on reading strategies. According to Ellis (2009), in language education "implicit instruction is directed at enabling learners to infer rules without awareness" (p. 16). Thus, a teacher directs students to the application of a rule or strategy without explicitly focusing the attention on the strategy itself. Second, teachers can provide instruction in an explicit way. During explicit instruction, in contrast to implicit strategy instruction, the teacher elaborates on the application or the benefit of a certain strategy, or encourages students to reflect on it.

Another mode to provide instruction to students is known as modelling behaviour. Modelling refers to the explicit application of a strategy by using a first-person view (e.g., "*Before I start reading this text, I am going to think of what I already know about this subject*"). Modelling behaviour, or providing instruction while thinking aloud, is viewed as an effective way to foster students' strategic ability (Dignath-van Ewijk et al., 2013; Okkinga, van Steensel, van Gelderen, van Schooten, et al., 2018). However, modelling strategy use is difficult for most teachers in secondary education, because they often lack background knowledge about the use of reading strategies (Okkinga, van Steensel, van Gelderen, & Sleegers, 2018).

Teachers' attitudes towards reading strategy instruction. The ways in which teachers think about and teach reading strategies influences their instructional behaviour (Hall, 2005). Following constructivist theories, a teacher does not only transmit information and knowledge, but also has to facilitate and coach the students' learning process (de Jager, Reezigt, & Creemers, 2002). Intervention effects in the field of reading comprehension are dependent on teacher knowledge, behaviour, and instructional quality (Okkinga, van Steensel, van Gelderen, & Sleegers, 2018), and there is strong variation between teachers (Staman, Visscher, & Luyten, 2014). A

micro-level study by Seymour and Osana (2003), in which they evaluated four training sessions on the implementation of reading strategy instruction with two middle-level teachers, showed that the teachers did not fully understand the definition of specific strategies and, therefore, faced problems during the implementation. Furthermore, teacher efficacy (i.e., teachers' feelings of competence towards instruction) also influences teachers' instructional behaviour, and is shown to relate positively to students' reading comprehension performance (Chambers Cantrell, Almasi, Carter, & Rintamaa, 2013).

With regard to instructional behaviour, previous research has shown that teachers' skills expand over time; for example, a teacher first needs sufficient classroom management skills to be able to provide reading strategy instruction (van de Grift, 2014). The instruction of strategies, including reading strategies, does not occur frequently during classroom observations in secondary education, since it is considered a complex and difficult form of instructional behaviour that requires many years of teaching experience (van de Grift, 2014; van der Scheer, Glas, & Visscher, 2017). Another explanation for the fact that strategy instruction does not occur often is because teachers' knowledge about reading strategies is inadequate (Seymour & Osana, 2003), or because a teacher does not feel skilled to do so or responsible for providing this type of instruction (Hall, 2005; Ness, 2016).

Using Student Data for Instructional Practices

Educational technology provides an opportunity to support teachers' instructional behaviour. In recent years, digital data have transformed instructional practices in secondary education (Hutchison & Colwell, 2014). Teachers nowadays can draw from a large source of data, such as formative assessment results in student monitoring systems, to prepare their lessons and to meet their students' instructional needs. The process of using student data to inform instructional practice is also known as Data-Based Decision Making (DBDM; Schildkamp, Lai, & Earl, 2013), and its use has been associated with increased student performance (Campbell & Levin, 2009). However, only providing teachers with data is not enough. Mandinach and Gummer (2016) argue that teachers need to be data literate, which means that they are able "to transform information into actionable instructional knowledge and practices by collecting, analysing, and interpreting all types of data … to help determine instructional steps" (p. 367).

DBDM can occur at the school, classroom, and student level. For example, a

teacher can collect formative assessment data at a classroom level to inform and adapt his or her instructional behaviour (Hoogland et al., 2016; Schildkamp et al., 2013). In their review of effective differentiation practices, Deunk, Smale-Jacobse, De Boer, Doolaard, and Bosker (2018) found that teachers using computerized systems as a differentiation tool had small to medium positive effects on students' performance in primary education (d = 0.29). Similarly, a study in secondary education in New Zealand showed that a DBDM intervention, in which teachers collaboratively practiced profiling based on student assessment data, had positive effects on students' reading comprehension performance (Lai, Wilson, McNaughton, & Hsiao, 2014). This indicates that students benefit when teachers apply educational technology and assessment data to facilitate differentiated instruction, for example by adapting their instruction to meet the needs of low, average, and high-performing students.

Educational technology is also capable of providing teachers with visualisations of formative assessment log data, allowing them to see at a glance how their students perform and which students face difficulties with certain skills or assignments. However, this data is often very extensive and only easily interpretable for skilled, well-informed teachers (Vanhoof, Verhaeghe, Van Petegem, & Valcke, 2013), and therefore not common practice in secondary education. A Dutch study by Kippers, Wolterinck, Schildkamp, Poortman, and Visscher (2018) showed that secondary teachers mostly use pen-and-paper tests instead of digital assignments, limiting the possibilities for using advanced analyses of results for subsequent teaching. In addition, questionnaire and interview data showed that teachers only made use of data for instructional purposes in 25–50% of the lessons. This made Kippers et al. (2018) emphasise the need for professional development for teachers in DBDM.

Professional Development Training

Although the use of computerized systems with student data is known to have positive effects on students' performance, this is likely to be influenced by the professional development (PD) practices accompanying the implementation of these systems (Deunk et al., 2018). Providing teachers with student data is only helpful when teachers know how to effectively interpret and use these data for their instructional practice. Although there seems to be scientific consensus about what effective instruction based on student data entails, the use of data to guide instructional practices receives little attention in pre-service teachers' education (Mandinach & Gummer, 2013). Similarly, recent studies have shown that in-service teachers experience challenges

in using data (Hoogland et al., 2016; Mandinach & Jimerson, 2016; Staman et al., 2014; Vanhoof et al., 2013). Therefore, professional training in using student data for instructional purposes is also an important prerequisite for effective instruction in digital learning environments.

The need for continuous learning with regard to data use in education is widely acknowledged (Mandinach & Jimerson, 2016; Poortman, Schildkamp, & Lai, 2016). It is insufficient to only help teachers develop the necessary skills to analyse data, as is the case in most PD programmes (Marsh, 2012). Rather, teachers need to be able to integrate data skills with subject matter content knowledge and pedagogical content knowledge (Staman et al., 2014); yet, many existing PD programmes lack such a triangulation (Mandinach & Gummer, 2016).

The implementation of new teacher behaviour is strongly promoted when the PD intervention includes the provision of concrete content materials for teachers. In the context of reading comprehension, for example, it is essential to incorporate content knowledge about effective generic and subject-specific reading strategies (de Jager et al., 2002; Lai et al., 2014). PD training focusing on content knowledge and pedagogical content knowledge in the field of reading strategy instruction has shown promising results. With regard to content knowledge (i.e., knowledge about reading strategies), studies in the review by Hall (2005) showed that providing teachers with courses in content area reading can help teachers understand the benefits of teaching reading strategies, as well as create a positive attitude towards reading. With regard to pedagogical content knowledge (i.e., knowledge about instructional methods), De Jager et al. (2002) showed that training teachers to apply two models of instruction, more specifically the cognitive apprenticeship model and the direct instruction model, led to successful changes in teachers' instructional behaviour. Furthermore, a study by Fisher, Frey, and Lapp (2011) showed that teachers who were trained and coached to use think-aloud strategies in the classroom improved their instructional behaviour, and that as a result, their students improved their performance on a standardised reading assessment.

Studying the effectiveness of PD interventions. Based on scientific consensus about the critical characteristics of professional development, Desimone (2009) proposed a core conceptual framework for studying the effects of PD on teachers and students which "allows for testing the theory of teacher change (e.g., that professional development alters teacher knowledge, beliefs, or practice)" (p. 185). According to this framework, a PD intervention is expected to influence teachers'

knowledge and skills and change teachers' attitudes and beliefs, which in turn leads to changes in instruction (see Figure 4.1). Subsequently, the changes in instruction are expected to lead to improved student performance.

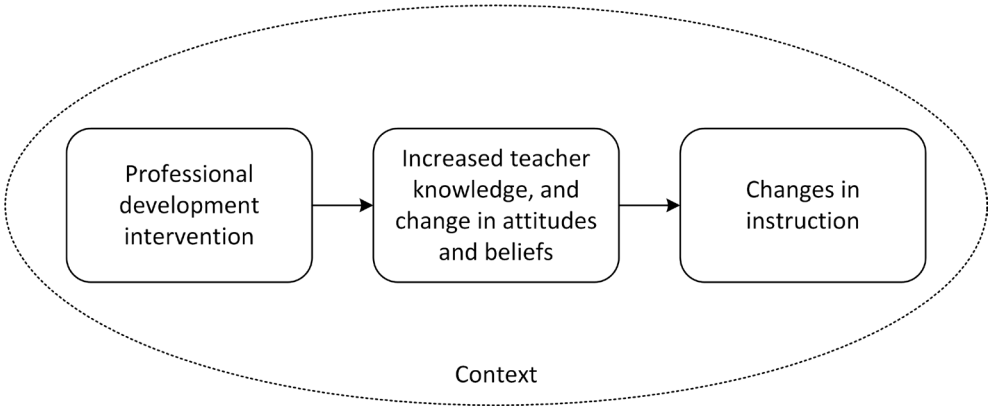


Figure 4.1 Conceptual framework for studying the effects of professional development on teachers (based on the framework by Desimone, 2009).

An essential component in Desimone's (2009) PD framework is the context in which the PD intervention occurs. The context is an important mediator and moderator in PD interventions; for example, interventions are influenced by teacher and student characteristics, and context factors at the school or classroom level. To provide a comprehensive account of effective components within PD intervention studies it is important to include the critical role of context factors.

Additionally, Desimone (2009) discusses appropriate strategies for measuring the effects of PD on changing teacher practice. She argues that mixed-method studies using interviews and classroom observations are "appropriate for providing narratives, examples, and anecdotes to answer research questions … describing and understanding the complexities of professional development in a specific context, how beliefs and attitudes change, and the processes through which teachers change their instruction" (p. 190). Additionally, surveys can obtain valid and reliable data on teachers' instructional practice, knowledge, and beliefs. However, the field of PD in education is still developing, and we need more extensive micro-level research to determine best practices.

Challenges of PD interventions. Although PD interventions in disciplinary literacy instruction or data use in education seem promising, there are many studies

reporting practical limitations and implementation challenges. Hoogland et al. (2016) highlight contextual factors such as the presence of a DBDM school culture, facilitation by means of time and resources, and PD as important prerequisites for successful data use in education. Staman, Timmermans, and Visscher (2017) found that teachers face difficulties when enacting data-based PD interventions designed by researchers. For most teachers, it is challenging to translate student data into differentiated instruction; they need professional training, clear goals, and professional guidance such as feedback on their teaching practice. According to O'Brien et al. (1995), the secondary school system with its content area divisions prevents both teachers and students from acknowledging the importance of disciplinary literacy as an essential part of the course itself. The authors argue that practical interventions should integrate literacy instruction in the regular curriculum in order to be effective. Lastly, Van Kuijk, Deunk, Bosker, and Ritzema (2016) found that behaviour of teachers participating in scientific research studies is probably influenced by the knowledge that their instruction is being studied, a phenomenon known as the Hawthorne effect. Therefore, it is challenging for researchers to determine whether effectiveness of PD interventions stems from the contents of the intervention itself.

Size and Scope of the Current Study

To our knowledge, there are currently no studies combining PD training in reading strategy instruction and the use of digitally visualised student performance data. The present study explores at a micro-level the effects of a PD training in providing reading strategy instruction enhanced by visualised student data. Interventions in the field of data use in education range from broad, comprehensive reform initiatives to narrowly focused interventions, such as local training programmes and workshops (Mandinach & Gummer, 2016; Marsh, 2012; Schildkamp et al., 2013; van der Scheer et al., 2017). Although PD in data use received a lot of attention from policy makers and researchers over the past two decades, most studies in this field only yield small effects. Therefore, Hill, Beisiegel, and Jacob (2013) advocate the use of rigorous research and explorative (i.e., single-site) analyses at the early stages of PD. The starting point of this type of research, or 'Stage 1', is a one-site pilot study in which the feasibility of the PD programme is analysed in practice using only a small sample of teachers (Hill et al., 2013), using rich, qualitative data. With regard to the length of the intervention, the authors suggest four to six sessions, which "could be undertaken within a single academic year" (pp. 479–480).

In the current study, multiple history teachers from four different secondary schools implemented the use of a digital learning environment (DLE) called 'Gazelle'[1] in which students read multiple expository history texts and answered multiple-choice questions about these texts (ter Beek, Spijkerboer, Brummer, & Opdenakker, 2018). Based on log file data from this DLE, we provided teachers with visualised student performance data and a single PD training session focused on the subject of strategic reading of history texts, followed by six accompanying lesson formats in a guiding manual.

The framework proposed by Hill et al. (2013) fits well within the current educational environment in the Netherlands. Dutch teachers in secondary education suffer from high levels of workload and stress, which often lead to burnouts or even attrition among teachers (Harmsen, Helms-Lorenz, Maulana, & van Veen, 2018). Many existing PD programmes cover longer periods from several weeks to an entire school year, requiring major time investments and efficient resource management from teachers and school management (Marsh, 2012; Okkinga, van Steensel, van Gelderen, van Schooten, et al., 2018; van der Scheer et al., 2017; van Kuijk et al., 2016). To test the feasibility of the present study design and to identify contextual factors that promote or impede the implementation of the PD intervention, we qualitatively explored teachers' personal experiences. By adopting a micro-level analysis design, this study can offer a unique insight in the combination of reading strategy instruction and data use in secondary history education.

Research Aims

This study explores to what extent the provision of visualised student data, combined with a professional development training in providing strategy instruction enhanced by analysing and using these data, affects history teachers' instructional practices in the context of reading strategy instruction. Additionally, it offers a valuable insight in how teachers experience this type of research and which difficulties they encounter. Inspired by Desimone's core PD framework (2009) and the guidelines of Hill et al. (2013) for studying PD, we focus on (a) how a small but focused PD training affects teachers' knowledge, attitudes, and beliefs with regard to teaching reading strategies, (b) how it affects the instructional methods teachers utilise, and (c) how teachers experience the use of a DLE with visualised data and a PD training. We will address the following research questions:

1 Gazelle is a Dutch acronym for 'Gemotiveerd en Actief Zelfstandig Lezen', which roughly translates into 'Motivated and Active Independent Reading'.

1. In what ways does a PD training in reading strategy instruction and data use affect teachers' knowledge, attitudes, and beliefs with regard to teaching reading strategies?
2. In what ways does a PD training in reading strategy instruction and data use affect teachers' instructional methods?
3. How do teachers experience the use of the provided DLE with visualised student data and the PD training in reading strategy instruction and data use, and which contextual factors promote or impede a successful implementation?

Method

Participants

In the school year of 2017–2018, nine history teachers from four Dutch secondary schools participated in a yearlong intervention study (cf. ter Beek et al., 2018). All teachers had a minimum of ten years' teaching experience. This was favourable because research has shown that novice teachers often still need to develop their basic teaching skills before they are able to master more complex skills, such as providing strategy instruction (van de Grift, 2014; van der Scheer et al., 2017). The teachers' mean age was 45.3 years (SD = 9.84) and on average they had 15.1 years' teaching experience (SD = 6.85; range 10–32; see Table 4.1). One of the history teachers was female (11.1%). All teachers taught history to seventh-grade students (M_{age} = 12.5 years, SD = 0.45), divided over 13 classrooms. From now on, we will refer to individual teachers using the pseudonyms mentioned in Table 4.1.

Teacher alteration. Due to workload issues, one classroom was divided over two teachers during the intervention. George taught this classroom during the first half of the school year, and Ian taught in the second half of the school year. Both teachers had comparable work experience (see Table 4.1).

Design and Context

We conducted an explorative mixed-method study with a quasi-experimental pretest-posttest design in an ecologically valid context. The study explores the effects of providing visualised data and a PD training on teachers' knowledge, attitudes, beliefs, and behaviour with regard to strategy instruction on a micro-level. Although the study

Table 4.1 Characteristics of participating history teachers and their classrooms (cf. ter Beek et al., 2018)

Teacher [a]	Gender	Age	Educational qualification	Years' work experience	Condition	Number of classrooms	Educational track	Class sizes
Alex	Male	38	Masters	14	Experimental A	2	Pre-university	26, 27
Barbara	Female	49	Masters	16	Experimental A	3	Pre-university	25, 26, 27
Chris	Male	59	Masters	32	Experimental A	2	Prevocational	28, 26
David	Male	61	Bachelors	12	Experimental A	1	Prevocational	21
Eric	Male	39	Masters	10	Experimental B	2	Pre-university	24, 29
Frank	Male	35	Masters	11	Experimental B	1	Pre-university	29
George	Male	50	Masters	12	Control	1 [b]	Pre-university	20
Ian	Male	36	Bachelors	11	Control			
Harry	Male	41	Bachelors	18	Control	1	Pre-university	22

Note. [a] All teacher names are pseudonyms. [b] George and Ian taught lessons to the same classroom; George during Phase 1, and Ian during Phase 2.

has been conducted at multiple schools, its explorative nature suits the 'Stage 1' type of research in the proposed PD research approach of Hill et al. (2013), because one of the goals is to analyse the feasibility of the PD programme design. All participating teachers had access to a digital learning environment (DLE, which is described in more detail below) with log file data output based on students' performance, but the data visualisations and additional PD training varied across conditions (ter Beek et al., 2018). We split the intervention into two phases to be able to analyse differences between and within conditions.

Research conditions. Four secondary schools participated in this intervention. Randomisation was carried out at the school level to ensure that all teachers within a school would be treated equally and to avoid contamination of the results among colleagues. This resulted in a quasi-experimental design with four teachers in Experimental group A, two teachers in Experimental group B, and three teachers in the control group (see Figure 4.2). We carried out the research in two consecutive phases, in which access to student data and the available support for teachers varied between conditions (see 'Phase 1' and 'Phase 2'). We conducted classroom observations during each phase and teacher questionnaires after each phase; these are referred to as T1 and T2 (see 'Procedure').

Data visualisations in the digital learning environment (DLE). Students from all history teachers' classrooms worked in a DLE in which students weekly read expository history texts about Greeks and Romans and subsequently answered ten multiple-choice questions about the text. All texts and questions were created by the researchers in cooperation with the participating teachers. Both Phases 1 and 2 consisted of six consecutive lessons. During lessons 2 to 5, students could consult supportive hints with cognitive and metacognitive strategy instruction in the DLE while reading the text and answering the questions. Cognitive hints focused on the content of the text (e.g., "A causal relation can be found in paragraph two: try to look for words like *because* or *therefore*"), while metacognitive hints aimed at students' regulation of their reading process (e.g., "Try to scan the text before reading to get an impression of what the text will be about"). After each multiple-choice question, students were asked to indicate their confidence in the correctness of their answer (i.e., judgement of learning) on a scale from 1 (*really unsure*) to 5 (*really sure*).

The DLE automatically translated log file data regarding students' performance, use of hints, time spent in the DLE, and judgement of learning into visualised data output for teachers. Furthermore, it presented individual and average classroom

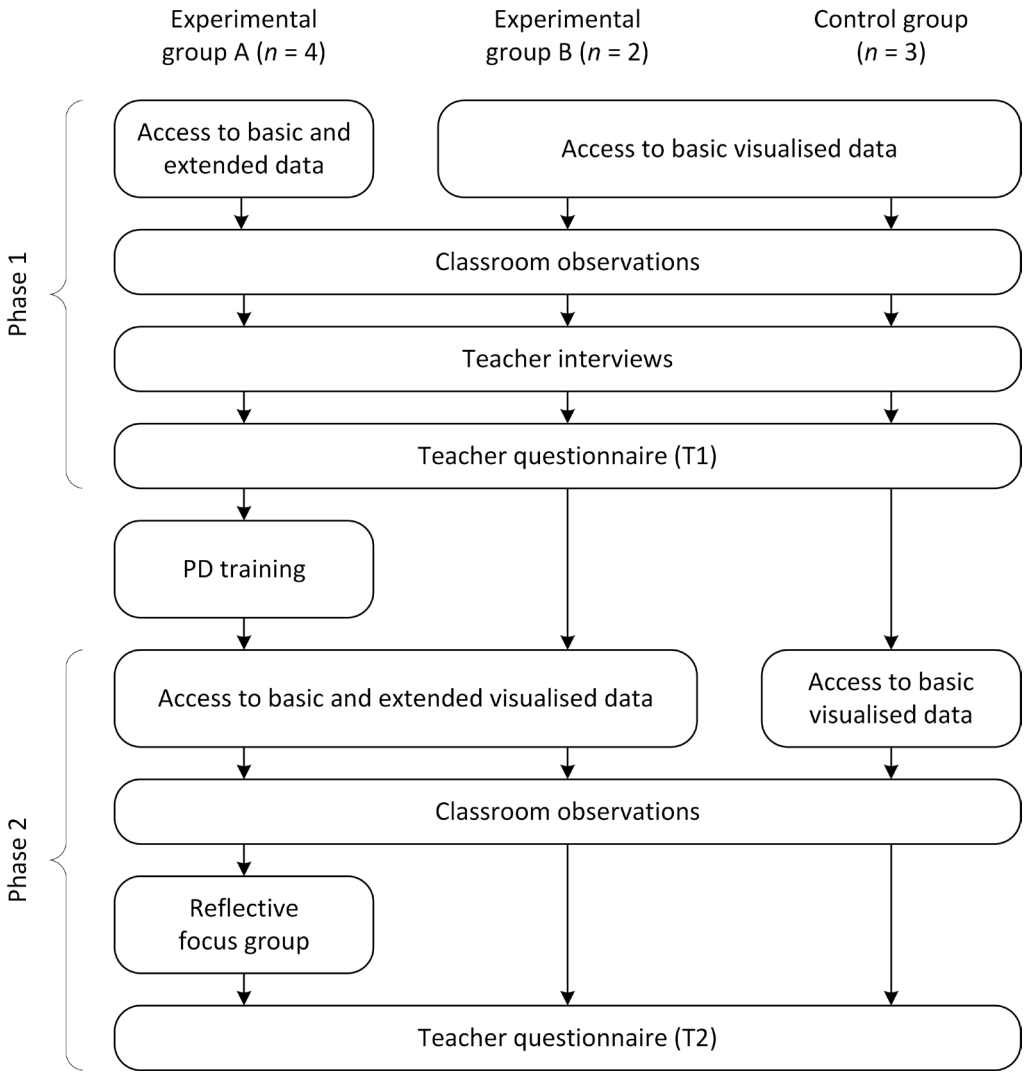


Figure 4.2 Research design and procedure for this study.

scores on various historical skills, such as recognising causal relationships or the chronological order of events, based on the students' performance on the multiple-choice questions. Lastly, the programme automatically assigned students to one of six possible profiles based on students' overall performance, hint use, and judgement of learning. Examples of these profiles are 'comprehensive readers' (i.e., students who have high reading comprehension scores and correctly judge their own learning), or 'inconsistent readers' (i.e., students who have low comprehension scores, but have

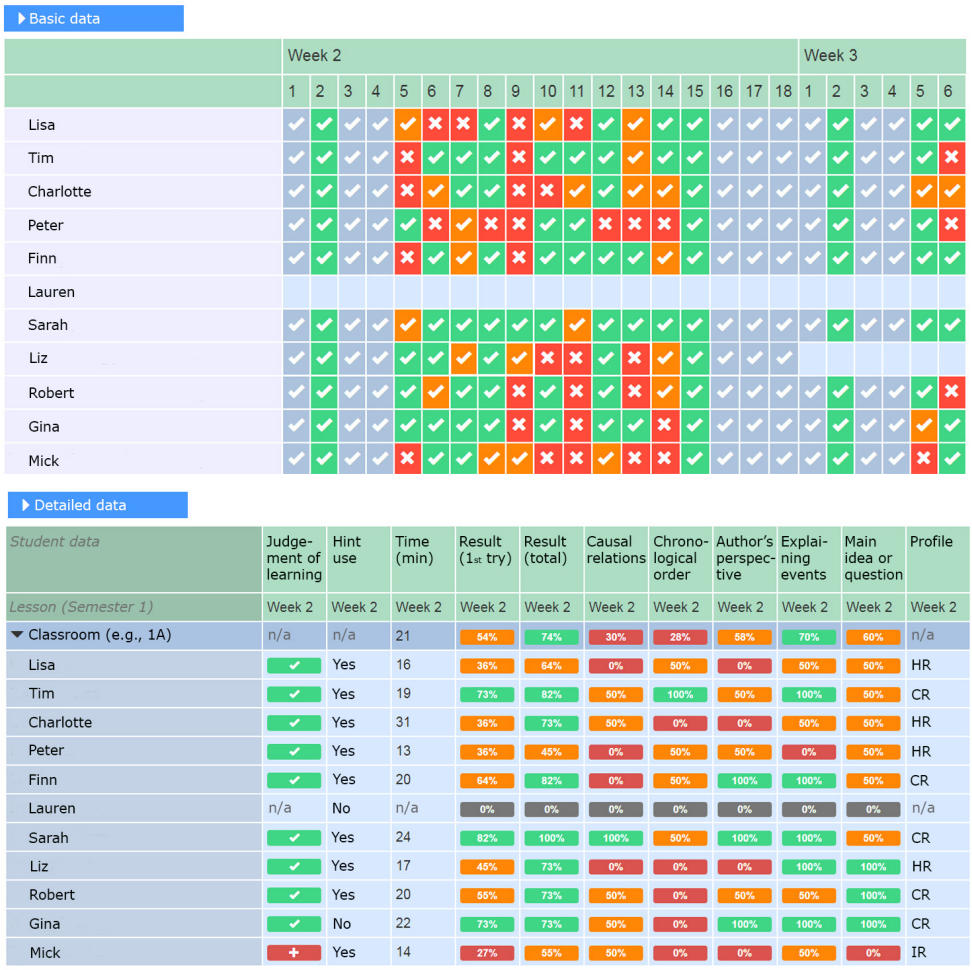
correct judgements, and use supportive hints – they are 'inconsistent' because one would not expect these students to have low performance scores). These extended, detailed data can help teachers to not only determine student differences and difficulties, but also the possible causes of it.

Figure 4.3 shows the different data visualisations from the DLE. For the basic visualised data (top), green check marks indicate that a students' answer was correct on the first try; orange check marks represent a corrected answer at the second try. A red cross indicates an erroneous answer at both first and second try. Grey squares with check marks resemble open-ended questions that were not automatically scored by the DLE. For the extended visualised data (bottom), a green check mark indicates correct estimation of performance; a red plus sign indicates overestimation, whereas a red minus sign indicates underestimation. Hint use is indicated with 'yes' (i.e., one or more hints used) or 'no'. Time spent in the DLE is shown in minutes per lesson. All performance scores between 1–33% are red; those between 34–66% are orange, and 67–100% are green. A grey score of 0% indicates that a student did not start the lesson yet.

Phase 1. During Phase 1, all teachers were provided with basic progress data, indicating whether students finished a lesson and whether their multiple-choice answers were correct or incorrect. Additionally, teachers in Experimental group A were provided with extended data visualisations about students' judgement of their own learning, time spent on task, use of supportive hints, and performance across various historical skills, such as recognising causal relationships or the chronological order of events. However, they did not receive any guidance or training to use or interpret the extended data during Phase 1.

Phase 2. Prior to Phase 2, the teachers from Experimental group A were provided with the same extended data visualisations as in Phase 1, in addition to the basic progress data. Additionally, they received a PD training with a guiding manual with instructions on how to implement explicit reading strategy instruction based on the provided visualised student data. The teachers in Experimental group B also received extended data visualisations in addition to the basic progress data during Phase 2, but no PD training. For teachers in the control group, the conditions were similar to Phase 1; they only received basic progress data.

The PD training and accompanying manual. To limit workload for the participating teachers in the context of high work pressure (Vanhoof et al., 2013), we



Student data	Judgement of learning	Hint use	Time (min)	Result (1st try)	Result (total)	Causal relations	Chronological order	Author's perspective	Explaining events	Main idea or question	Profile
Lesson (Semester 1)	Week 2	Week 2	Week 2	Week 2	Week 2	Week 2	Week 2	Week 2	Week 2	Week 2	Week 2
Classroom (e.g., 1A)	n/a	n/a	21	54%	74%	30%	28%	58%	70%	60%	n/a
Lisa	✓	Yes	16	36%	64%	0%	50%	0%	50%	50%	HR
Tim	✓	Yes	19	73%	82%	50%	100%	50%	100%	50%	CR
Charlotte	✓	Yes	31	36%	73%	50%	0%	0%	50%	50%	HR
Peter	✓	Yes	13	36%	45%	0%	50%	50%	0%	50%	HR
Finn	✓	Yes	20	64%	82%	0%	50%	100%	100%	50%	CR
Lauren	n/a	No	n/a	0%	0%	0%	0%	0%	0%	0%	n/a
Sarah	✓	Yes	24	82%	100%	100%	50%	100%	100%	50%	CR
Liz	✓	Yes	17	45%	73%	0%	0%	0%	100%	100%	HR
Robert	✓	Yes	20	55%	73%	50%	0%	50%	50%	100%	CR
Gina	✓	No	22	73%	73%	50%	0%	100%	100%	100%	CR
Mick	+	Yes	14	27%	55%	50%	0%	0%	50%	0%	IR

Figure 4.3 Examples for the basic visualised data (top) and extended, detailed visualised data (bottom). All student names are pseudonyms. HR = Help-seeking Reader; CR = Comprehensive Reader; IR = Inconsistent Reader.

carried out a short and single PD training at the school of Experimental group A. This 2-hour training covered several principles similar to reciprocal teaching (Palincsar & Brown, 1984), such as teaching comprehension-fostering reading strategies and expert modelling, and principles from structured or direct instruction (cf. de Jager et al., 2002), such as presentation of new content, guided and individual practice, and summarisation of content and evaluation. Additionally, the PD training covered both the use and the interpretation of the detailed visualised student data (Staman et al., 2014), as well as how teachers might integrate reading strategy use during their classroom instruction.

We supplemented the training with six 1-hour preparatory activities that teachers could complete individually in their own time. Inspired by the PD studies conducted by Ritzema (2015) and Van Kuijk et al. (2016), we constructed a guiding manual that involved three effective PD components: setting goals, acquiring relevant instructional skills in reading comprehension, and applying data use. These components were integrated in six guided lesson formats. For each of the six lesson formats, the first element in the manual is an informative text about the benefits and the use of two specific reading strategies (see Table 4.2). Secondly, three guiding questions help the teachers prepare their lesson by paying explicit attention to these strategies and by making use of the visualised data, for example to differentiate instruction for students who performed below average. Lastly, four reflective questions help the teacher to reflect on his or her lesson afterwards and to set new goals for lessons to come. At the end of the intervention, teachers had to evaluate the usefulness of the PD training and the accompanying manual using an evaluation form on the last page.

Table 4.2 Lessons, strategies, and visualised student data incorporated in the PD manual

Lesson	Central strategies to address by the teacher	Visualised student data to consult by the teacher
1	Motivating and Orienting	Previous comprehension performance in Lessons 1 to 6, Phase 1
2	Planning and Expecting	Time spent on Lesson 1, Phase 2
3	Structuring and Diagnosing	Summaries written in Lesson 1, Phase 2
4	Adjusting and Help Seeking	Hint use in Lessons 2 and 3, Phase 2
5	Evaluating and Reflecting	Tips formulated in Lessons 3 and 4, Phase 2
6	Subject-specific and Cross-subject strategies	Comprehension performance (per category) in Lessons 2 to 5, Phase 2

Procedure

With regard to the data collection procedure, this study has a mixed-method convergent parallel design (Dingyloudi & Strijbos, 2018). We collected all data sources separately but parallel; this enabled us to explore or confirm quantitative findings from the teacher questionnaires and classroom observations with qualitative,

in-depth interview data.

Classroom observations. During Phases 1 and 2, we conducted classroom observations to assess teachers' instructional behaviour and their variety in strategies used. We observed both regular and intervention lessons in weeks 2–5 to determine whether the training influenced teachers' general instructional behaviour. Since researchers or research assistants carried out all classroom observations individually, we recorded teachers' instruction using an audio recorder to be able to check the on-site coding afterwards. Six teachers gave consent to audio recordings of their lessons; three teachers (all in Experimental group A) only gave permission to observe the lesson without using a voice recorder. We observed at least two lessons in each classroom, resulting in 44 observations during Phase 1 and 40 during Phase 2. All observations comprised lessons of 50 minutes.

Teacher interviews. All participating teachers were willing to participate in an interview after Phase 1, and gave consent to record the interview with an audio recorder. The interviews were scheduled after students completed the last lesson in the DLE. We used a semi-structured interview format to ensure the uniformity of questions posed to every teacher, while at the same time allowing for flexibility with regard to teachers' remarks about the intervention. Each interview lasted approximately one hour.

Teacher questionnaire. We administered a self-report questionnaire about the teachers' knowledge, attitudes, and self-efficacy beliefs towards reading strategy instruction (i.e., T1). We presented this questionnaire to the participants after we conducted the interviews, to avoid influencing the interview results. The same questionnaire was administered after the reflective focus group meeting at the end of Phase 2 (i.e., T2).

PD training and reflective focus group meeting. Experimental group A received a 2-hour PD training prior to Phase 2. During this training, the teachers received a guiding manual to support their lesson preparation based on student data for all lessons in Phase 2. The PD training was not recorded. After Phase 2, all four teachers met in a reflective focus group setting, where we discussed the teachers' personal experiences with the PD training and the guiding manual. Similar to the interviews, we recorded this focus group meeting using an audio recorder.

Instruments

Observation instrument. Previous research on teachers' instruction of reading strategies in the classroom provided various reliable methods to observe classroom instruction (Dignath-van Ewijk et al., 2013; Linthorst & de Glopper, 2015; Smale-Jacobse, 2013; Smale-Jacobse & Timmerman, 2015). Because we used a digital programme specifically designed for this project, and to ensure ecological validity of our measurements, we composed a new observation instrument using elements from the aforementioned studies.

The observation instrument consisted of two parts. Part A assessed teachers' reading strategy instruction by focusing on the occurrence of 25 specific reading strategies, whose categorization (e.g., orienting, structuring, evaluating) was based on the effective learning strategies found in the meta-analysis of Donker et al. (2014). The individual items within these categories were based on the observed reading strategies in the studies by Linthorst and De Glopper (2015), Smale-Jacobse (2013), and Smale-Jacobse and Timmerman (2015). Strategy instruction could occur before reading a text (e.g., *Paying attention to specific elements of the text, such as illustrations or subheadings*), during text reading (e.g., *Monitoring text comprehension while reading*), or after reading (e.g., *Evaluating one's feelings or opinions about the text*).

The occurrence of reading strategies was to be scored categorically (i.e., did it occur, and if so, how?) instead of numerically (i.e., how many times a reading strategy occurred during the lesson). If a strategy occurred during the observed lesson, the observant had to indicate the mode in which it occurred: teacher explanation (E), questioning students (Q), or modelling the strategy (M). Moreover, explanations or questions could be specified as implicit (im) or explicit (ex) instruction (see Figure 4.4). Similar to the ATES instrument used by Dignath-van Ewijk et al. (2013), "implicit strategy instruction was coded every time teachers prompted the students for strategic behavior without addressing the strategic aspects of the behavior" (p. 343). Thus, implicit instruction does not focus specifically on the how or why of a certain strategy, but mentions it indirectly (e.g., "*Think of what you already know*").

In contrast to implicit instruction, explicit instruction includes step-by-step explanations of applying a strategy or elaborations on the usefulness of a certain strategy (e.g., "*Activation of prior knowledge before reading a text is a useful strategy for comprehensive reading, because it makes it easier for the brain to absorb new information*"). Whenever teachers elaborated on or questioned students about the

	E		Q		M	n/a
Structuring	im	ex	im	ex		
15. Summarising (parts of) the text.	☐	☐	☐	☐	☐	☐
Explanation:						
	im	ex	im	ex		
16. Indicating the main idea(s) of the text.	☐	☐	☐	☐	☐	☐
Explanation:						
	im	ex	im	ex		
17. Retelling in one's own words what the text is about.	☐	☐	☐	☐	☐	☐
Explanation:						

Figure 4.4 Excerpt from the classroom observation instrument (Part A)[2].

usefulness or the specific application of a particular strategy, we coded this as explicit instruction. For example, "*What do you do when you don't know the definition of a difficult word?*" is coded as explicit instruction by questioning students, whereas "*What does this word mean?*" is coded as implicit instruction. Due to the exceptionality of modelling behaviour, we conceptualised this as a separate mode of instruction and made no distinction between implicit and explicit instruction. Modelling behaviour was coded if a teacher phrased his or her instruction using a first-person view (e.g., "*I think that...*", "*I find this...*", or "*In my opinion...*"). Lastly, there was room for the observers to write down additional details, such as explanations for their choices made with regard to coding or general remarks about the observed instruction.

Part B of the observation instrument focused on the use of data output in the classroom. It indicated whether teachers referred to information from the DLE in their lessons (yes/no, e.g., "*I can see which students are lagging behind*"). It also focused on whether teachers paid specific attention to one of the variables in the detailed data visualisation, such as time spent on task or hint use – by mentioning something positive or something negative, by providing extra explanation, or by giving a specific assignment to their students.

Testing the observation instrument. The first author and four research assistants jointly trained the use of the observation instrument by coding fictional history lessons prior to the intervention. Unfortunately, there were no audio-visual materials available in which history teachers provided their students with reading strategy instruction. Therefore, to test interrater reliability, the first author drafted

2 The full instrument (in Dutch) is included in Ter Beek et al., 2018.

fictional utterances of teacher instruction[3]. Every observer had to assign 25 different fictional utterances to one of the 25 items on the observation instrument. After that, all had to indicate the mode of instruction. Since multiple observers coded the lessons and the items on the observation instrument were nominal, we used Krippendorff's alpha to establish reliability (Hayes & Krippendorff, 2007).

Analysis of the training scores yielded a Krippendorff's alpha reliability estimate of 0.92 for the scores on item level combined with the mode of instruction (E, Q, or M), which indicates good agreement between the observers. However, when analysing the scores including a more detailed distinction between implicit or explicit instruction (E_{im}, E_{ex}, Q_{im}, Q_{ex}, or M), the Krippendorff's alpha value was 0.66. Since a value of 0.66 is often considered the lower bound for reliability, the results of the analysis concerning the distinction between implicit and explicit instruction must be interpreted with caution (Hayes & Krippendorff, 2007; Krippendorff, 2004; Strijbos & Stahl, 2007).

Teacher questionnaire. We adapted three existing self-report questionnaires about teachers' knowledge, attitudes, or self-efficacy beliefs towards reading strategy instruction, and rephrased items to suit the context of secondary education. We only selected items that suited our research context, such as items focusing on comprehensive reading, to enhance the ecological validity of the instrument. Our sample size at T1 (n = 8) was too low to obtain good estimations of the reliability of scale scores; however, papers describing the original instruments report Cronbach's alphas ranging from .77 to .96 (see Table 4.3). Demographic items in the teacher questionnaire determined characteristics such as gender, age, and years of work experience.

The first scale, which we fully adopted from the instrument used by Dignath-van Ewijk and Van der Werf (2012), assesses teachers' knowledge on effective strategy instruction, based on the model of effective strategy instruction by Pressley, Harris, and Marks (1992). Teachers had to indicate the importance of certain ways of teaching strategies, each item starting with '*When teaching strategies, it is important to…*'. Eight items were measured on a 5-point Likert scale ranging from 1 (*totally disagree*) to 5 (*totally agree*).

The second scale, adapted from Meijer, Verloop, and Beijaard (2001), measures

3 The first author holds a Master's degree in history education; through practical experience, she could properly assess which forms of instruction often occur in classrooms. Therefore, training the coding of the lessons with fictional strategy instruction was the best possible option methodologically.

Table 4.3 Number of items, examples, and reliability indicators for the teacher questionnaire scales

Construct	Scale	*N* items	Example	Original instrument	Cronbach's α (original)
Knowledge	Teachers' strategy instruction knowledge	8	*'It is important to explain and model different learning strategies'*	Dignath-van Ewijk & van der Werf (2012)	.77 (8 items)
Attitude	Importance of reading strategies	5	*'Reading comprehension is important for students' general achievement'*	Meijer, Verloop, & Beijaard (2001)	.89 (10 items)
Self-efficacy beliefs	Teachers' sense of efficacy for literacy instruction	11	*'To what extent can you implement effective reading strategies in your classroom?'*	Tschannen-Moran & Johnson (2011)	.96 (22 items)

teachers' assumptions about the importance of comprehensive reading skills. We used this scale to indicate teachers' attitudes toward reading strategy instruction; if a teacher values students' comprehensive reading skills, we expect that they acknowledge the importance of reading strategy instruction. The original Importance Scale consists of ten items, equally divided over two segments: (a) the importance of

reading comprehension skills for student development and (b) the importance of goal setting in teaching reading comprehension. We only selected the first five items, since these were the most relevant for our study. Items were measured on a 5-point Likert scale ranging from 1 (*totally disagree*) to 5 (*totally agree*).

The third scale, based on the Teachers' Sense of Efficacy for Literacy Instruction Scale by Tschannen-Moran and Johnson (2011), assesses teachers' feelings of self-efficacy towards reading strategy instruction. We reduced the original 22 items to 11, omitting items focusing on instruction about strategies on other areas than comprehensive reading (i.e. writing, oral reading, and collaborative learning). All items started with '*To what extent are you able to…*' and were measured on a 5-point Likert scale ranging from 1 (*not at all*) to 5 (*very capable*).

Teacher interviews and reflective focus group meeting. We used a semi-structured interview format focused on four main topics: (a) questions about the importance of reading comprehension and how teachers provide regular reading instruction; (b) questions about the teachers' impression of their students; (c) questions about the contents of the DLE and the provided visualised data; and (d) a general evaluation of working with the DLE. The topics of the semi-structured interview format complemented the three components of the teacher questionnaire in a qualitative way, in line with the convergent parallel design of this study (Dingyloudi & Strijbos, 2018). During the reflective focus group meeting, similar topics were discussed, albeit in a more open-ended fashion.

Data Analysis

Quantitative data analysis. We used descriptive statistics using IBM SPSS Statistics 25 to analyse quantitative data from the teacher questionnaires and the classroom observations to answer research questions 1 and 2. Because of the small teacher sample, we analysed the questionnaire data on a descriptive and individual level. Since the teachers in our sample taught lessons to various classrooms and the number of observations varied between classrooms, we calculated the average number of different strategies observed (out of a maximum of 25) per lesson for each classroom. These data were also analysed descriptively to provide an explorative comparison of the observed instruction in Phases 1 and 2.

Qualitative data analysis. We used qualitative data from the teacher interviews and the reflective focus group meeting to answer the third research question. We

recorded and transcribed all utterances. Using Atlas.ti 8.3, the first author coded the interview data using emerging categories (Creswell, 2013) and four broad categories embedded in the semi-structured interview format: the importance of reading comprehension, teachers' impression of students, the provision of regular reading instruction, and the use of the visualised data in the DLE. The cross-tabulation option in Atlas.ti enabled us to create conceptually clustered matrices (Miles, Huberman, & Saldaña, 2014), which we used to analyse the different themes that emerged from the interview data. To illustrate our findings, we included verbatim quotations; for the sake of readability, we have removed any hitches or repetitions, provided that no important information was lost by doing so. The numbers following the quotations refer to the document and quotation number in the Atlas.ti dataset.

Results

Teachers' Knowledge, Attitudes, and Beliefs

The following results provide a descriptive overview of how the PD training in data use and reading strategy instruction affected teachers' knowledge, attitudes, and beliefs with regard to teaching reading strategies. Experimental group A consists of Alex, Barbara, Chris, and David; Experimental group B of Eric and Frank. George, Harry, and Ian belong to the control group. Unfortunately, T2 data are missing for Barbara. In addition, we were unable to make comparisons for George and Ian, since they both completed only one questionnaire (T1 and T2, respectively).

All participating teachers showed high levels of perceived knowledge about effective strategy instruction, with average scores ranging from 3.88 to 4.63 on T1 (see Table 4.4). After the PD training, the average scores of Alex and Chris decreased, while David's score increased. The average scores of Eric, Frank, and Harry decreased as well. Teachers' attitudes towards the importance of reading comprehension were very positive on T1, ranging from 4.00 to 4.80. Although for most of the teachers this average score remained high at T2, it slightly decreased for Alex and Chris. However, the differences were very small. The average self-efficacy beliefs towards reading strategy instruction on T1 ranged from 3.00 to 4.10. There is a slight increase visible at T2 for Alex, Chris, and David (Experimental group A), but also for Eric (Experimental group B). Frank's self-efficacy beliefs remained stable; Harry's score slightly decreased.

Table 4.4 Average scores on teachers' knowledge, attitude, and self-efficacy beliefs regarding reading strategy instruction at T1 and T2

Teacher	Condition	*M* knowledge		*M* attitude		*M* self-efficacy beliefs	
		T1	T2	T1	T2	T1	T2
Alex	Experimental A	4.50	4.00	4.80	4.60	3.00	3.45
Barbara	Experimental A	3.88	n/a	4.00	n/a	3.50	n/a
Chris	Experimental A	3.88	3.50	4.80	4.40	3.50	3.64
David	Experimental A	4.13	4.50	4.60	4.60	3.00	3.09
Eric	Experimental B	4.13	3.75	4.40	4.40	3.20	3.64
Frank	Experimental B	4.63	3.88	4.80	5.00	3.10	3.09
George	Control	4.25	n/a	4.80	n/a	3.10	n/a
Ian	Control	4.13	4.00	4.00	4.00	4.10	3.91
Harry	Control	n/a	3.50	n/a	5.00	n/a	2.91

Note. All teacher names are pseudonyms. T1 data are missing for Ian; T2 data are missing for Barbara and George.

Variety in Teachers' Instructional Behaviour

Phase 1. Teachers in all conditions did not differ descriptively in their average amount of various reading strategies used during Phase 1, with the exception of Barbara, who provided no reading strategy instruction at all. The average amount of various strategies used per lesson ranged from 0.00 to 3.25 (see Table 4.5). Detailed analysis of the observations revealed that strategies enacted in all conditions were mostly orienting strategies *before* reading, such as introducing the subject of the text or paying attention to specific elements of the text. During Phase 1, we never observed expectation or reflection strategies. More than half of the observed types of strategy instruction was implicit; with regard to the mode of instruction, explanation by the teacher was observed most often, followed by questioning students and modelling. Chris, George, and Harry did not model their instruction at all during the observations in Phase 1.

Phase 2. On average, teachers showed a higher variety of reading strategy instruction during Phase 2 compared to Phase 1, with the exception of Frank. The average amount of various strategies used per lesson ranged from 1.00 to 9.00 (see Table 4.5). There were no clear differences between the three research conditions. In Experimental group A, Alex and Chris showed a sharp increase in the average observed variety of strategies used. Overall, the various observed strategies were mostly orienting strategies, followed by adjustment strategies (e.g., control of text comprehension during reading). In addition, in Phase 2 we also observed motivating and reflecting strategies during classroom observations in Experimental group A, which was not the case for the other conditions. With regard to the mode of strategy instruction, there were no clear differences between Phases 1 and 2. However, Table 5 shows that Alex, Chris, David, and Harry modelled their instruction more often compared to Phase 1.

Use of the DLE. With regard to the use of the DLE in the classroom, we found that teachers occasionally discussed performance or progress results based on information provided by the DLE with their students during Phases 1 and 2, albeit mostly individually. Some teachers mentioned to their students at the start of the intervention in Phase 1 that they were able to consult information about students' progress. Additionally, most teachers encouraged their students to use supportive hints or to take their time when working in the DLE, but these comments were often quite general (e.g., "*Don't work too fast*" or "*You can click on hints if you like*"). Teachers

Table 4.5 Observed average strategy instruction per classroom per lesson in Phase 1 and Phase 2, specified by type and mode

Teacher	Class	Condition	*M* strategies (x/25)		% Explanation (E)		% Questioning (Q)		% Modelling (M)		% Explicit instruction	
			Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2
Alex	I	Exp. A	2.00	8.67	63.6	66.7	36.4	39.5	36.4	25.6	0.0	7.7
	II	Exp. A	2.00	7.00	60.0	65.1	22.2	22.5	10.0	34.9	20.0	0.0
Barbara	III	Exp. A	0.00	4.00	n/a	50.0	n/a	14.3	n/a	42.9	n/a	7.1
	IV	Exp. A	0.00	2.00	n/a	62.5	n/a	0.0	n/a	37.5	n/a	0.0
	V	Exp. A	0.00	2.00	n/a	66.7	n/a	0.0	n/a	33.3	n/a	0.0
Chris	VI	Exp. A	1.75	4.67	100.0	68.2	22.2	27.3	0.0	27.3	0.0	4.6
	VII	Exp. A	1.00	9.00	80.0	63.6	0.0	18.2	20.0	31.8	0.0	4.6
David	VIII	Exp. A	2.50	4.00	43.8	55.6	14.3	20.0	50.0	27.8	6.2	16.7
Eric	IX	Exp. B	3.25	6.75	63.2	64.3	10.5	10.3	31.6	35.7	5.3	0.0
	X	Exp. B	1.75	3.25	66.7	66.7	22.2	29.2	22.2	29.2	11.1	4.2
Frank	XI	Exp. B	2.50	1.00	61.5	60.0	16.7	0.0	30.8	40.0	7.7	0.0
George	XII	Control	3.00	n/a	70.0	n/a	0.0	n/a	30.0	n/a	0.0	n/a
Ian	XII	Control	n/a	1.67	n/a	71.4	n/a	0.0	n/a	28.6	n/a	0.0
Harry	XIII	Control	1.50	7.00	50.0	51.4	0.0	11.4	50.0	37.1	0.0	11.4

Note. All teacher names are pseudonyms. George and Ian taught lessons to the same classroom; George during Phase 1, and Ian during Phase 2.

in Experimental group A and B seldom referred to the detailed data visualisations. Harry's use of the DLE was exceptional; he used basic data from the DLE to grade his students, even though we explicitly demanded not to use the data for grading purposes[4]. Apart from that, there were no visible differences regarding the use of the DLE between teachers in different conditions in both Phases 1 and 2.

Teacher Experiences

Several themes concerning teachers' experiences with using the DLE in the classroom and consulting the visualised student data emerged during the coding of the interview data. Furthermore, during the focus groups, we also discussed the experiences with the PD training and the guided manual with teachers from Experimental group A (i.e., Alex, Barbara, Chris, and David), which led to the identification of practical and contextual barriers for implementation. This section describes the main findings, using quotes from the interviewed teachers to clarify and support them.

Importance of reading comprehension. Barbara, Chris, David, and Frank explicitly acknowledged the importance of reading comprehension for the subject of history, calling it "crucial" (Barbara, 4:66) and "part of my subject" (Chris, 5:62). Alex and Eric both mentioned the importance of identifying main ideas: "to grasp the storyline … and what is more and less important" (Alex, 3:45). Chris, David, and George focused in their interview on the importance of recognising causal relationships in a text, because "history tries to put the mush of the past in a logical order" (Chris, 5:57). Although most teachers were positive about this subject, George was the only one to explicitly state something negative about providing disciplinary literacy instruction: "For me, answering questions at the level of reading comprehension is not the same as [history] education. I struggle with that" (7:12).

Teachers' impression of students. During the interviews, all teachers expressed concerns about students' reading and concentration levels. Students' reading levels vary within classrooms and teachers tend to focus most on the students that have difficulties with reading comprehension. However, for Chris, his general impression of students' reading levels was low:

> "What they run into, not only in seventh grade, but also in eleventh grade, is that one does not know what is actually behind all the words, that you, as it were,

4 A research assistant discovered the use of student data for grading during one of the classroom observations. During the interview, Harry argued that he needed to do so "because otherwise, I cannot assess the ancient Greeks at all" (8:12) and because his students "are motivated by grades" (8:13).

> continue to swim on the surface, whilst the essence of a text lies at a depth of three meters." – Chris, 5:52

This was also the case for David, who mentioned that the reading levels of his students were "poor, I think. Yes, poor." (David, 6:28). It is important to note here that Chris and David taught lessons to prevocational students, who, on average, have lower comprehension performance scores. For students in pre-university education, teachers more often considered motivation problematic. For example, Eric was concerned about his students' motivation to work in the DLE: "The motivation for history is now associated with sitting down and reading texts for long periods of time – boring and much of the same" (Eric, 1:48). In addition, David, George, and Harry made remarks about the fact that their students are highly motivated by grades: "They like that reward structure" (Harry, 8:35). They were concerned that students' motivation to work in the DLE was low when they did not receive a grade afterwards.

Regular reading instruction. We encountered little reading strategy instruction during the classroom observations. Some teachers acknowledged that they did not provide extensive reading instruction during their lessons: "It is reading, and that is nice and all, but I really prefer telling stories" (Harry, 8:20). Even when teachers were aware of their students' reading problems, there was a preference for regular content instruction, as stated by Frank:

> "Yea, they just like it when I tell them something about history. When it is very teacher-driven. And we have received test results, and well… they just have problems with texts, with reading. They have the vocabulary of a cucumber." – Frank, 2:33

Alex, Barbara, Eric, and George noted that they sometimes embed strategy instruction in their history lessons, but this instruction is mostly focused on how to study instead of how to read. Alex also emphasised the fact that his students have different preferences, so they have to "decide for themselves which method suits them; for one, it is questioning the text, for the other, it is better to make a mind map, … and another student does not understand anything at all" (3:49).

Teachers' use of the DLE. When we asked teachers whether they consulted the visualised student data, most of them mentioned using only the basic information (i.e., whether students finished a lesson or not and the correctness of their answers), even the teachers who had access to detailed data as well. For example, Eric mentioned that he "looked at who was on schedule, but not if the answers were correct – that is

probably more relevant for you [researchers]" (Eric, 1:27). Teachers in Experimental group A did not often mention the detailed data visualisations, except for David:

> "I found it very interesting to see it for the first time, and that the results are built up. And the fact that a student ends up in a certain profile at a given time, which can change, and once that happens, you can do things with that. I really liked that." – David, 6:4

Barbara noted during the focus group meeting that the basic progress data supported her differentiation practices because the data visualisations in the DLE confirmed the image she already had of her students. David also mentioned that he looked at the data, but found it "a bit too premature" to adapt his instruction (David, 6:20). In contrast, Chris never consulted any type of data in the DLE during the intervention: "I was overwhelmed by this period … I just was not able to do it" (Chris, 9:18). This finding illustrates the discrepancy between the execution of the research project as planned by the researcher versus the actual execution by the participating teacher. Because of workload issues and time constraints, Chris decided not to use the data visualisations, despite the fact that teachers in this experimental condition were stimulated to do so through the PD training and the accompanying manual.

The guiding manual provided teachers with preparatory questions for which they had to consult the detailed data visualisations in the DLE. However, we were not able to analyse the actual use of the guiding manual. Because three teachers in Experimental group A did not return their manuals to the researchers as requested, it remains unclear if and to what extent teachers completed the assignments. David did return his manual, but only filled in Lessons 1 and 2 due to workload issues. For the third lesson, he only wrote, "I have no time to do this. Unfortunately." In fact, teachers in all conditions mentioned workload issues and other practical and contextual barriers, which we therefore decided to analyse and report separately.

Practical and contextual barriers. During the intervention, we noticed that not all teachers implemented the project lessons as planned. Through analysis of the interview and focus group data, it became clear that the intervention suffered from various individual and contextual problems. These factors did not only influence the outcomes of this study, but also probably play an important role in practice-oriented research in general. Insight into practical and contextual barriers is essential for designing and conducting research interventions in the field of PD training in education; therefore, we have enlisted them below.

False notion of the research project. The interview data indicated that three teachers in Experimental group A had a false notion of the independence of the DLE with regard to the regular curriculum and that they were afraid to interfere with the goals of the research project. For example, we did not observe any strategy instruction utterances in Barbara's lessons. During the interview, she stated:

> "I was under the impression that the research question concerned how [students] could independently learn to apply certain strategies within a self-contained environment. So, I have very deliberately disconnected my own lessons and just put them behind the computer. Everything they needed was in the environment." – Barbara, 4:1

Alex probably had the same impression, since he stated that "The idea of [the DLE] was not to, no interference; you just put them to work. You do not comment on anything" (Alex, 3:9). In addition, David remarked that he felt insecure because of the research aspect:

> "I did not trust myself in some things. There is this research project, so a lot is at stake. You need to behave precisely, or else… Whether I did the right things; that was a little painful." – David, 6:10

After the interview, we explained to these teachers that they were allowed to provide any form of instruction, but these false notions probably hampered their instructional behaviour during Phase 1. After Phase 2, Barbara and Alex mentioned that they integrated the contents of the DLE in their lessons more often. For example, Alex started each lesson with instruction about a certain reading strategy, such as orienting and summarising: "I have incorporated the lessons from [the DLE] in my own lessons. I have my own message, that I want to make clear, and I have connected that with the stories in [the DLE]" (Alex, 9:10).

Integration with the regular curriculum. Although we discussed the content and the order of the DLE texts with the teachers in advance, Alex, Barbara, Chris, and Harry complained that the contents of the programme did not fully align with the subjects being taught in the regular history lessons. In the Dutch seventh-grade history curriculum, regular methods consist of demarcated periods (e.g., 'the Time of Greeks and Romans'), and teachers can decide for themselves when they start a new period. Therefore, the teachers in this study differed in the subject they were working on with their students during Phases 1 and 2, despite the fact that we asked them to focus on the subject of Greeks and Romans during Phases 1 and 2.

> "For example, we were still working on hunter-gatherers, you know, in prevocational education. We have just started with Egypt and the pharaohs, and then they get a text about barbarians, and trade in ancient Greece, and Greek gods, those things. Well, I did not talk about that in class yet." – Chris, 5:23

Additionally, these teachers noticed during the interviews that they did not integrate the contents of the DLE and their regular lesson materials during Phase 1.

Time pressure and lesson preparation. Time pressure was often mentioned during the interviews and during the focus group. Alex, Barbara, Chris, and George complained about the density of the six-week lesson structure of the project and difficulties in combining the DLE with their regular curriculum:

> "I find the time pressure very high, so I would prefer not to say 'you have to finish everything within eight weeks', but you have to be able to spread things. So rather, 'you have ten weeks to complete eight lessons', something like that." – Alex, 3:27

The time pressure possibly also led to little preparation of the lessons. Teachers often did not prepare their lessons according to the guiding manual, and mentioned that they did not read the texts in the DLE beforehand, so they were not fully acquainted with the contents of the programme.

Logistic problems and IT facilities. In some classrooms, students worked on individual laptops, according to their school's bring-your-own-device policy. However, Alex, Barbara, Chris, George, Harry, and Ian had to make use of central computer rooms or laptop carts, which had to be reserved beforehand. This led to logistic problems and sometimes a slightly different implementation of the research project; for example, when teachers used the DLE lessons as homework assignments.

> "Well, this was very enlightening, because digitisation is high on the agenda at our school. Twenty-first-century skills. But this project alone already shows that there are still some limitations … we have laptop carts, but there are not enough laptops. There are always a few broken, and well, if you have a slightly larger classroom with 27 students, you already have a problem." - Barbara, 4:18

However, even in the schools where students had to bring their own devices, the teachers encountered problems:

> "Their Chromebooks are often not fully charged. If that is the case, I send them to the school's media library and I hope they will do everything there, and that they will return as soon as they are finished, but I lose sight … sometimes, the

> laptops from the media library are all lent. I think there are thirty, and a few are broken. So, students are dependent on their Chromebooks" - Eric, 1:23

If teachers encountered logistic problems during their lessons, it was difficult for them to focus their instruction on reading strategies or to use student data from the DLE to adapt their instruction. In some cases, instructional time was lost due to issues concerning the IT facilities. The problems with computers and laptops also caused irritation among some of the teachers, which did not benefit their motivation to work with the student data from the DLE.

Conclusions and discussion

This mixed-method study evaluated the extent to which the provision of visualised student data (with or without PD training in data use and reading strategy instruction) affected in-service history teachers' instructional knowledge, attitudes, self-efficacy beliefs, and behaviour using self-report questionnaires and classroom observations. Additionally, the teachers' personal experiences with the intervention were explored using qualitative interviews and a focus group meeting. By triangulating quantitative and qualitative data sources, we were able to explain our findings and highlight factors that might influence the implementation fidelity of practice-oriented, data-driven interventions.

Findings

In general, the history teachers reported having high levels of perceived strategy instruction knowledge, and they acknowledged the importance of reading comprehension skills. There were no visible differences between or within conditions after the PD training; the high scores on the pretest remained relatively high on the posttest. This finding can be explained by the fact that the scores on the pretest were already very high, possibly creating a ceiling effect (cf. Staman et al., 2014). Perceived knowledge of two teachers in Experimental group A decreased slightly after the training; these teachers might have realised after the training that they did not know as much as they initially thought they did. It is therefore also hard to link changes in teachers' knowledge, beliefs, and attitudes to changes in their instructional behaviour, as proposed in Desimone's (2009) framework. Furthermore, compared to their knowledge and attitudes, teachers' self-efficacy beliefs regarding reading strategy instruction were slightly less positive. This finding is in line with earlier research, which indicates that although teachers value reading instruction,

they often believe they do not have sufficient skills (Greenleaf et al., 2001; Hall, 2005). After the intervention, there was still a difference between teachers' attitudes and self-efficacy beliefs, indicating that teachers consider reading strategy instruction to be important, but do not always appear to consider themselves fully capable to provide this type of instruction.

With regard to the instructional behaviour, teachers who received a PD training and a guiding manual prior to Phase 2 employed a higher variety of reading strategies during their classroom instruction compared to Phase 1. However, it is important to note that two teachers from Experimental group A deliberately did not intervene with the programme during Phase 1, because they feared they would disturb the research project by doing so – a phenomenon known as the experimenter expectancy effect (Rosenthal, 1976). This explains that Barbara, for example, initially provided no reading strategy instruction at all: her false notion of the research project led her to provide *less* instruction during Phase 1, which presumably led to *more* observed strategy use during Phase 2 (i.e., it was not necessarily influenced by the PD training). In addition, teachers in Experimental group B and the control group also improved the variety of the instructional strategies used, making it hard to attribute this finding solely to the PD training. Although we focused on the variety of instructed reading strategies, the relatively low numbers resonate with earlier research on reading instruction in social studies classrooms (Linthorst & de Glopper, 2015; Ness, 2016).

The observed mode of instruction varied between teachers; however, most of the observed reading strategy instruction was provided as teacher explanation, followed by questioning students. Modelling behaviour occurred rarely as an instructional strategy; nevertheless, it is important to note that the teachers in Experimental group A practised modelling more often in Phase 2 compared to Phase 1, and compared to the other teachers (with the exception of Harry). As stated earlier, the results with regard to the type of reading strategy instruction (implicit, explicit) must be interpreted with caution due to the minimal interrater reliability of this aspect of the observation instrument. Nevertheless, we found that the reading strategy instruction provided also varied between teachers, but it was not influenced by the experimental conditions. For all teachers, the majority of the observed instruction was implicit; they mostly tell students what to do, instead of how or why they should do it.

Although research has shown that professional development in disciplinary literacy or data use is a long-term process that requires continuous commitment and a supportive school culture (cf. Hoogland et al., 2016; Moje, 2008; O'Brien et al.,

1995; Timperley, Wilson, Barrar, & Fung, 2007), it is difficult to implement long-lasting interventions. For practical and financial reasons, PD training initiatives to improve teacher effectiveness mostly occur on a small scale, for example locally (e.g., one school) or regionally (e.g., school district). These types of initiatives often result in small effects (Hill et al., 2013). Because of the small teacher sample in this study and the issues regarding implementation of the PD intervention, it was not feasible to calculate effect sizes for the different conditions. Teachers often did not fully or correctly implement the instructional practices from the PD training and the accompanying manual. For example, some teachers did not consult their students' data output before each lesson, while others never provided reflective strategy instruction after their students read a text. These implementation difficulties are similar to findings from previous research on data use and instruction (Dignath & Büttner, 2008; Donker et al., 2014; Kippers et al., 2018; Okkinga, van Steensel, van Gelderen, & Sleegers, 2018; Ritzema, 2015; Staman et al., 2017; Vanhoof et al., 2013; van Kuijk et al., 2016).

The practical and contextual barriers that teachers mentioned during the interviews and the reflective focus group caused implementation difficulties. We discovered that some of the participating teachers were restricted by a false notion of the research project, the integration of the contents of the DLE with the regular curriculum, little preparation of the lessons due to time pressure, and logistic problems such as the unavailability of computers. Although they cannot be directly linked to the results, it is important to acknowledge these factors in the context of the current study and future PD intervention studies.

Although we accept the fact that the practical and contextual barriers might have been frustrating for teachers to work with, we also noted that teachers did not provide us with practical solutions or strategic actions to address these problems. Similar to studies found in the review by Hall (2005), teachers blamed the curriculum, the textbook materials, or the students' reading motivation for not being able to provide effective strategy instruction. For example, Frank stated that his students "have the vocabulary of a cucumber" (2:33), but did not express any thoughts on how to improve his students' reading skills. It seems that a positive attitude towards reading strategy instruction, as displayed by all teachers in this study, is not enough to establish adequate reading strategy instruction in the history classroom. It is therefore important to continue the research on subject-specific literacy instruction and the factors that stimulate or hamper teachers' implementation of it.

Limitations

A methodological limitation of this study concerns the way in which we coded the lesson observations. Due to limited resources, lesson observations were done by individual researchers instead of researchers working in pairs. Therefore, we decided to adapt our observation instrument in such a way that the occurrence of reading strategies was to be scored categorically instead of numerically, to lower the risk of missed information during the observations. By doing so, we were unable to tell whether teachers provided *more* or *less* reading strategy instruction overall as well as for strategies specifically, as is often the case in this type of research (Dignath-van Ewijk et al., 2013; Smale-Jacobse, 2013; Smale-Jabobse & Timmerman, 2015). Nevertheless, our results provide valuable insights in teachers' variety of their instructional repertoire before and after a PD training.

Another limitation concerns the involvement of teachers in designing and preparing the research intervention. In our study, educational researchers designed the intervention and the contents of the PD training, while teachers conducted its practical application. Moreover, we discussed with teachers the reading problems they had identified in their classes, but did not test our assumptions about teachers' own PD needs prior to the training. Theories of practitioner research suggest that the validity of educational research increases when teachers are involved in designing and conducting rigorous research (Lai & Schildkamp, 2013; Robinson & Lai, 2006; Vanhoof et al., 2013). When teachers' theories of action and local knowledge are taken into account in the process, a PD intervention is expected to be tailored to the needs of the teacher, increasing its effectiveness. In addition, confusion about the research purpose or incorrect assumptions about researchers' expectations, such as the ones that occurred in the current study, might be prevented.

We observed both regular and intervention lessons to determine whether the provision of visualised data or a PD training influenced teachers' general instructional behaviour. However, we concluded that reading texts occurred sparingly during regular lessons (cf. Ness, 2016), resulting in a low occurrence of instructional behaviour with regard to reading strategies. Yet, this does not imply that the PD training had no effect. For example, in several lessons Barbara let her students work on creating a historical newspaper and mentioned that by doing so, the information from the PD training was less relevant to her regular lessons at that time. Due to the limited number of observed lessons per classroom, and the high variability among the teachers, the results with regard to teachers' instructional

behaviour must be interpreted with caution. We need more large-scale research to explore further the effects of PD trainings in reading strategy instruction and data use on teachers' instructional behaviour; hence, our explorative study might provide helpful suggestions.

Suggestions for Future Research

We identified several practical and contextual barriers that provided us with important suggestions for future practice-oriented research in the context of teaching reading strategies and using data in education. First, although PD research has been conducted in secondary education before, we noted that this educational context is challenging for practical research projects. During the interviews, many teachers mentioned issues concerning workload, time constraints, and other logistic problems (e.g., the availability of IT facilities, or the loss of lessons due to unexpected schedule changes). Combining the regular curriculum with the requirements of the intervention study, such as preparing the lessons based on student data and incorporating reading strategies during the lessons, might have been too complicated to perform in two fifty-minute lessons a week. With regard to the IT facilities, which Livingstone (2012) argues to be crucial for optimising learning with technology, it is important that future research avoids or reduces these contextual barriers as much as possible.

Second, in order to use digital tools effectively for whole-classroom instruction, good classroom management (i.e., applying basic instructional skills) is an important prerequisite (Okkinga, van Steensel, van Gelderen, & Sleegers, 2018; van de Grift, 2014). In our classroom observations, some teachers seemed to be more concerned with classroom management than with providing explicit reading instruction; teachers mainly used the available data visualisations to check whether students were on schedule, instead of using it for differentiation practices. This is in line with earlier research by Duffy (1982), who noted that with regard to reading instruction, "to the extent that interactive decisions are made, they seem to be associated more with management than with instruction" (p. 359). Future practical research on strategy instruction should include context factors like the classroom or individual students, to determine their role in the effectiveness of research interventions.

Lastly, the current study did not address the effects of teachers' altered instruction on students' academic performance, which is the last step in Desimone's (2009) PD framework. Future research should therefore assess whether and how

changes in teachers' instruction can be sustained on the long term. In addition, it is valuable to analyse the effects of teachers' strategy instruction on students' academic performance, self-regulated learning skills, or motivation to learn, to establish a comprehensive view of the effects of targeted PD training in data use. Finally, in line with Hill et al. (2013), we advocate the use of micro-level analysis in the field of PD interventions in education. The results from the present study illustrate the value of using micro-level analysis and including teachers' personal experiences in effectiveness research, which contributes to the development of future, large-scale PD design interventions.

5

Chapter 5

Fostering students' historical content knowledge and historical reasoning ability through subject-specific reading skills and reading engagement

This chapter has been submitted for publication as:

ter Beek, M., Opdenakker, M.-C., Deunk, M. I., & Strijbos, J. W. (2019). *The role of reading skills and engagement in history education.*

Abstract

The ability to apply various reading skills is an important prerequisite to comprehend expository texts commonly found in history textbooks, but it is unclear which specific skills contribute to students' historical content knowledge and historical reasoning abilities. This study used a Digital Learning Environment (DLE) to measure and support lower secondary students' subject-specific reading skills and engagement, and explored the relations with students' historical content knowledge and historical reasoning ability. Results showed that subject-specific reading skills, such as explaining historical events, correlated significantly with both historical content knowledge and historical reasoning ability, but not all skills were unique significant predictors. Moreover, students who showed high behavioural and cognitive engagement while reading performed significantly better on delayed content knowledge and reasoning tests compared to students with lower engagement. These findings indicate that to promote the advanced practice of historical reasoning, history education should pay attention to students' reading skills and engagement.

3

schools

197

students

2

research questions

Highlights

- Students' subject-specific reading skills are all somewhat related to each other, as well as to their historical content knowledge and historical reasoning ability.
- At a retention test 4–6 weeks after the intervention, students who showed high engagement while working in the DLE performed better than those who did not.
- Explaining historical events and generating historical questions are unique predictive skills for both students' historical content knowledge and their historical reasoning ability.

Introduction

Many authors concerned with history education have advocated for a specific disciplinary approach to literacy instruction in history classrooms, emphasising the importance of practices such as 'reading like a historian' and promoting historical thinking and reasoning (e.g., perspective-taking, contextualising, or sourcing information) for students' comprehension of historical texts (Moje, 2015; Monte-Sano, 2011; Reisman, 2012; Shanahan & Shanahan, 2008; Wineburg & Reisman, 2015). According to Wineburg and Reisman (2015), students who are only able to implement basic reading comprehension strategies "will remain spectators, passively gazing at the arena of knowledge production" (p. 636). Although we agree that a disciplinary literacy approach offers a valuable framework for sophisticated reading and understanding of historical texts, we also stress the importance of combining this approach with generic reading skills for reading and comprehending the expository format often found in regular history textbooks. Expository texts differ from the narrative texts that are more common in primary education, and the reading thereof can be challenging for adolescent students who recently transferred from primary to secondary education (Fry & Gosky, 2007) or for those who struggle with this type of texts (Faggella-Luby, Graner, Deshler, & Drew, 2012; Okkinga et al., 2018). In addition to reading skills, student engagement in reading is considered important for students' reading performance as well: Students do not only need to possess relevant reading skills, but must also actively apply these skills while reading (Guthrie & Klauda, 2016).

Reading Expository History Texts

In the lower secondary education history curriculum of many countries, the transfer of information relies heavily on the expository format commonly found in textbooks. More specifically, in the Dutch history curriculum, the entire world history—from prehistoric times until today's information age—is covered in a span of three years. As a result, history textbooks often include fact-dense texts that contain an abundance of novel concepts, perspectives, and vocabulary (Afflerbach & VanSledright, 2001; Mastropieri, Scruggs, & Graetz, 2003; Ramsay, Sperling, & Dornisch, 2010). Therefore, students who recently transferred from primary to secondary education will need to adapt their reading process accordingly and continue their development towards subject-specific reading proficiency for the subject of history (Alexander, 2003; VanSledright, 2004). In the following subsections, we will highlight three

literacy approaches that can support this developmental process: generic reading strategy instruction, disciplinary literacy practices, and the practice of focusing on subject-specific reading skills.

Reading strategy instruction. How students read and understand texts has been a vast area of educational research for decades. Since the 1980s, there has been an increasing emphasis on the idea that students' comprehension or knowledge of a subject can be fostered through both cognitive (i.e., how to read) and metacognitive reading strategy instruction (i.e., how to plan, monitor and evaluate your reading process). Strategies such as activating prior knowledge, identifying main ideas, and reflecting on the reading process are found to be effective for students' academic performance in general (Palincsar & Brown, 1984; National Reading Panel, 2000; Okkinga et al., 2018) as well as for the subject of history (McKeown, Beck, & Blake, 2009; Ramsay et al, 2010; Vaughn et al., 2013). A meta-analysis on whole-classroom reading strategy interventions showed the largest effect sizes for intervention studies conducted in grades 6–8 (Okkinga et al., 2018). However, in spite of the effectiveness of reading strategy instruction, researchers have recently adopted a more critical stance towards the instruction of cognitive and metacognitive reading strategies as a means unto itself, advocating instead for embedding reading strategy instruction in the curriculum of a specific discipline (McKeown et al., 2009; Moje, 2015). By doing so, it is assumed that reading texts will be more purposeful and effective.

Disciplinary literacy. The ability to apply relevant strategies when involved in reading texts for a specific school subject is commonly known as content-area literacy or disciplinary literacy. The concept of disciplinary literacy follows the assumption that students need to apply different reading strategies and heuristics for different text subjects (Goldman et al., 2016; Moje, 2008, 2015; Shanahan & Shanahan, 2008). For the subject of history, disciplinary literacy practices are often based on how historians (Wineburg, 1991, 1998) or expert readers (Shanahan, Shanahan, & Misischia, 2011) read and interpret historical texts. These texts often consist of primary or secondary source material, which enables students to apply expert reading practices such as sourcing (*where does this information come from?*), contextualising (*what were the characteristics of the time and society in which this was written?*), and corroborating (*is there similar or contradictory information available in other sources on this topic?*)—which are advanced skills that students need to develop throughout their academic career (Shanahan & Shanahan, 2008; Wineburg, 1991).

This focus on expert historians' reading processes led to the development of

several research programmes that focused on disciplinary literacy practices and their effectiveness for students' knowledge and comprehension of historical texts (Girard & McArthur Harris, 2012; Hynd, Holschuh, & Hubbard, 2004; Learned, 2018; Monte-Sano, 2011; Monte-Sano, De La Paz, & Felton, 2014; Nokes, Dole, & Hacker, 2007; Reisman, 2012; Wineburg & Reisman, 2015). For example, the text-based method called Questioning the Author, in which students had to identify and critically evaluate the author's background and stance, proved effective for students' comprehension and the self-monitoring thereof (McKeown et al., 2009). Although the disciplinary approach often has shown to be beneficial, some researchers dispute its value, emphasising the importance of generic strategy instruction for adolescents who struggle with reading (Faggella-Luby et al., 2012). Other research points to the possible barriers for students in (lower) secondary history classes, such as students' lack of background knowledge, lack of experience in using heuristics to reason critically about historical texts, or their susceptibility towards presentism—which occurs when people use their own contemporary frame of reference to explain events from the past (Duhaylongsod, Snow, Selman, & Donovan, 2015; Nokes, 2011; Perfetti, Britt, & Georgi, 1995; Wineburg, 2001).

Subject-specific reading skills. Irrespective of separate research foci in reading strategy instruction and disciplinary literacy practices, which might provide the impression that there is a sharp contrast between them (Learned, 2018), this is often not the case in daily educational practice. Unlike historians, lower secondary students do not focus on studying primary or secondary source material, but mostly read expository or informational texts in their textbooks.

There is a bidirectional relationship between the reading strategy instruction and disciplinary literacy. Mastering generic reading strategies, or knowing how to read, is an important prerequisite for the application of disciplinary literacy practices. Conversely, reading history texts provides students with the opportunity to develop their critical reading skills, such as determining an author's perspective. This implies that "getting good at history reading may significantly contribute to students' general ability to read critically" (Afflerbach & VanSledright, 2001, p. 697), a skill deemed important in the age of endless resources, information overload, and fake news. However, it is questionable whether an approach in which students simultaneously develop both generic reading comprehension skills and disciplinary literacy skills is fully attainable for 12- to 13-year-old students who recently transferred from primary to secondary education.

The current study operates in the grey area between generic reading comprehension strategies and disciplinary reading practices. It focuses on reading skills that are apparent in both reading strategy instruction and disciplinary literacy approaches, which we define as subject-specific reading skills. For example, recognising causal relations is a generic comprehension skill that enables students to identify how one sentence relates to another (e.g., by focusing on connectives), but for the specific subject of history it also enables students to reason causally about how certain historical events are related—an important skill in history education (Stoel, van Drie, & van Boxtel, 2015). The same accounts for skills such as finding explanations, generating questions, finding main ideas, and perspective-taking (e.g., contextualisation; Huijgen, van de Grift, van Boxtel, & Holthuis, 2018; van Boxtel & van Drie, 2018; Wineburg, 1991). These skills are also apparent in the practice known as historical reasoning.

Historical Content Knowledge and Historical Reasoning

Historical reasoning, often also referred to as historical thinking, encompasses a complex construct of the learning process students are involved in when confronted with historical texts or sources. The concept is closely related to disciplinary literacy, since expert historians often use heuristics to reason historically while reading texts (Nokes et al., 2007; Wineburg, 1991). According to Van Boxtel and Van Drie (2018; see also van Drie & van Boxtel, 2008), historical reasoning practices enhance students' knowledge and understanding of historical events, ideas, and developments. The core of their well-known historical reasoning framework consists of three elements that students can reason historically about: (I) continuity and change; (II) causes and consequences; and (III) similarities and differences. In addition, their framework defines six components of historical reasoning, which can be translated into concrete learning activities: (1) historical contextualisation; (2) using historical concepts; (3), using metahistorical concepts; (4) using historical sources; (5) providing (counter) arguments; and (6) asking historical questions. The first three components focus more on the construction of temporal-causal relations, whereas the latter three are more oriented towards argumentation and critical analysis.

Although this historical reasoning framework neither distinguishes difficulty levels of these components nor suggests a certain order in which they should be developed, some components are, in our view, more suitable for lower secondary education practice. Lee and Ashby (2000) investigated the development and changes

in students' ideas about the past, focusing on students between the ages of 7 and 14 years old. Their results show that the developmental progress of students' historical reasoning ability is rather complex. Even though older students were well able to reason historically about concepts such as 'evidence', which relates to the practice of argumentation and critical analysis in the historical reasoning framework of Van Boxtel and Van Drie (2018), there were major differences between and within different age groups. Lee and Ashby (2000) argue that the history curriculum should focus more on students' reasoning ability and thereby improve students' intellectual toolkit, but that this entails "a complex of multitrack understandings" (p. 216): Students need to learn how to make claims about the past, but also how to substantiate or overturn these claims.

While it is evident that historical reasoning and the reading of historical texts are closely related (van Boxtel & van Drie, 2018), the question remains how we can support students' historical reasoning ability through the reading of expository history texts in lower secondary education. Since students' (prior) knowledge of historical facts, concepts, and chronology is an important prerequisite for historical reasoning and learning in general, the components from the framework of Van Boxtel and Van Drie (2018) that focus on constructing temporal-causal relations (e.g., 'using historical concepts' and 'historical contextualisation') might offer a good starting point. In addition, 'asking historical questions' seems to be a component that some young adolescent students should be capable of, whereas 'providing (counter) arguments', 'using historical sources' and 'using metahistorical concepts' appear to be more advanced practices.

Using Digital Technology to Analyse Students' Engagement

Technology-enhanced learning environments are increasingly used to support students' reading and learning processes, including research focusing on history education (for a literature review, see O'Neill & Weiler, 2006 [on cognitive tools] and Poitras, Lajoie, & Hong, 2012 [on metacognitive tools]). These environments provide researchers with possibilities to mine and translate data to detect, analyse, and foster students' learning processes—a process commonly known as learning analytics (Azevedo & Gašević, 2019). The current study integrated the aforementioned subject-specific reading skills in a Digital Learning Environment (DLE) that was used by lower secondary students to read expository history texts. We subsequently used its log file data to analyse students' reading performance and engagement.

Engagement, for example in reading texts, is a difficult concept to grasp (Azevedo & Gašević, 2019). Fredricks, Blumenfeld, and Paris (2004) distinguished three types of engagement: behavioural engagement, which includes, for example, the time students spend on reading tasks; cognitive engagement, which relates to the quality of the reading process; and emotional engagement, which includes students' personal reactions to their reading task or classroom environment. Greene, Bolick, and Robertson (2010) showed that in hypermedia learning environments used to read historical texts students often found it difficult to engage in learning processes that could foster their historical knowledge.

In a previous study, based on the same dataset as the current study, we were able to identify different behavioural and cognitive engagement profiles with the application of learning analytics. Log file data about students' reading process provided several predictor variables, such as supportive hint use and time on task, which we subsequently used in latent profile analysis. We identified five engagement profiles, ranging from overall low engagement to moderate or high engagement, which differed significantly in terms of students' intrinsic motivation and text comprehension (ter Beek, Opdenakker, Deunk, & Strijbos, 2019a; see Chapter 3). These profiles will enable us to explore how engagement in reading texts affects students' historical content knowledge and historical reasoning ability in the current study.

The Current Study

Although there is an ongoing debate about what history education should entail, and whether teachers' instruction should focus more on substantive historical knowledge or on the ability to reason historically, it is important to note that these two concepts, in practice, cannot completely be separated from each other (Gestsdóttir, van Boxtel, & van Drie, 2018; Lee & Ashby, 2000). Students will need a sufficient prior knowledge base, including historical content knowledge, to be able to apply higher-order thinking skills such as reasoning historically about causes and effects (Kirschner, Sweller, & Clark, 2006; van Boxtel & van Drie, 2018). However, "in most conceptualizations of historical reasoning … the role of first-order knowledge is barely explicated" (van Boxtel & van Drie, 2018, p. 156). Therefore, the current study focuses on both students' historical content knowledge and historical reasoning ability, and investigates their relations with students' subject-specific reading skills.

Research aims and questions. The current study focuses on stimulating the

combined application of generic reading strategies as well as disciplinary literacy practices—which we refer to as subject-specific reading skills—in the context of history text reading in lower secondary education. More specifically, this study explores the contribution of subject-specific reading skills and students' engagement in the process of expository text reading to their historical content knowledge as well as their historical reasoning ability. We will address the following two research questions:

1. What is the relation between students' subject-specific reading skills and their historical content knowledge and historical reasoning ability?
2. Are there differences in historical content knowledge and historical reasoning ability between student profiles based on engagement in a digital learning environment?

We expect that the subject-specific reading skills will be positively related to students' historical content knowledge and historical reasoning ability. With regard to the engagement profiles, we expect that students with relatively higher engagement levels will perform better at the historical content knowledge and historical reasoning ability tests compared to students who, for example, spent less time on the reading tasks, thereby showing lower levels of engagement.

Method

Design and Context

In this study, students used a DLE called 'Gazelle'[1], which contained six expository texts for the subject of history (ter Beek, Spijkerboer, Brummer, & Opdenakker, 2018). The main theme of all texts was 'The time of Greeks and Romans' and each text contained approximately 550 words. We carefully analysed the contents of students' regular textbooks to prevent overlap or duplicate information in all texts. The lessons in which students read texts in the DLE replaced six regular lessons in the seventh-grade curriculum.

Procedure. The six-week intervention lasted from October until December, shortly after students entered secondary education. During the six lessons, which

1 Gazelle is a Dutch acronym for 'Gemotiveerd en Actief Zelfstandig Lezen', which roughly translates into 'Motivated and Active Independent Reading'.

each lasted approximately 50 minutes, students read an expository text about the ancient Greeks and completed several text-based assignments, including ten multiple-choice questions (see Figure 5.1). Students were able to correct their incorrect answers with a maximum of two attempts per question. The DLE offered support in the form of cognitive, metacognitive, and motivational hints, which students could deliberately access when needed. These hints provided students with both generic reading strategies and disciplinary literacy instruction. For example, it showed the reader that "causal relations can be found after the appearance of words such as *because* or *therefore*", but it also explained that "While you read, it is important to consider the society and time that the Spartans lived in because it was very different from our current society". The DLE recorded log file data of students' behavioural and cognitive engagement for each lesson in the intervention. A delayed Historical Content Knowledge test (HICK) was administered four to six weeks after completion of the last lesson in the DLE.

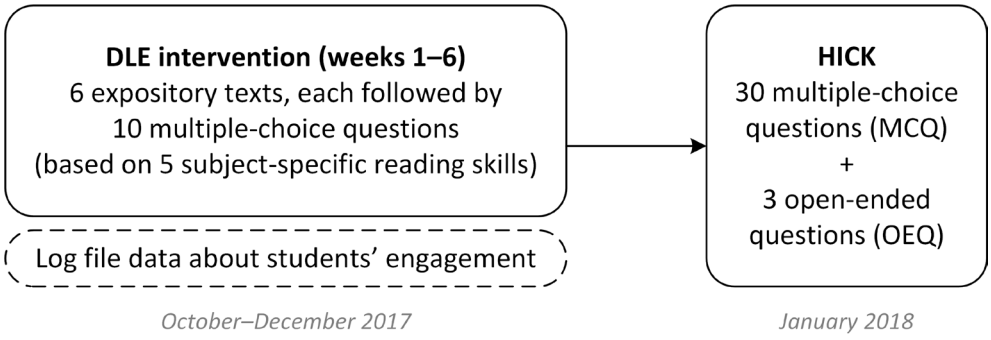


Figure 5.1 Timeline for the study and data collection. DLE = Digital Learning Environment; HICK = Historical Content Knowledge test.

Participants

Nine Dutch history teachers with 13 seventh-grade classrooms originally carried out the six-week research intervention. However, only five history teachers—who, combined, taught eight different classrooms—were willing to administer the additional HICK test after completing the lessons in the DLE. This was mainly due to contextual barriers such as teachers' workload, time constraints, and the availability of IT facilities (ter Beek, Opdenakker, Deunk, & Strijbos, 2019b; see Chapter 4). Therefore, a subset of participants with only students from these eight classrooms is included in the current study.

Parents or caretakers of all participating students were informed via a personal letter and were given the option to refuse the use of their child's data. This was the case for two students, whose data we deleted from all datasets. Therefore, at the start of the intervention, the total sample consisted of 197 students, of which 48.7% was female (n = 96) and 51.3% was male (n = 101); their average age was 12.5 years (SD = 0.44).

Measures

Subject-specific reading skills. Each of the six texts was accompanied by 10 multiple-choice questions. These questions were divided into five categories of subject-specific reading skills. Based on the official requirements for the state national history exam (College for Exams, 2014), we selected five skills that are of importance in the general domain of reading comprehension as well as for a disciplinary approach towards reading history texts: (1) recognising causal relations; (2) explaining historical events; (3) generating suitable research questions; (4) ordering of concepts; and (5) perspective-taking. Hence, students received two multiple-choice questions per skills category per week. Table 5.1 shows the five skills and exemplary multiple-choice questions. A closer look at the description of the skills defined in each question category reveals that the categories and the components from the historical reasoning framework by Van Drie and Van Boxtel (2008) are somewhat intertwined. For example, the included subject-specific reading skill defined as 'recognising causal relations' is also apparent at the core of this framework, which includes 'historical reasoning about causes and consequences' as one of its core elements. The log files from the DLE enabled us to compute students' average score on each of the five categories by using students' first attempt of answering the multiple-choice questions (0 = incorrect, 1 = correct) from weeks 1–6.

Engagement profiles. The profiles of students' behavioural and cognitive engagement while reading texts in the DLE have been identified in a previous study using latent profile analysis (LPA) based on log file data about students' time on task, cognitive hint use, metacognitive and motivational hint use, correct multiple-choice questions at first attempt, and judgment of learning accuracy (ter Beek et al., 2019a; see Chapter 3). A detailed description of these predictor variables as well as the LPA procedure and the quality of the profile solutions can be found in Chapter 3 of this dissertation.

Table 5.1 Subject-specific reading skills and examples of multiple-choice questions in the DLE

Category	Skill	Example
Cause and effect	Recognising (direct and indirect) causal relations	*"What can be considered a direct cause of the demise of the powerful city state of Sparta?"*
Explaining	Explaining historical events or developments	*"Explain why the 300 Spartan soldiers went into battle against 10,000 Persians."*
Generating questions	Generating or selecting suitable research questions	*"Imagine you are researching the status of women in ancient Greece. For which of the following questions can you find an answer in the current text?"*
Ordering of concepts	Identifying chronology or important text elements	*"Look at the following four elements from the text. Which of these are main ideas?"*
Perspective-taking	Contextualisation of concepts described in texts, or actors' point of view	*"What could have been a reason for the Spartans to leave sickly babies in the mountains to die?"*

Note. DLE = digital learning environment.

We identified five engagement profiles, which we labelled as 'types of readers' to distinguish the differences in students' engagement that these profiles represent. The *naïve readers* scored relatively low on all indicators of engagement. Moreover, these students had low performance, but did not appear to be (fully) aware of this. The *stubborn readers* also had relatively low scores on engagement indicators, but these students were more aware of their low performance. The profile of *help-seeking readers* consists of students who used significantly more supportive hints while reading. *Independent readers* scored relatively high on all indicators of engagement, except for their supportive hint use, indicating that these students were able to perform well at their first attempt of answering the multiple-choice questions without accessing the cognitive, metacognitive, and motivational hints. Finally, the *uncertain readers* had relatively high scores on almost all engagement indicators, especially time on task and hint use. However, they often misjudged the correctness of their answers. We used students' predicted membership for one of the five engagement profiles as an independent variable in the current study with regard to the second research question.

Historical content knowledge and reasoning ability. The first author[2] composed an instrument to measure students' historical content knowledge (HICK) based on the Assessment of Social Studies Knowledge (ASK) instrument by Vaughn et al. (2013). The original ASK instrument consists of two subtests: a content knowledge test comprised of 46 multiple-choice items, and a reading comprehension test with 21 multiple-choice items based on three text passages. The ASK instrument was modified so that its contents were related to the historical content covered in the current study, which was ancient Greece.

The final HICK instrument consisted of two components. The first component consisted of 30 multiple-choice questions (MCQ) about the literal contents of the six texts. It measured what students learned or remembered from reading these texts and answering the subject-specific assignments; therefore, it can be regarded as a measure of students' *historical content knowledge*. The second component included a short expository text about the Olympic Games followed by three open-ended questions (OEQ). Students were asked to identify differences between the Olympic Games in ancient Greece and the present day Olympic Games, explain why wars were paused during the ancient Olympic Games, and connect the Spartan and Athenian views on women in society to the fact that women were not allowed to compete in this event (i.e., contextualisation). By doing so, the OEQ component of the HICK incorporated elements from the historical reasoning framework by Van Boxtel and Van Drie (2018) and, therefore, it can be regarded as a measure of students' *historical reasoning ability*.

Appendix C contains sample questions of the MCQ and OEQ components of the HICK instrument. Prior to administration, three pre-service history teachers checked the instrument and found no major issues with regard to its contents. We administered the HICK approximately four to six weeks after completion of the last lesson. Subsequently, we analysed the internal consistency and reliability of the 30 MCQ items using the Kuder-Richardson Formula 20 (Kuder & Richardson, 1937). The yielded KR-20 value of 0.73 indicated reasonable reliability. Additionally, the first author coded the open questions following a predetermined answer model with a maximum of four points for Q1 and Q3 each, and two points for Q2, adding up to a maximum score of 10. To ensure interrater reliability, a research assistant also coded students' answers on all three open questions of one classroom. Cohen's Kappa was 0.71, indicating sufficient agreement (Cohen, 1960).

2 The first author holds a Master's degree in history education; through practical experience, she could properly assess which multiple-choice questions were suitable for the students participating in the current study.

Analyses

We analysed the data from the DLE using IBM SPSS Statistics 25. We used descriptive statistics, bivariate correlations, and multiple regression analysis with forced entry to explore the relations between subject-specific reading skills, historical content knowledge, and historical reasoning ability (RQ1). We used variance analysis with General Linear Models (GLM) and post hoc comparisons using Bonferroni adjustment to analyse the differences between the five engagement profiles (RQ2). Effect sizes are reported using partial eta squared, or partial η^2. We consider effect sizes as small when partial $\eta^2 < 0.06$, medium when $0.06 <$ partial $\eta^2 < 0.14$, and large when partial $\eta^2 > 0.14$ (cf. Cohen, 1988).

Attrition and missing data. Eight students did not complete all six lessons in the DLE, mostly due to sickness or classroom transfers. For these students, the average score on subject-specific reading skills as well as their engagement profile could not be based on all six lessons, which could distort the analyses. Therefore, we excluded these eight students (4.1% of the total sample) from all analyses, resulting in a sample of $N = 189$ students. Since the number of excluded students did not exceed 5% of the total sample, we found it acceptable to apply listwise deletion (Graham, 2009).

Due to absence, a further eight students missed the administration of the HICK questionnaire and an additional two students only completed the MCQ component of the HICK questionnaire, but not the OEQ component. These ten students did not significantly differ from the included students in terms of their scores on the five subject-specific reading skills, $p > .05$ for each skill, which substantiated our decision to exclude them in the analyses. Therefore, the final sample for the MCQ component consists of 181 students (91.8% of the total sample), whereas the final sample for the OEQ component consists of 179 students (90.9% of the total sample).

Results

Subject-Specific Reading Skills and Historical Content Knowledge and Historical Reasoning Ability

Table 5.2 shows the descriptive statistics and the correlations between the five subject-specific reading skills, and the MCQ and OEQ components of the HICK. For all subject-specific reading skills, with the exception of perspective-taking,

students on average answered half of the multiple-choice questions correctly at the first attempt. The average score on the MCQ component was 16.31 out of 30 and 4.64 out of 10 for the OEQ component. There were positive and moderate significant correlations ($p < .01$) between all five subject-specific reading skills as well as between each of these skills and students' performance on both MCQ and OEQ components of the HICK. It should be noted that the correlations with the MCQ component were somewhat stronger compared to the OEQ component.

Table 5.2 Descriptive statistics and bivariate correlations ($N = 181$)

Variable	1	2	3	4	5	6	7
1. Cause and effect	-						
2. Explaining	.41**	-					
3. Generating questions	.38**	.34**	-				
4. Ordering of concepts	.32**	.22**	.26**	-			
5. Perspective-taking	.36**	.43**	.37**	.28**	-		
6. HICK-MCQ	.38**	.37**	.41**	.31**	.34**	-	
7. HICK-OEQ	.28**	.35**	.32**	.19**	.30**	.45**	-
M	0.48	0.53	0.49	0.48	0.57	16.31	4.64
SD	0.15	0.16	0.16	0.17	0.18	4.76	2.09

Note. HICK = historical content knowledge; MCQ = multiple-choice questions; OEQ = open-ended questions. $N = 179$ for HICK-OEQ. *$p < .05$, two-tailed. **$p < .01$, two-tailed.

To test whether the subject-specific reading skills each add something unique to explain students' historical content knowledge and historical reasoning ability, we used multiple regression analysis with forced entry of the five predictive skills. Table 5.3 shows the results from these analyses and the predictors of the MCQ and OEQ components. For the MCQ component, the skills we defined as 'explaining historical events', 'generating historical questions', and 'ordering of concepts' were all significant unique predictors, $p = .036$, $p = .002$, and $p = .050$, respectively. For the OEQ component, only 'explaining' and 'generating questions' were significant unique predictors, $p = .012$ and $p = .041$. Although there were positive correlations with the MCQ and OEQ components, the multiple regression analysis showed that the subject-specific reading skills we defined as 'identifying

cause and effect' and 'perspective-taking' were neither uniquely predictive for students' historical content knowledge nor historical reasoning ability.

Table 5.3 Unique predictors of HICK-MCQ and HICK-OEQ components (N = 181)

	HICK-MCQ		HICK-OEQ	
Variable	*B*	95% CI	*B*	95% CI
Constant	4.87**	[2.06, 7.68]	0.62	[-0.69, 1.94]
Cause and effect	4.55	[-0.50, 9.14]	1.04	[-1.11, 3.20]
Explaining	4.67*	[0.32, 9.00]	2.63*	[0.59, 4.68]
Generating questions	6.87**	[2.49, 11.24]	2.14*	[0.09, 4.19]
Ordering of concepts	3.80*	[0.01, 7.58]	0.62	[-1.18, 2.42]
Perspective-taking	2.71	[-1.29, 6.70]	1.33	[-0.54, 3.20]
R^2	.28		.19	
F	13.86***		8.06***	

Note. HICK = historical content knowledge; MCQ = multiple-choice questions; OEQ = open-ended questions; CI = confidence interval. N = 179 for HICK-OEQ. *p < .05. **p < .01. ***p < .001.

Engagement Profiles and Historical Content Knowledge and Historical Reasoning Ability

Table 5.4 shows the proportion of students and the HICK scores of the five engagement profiles. The majority of students was assigned to either the 'naïve readers' profile (n = 63) or the 'stubborn readers' profile (n = 41). On average, independent readers scored highest on the MCQ component, whereas naïve readers scored the lowest. For the OEQ component, help-seeking readers had the highest average score, whereas naïve readers again scored the lowest.

Variance analysis using GLM yielded significant differences between the profiles' MCQ scores, $F(4, 176) = 6.15$, $p < .001$, partial $\eta^2 = .12$. Post hoc comparisons using Bonferroni adjustment showed that naïve and stubborn readers' MCQ scores differed significantly from those of both help-seeking and independent readers, in favour of the latter two. Similarly, there were significant differences between the profiles' OEQ scores, $F(4, 174) = 5.72$, $p < .001$, partial $\eta^2 = .12$. Post hoc comparisons showed that help-seeking readers performed significantly better on the OEQ component than the

Table 5.4 Proportion of students in and HICK scores of the engagement profiles

Characteristic	Total sample	Naïve readers	Stubborn readers	Help-seeking readers	Independent readers	Uncertain readers
% of students (number)	100 (181)	34.8 (63)	22.6 (41)	21.6 (39)	14.9 (27)	6.1 (11)
M HICK-MCQ (*SD*)	16.31 (4.76)	$14.83\ (4.30)_{a}$	$15.00\ (5.62)_{a}$	$18.28\ (4.10)_{b}$	$18.44\ (3.78)_{b}$	$17.55\ (4.03)_{a,b}$
M HICK-OEQ (*SD*)	4.64 (2.09)	$3.81\ (1.91)_{a}$	$4.56\ (1.98)_{a,b}$	$5.61\ (2.27)_{b}$	$5.04\ (1.71)_{a,b}$	$5.45\ (2.07)_{a,b}$

Note. HICK = Historical Content Knowledge test; MCQ = multiple-choice questions; OEQ = open-ended questions. The maximum HICK score was 30 for the MCQ component and 10 for the OEQ component. Within-row means with different subscripts differ significantly at $p < .05$.

naïve readers. Since we identified engagement profiles based on log file data such as time on task and hint use, this finding indicates that students' behavioural and cognitive engagement when reading expository texts in a DLE is positively related to both their historical content knowledge and their historical reasoning ability. This is especially the case for students who used relatively more supportive hints and were thus assigned to the 'help-seeking readers' profile.

Although the average HICK scores of the profiles differed significantly, it is important to keep in mind that this does not imply that individual students within these profiles are uniform in terms of historical content knowledge or historical reasoning ability. A student in the 'naïve readers' profile might score high on content knowledge but low on engagement, simply because he or she did not need to consult supportive hints or spent a lot of time on task. Conversely, a student in the 'independent readers' profile might have consulted many hints, indicating high engagement, but he or she may fail to answer questions on a delayed test correctly due to the lack of this support. Scatter plots of the HICK-MCQ and HICK-OEQ scores per engagement profile show that each profile includes students performing well above or below the average of the total sample (see Figures 5.2 and 5.3, respectively).

Discussion

Research has shown that both general reading skills and disciplinary literacy skills can contribute to students' understanding of text and content knowledge (Learned, 2018; Nokes et al., 2007). We refer to a combination of these two concepts as 'subject-specific reading skills'. The cognitive approach towards historical reasoning emphasises the role of mental resources, one of which is students' reading ability (van Boxtel & van Drie, 2018). The current study explored which subject-specific reading skills contribute to students' historical content knowledge and historical reasoning ability. Seventh-grade students read historical texts in a DLE, which provided us with log file data about their behavioural and cognitive engagement that allowed us to identify engagement profiles (ter Beek et al., 2019a; see Chapter 3). This approach allowed us to address whether the differences between these profiles relate to students' historical content knowledge and historical reasoning ability.

Summary of Findings

Bivariate correlations showed positive and moderate significant correlations between

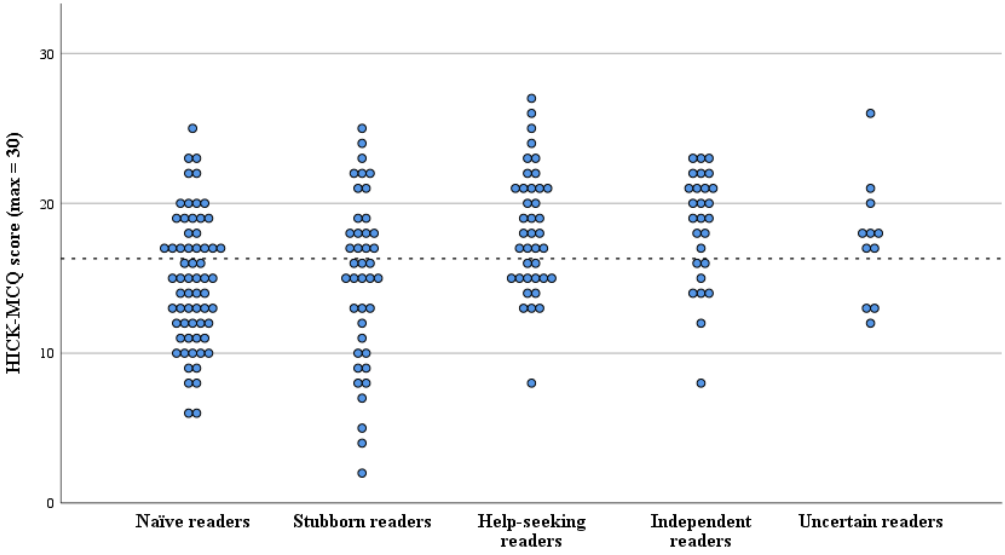


Figure 5.2. Scatter plot of students' HICK-MCQ scores per engagement profile. Each dot represents a single student (*N* = 181); the dotted line represents the total sample average.

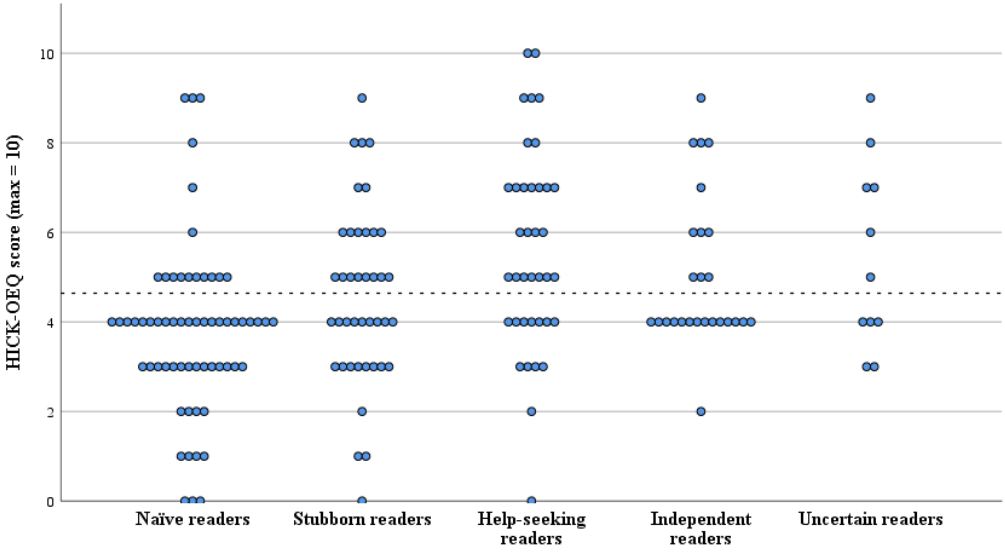


Figure 5.3 Scatter plot of students' HICK-OEQ scores per engagement profile. Each dot represents a single student (*N* = 179); the dotted line represents the total sample average.

all five subject-specific reading skills (i.e., cause and effect, explaining, generating questions, ordering of concepts, and perspective-taking). Although it could be expected that these skills would correlate, since they are all part of the general construct of reading comprehension, the individual subject-specific skills seemed to measure separate skills within the process of reading historical texts. Additionally, all five subject-specific reading skills correlated positively with students' historical content knowledge (i.e., the HICK-MCQ component) and historical reasoning ability (i.e., the HICK-OEQ component). These findings suggest that students' subject-specific reading skills contribute to students' historical content knowledge and historical reasoning ability, although the correlations with the latter were less strong.

Multiple regression analyses showed that the subject-specific reading skills 'explaining', 'generating questions', and 'ordering of concepts' were significant unique predictors of students' historical content knowledge, whereas only the first two skills were significant unique predictors of students' historical reasoning ability. Earlier findings indicate the importance of causal reasoning (Stoel et al., 2015) and contextualisation (Huijgen et al., 2018; van Boxtel & van Drie, 2018); however, the skills we defined as 'identifying cause and effect' and 'perspective-taking' did not significantly and uniquely predict historical content knowledge nor historical reasoning ability. A possible explanation might be that when all five subject-specific reading skills are combined in a reading task, the multiple-choice questions addressing 'explaining', 'generating questions', and 'ordering of concepts' require more higher-order thinking skills compared to, for example, recognising cause-and-effect relations in a text, and, thus, are more closely related to historical content knowledge and historical reasoning ability.

On average, students answered half of the multiple-choice questions related to the five subject-specific skills correctly. Similarly, students answered on average half of the items on the delayed historical content knowledge test correctly. One might consider these results poor, since in general in the Dutch educational system, a score of about 50% would result in an insufficient grade. However, given the fact that the HICK instrument was administered unannounced and four to six weeks after the last lesson in the DLE, one could also argue that these results are rather positive. Since there are no other studies that used the same HICK instrument to measure delayed historical content knowledge and historical reasoning ability, we cannot compare our results with earlier studies, and we do not know whether students improved their historical content knowledge or historical reasoning ability. Future studies might, for

example, adopt a quasi-experimental design that includes a pretest measurement of students' historical content knowledge and historical reasoning ability to analyse the effectiveness of students' subject-specific reading skills in more detail.

Profiles of students who scored high on several predictor variables related to engagement, such as time on task or supportive hint use, scored significantly higher on a delayed test of historical content knowledge and historical reasoning ability, indicating that engagement is positively related to students' learning process. However, the majority of students was still assigned to either the 'naïve readers' or the 'stubborn readers' profile, indicating that these students scored on average low on most engagement indicators, and, thus, did not perform well on both the historical content knowledge test and the historical reasoning components.

Limitations and Suggestions for Future Research

Due to the practice-oriented approach of this study, in which teachers used the DLE in authentic classroom settings, there are some limitations in relation to the statistical analyses. Although we carefully selected the contents of the texts and assignments in the DLE, we could not control for the possible influence of teachers' regular instruction on students' historical content knowledge. Therefore, it may be possible that students' performance on the historical content knowledge test (HICK) was not only related to the practising of their subject-specific reading skills in the DLE but also to the regular classroom instruction they received. Future research could control for this in statistical analyses by including quantified observations of teachers' instructional practices.

The HICK instrument used to measure both historical content knowledge and historical reasoning ability caused another limitation to the study at hand. It was specifically designed for the current study, but research has shown that constructing a valid and reliable instrument to measure historical reasoning processes is a highly challenging task (cf. Huijgen et al., 2018). Although the approach used in the current study provided us with relevant results, additional and robust tests, factor analyses, and larger samples are needed to further validate the instrument in order to more robustly determine the influence of subject-specific reading skills on students' historical content knowledge and historical reasoning ability. Moreover, even though the HICK was essentially based on the ASK instrument by Vaughn et al. (2013), the OEQ component aimed at measuring historical reasoning, whereas the original open-ended questions in the ASK instrument focused more on text comprehension.

The correlations between the five subject-specific reading skills and the OEQ were somewhat less strong compared to the correlations with the MCQ component, which is probably influenced by the fact that the MCQ component resembled the multiple-choice questions in the DLE. Moreover, the format of the OEQ also appealed to students' writing skills. Earlier studies have shown that students' initial writing ability is related to the quality of their written text, which might have led to better answers on the OEQ items, and, subsequently, higher scores (cf. De La Paz & Felton, 2014; van Drie, Braaksma, & van Boxtel, 2015). Future research could consider using sophisticated measures of students' historical reasoning skills that are not based solely on written answers, for example think-aloud protocols, to uncover students' reasoning process while or after reading historical texts.

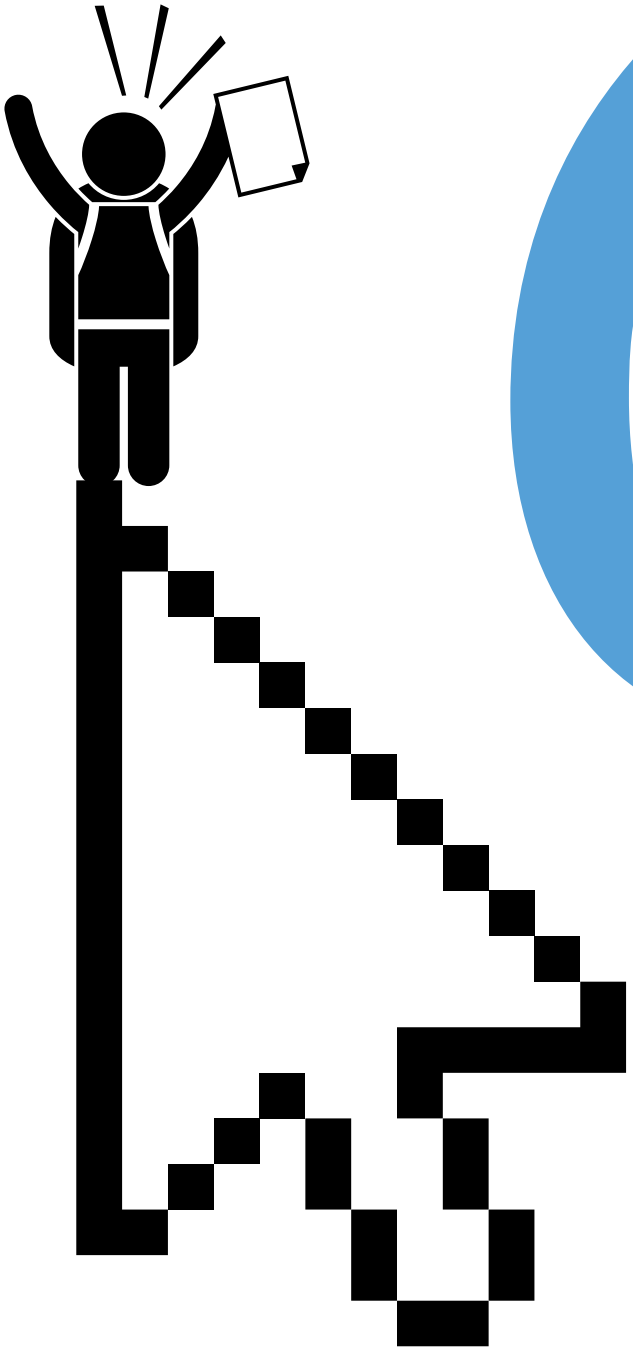
Scientific and Practical Implications

This study explored the grey area between generic reading strategy instruction and disciplinary literacy skills, which we defined as subject-specific reading skills. It shows the applicability of the historical reasoning framework by Van Boxtel and Van Drie (2018) within the context of reading historical texts in lower secondary education. When core components from this framework are translated into concrete reading skills, for example when students digitally practice with reading texts, it is possible to investigate which components are attainable for students of a specific age group. Our results show that the skills we defined as 'explaining historical events' and 'generating historical questions' contributed to both the historical content knowledge and the historical reasoning ability of seventh-grade students.

Moreover, the person-centred approach of identifying latent profiles from log file data showed that students' behavioural and cognitive engagement also seems to be an important factor for their (development of) historical content knowledge and historical reasoning ability. These findings indicate that the history curriculum in lower secondary education should stimulate both students' reading skills and engagement to promote students' historical content knowledge and historical reasoning ability throughout their academic career. For example, teachers could use expository texts as a basis for their instruction, followed by classroom discussions (cf. Wanzek, Swanson, Roberts, Vaughn, & Kent, 2015) about possible explanations for historical events, relevant historical questions to ask about a text, or different perspectives encountered in these texts (McKeown et al., 2009).

Furthermore, the differences between the five profiles provide teachers with

information about the general characteristics of students within these engagement profiles. For example, the fact that the majority of students in our sample belongs to the naïve and stubborn readers calls for attention towards students' engagement while reading texts in a DLE. However, these profiles do not offer a comprehensive image of individual students' learning processes. Teachers will need to determine both students' subject-specific reading skills as well as their engagement to be able to adapt their instruction to suit individual students' needs. Nevertheless, the DLE used in the current study may provide teachers with a useful first step towards informed practice with regard to fostering historical content knowledge and historical reasoning ability.



6

Chapter 6

General conclusion, discussion, and recommendations for future research and practice

Introduction

Reading comprehension is an essential skill for processing textual information and acquiring knowledge, especially for the subject of history given the abundance of texts students are required to read in their textbooks. The expository format of these texts is often challenging for many students in lower secondary education. As a result, there have been many studies on how to support students' reading comprehension, and over the last decade, an increasing number of studies used computer-supported or digital learning environments to achieve this goal. Previous research on cognitive and metacognitive support in digital learning environments has shown positive effects on students' reading performance in secondary education (Cheung & Slavin, 2012; Lan, Lo, & Hsu, 2014; Moran, Ferdig, Pearson, Wardrop, & Blomeyer, 2008). However, how instructional support in these kinds of learning environments adds to the students' learning process often remains unclear, since most studies only focus on performance or self-regulation as an outcome measure (Devolder, van Braak, & Tondeur, 2012; ter Beek, Brummer, Donker, & Opdenakker, 2018). Therefore, it is necessary to unravel how digital learning environments support or contribute to students' reading process and outcomes.

This dissertation focuses on a Digital Learning Environment (DLE) called "Gazelle", which was specifically designed for reading expository history texts in lower secondary education in the Netherlands (ter Beek, Spijkerboer, Brummer, & Opdenakker, 2018). The DLE aimed at stimulating students' reading comprehension, self-regulated learning, and motivation using cognitive, metacognitive, and motivational scaffolds called 'hints'. Hints were incorporated in the DLE, where students could access them when needed. In addition, the DLE provided teachers with visualised data output about students' reading performance. The DLE was used between 2016 and 2018 in various lower secondary classrooms in the Netherlands for the subjects of history and geography. The current dissertation focuses on the subject of history education.

The main aim of this dissertation was to analyse the practical implementation of the DLE, and its effects on students' (1) text comprehension, (2) self-regulation, (3) motivation, (4) engagement, (5) historical content knowledge, and (6) historical reasoning ability, as well as on teachers' use of data and their instructional practice in this field. By focusing on all these interrelated aspects, the results at hand provide a comprehensive overview of the use and

usefulness of a DLE to support students' reading comprehension in history education. Two main research questions were explored simultaneously:

- How do seventh-grade students and their history teachers use and experience a DLE enriched with strategic hints and visualised student data (i.e., how do they use it)?
- What are the effects of using the DLE on students' reading process and learning outcomes, and on history teachers' instructional practice (i.e., how useful is it)?

Figure 1.2 (see Chapter 1, p. 26) provides an overview of the concepts incorporated in this dissertation, as well as the studies and chapters in which they are included.

Summary of Main Findings

The first study (Chapter 2) focused on the provision of cognitive and metacognitive instructional scaffolding—in the form of hints—and its effects on students' text comprehension, self-regulated learning (SRL), reading strategy awareness, and motivation. During six weeks, 174 seventh-grade students from three different schools read expository texts in the DLE. Three different versions of support were offered. From weeks 2 to 5, students from Experimental group A could decide to access cognitive and metacognitive hints containing reading strategy instruction for the subject of history, whereas students from Experimental group B could decide to access similar hints for the subject of geography, but not for history. Students from the control group had no access to hints in either subject. Results showed that solely providing these hints did not lead to significant differences between the groups; nevertheless, students in both experimental groups significantly increased their levels of problem-solving strategy awareness. Although not every student in Experimental group A accessed at least one hint (contrary to what was expected), in-depth analyses comparing students who used one or more hints vs. students who did not use hints showed significantly better posttest text comprehension for the hint users. There were no significant differences with regard to students' SRL, awareness of reading strategies, and motivation (in terms of task value and self-efficacy). In general, posttest results showed no significant improvement in students' reading performance after six weeks; in fact, the average student performance decreased in all groups. Comparative analyses of all students, based on different initial reading levels (i.e., below-average, average, and above-average readers) showed that the performance of average and above-average readers significantly declined, but that

this was not the case for below-average readers. Furthermore, students' motivation significantly declined for the below-average readers in terms of self-efficacy beliefs, and for above-average readers in terms of task value.

In the second study (Chapter 3), behavioural and cognitive engagement profiles were distinguished using log file data from the DLE in the second year of the overarching research project. In contrast to the variable-centred approach of the first study, this study adopted a person-centred approach. By doing so, it was possible to explore the relation between students' individual activity in the DLE and their reading comprehension performance. In total, 327 seventh-grade students from four secondary schools worked in the DLE for six weeks. Five engagement profiles were identified based on predictor variables, which included students' time on task, hint use, average score at first try, and judgment of learning (JOL) accuracy. Latent profile analyses (LPA) showed that a five-profile solution had the best and most meaningful fit to the data. The first and largest profile, which was labelled as 'naïve readers', included students with relatively low scores on all predictor variables. The second profile, the 'stubborn readers', highly resembled the 'naïve readers' profile, with the exception of JOL accuracy; students from this profile seemed to be well aware of their low performance in the DLE but did not access hints to improve their performance. The third profile, the 'help-seeking readers', showed average scores on the predictor variables, but used significantly more hints than the first two profiles. The fourth profile, the 'independent readers', scored relatively high on all predictor variables, except for hint use, indicating that students from this profile obtained high performance scores even without accessing additional support in the form of hints. Lastly, the fifth and smallest profile, the 'uncertain readers', included students that scored high on all predictor variables, except for JOL accuracy, indicating that these students often underestimated themselves. Subsequently, the relations between these engagement profiles and students' motivation (i.e., students' task value, self-efficacy, and intrinsic motivation) and text comprehension performance were investigated. Results showed that highly engaged students initially had significantly higher task value and intrinsic motivation compared to students who showed little engagement. In addition, highly engaged students showed better text comprehension.

Although the first two studies showed the importance of analysing students' reading processes and outcomes when using a scaffolded DLE to read expository history texts, it is equally important to consider the role of teachers in this process. Earlier studies have shown that reading strategy instruction in history education

does not occur often (Linthorst & de Glopper, 2015; Ness, 2006). This might be explained by the fact that history teachers do not feel capable of providing this type of instruction, or that they lack the knowledge and motivation to do so (Greenleaf, Schoenbach, Cziko, & Mueller, 2001; Hall, 2005). Professional development (PD) training enables teachers to practice these skills, thereby improving their self-efficacy beliefs, knowledge, and motivation (Desimone, 2009). The third study of this dissertation (Chapter 4) focused on the participating history teachers and their beliefs, attitudes, knowledge, and instructional practices with regard to reading strategies. The study consisted of two phases. In the first phase, history teachers in Experimental condition A were provided with an extended visualisation of student performance data in the DLE and were observed during several lessons. In the second phase, these teachers received PD training and a guiding manual on how to translate these student performance data into structured, explicit reading strategy instruction. Teachers in Experimental condition B were provided with extended data visualisations in Phase 2, but received no PD training. Teachers in the control condition were only provided with basic visualisations of student performance data in both phases. The results showed that teachers in both the experimental conditions and the control condition reported high levels of perceived strategy instruction knowledge after each phase. Likewise, regardless of condition, teachers reported positive attitudes towards reading strategy instruction; however, their self-efficacy beliefs regarding this instruction were slightly less positive compared to their perceived knowledge. Teachers mainly explained strategies such as orienting on a text or adjusting reading strategies when encountering problems, whereas strategies with regard to reflection and evaluation occurred rarely during their instruction. Moreover, modelling behaviour was not observed often. Comparisons of the observational data in both phases showed that teachers in the experimental conditions used a wider range of strategies during their classroom instruction and used modelling behaviour more often after the PD training. Although these results seem promising, it is important to note that the overall variation in reading strategy instruction was relatively low in all conditions. To gain more insight in the quantitative data with regard to teachers' instruction, this mixed-method study was complemented with micro-level content analysis of qualitative interview and focus group data about teachers' experience with the DLE. All teachers mentioned contextual barriers that, in their opinion, hampered their use of the DLE, such as limited preparation time, high workload, unavailability of IT resources, and other pressing school matters.

The final study of this dissertation (Chapter 5) investigated the relations between subject-specific reading skills—defined as (1) recognising causal relations; (2) explaining historical events; (3) generating suitable research questions; (4) ordering of concepts; and (5) perspective-taking—and students' historical content knowledge and historical reasoning ability, to explore which specific reading skills are of importance. Furthermore, the identified profiles from Chapter 3 were used to explore the relations between students' behavioural and cognitive engagement in the DLE and their historical content knowledge and historical reasoning ability. The majority of the participants from the second year of the overarching research project completed an additional Historical Content Knowledge (HICK) test approximately four to six weeks after completion of the last lesson in the DLE. The HICK instrument consisted of two parts: the multiple-choice questions (MCQ) measured students' historical content knowledge, whereas the open-ended questions (OEQ) measured students' historical reasoning ability. Results showed that all subject-specific reading skills correlated significantly with both historical content knowledge and historical reasoning ability. However, multiple regression analysis showed that not all skills were unique significant predictors. The skills that were defined as 'explaining historical events', 'generating historical questions', and 'ordering of concepts' were all significant unique predictors for historical content knowledge. For historical reasoning ability, only 'explaining' and 'generating questions' were significant unique predictors. The subject-specific reading skills we defined as 'identifying cause and effect' and 'perspective-taking' correlated significantly with students' historical content knowledge and historical reasoning ability, but were not unique predictors. Furthermore, from analyses with regard to the engagement profiles it could be concluded that students who showed high behavioural and cognitive engagement while reading (i.e., the 'independent readers', 'help-seeking readers', and 'uncertain readers') performed significantly better on both components of the delayed HICK test compared to students with lower engagement (i.e., the 'naïve readers' and 'stubborn readers').

Integrative Findings and Experienced Challenges

Research on implementing educational technology in an ecologically valid, subject-specific context is often accompanied by various practical and scientific challenges. Similar to the context of PD intervention research, the combination of rigorous research and explorative analyses at the early stages of using an original DLE, including a detailed discussion of the integrative findings, may offer relevant starting

points for improvement (Hill, Beisiegel, & Jacob, 2013; see also Chapter 4). The following subsections shed light on aspects of the use and usefulness of the DLE that were apparent across the studies in this dissertation.

Students' use of the DLE. The overarching research project aimed to support students' self-regulation skills in a DLE in the context of reading comprehension of expository texts. Since students read the texts individually, it was expected that providing them with the choice to deliberately access supportive hints would invoke SRL processes. Following this approach, the support mechanism embedded in the DLE (in the form of cognitive, metacognitive, and motivational hints) was dependent on students' metacognitive monitoring, followed by (deliberately) accessing a hint by clicking on one of the hint buttons. However, in both years of the intervention students accessed, on average, few hints.

In the first year of the overarching research project, students from the Experimental groups A and B almost accessed no hints at all. In the first study (Chapter 2), half of the students from Experimental group A did not access any of the available hints and about 15% of the students only accessed a hint once. Metacognitive hints were accessed rarely overall. When asked for an explanation, students argued that they "did not know they were there", that to them, "using hints is like cheating", or that they ignored the hints on purpose, because "hints only contained even more text to read". In response to their explanations, the DLE was adapted in such a way that students were informed about the use and usefulness of hints via a pop-up screen at the beginning of each lesson, starting from the second phase of year 1. Nevertheless, the results from the subsequent interventions (i.e., Phase 2, year 1; both phases in year 2) did not show a substantial increase in students' overall hint use. The second study (Chapter 3) showed a high variation in hint use between the students, and more than half of the students belonged to the profiles in which overall hint use was relatively low.

Teachers' use of the DLE. With regard to the use of the DLE by teachers, all studies showed that teachers were willing to incorporate a DLE into their regular history lessons to let their students read expository texts. Although some teachers mentioned that the contents of the texts did not always align with the subject taught in the regular curriculum at that time in the school year, they implemented the DLE in their lessons, resulting in the majority of their students finishing all lessons in the DLE. Most teachers used the basic visualised data output in both years to track students' progress in the DLE (i.e., whether they finished a lesson) and performance

on their reading tasks; the teachers experienced this as a useful tool, even though they did not use it frequently.

However, even though the teachers emphasised the usefulness of the visualised data output, they did not use it to substantiate or adapt their regular instruction. During the lesson observations, the extended data visualisations were seldom used during teachers' instruction. Teachers indicated several contextual barriers for using the data output in their regular lessons, such as time pressure and high workload. Moreover, most of the teachers viewed the DLE as something separate from their regular lessons, for example by introducing it to their students as "the university's research", which might explain why they did not integrate the data output in their regular lessons. Furthermore, teachers did not always implement the use of the DLE as planned, which is a central issue in the discussion section of Chapter 4. These implementation difficulties have also been described by many other studies in this area (Dignath & Büttner, 2008; Staman, Timmermans, & Visscher, 2017; van Kuijk, Deunk, Bosker, & Ritzema, 2016).

The effects of using the DLE for students. In general, when comparing pre and posttest scores, students' text comprehension performance declined in each study, irrespective of their experimental condition, their initial reading level, or their engagement profile membership. This conclusion contradicts earlier findings by Cheung and Slavin (2012), who found that intensive reading interventions resulted in larger and positive effects on students' reading performance. Likewise, student motivation for the subject of history declined throughout the school year, which concurs with the findings of Azevedo, Cromley, and Seibert (2004), who found less stated interest by students when they were scaffolded with domain-specific hints. This finding can also be explained by the fact that student motivation is known to decrease after the transition from primary to secondary education and during a school year in general (Opdenakker, Maulana, & den Brok, 2012).

Nevertheless, in-depth analyses of students' hint use, as reported in Chapter 2, showed a positive and significant difference in posttest text comprehension between students who used hints and students who did not, in favour of the hint users. Comparisons between students with different initial reading comprehension levels showed that the performance scores of below-average students did not decrease significantly. Additionally, the findings with regard to the engagement profiles in Chapter 3 showed that students who used the hints frequently performed better on the reading comprehension posttest compared to students from the profiles who used few

hints, with the exception of the 'independent readers' profile. Combined, these findings indicate that for some students, in particular those who struggle with reading texts, using hints may contribute to their reading comprehension performance. For other students, in particular above-average readers (see Chapter 2) and the 'independent readers' profile (see Chapter 3) practicing reading comprehension in a non-adaptive DLE is not as effective and possibly even detrimental for their motivation in terms of task value. For these students, it might be more beneficial to work in a dynamic DLE that can adapt to students' individual needs. Hence, it is essential to carefully determine which students may benefit from the embedded support in a static DLE, and which students need different or more challenging reading tasks.

The effects of the DLE on teacher practices. Based on the available data, it is difficult to draw firm conclusions about the usefulness of the DLE for the teachers involved in this research. The implementation of the DLE often did not happen as planned, and the experimental group consisted of a small number of teachers. In addition, the participating teachers differed between the first and second year of the overarching research project. Since four new teachers were added to Experimental group A in the second year of the overarching research project (see Chapter 4), it was not possible to compare their experiences with the previous year.

Nevertheless, the qualitative teacher interviews and focus groups, as well as the experiences teachers communicated personally to the researchers throughout the overarching research project, indicated that teachers found it relevant to work with the DLE in their lessons. They were interested to see how their students performed and they were convinced about the potential usefulness of the basic and extended visualised student data. In general, their participation stimulated the teachers to reflect on their instructional practice with regard to reading strategy instruction, which can be considered a useful first step towards more and improved reading instruction in history education.

Limitations and Methodological Considerations

The overarching research project that formed the basis of this dissertation adopted a small-scale, practice-oriented approach. An advantage of this approach is that the research has been conducted in an ecologically valid context, which is relevant for educational practice. However, although the available data have been analysed carefully and extensively to provide both a qualitative and quantitative overview of the use and usefulness of implementing a DLE, there are several general limitations

that influence the results reported in this dissertation. These general limitations include methodological considerations such as the instruments used as well as several decisions made throughout the data analyses, which are outlined in the following subsections.

The influence of grouping procedures on study outcomes. All studies in the current dissertation have been involved in a process of choices and decisions with regard to grouping procedures. Like a kaleidoscope, using a certain lens or focusing on specific groups provides a unique picture of certain elements and outcomes in the process of students' self-regulated reading of texts in a DLE, but it is important to keep in mind that using a different lens might show different results. To determine the impact of the results found in the current studies, it is essential to reflect on these grouping procedures.

In Chapter 2, the distinction between below-average, average, and above-average students, based on means and standard deviations of the initial reading comprehension test scores, offers a basic indication of students' performance at that time. Likewise, the two operationalisations of hint-users and non-hint users were based on rather arbitrary cut-off points. The results showed that the operationalisation of 'hint users' as students who used one or more hints during the intervention led to significant differences on students' posttest text comprehension, whereas the operationalisation of 'hint users' as students who used *multiple* hints versus using a single hint, or no hints at all, did not yield any significant result. Including both operationalisations of the term 'hint user' showed the difficulty of determining why students used these hints and whether this use was effective for students' text comprehension. For example, students may have accessed a single hint out of curiosity, but this does not tell us anything about whether they used the supportive strategy information provided. Another possibility is that using a single hint was helpful to these students in such a way that they did not need to access another, similar hint.

In Chapters 3 and 5, groups of students (i.e., 'reader types') were created based on Latent Profile Analysis (LPA). The predictor variables used in the LPA, such as time on task and hint use, were based on the availability and suitability of log file data. Moreover, the determination of the best fitting profile solution was based on a combination of three criteria typically used in LPA research: statistical model fit, parsimony, and interpretability (Hickendorff, Edelsbrunner, McMullen, Schneider, & Trezise, 2018). Several indicators were used to determine the statistical model fit: Akaike information criterion (AIC), Bayesian information criterion (BIC), and the

entropy statistic. However, in light of the practice-oriented approach of the second study (Chapter 3), the interpretability and practical value of the final profile solution were the deciding factors in opting for the five-profile solution. The results reported in Chapters 3 and 5, for example those with regard to the differences between the profiles, are highly influenced by this choice, which should be taken into consideration when interpreting the results.

Measuring self-regulated learning using self-reports. In this dissertation, students' SRL was measured using self-report questionnaires, but there is an ongoing scientific debate about this approach (Azevedo, 2009; Schellings & van Hout-Wolters, 2011; Veenman, 2007). Self-report measures are widely used and easy to administer in large-scale testing, but they offer a subjective and personal interpretation (e.g., students report what they think they do). On the other hand, recent studies emphasise the temporal structure of SRL processes, and propose that multimodal data should be included more often to understand students' regulation of learning (Noroozi et al., 2019). Learning analytics provide the opportunity to measure and support students' SRL processes in real-time (e.g., what students actually do), using online trace data such as log files, eye-tracking, facial expressions, or even students' heart rate, perspiration, and electro-dermal activity (Bannert, Molenaar, Azevedo, Järvelä, & Gašević, 2017). It must be noted, however, that these multimodal data offer challenges as well, since these datasets are often extensive and hard to interpret.

Following this argument, the variables used in this dissertation do not provide a comprehensive image with regard to students' SRL process. Students' hint access is visible in the log file data, but it is unknown whether students considered the hints useful or whether they applied the strategy information provided in the hint. Moreover, hint use in itself is neither a good nor a bad SRL strategy. As described by Roll, Baker, Aleven, and Koedinger (2014), not using the hints can lead to what the authors call 'productive failure': avoiding help (and, to a certain extent, repeated failure) is often associated with better learning than seeking help when students encounter problems (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). Students may benefit from engaging in self-initiated solution attempts before they use the available support. Therefore, not using hints can also indicate the use of a deliberate SRL strategy, apart from the strategies that were measured with the self-report questionnaires. Therefore, it is essential to study the entire process from encountering a problem to finding a solution, which may or may not include the use of supportive hints. To this aim, qualitative research methods, such as think-aloud protocols,

process-mining techniques, or sequence analysis with regard to students' behaviour in a DLE could provide additional, more detailed information about students' self-regulation processes when working in digital environments.

Measuring reading comprehension, historical content knowledge, and historical reasoning ability using researcher-developed tests. Several outcome measures in this dissertation were developed specifically for this research project. By doing so, the results fit well within the research context. Since the contents of the texts and hints embedded in the DLE were created in cooperation with the participating teachers, it was important to align the corresponding multiple-choice questions with these contents. The multiple-choice questions in weeks 1 and 6 of each phase functioned as pretest and posttest measures of students' reading comprehension performance. However, these tests only contained ten multiple-choice items, and, therefore, did not fully resemble reading comprehension instruments frequently used in the educational literature. The use of researcher-developed tests to measure reading comprehension occurs regularly in practice-oriented research (cf. ter Beek, Brummer et al., 2018), since it can be challenging to fully capture students' comprehension in combination with the practical and contextual constraints of a regular lesson. This should be taken into consideration when interpreting the results with regard to students' text comprehension.

The instrument to measure students' historical content knowledge and historical reasoning ability was also created specifically for the fourth study (Chapter 5), even though it was based on the Assessment of Social Studies Knowledge (ASK) instrument of Vaughn et al. (2013). According to Reich (2009), historical content knowledge can be measured with multiple-choice questions, but this type of questions cannot fully capture students' historical reasoning ability. Therefore, three open-ended questions were created to measure students' historical reasoning ability, but this was still a written test. Although relevant results were found with regard to students' historical reasoning ability, oral and interactive methods, such as think-aloud protocols, student interviews, or classroom discussions, might give a more detailed overview of how students think and reason historically.

The possible barriers of practice-oriented research. Conducting educational research in an ecologically valid context is desirable in many ways, but it can also offer practical challenges for both researchers, teachers, and students. The studies in this dissertation attempted to find a balance between the intended scientific research design and its feasibility in daily educational practice. However, this led to several

limitations with regard to the available data and the context in which the studies were conducted.

Although the intervention underlying this dissertation lasted for two full years, we were only able to report student results with sufficient reliability for the first phase of each school year because there were many missing data for the second phase. Both teachers and students experienced more implementation difficulties in the second phase compared to the first phase; for teachers, this was mostly in terms of planning and availability of IT resources, while for students their motivation to work with the DLE was often an issue. Student evaluations and experiences from the lesson observations showed that a vast group of students experienced working in the DLE as boring and useless, and some teachers suggested that the lack of a reward (e.g. grades, bonus points, or other forms of extrinsic rewards) was the main cause of this negative perception. These findings show that it is challenging to implement a DLE, which should be conceived as a 'safe' practise environment due to the lack of performance pressure, in a school culture that is mainly focused on grades and student performance.

Recommendations for Future Research

The concept of blended learning, in which online educational materials and traditional classroom practice are combined, is becoming increasingly popular in the Dutch educational system. Recently, Boelens, De Wever, and Voet (2017) systematically analysed four key challenges for blended learning and found that flexibility, interaction, support of students' learning processes, and affect are four important components to take into account in research on blended learning environments. The results in the dissertation at hand show similarities with the challenges mentioned by Boelens et al. (2017). In fact, these four components, especially in light of the main findings of the current dissertation, provide useful suggestions for implementing a DLE to read texts in future research.

First, the authors mention the importance of incorporating flexibility. According to the self-determination theory of Ryan and Deci (2000), students' intrinsic motivation can be increased by stimulating feelings of autonomy. Although the access and use of hints in the DLE was optional, the texts and accompanying assignments were fixed and the same for all students. Since the DLE texts had to complement the topics of the regular history lessons, students were not offered the possibility to choose the subject of their interest or to work in the DLE at their own pace. Future

research could increase the levels of autonomy in a DLE and examine whether this adds to students' intrinsic motivation and reading performance. In addition, the repeated administration of the hefty MSLQ and MARSI questionnaires (four times per school year) led to satiation and sometimes even aversion among the students. When self-report questionnaires are to be used in future research, it is important to consider the length and flexibility of administering these questionnaires, while also considering the reliability of the scale scores of these instruments.

Second, Boelens et al. (2017) emphasise the need for facilitation of interaction in blended learning environments. Research on (computer-supported) collaborative learning has shown that student collaboration can enhance reading performance (Moeken, Kuiken, & Welie, 2016), intrinsic reading motivation (Guthrie, McRae, & Klauda, 2007), and even historical reasoning ability (van Drie, 2005). However, the design of the DLE used in this dissertation adopted an individual approach towards reading texts, because it aimed at measuring students' individual reading process. Therefore, it is only possible to draw conclusions about how students interacted with the environment, but not about the possible role of interaction between students. Future research could include a cooperative component in the DLE, such as chat options, to stimulate and investigate interactions and collaboration between students while reading (cf. van Drie, 2005).

Third, it is important to support students' SRL processes in terms of regulative and affective strategies. For example, Boelens and colleagues note that students performing below average in online environments may not yet possess the required SRL skills to learn independently. The support mechanisms embedded in the DLE are a central element of the majority of studies included in this dissertation, but more insight is needed into which specific types of strategy support are effective for students with different characteristics (ter Beek et al., 2018). The studies in this dissertation focused on the actual use of the hints in the DLE; however, they did not consider students' item-level navigation sequences or improvement of given answers after using the hints. Future research could include detailed analyses of students' hint using process to uncover whether the hints were helpful for a specific student at a specific time point in the reading process.

Fourth, fostering an affective learning climate—as recommended by Boelens et al. (2017)—is important for both students and teachers. When students and teachers feel safe, valued, and have positive attitudes towards the task at hand, this may lead to higher intrinsic motivation. As mentioned before, the intrinsic motivation of

students included in this dissertation decreased over time. To stimulate students to read texts, it is important to ascertain what drives these students. The same accounts for teachers' implementation of the DLE. As suggested by one of the participating teachers, scientific research in the field of educational technology should be aligned with teachers' practical needs, such as efficient grading or integrating training with existing PD programs, to ensure that the research project has practical value and that teachers are motivated to contribute. In addition, teachers might also be involved in developing the research design or analysing the data, to strengthen their interest and responsibility with regard to the research at hand. In the Dutch educational context, in which increasing time pressure and workload are predominant, researchers will have to coordinate the needs of all parties involved to bridge the gap between educational science and practice.

The dissertation at hand has shown promising results in the area of practice-oriented research with regard to using DLEs to support expository history text reading in lower secondary education. To deepen our knowledge on this topic, future research could also dive into the effectiveness of supportive hints by analysing different variations of the hints offered, such as hints that appear automatically after an incorrect answer, or hints that contain audio or video materials instead of written text. Additional trace data and log file data can be added to the analyses to explore the role of student engagement more in-depth. In addition, future research into the effectiveness of providing teachers with a PD training could be extended with more intensive program, enhancing teachers' involvement. Lastly, research on the effects of reading interventions on students' historical reasoning skills can be enriched by using even more components of the framework of historical reasoning of Van Boxtel and Van Drie (2018).

Practical Implications and Recommendations

For history education in lower secondary education, it is essential to include and combine reading strategy instruction and students' reading of relevant, domain-specific expository texts. History teachers play a major role in this process: It is essential that they acknowledge the relevance and importance of incorporating reading strategy instruction in their lessons. As Alexander and Kulikowich (1991) put it:

> To teach content information without incorporating instruction in strategic processing or to teach strategies in isolation of content information may

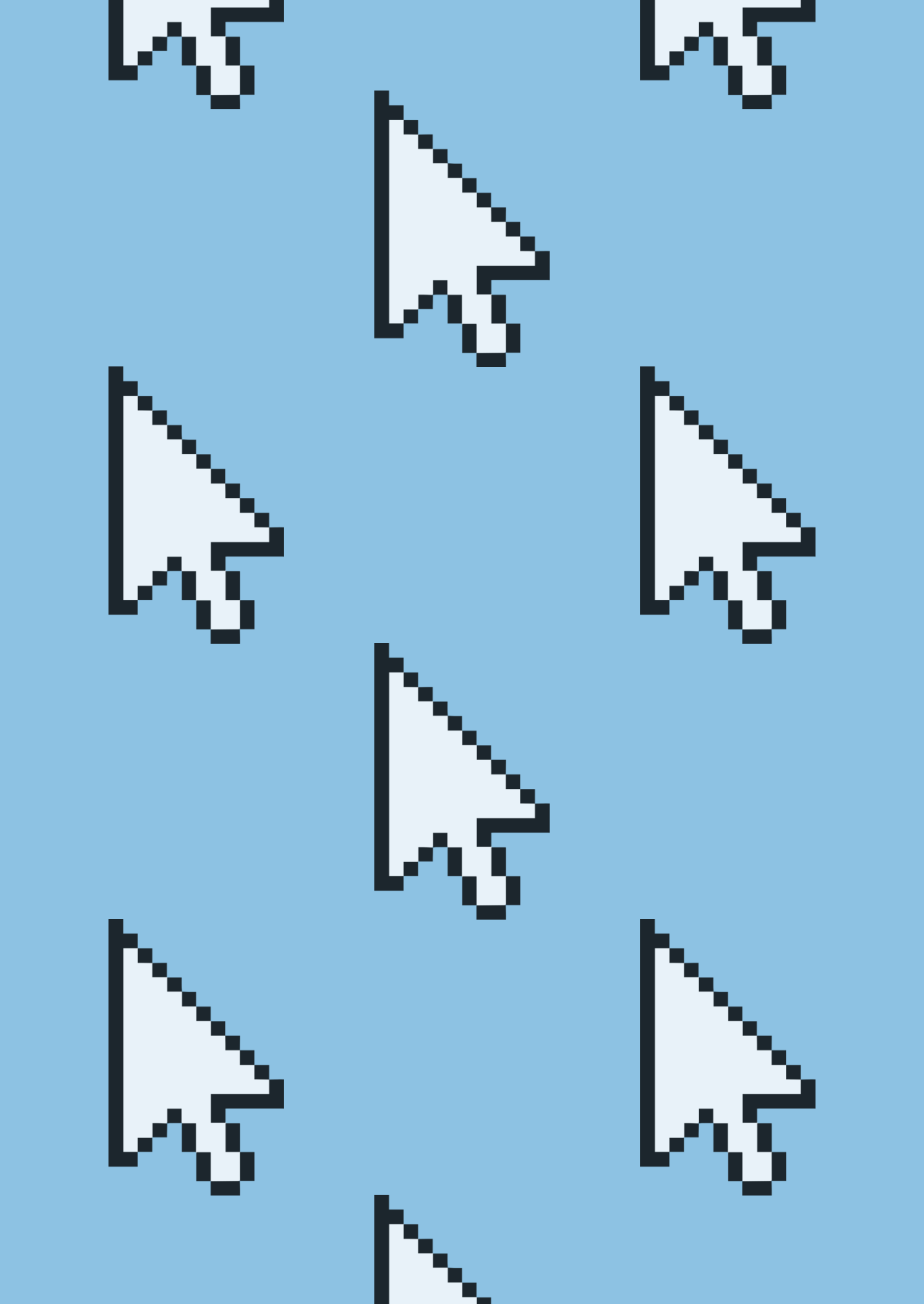
> contribute to segregation of these knowledges when integration is what is required. [...] Therefore, if a *marriage* of content and strategy knowledge is desired in our students, then teachers and teacher educators have a critical role to play in forging such a meaningful and long-lasting relationship. (p. 186, italics in the original)

The implementation of digital technology can offer support to both students and teachers to increase their knowledge and practice their skills with regard to reading strategies. Although the results did not show significant improvements in students' reading performance or motivation in general, it was found that below-average readers' performance did not significantly decrease, indicating that this group might benefit most from this type of practice. In addition, providing students with supportive hints improves their awareness of problem-solving strategies, regardless of the actual use of these hints. When students do use the hints provided, their performance has shown to be significantly better than that of students who do not use hints at all. Nevertheless, the current dissertation has shown that the possible benefits of using the DLE are dependent on students' behavioural and cognitive engagement as well as on teachers' implementation of the DLE in relation to the regular history curriculum. Students who are behaviourally and cognitively engaged when working in the DLE show higher motivation and performance levels compared to students who spend little time on a reading task, ignore available support, and do not take the assignment seriously. It is therefore important that teachers stimulate students' cognitive and behavioural engagement when reading texts. Moreover, it is helpful when teachers adopt a positive stance towards the use of technology for instruction in the classroom and are aware of the context factors that may impede successful implementation.

In general, the participating teachers expressed positive feelings towards using the DLE to read texts in their history lessons. Nevertheless, it is important to note that the application of digital technology by both teachers and students requires sufficient time and space for adaptation and acclimatisation. Rapid developments in educational technology provide teachers with a plethora of possibilities to substantiate and adapt their instruction. However, in the current context of high work pressure, teacher strikes, and alarming burnout rates in Dutch secondary education, considerable effort should be devoted by school boards or school leaders to increase the financial and developmental resources that can support the integration of digital technology in the (history) classroom.

Conclusion

The results presented in the current dissertation show that educational technology, particularly a DLE with supportive cognitive, metacognitive, and motivational hints for students, and visualised data output for teachers, can make a significant contribution to history education in terms of reading expository texts. This finding is relevant for educational practice since students' comprehension of texts is an indispensable skill for interpreting and understanding the past. For the current educational situation in the Netherlands, in which experts are reviewing existing educational curricula to take into account the knowledge and skills that suit our modern, 21st-century society, it is crucial to emphasise the role of reading comprehension and the stimulation thereof. This applies in particular to the subject of history in lower secondary education, for which the current dissertation has shown that language-oriented lessons in a scaffolded DLE can be meaningful.



Overview of appendices

Appendix A - Abbreviations and acronyms

		Chapter(s)
AIC	Akaike Information Criterion	3
AMSR	Adolescent Motivations for School Reading	3
ANOVA	Analysis Of Variance	2, 3
ANCOVA	Analysis Of Covariance	2, 3
ASK	Assessment of Social studies Knowledge	5
BIC	Bayesian Information Criterion	3
CR	Comprehensive Reader	4
DBDM	Data-Based Decision Making	4
DLE	Digital Learning Environment	1, 2, 3, 4, 5
E	Explanation (instructional strategy)	4
ER	Effort Regulation (MSLQ subscale)	2
ex	Explicit strategy instruction	4
Gazelle	Gemotiveerd, actief en zelfstandig lezen	1, 2, 3, 4, 5
GLM	General Linear Model	2, 3, 5
GLOB	Global reading strategies (MARSI subscale)	2
havo	Hoger algemeen voortgezet onderwijs	2, 3, 4, 5
HICK	Historical Content Knowledge	5
HR	Help-seeking Reader	4
IGO	Intrinsic Goal Orientation (MSLQ subscale)	2
IM	Intrinsic Motivation	3
im	Implicit strategy instruction	4
IR	Inconsistent Reader	4
JOL	Judgment Of Learning	3
LL	Log Likelihood	3
LPA	Latent Profile Analysis	3, 5
M	Mean value	2, 3, 4, 5
M	Modelling (instructional strategy)	4
MARSI	Metacognitive Awareness of Reading Strategies Inventory	2
MCAR	Missing Completely At Random	3
MCQ	Multiple-Choice Questions	3, 5

		Chapter(s)
MRIB-S	Motivations for Reading Information Books - School	3
MSLQ	Motivated Strategies for Learning Questionnaire	2, 3
MSR	Metacognitive Self-Regulation (MSLQ subscale)	2
N or *n*	Number of participants/members of a group	2, 3, 4, 5
Npar	Number of free parameters	3
NRO	Nationaal Regieorgaan Onderwijsonderzoek	1
NRO-PPO	NRO-Programmaraad voor Praktijkgericht Onderzoek	1
OEQ	Open-Ended Questions	5
PD	Professional Development	4
PIRLS	Progress in International Reading Literacy Study	1
PISA	Programme for International Student Assessment	1
PROB	Problem-solving strategies (MARSI subscale)	2
Q	Questioning students (instructional strategy)	4
RQ	Research Question	2, 3, 4, 5
SD	Standard Deviation	2, 3, 4, 5
SE	Self-Efficacy (MSLQ subscale)	2, 3
SRL	Self-Regulated Learning	2, 3
SUM	Summary/summaries	3
SUP	Support reading strategies (MARSI subscale)	2
T1/T2	Time points for measurements	2, 3, 4
TV	Task Value (MSLQ subscale)	2, 3, 5
vmbo	Voortgezet middelbaar beroepsonderwijs	2, 3, 4
vwo	Voorbereidend wetenschappelijk onderwijs	2, 3, 4, 5

Appendix B - Original and adapted IM scale items

The IM scale used in this study was composed by means of various items from two existing instruments used to measure students' reading motivation.

Original instrument: AMSR, intrinsic motivation scale (Coddington, 2009):

1. I enjoy reading for Language Arts/Reading class.
2. I enjoy it when reading materials for Language Arts/Reading makes me think.
3. I enjoy reading in my free time for Language Arts/Reading class.
4. I like to read for Language Arts/Reading class.
5. Reading for Language Arts/Reading class is boring to me.*
6. Reading for Language Arts/Reading class is a waste of time.*

Original instrument: MRIB-S, intrinsic motivation scale (Guthrie et al., 2009):

7. I read information books for school because it's fun.
8. The information books I read for school are interesting.

Adapted IM scale used in this study:

1. I enjoy reading texts for history class.
2. I enjoy it when reading texts for history make me think.
3. I enjoy reading texts in my free time for history class.
4. I like to read texts for history class.
5. Reading texts for history class is boring to me.*
6. Reading texts for history class is a waste of time.*
7. I read informational history texts because it's fun.
8. The informational texts I read for history are interesting.

Note. An asterisk denotes a reversed item. The adapted items were translated to Dutch before administration.

Appendix C - Examples of HICK MCQs and OEQs

Examples of multiple-choice and open-ended questions in the HICK instrument

The questions in the HICK instrument are inspired by the Assessment of Social Studies Knowledge (ASK) instrument (Vaughn et al., 2013).

Multiple-choice questions (MCQ) – Historical content knowledge (30 points in total)

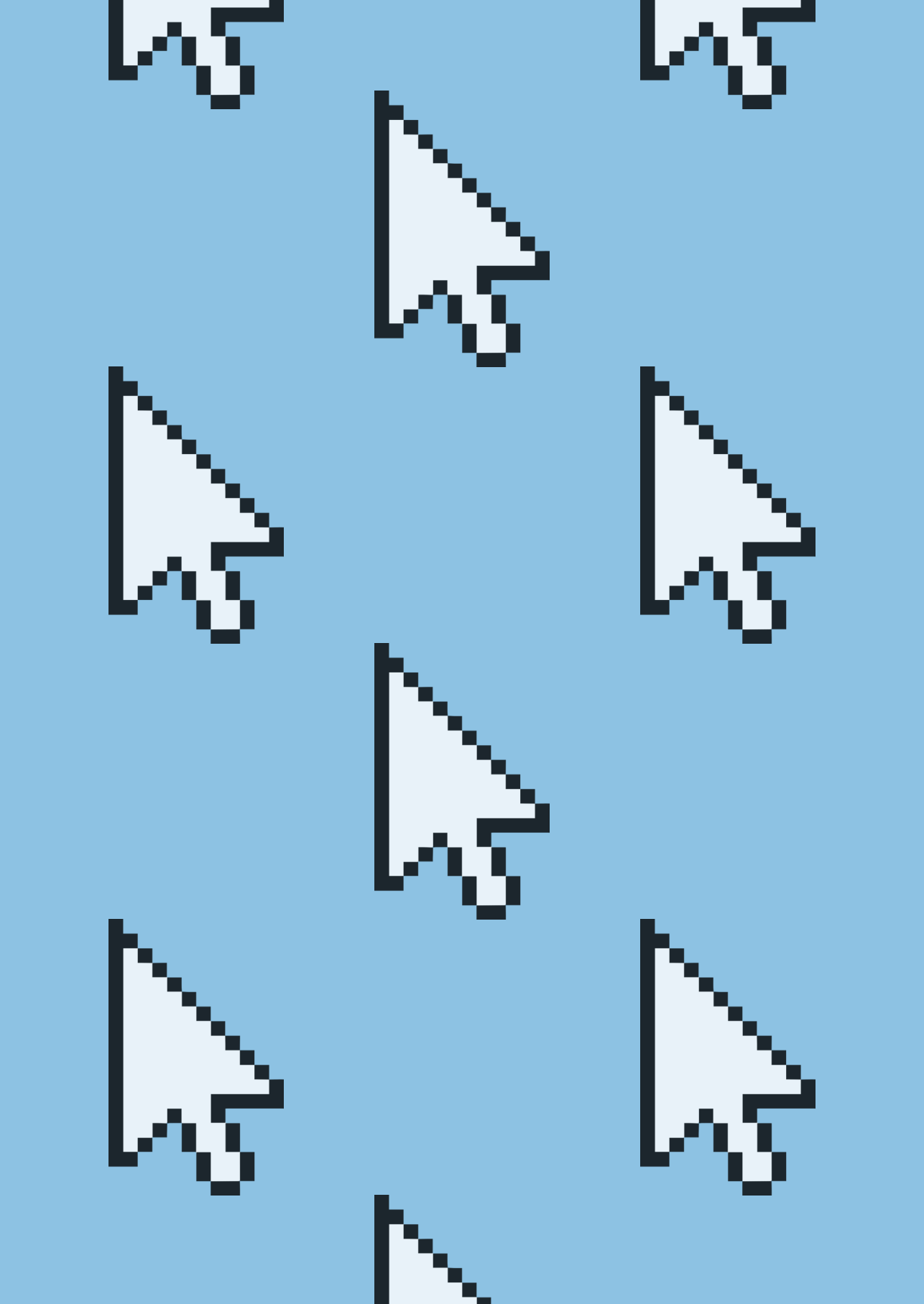
1. Which people defeated the Spartan army in the battle of Thermopylae?
    A. The Athenians
    B. The Persians
    C. The Ionians
    D. The Romans

2. A different name for the Greek term *ekklesia* is:
    A. Democracy
    B. Voting rights
    C. Legislation
    D. Public assembly

Open-ended questions (OEQ) – Historical reasoning ability (10 points in total)

1. Women were not allowed to compete in the Olympic Games. What would the Spartans have thought of that? And what about the Athenians? Explain your answer.
……
……
……
……
……

Note. The items in this appendix were translated from Dutch to English for readability purposes.



Nederlandstalige samenvatting (Dutch summary)

Ondersteuning van begrijpend lezen in het geschiedenisonderwijs

Het gebruik en de bruikbaarheid van een digitale leeromgeving

Begrijpend lezen is een essentiële vaardigheid voor het verwerken van tekstuele informatie en het verwerven van kennis. Het kunnen lezen en begrijpen van informatieve teksten is een cruciale voorwaarde voor de leerprestaties binnen vrijwel elk schoolvak, maar ook buiten het klaslokaal is lezen van groot belang: De overvloed aan informatie binnen onze moderne, digitale 21e-eeuwse samenleving vraagt om kritische, geïnformeerde en bekwame lezers (Alexander & The Disciplined Reading and Learning Research Laboratory, 2012; Kamil, Afflerbach, Pearson, & Moje, 2011). Tekstbegrip is essentieel voor algemene leestaken in het dagelijks leven, zoals het opvolgen van instructies in medische bijsluiters, het begrijpen van juridische voorwaarden, of het kunnen herkennen van *fake news* (Raad voor Cultuur & Onderwijsraad, 2019). Experts op het gebied van lezen bepleiten daarom dat instructie en oefening in begrijpend lezen een centraal onderdeel moeten vormen binnen het curriculum van elk schoolvak (Pereira & Nicolaas, 2019). Met name voor het vak geschiedenis is dit een relevante aanbeveling, omdat de lesmethodes vaak een grote hoeveelheid feitelijke, verklarende teksten bevatten.

Het belang van begrijpend lezen voor het vak geschiedenis

> "Het afzetgebied van de Griekse nijverheid werd veel groter door de kolonisatie. Athene exporteerde bijvoorbeeld vazen, juwelen, wijn, honing en natuurlijk olijven. Griekse handelaren brachten hun handelswaar naar alle kusten van de Middellandse Zee en de Zwarte Zee. Overal in dit gebied zijn resten van Grieks aardewerk teruggevonden."

Het bovenstaande fragment over de Griekse wereld is gebaseerd op een informatieve tekst uit een bestaande geschiedenismethode voor de brugklas havo/vwo. Op het eerste gezicht lijkt de tekst wellicht vrij eenvoudig, maar bij nader inzien heeft een leerling voldoende voorkennis en een grote woordenschat nodig om deze tekst echt te kunnen begrijpen. In tegenstelling tot de verhalende teksten die leerlingen gewend zijn vanuit het primair onderwijs, bevatten teksten uit lesmethodes in het

voortgezet onderwijs vaak veel vakspecifieke woorden en is er sprake van een hoge informatiedichtheid (Mastropieri, Scruggs, & Graetz, 2003; Ramsay, Sperling, & Dornisch, 2010; Swanson et al., 2016). Zo heeft een leerling in het bovenstaande voorbeeld de nodige voorkennis nodig over de geografische ligging van Athene, over wat het begrip 'kolonisatie' inhoudt, en hoe handel werd gedreven in deze tijdsperiode. Ook hun woordenschat is belangrijk om schooltaal of vakjargon, zoals 'afzetgebied', 'nijverheid', en 'exporteerde', te kunnen begrijpen. Tot slot moet een leerling inferenties kunnen maken: wat heeft het terugvinden van aardewerk te maken met handelaren? Naast het leggen van relaties tussen bepaalde zinnen, alinea's, of zelfs hele hoofdstukken, dienen leerlingen ook relevante leesstrategieën toe te kunnen passen op het moment dat zij een deel van de tekst niet begrijpen.

In de internationale onderzoeksliteratuur komt vaak de term *disciplinary literacy* terug, waarmee wordt bedoeld dat het leesonderwijs binnen specifieke schoolvakken gericht zou moeten zijn op relevante leesvaardigheden (Moje, 2015; Monte-Sano, 2011; Reisman, 2012; Shanahan & Shanahan, 2008; Wineburg & Reisman, 2015). Zo zouden leerlingen tijdens het lezen van teksten voor het vak geschiedenis historisch moeten kunnen redeneren en reflecteren op de onderwerpen en verschillende perspectieven die ze in de tekstboeken tegenkomen: dit weerspiegelt immers de manier waarop deskundige historici omgaan met primair of secundair bronnenmateriaal (Shanahan, Shanahan, & Misischia, 2011; van Boxtel & van Drie, 2018; van Drie & van Boxtel, 2008; Wineburg, 1991, 1998). Voor het vak geschiedenis in het voortgezet onderwijs is de toepassing van strategieën zoals *sourcing* (waar komt deze informatie vandaan?), *contextualisering* (wat is de historische context waarin dit werd geschreven?) en *corroborating* (tonen andere bronnen bevestigende of tegenstrijdige informatie?) gunstig gebleken voor het historisch begrip van leerlingen (Girard & McArthur Harris, 2012; Learned, 2018, Monte-Sano, 2011; Wineburg & Reisman, 2015).

Er zijn echter ook studies die aantonen dat deze disciplinaire aanpak voor veel (jonge) leerlingen in het voortgezet onderwijs nog wat te hoog gegrepen is, met name voor degenen die moeite hebben met begrijpend lezen in het algemeen (Duhaylongsod, Snow, Selman, & Donovan, 2015; Nokes, 2011; Perfetti, Britt, & Georgi, 1995). Faggella-Luby, Graner, Deshler en Drew (2012) stellen zelfs dat de voorgenoemde disciplinaire aanpak gebouwd is op drijfzand wanneer niet eerst voldoende aandacht wordt besteed aan algemene strategieën voor de bevordering van begrijpend lezen.

Het leesniveau en de leesmotivatie van Nederlandse jongeren

Recentelijk zijn er veel verontrustende berichten in de Nederlandse media verschenen over het begrijpend leesniveau en de leesmotivatie van Nederlandse jongeren. De Nederlandse Taalunie publiceerde hierover onlangs een uitgebreid rapport, waarin zij een oproep doen tot verandering (Pereira & Nicolaas, 2019). Uit internationale onderzoeken, zoals PIRLS (Progress in International Reading Literacy Study; Gubbels, Netten, & Verhoeven, 2017; Mullis, Martin, Foy, & Hooper, 2017) en PISA (Programme for International Student Assessment; Cito, 2012; Feskens, Kuhlemeier, & Limpens, 2016; Kordes, Bolsinova, Limpens, & Stolwijk, 2013; Organization for Economic Cooperation and Development [OECD], 2016a) is namelijk gebleken dat de leesvaardigheid van Nederlandse jongeren tussen de 10 en 15 jaar in de afgelopen twee decennia gestaag is gedaald. Recentelijk is er bij de PISA-resultaten van 2018 zelfs een sterk negatieve en significante daling van de leesvaardigheid tussen 2015 en 2018 vastgesteld, waarbij de gemiddelde Nederlandse leesvaardigheid met 485 punten het laagst was in de afgelopen 15 jaar en tevens lager was dan de gemiddelde leesvaardigheid van alle deelnemende landen (OECD, 2018). Een diepgaande analyse van de PISA-resultaten uit 2015 toonde aan dat bijna een op de vijf 15-jarigen dusdanig laag scoorde, dat zij als volwassene een hoog risico lopen op laaggeletterdheid. Dit heeft tot gevolg dat deze groep leerlingen problemen heeft bij hun ontwikkeling en hun functioneren in onze huidige informatiesamenleving (Feskens et al., 2016). De meest recente PISA-resultaten geven hetzelfde zorgelijke beeld: bijna een kwart van de 15-jarigen uit 2018 loopt een groot risico op laaggeletterdheid (Gubbels, van Langen, Maassen, & Meelissen, 2019).

Daarnaast blijkt uit internationale onderzoeken dat de leesmotivatie van Nederlandse leerlingen zwak genoemd kan worden. Bijna de helft van de 15-jarigen leest helemaal niet voor hun plezier, en zelfs de leerlingen die wel lezen, lezen slechts enkele minuten per dag (OESO, 2016b). Voor jongere leerlingen zijn de cijfers nog alarmerender: bijna een derde van alle Nederlandse 10-jarigen in het PIRLS-onderzoek uit 2016 gaf aan dat ze niet graag lezen, waardoor Nederland qua leesplezier onderaan de lijst van alle deelnemende landen staat (Mullis et al., 2017). Ook uit PISA-2018 bleek dat Nederlandse 15-jarigen zeer weinig plezier beleven aan lezen: 40% van de leerlingen gaf aan lezen te zien als tijdverspilling en 60% van hen leest enkel teksten als het echt moet, bijvoorbeeld om informatie op te zoeken (Gubbels et al., 2019). Het gebrek aan leesmotivatie kan mogelijk leiden tot lagere leerprestaties: onderzoek heeft aangetoond dat leesbegrip en leesmotivatie aan elkaar gerelateerd

zijn bij leerlingen in het lager voortgezet onderwijs (Guthrie, Klauda, & Ho, 2013).

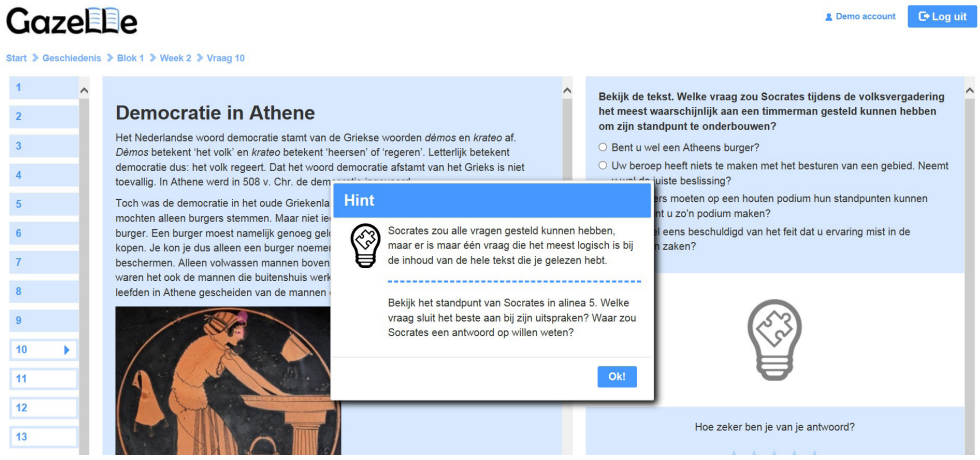
Stimuleren van tekstbegrip met educatieve technologie

In de afgelopen jaren zijn verschillende digitale leeromgevingen ontwikkeld om de lees- en leerprocessen van leerlingen te ondersteunen. Verschillende meta-analyses tonen aan dat deze digitale leeromgevingen vaak effectief zijn voor het stimuleren van begrijpend lezen in het algemeen (Cheung & Slavin, 2012; Lan, Lo, & Hsu, 2014; Moran, Ferdig, Pearson, Wardrop, & Blomeyer, 2008) of voor het vak geschiedenis in het bijzonder (O'Neill & Weiler, 2006; Poitras, Lajoie, & Hong, 2012). Zo is gebleken dat ondersteuning op het gebied van cognitie, metacognitie en motivatie—en dan vooral een combinatie van deze drie elementen—effectief kan zijn voor het tekstbegrip van leerlingen (Berthold, Nückles, & Renkl, 2007; Souvignier & Mokhlesgerami, 2006). Het is echter niet altijd duidelijk welke elementen uit deze digitale leeromgevingen leiden tot betere prestaties, omdat een beschrijving van de inhoud van de ondersteuning (wat voor informatie krijgen de leerlingen te zien?) vaak ontbreekt. Bovendien wordt tekstbegrip vaak wel gemeten, maar is er minder aandacht voor zelfregulerend leren of motivatie (ter Beek, Brummer, Donker, & Opdenakker, 2018). Het belang van begrijpend lezen bij het schoolvak geschiedenis, het feit dat de leesvaardigheid van leerlingen gestaag daalt en de veelbelovende resultaten uit onderzoek met digitale leeromgevingen vormden samen de aanleiding voor een onderzoek naar hoe het begrijpend lezen in het geschiedenisonderwijs digitaal ondersteund kan worden in de Nederlandse context.

Het onderliggende onderzoeksproject: Gazelle

Het belang van begrijpend lezen en de toenemende leesproblematiek in het voortgezet onderwijs vormden de aanleiding voor een driejarig onderzoeksproject (2015–2018), gesubsidieerd door het Nationaal Regieorgaan Onderwijsonderzoek (NRO-PPO: projectnummer 405-15-551; ter Beek, Spijkerboer, Brummer, & Opdenakker, 2018). Onder de werktitel 'Gemotiveerd, Actief en Zelfstandig Lezen' (kortweg: 'Gazelle') werd—in samenwerking met de deelnemende docenten—een programma ontwikkeld waarin brugklasleerlingen uit zowel havo/vwo- als vmbo-tl-klassen zelfstandig informatieve teksten voor de vakken geschiedenis en aardrijkskunde konden lezen, gevolgd door verschillende oefenopgaven. Hierbij kregen leerlingen de mogelijkheid om ondersteunende hints te raadplegen met daarin informatie over leesstrategieën die van belang zijn bij het lezen en begrijpen van teksten voor het vak geschiedenis. Onderzoek heeft namelijk aangetoond dat instructie over

leesstrategieën bij het lezen van geschiedenisteksten het tekstbegrip van leerlingen kan versterken (Vaughn et al., 2013). Enerzijds waren deze hints gericht op algemene leesstrategieën, zoals het afleiden van betekenissen uit de tekst. Zo luidde de hint bij vragen over oorzaak-gevolgrelaties in de tekst: 'Een oorzaak van een gebeurtenis kun je vaak afleiden door het gebruik van woorden als *vanwege* of *doordat*'. Anderzijds kwamen in de hints ook vakspecifieke leesstrategieën aan bod, zoals het bepalen van de standplaatsgebondenheid van een personage uit de tekst. Figuur 1 geeft een indruk van de inhoud van de digitale leeromgeving Gazelle.



Figuur 1 Screenshot uit de digitale leeromgeving 'Gazelle' met een cognitieve hint.

Naast ondersteuning van de leerlingen bood de digitale leeromgeving ook ondersteuning aan de docenten in de vorm van gevisualiseerde data-output. Op basis van logbestanden, die automatisch werden opgeslagen tijdens het werken in Gazelle, kon door middel van *learning analytics* een overzicht gegeven worden van het leesproces en de leesprestaties van de leerlingen. Zo konden docenten zien of leerlingen op specifieke leesvaardigheden, zoals 'oorzaak en gevolg herkennen', boven- of ondergemiddeld scoorden ten opzichte van de rest van de klas. Docenten kregen vervolgens een training aangeboden om deze data-output te vertalen naar effectieve instructie op het gebied van begrijpend lezen.

Doelen en inhoud van deze dissertatie

In deze dissertatie zijn zowel het gebruik als de bruikbaarheid van de digitale leeromgeving 'Gazelle' onderzocht. Deze digitale leeromgeving werd geïmplementeerd om enerzijds het tekstbegrip van brugklasleerlingen bij het vak

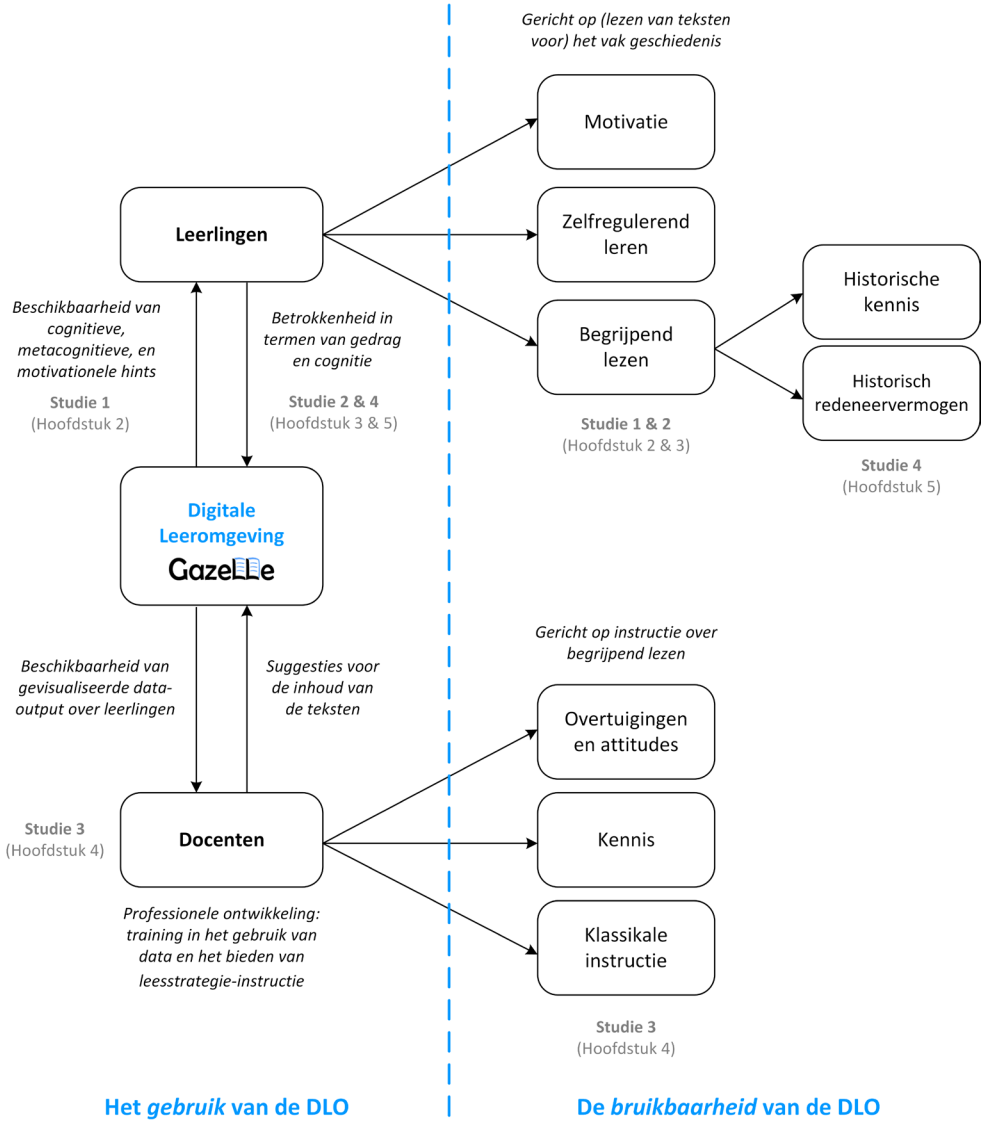
geschiedenis te ondersteunen, en anderzijds de geschiedenisdocenten te voorzien van bruikbare informatie over de prestaties en het leesproces van hun leerlingen. Daarom werden in deze dissertatie twee onderzoeksvragen parallel bestudeerd:

- Hoe gebruiken en ervaren leerlingen uit de brugklas en hun docenten een met strategische hints en gevisualiseerde data-output verrijkte digitale leeromgeving voor het vak geschiedenis (d.w.z., wat kunnen we zeggen over het *gebruik* van de digitale leeromgeving)?
- Wat zijn de effecten van het gebruik van deze digitale leeromgeving op enerzijds het leesproces en de leerresultaten van leerlingen en anderzijds op het instructiegedrag van geschiedenisdocenten (d.w.z., wat is de *bruikbaarheid* van de digitale leeromgeving)?

Elk hoofdstuk uit dit proefschrift beschrijft een aparte studie. Deze studies zijn allemaal uitgevoerd binnen de overkoepelende context van het vak geschiedenis in de brugklas van het voortgezet onderwijs. Figuur 2 geeft een overzicht van de concepten die in dit proefschrift zijn opgenomen, evenals de studies en hoofdstukken waarbinnen ze zijn opgenomen.

Resultaten van de verschillende studies

Studie 1 (hoofdstuk 2) was gericht op het aanbieden van cognitieve en metacognitieve hints in de digitale leeromgeving en het effect daarvan op het tekstbegrip, zelfregulerend leren en de motivatie van leerlingen. Gedurende zes weken lazen 174 brugklasleerlingen, afkomstig uit drie verschillende middelbare scholen, informatieve teksten in de digitale leeromgeving. De ondersteunende hints werden op drie verschillende manieren aangeboden. Van week 2 tot en met 5 kregen leerlingen uit Experimentele groep A de mogelijkheid om cognitieve en metacognitieve hints te openen tijdens het lezen van teksten voor geschiedenis. Leerlingen uit Experimentele groep B konden hints raadplegen voor het vak aardrijkskunde, maar niet voor geschiedenis. Leerlingen uit de controlegroep konden voor geen van beide vakken hints raadplegen tijdens het lezen. Uit de resultaten van deze studie bleek dat het *aanbieden* van de hints niet tot significante verschillen tussen de drie groepen leidde, maar leerlingen uit de twee experimentele groepen vertoonden wel een significante groei in hun kennis van probleemoplossende strategieën ten opzichte van leerlingen



Figuur 2 Overzicht van de concepten binnen de studies en van de hoofdstukken van deze dissertatie.

uit de controlegroep. In tegenstelling tot wat werd verwacht, werden de hints bij de geschiedenisteksten door ongeveer de helft van de leerlingen helemaal niet geraadpleegd. Uit aanvullende analyses bleek dat het tekstbegrip van leerlingen die een of meer hints gebruikten meer vooruitging dan het tekstbegrip van leerlingen die geen enkele hint gebruikten. Het aanbieden of gebruiken van hints in de DLE bleek

niet van invloed op het zelfregulerend leren of de motivatie van de leerlingen. Ook bleek het tekstbegrip van de gemiddelde en bovengemiddelde lezers te dalen na het werken in de DLE. De prestaties van leerlingen met een ondergemiddeld leesniveau bleven echter gelijk. Bij deze groep nam echter wel het zelfvertrouwen (*kan ik dit?*) af. Voor de bovengemiddelde lezers verloor het lezen van geschiedenisteksten na het werken in de DLE zijn waarde (*vind ik dit nuttig?*).

In tegenstelling tot de variabelegerichte benadering van de eerste studie werd in de tweede studie (hoofdstuk 3) gebruik gemaakt van een persoonsgerichte benadering. Met behulp van logbestandgegevens, die het gedrag en de prestaties van leerlingen nauwkeurig bijhielden tijdens het werken in de digitale leeromgeving, werden verschillende betrokkenheidsprofielen geïdentificeerd. Hierdoor was het mogelijk om de invloed van de betrokkenheid van leerlingen op hun leesvaardigheid te onderzoeken. In totaal werkten 327 brugklasleerlingen, afkomstig van vier middelbare scholen, zes weken lang in de digitale leeromgeving. Uit latente profielanalyse (LPA), op basis van verschillende voorspellende variabelen zoals tijdsbesteding, gebruik van hints, gemiddelde score bij de eerste antwoordpoging en inschattingsvermogen, bleek dat een oplossing met vijf verschillende betrokkenheidsprofielen de data het beste beschreef. Het eerste en grootste profiel, dat werd aangeduid als de 'naïeve lezers', bevatte leerlingen die relatief laag scoorden op alle voorspellende variabelen. Leerlingen uit het tweede profiel, de 'koppige lezers', leken sterk op de 'naïeve lezers', met uitzondering van hun inschattingsvermogen; leerlingen uit dit profiel leken zich terdege bewust te zijn van hun minder goede prestaties, maar raadpleegden geen hints om deze te verbeteren. Het derde profiel, de 'hulpzoekende lezers', bestond uit leerlingen met gemiddelde scores op de voorspellende variabelen, maar deze leerlingen gebruikten beduidend meer hints dan de eerste twee profielen. Leerlingen uit het vierde profiel, de 'onafhankelijke lezers', scoorden relatief hoog op alle voorspellende variabelen met uitzondering van hun hintgebruik, waardoor aangenomen kan worden dat leerlingen uit dit profiel geen ondersteunende hints hoefden te raadplegen om goed te presteren. Ten slotte omvatte het vijfde en kleinste profiel, de 'onzekere lezers', leerlingen die hoog scoorden op alle voorspellende variabelen, behalve op hun inschattingsvermogen, waaruit blijkt dat deze leerlingen zichzelf vaak onderschatten. Vervolgens werden de relaties tussen deze betrokkenheidsprofielen en de motivatie van leerlingen (d.w.z., de taakwaarde, het zelfvertrouwen en de intrinsieke motivatie) en tekstbegrip onderzocht. Resultaten van deze studie laten zien dat zeer betrokken leerlingen aanvankelijk een aanzienlijk hogere taakwaarde en intrinsieke motivatie

hadden in vergelijking met leerlingen die weinig betrokkenheid toonden. Bovendien toonden zeer betrokken leerlingen een beter tekstbegrip tijdens de nameting.

De derde studie (hoofdstuk 4) was gericht op de rol van docenten bij het gebruik van een digitale leeromgeving om het begrijpend lezen van leerlingen te ondersteunen. Eerdere studies over het aanbieden van leesstrategie-instructie in het geschiedenisonderwijs hebben aangetoond dat dit type instructie niet vaak voorkomt (Linthorst & de Glopper, 2015; Ness, 2006). Dit kan worden verklaard doordat geschiedenisdocenten vaak denken dat ze niet over voldoende vaardigheden beschikken om dit soort instructie te geven of doordat ze simpelweg de kennis en de motivatie missen om dit te doen (Greenleaf, Schoenbach, Cziko, & Mueller, 2001; Hall, 2005). Professionele ontwikkeling op dit vlak stelt docenten in staat om deze vaardigheden te trainen, waardoor hun zelfvertrouwen verbeterd kan worden (Desimone, 2009). Deze studie bestond uit twee fasen. In de eerste fase kregen geschiedenisdocenten uit Experimentele groep A een visualisatie van de prestaties van leerlingen in de digitale leeromgeving in de vorm van een uitgebreid resultatenrapport in de digitale leeromgeving. Daarnaast kregen zij in de tweede fase van het onderzoek een professionele training om deze gegevens te kunnen vertalen naar gestructureerde, expliciete leesstrategie-instructie. Docenten uit Experimentele groep B kregen de uitgebreide visualisaties in de tweede fase, maar geen bijbehorende professionele training. Docenten uit de controlegroep kregen alleen basale gegevens te zien over de voortgang van hun leerlingen. In beide fasen van deze studie werden de docenten geobserveerd tijdens hun geschiedenislessen. Resultaten van dit onderzoek toonden aan dat de docenten uit zowel de experimentele als de controlegroep naar hun eigen inschatting veel kennis hadden over het geven van leesinstructie en dat zij hier positief tegenover stonden. In beide groepen was het zelfvertrouwen wat betreft het daadwerkelijk geven van dit type instructie was echter relatief lager. Vergelijkend onderzoek naar de lesobservaties toonde aan dat docenten uit de experimentele groep de variatie van hun leesstrategie-instructie vergrootten door de professionele training. De totale variatie in leesstrategie-instructie bleef in beide groepen echter relatief laag. Door middel van interviews werden enkele verklaringen gevonden voor deze bevindingen, zoals beperkte voorbereidingstijd, een hoge werkdruk, of een gebrek aan beschikbare computers.

De laatste studie (hoofdstuk 5) onderzocht de relaties tussen vakspecifieke leesvaardigheid, zoals (1) het herkennen van causale relaties; (2) het verklaren van historische gebeurtenissen; (3) het formuleren van historische vragen; (4) het

ordenen van concepten; en (5) het standpunt van de auteur bepalen, en de historische inhoudelijke kennis en het historisch redeneervermogen van leerlingen, om te onderzoeken welke specifieke leesvaardigheden in dit proces van belang zijn. Verder werden in deze studie de geïdentificeerde betrokkenheidsprofielen uit hoofdstuk 3 gebruikt om de relaties te onderzoeken tussen de betrokkenheid van leerlingen in de digitale leeromgeving en de bovengenoemde uitkomstmaten. Het merendeel van de leerlingen uit de tweede studie (hoofdstuk 3) maakte een aanvullende toets, enkele weken na het voltooien van de lessen in de digitale leeromgeving. Met deze toets konden de historische kennis en het historisch redeneervermogen van deze leerlingen worden vastgesteld. Uit de analyses bleek dat alle vakspecifieke leesvaardigheden samenhingen met zowel historische kennis als met historisch redeneervermogen. Meervoudige regressieanalyse toonde echter aan dat niet alle vaardigheden unieke significante voorspellers waren. De vaardigheden die gedefinieerd werden als 'het verklaren van historische gebeurtenissen', 'het formuleren van historische vragen' en het 'het ordenen van concepten' waren allemaal belangrijke unieke voorspellers voor historische kennis. Voor het historisch redeneervermogen waren alleen 'verklaren' en 'vragen genereren' significante unieke voorspellers. De vakspecifieke leesvaardigheden die gedefinieerd werden als 'het herkennen van causale relaties' en 'het standpunt van de auteur bepalen' waren noch uniek voorspellend voor de historische kennis van leerlingen, noch voor hun historisch redeneervermogen. Verder kon uit analyses met betrekking tot de betrokkenheidsprofielen worden geconcludeerd dat leerlingen die tijdens het lezen een hoge betrokkenheid vertoonden (d.w.z., de 'onafhankelijke lezers', 'hulpzoekende lezers' en 'onzekere lezers') significant beter presteerden op de kennistoets vergeleken met leerlingen die een lagere betrokkenheid vertoonden (d.w.z., de 'naïeve lezers' en 'koppige lezers').

Reflecties op de beperkingen binnen de studies en suggesties voor toekomstig onderzoek

Het overkoepelende onderzoeksproject dat de basis van dit proefschrift vormde, paste een kleinschalige, praktijkgerichte onderzoeksaanpak toe. Hoewel de beschikbare data uitgebreid kwalitatief en kwantitatief zijn geanalyseerd om inzicht te geven in het gebruik en de bruikbaarheid van de digitale leeromgeving, zijn er verschillende algemene beperkingen die de resultaten uit dit proefschrift mogelijk beïnvloed hebben. Deze algemene beperkingen omvatten methodologische overwegingen zoals de gebruikte instrumenten en de beslissingen die gedurende het proces van data-analyse zijn genomen. In de onderstaande paragrafen zullen de belangrijkste drie

punten worden besproken.

Dit onderzoek is uitgevoerd in een ecologisch valide onderwijscontext. Voor de doelstellingen van praktijkgericht onderzoek is dit uiteraard wenselijk, maar het zorgt tegelijkertijd voor praktische en methodologische uitdagingen voor zowel de onderzoekers, de docenten, als de leerlingen. Tijdens de implementatie van de digitale leeromgeving in de praktijk bleken er verschillende factoren te zijn die het onderzoek hebben beïnvloed. Voor docenten was dit vooral in termen van planning en beschikbaarheid van computers, terwijl voor leerlingen hun motivatie om met de digitale leeromgeving te werken een kenmerkend probleem was in de tweede fase van elk schooljaar. Evaluaties en ervaringen vanuit de lesobservaties toonden aan dat een groot deel van de leerlingen het werken in de digitale leeromgeving als saai en nutteloos hebben ervaren. Hun docenten suggereerden dat het ontbreken van een beloning (bijv. cijfers, bonuspunten of andere vormen van extrinsieke motivatie) de belangrijkste oorzaak was van deze lage motivatie. Deze bevinding laat zien dat het een uitdaging is om een digitale leeromgeving zonder extrinsieke beloningen (wat zou moeten leiden tot een 'veilige' oefenomgeving voor de leerlingen) te realiseren in een schoolcultuur waarbinnen veel waarde wordt gehecht aan cijfers en prestatiematen.

Het tweede punt van discussie heeft betrekking op de manier waarop verschillende groepen binnen dit onderzoek met elkaar zijn vergeleken. Alle studies in het huidige proefschrift zijn betrokken geweest bij een proces van weloverwogen keuzes en beslissingen met betrekking tot groeperingsprocedures. Net als bij een caleidoscoop zorgt het gebruik van een bepaalde lens of een focus op specifieke groepen voor een specifiek beeld van bepaalde resultaten met betrekking tot het zelfregulerend lezen van teksten in een digitale leeromgeving. Het gebruik van een andere lens had echter waarschijnlijk tot andere uitkomsten geleid. Om de impact van de resultaten in de studies uit deze dissertatie te bepalen, is het essentieel om te reflecteren op deze groeperingsprocedures. Zo waren er in hoofdstuk 2 twee manieren waarop hintgebruikers en niet-hintgebruikers met elkaar werden vergeleken. De resultaten toonden aan dat de operationalisering van 'hintgebruikers' als leerlingen die tijdens de interventie een of meer hints gebruikten, leidde tot significante verschillen in tekstbegrip op de nameting, terwijl de operationalisering van 'hintgebruikers' als leerlingen die meerdere hints gebruikten versus het gebruik van een enkele hint, of helemaal geen hints, geen significante resultaten gaf. Deze bevindingen tonen aan dat het een uitdaging is om de effectiviteit van hintgebruik voor het tekstbegrip van leerlingen kwantitatief te benaderen. Hetzelfde geldt voor de verschillende

betrokkenheidsprofielen uit hoofdstuk 3 en 5: deze werden bepaald aan de hand van LPA, op basis van voorspellende variabelen. Deze variabelen waren gebaseerd op de beschikbaarheid van logbestanden. In het licht van de praktijkgerichte benadering van dit onderzoek waren de interpreteerbaarheid en de praktische waarde van de uiteindelijke profieloplossing de beslissende factoren bij de keuze voor de oplossing met vijf profielen. Deze informatie dient men in acht te nemen bij het interpreteren van de resultaten van deze studies.

Tot slot heeft de derde beperking binnen dit onderzoek te maken met de manier waarop de verschillende uitkomstmaten gemeten zijn. Zo werd het zelfregulerend leren van leerlingen gemeten met behulp van zelfrapportages, terwijl daar in de wetenschappelijke literatuur al lange tijd een discussie over bestaat (Azevedo, 2009; Schellings & van Hout-Wolters, 2011; Veenman, 2007). Zelfrapportages geven namelijk een subjectieve *interpretatie* van hoe leerlingen denken dat zij hebben geleerd, terwijl digitale leeromgevingen de mogelijkheid bieden om data te registreren die een beeld kunnen geven van wat leerlingen *daadwerkelijk* op dat moment doen. Zo kan een combinatie van aanvullende metingen, zoals logbestanden, oogbewegingen, gezichtsuitdrukkingen of zelfs de hartslag en transpiratie van leerlingen meer inzicht geven in hun zelfregulatie, hoewel ook deze gegevens niet altijd eenvoudig te interpreteren zijn (Bannert, Molenaar, Azevedo, Järvelä, & Gašević, 2017). Toekomstig onderzoek zou kunnen overwegen om *real-time* metingen van het leergedrag van leerlingen in een digitale leeromgevingen te gebruiken om hun zelfregulatievaardigheden te bepalen. Daarnaast zijn het tekstbegrip, de historische kennis en het historisch redeneervermogen van leerlingen in dit onderzoek gemeten met toetsen die door de onderzoekers zelf specifiek voor dit onderzoeksproject zijn ontwikkeld. Omdat de inhoud van de teksten en hints in de digitale leeromgeving zijn gemaakt in samenwerking met de deelnemende docenten, was het belangrijk om de meerkeuzevragen op deze inhoud af te stemmen. De meerkeuzevragen uit week 1 en 6 fungeerden als voor- en nametingen van de leesvaardigheid van leerlingen. Deze metingen bestonden echter uit slechts tien vragen en geven daardoor mogelijk een wat oppervlakkig beeld van het tekstbegrip van leerlingen. Andere methodes, zoals hardop denken, interviews met leerlingen of klassikale discussies, kunnen aanvullende inzichten bieden in hoe leerlingen lezen of historisch redeneren. Deze methodes zouden in toekomstig onderzoek gebruikt kunnen worden om een gedetailleerder beeld te geven van deze processen.

Praktische implicaties

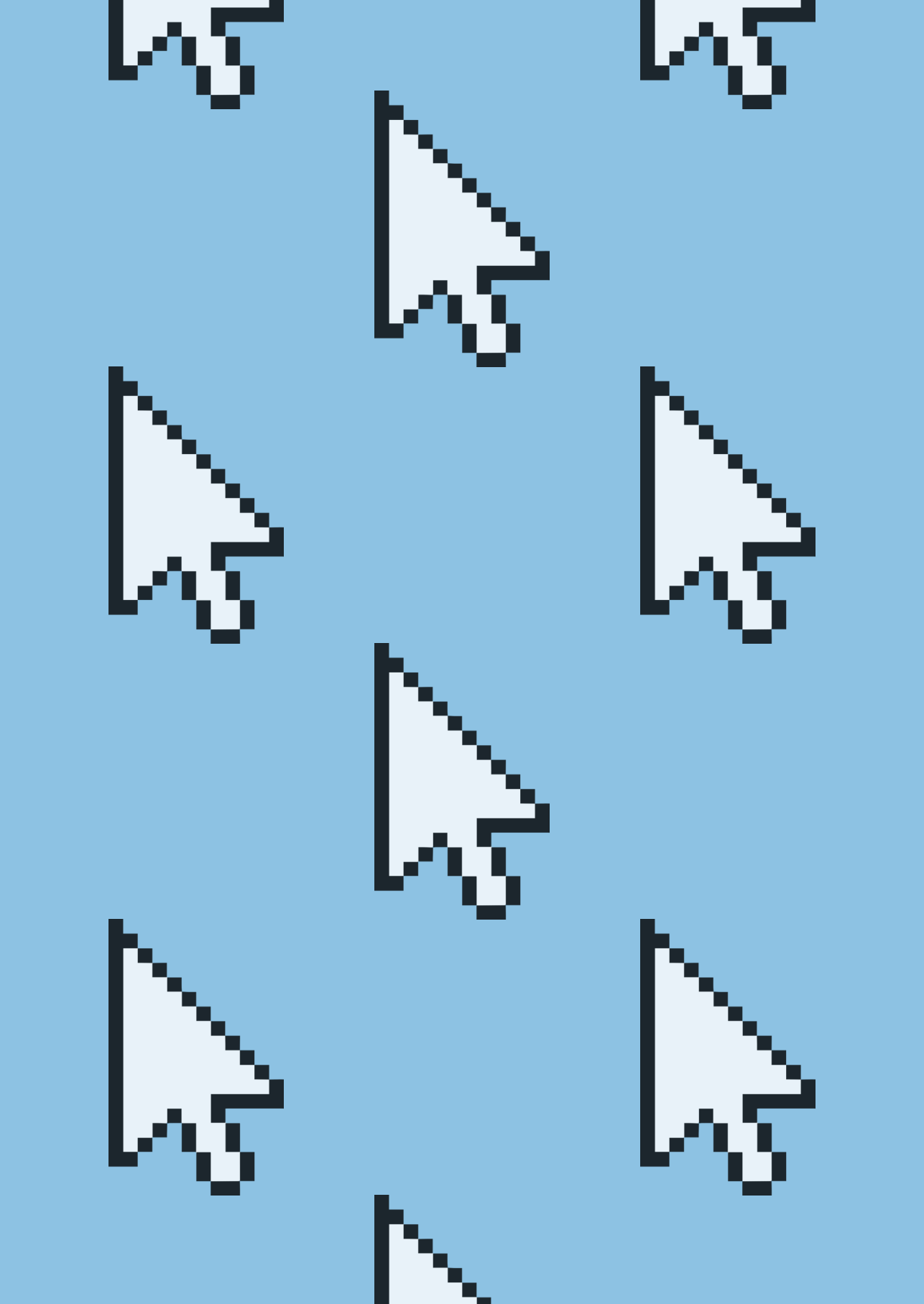
Uit dit onderzoek is gebleken dat digitale technologie, zoals in het geval van een digitale leeromgeving, ondersteuning kan bieden aan zowel leerlingen als docenten. In dit onderzoek werden echter geen significante verbeteringen gevonden wat betreft de leesprestaties of motivatie van leerlingen. Sterker nog, in alle groepen daalden de gemiddelde prestaties. Wel is er een hoopvolle bevinding gedaan met betrekking tot leerlingen die een ondergemiddeld leesniveau hebben: hun tekstbegrip daalde niet significant gedurende de interventie, terwijl dat bij de gemiddelde en bovengemiddelde lezers wel het geval was. Bovendien bleek dat het aanbieden van hints leidde tot een beter bewustzijn van probleemoplossende leesstrategieën, ongeacht het gebruik van deze hints. Indien leerlingen de hints daadwerkelijk gebruikten, was hun tekstbegrip significant beter ten opzichte van de leerlingen die geen enkele hint hadden aangeklikt. Desalniettemin tonen de resultaten van het huidige proefschrift aan dat de effectiviteit van het gebruik van de digitale leeromgeving grotendeels afhankelijk is van de betrokkenheid van leerlingen en van de implementatie door docenten. Leerlingen die betrokken zijn bij het werken in de digitale leeromgeving vertonen betere leesmotivatie en prestaties in vergelijking met leerlingen die weinig tijd besteden aan een leestaak, de beschikbare ondersteuning negeren en de opdrachten niet serieus nemen. Het is daarom belangrijk dat docenten de betrokkenheid van leerlingen stimuleren bij het lezen van teksten, bijvoorbeeld door in te zetten op klassikale gesprekken over het verklaren van historische gebeurtenissen, het stellen van goede historische onderzoeksvragen, of de mogelijke verschillen in standpunten van auteurs.

Daarnaast is het essentieel dat docenten die positief staan tegenover het gebruik van technologie in de klas, daartoe in de gelegenheid worden gesteld. Ook dienen zij zich bewust te zijn van de contextfactoren die de implementatie daarvan kunnen belemmeren. Over het algemeen hebben de deelnemende docenten aan dit onderzoeksproject zich positief uitgelaten over het gebruik van de digitale leeromgeving voor het lezen van informatieve teksten in hun geschiedenislessen. Het is echter belangrijk om op te merken dat de toepassing van digitale technologie voor zowel docenten als leerlingen voldoende tijd en ruimte vereist. De snelle ontwikkelingen op het gebied van onderwijstechnologie bieden docenten een overvloed aan mogelijkheden om hun instructie op te baseren. In de huidige context van hoge werkdruk, stakingen en alarmerende burn-outpercentages in het Nederlandse voortgezet onderwijs moet voldoende aandacht worden besteed aan de

(financiële) hulpmiddelen die een succesvolle integratie van digitale technologie in de klas mogelijk kunnen maken.

Conclusie

De resultaten uit dit proefschrift laten zien dat het gebruik van een digitale leeromgeving, met daarbinnen enerzijds cognitieve, metacognitieve en motiverende hints voor leerlingen en anderzijds gevisualiseerde data-output voor docenten, een bruikbare bijdrage kan leveren aan het huidige geschiedenisonderwijs als het gaat om het begrijpend lezen van informatieve teksten. Deze bevinding is relevant voor de onderwijspraktijk van veel geschiedenisdocenten, aangezien tekstbegrip voor elke leerling een onmisbare vaardigheid is voor het interpreteren en begrijpen van het verleden. Binnen de Nederlandse onderwijscontext worden momenteel bestaande onderwijscurricula herzien om zodoende rekening te kunnen houden met de kennis en vaardigheden die passen bij onze moderne, 21e-eeuwse samenleving. Deze dissertatie toont aan dat de rol van begrijpend lezen en de stimulering daarvan een cruciaal onderdeel hiervan zouden moeten uitmaken en dat dit relevant is voor het vak geschiedenis in de eerste jaren van het voortgezet onderwijs.



References

A

Aarnoutse, C. A. J. (1987). *Synoniementest, tegenstellingentest, voegwoordentest, hoofdgedachtetest. Tests voor begrijpend lezen bestemd voor groep 7 en 8 van het basisonderwijs. Handleiding en testmateriaal* [Synonyms test, antonyms test, conjunctions test, main ideas test. Tests for reading comprehension for the upper elementary grades. Manual and test equipment]. Nijmegen, The Netherlands: Berkhout.

Afflerbach, P., Pearson, P., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher*, *61*(5), 364–373. https://doi.org/10.1598/RT.61.5.1

Afflerbach, P., & VanSledright, B. (2001). Hath! Doth! What? Middle graders reading innovative history text. *Journal of Adolescent & Adult Literacy*, *44*(8), 696–707. https://www.jstor.org/stable/40018742

Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help? In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Proceedings of the 5th international conference on intelligent tutoring systems, ITS 2000* (pp. 292–303). Berlin: Springer Verlag. https://doi.org/10.1007/3-540-45108-0_33

Aleven, V. A. W. M. M., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, *26*(2), 147–179. https://doi.org/10.1207/s15516709cog2602_1

Alexander, P. A. (1998). The nature of disciplinary and domain learning: The knowledge, interest, and strategic dimensions of learning from subject matter text. In C. R. Hynd (Ed.), *Learning from texts across conceptual domains* (pp. 263–287). New York, NY: Routledge.

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, *32*(8), 10–14. https://doi.org/10.3102/0013189X032008010

Alexander, P. A. (2005). The path to competence: A lifespan developmental perspective on reading. *Journal of Literacy Research*, *37*(4), 413–436. https://doi.org/10.1207/s15548430jlr3704_1

Alexander, P. A., & Kulikowich, J. M. (1991). Domain knowledge and analogic reasoning ability as predictors of expository text comprehension. *Journal of Reading Behavior*, *23*(2), 165–190. https://doi.org/10.1080/10862969109547735

Alexander, P. A., & The Disciplined Reading and Learning Research Laboratory. (2012). Reading into the future: Competence for the 21st century. *Educational Psychologist*, *47*(4), 259–280. https://doi.org/10.1080/00461520.2012.722511

Anmarkrud, Ø., & Bråten, I. (2009). Motivation for reading comprehension. *Learning and Individual Differences*, *19*(2), 252–256. https://doi.org/10.1016/j.lindif.2008.09.002

Artelt, C., Schiefele, U., & Schneider, W. (2001). Predictors of reading literacy. *European Journal of Psychology of Education*, *16*(3), 363–383. https://doi.org/10.1007/BF03173188

Askell-Williams, H., Lawson, M. J., & Skrzypiec, G. (2012). Scaffolding cognitive and metacognitive strategy instruction in regular class lessons. *Instructional Science*, *40*(2), 413–443. https://doi.org/10.1007/s11251-011-9182-5

Azevedo, R. (2005). Computer environments as metacognitive tools for enhancing learning. *Educational Psychologist*, *40*(4), 193–197. https://doi.org/10.1207/s15326985ep4004_1

Azevedo, R. (2007). Understanding the complex nature of self-regulatory processes in learning with computer-based learning environments: An introduction. *Metacognition and Learning*, *2*(2-3), 57–65. https://doi.org/10.1007/s11409-007-9018-5

Azevedo, R. (2009). Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: A discussion. *Metacognition and Learning*, *4*(1), 87–95. https://doi.org/10.1007/s11409-009-9035-7

Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, *50*(1), 84–94. https://doi.org/10.1080/00461520.2015.1004069

Azevedo, R., Cromley, J. G., & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology*, *29*(3), 344–370. https://doi.org/10.1016/j.cedpsych.2003.09.002

Azevedo, R., & Gašević, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior*, *96*, 207–210. https://doi.org/10.1016/j.chb.2019.03.025

Azevedo, R., Harley, J., Trevors, G., Duffy, M., Feyzi-Behnagh, R., Bouchet, F., & Landis, R. (2013). Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 427–449). New York, NY: Springer. https://doi.org/10.1007/978-1-4419-5546-3_28

Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development*, *56*(1), 45–72. https://doi.org/10.1007/s11423-007-9067-0

B

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122–147. https://doi.org/10.1037/0003-066X.37.2.122

Bannert, M., Molenaar, I., Azevedo, R., Järvelä, S., & Gašević, D. (2017). Relevance of learning analytics to measure and support students' learning in adaptive educational technologies. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 568–569). New York, NY: ACM. https://doi.org/10.1145/3027385.3029463

Barab, S. A., Bowdish, B. E., & Lawless, K. A. (1997). Hypermedia navigation: Profiles of hypermedia users. *Educational Technology Research and Development*, *45*(3), 23–41. https://doi.org/10.1007/BF02299727

Barnes, M. A. (2015). What do models of reading comprehension and its development have to contribute to a science of comprehension instruction and assessment for adolescents? In K. L. Santi & D. K. Reed (Eds.), *Improving reading comprehension of middle and high school students* (pp. 1–18). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-14735-2_1

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction*, *17*(5), 564–577. https://doi.org/10.1016/j.learninstruc.2007.09.007

Berthold, K., Röder, H., Knörzer, D., Kessler, W., & Renkl, A. (2011). The double-edged effects of explanation prompts. *Computers in Human Behavior*, *27*(1), 69–75. https://doi.org/10.1016/j.chb.2010.05.025

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction*, *7*(2), 161–186. https://doi.org/10.1016/S0959-4752(96)00015-1

Boelens, R., De Wever, B., & Voet, M. (2017). Four key challenges to the design of blended learning: A systematic literature review. *Educational Research Review*, *22*, 1–18. https://doi.org/10.1016/j.edurev.2017.06.001

Bouchet, F., Harley, J. M., Trevors, G. J., & Azevedo, R. (2013). Clustering and profiling students according to their interactions with an intelligent tutoring system fostering self-regulated learning. *Journal of Educational Data Mining*, *5*(1), 104–146. Retrieved on April 9th, 2019 from https://jedm.educationaldatamining.org/index.php/JEDM/article/view/32

C

Campbell, C., & Levin, B. (2009). Using data to support educational improvement. *Educational Assessment, Evaluation and Accountability*, *21*, 47–65. https://doi.org/10.1007/s11092-008-9063-x

Chambers Cantrell, S., Almasi, J. F., Carter, J. C., & Rintamaa, M. (2013). Reading intervention in middle and high schools: Implementation fidelity, teacher efficacy, and student achievement. *Reading Psychology*, *34*(1), 26–58. https://doi.org/10.1080/02702711.2011.577695

Cheung, A. C., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, *7*(3), 198–215. https://doi.org/10.1016/j.edurev.2012.05.002

Cito. (2012). *Leesmotivatie, leesgedrag en leesvaardigheid van Nederlandse 15-jarigen. Aanvullende analyses op basis van PISA-2009* [Reading motivation, reading behaviour, and reading skills of Dutch 15-year-olds. Additional analyses based on PISA-2009]. Arnhem, the Netherlands: Cito.

Coddington, C. S. (2009). A*dolescent motivations for school reading questionnaire (AMSR)*. Retrieved on March 30th, 2017 from http://www.cori.umd.edu/measures/AMSR.pdf

Cohen, J. (1960). A coefficient of agreement for nominal scales. E*ducational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

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

College for Exams. (2014). *Geschiedenis vwo: Syllabus centraal examen 2016 op basis van domein A en B van het examenprogramma (dus geen tijdelijke afwijking meer)* [Dutch central state exam syllabus, 2016]. Retrieved from https://www.examenblad.nl/examenstof/syllabus-2016-geschiedenis-vwo/2016/vwo/f=/geschiedenis_vwo_def_versie_2016_aanpassing_Verona.pdf

College voor Toetsen en Examens, & Cito (2011). *ToetsSpecial: Normering centrale examens voortgezet onderwijs* [TestSpecial: Standardisation of central secondary education exams]. Retrieved from http://www.toetsspecials.nl/html/normering/equivalering.shtm

Collins, I. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis. With applications in the social, behavioral, and health sciences.* Hoboken, NJ: John Wiley & Sons.

Creswell, J. W. (2013). *Qualitative inquiry & research design. Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: SAGE.

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, *99*(2), 311–325. https://doi.org/10.1037/0022-0663.99.2.311

D

de Boer, S. E., Hagenbeek, A., de Waal, M., Weening, R. W., & Admiraal, W. F. (2013). *Meer motivatie door talentontwikkeling bij leerlingen in de bovenbouw van het VWO* [More motivation through talent development among students in upper secondary education]. Harderwijk, the Netherlands: RSG Slingerbos | Levant.

de Jager, B., Reezigt, G. J., & Creemers, B. P. M. (2002). The effects of teacher training on new instructional behaviour in reading comprehension. *Teaching and Teacher Education*, *18*(7), 831–842. https://doi.org/10.1016/S0742-051X(02)00046-X

de Kock, W. D. (2016). *The effectiveness of hints during computer supported word problem solving* (Doctoral dissertation, University of Groningen, the Netherlands). Retrieved from http://hdl.handle.net/11370/d7f929d9-810c-45e2-81db-a83c7d9d989e

De La Paz, S., & Felton, M. K. (2010). Reading and writing from multiple source documents in history: Effects of strategy instruction with low to average high school writers. *Contemporary Educational Psychology, 35*(3), 174–192. https://doi.org/10.1016/j.cedpsych.2010.03.001

De Naeghel, J., Van Keer, H., Vansteenkiste, M., & Rosseel, Y. (2012). The relation between elementary students' recreational and academic reading motivation, reading frequency, engagement, and comprehension: A self-determination theory perspective. *Journal of Educational Psychology*, *104*(4), 1006–1021. https://doi.org/10.1037/a0027800

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Deunk, M. I., Smale-Jacobse, A. E., de Boer, H., Doolaard, S., & Bosker, R. J. (2018). Effective differentiation practices: A systematic review and meta-analysis of studies on the cognitive effects of differentiation practices in primary education. *Educational Research Review*, *24*, 31–54. https://doi.org/10.1016/j.edurev.2018.02.002

Devolder, A., van Braak, J., & Tondeur, J. (2012). Supporting self-regulated learning in computer-based learning environments: Systematic review of effects of scaffolding in the domain of science education. *Journal of Computer Assisted Learning*, *28*(6), 557–573. https://doi.org/10.1111/j.1365-2729.2011.00476.x

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, *3*(3), 231–264. https://doi.org/10.1007/s11409-008-9029-x

Dignath-van Ewijk, C., Dickhäuser, O., & Büttner, G. (2013). Assessing how teachers enhance self-regulated learning: A multiperspective approach. *Journal of Cognitive Education and Psychology*, *12*(3), 338–358. https://doi.org/10.1891/1945-8959.12.3.338

Dignath-van Ewijk, C., & van der Werf, G. (2012). What teachers think about self-regulated learning: Investigating teacher beliefs and teacher behavior of enhancing students' self-regulation. *Education Research International*, *2012*(741713), 1–10. https://doi.org/10.1155/2012/741713

Dingyloudi, F., & Strijbos, J. W. (2018). Mixed methods research as a pragmatic toolkit: Understanding versus fixing complexity in the learning sciences. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (p. 444–454). New York, NY: Routledge. https://doi.org/10.4324/9781315617572

Donker, A. S., de Boer, H., Kostons, D., Dignath-van Ewijk, C. C., & van der Werf, M. P. C. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review*, *11*, 1–26. https://doi.org/10.1016/j.edurev.2013.11.002

Drenth, P. J. D., & Sijtsma, K. (Eds.) (2005). *Testtheorie. Inleiding in de theorie van de psychologische test en zijn toepassingen* [Test theory. An introduction to the theory of the psychological test and its applications]. Houten, the Netherlands: Bohn Stafleu van Loghum.

Duffy, G. G. (1982). Fighting off the alligators: What research in real classrooms has to say about reading instruction. *Journal of Reading Behavior*, *14*(4), 357–373. https://doi.org/10.1080/10862968209547464

Duffy, G. G., Roehler, L. R., Sivan, E., Rackliffe, G., Book, C., Meloth, M. S., ..., Bassiri, D. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, *22*(3), 347–368. https://doi.org/10.2307/747973

Duhaylongsod, L., Snow, C. E., Selman, R. L., & Donovan, M. S. (2015). Toward disciplinary literacy: Dilemmas and challenges in designing history curriculum to support middle school students. *Harvard Educational Review*, *85*(4), 587–608. https://doi.org/10.17763/0017-8055.85.4.587

E

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3–25). Bristol, UK: Multilingual Matters.

F

Faggella-Luby, M. N., Graner, P. S., Deshler, D. D., & Drew, S. V. (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

Feskens, R., Kuhlemeier, H., & Limpens, G. (2016). *Resultaten PISA-2015. Praktische kennis en vaardigheden van 15-jarigen* [PISA results 2015: Practical knowledge and skills of 15-year-olds]. Arnhem, the Netherlands: Cito. Retrieved from https://www.cito.nl/-/media/files/kennis-en-innovatie-onderzoek/pisa/cito-pisa-rapport-2015.pdf?la=nl-NL

Fisher, D., & Frey, N. (2008). Student and teacher perspectives on the usefulness of content literacy strategies. *Literacy Research and Instruction*, *47*(4), 246–263. https://doi.org/10.1080/19388070802300330

Fisher, D., Frey, N., & Lapp, D. (2011). Coaching middle-level teachers to think aloud improves comprehension instruction and student reading achievement. *The Teacher Educator*, *46*(3), 231–243. https://doi.org/10.1080/08878730.2011.580043

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Flunger, B., Trautwein, U., Nagengast, B., Lüdtke, O, Niggli, A., & Schnyder, I. (2015). The Janus-faced nature of time spent on homework: Using latent profile analyses to predict academic achievement over a school year. *Learning and Instruction*, *39*, 97–106. https://doi.org/10.1016/j.learninstruc.2015.05.008

Flunger, B., Trautwein, U., Nagengast, B., Lüdtke, O, Niggli, A., & Schnyder, I. (2017). A person-centered approach to homework behavior: Students' characteristics predict their homework learning type. *Contemporary Educational Psychology*, *48*, 1–15. https://doi.org/10.1016/j.cedpsych.2016.07.002

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Fry, S. W., & Gosky, R. (2007). Supporting social studies reading comprehension with an electronic pop-up dictionary. *Journal of Research on Technology in Education*, *40*(2), 127–139. https://doi.org/10.1080/15391523.2007.10782501

G

García Duncan, T., & McKeachie, W. J. (2005). The making of the motivated strategies for learning questionnaire. *Educational Psychologist*, *40*(2), 117–128. https://doi.org/10.1207/s15326985ep4002_6

Girard, B., & McArthur Harris, L. (2012). Striving for disciplinary literacy instruction: Cognitive tools in a world history course. *Theory & Research in Social Education*, *40*(3), 230–259. https://doi.org/10.1080/00933104.2012.705183

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C. … & Project READI. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, *51*(2), 219–246. https://doi.org/10.1080/00461520.2016.1168741

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Greene, J. A., Bolick, C. M., & Robertson, J. (2010). Fostering historical knowledge and thinking skills using hypermedia learning environments: The role of self-regulated learning. *Computers & Education*, *54*(1), 230–243. https://doi.org/10.1016/j.compedu.2009.08.006

Greenleaf, C. L., Schoenbach, R., Cziko, C., & Mueller, F. L. (2001). Apprenticing adolescent readers to academic literacy. *Harvard Educational Review*, *71*(1), 79–127. https://doi.org/10.17763/haer.71.1.q811712577334038

Gubbels, J., Netten, A., & Verhoeven, L. (2017). *Vijftien jaar leesprestaties in Nederland. PIRLS-2016* [Fifteen years of reading performance in the Netherlands. PIRLS-2016]. Nijmegen, the Netherlands: Radboud Universiteit, Expertisecentrum Nederlands.

Gubbels, J., van Langen, A. M. L., Maassen, N. A. M., & Meelissen, M. R. M. (2019). *Resultaten PISA-2018 in vogelvlucht* [The results of PISA-2018 in a nutshell]. Enschede, the Netherlands: Universiteit Twente. https://doi.org/10.3990/1.9789036549226

Guthrie, J. T., & Davis, M. H. (2003). Motivating struggling readers in middle school through an engagement model of classroom practice. *Reading & Writing Quarterly*, *19*(1), 59–85. https://doi.org/10.1080/10573560308203

Guthrie, J. T., & Klauda, S. L. (2016). Engagement and motivational processes in reading. In P. Afflerbach (Ed.), *Handbook of individual differences in reading: Reader, text, and context* (pp. 41–53). New York: Routledge.

Guthrie, J. T., Klauda, S. L., & Ho, A. N. (2013). Modeling the relationships among reading instruction, motivation, engagement, and achievement for adolescents. *Reading Research Quarterly*, *48*(1), 9–26. https://doi.org/10.1002/rrq.035

Guthrie, J. T., McRae, A., & Klauda, S. L. (2007). Contributions of Concept-Oriented Reading Instruction to knowledge about interventions for motivations in reading. *Educational Psychologist*, *42*(4), 237–250. https://doi.org/10.1080/00461520701621087

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (3rd ed., pp. 403–422). New York: Longman.

Guthrie, J. T., & Wigfield, A. (2017). Literacy engagement and motivation: Rationale, research, teaching, and assessment. In D. Lapp & D. Fisher (Eds.), *Handbook of research on teaching the English language arts* (4th ed., pp. 57–84). New York, NY: Routledge. https://doi.org/10.4324/9781315650555

Guthrie, J. T., Wigfield, A., Cambria, J., Coddington, C. S., Klauda, S. L., & Morrison, D. A. (2009). *Motivations for reading information books among adolescent students.* Retrieved on March 30th, 2017 from http://cori.umd.edu/research-projects/real/research/Motivation_ReportFINAL.pdf

Guthrie, J. T., Wigfield, A., & Klauda, S. L. (2012). *Adolescents' engagement in academic literacy.* Retrieved on April 23rd, 2019 from http://www.corilearning.com/research-publications/2012_adolescents_engagement_ebook.pdf

Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (1999). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading*, *3*(3), 231–256. https://doi.org/10.1207/s1532799xssr0303_3

H

Hacquebord, H. I., Linthorst, T. R., Stellingwerf, B. P., & de Zeeuw, M. (2004). *Voortgezet taalvaardig. Een onderzoek naar tekstbegrip en woordkennis en naar de taalproblemen en taalbehoeften van brugklasleerlingen in het schooljaar 2002–2003* [A study into text comprehension and word knowledge and into the language problems and language needs of seventh-grade pupils in the school year 2002–2003]. Groningen, the Netherlands: Expertisecentrum taal, onderwijs en communicatie.

Hall, L. A. (2005). Teachers and content area reading: Attitudes, beliefs and change. *Teaching and Teacher Education*, *21*(4), 403–414. https://doi.org/10.1016/j.tate.2005.01.009

Harmsen, R., Helms-Lorenz, M., Maulana, R., & van Veen, K. (2018). The relationship between beginning teachers' stress causes, stress responses, teaching behaviour and attrition. *Teachers and Teaching: Theory and Practice*, *24*(6), 626–643. https://doi.org/10.1080/13540602.2018.1465404

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*(1), 77–89. https://doi.org/10.1080/19312450709336664

Heller, R., & Greenleaf, C. L. (2007). *Literacy instruction in the content areas: Getting to the core of middle and high school improvement*. Washington, D.C.: Alliance for Excellent Education.

Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., & Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences*, *66*, 4–15. https://doi.org/10.1016/j.lindif.2017.11.001

Hill, H. C., Beisiegel, M., & Jacob, R. (2013). Professional development research: Consensus, crossroads and challenges. *Educational Researcher*, *42*(9), 476–487. https://doi.org/10.3102/0013189X13512674

Hoogland, I., Schildkamp, K., van der Kleij, F., Heitink, M., Kippers, W., Veldkamp, B., & Dijkstra, A. M. (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. *Teaching and Teacher Education*, *60*, 377–386. https://doi.org/10.1016/j.tate.2016.07.012

Huijgen, T., van de Grift, W., van Boxtel, C., & Holthuis, P. (2018). Promoting historical contextualization: The development and testing of a pedagogy. *Journal of Curriculum Studies*, *50*(3), 410–434. https://doi.org/10.1080/00220272.2018.1435724

Hutchison, A. C., & Colwell, J. (2014). The potential of digital technologies to support literacy instruction relevant to the Common Core State Standards. *Journal of Adolescent & Adult Literacy*, *58*(2), 147–156. https://doi.org/10.1002/jaal.335

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

J

Jetton, T. L., & Lee, R. (2012). Learning from text: Adolescent literacy from the past decade. In T. L. Jetton & C. Shanahan (Eds.), *Adolescent literacy in the academic disciplines: General principles and practical strategies* (pp. 1–33). New York: Guilford Press.

Johnson-Glenberg, M. C. (2005). Web-based training of metacognitive strategies for text comprehension: Focus on poor comprehenders. *Reading and Writing*, *18*(7–9), 755–786. https://doi.org/10.1007/s11145-005-0956-5

Jülicher, T. (2018). Education 2.0: Learning analytics, educational data mining and co.. In T. Hoeren & B. Kolany-Raiser (Eds.), *Big data in context. Legal, social and technological insights* (pp. 47–54). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-62461-7_6

K

Kamil, M. L., Afflerbach, P., Pearson, P. D., & Moje, E. B. (2011). Reading research in a changing era: An introduction to the handbook of reading research. In M. L. Kamil, P. Afflerbach, P. D. Pearson, & E. B. Moje (Eds.), *Handbook of reading research* (Vol. IV, pp. xiii–xxvi). New York, NY: Routledge.

Kendeou, P., van den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research & Practice*, *29*(1), 10–16. https://doi.org/10.1111/ldrp.12025

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 209–226). Malden, MA: Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch12

Kintsch, W., & van Dijk, T. A. (1978). Towards a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education*, *75*, 199–213. https://doi.org/10.1016/j.tate.2018.06.015

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction*, *12*(1), 1–10. https://doi.org/10.1016/S0959-4752(01)00014-7

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Kordes, J., Bolsinova, M., Limpens, G., & Stolwijk, R. (2013). *Resultaten PISA-2012. Praktische kennis en vaardigheden van 15-jarigen* [PISA results 2012: Practical knowledge and skills of 15-year-olds]. Arnhem, the Netherlands: Cito. Retrieved from https://www.cito.nl/-/media/files/kennis-en-innovatie-onderzoek/pisa/cito_pisa_2012.pdf?la=nl-NL

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology.* Thousand Oaks, CA: Sage.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160. https://doi.org/10.1007/BF02288391

L

Lajoie, S. P. (2005). Extending the scaffolding metaphor. *Instructional Science*, *33*, 541–557. https://doi.org/10.1007/s11251-005-1279-2

Lan, Y.-C., Lo, Y.-L., & Hsu, Y.-S. (2014). The effects of meta-cognitive instruction on students' reading comprehension in computerized reading contexts: A quantitative meta-analysis. *Journal of Educational Technology & Society*, *17*(4), 186–202. https://www.jstor.org/stable/jeductechsoci.17.4.186

Lai, M. K., & Schildkamp, K. (2013). Data-based decision making: An overview. In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in education: Challenges and opportunities* (pp. 9–21). Dordrecht, the Netherlands: Springer. https://doi.org/10.1007/978-94-007-4816-3_2

Lai, M. K., Wilson, A., McNaughton, S., & Hsiao, S. (2014). Improving achievement in secondary schools: Impact of a literacy project on reading comprehension and secondary school qualifications. *Reading Research Quarterly*, *49*(3), 305–334. https://doi.org/10.1002/rrq.73

Learned, J. E. (2018). Doing history: A study of disciplinary literacy and readers labeled as struggling. *Journal of Literacy Research*, *50*(2), 190–216. https://doi.org/10.1177/1086296X17746446

Lee, M. J. W., Kirschner, P. A., & Kester, L. (2016). Learning analytics in massively multi-user virtual environments and courses. *Journal of Computer Assisted Learning*, *32*(3), 187–189. https://doi.org/10.1111/jcal.12139

Lee, P., & Ashby, R. (2000). Progression in historical understanding among students aged 7–14. In P. N. Stearns, P. Seixas, & S. Wineburg (Eds.), *Knowing, teaching, and learning history: National and international perspectives* (pp. 199–222). New York, NY: New York University Press.

Linthorst, R., & de Glopper, K. (2015). *De didactiek van begrijpend lezen in het voortgezet onderwijs: Lesobservaties bij Nederlands en zaakvakken* [Didactics of reading comprehension in secondary education: Lesson observations in Dutch language and social studies classrooms]. Pedagogische Studiën, 92(2), 150–166. Retrieved on January 30th, 2019 from http://pedagogischestudien.nl/download?type=document&identifier=616665

Livingstone, S. (2012). Critical reflections on the benefits of ICT in education. *Oxford Review of Education*, *38*(1), 9–24. https://doi.org/10.1080/03054985.2011.577938

Lynch, L., Fawcett, A. J., & Nicolson, R. I. (2000). Computer-assisted reading intervention in a secondary school: An evaluation study. *British Journal of Educational Technology*, *31*(4), 333–348. https://doi.org/10.1111/1467-8535.00166

Lysenko, L. V., & Abrami, P. C. (2014). Promoting reading comprehension with the use of technology. *Computers & Education*, *75*, 162–172. https://doi.org/10.1016/j.compedu.2014.01.010

M

Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, *60*, 366–376. https://doi.org/10.1016/j.tate.2016.07.011

Mandinach, E. B., & Jimerson, J. B. (2016). Teachers learning how to use data: A synthesis of the issues and what is known. *Teaching and Teacher Education*, *60*, 452–457. https://doi.org/10.1016/j.tate.2016.07.009

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record*, *114*(11), 1–48. Retrieved on May 24th, 2019 from https://www.rdc.udel.edu/wp-content/uploads/2015/04/Interventions-Promoting-Educators.pdf

Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling*, *16*(2), 191–225. https://doi.org/10.1080/10705510902751010

Mason, L. H. (2013). Teaching students who struggle with learning to think before, while, and after reading: Effects of self-regulated strategy development instruction. *Reading & Writing Quarterly*, *29*(2), 124–144. https://doi.org/10.1080/10573569.2013.758561

Mastropieri, M. A., Scruggs, T. E., & Graetz, J. E. (2003). Reading comprehension instruction for secondary students: Challenges for struggling students and teachers. *Learning Disability Quarterly*, *26*(2), 103–116. https://doi.org/10.2307/1593593

Masyn, K. E. (2013). Latent class analysis and finite mixture modelling. In T. D. Little (Ed.), *Oxford library of psychology. The Oxford handbook of quantitative methods: Statistical analysis* (pp. 551–611). New York, NY: Oxford University Press.

McKeown, M. G., & Beck, I. L. (2009). The role of metacognition in understanding and supporting reading comprehension. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 7–25). New York: Routledge.

McKeown, M. G., Beck, I. L., & Blake, R. G. K. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly*, *44*(3), 218–253. https://doi.org/10.1598/RRQ.44.3.1

McNamara, D. S. (Ed.) (2007). *Reading comprehension strategies: Theories, interventions, and technologies.* New York: Erlbaum.

Meijer, P. C., Verloop, N., & Beijaard, D. (2001). Similarities and differences in teachers' practical knowledge about teaching reading comprehension. *The Journal of Educational Research*, *94*(3), 171–184. https://doi.org/10.1080/00220670109599914

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.

Moeken, N., Kuiken, F., & Welie, C. (2016). Samen werken aan een betere leesvaardigheid: Een vakoverstijgende aanpak voor samenwerkend lezen in het voortgezet onderwijs [Working together for better reading skills: A cross-curricular approach for collaborative reading in secondary education]. *Levende Talen Magazine*, *103*(8), 18–23. Retrieved from http://lt-tijdschriften.nl/ojs/index.php/ltm/article/view/1638/1246

Moje, E. B. (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. B. (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

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, *94*(2), 249–259. https://doi.org/10.1037/0022-0663.94.2.249

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*(2), 267–296. https://doi.org/10.1037/A0021890

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. https://doi.org/10.1111/j.1467-873X.2011.00547.x

Monte-Sano, C., De La Paz, S., & Felton, M. (2014). Implementing a disciplinary-literacy curriculum for US history: Learning from expert middle school teachers in diverse classrooms. *Journal of Curriculum Studies*, *46*(4), 540–575. https://doi.org/10.1080/00220272.2014.904444

Moran, J., Ferdig, R. E., Pearson, P. D., Wardrop, J., & Blomeyer, R. L., Jr. (2008). Technology and reading performance in the middle-school grades: A meta-analysis with recommendations for policy and practice. *Journal of Literacy Research*, *40*(1), 6–58. https://doi.org/10.1080/10862960802070483

Morgan, P. L., & Fuchs, D. (2007). Is there a bidirectional relationship between children's reading skills and reading motivation? *Exceptional Children*, *73*(2), 165–183. https://doi.org/10.1177/001440290707300203

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 International Results in Reading.* Boston, United States: TIMSS & PIRLS International Study Center. Retrieved from http://timssandpirls.bc.edu/pirls2016/international-results/

N

Narciss, S., Proske, A., & Koerndle, H. (2007). Promoting self-regulated learning in web-based learning environments. *Computers in Human Behavior*, *23*(3), 1126–1144. https://doi.org/10.1016/j.chb.2006.10.006

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development.

Ness, M. (2016). Reading comprehension strategies in secondary content area classrooms: Teacher use of and attitudes towards reading comprehension instruction. *Reading Horizons*, *55*(1), 59–85. Retrieved on February 4th, 2019 from http://scholarworks.wmich.edu/reading_horizons/vol55/iss1/5

Nokes, J. D. (2011). Recognizing and addressing the barriers to adolescents' "reading like historians". *The History Teacher*, *44*(3), 379–404. https://www.jstor.org/stable/41303991

Nokes, J. D., Dole, J. A., & Hacker, D. J. (2007). Teaching high school students to use heuristics while reading historical texts. *Journal of Educational Psychology*, *99*(3), 492–504. https://doi.org/10.1037/0022-0663.99.3.492

Noroozi, O., Alikhani, I., Järvelä, S., Kirschner, P. A., Juuso, I., & Seppänen, T. (2019). Multimodal data to design visual learning analytics for understanding regulation of learning. *Computers in Human Behavior*, *100*, 298–304. https://doi.org/10.1016/j.chb.2018.12.019

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

O

O'Brien, D. G., Stewart, R. A., & Moje, E. B. (1995). Why content literacy is difficult to infuse into the secondary school: Complexities of curriculum, pedagogy, and school culture. *Reading Research Quarterly*, *30*(3), 442–463. https://www.jstor.org/stable/747625

Okkinga, M., van Steensel, R., van Gelderen, A. J. S., & Sleegers, P. J. C. (2018). Effects of reciprocal teaching on reading comprehension of low-achieving adolescents. The importance of specific teacher skills. *Journal of Research in Reading*, *41*(1), 20–41. https://doi.org/10.1111/1467-9817.12082

Okkinga, M., van Steensel, R., van Gelderen, A. J. S., van Schooten, E., Sleegers, P. J. C., & Arends, L. R. (2018). Effectiveness of reading-strategy interventions in whole classrooms: A meta-analysis. *Educational Psychology Review*, *30*(4), 1215–1239. https://doi.org/10.1007/s10648-018-9445-7

O'Neill, D. K., & Weiler, M. J. (2006). Cognitive tools for understanding history: What more do we need? *Journal of Educational Computing Research*, *35*(2), 181–197. https://doi.org/10.2190/H22P-7718-81G5-0723

Opdenakker, M.-C., Maulana, R., & den Brok, P. (2012). Teacher-student interpersonal relationships and academic motivation within one school year: Developmental changes and linkage. *School Effectiveness and School Improvement*, *23*(1), 95–119. https://doi.org/10.1080/09243453.2011.619198

Organization for Economic Cooperation and Development [OECD]. (2016a). *PISA 2015 results (Volume I): Excellence and equity in education.* PISA. Paris, France: OECD. https://doi.org/10.1787/9789264266490-en

Organization for Economic Cooperation and Development [OECD]. (2016b). *Netherlands 2016: Foundations for the future. Reviews of National Policies for Education*. Paris, France: OECD. https://doi.org/10.1787/9789264257658-en

Organization for Economic Cooperation and Development [OECD]. (2018). *PISA 2018 results (Volume I): What students know and can do*. PISA. Paris, France: OECD. https://doi.org/10.1787/5f07c754-en

P

Pajares, F. (2008). Motivational role of self-efficacy beliefs in self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning. Theory, research, and applications* (pp. 111–140). New York: Erlbaum. https://doi.org/10.4324/9780203831076

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*(2), 117–175. https://doi.org/10.1207/s1532690xci0102_1

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*(422), 1–28. https://doi.org/10.3389/fpsyg.2017.00422

Paris, S. G., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology*, *76*(6), 1239–1252. https://doi.org/10.1037/0022-0663.76.6.1239

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related concepts for learning, education and human activity. *Journal of the Learning Sciences*, *13*(3), 423–451. https://doi.org/10.1207/s15327809jls1303_6

Pereira, C., & Nicolaas, M. (2019). *Effectief onderwijs in begrijpend lezen. Acties voor beter leesbegrip en meer leesmotivatie.* [Effective reading comprehension education. Actions for better reading comprehension and more reading motivation]. Den Haag, the Netherlands: Taalunie. Retrieved on August 30th, 2019 from http://taalunieversum.org/sites/tuv/files/downloads/Actieplan_Effectief%20onderwijs%20in%20begrijpend%20lezen.pdf

Perfetti, C. A., Britt, M. A., & Georgi, M. C. (1995). *Text-based learning and reasoning: Studies in history*. Hillsdale, NJ: Erlbaum.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Malden, MA: Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch13

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In: M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 451–502). San Diego, CA: Academic Press.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Pintrich, P. R., Smith, D. A. F., García, T., & McKeachie, W. J. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ).* Ann Arbor, MI: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

Poitras, E., Lajoie, S., & Hong, Y-J. (2012). The design of technology-rich learning environments as metacognitive tools in history education. *Instructional Science*, *40*(6), 1033–1061. https://doi.org/10.1007/s11251-011-9194-1

Poortman, C. L., Schildkamp, K., & Lai, M. K. (2016). Professional development in data use: An international perspective on conditions, models, and effects. *Teaching and Teacher Education*, *60*, 363–365. https://doi.org/10.1016/j.tate.2016.07.029

Pressley, M., Harris, K. R., & Marks, M. B. (1992). But good strategy instructors are constructivists! *Educational Psychology Review*, *4*(1), 3–31. https://doi.org/10.1007/BF01322393

Puntambekar, S., & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, *40*(1), 1–12. https://doi.org/10.1207/s15326985ep4001_1

Puntambekar, S., & Stylianou, A. (2005). Designing navigation support in hypertext systems based on navigation patterns. *Instructional Science*, *33*(5–6), 451–481. https://doi.org/10.1007/s11251-005-1276-5

R

Raad voor Cultuur, & Onderwijsraad. (2019). *Lees! Een oproep tot een leesoffensief* [Read! A call for a reading offensive]. Den Haag, the Netherlands: Onderwijsraad. Retrieved from https://www.onderwijsraad.nl/upload/documents/publicaties/volledig/ONLINE-Lees-Een-oproep-tot-een-leesoffensief.pdf

Ramsay, C. M., Sperling, R. A., & Dornisch, M. M. (2010). A comparison of the effects of students' expository text comprehension strategies. *Instructional Science*, *38*(6), 551–570. https://doi.org/10.1007/s11251-008-9081-6

Rashid, T., & Asghar, H. M. (2016). Technology use, self-directed learning, student engagement and academic performance: Examining the interrelations. *Computers in Human Behavior*, *63*, 604–612. https://doi.org/10.1016/j.chb.2016.05.084

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

Reisman, A. (2012). Reading like a historian: A document-based history curriculum intervention in urban high schools. *Cognition and Instruction*, *30*(1), 86–112. https://doi.org/10.1080/07370008.2011.634081

Retelsdorf, J., Köller, O., & Möller, J. (2011). On the effects of motivation on reading performance growth in secondary school. *Learning and Instruction*, *21*(4), 550–559. https://doi.org/10.1016/j.learninstruc.2010.11.001

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Ritzema, E. S. (2015). *Professional development in data use: The effects of primary school teacher training on teaching practices and students' mathematical proficiency* (Doctoral dissertation, University of Groningen, the Netherlands). Retrieved on February 4th, 2019 from http://hdl.handle.net/11370/0a96bd4b-2aea-4821-9ea0-100dd673ca7f

Robinson, V. M. J., & Lai, M. K. (2006). *Practitioner research for educators: A guide to improving classrooms and schools.* Thousand Oaks, CA: Corwin Press.

Roll, I., Baker, R. J. S. d., Aleven, V., & Koedinger, K. R. (2014). On the benefits of seeking (and avoiding) help in online problem-solving environments. *Journal of the Learning Sciences*, *23*(4), 537–560. https://doi.org/10.1080/10508406.2014.883977

Rosenthal, R. (1976). *Experimenter effects in behavioral research*. New York, NY: Irvington.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

S

Schellings, G., & van Hout-Wolters, B. (2011). Measuring strategy use with self-report instruments: Theoretical and empirical considerations. *Metacognition and Learning*, *6*(2), 83–90. https://doi.org/10.1007/s11409-011-9081-9

Schiefele, U., & Löweke, S. (2018). The nature, development, and effects of elementary students' reading motivation profiles. *Reading Research Quarterly*, *53*(4), 405–421. https://doi.org/10.1002/rrq.201

Schiefele, U., Schaffner, E., Möller, J., & Wigfield, A. (2012). Dimensions of reading motivation and their relation to reading behavior and competence. *Reading Research Quarterly*, *47*(4), 427–463. https://doi.org/10.1002/RRQ.030

Schildkamp, K., Lai, M. K., & Earl, L. (2013). *Data-based decision making in education: Challenges and opportunities.* Dordrecht, the Netherlands: Springer. https://doi.org/10.1007/978-94-007-4816-3

Schraw, G. (2009). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 415–429). New York: Routledge.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, *19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033

Schunk, D. H., & Zimmerman, B. J. (Eds.) (2008). *Motivation and self-regulated learning. Theory, research, and applications.* New York: Erlbaum. https://doi.org/10.4324/9780203831076

Senechal, M. (2006). Testing the home literacy model: Parent involvement in kindergarten is differentially related to grade 4 reading comprehension, fluency, spelling, and reading for pleasure. *Scientific Studies of Reading*, *10*(1), 59–87. https://doi.org/10.1207/s1532799xssr1001_4

Seymour, J. R., & Osana, H. P. (2003). Reciprocal teaching procedures and principles: Two teachers' developing understanding. *Teaching and Teacher Education*, *19*(3), 325–344. https://doi.org/10.1016/S0742-051X(03)00018-0

Shanahan, C., Shanahan, T., & Misischia, C. (2011). Analysis of expert readers in three disciplines: History, mathematics, and chemistry. *Journal of Literacy Research*, *43*(4), 393–429. https://doi.org/10.1177/1086296X11424071

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

Sheard, J. (2010). Basics of statistical analysis of interactions data from web-based learning environments. In C. Romero, S. Ventura, M. Pechenizkiy & R. S. J. d. Baker (Eds.), *Handbook of Educational Data Mining* (pp. 27–42). Boca Raton, FL: CRC Press.

Siemens, G., & Baker, R. S. J. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252–254). New York, NY: ACM Publishing. https://doi.org/10.1145/2330601.2330661

Smale-Jacobse, A. E. (2013). *De effecten van docenttraining gericht op directe instructie van strategiegebruik bij begrijpend lezen* [Effects of teacher training aimed at direct instruction of strategy use during comprehensive reading]. Groningen, the Netherlands: GION. Retrieved on January 30th, 2019 from https://www.rug.nl/research/portal/files/14425915/Trainendocenten.pdf

Smale-Jacobse, A. E., & Timmerman, J. (2015). *Lezen met begrip in leeslessen en geschiedenislessen* [Reading comprehensively during reading lessons and history lessons]. Groningen, the Netherlands: GION. Retrieved on January 30th, 2019 from https://www.nro.nl/wp-content/uploads/2014/09/brochure-Lezen-met-begrip-in-leeslessen-en-geschiedenislessen-Smale-Jacobse.pdf

Souvignier, E., & Mokhlesgerami, J. (2006). Using self-regulation as a framework for implementing strategy instruction to foster reading comprehension. *Learning and Instruction*, *16*(1), 57–71. https://doi.org/10.1016/j.learninstruc.2005.12.006

Staman, L., Timmermans, A. C., & Visscher, A. J. (2017). Effects of a data-based decision making intervention on student achievement. *Studies in Educational Evaluation*, *55*, 58–67. https://doi.org/10.1016/j.stueduc.2017.07.002

Staman, L., Visscher, A., & Luyten, H. (2014). The effects of professional development on the attitudes, knowledge and skills for data-driven decision making. *Studies in Educational Evaluation*, *42*, 79–90. https://doi.org/10.1016/j.stueduc.2013.11.002

Stoel, G. L., van Drie, J. P., & van Boxtel, C. A. M. (2015). Teaching towards historical expertise. Developing a pedagogy for fostering causal reasoning in history. *Journal of Curriculum Studies*, *47*(1), 49–76. https://doi.org/10.1080/00220272.2014.968212

Strijbos, J. W., & Stahl, G. (2007). Methodological issues in developing a multi-dimensional coding procedure for small-group chat communication. *Learning and Instruction*, *17*(4), 394–404. https://doi.org/10.1016/j.learninstruc.2007.03.005

Swanson, E., Hairrell, A., Kent, S., Ciullo, S., Wanzek, J. A., & Vaughn, S. (2014). A synthesis and meta-analysis of reading interventions using social studies content for students with learning disabilities. *Journal of Learning Disabilities*, *47*(2), 178–195. https://doi.org/10.1177/0022219412451131

Swanson, E., Wanzek, J., Vaughn, S., Fall, A-M., Roberts, G., Hall, C., & Miller, V. L. (2016). Middle school reading comprehension and content learning intervention for below-average readers. *Reading & Writing Quarterly*, *33*(1), 37–53. https://doi.org/10.1080/10573569.2015.1072068

T

Taboada, A., Tonks, S. M., Wigfield, A., & Guthrie, J. T. (2009). Effects of motivational and cognitive variables on reading comprehension. *Journal of Reading and Writing*, *22*, 85–106. https://doi.org/10.1007/s11145-008-9133-y

Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling*, *20*(4), 640–657. https://doi.org/10.1080/10705511.2013.824781

ter Beek, M., Brummer, L., Donker, A. S., & Opdenakker, M.-C. J. L. (2018). Supporting secondary school students' reading comprehension in computer environments: A systematic review. *Journal of Computer Assisted Learning*, *34*(5), 557–566. https://doi.org/10.1111/jcal.12260

ter Beek, M., Spijkerboer, A. W., Brummer, L., & Opdenakker, M.-C. (2018). *Gemotiveerd, actief en zelfstandig lezen: Hoe een digitale leeromgeving zowel de leerling als de docent kan ondersteunen bij het begrijpend lezen van informatieve zaakvakteksten in het voortgezet onderwijs* [Motivated, active, and independent reading: How a digital learning environment can support both students and teachers in the context of comprehensive reading of expository content area texts in secondary education]. Groningen, the Netherlands: Rijksuniversiteit Groningen. Retrieved on June 3rd, 2019 from https://www.rug.nl/research/portal/files/777.2353/Gazelle_rapport_digitaal.pdf

ter Beek, M., Opdenakker, M.-C., Deunk, M. I., & Strijbos, J. W. (2019a; submitted). *Using learning analytics and latent profile analysis to explore the relations between reading engagement, motivation, and comprehension.*

ter Beek, M., Opdenakker, M.-C., Deunk, M. I., & Strijbos, J. W. (2019b). Teaching reading strategies in history lessons: A micro-level analysis of professional development training and its practical challenges. *Studies in Educational Evaluation*, *63*, 26–40. https://doi.org/10.1016/j.stueduc.2019.07.003

ter Beek, M., Opdenakker, M.-C., Spijkerboer, A. W., Brummer, L., Ozinga, H. W., & Strijbos, J. W. (2019). Scaffolding expository history text reading: Effects on adolescents' comprehension, self-regulation, and motivation. *Learning and Individual Differences*, *74*, 101749. https://doi.org/10.1016/j.lindif.2019.06.003

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis iteration.* Wellington, New Zealand: Ministry of Education. Retrieved on May 17th, 2019 from http://www.oecd.org/education/school/48727127.pdf

Tschannen-Moran, M., & Johnson, D. (2011). Exploring literacy teachers' self-efficacy beliefs: Potential sources at play. *Teaching and Teacher Education*, *27*(4), 751–761. https://doi.org/10.1016/j.tate.2010.12.005

Tze, V. M. C., Daniels, L. M., Buhr, E., & Le, L. (2017). Affective profiles in a massive open online course and their relationship with engagement. *Frontiers in Education*, *2*(65). https://doi.org/10.3389/feduc.2017.00065

U

Unrau, N., & Schlackman, J. (2006). Motivation and its relationship with reading achievement in an urban middle school. *The Journal of Educational Research*, *100*(2), 81–101. https://doi.org/10.3200/JOER.100.2.81-101

V

van Boxtel, C., & van Drie, J. (2018). Historical reasoning: Conceptualizations and educational applications. In S. A. Metzger & L. McArthur Harris (Eds.), *The Wiley international handbook of history teaching and learning* (pp. 149–176). Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/9781119100812.ch6

van de Grift, W. J. C. M. (2014). Measuring teaching quality in several European countries. *School Effectiveness and School Improvement*, *25*(3), 295–311. https://doi.org/10.1080/09243453.2013.794845

van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science*, *328*(5977), 453–456. https://doi.org/10.1126/science.1182594

van der Scheer, E. A., Glas, C. A. W., & Visscher, A. J. (2017). Changes in teachers' instructional skills during an intensive data-based decision making intervention. *Teaching and Teacher Education*, *65*, 171–182. https://doi.org/10.1016/j.tate.2017.02.018

van Drie, J. P. (2005). *Learning about the past with new technologies. Fostering historical reasoning in computer-supported collaborative learning.* (Doctoral dissertation, University of Utrecht, the Netherlands). Retrieved from https://dspace.library.uu.nl/handle/1874/7423

van Drie, J., Braaksma, M., & van Boxtel, C. (2015). Writing in history: Effects of writing instruction on historical reasoning and text quality. *Journal of Writing Research*, *7*(1), 123–156. https://doi.org/10.17239/jowr-2015.07.01.06

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

Vanhoof, J., Verhaeghe, G., Van Petegem, P., & Valcke, M. (2013). Improving data literacy in schools: Lessons from the school feedback project. In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in education: Challenges and opportunities* (pp. 113–134). Dordrecht, the Netherlands: Springer. https://doi.org/10.1007/978-94-007-4816-3_7

van Kuijk, M. F., Deunk, M. I., Bosker, R. J., & Ritzema, E. S. (2016). Goals, data use, and instruction: The effect of a teacher professional development program on reading achievement. *School Effectiveness and School Improvement*, *27*(2), 135–156. https://doi.org/10.1080/09243453.2015.1026268

Vanslambrouck, S., Zhu, C., Pynoo, B., Lombaerts, K., Tondeur, J., & Scherer, R. (2019). A latent profile analysis of adult students' online self-regulation in blended learning environments. *Computers in Human Behavior*, *99*, 126–136. https://doi.org/10.1016/j.chb.2019.05.021

VanSledright, B. A. (2004). What does it mean to read history? Fertile ground for cross-disciplinary collaborations? *Reading Research Quarterly*, *39*(3), 342–346. https://www.jstor.org/stable/4151775

van Steensel, R., van der Sande, L., Bramer, W., & Arends, L. (2016). *Effecten van leesmotivatie-interventies: Uitkomsten van een meta-analyse* [Effects of reading motivation interventions: Results of a meta-analysis]. Rotterdam: Erasmus University.

Vaughn, S., Roberts, G., Swanson, E. A., Wanzek, J., Fall, A-M., & Stillman-Spisak, S. J. (2015). Improving middle-school students' knowledge and comprehension in social studies: A replication. *Educational Psychology Review*, *27*(1), 31–50. https://doi.org/10.1007/s10648-014-9274-2

Vaughn, S., Swanson, E. A., Roberts, G., Wanzek, J., Stillman-Spisak, S. J., Solis, M., & Simmons, D. (2013). Improving reading comprehension and social studies knowledge in middle school. *Reading Research Quarterly*, *48*(1), 77–93. https://doi.org/10.1002/rrq.039

Veenman, M. V. J. (2007). The assessment and instruction of self-regulation in computer-based environments: A discussion. *Metacognition and Learning*, *2*(2–3), 177–183. https://doi.org/10.1007/s11409-007-9017-6

Veenman, M. V. J. (2016). Metacognition. In P. Afflerbach (Ed.), *Handbook of individual differences in reading: Reader, text, and context* (pp. 26–40). New York: Routledge.

Vermunt, J. K., & Magidson, J. (2013). *Latent GOLD 5.0 upgrade manual.* Belmont: MA: Statistical Innovations Inc. Retrieved on April 25th, 2019 from https://www.statisticalinnovations.com/wp-content/uploads/LG5manual.pdf

W

Wanzek, J., Swanson, E. A., Roberts, G., Vaughn, S., & Kent, S. C. (2015). Promoting acceleration of comprehension and content through text in high school social studies classes. *Journal of Research on Educational Effectiveness*, *8*(2), 169–188. https://doi.org/10.1080/19345747.2014.906011

Welie, C. J. M. (2017). *Individual differences in reading comprehension. A componential approach to eighth graders' expository text comprehension*. (Doctoral dissertation, University of Amsterdam, the Netherlands). Retrieved from https://pure.uva.nl/ws/files/9259065/Welie_Thesis_complete.pdf

Welie, C., Schoonen, R., Kuiken, F., & van den Bergh, H. (2017). Expository text comprehension in secondary school: For which readers does knowledge of connectives contribute the most? *Journal of Research in Reading*, *40*(1), 42–65. https://doi.org/10.1111/1467-9817.12090

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., & Barbosa, P. (2008). Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools*, *45*(5), 432–445. https://doi.org/10.1002/pits.20307

Wilschut, A., & Bitter, M. (1992). Geschiedenisonderwijs en computers [History education and computers]. In O. Boonstra, L. Breure, & P. Doorn (Eds.) *Historische informatiekunde. Inleiding tot het gebruik van de computer bij historische studies* (pp. 359–377). Hilversum, the Netherlands: Uitgeverij Verloren.

Wineburg, S. S. (1991). On the reading of historical texts: Notes on the breach between school and academy. *American Educational Research Journal*, *28*(3), 495–519. https://doi.org/10.3102/00028312028003495

Wineburg, S. S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science*, *22*(3), 319–346. https://doi.org/10.1207/s15516709cog2203_3

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

Wineburg, S., & Reisman, A. (2015). Disciplinary literacy in history: A toolkit for digital citizenship. *Journal of Adolescent & Adult Literacy*, *58*(8), 636–639. https://doi.org/10.1002/jaal.410

Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning. Theory, research, and applications* (pp. 297–314). New York: Erlbaum. https://doi.org/10.4324/9780203831076

Wolters, C. A., Barnes, M. A., Kulesz, P. A., York, M., & Francis, D. J. (2017). Examining a motivational treatment and its impact on adolescents' reading comprehension and fluency. *The Journal of Educational Research*, *110*(1), 98–109. https://doi.org/10.1080/00220671.2015.1048503

Z

Zepeda, C. D., Richey, J. E., Ronevich, P., & Nokes-Malach, T. J. (2015). Direct instruction of metacognition benefits adolescent science learning, transfer, and motivation: An in vivo study. *Journal of Educational Psychology*, *107*(4), 954–970. https://doi.org/10.1037/edu0000022

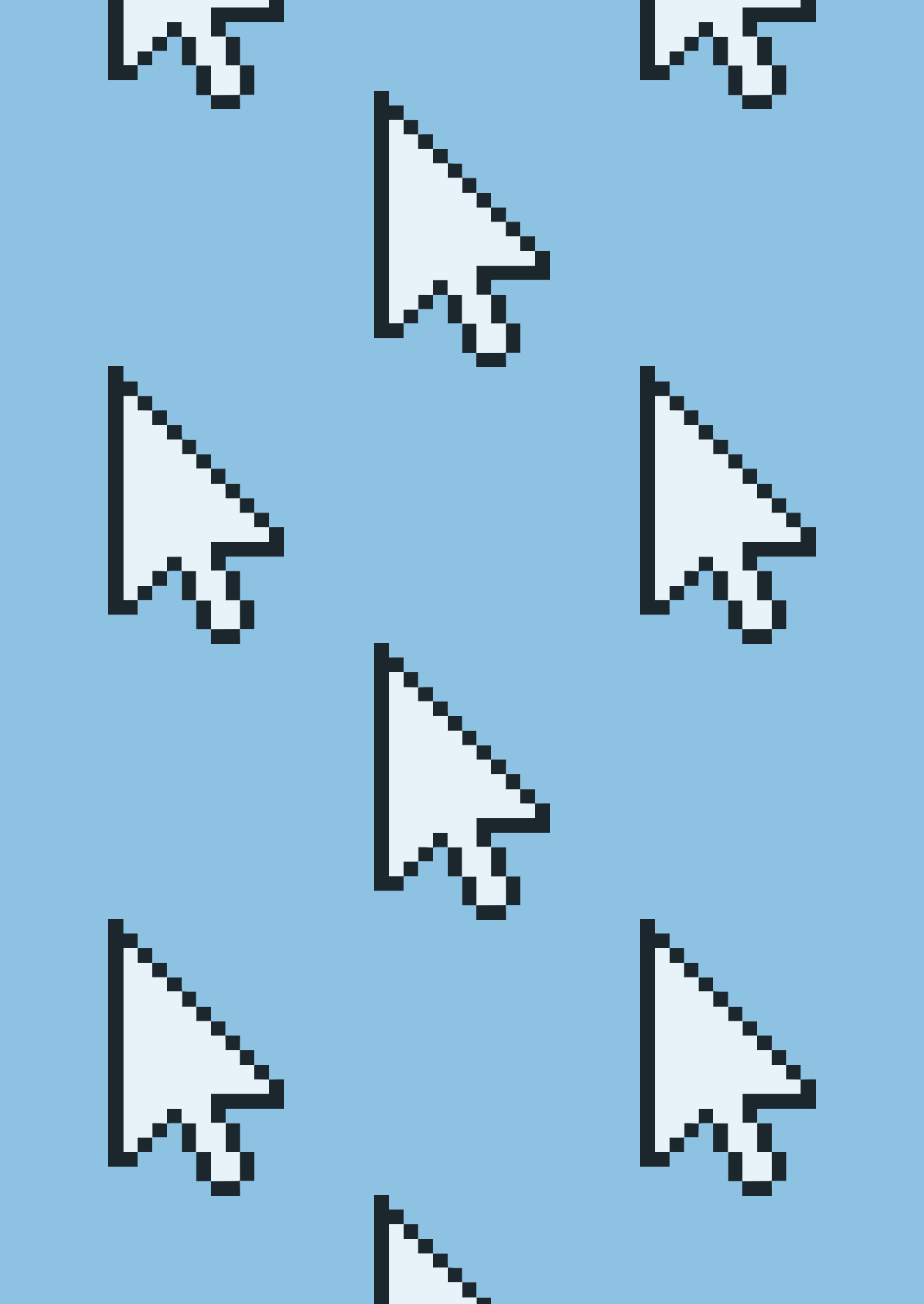
Zheng, L. (2016). The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: A meta-analysis. *Asia Pacific Education Review*, *17*, 187–202. https://doi.org/10.1007/s12564-016-9426-9

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–40). San Diego, CA: Academic Press. https://doi.org/10.1016/b978-012109890-2/50031-7

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166–183. https://doi.org/10.3102/0002831207312909

Zimmerman, B. J. (2011). Motivational sources and outcomes of self-regulated learning and performance. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 49–64). New York: Routledge.

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299–315). New York: Routledge.



About the author

Marlies ter Beek was born on June 30th, 1990 in Enschede, the Netherlands. After moving to Groningen to study History, she received her Bachelor's degree in 2011 and her Master's degree in History Today (*cum laude*) in 2012 at the University of Groningen. In 2013, she enrolled in the Postmaster Teacher Training track in History and Governance at the University of Groningen, during which she taught history lessons to multiple classrooms at secondary schools in Groningen and Friesland. She received her second Master's degree, including a first-degree teaching qualification, in 2014.

In November 2015, she started working as a junior researcher in the NRO-PPO project 'Cognitive, metacognitive, and motivational hints to promote self-regulated learning in secondary education: Researching the effectiveness of a supportive digital learning environment' at the Groningen Institute for Educational Research (GION). Marlies started her PhD trajectory, which originated from the NRO project, in February 2017. During her work as a PhD student, she was a member of the GION communication committee, advocating visibility of scientific research output through online media. Furthermore, she volunteered to answer practical educational research questions for the NRO Kennisrotonde. During the last year of her PhD trajectory, Marlies was involved in teaching courses for the Bachelor and Master programmes of pedagogical sciences (e.g., 'Educational Design' and 'Learning in Interaction with Artefacts') and obtained her University Teaching Qualification in October 2019.

Marlies is currently employed at SURFnet and the University of Twente, researching professional development opportunities for teachers' use of educational technology in the context of the Dutch 'Versnellingsplan Onderwijsinnovatie met ICT' (Acceleration plan Educational innovation with ICT). Additionally, she is still employed part-time at the GION until May 1st, 2020, to support her colleagues in various educational and research activities.

Publications

2019

ter Beek, M., Opdenakker, M.-C., Deunk, M. I., & Strijbos, J. W. (2019). Teaching reading strategies in history lessons: A micro-level analysis of professional development training and its practical challenges. *Studies in Educational Evaluation*, *63*, 26–40. https://doi.org/10.1016/j.stueduc.2019.07.003

ter Beek, M., Opdenakker, M.-C., Spijkerboer, A. W., Brummer, L., Ozinga, H. W., & Strijbos, J. W. (2019). Scaffolding expository history text reading: Effects on adolescents' comprehension, self-regulation, and motivation. *Learning and Individual Differences*, *74*(101749), 1–12. https://doi.org/10.1016/j.lindif.2019.06.003

ter Beek, M., Opdenakker, M.-C., Deunk, M. I., & Strijbos, J. W. (2019). Weten wat er staat: Onderzoek: digitaal oefenen kan begrijpend lezen verbeteren. *Kleio*, *60*(5), 54–56.

ter Beek, M., Opdenakker, M.-C., Deunk, M. I., & Strijbos, J. W. (2019). Begrijpend lezen is je vak. *Didactief*, *49*(9), 20.

2018

ter Beek, M., Spijkerboer, A. W., Brummer, L., & Opdenakker, M.-C. (2018). *Gemotiveerd, actief en zelfstandig lezen. Hoe een digitale leeromgeving zowel de leerling als de docent kan ondersteunen bij het begrijpend lezen van informatieve zaakvakteksten in het voortgezet onderwijs.* Groningen: Rijksuniversiteit Groningen.

ter Beek, M., Brummer, L., Donker, A. S., & Opdenakker, M.-C. J. L. (2018). Supporting secondary school students' reading comprehension in computer environments: A systematic review. *Journal of Computer Assisted Learning*, *34*(5), 557–566. https://doi.org/10.1111/jcal.12260

Submitted

ter Beek, M., Opdenakker, M.-C., Deunk, M. I., & Strijbos, J. W. (2019). *Using learning analytics and latent profile analysis to explore the relations between reading engagement, motivation, and comprehension.*

ter Beek, M., Opdenakker, M.-C., Deunk, M. I., & Strijbos, J. W. (2019). *The role of reading skills and engagement in history education.*

Presentations

2019

ter Beek, M., Opdenakker, M.-C., Deunk, M. I., & Strijbos, J. W. (2019). *Exploring the impact of a PD training on history teachers' reading strategy instruction.* Poster presented during EARLI 2019, Aachen, Duitsland.

ter Beek, M. (2019). *Begrijpend lezen bij brugklassers: Hoe een digitale leeromgeving zowel de leerling als de docent kan ondersteunen.* Poster gepresenteerd tijdens het Kennisfestival van het Ministerie van OCW, Den Haag.

2018

ter Beek, M., & Opdenakker, M.-C. (2018). *Begrijpend lezen bij brugklassers: Hoe een digitale leeromgeving zowel de leerling als de docent kan ondersteunen.* Presentatie verzorgd tijdens het NRO-congres 'Samen op Koers', Utrecht.

ter Beek, M., & Opdenakker, M.-C. (2018). *Motiverende opdrachten en hints in een digitale leeromgeving: Effecten op leerprestaties en motivatie van leerlingen.* Symposiumbijdrage tijdens de Onderwijs Research Dagen (ORD), Nijmegen.

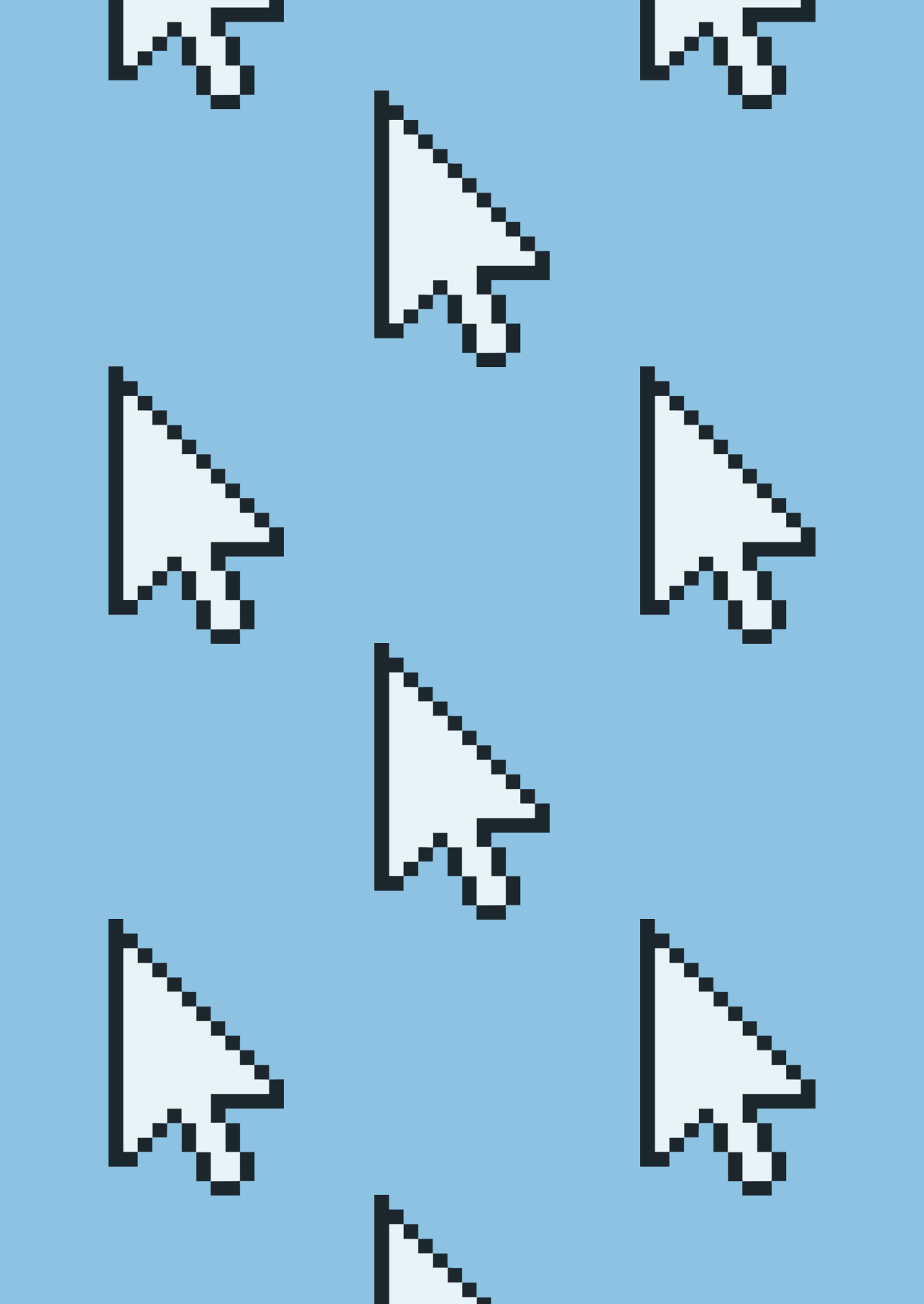
ter Beek, M. (2018). *Reading 2.0: Benefits for students.* Presentatie in het kader van de Three Minute Thesis competition van de Rijksuniversiteit Groningen. De video van deze presentatie is te zien via https://www.youtube.com/watch?v=VzKsC2bGlDI

ter Beek, M. (2018). *Gazelle: Gemotiveerd en actief zelfstandig lezen in een digitale leeromgeving.* Presentatie verzorgd tijdens IPON 2018, Utrecht.

2017

ter Beek, M., Brummer, L., Donker, A., & Opdenakker, M.-C. (2017). *Supportive hints in a digital learning environment: Effects on students' motivation.* Poster presented during JURE/EARLI 2017, Tampere, Finland.

ter Beek, M., Brummer, L., Opdenakker, M.-C., & Donker, A. (2017). *Digitale ondersteuning bij het lezen van geschiedenisteksten in het VO: Welk effect hebben hints op de prestaties van leerlingen?* Symposiumbijdrage tijdens de Onderwijs Research Dagen (ORD), Antwerpen, Belgium.



Dankwoord

Ze zeggen wel eens dat clichés een grote kern van waarheid bevatten. Voor de volgende uitspraak is dat zeker van toepassing: een proefschrift schrijf je niet alleen. Nu moet je jezelf natuurlijk niet vergeten (lekker bezig Marlies), maar op deze plek wil ik toch voornamelijk een woord van dank uitspreken aan alle andere mensen die hebben bijgedragen aan dit geheel.

Allereerst een groot woord van dank aan mijn promotor, Jan-Willem. Ik heb ontzettend veel geleerd van al jouw input over het opzetten van de hoofdstukken, het analyseren van alle data en het reageren op anonieme reviewers. Je gaf me ooit een compliment over mijn heldere schrijfstijl, maar jouw ontelbare suggesties in mijn Word-bestanden maakten dat mijn manuscripten veel wetenschappelijker geformuleerd werden. En ondanks dat we gekscherend wel eens gedacht hebben dat jij met de *APA manual* onder je kussen slaapt, wist je altijd wèl (tot op de schuingedrukte komma) de onregelmatigheden uit mijn literatuurlijst te vissen.

Ook wil ik hier mijn copromotor en tevens projectleider van het NRO-project, Marie-Christine, bedanken. Van begin af aan heb jij meegedacht over het design, de analyses en de focus van het onderzoeksproject. Je uitgebreide kennis en expertise hebben ervoor gezorgd dat mijn stukken een gedegen theoretische onderbouwing kregen. Daarnaast wist je altijd een kritische blik op mijn teksten te werpen; zelfs in het laatste jaar van mijn traject heb je met grote regelmaat op afstand mijn stukken gelezen en van waardevol commentaar voorzien. Ik waardeer het zeer dat je op deze manier betrokken bent gebleven bij de inhoud van mijn proefschrift.

Marjolein, jij bent in het laatste jaar van mijn promotietraject als copromotor aan mijn begeleidingsteam toegevoegd, waarbij je vanaf ons eerste overleg uitvoerig hebt meegedacht over het eindproduct dat hier nu ligt. Dankzij jouw 'blik van buitenaf' en je creatieve, reflectieve vragen wist je altijd de relevantie van mijn resultaten te benadrukken. Je bent bovendien ontzettend begaan met je collega's en de wereld om je heen, wat ervoor gezorgd heeft dat mijn (zelf)vertrouwen en optimisme overeind bleven tijdens het schrijfproces. Heel veel dank daarvoor.

Ook wil ik hier de leden van de leescommissie, Prof. dr. Carla van Boxtel, Prof. dr. Eliane Segers en Prof. dr. Adrie Visscher bedanken voor het lezen en beoordelen van mijn proefschrift.

Voor wat betreft het onderzoeksproject wil ik op deze plek graag het NRO bedanken voor de financiering: zonder deze subsidie was ik waarschijnlijk nooit op deze vacature gestuit en in de wereld van praktijkgericht onderwijsonderzoek terechtgekomen. Ook wil ik graag de coördinatoren, docenten en de meer dan 500 leerlingen op de deelnemende scholen bedanken. Dankzij jullie enthousiasme en inzet heeft dit proefschrift uiteindelijk tot stand kunnen komen.

Daarnaast een woord van dank aan de leden van het Gazelle-projectteam, die allemaal op verschillende momenten tijdens het onderzoek een bijdrage hebben geleverd en met wie ik altijd kon sparren over zowel inhoudelijke als persoonlijke onderwerpen. Alma, bedankt voor je hulp en de goede gesprekken tijdens het schrijven van het projectrapport. Anouk, Danny en Leonie: dank jullie wel voor de samenwerking in de eerste jaren van het Gazelle-project. Anouk, ik vond het fijn dat je zo af en toe even inventariseerde hoe het me allemaal verging en altijd de tijd nam om mijn e-mails (of zeg maar gerust epistels) te beantwoorden. Veel dank ook aan onze studentassistenten Eva, Hidde, Marida, Marrit en Tessa, voor al jullie dataverwerking, lesobservaties en de fijne samenwerking! Tot slot bedank ik hier graag Robbert, onze 'ict-man', die het Gazelle-programma heeft vormgegeven en wie we altijd konden mailen met last-minute verzoekjes.

Ik heb tijdens mijn promotietraject meerdere kamers en kamergenoten versleten. De beruchte 'kamer 205' mag daarbij zeker niet vergeten worden: speciale dank aan mijn voormalig kamergenoten Fabiola, Sanne, Hajo, Jochem en Hanneke. Onze goede gesprekken, woordgrappen, bordtekeningen, YouTubevideo's, et cetera <voeg hier zelf nog meer ongein in> hebben ervoor gezorgd dat ik elke dag met een grote glimlach thuiskwam. *We bellen!* Alle andere promovendi van het GION wil ik hier ook graag noemen. Alexandra, Anne, Arjan, Edwin, Inge, Mariëtte, Marij en Marinda: ik vond het heel fijn dat ik al die tijd zo'n fijne groep collega's om me heen had. De gezellige lunches, borrels, wandelingen, congresbezoeken, Sinterklaasvieringen en pubquizzen heb ik altijd heel erg gewaardeerd.

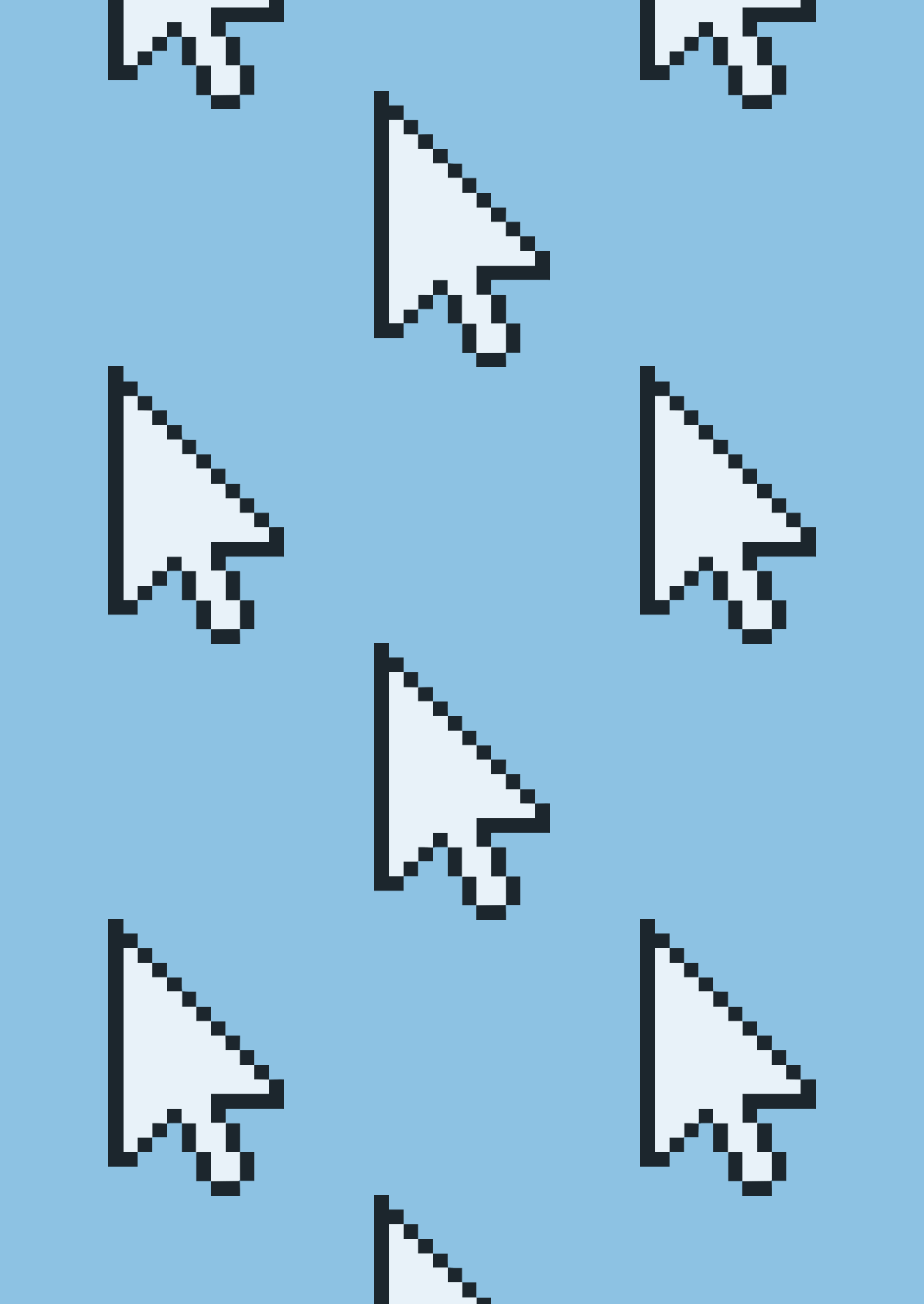
Tijdens mijn promotietraject heb ik ook onderwijs mogen verzorgen; op deze plaats wil ik daarom graag Jasperina en Jolien bedanken voor de fijne samenwerking hierin. Ook wil ik Matthijs hier bedanken, bij wie ik altijd mocht aankloppen voor statistiekadvies, en natuurlijk Sonja, Stephanie, Liesbeth en Nienke voor de nodige secretariële ondersteuning. Daarnaast ook aan alle andere collega's van het GION, die ik hier niet allemaal bij naam kan noemen: bedankt!

In de categorie collega's wil ik ten slotte Monique en Annegien in het zonnetje zetten. Monique, jij hebt vanaf het begin van het Gazelle-project alle ups en downs van het uitvoeren van praktijkonderzoek meegekregen. Ondanks dat (of moet ik zeggen: dankzij dit?) ben je ook als promovenda aan de slag gegaan bij het GION en heb ik daardoor al jaren een geïnteresseerde en betrokken collega. Dank daarvoor! Annegien, jij en ik hebben in het gevreesde laatste jaar een kantoor mogen delen, maar juist die combinatie heeft ervoor gezorgd dat ik het uiteindelijk tot een goed einde wist te brengen. Zonder de relativerende gesprekken en aandacht voor al het niet-werkgerelateerde was dit me niet gelukt. Gedeelde smart is halve smart, zullen we maar zeggen (nog zo'n waar cliché!) en ik vind het dan ook heel fijn en gezellig dat jullie tijdens mijn promotie als paranimfen aan mijn zijde staan.

Dan nog wat mensen die ik hier niet wil vergeten, omdat ze allen op een bijzondere manier hebben bijgedragen: mijn lieve vriendinnen 'uit het westen', bij wie ik altijd mijn verhaal kon doen, en tot slot natuurlijk Exstudiantes Dames 6, met wie ik (indien nodig) mijn sores kon wegslaan op het volleybalveld. Bedankt allemaal!

Naast collega's en vrienden is er natuurlijk ook nog mijn (schoon)familie, die ik voor zoveel meer kan bedanken dan wat hier nu op papier staat. Familie Venhuizen, bedankt voor jullie interesse in mijn onderzoek en de ontspannen weken in Oostenrijk waarin ik weer helemaal kon opladen. *Hail nuver boukje* dit, nietwaar? Ook Sander, mijn grote kleine broertje, bedankt voor alles: ik bewonder het dat jij altijd zo goed naar je eigen gevoel luistert en je eigen geluk vooropstelt. Daar kan ik soms een voorbeeld aan nemen. Papa en mama, dankzij jullie heb ik mijn eigen leer- en leesvaardigheid al op jonge leeftijd ontwikkeld, waardoor ik uiteindelijk een universitaire studie heb kunnen volgen. Jullie hebben me altijd aangespoord om te doen wat ik zelf interessant vond en geen seconde aan getwijfeld aan mijn kunnen. Wat ik ook doe, jullie zullen altijd trots op me zijn; daar ben ik jullie enorm dankbaar voor.

In dit dankwoord mag tot slot mijn lieve Ido uiteraard niet ontbreken. Ido, jouw nuchtere, wijze adviezen hebben meerdere malen ervoor gezorgd dat ik dit proefschrift heb kunnen (en willen) voltooien. Jij weet als geen ander hoe zwaar ik het af en toe vond, maar je had genoeg vertrouwen en wist me altijd weer moed in te spreken. En nog bedankt voor de afleiding: ik vond het heerlijk om tijdens het laatste jaar naast al het schrijven ook aan project 'bruiloft' te kunnen werken. Ik ben blij en vereerd dat jij al zoveel jaren de mijne bent.



ICO Dissertation Series

In the ICO Dissertation Series dissertations are published of graduate students from faculties and institutes on educational research within the ICO Partner Universities: Eindhoven University of Technology, Leiden University, Maastricht University, Open University of the Netherlands, Radboud University Nijmegen, University of Amsterdam, University of Antwerp, University of Ghent, University of Groningen, University of Twente, Utrecht University, Vrije Universiteit Amsterdam, and Wageningen University, and formerly Tilburg University (until 2002). The University of Groningen and the Erasmus University Rotterdam have been 'ICO 'Network partner' in 2010 and 2011. From 2012 onwards, these ICO Network partners are full ICO partners, and from that period their dissertations will be added to this dissertation series. (List update February 4, 2020)

401. Fix, G.M. (20-12-2018). *The football stadium as classroom. Exploring a program for at-risk students in secondary vocational education.* Enschede: University of Twente.

402. Gast, I. (13-12-2018). *Team-Based Professional Development – Possibilities and challenges of collaborative curriculum design in higher education*. Enschede: University of Twente.

403. Wijnen, M. (01-02-2019). *Introduction of problem-based learning at the Erasmus School of Law: Influences on study processes and outcomes.* Rotterdam: Erasmus University Rotterdam.

404. Dobbelaer, M.J. (22-02-2019). *The quality and qualities of classroom observation systems.* Enschede: University of Twente.

405. van der Meulen, A.N. (28-02-2019). *Social cognition of children and young adults in context.* Amsterdam: Vrije Universiteit Amsterdam.

406. Schep, M. (06-03-2019). *Guidance for guiding. Professionalization of guides in museums of art and history.* Amsterdam: University of Amsterdam.

407. Jonker, H.M. (09-04-2019). *Teachers' perceptions of the collaborative design and implementation of flexibility in a blended curriculum.* Amsterdam: University of Amsterdam.

408. Wanders, F.H.K. (03-05-2019). *The contribution of schools to societal participation of young adults: The role of teachers, parents, and friends in stimulating societal interest and societal involvement during adolescence.* Amsterdam: University of Amsterdam.

409. Schrijvers, M.S.T. (03-05-2019). *The story, the self, the other. Developing insight into human nature in the literature classroom.* Amsterdam: University of Amsterdam.

410. Degrande, T. (08-05-2019). *To add or to multiply? An investigation of children's preference for additive or multiplicative relations.* Leuven: KU Leuven.

411. Filius, R.M. (23-05-2019). *Peer feedback to promote deep learning in online education. Unravelling the process.* Utrecht: Utrecht University.

412. Woldman, N. (24-05-2019). *Competence development of temporary agency workers.* Wageningen: Wageningen University.

413. Donszelman, S. (06-06-2019). *Doeltaal-leertaal didactiek, professionalisering en leereffecten.* Amsterdam: Vrije Universiteit Amsterdam.

414. van Oeveren, C.D.P. (12-06-2019). *ITHAKA gaf je de reis.* Amsterdam: Vrije Universiteit Amsterdam.

415. Agricola, B.T. (21-06-2019). *Who's in control? Finding balance in student-teacher interactions.* Utrecht: Utrecht University.

416. Cuyvers, K. (28-08-2019). *Unravelling medical specialists' self-regulated learning in the clinical environment.* Antwerp: University of Antwerp.

417. Vossen, T.E. (04-09-2019). *Research and design in STEM education.* Leiden: Leiden University.

418. van Kampen, E. (05-09-2019). *What's CLIL about bilingual education?* Leiden: Leiden University.

419. Henderikx, M.A. (06-09-2019). *Mind the Gap: Unravelling learner success and behaviour in Massive Open Online Courses.* Heerlen: Open University of the Netherlands.

420. Liu, M. (13-09-2019). *Exploring culture-related values in Chinese student teachers' professional self-understanding and teaching experiences.* Utrecht: Utrecht University.

421. Sun, X. (13-09-2019). *Teacher-student interpersonal relationships in Chinese secondary education classrooms.* Utrecht: Utrecht University.

422. Wu, Q. (02-10-2019). *Making Construct-Irrelevant Variance Relevant: Modelling item position effects and response behaviors on multiple-choice tests.* Leuven: KU Leuven.

423. Jansen, R.S. (11-10-2019). *Dealing with autonomy: Self-regulated learning in open online education.* Utrecht: Utrecht University.

424. van Ginkel, S.O. (23-10-2019). *Fostering oral presentation competence in higher education.* Wageningen: Wageningen University.

425. van der Zanden, P. (05-11-2019). *First-year student success at university: Domains, predictors, and preperation in secondary education.* Nijmegen: Radboud University Nijmegen.

426. de Bruijn, A.G.M. (14-11-2019). *The brain in motion: Effects of different types of physical activity on primary school children's academic achievement and brain activation.* Groningen: University of Groningen.

427. Hopster-Den Otter, D. (28-11-2019). *Formative assessment design: A balancing act.* Enschede: University of Twente.

428. Harmsen, R. (10-12-2019). *Let's talk about stress. Beginning secondary school teachers' stress in the context of induction programmes.* Groningen: University of Groningen.

429. Post, T. (11-12-2019). *Fostering inquiry-based pedagogy in primary school: a longitudinal study into the effects of a two-year school improvement project.* Enschede: University of Twente.

430. Ackermans, K. (20-12-2019). *Designing video-enhanced rubrics to master complex skills.* Heerlen: Open University of the Netherlands.