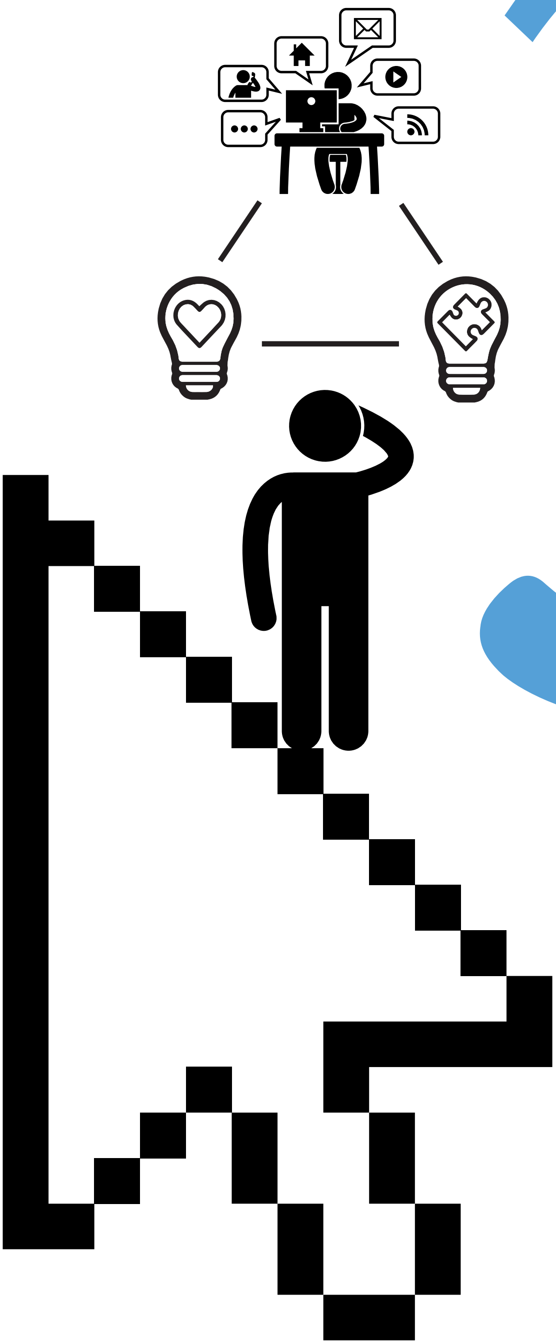
Supporting reading comprehension in history education

Beek, ter, Marlies

DOI:
10.33612/diss.121518620

Publication date:
2020



3

Chapter 3

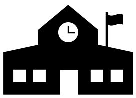
Using learning analytics and latent profile analysis to explore the relations between reading engagement, motivation, and comprehension

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.*

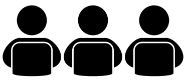
Abstract

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).

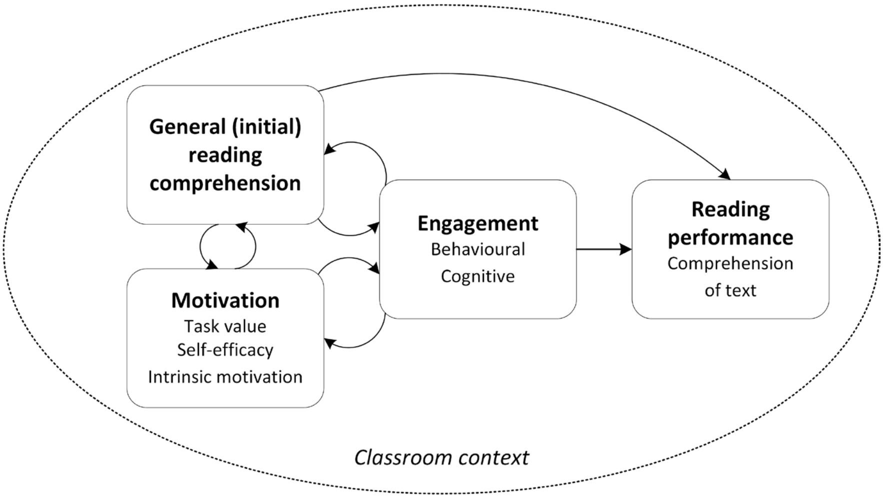


Figure 3.1 Conceptual framework for studying the relations between motivation, engagement, and reading comprehension.

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.

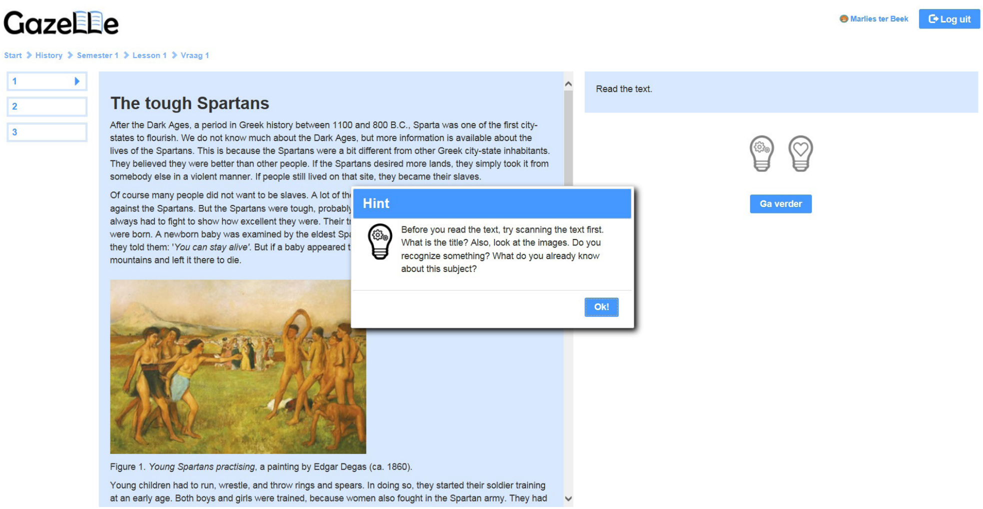


Figure 3.2 Screenshot of the Gazelle-program showing the contents of a supportive hint (translated from Dutch to English).

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).

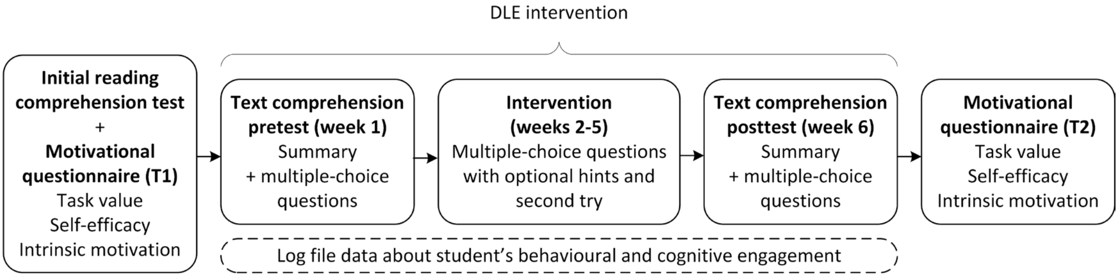


Figure 3.3 Timeline for the study procedure and data collection.

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, χ^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.

3

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.

3

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 ($p < .05$).

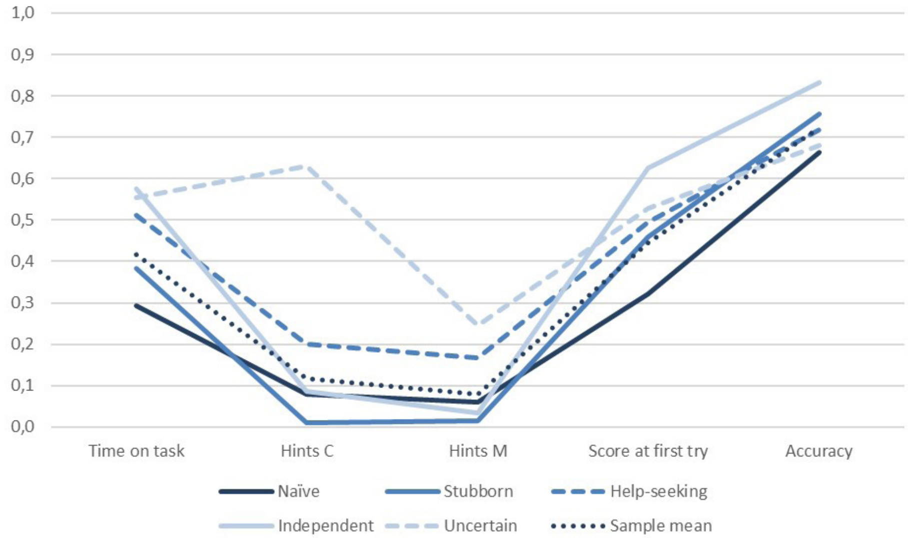


Figure 3.4 Normalised means [0–1] plot for the five latent profiles and the sample mean.

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 $p < .05$.

3

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.