FLIPPED LEARNING IN SECONDARY EDUCATION HISTORY CLASSROOMS

What are the Effects and
What is the Role of Self-Regulated Learning?

Flipped Learning in Secondary Education History Classrooms

What are the Effects and What is the
Role of Self-Regulated Learning?

David C.D. van Alten

Leden beoordelingscommissie:

Prof. dr. P.H.M. Drijvers

Prof. dr. D. Gijbels

Prof. dr. T. van Gog

Prof. dr. P.A. Kirschner

Prof. dr. J.W.F. van Tartwijk

ISBN: 978-94-6416-412-1

Cover design en layout: Publiss | www.publiss.nl

Printed by: Ridderprint | www.ridderprint.nl

© David van Alten, 2021, the Netherlands

All rights reserved. No part of this dissertation may be reproduced in any form without written permission from the author or, when appropriate, of the publishers of the publication.

Flipped Learning in Secondary Education History Classrooms

What are the Effects and What is the Role of Self-Regulated Learning?

Flipped Learning in het voortgezet onderwijs bij het vak geschiedenis
Wat zijn de effecten en wat is de rol van zelfregulerend leren?
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 19 maart 2021 des ochtends te 10.30 uur

door

David Cornelis Diemer van Alten

geboren op 2 februari 1990
te Rotterdam

Promotor:

Prof. dr. L. Kester

Copromotoren:

Dr. J.J.H.M. Janssen

Dr. C. Phielix

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van het Ministerie van Onderwijs, Cultuur, en Wetenschap (OCW/ PromoDoc/1065001).

Table of Contents

Preface	6
Chapter 1	9
General Introduction	
Chapter 2	23
Effects of Flipped Learning on Learning Outcomes and Satisfaction: a Meta-Analysis	
Chapter 3	59
Effects of Self-Regulated Learning Prompts in a Flipped History Classroom	
Chapter 4	93
Self-Regulated Learning Support in Flipped Learning Videos Enhances Learning Outcomes	
Chapter 5	125
Secondary Students' Online Self-Regulated Learning During Flipped Learning: A Latent Profile Analysis	
Chapter 6	157
General Discussion	
References	166
List of Supplementary Materials	184
Nederlandse Samenvatting (Summary)	186
Curriculum Vitae	193
List of Publications and Professional Contributions	194
ICO Disseration Series	196
Dankwoord (Acknowledgements)	200

Preface

For, and this makes us unique, we regard a person who takes no part in politics not as unambitious but as useless, and we are able to make informed decisions at all event, as far as we do not come up with constructive ideas ourselves; Instead of looking on discussion as a stumbling block in the way of action, we think it an indispensable preliminary to any wise action at all.

Here too we differ from others: in our enterprises we can show the greatest courage and at the same time the highest deliberation. For others, fearlessness is motivated by ignorance, and deliberation leads to hesitation. But the palm of courage will surely be adjudged most justly to those, who best know the difference between danger and comfort and yet are never tempted to shrink from danger.

Thucydides, *History of the Peloponnesian War* 40.2-40.3; Pericles' funeral oration.¹

This dissertation is the result of an ancient historian who started an adventure in 2015, *joined the social side*, and became an educational scientist. Still, I could not resist to begin this dissertation with an ancient historical fragment. This historical text was written in the fifth century BC by Thucydides. The so-called father of scientific history describes how Pericles, one of the prominent Athenian politicians of his time, publicly honors those who died in the first year of the Peloponnesian War with Sparta (460 – 445 BC). I am well aware that linking ancient and modern concepts is not without its risks. Nonetheless, this fragment represents one of the main themes in this dissertation and symbolizes various aspects for me.

In this text, Pericles defends Athenian democracy against other forms of government where only a small group or a single leader makes (quick) decisions. In these governments, in contrast to Athens, a political attitude of the citizens who reflect on these decisions is not desired. He argues that the power of Athens lies in the ability that their citizens are able (and allowed) to discuss important decisions, before taking action. The more information that is available to everyone, the better the discussion, and subsequently decision, will be. To describe this ability, Thucydides uses *λόγος* and *λογισμός*. These concepts can mean different things in ancient Greek, but in this context can best be translated as reasoning, deliberation, and reflection. For me, this ability to reflect on something before you take action is closely related to self-regulation, one of the main themes in this dissertation. In the second part of this dissertation, we investigate how to support students' thinking before (and about) learning in a flipped learning setting. A good self-regulator is someone who reflects and discusses before taking action. Pericles' words show to me that even thousands of years ago, the ability to self-regulate was discussed and seen as beneficial for future actions.

From a personal perspective, the concept of reflection as discussed by Pericles also symbolizes the process that resulted in this dissertation. I enjoyed extensive reflection on conducting social scientific research with so many great colleagues from Utrecht University, and in particular Liesbeth, Jeroen, and Chris. Looking back I conclude, just

as Pericles, that extensive reflection and deliberation rather have been a condition for the completion of this dissertation, than a stumbling block in our way.

1

Chapter 1

General Introduction

Flipped learning (FL) involves students preparing learning materials before class (e.g., by watching a lecture video), and applying the learning materials during class (e.g., by working on problem solving assignments). FL has become a widespread teaching method in all levels of education. As can be seen in Figure 1, research about FL has taken off in the last ten years. One of the first articles mentioning the phenomenon of inverted classrooms as “inverting events traditionally taken place inside the classroom now take place outside and vice versa” was already published in the year 2000 (Lage et al., 2000). Lage et al. illustrated how to let the students prepare learning materials by lectures outside of the classroom, to be able to deepen the learning materials in class with more interaction between students, and with their teacher. They were also one of the first to mention the increased responsibilities for students, the additional costs, and the positive perception of participating students. However, the underlying concept of moving instruction before class to extensively apply and discuss learning material during class was not new. For example, in the context of exploring the added value of computers in education and the possibility to use more effective learning activities during class time (Mazur, 1991). Or, from a constructivist perspective, researchers suggested that students learn better if they actively process information and apply it in new contexts, rather than getting information transmitted from instructors in the form of lectures (King, 1993).

In the direct years that followed the first scientific publication of inverted learning, the concept appears not to be studied widely. This changed at the beginning of the 2010s, through the works of the secondary school teachers Bergmann and Sams (2012) from the United States. It seems that from this period the popularity of FL arose mainly from practice, quickly followed by a steep increase in scientific research.

What is Flipped Learning?

Research into FL initially focused on what defines FL and what not. One of the issues in defining the concept is whether the use of videos and/or computer technology defines FL (Bernard, 2015). Some researchers define FL as students watching online instructional videos before class (Cheng et al., 2019; Hew & Lo, 2018; Lo et al., 2017; Scott et al., 2016; Strayer, 2012), or receiving computer-based individual instruction outside the classroom (Bishop & Verleger, 2013; Lo & Hew, 2017). In these definitions, FL can be considered as a particular example of blended learning (i.e., combining online and face-to-face education). Other researchers adopt a more generic definition of FL, in which neither videos nor technology necessarily define FL (Abeysekera & Dawson, 2015; Akçayır & Akçayır, 2018; Kim et al., 2014; Lage et al., 2000; Mullen & Sullivan Jr., 2015; O’Flaherty & Phillips, 2015; Peterson, 2015). In this view of FL, teacher guided activities before class such as reading material are also considered to be a part of FL. We adhere to the more broad definition of FL, as in our view the

medium students use before class in itself does not affect learning in general (Clark, 1994; DeLozier & Rhodes, 2017). In practice, most applications of FL are a specific type of blended learning, but this is not a defining aspect.

Figure 1

Amount of Records in the Web of Science Database (Title, Abstract, Keywords) per Year for Flipped Classroom, Flipped Learning, and Inverted Classroom. It Is Possible that one Record Contains Multiple Keywords

Another issue with the definition of FL is that it has become known under various concepts. Already in 2015, a large variety of terminologies and definitions of FL were found (Abeysekera & Dawson, 2015). As can be seen in Figure 1, inverted learning was the first term, that was mentioned in 2000, but the concept became widely known

as flipped classroom from 2012 onwards. In addition, flipped learning became an increasingly popular term. This was partly due the publications of Bergmann and Sams (2014), who earlier popularized the term flipped classroom and co-founded the Flipped Learning Network (2014). In their view, a flipped classroom is a simple flip by letting students prepare classes by reading texts or watching videos and apply learning materials inside the class. The flipped classroom *could* lead to flipped learning, which encompasses more engagement and a focus on the 'four pillars of F-L-I-P': Flexible environment, Learning culture, Intentional content, and Professional educator.

However, as seen in Figure 1, the terms flipped classroom and flipped learning seem to be used interchangeably in scientific publications, where flipped classroom remains the dominant term. In our view, both terms can be used to describe the situation in which students prepare learning material before class (e.g., by watching a lecture video), and applying the learning material during class (e.g., by working on problem-based assignments). We prefer the term FL, as it has less connotation of the requirement of a physical classroom. Although students engaging in the application of the learning material with their teacher and peers is an important aspect of FL, this can also be accomplished in an online class.

Why Would Flipped Learning Work?

The next question to answer is why FL could be an effective teaching method. Review studies analyzed the most commonly named benefits and challenges of FL in educational research, as presented in Table 1 (Akçayır & Akçayır, 2018; Lo et al., 2017). Besides, there are several theoretical suppositions why FL would work. From an educational psychological perspective, conveying information and instruction through conventional lectures in the traditional classroom has its shortcomings. The teacher delivers the instructional message in a one-size-fits-all manner and the possibilities for the students to interact with the teacher during the lecture (e.g., to ask or answer questions) are limited, which could hamper students' engagement and in turn their learning (Chi & Wylie, 2014; Schmidt et al., 2015). In addition, during FL students receive instruction before class, for example, through video lectures. This has the advantage that students are able to control the viewing frequency and viewing pace of the instruction material before class (Abeysekera & Dawson, 2015). In this way, cognitive load could be reduced as learner-controlled video segments are better able to support students processing the learning material (i.e., segmentation effect, see: Clark et al., 2005).

Second, students in traditional classes work independently on homework assignments to apply the information and instruction that was presented to them in lectures. From a cognitivist perspective, however, a lack of direct instructional guidance during

application might result in cognitive overload and hinders students to store knowledge in their long-term memory (Kirschner et al., 2006). In FL, however, students devote class time to learning activities focused on applying the instructional material with the guidance of their teacher (Roehl et al., 2013). Students’ learning processes benefit from a teacher that guides and facilitates active learning (i.e., a class in which students engage in learning activities instead of passively listening) by providing them direct feedback (van den Bergh et al., 2014). Moreover, from an educational psychological perspective, this is beneficial for learning because learning happens when the learner is actively generating meaning, instead of passively receiving information (Wittrock, 1992). Our mind actively and dynamically constructs meaning by building relations with prior knowledge (Wittrock, 1992). In the FL, this learning process is guided by the teachers.

Table 1
Major Benefits and Challenges of Flipped Learning, Adapted from Akçayır and Akçayır (2018) and Lo and Hew (2017)

Benefits of FL	Challenges of FL
Better learning retention	Disengagement of students during pre-class video instruction
Differentiated instruction / activities based on students' pre-class learning	Effects of the flipped method are not long lasting
Enhances ICT skills (teachers, students)	Implementation issues
Enhances learning outcomes	Increases workload (teachers, students)
Enhances students' satisfaction	Lacking ICT skills (teachers, students)
Flexible learning	Not enhancing learning outcomes
Fosters student motivation / engagement	Students' not able to ask questions during pre-class instruction
Fosters students' autonomy	Resistance to change (teachers, students)
Improves students' attendance	Start-up effort and costs for institutions
Increases students' study effort	Students dislike the method
More efficient class time	Students' lacking IT resources
More feedback of the teacher during class	Students' lacking SRL skills
More peer-interaction and collaborative activities during class	Unfamiliarity the method (teachers, students)
On-demand access to video instruction	Unpreparedness of students in class
Promotes creativity	
Students are better prepared before class	

What are the Effects of Flipped Learning?

Based on these theoretical suppositions, FL researchers increasingly began to explore the effects of implementing the FL method compared with non-flipped (e.g., more traditional) methods. As shown in Figure 1, more and more studies were published during the 2010s that measured effects of FL in a diverse set of educational contexts. Narrative reviews of FL provided general hypotheses and views on the effects of FL (e.g., Abeysekera & Dawson, 2015; Akçayır & Akçayır, 2018; Betihavas et al., 2016; Bishop & Verleger, 2013; Bond, 2020; DeLozier & Rhodes, 2017; Lo & Hew, 2017; Lundin et al., 2018; O'Flaherty & Phillips, 2015; Seery, 2015). However, these reviews need to be interpreted with caution. For instance, generalization of their conclusions is difficult because the number of studies considered in these reviews was small and the included studies often lacked the use of a control group (Lundin et al., 2018; O'Flaherty & Phillips, 2015). Because of the limitations of the previous reviews, we take a meta-perspective to provide an overview of the current state of research. Therefore, in **Chapter 2** we answer the question: What is the effect of FL on assessed and perceived learning outcomes and student satisfaction, compared with traditional teaching methods? We conduct a systematic review including meta-analyses on 114 available empirical studies with the aim of drawing more general conclusions about the effects of FL on student learning outcomes and satisfaction. As we distinguish learning outcomes into assessed (i.e., objectively measured) and perceived (i.e., self-reported) learning outcomes, we conduct three separate meta-analyses.

As important as it is to provide a general conclusion on the basis of the average effect of all the included studies, this does not tell us much about the circumstances in which FL works better, or which design is more effective. Therefore, we also examine if instructional design characteristics, educational context characteristics, or study quality characteristics explain variety in the effects of FL. We answer this second research question by making use of moderator analyses: Does the effect of FL on assessed and perceived learning outcomes and student satisfaction depend on instructional design characteristics, educational context characteristics, or study quality characteristics?

Flipped Learning and the Role of Self-Regulated Learning

We noticed that in the initial stage of FL research, the role of Self-Regulated Learning (SRL) skills in FL was notably absent in the scientific literature (Chen et al., 2018; Cheng et al., 2019; Låg & Sæle, 2019; Zheng et al., 2020). However, one of the possible explanations for the substantial variation in effect sizes of FL is the variation of students' SRL skills. In fact, to benefit from FL, certain SRL behavior (e.g., monitoring or planning) is required (He et al., 2016; Shih & Huang, 2019). In addition, it was not clear if and how SRL should be supported in FL environments. Therefore, we dedicate

the second part of this dissertation to this knowledge gap and aim to add more knowledge to the field of SRL support research during FL.

Defining Self-Regulated Learning

SRL has been conceptualized in different ways (Panadero, 2017). A common ground is that self-regulated learners metacognitively, behaviorally, and motivationally monitor their own learning and subsequently control their learning activities accordingly. SRL occurs at different levels. For example, students can control how long they study, if they want to restudy, what information they want to study, and what learning tasks they want to use (Kostons et al., 2012). SRL activities are usually covert actions in the minds of students (Veenman, 2016). The monitoring of self-regulated students occurs at the meta-level, which leads students to control (unconcealed) learning activities at the object-level (Nelson & Narens, 1990). As SRL activities cannot be directly assessed by observation, they have to be inferred from, for example, self-reports or behavioral traces (Veenman, 2016; Veenman et al., 2006).

In a common theoretical model, SRL is defined as a cyclical phase process in which students subsequently move from the forethought phase, to the performance phase, and to the self-reflection phase (Zimmerman, 2000; Zimmerman & Moylan, 2009). In the forethought phase, students analyze the task (e.g., goal setting and planning). Motivational beliefs such as self-efficacy, task value and goal orientation, affect students' task analysis and the activation of learning strategies (Panadero, 2017). In the performance phase, students work on the learning tasks, while actively metacognitively monitoring their progress. They perform self-control strategies such as time management and help seeking. Metacognition, which can be defined as knowledge, awareness, and regulation of one's learning, is one of the important skills for students to guide them through the SRL phases. In the self-reflection phase, students evaluate their performance and make attributions to adapt their approach to future performance.

In general, there is a clear relationship between SRL strategies and higher learning outcomes (e.g., Bjork et al., 2013; Dent & Koenka, 2016; Dignath & Büttner, 2008; Donker et al., 2014). One explanation is that students applying SRL activities make better use of cognitive activities and, in turn, improve their learning (Nelson & Narens, 1990). As students often struggle to adequately regulate their learning, previous research has shown that supporting SRL by means of SRL interventions (e.g., by means of SRL instruction or SRL prompts) improve students' SRL skills and in turn (at least partly) enhances learning outcomes (Dent & Koenka, 2016; Dignath & Büttner, 2008; Donker et al., 2014; Jansen et al., 2019; Zheng, 2016). Supporting students' SRL skills is even more important in the situation in which they have more autonomy over their learning process, as is the case in FL.

Flipped Learning Demands More Self-Regulated Learning Behavior of Students

Students' SRL skills are addressed when they have to regulate where, when, and how they study learning material outside of the classroom, in particular in an online learning environment (He et al., 2016; Kim et al., 2014; Lee & Tsai, 2011; Shih & Huang, 2019; Sletten, 2015; Tan et al., 2017). The extent to which students are given more autonomy in the FL setting differs according to the type of implementation (Van Laer & Elen, 2017). First, students can perceive different levels of autonomy, depending on how much autonomy they are given for studying learning material (Bouwmeester et al., 2019; Shih & Huang, 2019). Second, students' autonomy is also affected by the degree to which students can manipulate the pace and sequencing (e.g., pause and rewind the instructional videos) of the learning content (Abeysekera & Dawson, 2015; Van Laer & Elen, 2017), but also in their decision where (*environmental structuring*) and when (planning) to learn (Jansen et al., 2018). Consequently, when students are able to regulate their own learning adequately, they achieve higher learning outcomes in FL in contrast to students who are less able to regulate their own learning (Lai & Hwang, 2016; Lee & Choi, 2019; Shibukawa & Taguchi, 2019). Students with lesser SRL skills could show ineffective study behavior when they are given more autonomy. For example, because inadequate time management could lead to problematic study behavior if there is freedom of navigation and sequencing of instruction in an online learning environment (e.g., Butzler, 2016; Lee et al., 2010; Sletten, 2015). Moreover, even if students understand that different study behavior is necessary to benefit from FL, students do not always change their learning behavior, or, they may even show resistance to do so (Boevé et al., 2017).

In addition to students' SRL skills being an important condition for a successful implementation of FL, researchers have also asked whether participating in FL improve their SRL skills due to this increase in autonomy. In general, there is strong evidence for the recommendation to set an appropriate level of challenge in learning environments to develop students' SRL skills (Quigley et al., 2018). However, there are inconsistencies in the scientific literature of FL with regard to SRL skill acquisition. Meta-analytic research has found that students in FL improve in terms of SRL in comparison to students in a traditional setting (Tan et al., 2017). In contrast, Lape et al. (2014) found no significant improvements in SRL skills for students in FL compared with students in a traditional classroom.

Lastly, as the value of SRL support in implementing FL has become clear, this seems even more important for younger students in secondary education as SRL development it is related to age. First, although SRL skills begin to develop during pre-school and early-school years, they tend to strengthen and continuously develop during primary and secondary education (Muijs & Bokhove, 2020; Veenman et al., 2006). However, it seems that secondary school students' have less confidence in their SRL skills in comparison

to primary school pupils (Wigfield et al., 2006, 2011). This could mean that students in secondary education, with less confidence in their ability to regulate their learning, set easier goals which could lead to a negative influence on learning outcomes (Wigfield et al., 2011). Second, children's SRL skills also develop through maturation, interaction, and imitation of older peers, regardless of SRL instruction (Muijs & Bokhove, 2020). This implies that children develop SRL skills differently, according to the opportunities they get in their home and school context, in which social background potentially also plays a role (Muijs & Bokhove, 2020; Veenman et al., 2006).

To enhance students' SRL skills and confidence in performing SRL activities, SRL support at school can play an important role. Research has shown that SRL instruction can help young students to develop SRL skills faster and in a more structured manner (Dignath & Büttner, 2008, 2018; Muijs & Bokhove, 2020). This could be particularly beneficial for those students with a social background that offers fewer opportunities at home (Muijs & Bokhove, 2020; Veenman et al., 2006). In addition, SRL support has an impact on learning outcomes in secondary education and at higher ages, for example, if this support includes elaboration of the application of SRL strategies that students already acquired (Dent & Koenka, 2016; Dignath & Büttner, 2008). In general, meta-analytical research on SRL has shown that supporting SRL in primary and secondary education resulted in better learning outcomes (Dent & Koenka, 2016; Dignath & Büttner, 2018; Donker et al., 2014). There is also plenty of evidence from Hypermedia Learning and Multimedia Learning that shows that receiving SRL support enhances students' learning outcomes (Azevedo & Cromley, 2004; Azevedo & Hadwin, 2005; Bannert et al., 2009, 2015; Kauffman, 2004). As research on supporting SRL in FL contexts in secondary education are rare, however, our focus in the rest of the dissertation is on this particular educational context.

Supporting Self-Regulated Learning in Flipped Online Environments in Secondary Education

Video-Embedded SRL Prompts

How can we support secondary education students' SRL skills in an FL learning environment? For secondary education students, it is known that providing them with SRL prompts could lead to the internalization of SRL activities and, as a result, improve learning (Greene et al., 2015). Prompts are usually given to students in the form of questions and encouragements (e.g., *What have you learned?*, *Explain what you have learned*) to incite students to engage in SRL activities. Providing students with SRL prompts in a computer-based learning environment is one example of an effective strategy for improving students' SRL and learning outcomes (Devolder et al., 2012; Sitzmann et al., 2009; Wong et al., 2019; Zheng, 2016). In FL this online SRL support offers help to students when they need it the most: While they learn at home and their SRL skills are tested. In the context

of FL in primary and higher education, previous research has shown that embedded SRL prompts in instructional videos resulted in better learning outcomes, and students performing more SRL activities, compared with students who did not receive SRL prompts (Lai & Hwang, 2016; Moos & Bonde, 2016; Shyr & Chen, 2018; Yilmaz et al., 2018).

Therefore, in **Chapter 3**, we study an SRL support intervention in an authentic six-weeks flipped history classroom including 154 eight-grade students. The main research question we ask in this mixed-method study is: What is the effect of video embedded SRL prompts on secondary education students' (a) SRL self-reports, (b) online SRL activity, and (c) learning outcomes? We allocate students to an SRL-prompts condition or a no-SRL-prompts condition, using students' average scores for History to match students within classes to make sure both groups were comparable in terms of performance level. The students in the SRL-prompts condition receive three to four SRL prompts per video (similar to Moos & Bonde, 2016). Each prompt is designed to support students' SRL activities in one of the three SRL phases (Zimmerman & Moylan, 2009). For example, in the forethought phase one of our SRL prompts is: *How do you make sure that you will be able to recall all the learning material from this instructional video?*

Video-Embedded SRL Prompts and Self-Regulated Learning Instruction

Previous research into improving students' SRL skills supports the idea that adding explicit instruction of SRL in an online learning environment could be a beneficial teaching strategy in higher education (Bannert et al., 2009; Schwonke et al., 2013). For secondary education, it is known that teachers rarely explicitly support SRL by means of instruction in the classroom (Kistner et al., 2010; Moos & Ringdal, 2012). Still, this is beneficial for improving SRL skills and enhance learning outcomes (Kistner et al., 2010; Kramarski et al., 2013; Paris & Paris, 2001). In online secondary educational contexts, there are also indications that providing students with SRL prompts (aimed at improving all aspects of SRL) and explicit instruction about SRL is beneficial for students' learning (Dent & Koenka, 2016; Greene et al., 2010, 2015; Schwonke et al., 2013).

Therefore, the intervention in **Chapter 4** adds explicit SRL instruction to the SRL prompts. Before or after the instructional video, we shortly explain to students how they could effectively perform SRL activities, and why this is important for learning. We use the same measurements and learning materials as the study in **Chapter 3**. In the same educational context, with 156 eighth-grade students, we investigate: What is the effect of video embedded SRL support on students' (a) SRL self-reports, (b) online SRL activity, and (c) learning outcomes? In this way, we are able to answer the question if adding explicit SRL instruction to SRL prompts in an online FL environment improves the design of SRL support. In addition, we include two questions regarding student satisfaction. First, what is the effect of SRL support on students' satisfaction about the online learning environment? Second, how do students in the SRL-support condition value the SRL support?

A Person-Centered Approach to Online Self-Regulated Learning Behavior

Whereas the studies in **Chapters 3 – 4** take a *variable-centered approach*, the study in **Chapter 5** will take a *person-centered approach*. A variable centered approach describes associations among variables and assumes that the students are homogeneous with respect to how the predictors operate on the outcomes (Laursen & Hoff, 2006). It explains the relative importance of predictor variables explaining variance amongst the included students in outcome variables (Laursen & Hoff, 2006). This approach is helpful to demonstrate the main workings of SRL support on the average SRL behavior and learning outcomes of the student population.

A *person-centered approach* allows for a distinction of different subgroups amongst this population and assumes that the population is heterogeneous with respect to how the predictors operate on the outcomes (Laursen & Hoff, 2006). This approach describes differences among individuals in how variables are related to each other and identify groups of individuals who share particular attributes or relations among attributes (Laursen & Hoff, 2006). In SRL research, such an approach adds valuable insights as it investigates the possibility that some subgroups benefit more from SRL support than others (Barnard-Brak et al., 2011; Broadbent & Fuller-Tyszkiewicz, 2018; Kizilcec et al., 2017; Laursen & Hoff, 2006; Ning & Downing, 2015).

However, this has not been investigated in the context of FL in secondary education. We aim to address this knowledge gap by answering the following research question: Which SRL profiles can be identified for secondary education students during FL according to their online SRL activities? In online learning environments, trace data that indicates SRL behavior can be gathered automatically and unobtrusively. Previous research has shown the potential of data-mining processes (e.g., sequential pattern mining) to gather these traces of students' SRL from online learning environments (Jovanović et al., 2017; Kovanović et al., 2015; Maldonado-Mahauad et al., 2018). This type of SRL measurement is an important addition to self-reported SRL. For example, because students' SRL behavior changes throughout a course and this cannot be captured in a time-fixed questionnaire (Wang, 2019). Self-reports also stimulate students to think about their own learning and possibly also affect their SRL behavior (Greene & Azevedo, 2010). In addition, students can find it difficult to accurately assess their own SRL activity (e.g., due to overconfidence) or simply are subject to response sets and memory biases (Li et al., 2020; Pekrun, 2020; Wang, 2019).

It can be hypothesized that students with better SRL skills engage in online SRL activities more often, which in turn will predict their learning outcomes (Kizilcec et al., 2017; You, 2016). In this view, undertaking more SRL activities is an indication of better SRL skills, which are related to higher learning outcomes. However this quantitative

interpretation of SRL activities could fall short, as students with worse SRL skills may undertake more, but inadequate SRL activities and students with better SRL skills may undertake less, but more efficient SRL activities (Vanslambrouck et al., 2019; Wormington et al., 2012). To address this knowledge gap, we also relate the latent SRL profiles and their distinct SRL behaviors to their learning outcomes.

Overview of this Dissertation

Finally, in **Chapter 6**, we discuss the overall scientific implications of our results. We discuss methodological limitations and present guidance for future research. In addition, we reflect on practical insights that emerged from our FL interventions in an ecologically valid setting. Summarized, the current dissertation is divided in three parts and builds up chronologically. We provide a *history* of the rise of FL in **Chapters 1 – 2** and synthesize the current scientific literature about the effects of FL. We provide the *present* state of FL research in **Chapters 2 – 5** where we investigate the effects of SRL support embedded in flipped video instructions on SRL and learning outcomes, and explore the desirability of adapting SRL support to differences in online SRL behavior. Finally, these insights from our studies give rise to discussion of *future* prospects in **Chapter 6**.

2

Chapter 2

Effects of Flipped Learning on Learning Outcomes and Satisfaction: a Meta-Analysis

This chapter is based on:

van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review*, 28, 1–18. <https://doi.org/10.1016/j.edurev.2019.05.003>

The authors would like to thank E. Papangeli (EP) and J. Teunissen (JT) for their help with coding.

Acknowledgement of author contributions:

DvA, CP, JJ, and LK designed the study; DvA, CP, and JJ developed the coding schemes; DvA collected the data; DvA, CP, JJ, EP, and JT coded the included articles; DvA, CP, JJ, and LK planned the data analysis; DvA and JJ analyzed the data; DvA drafted the manuscript; DvA, CP, JJ, and LK contributed to critical revision of the manuscript; CP, JJ, and LK supervised the study.

Supplementary material to this chapter can be found on:

<https://doi.org/10.34894/NSXMET>

Abstract

During flipped learning, students study instructional material before class and apply this material during class. To provide a statistical synthesis of current research on effects of flipped learning, we conducted meta-analyses that included 114 studies which compared flipped and non-flipped classrooms in secondary and postsecondary education. We found a small positive effect on learning outcomes, but no effect was found on student satisfaction regarding the learning environment. In addition, we found considerable heterogeneity between studies. Moderator analyses showed that students in flipped classrooms achieve higher learning outcomes when the face-to-face class time was not reduced compared to non-flipped classrooms, or when quizzes were added to the flipped classrooms. We conclude that a flipped learning approach is a promising pedagogical approach when appropriately designed. Our results provide insights into effective instructional flipped learning design characteristics that support an evidence-informed application of flipped learning.

Introduction

In a *flipped learning* (FL) approach, students study instructional material before class (e.g., by watching online lectures) and apply the learning material during class. The popularity of this pedagogical approach, also known as *flipped classroom* or the *inverted classroom*, has been growing rapidly during the last decade and has been applied and investigated in a wide variety of educational contexts (Bergmann & Sams, 2012). However, an overarching quantitative synthesis of existing empirical research into the effects of FL, which would make it possible to draw general conclusions about FL, is lacking. Therefore, we present a meta-analysis investigating the effects of FL in secondary and postsecondary education on learning outcomes and student satisfaction, in comparison with a traditional (i.e., non-flipped) classroom in which students spend most classroom time on acquiring knowledge through lecture activities. We also provide valuable insights into possible differential effects of FL by including several instructional design characteristics of FL (such as group assignments, quizzes, and lecture activities) in a moderator analysis. The outcomes of this study is relevant to a wide variety of stakeholders such as teachers, educational policy makers and scholars who are interested in an evidence-informed application of FL.

Previous Research

In the current literature on FL, a large variety of definitions of FL can be found (Abeysekera & Dawson, 2015). We define FL as students preparing instructional material before class (e.g., by watching a lecture video), and applying the instructional material during class (e.g., by working on problem solving assignments). An issue related to the definition of FL seems to be whether the use of videos and/or computer technology before class defines FL (see also Bernard, 2015). Some researchers define FL as students watching online instructional videos before class (Cheng et al., 2019; Hew & Lo, 2018; Lo et al., 2017; Scott et al., 2016; Strayer, 2012), or receiving computer-based individual instruction outside the classroom (Bishop & Verleger, 2013; Lo & Hew, 2017). Other researchers, however, adopt a more generic definition of FL, in which neither videos nor technology necessarily define FL (Abeysekera & Dawson, 2015; Akçayır & Akçayır, 2018; Kim, et al., 2014; Lage et al., 2000; Mullen & Sullivan Jr., 2015; O'Flaherty & Phillips, 2015; Peterson, 2015). In this broader definition, activities before class such as reading material are also considered to be a part of FL. We chose to adhere to the last definition of FL, because in our view the medium students use before class in itself does not affect learning in general (Clark, 1994; DeLozier & Rhodes, 2017). Moreover, our aim is to conduct a comprehensive review study of the effects of FL without excluding certain design characteristics of FL in advance.

Important benefits and challenges of FL are provided by narrative review studies (e.g., Akçayır & Akçayır, 2018; Lo et al., 2017, see also Table 1 in Chapter 1). In addition, the application of FL can also be evaluated by the following suppositions from an educational psychological perspective. First, conveying information and instruction through conventional lectures in the traditional classroom has its shortcomings, because the teacher delivers the instructional message in a one-size-fits-all manner and the possibilities for the students to interact with the teacher during the lecture (e.g., to ask or answer questions) are limited. This might lead to a lack of student engagement and it might hamper students to actively construct knowledge (Schmidt et al., 2015). However, this should not imply that teachers should stop giving direct instruction. In an FL approach, therefore, students receive information and instruction before class, for example, through video lectures. This has the advantage that students are able to control the viewing frequency and viewing pace of the instruction material before class (Abeysekera & Dawson, 2015). In this way, cognitive load could be reduced as learner-controlled video segments are better able to support students processing the learning material (i.e., segmentation effect, see Clark et al., 2005).

It is important to consider that FL could induce an additional demand on students as it increases the appeal on the self-regulated learning capabilities of students. While some researchers presume that this may lead to positive effects in terms of learning outcomes and self-regulated learning capabilities of students (e.g., Lape et al., 2014), others hypothesize that student's lacking self-regulated learning capabilities could experience disadvantages in a flipped classroom (e.g., Lai & Hwang, 2016).

Second, students in traditional classrooms independently work on homework assignments to apply the information and instruction that was presented to them in lectures. From a cognitivist perspective, however, a lack of direct instructional guidance during application might result in cognitive overload and hinders students to store knowledge in their long-term memory (Kirschner et al., 2006). During FL, however, students can devote classroom time to learning activities focused on applying the instructional material (such as working through problems and engaging in collaborative learning) with the guidance of the teacher (Roehl et al., 2013). From an educational psychological perspective, this is beneficial for learning because learning happens when the learner is actively generating meaning, instead of passively receiving information (Wittrock, 1992). Our mind actively and dynamically constructs meaning by building relations with prior knowledge (Wittrock, 1992). In the flipped classroom, this learning process is guided by the teachers. Students in a traditional classroom lack this guidance when they engage in similar activities during homework assignments after class.

Third, according to Chi and Wylie (2014), students can achieve deeper understanding of the learning material as they become more engaged. Chi and Wylie (2014) distinguish

the following modes of engagement: *passive* (e.g., receiving knowledge, listening to a lecture and being able to recall information), *active* (e.g., manipulating knowledge, taking notes and being able to apply knowledge to similar contexts), *constructive* (e.g., generating knowledge, comparing and contrasting information and being able to transfer knowledge or procedures), and *interactive* (e.g., dialoguing, discussing with peers, being able to co-create knowledge) as modes of engagement. In an FL approach, students are better able to achieve higher learning outcomes, as there is more classroom time available for learning activities that foster active, constructive, and interactive engagement modes. In contrast, in a traditional classroom most classroom time is devoted to activities such as lectures, with mostly a passive mode of engagement.

Fourth, as compared to traditional classrooms, FL allows more classroom time for students to interact with their teacher(s) and peers, and to engage in learning activities that are differentiated based on their capabilities and needs (Bergmann & Sams, 2012). Classroom interaction enhances student relatedness (Abeysekera & Dawson, 2015), stimulates collaborative learning (DeLozier & Rhodes, 2017), and offers students the opportunity to receive support from their teacher or peers when applying knowledge during class. For instance, an essential component of a teacher's role in an active learning classroom (i.e., a classroom in which students engage in learning activities instead of passively listening) is to provide feedback to students to guide and facilitate students' learning processes (van den Bergh et al., 2014). Thus, in a flipped classroom, there may be more room for students to receive effective feedback and differentiated instruction from their teacher.

Finally, FL is often assumed to be a promising pedagogical approach that increases student satisfaction about the learning environment (O'Flaherty & Phillips, 2015; Seery, 2015). However, previous research found mixed results about students' attitudes towards FL. In general, students were positive about the learning environment, but a few studies showed opposite results (Betihavas et al., 2016; Lo & Hew, 2017). From a theoretical perspective, Abeysekera and Dawson (2015) hypothesize that an FL approach is able to satisfy students' needs for competence, autonomy and relatedness and, thus, entice greater levels of motivation (according to self-determination theory, see: Ryan & Deci, 2000).

Recent narrative reviews of FL provided general hypotheses and views on the effects of FL (e.g., Abeysekera & Dawson, 2015; Akçayır & Akçayır, 2018; Betihavas et al., 2016; Bishop & Verleger, 2013; DeLozier & Rhodes, 2017; Lo & Hew, 2017; Lundin et al., 2018; O'Flaherty & Phillips, 2015; Seery, 2015). However, these reviews need to be interpreted with caution. Generalization of their conclusions is difficult because the number of studies considered in these reviews was small, and the included studies

often lack the use of a control group (Lundin et al., 2018; O'Flaherty & Phillips, 2015). For these reasons, we conducted a meta-analysis, as a comprehensive synthesis of empirical studies that is able to draw more general conclusions about the effects of FL on student learning outcomes and satisfaction.

Recently published meta-analyses already showed the potential of FL in comparison with a more traditional approach in terms of student achievement. Lo et al. (2017) analysed 21 studies in mathematics education and found a small effect ($g = 0.30$), and Hew and Lo (2018) analysed 28 studies in health professions education and found a similar small effect ($g = 0.33$). Chen et al. (2018) compared 32 studies from health science and 14 non-health science studies and found an effect size of 0.47, while Cheng et al. (2019) included 55 studies from every domain and found an average effect of 0.19.

The current meta-analysis is able to add something to the scientific and practical relevance of these studies for the following reasons. First, we made no distinction in study domain or publication type, and were therefore able to include more studies ($k = 114$) to present a comprehensive meta-analysis on the effects of FL. Second, our meta-analysis includes more outcome variables than student achievement, as we analyzed the effects of FL on three types of outcome: assessed learning outcomes (determined by assessments such as exams), perceived learning outcomes (determined by students themselves in questionnaires), and student satisfaction. Third, we conducted extensive moderator analyses to examine if deeper lying instructional design characteristics, educational context characteristics, or study quality characteristics affect the effectiveness and attractiveness of FL. We used meta-regression models to better explain the variation between studies that occurs if the heterogeneity between effect sizes is larger than can be expected by sampling error only (Borenstein et al., 2009).

Rationale Moderator Variables

There is still little empirical evidence that provides insights into the instructional design characteristics of an effective FL approach, and therefore we examine if different implementations of FL have different effects. We selected moderator variables based on previous research. First, we consulted FL review studies to determine which theoretical grounding is used to explain variance of effects. Second, we conducted a preliminary literature search to determine which variables are frequently reported in FL intervention studies. We present our rationale for each included moderator variable in three categories: design characteristics, educational context, and study quality.

Design Characteristics

The following moderator variables have been selected as FL instructional design

characteristics: *adding quizzes*, *adding small group assignments*, *face-to-face classroom time*, and *lecture activities*. Quizzes and clicker questions are frequently implemented in flipped classrooms, for example, to determine to what extent students understand the instructional material provided before class (DeLozier & Rhodes, 2017). Previous research has shown that quizzes have a positive effect on learning outcomes because of the testing effect (Bangert-Drowns et al., 1991; Dirkx et al., 2014; Roediger & Karpicke, 2006). In addition, Spanjers et al. (2015) found in their meta-analysis on blended learning that the inclusion of quizzes positively moderated the effects of blended learning on learning outcomes and satisfaction. As FL can be regarded as a particular example of blended learning, defined as the combination between online and face-to-face education (Graham, 2006), we expected to find similar results. Hew and Lo (2018) and Lo et al. (2017) also found that instructors in a flipped classroom who employed quizzes at the start of their classes yielded higher learning outcomes as opposed to instructors who did not.

Another instructional design feature that is often applied during FL is the addition of *small group assignments*. DeLozier and Rhodes (2017) showed that learning in small groups with learning activities such as pair-and-share, paired problem-solving assignments, and group discussions in the FL context enhances learning outcomes. Research has shown positive effects of cooperative learning on achievement and attitudes (Kyndt et al., 2013; Slavin, 1991). Lo et al. (2017) also suggest that small-group learning activities could enhance the effectivity of a flipped (mathematics) classroom.

Furthermore, we examined if the amount of *face-to-face classroom time* might explain variance of the effects between the studies. It is sometimes argued that face-to-face classroom time can be reduced in a flipped classroom because students spend more time on learning activities before class. For example, a study on reducing classroom time in higher education by two-thirds through implementing FL showed that students in the flipped classroom achieved learning outcomes that were at least as good, and in one comparison even significantly better than the learning outcomes of students in the traditional classroom (Baepler et al., 2014). So, the authors concluded that active learning classrooms (blended or flipped) with less face-to-face time are at least as effective as traditional classrooms, and are thus more cost-efficient. In contrast, in a study that reduced face-to-face classroom time by making use of FL in order to help solve the shortage of teachers in secondary education (Heyma et al., 2015), it appeared that students in the flipped classroom performed significantly worse in terms of learning outcomes. The authors suggested this could be due the students inability to deal with more responsibility to regulate their own learning (Heyma et al., 2015).

Finally, we analyzed the use of *lecture activities* during the flipped classroom as a potential moderator variable. Although it is the aim of FL to reduce in class lectures

and present instructional material before class, Walker et al. (2008) concluded that students still highly value the integration of in class microlectures in a flipped classroom. Moreover, McLaughlin et al. (2013) suggested that in class microlectures in a flipped classroom can be used to address misunderstandings or gaps in student knowledge. Still, large-group lectures have their shortcomings in terms of effectiveness and engagement (Schmidt et al., 2015).

Educational Context

The following moderator variables with regard to the educational context were seen as valuable: *academic domain*, *educational level*, and *intervention duration*. We examined the possible influence of the *academic domain* in which FL is applied (with humanities, social sciences, natural sciences and formal sciences as categories). In this way, we wanted to investigate whether FL might be more effective in different academic domains. Only a small number of studies investigated possible differential effects of FL in particular academic domains (Bernard, 2015; Hew & Lo, 2018; Lo et al., 2017). O'Flaherty and Phillips (2015) suggested that some academic domains may be better suited to apply FL, as studies in different domains reported positive and negative student perceptions of FL. In other research, academic domain sometimes appeared to significantly moderate the findings as well. For example, in a meta-analysis on the effects of technology use in postsecondary education (Schmid et al., 2014) and in a meta-analysis on the effects of cooperative learning (Kyndt et al., 2013).

Moreover, we included *educational level* (e.g., secondary and post-secondary education) in the moderator analyses to see whether FL is as effective in higher education as in lower levels of education (Lo & Hew, 2017). FL, as a more student-centered approach (Calimeris & Sauer, 2015), might work differently for students of different age groups because of differences in self-regulated learning abilities which tend to increase at higher ages (Wigfield et al., 2011; Zimmerman & Martinez-Pons, 1990). We acknowledge that FL could have different meanings in secondary and higher education and that the two settings can sometimes be very different in the way they are structured. It seems safe to assume, however, that the main working mechanism of FL is similar in both secondary and higher education, as it has similar design characteristics. As we do not want to underestimate the difference in both educational contexts, we added educational type as a possible moderator variable to research if studies conducted in both contexts yielded differential effects.

Lastly, we considered *intervention duration* as moderator variable. Effects of FL may be affected by a novelty effect, which means that just because FL is a new approach, it affects the learning outcomes or satisfaction of students positively or negatively (Chandra & Lloyd, 2008). Furthermore, intervention duration might moderate the results if students need to get used for a longer period of time in order for a flipped

classroom to work effectively.

Study Quality

It is also possible that studies with a more rigorous experimental study design yielded significantly different results than studies with less quality (Cheung & Slavin, 2016). For example, in a meta-analysis by Lazonder and Harmsen (2016), it appeared that randomized experiments yielded significantly higher effects of guidance in inquiry-based learning on performance success than quasi-experiments. In previous narrative reviews on FL it seemed that in the initial phase of research the methodological quality of research could be improved as, for instance, only a small proportion of the studies used a randomized design with a control group (Bishop & Verleger, 2013; DeLozier & Rhodes, 2017). Consequently, we included the following moderator variables concerning the study quality, to investigate if these moderator variables have an impact on the effects of FL to estimate the magnitude of this problem. *Allocation type* was coded to distinguish studies that randomly allocated students to the flipped or control condition, randomly allocated pre-existing groups (e.g., there are eight classes in one school and four of them were randomly allocated to the flipped condition), and studies that were not able to apply randomization. *Control group type* constitutes the difference between studies that used a previous cohort as control group and studies that simultaneously studied a flipped and control condition (Cheung & Slavin, 2016). *Group equivalence test* means the manner in which studies assessed the comparability of students in the flipped and control conditions before the intervention in terms of prior knowledge (e.g., statistically tested prior knowledge in the form of a SAT math score or GPA, providing a descriptive statement, or not mentioned). Lastly, we examined *study type* as a possible moderator to investigate whether there is a difference in effect sizes reported in journals, dissertations, and conference papers. In this way, we could evaluate the presence of publication bias in our data, which means that published studies contain a significant higher effect size than non-published studies (Torgerson, 2006).

Research Objectives

Our first aim was to conduct three meta-analyses to examine the average effect of FL on assessed learning outcomes, perceived learning outcomes, and student satisfaction in comparison with a traditional classroom. We separately analyzed assessed and perceived learning outcomes as dependent variables, to make a distinction between learning outcomes determined by assessments (e.g., exams) and students themselves (e.g., questionnaire). As DeLozier and Rhodes (2017) showed, initial (qualitative) research attempts on FL were mainly focused on how students perceived their own learning in a flipped classroom, and they convincingly argued that students are often unable to correctly assess their own learning gains (see: Kruger & Dunning, 1999). Thus, there seems to be a fundamental difference between assessed learning outcomes determined by someone else than the student and perceived learning outcomes determined by the student and usually measured with questionnaires on how students perceive their own learning gains. Students' satisfaction of the learning environment, as third dependent variable in our study, was chosen to examine the question if participants of a flipped classroom give higher satisfaction scores than students in a non-flipped classroom. As Eichler and Peebles (2016) showed, student evaluations are often taken into account by policy makers in their evaluation of flipped classrooms. Student satisfaction is determined by the students and usually measured by self-reports on how satisfied they were with the learning environment.

Our second aim was to conduct moderator analyses to investigate if variety of effects between studies can be explained by instructional design characteristics, educational context, or study quality characteristics. We expected a significant amount of heterogeneity of the effects that cannot only be explained by sampling error, because FL implementations vary from classroom to classroom and educational context to context. In short, the current study answers the following research questions: (a) what is the effectiveness of FL on assessed and perceived learning outcomes and student satisfaction about the learning environment in comparison with a traditional classroom?; and (b) does the effect of FL on assessed and perceived learning outcomes and student satisfaction depend on instructional design characteristics, educational context characteristics, or study quality characteristics?

Method

Literature Search

The results of the literature search and the selection of articles can be consulted in Figure 1, based on the PRISMA flow diagram (Moher et al., 2009). We decided to

include unpublished studies (e.g., conference papers, and dissertations) to take into account the possibility of publication bias (Polanin et al., 2016). Our search process consisted of three parts. At the end of May 2016, we first consulted the databases OVID (MEDLINE® and PsycINFO), Web of Science, ERIC, and Scopus. As our goal was to retrieve all the available research on the phenomenon of flipped learning (including synonyms such as inverted classroom and flipping the classroom), the following search parameters were used in titles and abstracts: (Flip* OR invert*) AND (class* OR course* OR learning). A time period restriction was set to finding records after 1990, an ample margin since the concept of FL appeared in the scientific literature at the end of that decade (i.e., Lage et al., 2000). Second, we manually screened Google Scholar with the same combination of search terms. Third, we manually checked the references of all the known (narrative) reviews of FL. This yielded 22 additional records, of which three eventually were included. Two of these studies did not mention FL, and one was an unpublished study that was not included in the consulted databases. We decided to include these three studies because their FL interventions met our definition of FL.

Figure 1
Flow Diagram of Literature Search and Processing of Records

Inclusion of Studies

Of the 2,904 retrieved records in total, the first author removed 982 duplicates (see Figure 1). Then, during the first screening of titles and abstracts, 678 studies were excluded if a study's topic was clearly not about FL (such as physics studies which mentioned *classes* of macromolecules in combination with *inverted* micellar behavior). To establish reliability of the screening process, every record was independently screened by at least two people (the first author screened every record, and the second and third author each screened half of the records). Disagreements between coders were resolved by discussion.

Next, the first three authors assessed the full texts of the remaining 1,242 records for eligibility by applying seven inclusion and exclusion criteria presented in Table 1. We excluded articles if the language was not English, if no quantitative effects of FL were measured, and if studies did not make use of a between subjects design with a traditional classroom as control group (i.e., two-group RCT or quasi-experiments). In addition, studies which measured other dependent variables than learning outcomes or student satisfaction were excluded (e.g., they examined the opinions of involved teachers or directors, interaction in the classroom, self-regulated learning, collaborative learning, or critical thinking). In case the full-text of a record was not available via our own or other library systems, we consulted the websites of the scholars involved and sought contact with the first author to try to obtain the manuscript before we excluded a record.

We developed Table S1 (online only) with additional information of the 54 records we excluded, because they presented insufficient statistical information to calculate an effect size analysis (a complete list of references of these excluded studies is presented in appendix A, online only). In this table we provide an explanation of the insufficient data for each excluded study, together with parameters such as the authors' conclusion with regard to assessed and perceived learning outcome and satisfaction, the publication type, and the total sample size.

Lastly, we checked our search database for cases in which a journal article was also retrieved as dissertation or conference paper (where the article was based on). This was also done for records that were excluded because insufficient statistical information was provided. This yielded one dissertation that was used to retrieve additional information in the coding process for that particular study. In total, 114 studies were included in our meta-analysis (a complete list of references of these included studies is presented in appendix B, online only).

Table 1
Inclusion and Exclusion Criteria and Numbers of Studies Excluded

Criterion	Inclusion	Exclusion	Excluded studies
Language	English	Non-English language	14
Impact study	Studies investigating effects of FL	All other non-impact studies	709
Control group	Studies making use of a traditional control group	Studies not making use of a traditional control group	290
Dependent variable	Studies investigating assessed learning outcomes, perceived learning outcomes or student satisfaction	All other studies investigating other dependent variables	34
Availability	The full-text of the study must be available to consult via University library systems or the internet	Studies of which the full-text was not available to consult	27
Statistical information	Studies reporting sufficient statistical information to calculate an effect size	Studies not reporting sufficient statistical information	54

Note. Articles can be excluded based on multiple criteria, but in this table for each study only the first encountered exclusion criterion is given.

Coding of Study Characteristics and Moderator Variables

The included studies were systematically analyzed by making use of two coding schemes. One scheme on the study level and one scheme on the effect size level (Lipsey & Wilson, 2001). In the first coding scheme, general study information (e.g., study type and publication year), sample characteristics (e.g., gender and age), educational context features (e.g., duration of the intervention, academic domain, and region), and methodological characteristics (e.g., type of comparison group, allocation, ratios face-to-face learning time between the two groups, and assessment of group equivalence) were coded for each study. In order to be able to compare possible effects of the various types of classroom activities, the following categories were used: lecture activities, individual assignments, group assignments, and quizzes (for similar categorizations of in class activities see DeLozier & Rhodes, 2017; Lo & Hew, 2017). For both the flipped and the control condition, it was coded whether each of these type of learning activities occurred during class, or outside class. We assumed that a particular classroom activity was not applied if it was not mentioned in the study (e.g., when a study mentions that group assignments were given to students in the flipped condition, but information about group work in the traditional condition was absent). Similarly, if studies did not explicitly state that face-to-face classroom time was reduced, sustained, or increased in the flipped condition, we assumed it was kept constant with the traditional condition. In the second coding scheme on effect size level, we gathered information about the type of dependent variable (e.g., assessed learning outcome) and the raw data that was used to calculate the particular effect size.

After three rounds of testing and improving the coding schemes, the first three authors and two trained research assistants independently double-coded a set of 22 randomly chosen articles (about 18% of the included studies). This resulted in 87% total agreement, and a substantial interrater-reliability with an average Cohen's κ of .77 (range .45 - 1, $p < .001$). In some cases an unrepresentatively low Cohen's κ could be explained because some categories were overrepresented while there was high agreement between the coders (e.g., the variable *region* in which the category North America was coded in 80% of the cases; see Feinstein & Cicchetti, 1990). Nevertheless, the first three authors independently double-coded all variables that had a Cohen's $\kappa < .80$ (i.e., educational level, gender, allocation, group equivalence test, face-to-face learning time, student experience, group assignments, and quizzes). Remaining disagreements were resolved by discussion amongst the first three authors.

Computation of Effect Sizes

We used the standardized mean difference as an effect size metric, calculated as the mean differences between the mean scores of the flipped and control groups divided by the pooled standard deviation (Borenstein et al., 2009). A positive effect size in this study indicates that students in the flipped classroom (i.e., the experimental group) outperformed the students in the traditional classroom (i.e., the control group), students in the flipped classroom perceived they learned more than the students in the traditional classroom, or students in the flipped classroom were more satisfied with the learning environment. Hedges' g was used as effect size unit, which is a common adjustment of Cohen's d for studies with a small sample size (Hedges, 1981).

A total of 148 effect sizes were calculated on the basis of the following order of preference, because this hierarchical order maximizes the precision of the meta-analyses. The first preferred option was to calculate an effect size from adjusted posttest means that took potential pretreatment differences into account, either by regression analysis or covariance ($k = 8$). The second preferred method was to calculate effect sizes based on a pretest-posttest-control group design ($k = 18$). Morris (2008) demonstrated that an effect size based on the mean pre-post change in the intervention group minus the mean pre-post change in the control group divided by the pooled pretest standard deviation leads to the least biased estimate of treatment effects in pretest-posttest-control group designs. Formulas 8, 9 and 10 from Morris (2008) were used to calculate effect sizes, and formula 25 was used to calculate variances (formulas are provided in appendix C, online only). As only one study reported a correlation between the pretest and posttest ($r = 0.85$; Reza & Ijaz Baig, 2015) a sensitivity analysis was not possible (Borenstein et al., 2009). Therefore, we used a conservative estimation of $r = 0.70$, as recommended by Rosenthal (1991), for studies that did not report the correlation between pre- and posttest. In the

case these two options to calculate effect sizes were not possible, effect sizes were calculated on the basis of the raw means and standard deviations from the post-test ($k = 93$). When these were not reported, the least preferred method was the use of inferential statistics (such as the t -test, $k = 20$ and the F -test, $k = 9$). The online meta-analysis effect size calculator by Lipsey and Wilson (2001) was used in the case of adjusted post-test means and when studies reported raw means and standard errors. The metafor package (Viechtbauer, 2010) for the software program R (R Development Core Team, 2008) was used to calculate effect sizes from a pretest-posttest-control group design. In other cases, the software program Comprehensive Meta-Analysis (CMA, Version 3.3.070) was used.

Non-Independence

Eight studies reported the data of multiple flipped or traditional classroom conditions. In these cases, the means and standard deviations were pooled for that particular condition (following the formulas provided in Spanjers et al., 2015). When multiple effect sizes of one dependent variable in one study could be calculated (e.g., from two midterm exams and one final exam), we chose to only include the measuring point at the end of the intervention to be able to make a fair comparison between all included studies. Ten studies presented multiple post-test results that matched one of our dependent variables (e.g., objective learning outcomes). An example would be a study that presented results for three different post-tests of objective learning outcomes: a recall, an application, and a comprehension test. In those cases individual fixed effect meta-analyses were carried out and the results were included as a composite effect size in the main meta-analysis (Borenstein et al., 2009). In addition, one study reported two different interventions (with different samples). We treated this study as if it were two independent studies.

Data Analysis

Three univariate meta-analyses were conducted for each dependent variable using the metafor package for R (Viechtbauer, 2010). We used the random-effects model, because we do not assume that there is one general effect size of FL that is true for all studies (Borenstein et al., 2009). The study's variance consisted of a within study variance and between studies variance component, and studies were weighted according to the extent of variation (Borenstein et al., 2009). The effect of possible outliers in the data was checked by using forest and funnel plots and one-study-removed analysis in metafor for R (Viechtbauer, 2010).

We used the restricted maximum likelihood estimation method to fit all models. Besides standard error and 95% confidence intervals, we reported results of

the Cochran Q -test for statistical heterogeneity. The I^2 statistic indicates that the proportion of the variation in effect sizes is due to true between-study heterogeneity rather than sampling error, and $> 75\%$ represents considerable heterogeneity. The τ^2 statistic indicates the extent of between-study variance. Using the Knapp and Hartung (2003) adjustment to the standard errors of the estimated coefficients, we accounted for the uncertainty in the estimate of residual heterogeneity. With this adjustment, individual coefficients and confidence intervals use the t -distribution with $k - p$ (p being the number of model coefficients) degrees of freedom.

We conducted moderator analyses with a mixed-effects meta-regression, allowing for between-study variation, using restricted maximum likelihood to estimate the amount of residual heterogeneity (τ^2_{res}). Moderator variables are added as (study level) covariates in the linear meta-regression model to examine systematic differences in the strength of the effect (Borenstein et al., 2009). For categorical moderator variables dummy variables were created and continuous variables were centered around the mean. With the Knapp and Hartung (2003) adjustment, the omnibus test of the moderator analysis uses an F -distribution with m and $k - p$ degrees of freedom. I^2_{res} is given as a percentage of unaccounted variation between studies that is due to heterogeneity rather than sampling error and R^2 as the proportion of amount of heterogeneity accounted for by the moderator variable, often called the model predictive power (López-López et al., 2014).

Moderator analyses in a meta-analysis have certain statistical limitations (Hempel et al., 2013). Polanin and Pigott (2015) made clear that current practice in educational and psychological meta-analyses with regard to significance testing could be much improved. Moderator analyses are prone to Type I errors (incorrect rejection of a true null hypothesis leading to falsely identified moderator effects) if multiple (i.e. univariate) tests are conducted. Therefore, we used meta-regression models to account for multiple testing (Polanin & Pigott, 2015). Type II errors are also common in moderator analyses, because of low power (Hempel et al., 2013). To aid the reader in understanding the magnitude of this problem, we provided retrospective power analysis where appropriate (Hedges & Pigott, 2004; Hempel et al., 2013). While it might be true that retrospective power analysis for non-significant results are uninformative if they only provide observed power (Valentine et al., 2010), we will use these (two-sided) power calculations to estimate the power to detect what we assume to be the smallest important effect size (i.e., $g = 0.20$) given the number of studies we found, and the average within study sample size in those studies (Hempel et al., 2013; Valentine et al., 2010).

Publication bias

We included unpublished studies to reduce the effect of publication bias on our data and added publication type as a moderator variable to explore the possibility that published journals yielded significantly higher effect sizes than unpublished papers (Banks et al., 2012). In addition, we performed a cumulative meta-analysis sorted on precision (i.e., increasing standard error, see Borenstein et al., 2009). Funnel plots were produced using the metafor package to inspect asymmetry and heterogeneity. We also computed a Pearson's correlation test between effect size and sample size, as a negative correlation could indicate publication bias (Levine et al., 2009).

Results

Descriptive Findings

The included studies fell into the time span from the first included article published in 2006, until the end of May 2016, which was the final date of the literature search. In Table S2 (online only), we present an overview of the variables publication type, educational level, academic domain, study design, outcome that was measured, the effect sizes, and sample sizes for each included study. In this Table S2 (online only), the design of each study is summarized in terms of the type of control group (simultaneous vs. previous cohort) and the type of allocation (completely randomized each individual student, randomized pre-existing classes, or no randomization to groups). Next, in Table 2 we provide the frequencies of a selection of variables we coded, for categories that contained at least one study. This can be a helpful way of getting an impression of the characteristics of our sample. In addition to the information in Table 2, we found that the years of experience of the teacher(s) with FL was not reported in 50% of the studies, and in 41% of the studies it was their first experience with FL. Age was reported in 20% of the studies ($M = 20.40$, $SD = 3.09$). It should also be noted that from the 12 studies that appeared not to have equal student populations in both research conditions after a group equivalence test (e.g., a t -test on GPA), as shown in Table 2, ten of these studies took this difference into account by means of, for example, an adjusted post-test means regression analysis.

In Figure 2, a frequency distribution (in percentages) of the learning activities used in both the flipped and traditional classrooms is shown. About half of the flipped classrooms (53%) still contained some lecture activities inside the classroom, usually in the form of microlectures or just-in-time lectures (e.g., Eichler & Peebles, 2016). These microlectures are often tailored to the results of quizzes or other forms of formative assessment made by students. Furthermore, a clear distinction can be seen in the types and frequency of the different learning activity categories, where flipped classes

clearly contained more active learning components than the traditional classes, such as quizzes (both before and during class) and group activities.

Regarding the possibility of publication bias affecting our data, funnel plots for each dependent variable were examined for asymmetry, as presented in Figures S1, S2, S3 (online only). In addition, we conducted three cumulative meta-analyses by precision, presented in Figures S4, S5 and S6 (online only). They showed no positive drift towards the end as would be the case in the presence of a strong publication bias (Borenstein et al., 2009). Lastly, we found no strong evidence for a negative correlation between sample size and effect size in the data from assessed learning outcome, $r = -.15$, $p = .494$, perceived learning outcome, $r = -.03$, $p = .945$, and satisfaction, $r = -.10$, $p = .674$, as it would be the case when there is a bias against nonsignificant findings (Levine et al., 2009).

Figure 2

Percentage Frequency Distribution of Learning Activities in Both Conditions

Table 2*Frequencies of a Selection of the Coded Variables and Categories*

Variable	Category	Total studies (k)
Academic domain	Humanities	9
	Social sciences	15
	Natural sciences	51
	Formal sciences	40
Allocation	Randomized on individual level	11
	Randomized pre-existing groups	28
	No randomization	76
Educational level	Secondary education	11
	Higher education	104
Face-to-face time	Equal in both conditions	97
	Reduced in the flipped classroom	14
	Increased in the flipped classroom	4
Intervention and control group equivalence assessment	Statistically demonstrated	54
	Descriptive statement	18
	Only assumed to be equal	31
	Groups were not equal	12
Intervention duration	1-10 weeks	25
	> 10 weeks	90
Outcome measurement	Standardized test	22
	Non-standardized test (e.g., teacher made)	93
Region	Asia	13
	North America	91
	Europe	6
	Middle East	1
	Oceania	1
	The Caribbean	3
Study type	Journal publication	86
	Conference paper	23
	Doctoral dissertation	5
	Master thesis	1
Teacher	Same for both conditions	60
	Different for both conditions	21
Type of control group	Simultaneous group	57
	Previous cohort	58
Video-based instructional activities before class	Yes	110
	No	5

Note: In total 114 studies were included, but one study reported two independent interventions. If the categories for one variable do not add up to the total of 115 included interventions, it means that for this number of studies no information about this variable was reported.

Effects of Flipped Learning on Assessed And Perceived Learning Outcomes and Student Satisfaction

In Table 3, the results of the average weighted Hedges' g for assessed learning outcomes, perceived learning outcomes, and satisfaction are presented, including the 95% prediction intervals, the Q -test for heterogeneity, the between-study variance, and the percentage of variation between studies that is due to heterogeneity rather than sampling error. The average effect size for *assessed learning outcomes* ($g = 0.36$) was found to be significant. Given the number of studies and the average within study sample size in those studies, power to detect what we assume to be the smallest important effect ($g = 0.20$) was very high (0.99). The average effect on *perceived learning outcomes* was nearly identical ($g = 0.36$), but not significant ($p = .13$). However, the observed power was low (.44), as was the power (.17) to detect the smallest important effect ($g = 0.20$). For *student satisfaction*, a trivial and non-significant effect size ($g = 0.05$) was found. However, the power to detect an effect of $g = 0.20$ was low (.40). The Q -test was significant for all three dependent variables, the distribution of effect sizes is considered heterogeneous with $I^2 > 75\%$ indicating that a large proportion of the variability appears to be true variance. Forest plots of the dependent variables assessed and perceived learning outcomes and student satisfaction sorted on highest effect size can be consulted in Figures S7, S8, and S9 (online only). The average effect size does not seem to be heavily affected by outliers, as was visible in the one-study-removed analyses in R. The average effect of assessed learning outcomes without the most influential study was $g = 0.35$, $p < .001$, 95% CI [.27, .43] $I^2 = 86\%$ (Viechtbauer & Cheung, 2010).

Moderator Analyses

In Tables 4 and 5, the results of the meta-regression moderator analyses are presented for the effect on *assessed learning outcomes* and *satisfaction*. For *perceived learning outcomes*, moderator analyses were not regarded as meaningful, as the total number of studies was eight and thus most moderator categories could not be meaningfully compared. We conducted meta-regression analyses for each group of moderator variables separately (i.e., design characteristics, educational context characteristics, and study quality characteristics). First, this was necessary to account for multicollinearity, as we saw that some moderator variables correlated with each other. Thus, running a moderator analysis with all moderators in one model would have distorted the results. Second, we consider this approach as the best possible balance between the risk of type I errors (univariate moderator analysis) and type II errors (all variables in one model).

For *assessed learning outcome*, design characteristics moderated the findings as the omnibus test of all regression coefficients was significant ($p = .016$) and accounted

for 10% of the amount of heterogeneity. It appeared that studies that shortened the classroom time of the flipped condition had a significantly lower ($p = .027$) average effect than studies in which the classroom time in both conditions was equal (with a difference of $g = -0.26$, while accounting for all the other variables in the model). In addition, adding quizzes in the flipped condition also showed a significant ($p = .044$) difference with studies where quizzes were not added or already applied in the traditional condition (with a difference of $g = 0.19$, while accounting for all the other variables in the model).

The omnibus tests of all the regression coefficients in the educational context characteristics model ($p = .058$) and the study quality characteristics ($p = .161$) were not significant. We decided to drop type of control group as study quality moderator from the model, as it highly correlated with allocation type and therefore distorted the findings ($r = -.72$). The interpretation of both moderators is fairly similar, as for instance, a study with a previous cohort design is not able to randomly allocate students to conditions. According to the meta-regression power analysis simulation study by Hempel et al. (2013), the power of our *assessed learning outcomes* moderator models to detect a moderator effect of 0.2 was close to 80% (with 114 included studies, an average sample size of 180 students per study, and residual heterogeneity of 0.1).

For *satisfaction*, none of the moderator meta-regression models was significant. However, the power of these moderator models to detect a moderator effect of 0.2 was below 20% (with 22 included studies, an average sample size of 160 students per study, and residual heterogeneity of 0.3). Face-to-face ratio between conditions was dropped from the design characteristics model due to the unequal distribution of studies in the moderator categories. For the study quality model we also had to choose between the highly correlated variables allocation and type of control group. Here, we dropped allocation from the model, as it improved the R^2 , and the distribution of studies in the categories was more evenly in the type of control group variable.

Table 3

Results of the Univariate Random-Effects Meta-Analyses

Dependent variable	k (#students)	g (p)	SE	95% CI	Q (p)	df _g	τ ² (SE)	I ²
Assessed learning outcomes	114 (20318)	0.36 (< .001)	0.04	[0.28, 0.44]	1221.86 (< .001)	113	0.14 (0.02)	88%
Perceived learning outcomes	8 (953)	0.36 (.13)	0.21	[-0.13, 0.85]	39.45 (< .001)	7	0.28 (0.18)	87%
Student satisfaction	22 (3501)	0.05 (.73)	0.13	[-0.23, 0.32]	181.99 (< .001)	21	0.33 (0.12)	92%

Note. k = number of studies; # students = total number of participants; g = mean weighted effect size in Hedges' g; SE = standard error; CI = confidence interval; Q = Cochran's heterogeneity test; df= degrees of freedom Q-test; τ²= between-study variance; I²= percentage of variation between studies that is due to heterogeneity rather than sampling error.

Table 4

Results of the Moderator Analyses of Assessed Learning Outcomes

Moderator Variable	k	Estimate (SE)	p	95% CI	F (df1, df2) ^a	p ^b	τ ² _{res} (SE)	I ² _{res}	R ²
Design characteristics					F(5, 108) = 2.93	.016	0.10 (0.02)	86%	.10
Intercept	114	0.302 (0.07)	< .001	[0.17, 0.44]					
<i>Face-to-face time</i>									
FC = TC (ref)	97								
FC < TC	14	-0.260 (0.12)	.027	[-0.49, -0.03]					
FC > TC	4	-0.237 (0.21)	.271	[-0.66, 0.18]					
Group assignments									
No addition in FC (ref)	72								
Addition in FC	42	0.088 (0.07)	.310	[-0.08, 0.26]					
<i>Lectures during class</i>									
No lectures in FC (ref)	53								
Lectures in FC	61	0.028 (0.08)	.731	[-0.13, 0.19]					

Moderator Variable	k	Estimate (SE)	p	95% CI	F (df1, df2) ^a	p ^b	τ^2_{res} (SE)	I ² _{res}	R ²
Quizzes									
No addition in FC (ref)	78								
Addition in FC	36	0.189 (0.09)	.044	[0.01, 0.37]					
Educational context characteristics									
Intercept	114	0.309 (0.06)	< .001	[0.19, 0.42]	F(5, 108) = 2.22	.058	0.13 (0.02)	86%	.07
Academic domain									
Natural sciences (ref)	51								
Humanities	9	0.162 (0.16)	.316	[-0.16, 0.48]					
Social sciences	15	0.283 (0.12)	.022	[0.04, 0.52]					
Formal sciences	39	-0.035 (0.09)	.698	[-0.22, 0.15]					
Educational level									
Higher education (ref)	103								
Secondary education	11	0.179 (0.14)	.215	[-0.11, 0.46]					
Intervention duration	114	-0.012 (0.01)	.142	[-0.03, 0.01]					
Study quality characteristics									
Intercept	114	0.370 (0.07)	< .001	[0.23, 0.51]	F(8, 105) = 1.51	.161	0.13 (0.02)	86%	.05
Allocation type									
Not random (ref)	75								
Pre-existing groups	28	0.198 (0.10)	.056	[-0.01, 0.40]					
Individual allocation	11	0.161 (0.14)	.258	[-0.12, 0.44]					
Group equivalence test									
Tested, equal (ref)	54								
Tested, not equal	12	-0.020 (0.13)	.884	[-0.29, 0.25]					
Not tested, descriptive statement	18	-0.225 (0.12)	.067	[-0.47, 0.02]					
Not tested, no descriptive statement	30	0.008 (0.13)	.938	[-0.20, 0.21]					

Moderator Variable	k	Estimate (SE)	p	95% CI	F (df1, df2) ^a	p ^b	τ^2_{res} (SE)	I ² _{res}	R ²
<i>Study type</i>									
Journal publication (ref)	86								
Conference	22	-0.179 (0.10)	.092	[-0.39, 0.03]					
Dissertation	5	-0.041 (0.21)	.849	[-0.46, 0.38]					
Thesis	1	0.388 (0.45)	.394	[-0.51, 1.29]					

Note. Ref = reference category; k = total number of studies, as well as effect sizes; SE = standard error; CI = confidence interval; τ^2_{res} = estimated amount of residual heterogeneity; I² = percentage of unaccounted variation between studies that is due to heterogeneity rather than sampling error; R² = proportion of amount of heterogeneity accounted for; FC = flipped classroom, TC = traditional classroom.

^a Omnibus test of all regression coefficients of the moderators in the model.

^b p-Value of the omnibus test.

Table 5
Results of the Moderator Analyses of Satisfaction

Moderator Variable	k	Estimate (SE)	p	95% CI	F (df1, df2) ^a	p ^b	τ^2_{res} (SE)	I ² _{res}	R ²
Design characteristics									
Intercept	22	-0.520 (0.30)	.099	[-1.15, 0.11]	F(3, 18) = 1.76				
Group assignments									
No addition in FC (ref)	10								
Addition in FC	12	0.395 (0.31)	.221	[-0.26, 1.05]					
Lectures during class									
No lectures in FC (ref)	10								
Lectures in FC	12	0.554 (0.31)	.090	[-0.10, 1.20]					
Quizzes									
No addition in FC (ref)	15								
Addition in FC	7	0.156 (0.30)	.612	[-0.48, 0.79]					
Educational context characteristics									
Intercept	22	-0.105 (0.25)	.681	[-0.64, 0.43]	F(5, 16) = 0.36				
						.867	0.40 (0.15)	92%	.00

Moderator Variable	k	Estimate (SE)	p	95% CI	F (df1, df2) ^a	p ^b	τ^2_{res} (SE)	I ² _{res}	R ²
<i>Academic domain</i>									
Natural sciences (ref)	8								
Humanities	3	0.287 (0.47)	.548	[-0.70, 1.28]					
Social sciences	2	0.376 (0.56)	.515	[-0.82, 1.57]					
Formal sciences	9	0.086 (0.51)	.804	[-0.62, 0.78]					
<i>Educational level</i>									
Higher education (ref)	20								
Secondary education	2	0.086 (0.51)	.869	[-1.00, 1.17]					
Intervention duration	22	-0.039 (0.03)	.258	[-0.11, 0.03]					
<i>Study quality characteristics</i>									
Intercept	22	0.552 (0.24)	.037	[0.04, 1.07]	F(6, 15) = 1.45	.259	0.30 (0.13)	88%	.09
<i>Control group type</i>									
Previous cohort (ref)	12								
Simultaneous	10	-0.211 (0.28)	.458	[-0.80, 0.38]					
<i>Group equivalence test</i>									
Tested, equal (ref)	11								
Tested, not equal	2	-0.862 (0.45)	.075	[-1.82, 0.10]					
Not tested, descriptive statement	4	-0.539 (0.37)	.162	[-1.32, 0.24]					
Not tested, no descriptive statement	5	-0.770 (0.37)	.056	[-1.56, 0.02]					
<i>Study type</i>									
Journal publication (ref)	16								
Conference	4	-0.095 (0.40)	.816	[-0.95, 0.76]					
Dissertation	2	-0.475 (0.48)	.341	[-1.51, 0.56]					

Note. Ref = reference category; k = total number of studies, as well as effect sizes; SE = standard error; CI = confidence interval; τ^2_{res} = estimated amount of residual heterogeneity; I² = percentage of unaccounted variation between studies that is due to heterogeneity rather than sampling error; R² = proportion of amount of heterogeneity accounted for; FC = flipped classroom, TC = traditional classroom.

^a Omnibus test of all regression coefficients of the moderators in the model.

^b p-Value of the omnibus test.

Discussion

Is Flipped Learning Effective to Enhance Learning Outcomes and Student Satisfaction?

The first research question we addressed in the present study was about the effects of FL on assessed and perceived learning outcomes and student satisfaction. First, we found a significant small effect ($g > 0.2$; Cohen, 1988) of FL on assessed learning outcomes. According to Hattie (2012), this effect is close to the hinge point (> 0.4) of average effects of educational interventions where teachers and researchers should strive for. However, while these two interpretative rules of thumb are often used to interpret effect sizes, what do they mean in terms of the underlying dependent variable assessed learning outcome? For instance, on a Math exam score with $M = 550$ and $SD = 100$, an effect size of 0.36 means that the average score of a student in the flipped classroom is 0.36 standard deviations above the average student in the traditional classroom (e.g., a score of 586 vs. 550). In other words, 64% of the students in the flipped classroom will be above the mean of the students in the traditional classroom. Thus, although the effect on assessed learning outcome may be regarded as small, in the context of education it seems meaningful.

This is also true when we compare similar types of educational interventions in secondary education and other benchmarks. An effect of 0.36 is large if compared to the average annual gain in effect size ($g < 0.24$) from nationally normed tests on reading and math in the higher grades of secondary school (i.e., grade > 9 where FL is mostly applied; Hill et al., 2008), or if compared to the mean effect size from meta-analyses with comparable interventions in high school ($k = 28$, $d = 0.24$). For higher education, the meta-analyses by Schneider and Preckel (2017) provide an insight into comparable meta-analyses to gauge the relative size of effects. For example, on the basis of their ranking of 105 variables ordered by strength of their association with achievement, a 0.36 effect of FL would be comparable to other interventions on the instruction variable technology such as Intelligent tutoring systems (0.35, rank 47) and blended learning (0.33, rank 52).

Our main results are comparable with previous (smaller scale) meta-analyses of FL who found small average effect sizes ranged from 0.19 to 0.47 (Chen et al., 2018; Cheng et al., 2019; Hew & Lo, 2018; Lo et al., 2017). Together, these meta-analyses form a generalizable synthesis of the first generation of FL research: the combined remarks and recommendations could positively affect future research. In a broader context, the results are also in line with a meta-analysis on blended learning by Spanjers et al. (2015), which found a comparable significant and small effect ($g = 0.34$).

of blended learning on assessed learning outcomes. Even though FL is a particular example of blended learning, our meta-analyses did not contain the same studies in their sample.

Second, we found a non-significant average effect of FL on perceived learning outcomes. We chose to analyze assessed and perceived learning outcomes separately, as research has shown that learning outcomes determined by the subject itself are usually inaccurate (Kruger & Dunning, 1999). DeLozier and Rhodes (2017) showed that during initial FL research mostly students' perceptions of their own learning were measured, but we had to exclude many of these studies because they did not make use of a comparison group. It is possible that the effect is not significant due to the small sample size and therefore low power. However, it is notable that the average effect size is comparable to the effect on assessed learning outcome. More research should be conducted to be able to draw a more confident conclusion, although we regard assessed learning outcome as a better and more reliable measurement of learning outcome.

Third, we found a non-significant effect of FL on student satisfaction, with an effect close to zero. We can conclude that on average, students are equally satisfied with flipped classrooms as traditional classrooms. These findings are in line with the meta-analysis of Spanjers et al. (2015) on blended learning, in which they found a non-significant trivial effect size ($g = 0.11$) for student satisfaction. They speculate that this might relate to the suggestion made by Sitzmann et al. (2006) that students perceive web-based instruction as more demanding than classroom instruction in terms of time commitment. Tune et al. (2013) also report that an increase in workload was the most cited comment by students during FL. Thus, in contrast with preliminary conclusions made in qualitative reviews on FL, we conclude that there is no evidence that FL leads to higher student satisfaction about the learning environment. As the variance between the studies was large and in some studies students during FL were very positive and in other studies very negative in comparison with a traditional classroom, the design and educational context of FL should be carefully planned.

Do Instructional Design, Educational Context, or Study Quality Characteristics Explain Differential Effects of Flipped Learning?

The second research question we addressed was about the possible moderator effects of instructional design, sample, and study quality characteristics on the effectiveness of FL. For all meta-analyses, we found significant heterogeneity in effect sizes between studies, which is caused by more than random sampling alone. It means that in some studies, students during FL performed significantly worse than students in the traditional classroom (or did not perceive they learned better and

vice versa). For student satisfaction, this means that in some studies students in the flipped classroom were more satisfied than students in the traditional classroom, and vice versa as the average is close to zero.

It should first be noted that we should be careful with the interpretation of the moderator analyses and not draw too major conclusions. The proportion of amount of heterogeneity accounted for in all the meta-regression models was not higher than 10%. In addition, while the meta-regression models on assessed learning outcomes were around 80% power, the meta-regression models on student satisfaction had a very low power of 20%. Nevertheless, it is valuable to investigate patterns in the moderator analyses data which could contain indications of possible explanations for the large heterogeneity and provide helpful directions for further research.

We found that students in flipped classrooms in which the face-to-face time was not reduced performed significantly better compared to flipped classrooms that reduced classroom time. In most of the studies in which face-to-face classroom time in the flipped classroom was reduced, it was a deliberate decision (e.g., Baepler et al., 2014; Bossaer et al., 2016; Gross et al., 2015; Hibbard et al., 2016; Koo et al., 2016). Our finding contradicts the conclusion of Baepler et al. (2014) that flipped classrooms with less face-to-face time are at least as effective as traditional classrooms. Our results show that sustaining face-to-face time is a critical feature for a successful implementation of a flipped classroom. This is especially important for policymakers and curriculum designers, as it is sometimes argued that a flipped classroom approach is implemented in order to reduce costs (i.e., reducing face-to-face time, see: Asef-Vaziri, 2015).

The addition of quizzes during FL also seems to positively affect assessed learning outcomes. This is in line with research on the effectiveness of the testing effect (e.g., McDaniel et al., 2007). It is also consistent with the study by Spanjers et al. (2015), which also found that quizzes were a significant moderator that affected the attractiveness and effectiveness of blended learning.

Other moderator effects were not found. Taking into account the low power for the moderator analysis (especially for student satisfaction), this should not be interpreted as strong evidence for the absence of moderator effects. For example, given the results of the omnibus tests in the educational context characteristics model ($p = .058$) model for assessed learning outcome, it might be justified to further explore patterns in the moderator analysis data such as the difference between studies from social sciences who had an higher average effect of 0.28 than studies from natural sciences ($p = .022$). This is in line with another FL meta-analysis which found that subject area significantly moderated their results (Cheng et al., 2019).

Still, some additional context can be helpful to interpret the non-significant results. Adding small group assignments was not a significant moderator, although we hypothesized that this could be the case. Research has shown that, for example, cooperative learning (i.e., an instructional technique in which students work together in small groups on a structured learning task, in which positive interdependence and individual accountability are important factors) has positive effects on achievement and attitudes (Kyndt et al., 2013; Slavin, 1991). An explanation for our finding might be that we were not able to verify if the group assignments included in our sample met the requirements of, for example, cooperative learning, due to the broad coding categories (e.g., quizzes, group assignments, lecture activities). Therefore, we recommend future research to describe their design of in classroom (group) activities in greater detail. However, group assignments did not negatively moderate the effect sizes, which means that including group assignments in an FL implementation will not hamper learning outcomes either.

The same is true for lecture activities, as 53% of the flipped interventions still implemented some form of lecture activities in the classroom and our moderator analysis shows no negative result. In practice, adding microlectures can be used to address misunderstandings, for example on the basis of the outcomes of formative quizzes before class or at the beginning of a class (McLaughlin et al., 2013). Qualitative data from the study by Hagen and Fratta (2014) also showed that some students criticize the flipped classroom for a lack of lecture activities from their professor. This is in line with the conclusion by Walker et al. (2008) that students still value the integration of microlectures during FL. The result of our moderator-analysis on student satisfaction seems to fit into this context. Although adding microlectures to the flipped classroom was not a significant moderator, we noticed a trend that student satisfaction was higher in flipped classrooms that included lecture activities, in comparison to flipped classrooms without lecture activities (with a difference of $g = 0.55$, $p = .090$). If we take in to account the low power of this particular meta-regression model, it is at least an interesting pattern which is worth to investigate in future research.

Lastly, the moderator variable educational level was unequally distributed in our data ($k = 11$ in secondary education and $k = 103$ higher education). It seems that, while FL is also very widespread in secondary education (Bergmann & Sams, 2012), it is necessary to conduct more research in this context with more robust research designs (as many studies in secondary education were excluded because they did not make use of a control group). This is especially needed if one hypothesizes that FL can work differently for students of different age groups (for example, because of differences in self-regulated learning abilities, see Wigfield et al., 2011; Zimmerman & Martinez-Pons, 1990). To evaluate if the applications of FL in both contexts contained

different design characteristics, we checked the distribution of the features we coded in both contexts. It appeared that quizzes, group assignments, and lecture activities in the flipped classroom evenly occurred.

Other Potential Sources of Variation in Effects of Flipped Learning

Although the moderator analyses are a valuable start to explore possible patterns in the heterogeneity of the included studies, there are other important potential sources of variation in effects of FL to consider. Previous research has addressed the issue of a digital divide. This means that there are differences amongst subgroups such as gender, ethnicity, and socioeconomic background and how they have access to technology and how these differences might have an impact on a student's ability to benefit in terms of learning with technology (Jackson et al., 2008). In the context of FL, where students are usually required to watch online videos in their own time, it is necessary to evaluate the possible consequences of a digital divide in all its manifestations.

It seems that there is no direct evidence for strong differential effects due to the digital divide in our data, as most studies report that there are no differences in the effect of FL on learning outcomes for gender (Adams et al., 2016; Scott et al., 2016; Wasserman et al., 2015), no differences for gender and race (Lancellotti et al., 2016; Olitsky & Cosgrove, 2016), and no differences for gender, race, and socioeconomic status (Calimeris & Sauer, 2015). There was one study in which more women reported they watched instruction videos before class than men (Ossman & Warren, 2014).

Another potential source of variation in effects of FL is student's self-regulated learning (SRL) capability. We noticed that only a few of the included studies refer to SRL as an important variable in understanding the working principle of FL and even fewer studies measure SRL as a variable. Lape et al. (2014) conclude that there were no significant metacognitive gains visible after the intervention for students both in the flipped and control condition. This question presumes that students working in a flipped classroom will improve their SRL capabilities more than students in a traditional classroom (mediation effect). Evidence for this is found in a meta-analytical FL study within nursing education which included 29 Chinese studies, which found that self-learning abilities improved significantly better by students in flipped classrooms compared with traditional classrooms (Tan et al., 2017). However, it can also be hypothesized that SRL is an important moderator of the success of FL. As students with better SRL capabilities can benefit more from FL, this would presumably lead to higher learning outcomes (Lim & Morris, 2009). Then, teachers and researchers should consider to what extent their flipped classroom should include SRL support for students, as there is evidence that flipped classrooms with SRL support achieve

higher learning outcomes in comparison with flipped classrooms without SRL support (Moos & Bonde, 2016; Shyr & Chen, 2018).

The same is true for motivation, as the effect of FL can be dependent on how well students are motivated to study in advance. For instance, students' motivation during FL could be positively or negatively affected by their preference for this method (e.g., Asiksoy & Özdamli, 2016; Hibbard et al., 2016). If a student is not motivated to complete preparatory activities before class, this could lead to negative effects, as that student will probably experience difficulty applying learning materials during class. Therefore, it seems important to take students' motivation into account when designing a FL course. For example, instructing students about the benefits of FL could increase motivation to study before class, while a too precise in-class repetition of content from the instructional video could hamper the need for students to study before class. Unfortunately, we could not investigate motivation as moderator variable as few studies measured students' motivation.

Limitations

The following limitations in our review should be kept in mind when interpreting our results. We compared flipped classrooms with traditional classrooms. The frequency distribution of learning activities in Figure 2 shows that flipped classrooms are comparable with each other, as well as traditional classrooms. While we were only able to compare these two main categories, we still acknowledge that flipped and traditional classrooms can differ in practice. We recommend future research to explicitly describe both research condition in more detail. In addition, there are a few variables which contain important information about variation in effects but are not systematically reported in current studies on FL, such as age, gender, class size, and experience with a flipped classroom of both teachers and students. Some variables that we coded could not be used in the moderator analyses, because one category was heavily underrepresented in the data (e.g., 109 included studies did make use of online videos and only five studies did not). For example, the effect of the medium of instruction before class in the flipped condition could not be meaningfully compared, as 109 included studies did make use of online videos and only five studies did not. Lastly, most studies were conducted in the region North America (mainly the United States) and in the context of post-secondary education educational. This means that these two contexts are overrepresented in current research on FL and in our data.

The information about costs of an FL implementation was also underrepresented in our data. Although we argue that FL is not inherently dependent on the use of technology, we see that in most implementations several high- and low-cost technologies are used. For instance to record instruction video's, to design online

learning environments where video's and additional learning assignments are distributed, and sometimes technologies to apply the learning material in class. Besides technology costs, implementing a flipped classroom also means an initial huge time investment (e.g., Talbert, 2015), even if one decide to find free to use online videos. One might wonder if the costs of implementing FL is justified given the results we presented in this meta-analysis. We are, however, unable to answer this question, as this is heavily context dependent and the included studies often did not report details about costs. Nevertheless, we would like to point out some interesting studies that might be helpful for teachers and policy-makers. Touchton (2015) showed that initial implementation costs can be high, but that this investment can pay off on the long term. Peterson (2015), Hudson et al. (2015), and O'Flaherty and Phillips (2015) note that IT support can be critical for a successful implementation of FL. Asef-Vaziri (2015) shows that some institutions are interested in FL in order to reduce costs, when a significant portion of face-to-face classroom time is replaced by online learning. This last idea should be explored with caution, since our moderator analysis shows that sustaining face-to-face time leads to significant higher effects on assessed learning outcome. Still, some included studies report a successful cost reduction by implementing a flipped classroom while at the same time student learning outcomes were at least equal to the previous traditional classroom (Hudson et al., 2015; Olitsky & Cosgrove, 2016).

Besides, there are a few biases related to the average duration of the interventions. First, it usually was the first experience with FL for teachers. Although our results show that these first implementations of FL were on average enhancing assessed learning outcomes, we think that such big changes in a curriculum become even more effective after (minor) adjustments based on evaluation of the first implementation. For instance, the study by Hudson et al. (2015) showed that student satisfaction in their flipped classroom only increased after several semesters of redesigning the flipped course. In addition, the behavior of teachers during class changes with a flipped approach, but research has shown that, in general, training and time is needed to achieve effective change in teachers' behavior (Niemi, 2002; van den Bergh et al., 2014). Second, it seems that students in the flipped classroom seem to need an adaption time to get used to it that can take several weeks to a full semester (Hotle & Garrow, 2015).

Another potential bias is a combination of effects that are almost inherent to educational research. Some students may have been aware they are participating in a research condition which may affect their expectations and behavior (Hawthorne-effect), and some students in the flipped classroom could have shown better performance than students in a traditional classroom from the expectation of teachers that this innovation would lead to better learning outcomes alone (Pygmalion-effect). It turned

out that 60 studies in our sample had the same teacher for both conditions (in which a teacher's preference and experience could be a bias) and 21 studies had different teachers for both condition (which is a bias on itself). Therefore, we cannot exclude the influence of these effects nor control for them in our meta-analysis. However, there are reasons to assume that the impact of these underlying effects is not high. First, it is clear that not all students in the flipped classrooms were satisfied with their learning environment in comparison to students in a traditional classroom: the effect size for satisfaction was $g = 0.05$, and heterogeneity was large and significant. In addition, Figure S9 (online only) shows that 12 out of 22 studies reported a negative or neutral effect size. Second, we separately analyzed assessed and perceived learning outcomes to establish a possible difference between what students think they have learned, and what was measured in achievement tests. Still, we recommend future research into the effect of FL in comparison with non-flipped classrooms to better account for the possibility of a Hawthorne-effect, by, for example, use a between and within subjects research design that switches both methods (preferably also several times) between independent groups of students.

In addition, we had to exclude 54 studies because they did not report enough statistical information to calculate effect sizes. If we compare the results from the meta-analyses with the results of studies that were excluded because they lacked sufficient statistical information as qualitatively analyzed and shown in Table S1 (online only), it seems that the reason they contain non-significant results does not occur more often than that the authors report a positive finding. Because of the largely equal distribution of positive and neutral reported outcomes, we consider that excluding these studies did not have a great impact on our general findings. However, the statistical power of the meta-analyses as well as the moderator analyses could therefore have been reduced.

The influence of publication bias should be never neglected in any meta-analysis. However, results of the tests we performed to estimate the effect of publication bias seem to indicate that it did not have a major impact on our results (e.g., funnel plots, cumulative analyses, correlation test between effect size and sample size). In addition, we also included non-published articles, and we did not find a moderator effect for publication type.

Conclusion

The present study quantitatively synthesized the results of 114 studies about the effects of FL on learning outcomes and student satisfaction. With respect to the issue of defining FL, we found that 95% of the included flipped interventions used video based instruction activities. Therefore, we continue to follow Abeysekera and Dawson

(2015) in their recommendation to adopt a high-level uniform and broad definition of FL, provided that researchers are very specific about the operationalization of the definition.

In general, we can conclude that students during FL achieve significantly higher assessed learning outcomes than students in traditional classrooms, and are equally satisfied with the learning environment. The main implication following our results is that flipped classrooms are worth implementing. Careful attention should be paid, however, to the design of the flipped classroom as simply flipping before and during classroom activities might be not enough. The significant heterogeneity in the effect sizes of the studies means that it matters how flipped classrooms are implemented. Our study provides a few general directions to effective design of a flipped classroom, as sustaining face-to-face time and adding quizzes are critical features for a successful implementation of a flipped classroom.

3

Chapter 3

Effects of Self-Regulated Learning Prompts in a Flipped History Classroom

This chapter is based on:

van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2020). Effects of self-regulated learning prompts in a flipped history classroom. *Computers in Human Behavior*, 108, 106318. <https://doi.org/10.1016/j.chb.2020.106318>

Acknowledgement of author contributions:

DvA, CP, JJ, and LK designed the study; DvA collected the data; DvA, CP, JJ, and LK planned the data analysis; DvA and JJ analyzed the data; DvA drafted the manuscript; DvA, CP, JJ, and LK contributed to critical revision of the manuscript; CP, JJ, and LK supervised the study.

Supplementary material to this chapter can be found on:

<https://doi.org/10.34894/C5SRAZ>

Abstract

Flipped learning (FL) is a didactical approach aimed at letting students come to class prepared and apply the learning material actively during class. As FL places a higher demand on students' self-regulated learning (SRL) skills, our goal in the current study was to research the effects of SRL support in a flipped classroom on students' SRL (self-reports and online activities) and learning outcomes. Previous research showed that video embedded SRL prompts enhances students' SRL and learning outcomes. We measured the effects of SRL prompts in a quasi-experimental design in six flipped History classrooms in secondary education where 154 students were randomly assigned to the SRL prompts or no-SRL-prompts condition. We found positive effects of the SRL prompts for the completion rate of the instructional videos (i.e., students in the SRL-prompts condition watched more videos), but not for other indicators of SRL or learning outcomes. Thus, in contrast to previous research from higher education, our results show that implementing SRL prompts in a flipped classroom is not directly effective in secondary education. We address potential explanations for the absence of effects of the SRL prompts for theory and practice into SRL support in flipped classrooms.

Introduction

In a flipped learning (FL) approach, students study learning materials before class (e.g., by watching instructional videos) and apply the content of the learning materials during class. Meta-analyses in which the effect of flipped classrooms was compared to traditional classrooms showed small positive effects on learning outcomes (van Alten et al., 2019; Chen et al., 2018; Cheng et al., 2019; Låg & Sæle, 2019). The meta-analyses also showed substantial and significant variation in the effects of FL on learning outcomes. These studies revealed possible moderating variables affecting the effectivity of FL, such as adding quizzes to the flipped classroom, not reducing face-to-face time, and academic subject.

What is not yet known about FL, is the role that students' self-regulation plays during flipped learning. This is surprising, because self-regulated learning (SRL) seems to be a prerequisite for the success of FL as the individual learning phase before class places a higher demand on students' SRL skills and self-discipline (He et al., 2016). In addition, a meta-analysis found that students' SRL skills improved significantly better in flipped classrooms compared to traditional classrooms (Tan et al., 2017). As flipped classrooms seem to both depend on students' SRL skills and potentially also enhance them, it is worthwhile to investigate to what extent SRL skills should explicitly be supported in a flipped classroom, in which educational contexts, and what kind of support is effective.

In general, supporting SRL is a well-established method to improve student learning outcomes (Dent & Koenka, 2016). Providing students with SRL *prompts* in a computer-based learning environment is one example of an effective strategy to improve students' SRL (Devolder et al., 2012; Wong et al., 2019). In the context of flipped classrooms in higher education, embedded SRL prompts in an instructional video resulted in better learning outcomes and students performing more SRL activities, compared to students who did not receive SRL prompts (Moos & Bonde, 2016). For secondary education students, it is known that providing them with SRL prompts could lead to the internalization of SRL activities and, as a result, improve learning (Greene et al., 2015). In fact, as SRL skills increase at higher ages, it is highly recommended to support secondary education students' SRL to gradually develop their skills (Ramdass & Zimmerman, 2011; Wigfield et al., 2011). However, it is not yet fully known if SRL prompts in flipped classroom videos are as beneficial for secondary school students as for higher education students. Therefore, the current study further investigates the effects of SRL prompts in flipped classroom videos, and aims to reproduce earlier found effects in higher education in the context of secondary education.

Self-Regulated Learning in Flipped Classrooms: Theoretical Perspectives

The common ground of SRL definitions is that students are actively regulating and monitoring their cognition, metacognition, learning behavior, and motivation while learning, and that this involves strategies and goal processes (Panadero, 2017; Puustinen & Pulkkinen, 2001). Metacognition involves knowledge, awareness, and regulation of one's thinking and is one of the central aspects of SRL (Zimmerman & Moylan, 2009). Self-regulated learners have the ability and motivation to apply metacognitive processes such as planning, monitoring, and controlling their own learning and reflect thereon. In the theoretical model we use, SRL is presented as a cyclic model in which three phases (forethought, performance, self-reflection) are distinguished and learners apply metacognitive strategies such as goal setting and planning in the forethought phase, monitoring and self-control in the performance phase, and evaluation and adaption in the self-reflection phase (Zimmerman & Moylan, 2009).

SRL is a relevant addition to current research on FL, as a flipped classroom places a strong demand on students' SRL skills (He et al., 2016) and self-directness (Lee & Choi, 2019) when studying outside the classroom. First, general meta-analytical research on SRL has shown that supporting SRL in primary and secondary education resulted in better learning outcomes (Dent & Koenka, 2016; Dignath & Büttner, 2008; Donker et al., 2014). In addition, there is plenty of evidence from Hypermedia Learning and Multimedia Learning that shows that receiving SRL support enhances students' learning outcomes (Azevedo & Cromley, 2004; Azevedo & Hadwin, 2005; Bannert et al., 2009; Bannert et al., 2015; Kauffman, 2004). Therefore, as most flipped classrooms contain technology-based learning environments, it seems worthwhile to add SRL support in flipped classrooms to enhance learning.

Second, students appear to utilize SRL strategies in online learning contexts better than in traditional learning contexts (Lee & Tsai, 2011; Sletten, 2017). On the positive side, this implies advantages for students in a flipped classroom, as students are able to control the frequency and pace of the instructional videos before class. This could reduce cognitive load which could be beneficial for learning (Clark et al., 2005) and increase student's autonomy in their own learning process which could be beneficial for motivation (Hsu et al., 2019) in comparison to learning material that is presented in a non-manipulable lecture format. In addition, research has found that students who are well able to regulate their own learning process benefit more from a flipped classroom in terms of higher learning outcomes than student's who lack these SRL skills (Lai & Hwang, 2016; Lee & Choi, 2019; Lim & Morris, 2009; Shibukawa & Taguchi, 2019). On the negative side, however, it could be challenging for students to regulate their own learning while preparing the instructional material before class (Sletten, 2017; Wolters et al., 2005). For example, Butzler (2016) illustrated that most students

in a flipped classroom lacked SRL skills (such as reflecting on their own learning) to fully benefit from the flipped classroom structure. Other researchers also found that the quality of students' preparations before class, and with it the engagement of the students during in-class activities, depends on self-regulated and self-directed skills (e.g., time management and self-discipline; He et al., 2016; Lai & Hwang, 2016). SRL skills of students can thus also be seen as an important prerequisite skill that students need to acquire and apply to benefit from the advantages of FL.

Third, SRL skills of students who participate in a flipped classroom seem to improve, because FL places a higher demand on students' SRL skills (Tan et al., 2017). In contrast, Lape et al. (2014) found no significant gains in SRL skills for students in the flipped classroom compared to a traditional classroom. Thus, these inconsistencies in the literature raise the question to what extent teachers and curriculum designers should explicitly support SRL in flipped classrooms in order to make them more effective, in terms of increased SRL skills and better learning outcomes.

Previous Research on Self-Regulated Learning Prompts in Flipped Learning

As it is demonstrated that FL places a higher demand on students' SRL skills while they learn at home, we aim to support their SRL at that particular learning phase (i.e., during the online videos). One effective way to effectively support SRL is to make use of prompts (Zheng, 2016). Prompts are usually given to students in the form of questions and encouragements (e.g., *What have you learned? Explain what you have learned*) to incite students to engage in SRL activities. Providing students with SRL prompts in a computer-based learning environment is an effective strategy to improve students' SRL and learning outcomes (Devolder et al., 2012; Sitzmann et al., 2009; Wong et al., 2019; Zheng, 2016). As online instructional videos are frequently used in a flipped classroom, SRL prompts in these videos could increase students' SRL activity which – in turn – could lead to higher learning outcomes in a flipped classroom.

It should be noted, however, that there could be possible moderators which means that different students groups benefit differently from SRL prompts. For example, research has shown that effects of SRL prompts can be different according to students' cognitive ability (Sitzmann et al., 2009; Wong et al., 2019). It is also known that students with certain cognitive or metacognitive deficits experience difficulties in completing homework (Bryan et al., 2001). Yeh et al. (2010) found an interaction effect between cognitive ability and prompt type, indicating that higher and lower-knowledge learners benefit differently from SRL prompts.

Only a limited number of studies have tested the implementation of SRL support in a flipped classroom. Butzler (2016) found that students worked inefficiently and felt

incompetent (e.g., not knowing how to take notes from instructional videos) when SRL was not explicitly supported. Prompts with specific learning strategies at each learning phase in a flipped classroom could help students to actively plan, monitor, and reflect on their learning (Butzler, 2016; Jovanović et al., 2017).

A study conducted by Moos and Bonde (2016) found that embedded SRL prompts in an instructional video resulted in better learning outcomes and students performing more SRL activities when learning from the video, in contrast to students who did not receive SRL prompts. In addition, no effect of prompts was found on students' perceived mental effort. Moos and Bonde (2016) prompted students in the forethought phase to set goals and strategically plan their learning, in the performance phase to monitor and control their learning, and in the self-reflection phase to evaluate their own learning and set goals. A typical forethought prompt was, for instance: *What strategies do you think will be effective while learning about [the topic of this video]?*; a performance prompt: *What information have you learned so far?*; and a self-reflection prompt: *What could you have done differently while learning about [the topic of this video]?* With these SRL prompts, both the quantity and quality of SRL activities were enhanced by informing students about effective strategies and prompting them at the forethought, performance and self-reflection phase of learning (Moos & Bonde, 2016; Zimmerman & Moylan, 2009).

In addition, two studies showed that students achieved higher learning outcomes when they had access to an additional SRL support learning system, in contrast to students who did not have access. First, Lai and Hwang (2016) compared a flipped classroom in which students were supposed to use a self-regulated monitoring system to set goals and evaluate their own learning with a regular flipped classroom. The students in the SRL supported flipped classroom improved significantly in terms of self-efficacy, strategies of planning study time, and learning outcomes, in contrast to students in the regular flipped classroom. Second, Shyr and Chen (2018) showed that a technology-enhanced language learning system that scaffolded students' SRL was able to increase students' learning outcomes and self-reported SRL as compared to students in a regular flipped classroom. Students in the intervention group who had access to this language learning system received SRL prompts while they selected learning tasks, and by scaffolding, students were supported on orientation, goal-setting, planning, monitoring, and assessing their own learning tasks.

For the following reasons, our study extends the previous research. First, earlier studies that yielded positive effects of SRL support in a flipped classroom compared two flipped groups, where the intervention group had access to an additional SRL support system (e.g., Lai & Hwang, 2016; Shyr & Chen, 2018). However, it is possible that the results they found were due to differences in time-on-task between the

groups with or without the SRL support caused by the additional support system. Therefore, in this study, we investigated the effects of SRL support while keeping *time-on-task* between groups as equal as possible.

Second, there is a lack of data regarding FL in *secondary education* as most research on SRL support in FL is conducted in primary education (e.g., Lai & Hwang, 2016) or higher education (e.g., Moos & Bonde, 2016). It is important to know if earlier results of explicit support of SRL by prompts apply to secondary education. It is clear that students of different age differ in SRL skills, and that SRL skills increase at higher ages (Dent & Koenka, 2016; Wigfield et al., 2011; Zimmerman & Martinez-Pons, 1990). The fact that students' confidence in their SRL skills from primary to secondary education decreases (Wigfield et al., 2011), indicates extra importance to apply SRL support in secondary education. Similarly, Ramdass and Zimmerman (2011) argued that it is important that students from all ages gradually learn to develop SRL by repeated practice, as it enhances learning. Therefore, we conducted our study in secondary education as it seems even more important that SRL of younger students in flipped classrooms is supported.

Third, the current study took place in an *ecologically valid classroom setting* and aimed to investigate effects over the course of six weeks. Previous research often focuses on short term interventions of SRL prompts and sometimes in an out-of-school, laboratorial learning environment with a restricted learning time (cf. Bannert et al., 2015; Lee et al., 2010; Moos & Bonde, 2016).

Present Study

In the present study, we investigated the effect of explicit support of SRL with prompts in a flipped classroom, on students' SRL self-reports, online SRL activity, and learning outcomes. We answer the following research questions:

1. What is the effect of SRL prompts in a flipped classroom on secondary education students' (a) SRL self-reports and (b) online SRL?
2. What is the effect of SRL prompts in a flipped classroom on learning outcomes of students in secondary education?
3. Does students' performance level moderate the effects of SRL prompts?

With regard to the first research question, we operationalized SRL in two ways: (a) as a self-reported construct and (b) as online log data measures of SRL activities (Panadero et al., 2016; Rovers et al., 2019). Thus, we measured SRL in two ways: by asking students how they perceived their own SRL and by investigating SRL activities that are traceable in online log data. Jovanović et al. (2017) showed that it is possible

to identify actions in students' learning behavior as manifestations of their learning strategies. This is considered to be an important addition to self-reports, as it aims to measure interaction traces of students' SRL behavior in a natural setting (Jovanović et al., 2017; Panadero et al., 2016; Winne, 2014). For the second research question, we defined learning outcomes as students' score on a learning outcome test similar to tests that are regularly used in secondary education to assess their learning. With regard to research question three, we used performance level in a moderator analysis to investigate a possible interaction effect between performance level and condition. We defined performance level as the students' average scores for History that current schoolyear. An interaction between performance level and the research condition would mean that SRL prompts work better for students with a higher, or lower, performance level. Lastly, we evaluated students' answers on the satisfaction questionnaire and their answers on a semi-structured focus group interview to better interpret our findings with regards to their experiences in our research intervention.

All in all, our hypothesis is that students who are regularly prompted to perform SRL activities during their learning process, will (a) report higher SRL self-report values, (b) show more SRL activities during online learning, and (c) will achieve higher learning outcomes, in comparison with students that are not provided with SRL prompts.

Method

Participants

A total of 169 second year students of (1) senior general and (2) pre-university level within Dutch secondary education were asked to participate in this study (including opt-out consent from their parents). The Netherlands has a tracked secondary education system (from grade 7 onwards) and these are the two highest tracks. Ethical approval was obtained from the Faculty Ethics Review Board in April 2018. The study took place in one school and the intervention was part of the students' regular school curriculum. Students who decided not to participate in this research still had to attend the lessons and complete the final test, but their results were not analyzed by the researchers. One voucher worth of 10 euros was raffled amongst the students in each participating class, while every student in every classroom received a small gift at the end of the study. Eight students were excluded because consent was not obtained and seven students were excluded because they missed more than half of the lessons and/or they were absent when the measurements were taken. Consequently, our sample consisted of 154 students ($M_{\text{age}} = 13.49$ years, $SD_{\text{age}} = 0.50$; 84 boys and 70 girls). The sample came from six pre-existing classrooms: two senior

general level classrooms ($n = 44$) and four pre-university level classrooms ($n = 110$). A priori power analyses indicated that, for sufficient power of 80% and detecting medium to large effects as expected by previous research, we needed to have a minimum sample size of 152 students. Thus, our final sample size of 154 students was adequate.

The first author was one of the four teachers. The other three teachers volunteered to participate and each received a voucher worth of 25 euros at the end of the study. The average teaching experience of the participating teachers was 4 years, and their experience with teaching flipped classrooms was comparably little.

Design

The quasi-experiment used a between-subjects design with embedded SRL prompts in the instructional videos in the intervention group, and no embedded SRL prompts in the control group. The experiment consisted of eight lessons (and seven videos) over the course of six weeks. Stratified randomization was used to allocate students within every class to the SRL-prompts condition ($n = 74$) or the no-SRL-prompts condition ($n = 80$). The students' average scores for History (recalculated in a standardized rank score within each classroom) were used to match students and make sure both groups were comparable in terms of performance level. The rationale for this stratified randomization within each classroom was to control for possible differential effects on the classroom level (e.g., teacher effect, cf. Nye et al., 2004).

Online Learning Environment

Edpuzzle (<https://edpuzzle.com/>) was used as online learning environment to share the videos with the students. We provided each student with a personal login that was linked to their corresponding research condition (video with or without SRL prompts). The instructional videos were identical for both groups, had an average length of seven minutes and were developed and recorded by the first author. The school's digital homework system was used to announce which video had to be viewed in preparation before class.

In Class Learning Materials

The subject of the lessons developed for this intervention was the industrial revolution during the late 19th and early 20th century, a compulsory part of the curriculum at the senior general and pre-university educational levels. The in-class learning materials were mostly based on the schoolbooks that are used for teaching History in the participating school. In addition, every student was provided with a physical workbook which included instructions about Edpuzzle, lesson procedures, and three to four corresponding cognitive questions (in both conditions) for each instructional

video that was required to complete before class. Teachers discussed the cognitive questions at the start of each lesson to recall the content of the instructional videos. Additionally, we developed assignments in which students were actively engaged to apply the content of the learning materials (e.g., analyzing historical sources and creating advertisement leaflets to explain the benefits of 19th century industrial inventions). Participating teachers were able to give feedback on all the learning materials before the intervention started. As the students were allocated to research conditions within classrooms, all the in-class learning materials were exactly the same for both research conditions.

Self-Regulated Learning Prompts

The students in the SRL-prompts condition received three to four SRL prompts per video (similar to Moos & Bonde, 2016, see Appendix A) according to the three SRL phases of forethought, performance, and self-reflection (Zimmerman & Moylan, 2009). For example, in the forethought phase one of our SRL prompts was: *How do you make sure that you will be able to recall all the learning material from this instructional video?* The practical example that students received after answering this prompt was: *For example, you can regularly pause the video and make notes to summarize the learning material.* An example of a SRL prompt in the performance phase is: *Do you understand everything you heard so far? If not, what can you do to fix it?* Lastly, a typical SRL prompt in the self-reflection phase at the end of the instructional video looked like this: *Did you achieve your goal(s) you set at the beginning of this instructional video? If not, what do you need to complete it?* An overview of all the SRL prompts per video, focusing on the whole process of SRL in each SRL phase is presented in Appendix A (cf. Zheng, 2016 on supporting each aspect of SRL instead of one specific learning phase). The prompts were presented to students as forced open question stops within the video. Students were required to think about the SRL prompts and answer them to continue the video. In around half of the SRL prompt, a specific hint popped up after a student's answer to show a practical example of that SRL behavior (see Appendix A). Students in both research conditions were not able to fast forward the video, but were allowed to rewind and re-watch the videos (or portions of the videos) unlimited.

Measurements

Self-Regulated Learning Log Data

Edpuzzle was also used to collect SRL log data. We created the following four variables as indications of SRL activity: Students' *video completion rate*, *video watch time*, *video portions viewed*, and students' *rewind actions* (i.e., how many times a student reviewed a portion of the video). *Video completion rate* and *video watch time* can be seen as traces of SRL activity, as Cicchinelli et al. (2018) showed that students with high SRL

accessed online learning platforms more and spent more time on it (cf. Sitzmann & Ely, 2010). *Video portions viewed* and *rewind actions* can be seen as manifestations of learning regulation activities, as students take action to acquire or reinforce knowledge (Bannert et al., 2014; Cicchinelli et al., 2018; Kizilcec et al., 2017; Maldonado-Mahauad et al., 2018).

In addition, to evaluate if the students in the SRL-prompts condition properly worked with the prompts, we collected their answers that they had to provide to continue the video. We counted how many valid (e.g., serious) answers were given by students who watched the videos and calculated an average per video to assess *SRL-prompt compliance*.

Self-Regulated Learning Questionnaire

We measured SRL also by a contextualized version of the revised version of the self-regulated online learning questionnaire (SOL-Q-R) with a pretest-posttest design (Jansen et al., 2018; Jansen et al., 2017). The revised version of this questionnaire contains the following seven scales: *metacognitive activities before, during, and after learning*, *time management*, *environmental structuring*, *persistence*, and *help seeking*.

We translated the SOL-Q-R, that was originally constructed for online learning in Dutch, with a student population aged 14-15 years in mind. Potential difficult words (such as: *to reflect*) were explained at the bottom of the questionnaire. We substituted references to *online course* in the questionnaire into *homework for History* (e.g., I think about what I have learned after I finish working on homework for History), to adjust it to the educational context and to be able to use it as pretest as well. As can be seen in Table 1, the scales of our contextualized version of the SOL-Q-R were reliable, and, in addition, the reliabilities were very similar to the three datasets used to validate the original questionnaire in the context of blended learning in higher education (Jansen et al., 2018).

Table 1*Internal-Consistency Reliabilities of the Contextualized SOL-Q-R Scales*

Scale	<i>n</i>	pretest		<i>n</i>	posttest	
		Items	α		Items	α
Metacognitive activities before	146	7	.78	131	7	.80
Metacognitive activities during	149	7	.74	136	7	.80
Metacognitive activities after	142	6	.83	133	6	.85
Persistence	142	7	.90	136	7	.90
Help seeking	149	6	.74	130	6	.82
Environmental structuring	151	4	.77	133	4	.79
Time management	146	5	.74	129	5	.67

Motivation Questionnaire

Motivation was measured as a pretest only to investigate if the students in the research conditions were equal in terms of motivation. We used four relevant motivation scales from the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al., 1991). We also contextualized this questionnaire to the context of completing homework for History, to be able to measure students' motivation in this particular context as precise as possible. *Intrinsic goal orientation*, *extrinsic goal orientation*, and *task value* were used to measure how students valued and perceived their engagement into a particular learning task (i.e., homework for History). *Self-efficacy for learning* was included to measure students' expectancy beliefs about how they complete their History homework. As can be seen in Table 2, the scales of our translated and contextualized version were acceptable except for the *extrinsic goal orientation* scale. We nevertheless kept this scale because the value is still above .5 (which would be unacceptable) and the scale is only used to check for research group equivalence.

Table 2*Internal-Consistency Reliabilities of the Contextualized MSLQ Scales*

Scale	<i>n</i>	Items	α
Intrinsic goal orientation	151	4	.76
Extrinsic goal orientation	149	4	.63
Task value	141	6	.91
Expectancy: self-efficacy for learning	141	8	.89

Learning Outcome Test

The researchers and involved teachers developed a learning outcome test based on the learning materials that are used for teaching History in that school and comparable to what students regularly have to complete during the schoolyear. Every participating student received the same test. The test consisted of 25 items, of

which five were multiple choice (four answer options each). The 20 open questions contained both recall (i.e., reproduction), comprehension (i.e., understanding), and transfer (i.e. applying) assignments. For example, a recall question was: *Explain the following concepts: Cottage industry, urbanization and capitalism.* A comprehension question was: *Explain why a rapid need arose for the Spinning Jenny during the Industrial Revolution.* A transfer level question was, for instance: *Give an argument why the following source represents the working conditions in the 19th century.* In general, the test was found to be reliable (25 items; $\alpha = .79$).

Student Satisfaction

We evaluated how the students valued the flipped classroom method as a means to better interpret our findings. We added seven Likert-items (range from 1 to 7) to the posttest as an anonymized satisfaction questionnaire (e.g., I think the instructional videos are boring; I think the instructional videos are instructive; I believe that the instructional videos are designed in a clear and understandable way). The satisfaction questionnaire was found to be reliable (7 items; $\alpha = .87$). Besides, we asked students suggestions for improvements in the videos, and if they want to share any other compliment or complaint about the FL method, in two open answer questions.

In addition, we also conducted small-scale semi-structured focus group interviews to have a better understanding of experiences of students and to aid the interpretation of our results. The first author conducted three semi-structured interviews with three randomly chosen students (on voluntary basis). Two interview sessions were done with three students from the SRL-prompts condition each, and one session with three students from the no-SRL-prompts condition). Each focus group interview lasted around 10 minutes and students were asked to evaluate their experiences in this flipped classroom, the in-class materials, and the homework (instructional videos and accompanying questions). In addition, students from the SRL-prompts condition were specifically asked how they valued the SRL prompts.

Procedure

Before the flipped intervention started, the students completed the pre-test questionnaire with the SRL and motivation items during regular class time. At the start of the intervention, students received a brief instruction about working in a flipped classroom and a small rationale behind it, as for most students it was their first experience. In addition, students received a classroom instruction on how to access the videos at home.

Students were informed that their own teacher could track their progress. Teachers were instructed not to change their usual teaching interaction with their students,

and received instructions and training on how to apply the learning materials each lesson. The teachers used the school's digital homework system to inform students of their own class about the required preparations for each lesson (one instructional video and the accompanied content questions in their workbooks). Each lesson of 65 minutes started with retrieval practice activities about the content of the videos and a discussion of the content questions from the workbooks. This was usually followed by a complementary micro lecture of approximately 10 minutes to discuss not yet fully understood parts and more in-depth aspects of the learning materials. The largest and remaining part of the lesson was dedicated to engaging learning activities to apply the learning materials. After eight lessons, students completed the posttest SRL and satisfaction questionnaire during regular class time. In the subsequent lesson, students had 65 minutes to complete the learning outcome test.

Data Analysis

For the SOL-Q-R and satisfaction questionnaires we calculated scale means of the relevant items, where all students were included if they filled in at least $n - 1$ items of that scale. Log data from Edpuzzle was used to calculate four SRL activity variables. For *completion rate* of the videos, we calculated the mean of the video completion rate of each individual student. For students' *video watch time* on and *video portions viewed*, we added up the individual scores of each student on these variables. For students' *rewind actions*, we counted and added up all *rewind actions* from all videos.

We collected the students' average scores for History that current schoolyear as indication of performance level and recalculated that in a standardized rank score within each classroom. To include performance level as covariate in our analyses, we recalculated rank scores into standardized Z-scores within each classroom to be able to make a fair comparison between performance level of individual students.

For the learning outcome test a comprehensive assessment scheme was validated by the researchers and teachers to score the students' tests. Each teacher scored the tests of their own students and we created a moment early in the scoring phase in which the teachers discussed cases where they doubted their scoring of an answer to an open question. To assess the interrater reliability of the open question scores, the first author independently scored a random sample of five unscored tests from the other classrooms (20 tests in total). We compared the scores of the first author (400 items in total) with the scores of the corresponding teacher from that classroom. After evaluating the Cohen's κ for each item, the agreement percentage, and the item total correlation, we deleted five items because of low reliability (i.e., $\kappa < .55$). This resulted in 15 remaining open questions with an average total agreement of 85%, and a moderate to substantial interrater-reliability with an average Cohen's κ of .74

(range $\kappa = .55$ to 1 , $p < .001$). Finally, we created a sum score of the 15 remaining open answer questions and the 5 multiple choice questions, resulting in a maximum score of 24 points for the learning outcome test (as some open answer questions had a greater weight on the test).

Finally, to answer the first two research questions (i.e., is there a difference between the SRL prompts and no-SRL-prompts conditions on SRL self-reports, SRL activities, and learning outcome), we checked the data for normality and outliers. Due to the linear relationships between the dependent variables, as shown in the correlation matrix in Table C.1 (Appendix C), we analyzed the differences between both research *conditions* by two separate MANOVAs (i.e., SRL self-report scales and SRL log data) and one ANOVA (i.e., learning outcomes) in SPSS. As the SRL log data variables were not normally distributed, we performed a bootstrapping method (1000 Bootstrap samplings with replacement, cf. Field & Wilcox, 2017). In addition, we included educational level (i.e., senior general or pre-university level) in the model as controlling factor.

To answer the third research question and examine if performance level moderated the findings, we added students' performance level as covariate and checked for possible interaction effects between performance level and condition, which would mean that the SRL prompts had a differential effect according to how well students usually performed in History. We also included educational level in this model as controlling factor.

Lastly, to better interpret our findings, we compared the answers on the satisfaction questionnaire with the answers on the semi-structured focus group interviews. These interviews were transcribed and subsequently analyzed by recurring themes such as students' experiences of the flipped classroom in terms of advantages and disadvantages, learning from the videos before class, and, for the SRL-prompts condition, how they valued the prompts.

Results

Independent samples *t*-tests revealed no significant difference ($p > .05$) between the students in the SRL prompts and no-SRL-prompts conditions *before* the intervention for the pretest SOL-Q-R scales, the MSLQ scales, and students' performance level (standardized average of students' current History scores).

In our flipped classroom and research design, videos were used to convey important subject matter before class. Students were instructed to independently watch these

videos out of the classroom. It is important to know to what extent the students actually watched these videos. First, in order to know the amount of subject matter they were exposed to before class. Second, and more importantly, to evaluate the effect of non-compliance to the intervention on our results (cf. He et al., 2016; Müller & Seufert, 2018). For example, non-compliance of participants in a research intervention could hamper statistical power to detect intervention effects (Jo, 2002). In our case, a low *video completion rate* for the students in the SRL-prompts condition means that they were not substantially exposed to our intervention. In fact, it would mean that the students in the SRL-prompts condition who did not watch the videos are more similar to students in the control group in the sense that they also did not receive the SRL prompts. To summarize, students who had a low *video completion rate* could hinder a fair comparison between both conditions.

The *video completion rate* variable showed that 17 students did not watch any videos before class, 67 students watched all videos, and 105 students (77%) watched at least five out of seven videos. We decided to answer our research questions and perform our analyses on the group of students (both research conditions) who watched a substantial amount of the videos (more than two-thirds of the videos). We chose five out of seven videos as our threshold as this is comparable to the 80% threshold in other research dealing with non-compliance (Armijo-Olivo et al., 2009). In Table 3, we present the descriptive statistics for all the dependent variables for this group of students. Because the exclusion of non-compliance students could introduce a selection bias (Armijo-Olivo et al., 2009; Fergusson et al., 2002), we present our results on the entire group of students in Appendix B for transparency purposes.

Lastly, we assessed *SRL-prompt compliance* by the average amount of valid answers students provided in response to the prompts in order to continue the video. The average valid answers per video was 90% and ranged from 85% (video 7) to 95% (video 5). This indicates that students in general properly worked with the prompts.

Effects of Self-Regulated Learning Prompts

Self-Regulated Learning Self-Reports

First, a MANOVA that included condition (SRL-prompts versus no-SRL-prompts) and educational level (senior general versus pre-university) as factors and the seven SOL-Q-R scales as dependent variables was used to investigate if students who received SRL prompts reported higher SRL self-reports than students who did not receive SRL prompts (RQ1a). We found no significant main effect of conditions on the SOL-Q-R scales, $F(7, 91) = 2.01$, $p = .062$; Wilk's $\Lambda = 0.866$, partial $\eta^2 = .13$. We are aware that this omnibus test is not significant but we decided to present the results of the follow up ANOVA for completeness reasons. Due to non-compliance

exclusion, our analyses lost some power and these post-hoc tests have more power to detect differences. These results, nevertheless, should be interpreted with caution and seen as explorative result interesting for future research. ANOVA revealed a significant difference between conditions on the SOL-Q-R *help seeking* scale, $F(1, 97) = 6.22, p = .014$, partial $\eta^2 = .06$. Students who received SRL prompts reported that they performed more *help seeking* activities while completing their homework than students who did not receive the prompts (see Table 3 for the descriptive findings).

Furthermore, we found a significant effect of educational level on the SOL-Q-R scales, $F(7, 91) = 3.31, p = .003$; Wilk's $\lambda = 0.797$, partial $\eta^2 = .20$. Follow up ANOVAs showed that for the SOL-Q-R scales *help seeking* ($p = .014$, partial $\eta^2 = .04$) and *time management* ($p = .002$, partial $\eta^2 = .09$), students in the pre-university level scored significantly higher (*help seeking*: $M = 3.80, SD = 1.22$; *time management*: $M = 4.93, SD = 0.99, n = 80$) than students in the higher senior level (*help seeking*: $M = 3.17, SD = 1.03$; *time management*: $M = 4.16, SD = 0.91, n = 21$).

Self-Regulated Learning Online Activity

Second, a MANOVA with condition and educational level as factors and *video completion rate* and *rewind actions* as online SRL activity dependent variables was used to investigate if students who received SRL prompts performed more online SRL activities than students who did not receive SRL prompts (RQ1b). We dropped *video watch time* and *portions viewed* from the analysis because of a very high correlation ($R > .96, p < 0.01$, see Appendix C) and the problem of multicollinearity. We found a significant main effect of conditions on online SRL activity, $F(2, 100) = 3.41, p = .037$; Wilk's $\lambda = 0.936$, partial $\eta^2 = .06$. Results of the follow up ANOVAs determined a significant difference for the SRL prompt and no-SRL-prompts conditions on the online SRL activity variable *video completion rate*, as the students in the SRL prompts had higher *video completion rates* than students who did not receive SRL prompts, $F(1, 101) = 6.88, p = .010$, partial $\eta^2 = .06$. We found no significant main effect of educational level on online SRL activity, $F(2, 100) = 1.35, p = .264$; Wilk's $\lambda = 0.974$, partial $\eta^2 = .03$.

Table 3

Mean and Standard Deviation for the Dependent Variables (Included Students who Watched at least 5 out of 7 Videos)

Dependent variables (posttest)	SRL-prompts condition			No-SRL-prompts condition		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
<i>SRL questionnaire</i>						
Metacognitive activities before	51	3.51	1.17	51	3.45	1.13
Metacognitive activities during	51	3.69	1.17	52	3.50	1.14
Metacognitive activities after	51	3.10	1.23	52	2.80	1.17
Persistence	51	4.87	1.18	52	4.70	1.34
Help seeking	51	3.97	1.19	52	3.35	1.14
Environmental structuring	51	5.37	1.40	51	5.41	1.41
Time management	51	4.92	0.96	52	4.64	1.04
<i>SRL activity</i>						
Video completion rate	51	95.66	7.37	54	92.67	10.44
Video watch time	51	68.24	22.56	54	66.35	21.46
Video portions viewed	51	92.27	29.80	54	85.07	27.86
Rewind actions	51	25.18	27.96	54	20.17	24.66
<i>Learning outcome</i>						
Learning outcome test	51	16.72	3.87	54	16.13	3.91

Learning Outcomes

Third, an ANOVA on the learning outcome test scores as dependent variable with condition and educational level as factors was used to investigate if students who received SRL prompts obtained higher learning outcomes than students who did not receive SRL prompts (RQ2). We found no significant main effect of condition on learning outcome; $F(1, 101) = 0.33, p = .569$, partial $\eta^2 = < .01$, but we did find a significant main effect of educational level on learning outcome; $F(1, 101) = 7.71, p = .007$, partial $\eta^2 = < .07$. Students in the senior general level scored 12.92 points on average (12 points required to get a passing grade, $SD = 3.40$) and students in pre-university level scored on average 16.20 points (13.2 points required for a passing grade, $SD = 4.07$). The total fail rate of the test was 30%, which indicates that the test was not too difficult nor too easy (e.g., an easy test could possibly result in a ceiling effect).

Moderation of The Effects of Self-Regulated Learning-Prompts by Performance Level

While educational level contains information about students' abilities on the classroom level, performance level (i.e., the students' average scores for History that current schoolyear) contains information about students' abilities on the individual level for a particular subject matter. Due to matching, both research conditions contained students of equal performance levels. The possible interaction effect between performance level and condition on the dependent variables would indicate that the SRL prompts have different effects for students with a higher or lower performance level.

Thus, we performed two separate MANCOVAs and one ANCOVA with performance level as a covariate, condition type and educational level as factors, and the SRL questionnaire scales, the SRL online activity variables, and the learning outcome test as dependent variables to investigate if performance level moderated the effects of SRL prompts (RQ3). In Table C.2 (Appendix C) we present the correlations between performance level and the dependent variables. We found no significant interaction effects between performance level and condition on the SRL self-reports, $F(7, 90) = 1.06, p = .394$; Wilk's $\lambda = 0.924$, partial $\eta^2 = .08$; no significant interaction effects between performance level and condition for the SRL online activities, $F(2, 99) = 0.34, p = .713$; Wilk's $\lambda = 0.993$, partial $\eta^2 = .01$; and no significant interaction effects between performance level and condition for learning outcome; $F(1, 100) = 1.33, p = .252$, partial $\eta^2 = < .01$. This shows that the effect of the SRL prompts is not different for students' performance level.

Student Satisfaction in the Flipped Classroom

The effectiveness of the SRL prompts is related to the design of the flipped classroom and how students valued it in general. Therefore, we evaluated the results from the satisfaction questionnaire to interpret students' evaluations of our flipped classroom implementation, as negative evaluations of our research setting could potentially affect the outcomes.

We decided to include all students (including the non-compliant students) in the analysis of satisfaction, as it could be that *video completion rate* was dependent on students' satisfaction about the flipped classroom. The correlation between *video completion rate* and the satisfaction questionnaire was $r(140) = .23, p = .006$. We performed an explorative ANCOVA with condition and educational level as factors and controlling for *video completion rate* as covariate on the total score of the satisfaction questionnaire. After controlling for *video completion rate*, the ANCOVA showed no main effect of condition on satisfaction, $F(1, 136) = 0.04, p = .833$, partial $\eta^2 < .01$, but a significant main effect of educational level on satisfaction, $F(1, 136) = 17.59$,

$p < .001$, partial $\eta^2 < .12$. Students in the pre-university level ($M = 36.17$, $SD = 8.23$, $n = 103$) valued the flipped classroom and the instructional videos significantly more positively than the students in the senior general level ($M = 28.62$, $SD = 8.47$, $n = 37$), with a maximum score of 49. All in all, students were generally positive about the flipped classroom. This is further illustrated with the findings presented in Table 4. It shows the results of the open questions in which all the students were asked for improvements in the videos, or provide any other compliment or complaint about the FL method. The majority of students answered that they had no specific feedback and relatively few students provided feedback.

Finally, the semi-structured focus group interviews showed that students in general were satisfied with the flipped course and would like to use this method more often. For example, they valued the variety this method adds to regular teaching methods and the quality and organization of the course. Some students expressed that the videos helped them understand the learning material better, and that it reduced the time they usually spend to History homework. In contrast, another recurring theme was the recommendation to shorten the videos, and add some variety in the style of the videos. When asked about their experience of the accompanying lessons, some students remarked that the lessons contained a high amount of learning activities where they had to actively apply the learning material, but that these activities were also present in the usual lessons before the flipped course. Students who admitted that they did not watch (all of the) online instruction videos gave as reason that they usually lack motivation to complete homework, also for their homework of other subjects. Some students from the SRL-prompts condition complained about the SRL prompts. They felt that these prompts distracted them from learning the content. When asked for clarification, they said that they are not used to think about regulating their own learning, and that they do not see the added value of this activity. The content questions in their workbooks were positively valued, as they liked to use them to make sure if they understood the learning content.

Table 4

Results of the Open Answers from the Satisfaction Questionnaire, n = 154 Students. In Parentheses is Given how Often a Comparable Answer was Given

Is there something that can be improved in the videos? (e.g., anything disturbing, unclear?)	Do you have any other compliments or complains?
Too long and/or boring (18)	Videos were instructive and/or comprehensible (12)
Face cam distracting (14)	Better than usual teaching method (12)
Negative comment about the prompts (9)	Nice method (10)
Lacks entertainment (7)	Questions accompanying the videos help me to focus (2)
Quality of the microphone (5)	Method motivates me to complete my homework (2)
Simpler language (3)	Prefer the usual teaching method (2)
Not able to fast forward videos (2)	Well organized (2)
	Method requires less time to learn (2)

3

Discussion

The main purpose of the current study was to investigate the effect of video embedded SRL prompts on students' SRL (self-reports and online activities) in a flipped history classroom over a period of six weeks. We found that students provided with SRL prompts did not significantly improve in terms of SRL and learning outcomes in comparison with students who were not provided with the SRL prompts. These results could have implications for both theory and practice, as previous research in higher education found large effects on students' SRL and learning outcomes by implementing SRL support (e.g., with video embedded prompts) in FL (cf. Moos & Bonde, 2016).

However, our results show that SRL support does not yield similar results in the context of secondary History education. We will address possible explanations to explain these findings related to, for example, the research context, type of prompt implementation, methodological aspects, or a combination of them.

Did the Self-Regulated Learning-Prompts Affect Self-Regulated Learning Self-Reports and Self-Regulated Learning Activity?

Self-Regulated Learning Self-Reports

In contrast to our expectations based on previous research (Lai & Hwang, 2016; Lee et al., 2010; Moos & Bonde, 2016; Shyr & Chen, 2018), the results on the SOL-Q-R posttest showed no significant effects of the prompts. Although we need to be

cautious to interpret this finding, as the MANOVA omnibus test was not statistically significant ($p = .06$), we saw a significant difference between the SRL and no-SRL-prompts conditions on the *help seeking* scale. It can therefore not be ruled out that this finding can be seen as a false positive. However, as we lost some power due to the exclusion of 23% of the students due to non-compliance, it is worth to at least discuss this difference to direct further research. For example, Sun et al. (2016) investigated the effect of FL (in contrast to non-FL) on SRL, and also only found an effect on *help seeking*. In our intervention, it would mean that students in the SRL-prompts condition were more inclined to ask questions to others during their learning process than students without the prompts. As can be seen in the overview of prompts (Appendix A), a few prompts explicitly prompted students to ask questions to others when they felt they needed help. It could be possible that students without the SRL prompts paradoxically perceived more control and competence while learning from the videos, which gave them less urge to ask questions. It is also possible that due to the Dunning-Kruger effect students in the no SRL prompts group overestimated their own learning (and asked less questions), because they were less aware of their own learning process as they did not receive SRL prompts. Students in the SRL-prompts condition, in contrast, were explicitly prompted to think about their learning process and were therefore probably more aware of what they did not understand, and possibly applied the help seeking strategy more often.

Self-Regulated Learning Online Activity

The results of students' online SRL activity only showed a significant difference in *video completion rate*. This means that students in the prompts condition watched more videos than students in the no prompts condition. This effect could indicate that the SRL prompts made students more aware of their learning process and that they therefore completed their homework more consistently than students without the SRL prompts. This could, for instance, be caused by SRL prompts that explicitly connected the coherence of the content across the homework videos.

However, we also expected to find differences in the other SRL activity indications we measured (such as *video watch time*, *video portions viewed*, and *rewind actions*; Bannert et al., 2015). In previous research, for example, Sitzmann and Ely (2010) found that positive effects of SRL prompts on learning outcomes were mediated via the total time students spent on learning. Our results did not show any significant difference for *rewind actions*. Our SRL prompts explicitly encouraged students to monitor their understanding of the content, and were frequently prompted to rewind video portions (there were 70 video portions to be viewed in total) when they would encounter difficulties. In our results, we see high standard deviations for the *rewind actions* for both conditions, and that students with SRL prompts on average rewound five more video portions than students without the prompts. This indicates that, independent

from the research condition, there is a group of students who hardly rewind video portions, in contrast to a group of students who rewinds video portions multiple times. Although we did not find any statistically significant effects of the SRL prompts on these data, we can carefully draw the conclusion that our FL intervention (even without the SRL support) stimulated (at least a part of) the students to monitor their own learning and apply SRL. It is worthwhile for future research to investigate this contrast in learning behavior on the student level, to investigate potential explaining factors at this level.

It should also be noted that *video watch time*, *video portions viewed*, and *rewind actions* are very highly correlated $R > .96$ (see Appendix C). This can be explained as the *video watch time* is only increased if a student rewound a portion of a video. The fact that we had to drop two of these variables from the analysis due to multicollinearity seems a minor problem, as it appears to measure one underlying SRL concept of deliberately acquire or reinforcing knowledge (Bannert et al., 2014).

Did the Self-Regulated Learning-Prompts Enhance Learning Outcomes?

In contrast to previous research (Bannert et al., 2009; Lai & Hwang, 2016; Moos & Bonde, 2016; Sitzmann et al., 2009; Sitzmann & Ely, 2010), we did not find differences in learning outcomes between the students in the conditions with and without SRL prompts. First, Müller and Seufert (2018) suppose that the lack of effects on learning outcomes could be due to student compliance. As example, they show that student compliance is crucial for the effects of prompts, and, that this compliance can be increased when students become more involved (e.g., they can choose the best applicable prompts in their situation) in the design of the prompts (Bannert et al., 2015). We cannot rule out that this could have affected our study results, as our students all received the same prompts at the same time. Besides, in our subsample of students who completed most of the videos, we can only ascertain that students at least saw the SRL prompts, but we did not have control over the extent to which they really applied the SRL prompts in their learning process.

Second, it could be the case that the effect of SRL prompts is dependent on the type of learning outcome that is measured. Previous SRL prompt research, where a clear distinction is made between recall, comprehension, and transfer (i.e., applying the learning material to new contexts), usually find significant differences in learning outcomes for transfer (Bannert et al., 2009, 2015; Müller & Seufert, 2018). In our present study, we made use of an ecologically valid and generic test, in which this categorically difference is hard to make for individual items.

Did Performance Level Moderate the Effect of the Self-Regulated Learning Prompts?

Our last research question was if students' performance level (i.e., a students' average score for History that current schoolyear) moderated the effects of SRL prompts. Although we matched students according to their individual performance level to make comparable research conditions, we hypothesized that students with different performance levels within the research conditions could experience differential benefits of SRL prompts. In contrast to previous research (Sitzmann et al., 2009; Wong et al., 2019; Yeh et al., 2010), we did not find that the SRL prompts had different effects for students with different performance levels in our context. It could be possible, however, that other measurements (e.g., intelligence, pre-knowledge) are more suitable to investigate individual differences on cognitive abilities. For example, research has found that students with higher prior knowledge perform more online SRL activities than students with lower prior knowledge (Taub et al., 2014).

However, we did find differential effects for the different educational levels on SRL self-reports and learning outcomes. While performance level indicates an individual student's performance in History, educational level is a more generic indication of a student's ability. We found that students in the pre-university level scored higher on the *help seeking* and *time management* scale, on the learning outcome test, and were more satisfied with the flipped environment than students in the senior general level. We also found that students in the senior general level on average watched 55% homework videos compared to 80% of the videos watched by pre-university students. The findings are in line with research that shows that SRL and motivation are important predictors of school achievement for the current group of students (Pintrich & de Groot, 1990; Vukman & Licardo, 2010). It shows that students in the highest educational level not only perform better in the learning test because of higher cognitive abilities, but it also explains the higher SRL self-report scores and the higher persistency and motivation to complete the homework videos. Furthermore, as these higher achievers have more SRL skills, they are likely to be more satisfied in a flipped classroom that makes greater use of their ability to self-regulate their learning. We recommend to take this difference in educational level into account in future research on SRL prompts in a flipped classroom setting.

Strengths and Limitations: Implications for Theory and Practice

First, an important difference between our study and previous research is the educational context (student level and age). Previous research in primary and higher education showed positive effects of SRL prompts in (FL) learning environments, but we did not find such effects in secondary education. We know that students' age is an import factor in the development of SRL skills (Wigfield et al., 2011; Zimmerman

& Martinez-Pons, 1990). It could be possible that some of the younger students lack certain metacognitive skills, or do not have the motivation to apply them. Also, additional SRL strategy instruction is needed for these students in secondary education to effectively work with the SRL prompts. For example, nine students from the SRL-prompts condition expressed negative comments about the prompts when asked about disadvantages in the satisfaction questionnaire. In the focus group interviews, some students from the SRL-prompts condition explained that they sometimes felt that these prompts distracted them from learning the content and that they do not see the added value of this activity. It could be that they are not aware that they are actually applying SRL methods, do not fully understand the connection with the prompts, and sometimes even see them as additional (and distracting) workload. The fact that the interviewed students are clearly positive about the content questions accompanying the videos in their workbook, and a large group of students in general rewind video parts relatively often, illustrates that students did put effort in monitoring their learning. This implies that making the *why* of the SRL prompts explicit to students (e.g., by adding instruction or training) could be an important point of improvement for future research (cf. Jansen et al., 2020). The introduction about working in a flipped classroom in the current intervention could, for instance, be supplemented with an in-class training on SRL skills to make sure the students have a basic level and understanding of SRL skills in this particular setting (Bannert et al., 2009; Bol et al., 2016; Kistner et al., 2010).

Second, the analysis of students' satisfaction showed that there was a difference for educational level on the satisfaction questionnaire, as students in the higher educational level valued the flipped classroom learning environment better. The average satisfaction of the students from the senior general level, however, was still acceptable (with a mean score of 29 out of 49). Given the positive results of the satisfaction questionnaire and the focus group interviews, we conclude that students' general experiences with the flipped classroom have not been the cause of the absence of effects of the SRL prompts.

Third, our methodological design contains both strengths and weaknesses that could have affected the results. In contrast to previous research, we developed an intervention of six weeks in an ecologically valid classroom setting, to further test effects that were found in short term research designs. The advantage of randomization within groups was a likely elimination of between group differences and confounders (e.g., teacher, classroom, educational level, and other circumstantial effects). A possible limitation of this is that we had less control over the contact participants from different research conditions possibly had with each other. However, this can never be fully prevented in quasi-experiments of multiple weeks in an ecologically valid environment. Ideally, future research will also include designs that focus on improving students' SRL during a semester or school year to evaluate longer term effects.

In addition, measurement type of SRL could also have been a factor. By measuring SRL with pretest and posttest questionnaires and by collecting online expressions of SRL activities, we followed recent recommendation to measure different levels of SRL and including measurements that focus on both the quality and quantity of SRL (Dent & Koenka, 2016; Panadero et al., 2016; Rovers et al., 2019). However, a recent meta-analysis found that effects of interventions on SRL depend on the measurement type, as effects were stronger when SRL activities were measured with a quantitative measurements (e.g., Moos & Bonde, 2016), compared to questionnaires (Jansen et al., 2019).

Fourth, while we used power analysis to calculate the necessary sample size beforehand to find effect sizes comparable to previous research, we lost some power due to non-compliance of 23% of the students. We decided to exclude non-compliance students who watched less than five (of seven) videos (from both conditions), because it means that students in the SRL-prompts condition were not sufficiently exposed to the SRL prompts and are in fact more comparable to students in the control group. Therefore, conclusions on the non-significant results should be interpreted with caution. As we are aware that this could introduce a selection bias to our results, we also provided the results of the analyses on all the students in Appendix B. It shows that the results from the analyses on all the students do not lead to substantially different conclusions.

Fifth, the online SRL measurement *video watch time* is our best indication of time-on-task. It should be noted, however, that this measurement only measured actual playtime of the video: when students paused the video, the *video watch time* did not increase. We could also not measure the time that students spent on learning and applying SRL strategies when they learned material at home before and after the video. The average *SRL-prompt compliance* rate of 90% indicates that students in the SRL condition generally followed the instruction. The actual rate could be slightly higher or lower, as it does not mean that when a student did not provide a valid answer (e.g., blank) on a particular prompt, the student did do something outside of the online learning environment with the SRL instruction (and vice versa).

Conclusion

When curriculum developers and teachers decide to apply FL, they should ask the question if it is worth to invest in the support of SRL. General consensus from previous research shows that supporting SRL could be an important addition or even precondition for a FL. One effective way to do this in higher education is embedding SRL prompts in instructional videos. Our study shows that this is not directly applicable in the educational context of a History course in secondary education. Our

findings indicate that the current flipped intervention in general stimulates students to perform SRL activities while learning from the videos, and we found an effect of the SRL prompts on the amount of videos the students completed. Future research should investigate how SRL can be optimally supported in flipped secondary education. The results from the small scale focus group interviews suggest that making explicitly clear to students why SRL prompts are useful for their learning could be an important addition to improve the effect of SRL prompts.

Appendix A

Overview of all the video embedded SRL prompts for each video (knowledge clip) and each SRL phase. Hints (in italics) were provided after students answered a that particular SRL prompt.

Video	Forethought prompts (start of video)	Performance prompts (middle of video)	Self-reflection prompts (end of video)
1	<p>What do you think will be the best way to learn from this knowledge clip?</p> <p><i>Hints: You can think of answering the quizzes and making notes in your workbook. And how will you make sure you are concentrated to learn?</i></p>	<p>Which information have you learned up until now about the agricultural revolution?</p> <p><i>Hint: You could rewind difficult parts of the video, or read paragraph 5.2 in your textbook.</i></p>	<p>(1) Did your approach to learn from this knowledge clip work for you? Or is it necessary to make some changes in your strategy?</p> <p><i>Hint: Do you eliminate distraction when you learn?</i></p> <p>(2) Do you understand the learning content from the video? If not, will you watch certain parts again?</p>
2	<p>What do you already know about the emergence of factories?</p>	<p>Is there any information so far that you did not understand?</p> <p><i>Hint: You could rewind difficult parts of the video, or read paragraph 5.2 in your textbook.</i></p>	<p>What have you learned about the emergence of factories?</p>
3	<p>Answering the quizzes in your workbook helps you to understand the learning content. How do you make sure you will also understand the content outside the quizzes?</p> <p><i>Hint: Make notes, pause regularly and summarize the content for yourself, keep asking yourself questions if you understand the learning content.</i></p>	<p>What have you learned about steam power? What can you do if you experience difficulties?</p>	<p>If you think about the way you are learning from the videos right now, do you need to adjust your strategies?</p> <p><i>Hint: Make your current learning strategy explicit for yourself. What else could work good for you? How can you know if it works for you?</i></p>
4	<p>Before you start this knowledge clip, do you already have questions about the working conditions of the working class in the factories?</p> <p><i>Hint: Try to ask yourself 2-3 questions and put them here. For example: are the working conditions the same in the 17th and 19th century?</i></p>	<p>Is there information about the working environments that you did not understand? If yes, what could you do now?</p> <p><i>Hint: Do you ask questions to someone nearby, or a classmate, or your teacher? Are there other options for help that you approach (for example your textbook)?</i></p>	<p>Is it clear for you what the most important learning content was in this video?</p> <p><i>Hint: Do you look at the heading titles in the video to understand this? Or do you use the quizzes in your workbook to evaluate that?</i></p>

Video	Forethought prompts (start of video)	Performance prompts (middle of video)	Self-reflection prompts (end of video)
5	<p>(1) Do you set goals for yourself that would help you to learn the learning content in this video? Why?</p> <p><i>Hints: for example, you could set goals like:</i></p> <ul style="list-style-type: none"> -I want to be able to explain difficult concepts in my own words -I want to answer the quizzes correctly -I want to spend 15 minutes to my history home work. <p>(2) This knowledge clip is about urbanization: cities that will grow larger and larger. Before you watch the video, do you already have ideas what this has to do with factories?</p>	<p>Are you able to learn from this video in an effective way? How is your concentration, do you experience distraction? Do you work on a quiet place? If not, what can you change?</p>	<p>Do you know why it is more effective to answer the quizzes thoughtfully, and to make notes in your workbook?</p>
6	<p><i>Hint: this video is about the question who will stick up for the working class. Why is this necessary if you think about the previous videos?</i></p>	<p>The next section of this video will be about three different political views. It is important to know the differences and similarities. How will you make sure to understand and remember them?</p>	<p>Did you understand all the information about the social issue? If not, will you rewind the most important parts?</p>
7	<p>Think about two goals that you want to accomplish by watching this knowledge clip. For example about your planning, or the way you work.</p>	-	<p>Did you accomplish the goals you have set at the beginning of this knowledge clip? If not, what can you still do to reach them?</p>

Appendix B

Presentation of the analyses on data from all the students (including non-compliance). In Table B.1, we present the descriptive statistics for all the dependent variables.

Effects of *Self-Regulated Learning* -prompts

Self-Regulated Learning Self-Reports

A MANOVA that included condition (SRL prompts versus no SRL prompts) and educational level (senior general versus pre-university) as factors and the seven SOL-Q-R scales as dependent variables was used to investigate if students who received SRL prompts reported higher SRL self-reports than students who did not receive SRL prompts (RQ1a). We found no main effect of condition on the SRL self-reports scales, $F(7, 132) = 0.44, p = .879$; Wilk's $\lambda = 0.977$, partial $\eta^2 = .02$, but we did find a main effect of educational level on the SRL self-reports scales, $F(7, 132) = 3.25, p = .003$; Wilk's $\lambda = 0.853$, partial $\eta^2 = .15$. Follow up ANOVAs showed that the pre-university level ($M = 4.73, SD = 1.11, n = 104$) scored significantly higher ($p = .005$, partial $\eta^2 = .06$) on the *time management* scale than the senior general level ($M = 4.16, SD = 0.84, n = 38$).

Self-Regulated Learning Online Activity

We dropped *video watch time* and *portions viewed* from the analysis because of very high correlations with *rewind actions* and the problem of multicollinearity. A MANOVA with condition and educational level (pre-university, senior general) as factors and the *completion rate* and *rewind actions* as online SRL activity variables as dependent variables was used to investigate if students who received SRL prompts performed more online SRL activities than students who did not receive SRL prompts (RQ1b). We found no significant main effect for condition on online SRL activity, $F(2, 148) = 0.18, p = .837$; Wilk's $\lambda = 0.998$, partial $\eta^2 < .01$. We found a significant main effect for educational level on online SRL activity, $F(2, 148) = 9.42, p = < .001$; Wilk's $\lambda = 0.887$, partial $\eta^2 = .11$. Follow up ANOVAs showed a significant difference in means ($p < .001$, partial $\eta^2 = .11$) of the online SRL activity variable *completion rate* between educational levels: students in the senior general level on average watched 55% of the videos ($SD = 30.49, n = 44$) and students in pre-university on average 80% ($SD = 30.49, n = 109$).

Learning outcomes

An ANOVA on the final recall and comprehension test with condition and educational level as factors was used to investigate if students who received SRL prompts obtained higher learning outcomes than students who did not receive SRL prompts (RQ2). We found no significant effect of condition on learning outcome; $F(1, 148) = 0.45, p = .505$,

partial $\eta^2 = < .01$, but we did find a significant effect of educational level on learning outcome, $F(1, 148) = 21.30, p = < .001$, partial $\eta^2 = < .13$.

Table B.1

Mean and Standard Deviation for the Dependent Variables (all Students Included)

Dependent variables (posttest)	SRL-prompts condition			No-SRL-prompts condition		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Metacognitive activities before	71	3.50	1.18	74	3.50	1.15
Metacognitive activities during	71	3.58	1.17	75	3.63	1.11
Metacognitive activities after	71	2.99	1.21	74	2.96	1.17
Persistence	70	4.55	1.31	75	4.59	1.29
Help seeking	70	3.84	1.14	75	3.53	1.14
Environmental structuring	71	5.26	1.43	74	5.36	1.34
Time management	71	4.55	1.10	75	4.59	1.04
Video completion rate	73	73.89	36.43	80	72.25	34.06
Video watch time	73	52.40	32.00	80	52.13	30.03
Video portions viewed	73	70.34	43.35	80	65.91	39.02
Rewind actions	73	18.51	25.63	80	15.26	22.71
Learning outcome test	74	15.52	4.22	78	15.03	4.11

Moderation of the Effects of Self-Regulated Learning-Prompts by Performance Level

We performed two separate MANCOVAs and one ANCOVA with performance level (i.e., the students' average scores for History that current schoolyear) as a covariate, condition type and educational level as factors, and the SRL questionnaire scales, the SRL online activity variables, and the learning outcome test as dependent variables to investigate if performance level moderated the effects of SRL prompts (RQ3). We found no significant interaction effects between performance level and condition for the SRL self-reports, $F(7, 131) = 1.27, p = .269$; Wilk's $\Lambda = 0.936$, partial $\eta^2 = .06$; no significant interaction effects between performance level and condition for the SRL online activities, $F(2, 147) = 0.44, p = .645$; Wilk's $\Lambda = 0.994$, partial $\eta^2 = .01$; and no significant interaction effects between performance level and condition for learning outcome; $F(1, 147) = 0.52, p = .470$, partial $\eta^2 = < .01$. This shows that the effect of the SRL prompts is not different for students' performance level.

Appendix C

Table C.1
Pearson Correlation Matrix for all the Dependent Variables (N between 102 and 105, only students who watched at least five out of seven videos)

Dependent variable	MAB	MAD	MAa	PER	HS	ES	TM	VCR	VWT	VPV	RA
Metacognitive activities before (MAB)	-										
Metacognitive activities during (MAD)	.715**	-									
Metacognitive activities after (MAa)	.768**	.799**	-								
Persistence (PER)	.357**	.382**	.350**	-							
Help seeking (HS)	.432**	.485**	.362**	.160	-						
Environmental structuring (ES)	.200*	.413**	.198*	.299**	.298**	-					
Time management (TM)	.248*	.334**	.294**	.684**	.281**	.422**	-				
Video completion rate (VCR)	.070	.048	.076	.183	.014	-.081	.151	-			
Video watch time (VWT)	.021	.027	.073	.088	.050	-.108	.053	.481**	-		
Video portions viewed (VPV)	.026	.035	.085	.086	.078	-.110	.076	.496**	.987**	-	
Rewind actions (RA)	.014	.027	.075	.053	.083	-.101	.048	.306**	.968**	.978**	-
Learning outcome test (LO)	.176	.164	.156	.328**	.104	.209*	.374**	.303**	.217*	.227*	.175

* $p < 0.05$ (1-tailed).
** $p < 0.01$ (1-tailed).

Table C.2

Pearson Correlation Matrix for Performance Level and Dependent Variables (N between 102 and 105, only students who watched at least five out of seven videos)

Dependent variable	Performance level
Metacognitive activities before	.180*
Metacognitive activities during	.056
Metacognitive activities after	.185*
Persistence	.391**
Help seeking	-.008
Environmental structuring	.169*
Time management	.357**
Video completion rate	.074
Video watch time	.069
Video portions viewed	.071
Rewind actions	.060
Learning outcome test	.556**

* $p < 0.05$ (1-tailed).

** $p < 0.01$ (1-tailed).

4

Chapter 4

Self-Regulated Learning Support in Flipped Learning Videos Enhances Learning Outcomes

This chapter is based on:

van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2020). Self-regulated learning support in flipped learning videos enhances learning outcomes. *Computers and Education*, 158, 104000. <https://doi.org/10.1016/j.compedu.2020.104000>

Acknowledgement of author contributions:

DvA, CP, JJ, and LK designed the study; DvA collected the data; DvA, CP, JJ, and LK planned the data analysis; DvA and JJ analyzed the data; DvA drafted the manuscript; DvA, CP, JJ, and LK contributed to critical revision of the manuscript; CP, JJ, and LK supervised the study.

Supplementary material to this chapter can be found on:

<https://doi.org/10.34894/YIOF09>

Abstract

In flipped learning, students study learning material before class and apply the content of the learning material during class. This requires self-regulated learning (SRL) behavior due to the increased autonomy in this instructional approach. Providing students with video-embedded SRL support (i.e., prompts and explicit instruction) during the learning activities before class has proven to be an effective strategy in primary and higher education to enhance students' SRL and learning outcomes. The current study aims to investigate the effect of SRL support in a flipped class in secondary education over the course of eight weeks. In total, 115 eighth-grade students from five classes participated in a quasi-experimental study, which measured the effects of SRL support on students' SRL (self-reports and online activities), learning outcomes, and satisfaction. We found a positive effect of SRL support on learning outcomes, but we could not explain this by differences in students' SRL. Although all the students were generally positive about the flipped learning environment, some students clearly disliked the SRL instruction. We conclude that SRL support is beneficial for students' learning but that it should be carefully designed to avoid students' dissatisfaction, which could potentially nullify these beneficial effects on learning.

Introduction

Flipped learning (FL) is an effective approach to increasing students' learning outcomes, in which they study learning material before class (e.g., by watching instructional videos) and apply the content of the learning material during class (van Alten et al., 2019; Cheng et al., 2019; Låg & Sæle, 2019). To benefit from FL, self-regulated learning (SRL) behavior (e.g., monitoring) is required (He et al., 2016; Shih & Huang, 2019). However, students do not always use optimal learning strategies during self-study (Dirkx et al., 2019), and can even resist changing their study behavior during FL (Boevé et al., 2017). Providing students with SRL support (e.g., by prompts) during their learning activities before class has proven to be an effective strategy for enhancing students' SRL and learning outcomes during short-term FL interventions (e.g., one session or a few weeks) in primary education (Lai & Hwang, 2016) and higher education (Moos & Bonde, 2016). The current study investigates the effects of SRL support during FL in secondary education over the course of eight weeks.

Higher Demand on Self-Regulated Learning Skills in Flipped Learning

Students that show SRL behavior actively regulate and monitor their cognition, behavior, and motivation while learning, and this involves strategies and goal processes (Panadero, 2017; Zimmerman & Moylan, 2009). Self-regulated learners have the ability and motivation to think about how, what and why they are learning (i.e., metacognition) and thereby control their learning behavior (i.e., self-regulation). Self-regulated learning has been theorized as a cyclical model in which SRL occurs consecutively in a forethought phase (e.g., task analysis), a performance phase (e.g., self-control and monitoring), and a self-reflection phase (e.g., self-judgment and change behavior accordingly) while learning (Zimmerman & Moylan, 2009).

In the case of FL, which is an example of a blended teaching strategy, students are given more autonomy over their learning process. This puts a higher demand on students' SRL skills (He et al., 2016; Lee & Tsai, 2011; Shih & Huang, 2019; Tan et al., 2017). The extent to which students are given autonomy differs according to the implementation of FL (Van Laer & Elen, 2017). First, students can perceive different levels of autonomy, depending on how much control they are given over studying various learning material (Bouwmeester et al., 2019; Shih & Huang, 2019). Second, students' autonomy is also affected by the degree to which students can manipulate the pace and sequencing (e.g., by pausing and rewinding the instructional videos) of the learning content (Abeysekera & Dawson, 2015; Van Laer & Elen, 2017), and also by their decisions as to where (environmental structuring) and when (planning) to learn (Jansen et al., 2018). Consequently, when students are able to regulate their

own learning adequately and according to their increased autonomy, they achieve higher learning outcomes during FL in comparison with students who are not able to do so (Lai & Hwang, 2016; Lee & Choi, 2019; Shibukawa & Taguchi, 2019).

However, this high level of autonomy and the consequent demand on students' SRL skills in FL could pose problems for students (Butzler, 2016; Heyma et al., 2015; Lee et al., 2010; Sletten, 2015; Wolters et al., 2005). Even if students understand that different study behavior is necessary to benefit from FL, they do not always change their learning behavior, and they may even resist doing so (Boevé et al., 2017). In addition, research shows that students' knowledge about cognitive learning strategies is lacking, which in turn hampers metacognition during self-study, as students do not always use optimal learning strategies (Blasiman et al., 2017; Dirkx et al., 2019). On the one hand, this could be explained by the lack of sufficient training in effective study strategies such as distributed practice and retrieval practice (Dignath & Büttner, 2018; Surma et al., 2018). On the other hand, this could be explained by age, as younger students have fewer SRL skills (Dent & Koenka, 2016; Wigfield et al., 2011). This could be a challenge for students applying FL, as they could lack sufficient metacognitive skills and cognitive learning strategies to fully benefit from the increased autonomy (e.g., Heyma et al., 2015).

Self-Regulated Learning Support in Online Flipped Learning Environments

As students may struggle to apply SRL to fully benefit from FL, supporting students' SRL in an FL environment seems worthwhile. Previous research has shown that supporting SRL by means of SRL interventions (e.g., SRL instruction and SRL prompts) enhances students' SRL and in turn (at least partly) enhances learning outcomes (Dent & Koenka, 2016; Dignath & Büttner, 2008; Donker et al., 2014; Jansen et al., 2019; Zheng, 2016). In general, in the context of blended learning, Van Laer and Elen (2017) found seven key attributes that can support SRL: authenticity, personalization, learner-control, scaffolding, interaction, cues for reflection, and cues for calibration. The cues for reflection (defined as SRL prompts) are known to be effective in supporting students' SRL (Devolder et al., 2012) and have also been studied in the particular context of FL. For example, SRL prompts are given to students learning from instructional videos, in the form of encouragement and questions (e.g., 'Explain what you have learned' and 'What have you learned?') to motivate them to engage in SRL activities.

The theory behind the effectiveness of SRL support is that students are informed about and encouraged to engage in SRL activities in each phase of learning: to set goals and plan in the forethought phase, to monitor and control their learning in

the performance phase, and to evaluate their own learning and set goals in the self-reflection phase (Moos & Bonde, 2016; Zimmerman & Moylan, 2009). A substantial number of studies have concluded that providing students with this form of SRL support during online learning activities before class leads to higher SRL (measured through self-reports or online activities) and learning outcomes (Butzler, 2016; Jovanović et al., 2017; Jovanovic et al., 2019; Lai & Hwang, 2016; Moos & Bonde, 2016; Shyr & Chen, 2018; Sletten, 2017; Sun et al., 2018; Yilmaz et al., 2018). In addition, we know from research that SRL support should be embedded in the learning environment to stimulate students' optimal use of the support (Clarebout et al., 2010) and is best implemented in a context-specific manner and not as an independent form of instruction (e.g., Paris & Paris, 2001).

Nevertheless, there are still knowledge gaps in the scientific literature that we address in the current study. First, previous research into SRL support in FL focused specifically on (video-embedded) SRL prompts (Devolder et al., 2012). Prompting SRL has proven to be effective in enhancing learning outcomes in primary education (Lai & Hwang, 2016) and higher education (Moos & Bonde, 2016). The theoretical novelty in the current study is that we added explicit SRL instruction to this type of intervention to investigate whether it improves SRL support in secondary education. Previous research into improving students' SRL skills in other online learning contexts has supported the idea that adding explicit instruction of SRL in an online learning environment could be a beneficial teaching strategy for enhancing students' SRL and learning outcomes (Bannert et al., 2009; Jansen et al., 2020; Kistner et al., 2010; Kramarski et al., 2013; Moos & Ringdal, 2012; Paris & Paris, 2001). Therefore, the current study combines two effective SRL support strategies: not only prompting students *when* SRL behavior is beneficial, but also demonstrating *how* they could do it, and *why* it is important.

Second, there is a lack of data regarding the effectivity of SRL support in ecologically valid FL settings (i.e., over the course of multiple weeks in a real classroom, rather than in a laboratory setting), and in particular in secondary education (Greene et al., 2015). In contrast to previous research in higher education, one of the first SRL prompt interventions within secondary education found no differences in terms of learning outcomes between students who received SRL prompts and students who did not (van Alten et al., 2020a). In addition, previous research has shown the importance of SRL support for the SRL development of primary and secondary education students (Dignath & Büttner, 2008; Muijs & Bokhove, 2020; Veenman et al., 2006; Wigfield et al., 2011). Therefore, more research is necessary to determine whether what we know from research in the context of higher education is applicable to secondary education.

Third, previous research showed that students with high prior knowledge of a certain

topic perform more online SRL activities than students with low prior knowledge (Moos & Azevedo, 2009; Taub et al., 2014). It could be possible that the effect of SRL support on SRL is dependent on students' prior knowledge (Kramarski et al., 2013; Lim et al., 2009; Sergis et al., 2018; Yeh et al., 2010). For example, students with low prior knowledge could find benefiting from SRL support difficult as it is cognitively demanding (Schwonke et al., 2013). In contrast, students with low prior knowledge have been found to benefit more from SRL support, for example because they are more dependent on guidance in effective learning strategies and resource management in comparison with students with high prior knowledge who are better able to assess the effectiveness of their own learning (Schwonke et al., 2013). The moderating role of prior knowledge has not yet been investigated in SRL support research in FL.

Fourth, little research has been carried out to investigate the relationship between students' SRL activity and students' satisfaction. Research has shown that students struggle with effectively applying SRL during self-study (e.g., Blasiman et al., 2017; Dirxx et al., 2019). This can be related, for example, to their views on FL (Boév et al., 2017), or to their appreciation of the SRL support (van Alten et al., 2020a). In general, students' satisfaction about the learning environment is an important condition for students to perform and develop SRL in higher education (Shih et al., 2018) and secondary education (Schuitema et al., 2012). Hence, if students are not satisfied with the FL environment, it could hinder their participation and their use of the SRL support. Therefore, it is relevant to investigate students' satisfaction with the online learning environment and about the SRL support in addition to the effects of the SRL prompts, in order to better explain the effect of SRL support on students' learning.

Present Study

Consequently, we answer the following research questions (RQs) in the present study: What is the effect of SRL support during FL on secondary education students' SRL self-reports (RQ1a) and online SRL activity (RQ1b)? What is the effect of SRL support during FL on learning outcomes of secondary education students (RQ2)? For these first two research questions, we hypothesize that students who are regularly supported in performing SRL activities during their learning process, and who are provided with instruction about how to do it and why it is important (SRL-support condition), outperform students without SRL support (no-SRL-support condition).

In addition, we investigate if students' prior knowledge moderates the effects of SRL support (RQ3). An interaction between prior knowledge and the research conditions on SRL or learning outcomes, would indicate that SRL support works better for students with a higher (or lower) prior knowledge.

Finally, we answer two questions regarding student satisfaction: what is the effect of SRL support during FL on secondary education students' satisfaction about the online learning environment (RQ4a) and how do students in the SRL-support condition value the SRL support (RQ4b)?

Method

Design

We used a quasi-experimental between-subjects design with SRL support (instruction and prompts) in the instructional videos for the intervention condition, and no SRL support in videos for the control condition. The intervention consisted of 10 lessons (and eight videos) over the course of eight weeks. In this quasi-experimental design, three teachers each taught two different pre-existing classes. For each teacher, we randomly allocated one of their classrooms to the intervention condition and the other to the control condition to balance the impact of teacher effects in our design.

Participants

A total of 156 eighth-grade students in their second year of Dutch secondary education (13–14 years old) and their parents were asked for their consent to participate in this study. Ethical approval was obtained from the Faculty Ethics Review Board. All students were compensated with a small gift after the intervention, and participants had a chance of winning one of six vouchers (10 euros each) that were raffled in each class after the intervention.

The study took place in a large urban school (2200 students) and the intervention was part of the students' regular school curriculum in 2019 from May to July. The Netherlands has a tracked secondary education system (from Grade 7 onwards), and the school offers the two highest tracks: *senior general* and *pre-university*. *Pre-university* consists of both *atheneum* and *gymnasium* levels the latter including Latin and Greek as compulsory courses. The sample consisted of six pre-existing classes and three teachers, including the first author, who each taught two classes.

Five students were excluded because consent was not obtained, and one student because of missing data. After *a priori* equivalent tests, it appeared that there were significant ($p < .05$) differences between conditions in terms of prior knowledge, self-reported SRL skills and motivation. These differences were due to the inclusion of different educational levels and an unequal distribution across conditions: *senior general* level ($n = 21$), *atheneum* level ($n = 119$), and *gymnasium* level ($n = 10$). Therefore,

we excluded the students from the *senior general* and *gymnasium* levels to minimize *a priori* differences between conditions and enable a fair comparison. Thus, the participants of the current intervention were representative of approximately 20% of Dutch secondary education students following the *pre-university* track.

Finally, we excluded four students due to non-compliance, as they did not watch at least 70% of the instructional videos. For the students in the SRL-support condition, non-compliance indicated that they were not sufficiently exposed to our intervention (Jo, 2002). The four excluded students were allocated to the SRL-support condition. Consequently, our sample consisted of 115 students (50 in the SRL-support condition and 65 in the no-SRL-support condition) from five classes. *A priori* power analyses indicated that a sample size of 152 students was required for sufficient power of 80% and for detecting medium effect sizes as expected by previous research. The current included sample provides a power of 50-60% to detect medium effect sizes.

Materials

Learning Materials

Edpuzzle (<https://edpuzzle.com/>) was used as an online learning environment for sharing the instructional videos with the students. We provided each student with a personal login that was linked to their corresponding research condition. The videos were developed and recorded by the first author. The eight videos were identical for both research conditions, they had an average length of seven minutes, and fast forwarding was disabled when watching the video for the first time.

The learning materials used for the in-class lessons were integrated into the school's regular History curriculum (the Industrial Revolution during the late 19th and early 20th centuries) and were developed in collaboration with participating teachers, using teaching materials that were already used in the participating school. Every student was provided with a workbook, which included instructions about Edpuzzle, lesson procedures, and three corresponding cognitive questions (in both conditions) corresponding to each instructional video (e.g., 'What are advantages of building steam factories in the city?'). Teachers discussed the cognitive questions at the start of each lesson to recall the content of the instructional videos. During the lessons, students were actively engaged in applying the content of the instructional videos (e.g., analyzing historical sources and creating advertisement leaflets to explain the benefits of 19th century industrial inventions).

Self-Regulated Learning Support

The difference between the SRL-support condition and the no-SRL-support condition was twofold. A detailed overview of all the SRL support (prompts and instruction) for

each video is presented in Table A.1. First, the students in the SRL-support condition received three to four SRL prompts per video, corresponding to the three SRL phases (forethought, performance, and self-reflection; Zimmerman & Moylan, 2009). The SRL prompts in this research were based on the study of Moos and Bonde (2016) and focused on every aspect of SRL (e.g., *help seeking*, *metacognitive activities*, and *time management*). Students were required to think about the SRL prompts and answer them to continue the video.

Second, a brief SRL instruction was added at the beginning or at the end of the video. We presented the SRL instruction to students with a gradual reduction in content (i.e., fading or scaffolding) with the aim that students would internalize the SRL strategies instructed (Kollar & Fischer, 2006). The SRL instruction focused mainly on providing students with explicit reasons for thinking about their own learning and why this is useful. We also provided them with proper metacognitive strategies (e.g., *time management*, *environmental structuring*, *help seeking*, and monitoring understanding) and cognitive learning strategies (e.g., spaced practice and how to make notes). Most SRL prompts were built upon the SRL instruction at the beginning of the video (see Table A.1). Monitoring prompts sometimes referred to the quiz questions in the workbooks, followed up by explicit instruction about why this is an effective monitoring strategy.

Measurements

Self-Regulated Learning Log Data

Edpuzzle was used to collect SRL log data. We used the following three variables as indications of SRL activity: students' *video timing*, *video completion rate*, and *rewind actions* (cf. Bannert et al., 2014; Maldonado-Mahauad et al., 2018). For *video timing*, log data included whether a student completed the instructional video on time (2), too late (i.e., after the deadline, which was the beginning of class; 1), or did not complete (0). We calculated the mean video timing score of all the videos. For *video completion rate*, we calculated the percentage of the *video completion rate* of each individual student (e.g., on a score from 0 to 100%). *Video timing* and *completion rate* can be considered to be SRL activities in the *forethought* phase (e.g., planning) and *performance* phase (e.g., *time management* and *persistence*), respectively, according to the SRL model by Zimmerman and Moylan (2009). For *rewind actions*, we counted all *rewind actions* (i.e., when students watched a portion of a video more than once) and calculated a sum score. Every video was automatically divided into 10 portions, and log data were available about how often a student watched a particular portion. *Rewind actions* is an indication of SRL activity in the *performance* and *self-reflection* phases, in which students monitor and evaluate their understanding of the learning content and consciously choose to repeat a part of the instruction.

In addition, the first author coded how many valid (e.g., serious) online answers to the SRL prompts were given by students in the SRL-support condition. The resulting average per video is an indication of *SRL-prompt compliance* which we used to evaluate to what extent the students in the SRL-support condition properly worked with the prompts.

Self-Regulated Learning Questionnaire

The revised self-regulated online learning questionnaire (SOL-Q-R) was used to measure students' SRL self-reports (Jansen et al., 2018). This questionnaire contains the following seven scales for which we calculated scale means: *metacognitive activities before, during, and after learning*, *time management*, *environmental structuring*, *persistence*, and *help seeking*. We translated and contextualized the SOL-Q-R, originally constructed for online learning, to the Dutch student population of 14–15 years (e.g., *'I think about what I have learned after I finish working on homework for History'*) and deleted one to two items in each scale to reduce the duration and avoid students providing non-serious answers. As can be seen in Table B.1, the reliability of the scales in general was acceptable. However, due to the low Cronbach's α for *environmental structuring* (.36), we dropped this scale from further analysis.

Motivation Questionnaire

Motivation was measured only as a pre-test, to investigate if the students in the research conditions were equal in terms of motivation. We used the Motivated Strategies for Learning Questionnaire (MSLQ) and calculated scale means (Pintrich et al., 1991). The MSLQ is the most widely used questionnaire that has proven to be a reliable and useful tool for measuring motivational and SRL constructs (Duncan & McKeachie, 2005). We used the following motivational constructs, as they cover all aspects relevant to SRL. *Intrinsic goal orientation*, *extrinsic goal orientation*, and *task value* were used to measure how students valued and perceived their engagement in a particular learning task (e.g., homework for History). *Self-efficacy for learning* was included to measure students' expectancy beliefs about how they complete their History homework. As can be seen in Table B.2, the scales for our translated and contextualized version were reliable.

Prior Knowledge Test

We developed a prior knowledge test about the content of the lessons (i.e., the Industrial Revolution). We used the test to detect possible differences between research conditions before the intervention. In addition, we aimed to answer RQ3 to determine whether prior knowledge moderates the effects of SRL support on SRL and learning outcomes. The test was presented to students as a quiz on paper and completed during class before the intervention started. It contained seven multiple-choice questions and two open questions about a historical source. The first author

scored all the tests. We calculated a sum score for the prior knowledge test. The reliability of the test was low in terms of Cronbach's α (9 items; $\alpha = .26$), but this is an expected value when using (pre)knowledge tests that usually do not measure the same underlying construct (Taber, 2018).

History Performance Level

The intervention took place at the end of the school year. To further assess *a priori* differences between the research conditions, we therefore created history performance level as an additional measure that indicated a student's performance for this subject in that particular year. We calculated a mean score for all the students' test scores for History in the current schoolyear. This score is commonly used by schools to evaluate whether a student meets the criteria to pass to the next grade.

Learning Outcome Test

The learning outcome test was developed by the researchers and participating teachers and was based on learning materials and was comparable to tests which students usually complete during the schoolyear. Every participating student received the same test in class during a regular test week at the end of the school year. The test consisted of 24 items, of which four were multiple-choice questions. The 20 open questions contained recall (e.g., give the definition of a concept), comprehension (e.g., explain the impact of a concept), and transfer (e.g. relate a historical source to a concept) assignments. Each teacher made use of a comprehensive assessment scheme (validated by the researchers and teachers) to score the tests of their own students. The inter-rater reliability of the open-question scores was assessed by the first author, who independently scored a random sample of five unscored tests from the other classes and compared these with the other teachers' scores (20 tests and 400 items in total). We deleted one item for which Cohen's κ was too low ($\kappa = .30$, 60% agreement rate). The Cohen's κ values for the remaining items were substantial with an average Cohen's κ of .80 (SD = 0.14, range $\kappa = .53$ to 1, $p < .001$) and an average total agreement of 85%. In general, the remaining items (open and closed questions) together were found to be reliable (24 items; $\alpha = .83$), and we calculated a sum score as an indication of students' learning outcomes.

Student Satisfaction

To measure student satisfaction with the learning environment (i.e., the online videos), we added six Likert-items (range from 1 to 7; e.g., '*I think the instructional videos are boring*'; '*I think the instructional videos are instructive*'; '*I believe that the instructional videos are designed in a clear and understandable way*') to the post-test questionnaire to create a mean score. In general, the satisfaction questionnaire was found to be reliable ($\alpha = .85$). In addition, we performed a confirmatory factor analysis in Mplus

(version 8) to verify the underlying one-factor structure. As shown in Table B.3, all goodness-of-fit indicators showed an acceptable model fit (Kline, 2005; Kyriazos, 2018). In addition, students in the SRL-support condition received four additional Likert-items about their experience with the SRL support which we used to describe their general satisfaction about the design of the SRL support (e.g., *'The learning tips I received at the beginning/end of the videos were useful to learn better'*; *'The questions that were part of the videos annoyed me'*).

Finally, we added several evaluative questions to the post-test questionnaire to better interpret students' behavior in our FL experiment. First, we asked all students two open-answer questions about the general benefits of or points of improvements for FL. Second, we provided students with four statements with possible reasons why they watched the videos and five statements why they did not watch the videos. Students were free to select unlimited motives for (not) watching the videos.

Procedure

Before the flipped intervention started, the students separately completed the pre-test questionnaire, with the SRL and motivation items, and the prior knowledge test during regular class time. Students were briefed about working in an FL class and the procedures which they were expected to meet. In addition, students received classroom instruction on how to access the videos at home. They were also informed that their own teacher could track their progress. To stimulate students to comply with the instructions and complete their homework, we informed students that watching the videos in time would be rewarded with a bonus point (which was not included in our analysis) in their final test score.

Teachers were told not to change their usual teaching interaction with their students, and received instructions and training on how to apply the learning materials for each lesson. The teachers used the school's digital homework system to inform students of their own class about the required preparations for each lesson (one instructional video and the accompanying quiz questions about the learning content in their workbooks). Each lesson of 65 minutes started with retrieval practice activities about the content of the videos and a discussion of the quiz questions from the workbooks. This was usually followed by a complementary micro-lecture of approximately 10 minutes to discuss parts that were not yet fully understood as well as more in-depth aspects of the learning material. The remaining part of the lesson was dedicated to engaging learning activities to apply the learning material (e.g., creating advertisement posters, writing essays, analyzing historical sources). After 10 lessons, students completed the posttest SRL and satisfaction questionnaire during regular class time. In the subsequent test week, students had 65 minutes to complete the learning outcome test.

Data Analysis

We first checked the data distribution for normality and potential outliers. In the event that dependent variables were not normally distributed (e.g., *completion rate*), we performed a bootstrapping method (2000 Bootstrap samplings) with replacement to estimate the impact of non-normality (Adèr et al., 2008; Field & Wilcox, 2017). In case of an *a priori* difference between the research conditions for one of the pre-test variables (i.e., *task value*), we included the respective variable as covariate in the models (Steiner et al., 2009).

Next, we analyzed the differences between the SRL-support and the no-SRL-support conditions for the variables of each research question. Due to the linear relationships between the dependent variables (see Table B.4), we decided not to conduct one MANCOVA. Instead, we performed two separate MANCOVAs (with the SRL self-report scales and the SRL activity variables as dependent variables respectively), and two ANCOVAs (i.e., learning outcome test and satisfaction as dependent variables respectively) in SPSS with condition as independent variable and *task value* as a covariate.

To examine if prior knowledge moderated the findings, we added students' prior knowledge as a covariate to the existing models and checked for a possible interaction effect between performance level and condition, which would suggest that the SRL support had a differential effect according to students' prior knowledge. To answer the final research question about the students' satisfaction with the learning environment, we analyzed students' mean scores on the relevant post-test questions.

Results

Evaluation of a Priori Differences Between Conditions

As part of the quasi-experimental protocol, we checked for *a priori* differences between the students in the SRL-support and no-SRL-support conditions before the intervention. In Table B.5, we present the descriptive statistics for all the pre-test dependent variables for both research conditions.

We conducted three separate MANOVAs, in which we included condition as a factor and the related pre-test measures (i.e., SRL self-reports, motivation, and cognitive abilities) as dependent variables. We found no significant difference between conditions in a MANOVA including the six SRL self-reported pre-test scales, $F(6, 108) = 0.74, p = .619$; Wilk's $\lambda = 0.961$.

For students' motivation, we found a significant difference between conditions in a MANOVA with the four motivation scales, $F(4, 110) = 2.84, p = .028$; Wilk's $\Lambda = 0.906$. Follow-up ANOVAs showed that, for the motivation scale *task value* ($p = .009$, partial $\eta^2 = .06$), the students in the SRL-support condition scored significantly higher ($M = 4.27, SD = 1.52, n = 50$) than those in the no-SRL-support condition ($M = 3.53, SD = 1.44, n = 65$), but not for the other three motivation scales.

With regard to student's cognitive abilities, we found no significant differences between conditions in a MANOVA including prior knowledge and history performance level as dependent variables and condition as a factor, $F(2, 104) = 1.10, p = .337$; Wilk's $\Lambda = 0.979$.

Overall, there were no *a priori* differences between students from both research conditions except for *task value*. In the following results, we included *task value* in each subsequent analysis as a covariate (Steiner et al., 2009). Correlations between *task value* and most of the dependent variables were $r > .36, p < .001$.

Effects of Self-Regulated Learning Support

In Table 1, we present the descriptive statistics for all the dependent variables of both research conditions. We also assessed *SRL-prompt compliance* to estimate how effectively students in the SRL-support condition have worked with our SRL intervention. The average of valid (i.e., serious) student answers per video was 86% and ranged from 79% (video 4) to 93% (video 2). We concluded that, in general, students worked properly with the SRL prompts.

Self-Regulated Learning Self-Reports

To assess whether SRL support enhances self-reported SRL (RQ1a), we performed a MANCOVA with condition (SRL-support versus no-SRL-support) as a factor and the six SRL self-report scales as dependent variables. We found no significant main effect of condition on SRL self-reports, $F(6, 107) = 0.82, p = .555$; Wilk's $\Lambda = 0.956$, partial $\eta^2 = .04$.

Self-Regulated Learning Online Activity

To assess whether SRL support increases online SRL activity (RQ1b), we performed a MANCOVA with condition as a factor, *completion rate*, *rewind actions*, and *video timing* as dependent variables. We found no significant main effect of condition on SRL online activities, $F(3, 110) = 1.52, p = .214$; Wilk's $\Lambda = 0.960$, partial $\eta^2 = .04$.

Learning Outcomes

To evaluate whether SRL support increases learning outcomes (RQ2), we performed an ANCOVA with condition as factor and the learning outcome test score as the

dependent variable. We found a significant main effect of condition on learning outcome, $F(1, 111) = 5.13, p = .025$, partial $\eta^2 = .04$. As can be seen in Table 1, students in the SRL-support condition scored higher than those who did not receive SRL support.

Prior Knowledge as Possible Moderator

To assess whether students' prior knowledge moderated the effects of SRL support (RQ3), we performed two separate MANCOVAs and one ANCOVA. We added prior knowledge as a covariate, condition as a factor, and the SRL questionnaire scales, the SRL online activity variables, and the learning outcome test as dependent variables to the separate models.

Table 1

Mean and Standard Deviation for the Dependent Variables per Condition and P-values of the Analyses

Dependent variables (post-test)	SRL-support condition			No-SRL-support condition		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
<i>SRL self-report scales ($p = .555$)</i>						
Metacognitive activities before	50	3.25	1.17	65	3.21	1.00
Metacognitive activities during	50	3.95	1.01	65	3.68	1.01
Metacognitive activities after	50	3.16	1.30	65	3.04	1.18
Persistence	50	4.82	1.05	65	4.55	1.28
Help seeking	50	4.29	1.30	65	4.34	1.13
Time management	50	4.56	1.14	65	4.55	1.04
<i>SRL online activity ($p = .214$)</i>						
Video completion rate	50	97.38	6.45	65	99.04	4.01
Rewind actions	50	48.40	47.17	65	42.32	48.95
Video timing	50	1.85	0.25	65	1.93	0.14
<i>Learning outcome ($p = .025$)</i>						
Learning outcome test	50	20.47	4.92	65	17.65	5.98
<i>Satisfaction ($p = .184$)</i>						
Satisfaction with learning environment	50	5.39	0.88	65	5.41	1.26

Note: Provided *p*-values are from the separate (M)ANCOVAs.

We found no significant interaction effects between prior knowledge and condition on the SRL self-reports, $F(6, 97) = 1.17, p = .331$; Wilk's $\lambda = 0.933$, partial $\eta^2 = .07$; no significant interaction effects between prior knowledge and condition were found for the SRL online activities, $F(3, 100) = 9.89, p = .402$; Wilk's $\lambda = 0.971$, partial $\eta^2 = .03$; and no significant interaction effects between performance level and condition were

found for learning outcomes; $F(1, 101) = 0.02$, $p = .892$, partial $\eta^2 = < .01$. Overall, the effects of SRL support on SRL (self-reported and online activity) and learning outcomes did not depend on the level of students' prior knowledge.

Satisfaction

First, we examined whether students who received SRL support valued the instructional videos differently than students who did not receive SRL support (RQ4a). We performed an ANCOVA with condition as factor and the satisfaction score as a dependent variable. We found no significant main effect of conditions on satisfaction, $F(1, 110) = 1.79$, $p = .184$, partial $\eta^2 = .02$.

Second, we evaluated how students in the SRL-support condition valued the SRL support in the videos (RQ4b). We provide descriptive means and standard deviations in Table 2. These demonstrate that students in the SRL-support condition were on average neutral about the SRL prompts and negatively valued the SRL instruction in the videos. In addition, the students did not perceive that the SRL support caused a change in their metacognitive learning behavior. It should be noted that the standard deviations for all evaluative questions were large, which indicates that the satisfaction of students varied greatly.

Table 2

Mean and Standard Deviation for Satisfaction about the SRL Support on a Scale from 1 to 7 (1 = totally disagree and 7 = totally agree)

Question	<i>n</i>	<i>M</i>	<i>SD</i>
The questions that were part of the videos were useful to learn better	49	4.37	1.91
The questions that were part of the videos annoyed me	48	4.42	1.97
The learning tips I received at the beginning/end of the videos were useful to learn better	49	2.31	1.48
The learning tips I received at the beginning/end of the videos annoyed me	48	5.17	2.01

Third, to better interpret our findings, we added evaluative questions about how students experienced our FL experiment. When asked how seriously they watched the instructional videos to understand the learning material, students on average answered that they deliberately watched them ($M = 5.50$, $SD = 1.41$, $n = 112$, scale from 1 to 7). Furthermore, in Table 3 we present an overview of students' open answers to the questions asking what they valued positively about FL, and what could be improved. The results confirm the earlier reported findings that students were generally positive about their experience in our FL intervention and that a proportion of the students complained about the SRL support. In addition, students indicated why they did or did not watch the videos and *t*-tests showed no significant differences

between the two conditions for each of the motives, as presented in Figure 1. It appears that, for more than 80% of the students, the bonus point was an important incentive to watch the videos, and the majority watched the videos because they liked the FL method.

Table 3

Results What Students Valued about the FL Method and What Could be Improved, n = 115 Students

What did you like about the FL method?	n	Do you have any points of improvements for the FL method?	n
I learned better	36	Negative comment about the SRL instruction (e.g., distracting, annoying, could be reduced)	16
I spent less time on homework	30	Negative comment about the SRL prompts (e.g., distracting, could be reduced)	13
I liked it compared with usual homework	14	Videos lacked entertainment	12
Being able to pause and rewind the instruction	12		

Figure 1

Percentage of Students in the SRL-Support (n = 50) and the No-SRL-Support (n = 65) Conditions who Agreed with the Statements About Reasons why they Watched (first four Statements) or did not Watch (last five statements) the Instructional Videos

Discussion

Effects of Self-Regulated Learning Support on Self-Regulated Learning, Learning Outcomes, and Satisfaction

In the current research intervention in an eight-week flipped History class, eighth-grade students in an SRL-support condition received video-embedded SRL prompts and instruction. We expected that the SRL support would help students to benefit from the FL approach, because the increased learner autonomy that comes with this approach demands SRL skills that most students do not possess. As previous research in the same educational context showed that SRL prompts did not enhance learning outcomes (van Alten et al., 2020a), we added explicit SRL instruction to the design of embedded SRL support to explain to students why SRL activities are beneficial, and how they can be performed. We hypothesized that the embedded SRL support enhances students' SRL skills, compared with students in the condition without SRL support and, consequently, enhances their learning outcomes. We found that students in the SRL-support condition achieved higher learning outcomes, but, in contrast to our hypothesis, we found no difference in SRL skills as indicated by the SRL self-reports and the SRL online activity variables. We provide several reasons that could explain these findings and would be interesting for future research.

Self-Regulated Learning Support Enhances Learning Outcomes, but not via (the Number of) Self-Regulated Learning Activities

Previous research has shown that SRL interventions and students' SRL skills are positively related to higher learning outcomes (Dent & Koenka, 2016; Donker et al., 2014). Therefore, studies that test and show positive effects of SRL interventions on learning outcomes assume that these are caused by a positive mediation effect of SRL interventions on learning outcomes by SRL activities (Jansen et al., 2019). However, a meta-analysis showed that SRL activity only partially mediates the effects of SRL interventions on learning outcomes (Jansen et al., 2019). This suggests that there are other factors leading to enhanced learning outcomes resulting from SRL support, apart from improving students' SRL self-reports and activities.

First, this could be explained by an increase in time-on-task for students whose SRL was supported. Due to the SRL support, students could be more aware in the monitoring phase that they had to spend more time on learning because they did not understand parts of the learning material (Belski & Belski, 2014; Jansen et al., 2019). Unfortunately, we could not measure these corrective learning activities outside the online learning environment. It is also possible that some students who received SRL support spend less time on learning and performing SRL activities, because the SRL

support enabled them to learn more efficiently. Indeed, some students reported that they valued the FL method because they spend less time on their homework than usual (see Table 3).

Second, the strategy behind our SRL support was to enhance both the quality (i.e., SRL instruction) and the quantity (i.e., SRL prompts) of students' SRL activities (Panadero et al., 2016). However, the extent to which we measured SRL (i.e., log data and self-reports) mainly provides quantitative SRL indications. While we did not find that the effect of SRL support on learning outcomes was due to an increase in the number of SRL activities, it could be possible that our intervention improved the quality of students' SRL activities.

Students Were Satisfied With Flipped Learning, but not With The Self-Regulated Learning Support

Our results show that we can rule out the possibility that we found no effects on SRL because the students lacked the level of satisfaction required to perform SRL activities due to the learning environment (cf. Schuitema et al., 2012). Students from both research conditions were generally satisfied with the FL environment. The mean score on the satisfaction questionnaire showed a positive value (Table 1), and students from both conditions seem to have similar reasons to watch or not watch the videos (Figure 1). In addition, the students stated that their effort to learn from the videos was deliberate. The *SRL-prompt compliance* rate also suggests that the students from the SRL-support condition used the prompts properly.

However, students from the SRL-support condition expressed their dissatisfaction with the SRL support in the videos (Tables 2 and 3). The opinions about the SRL prompts were neutral on average, but varied greatly. The opinions about the SRL instructions were clearly negative, although these also varied greatly. In contrast to previous research in the same educational context (van Alten et al., 2020a), SRL prompts alone did not enhance learning outcomes. In the current study, we found an effect on learning outcomes when SRL instruction was added, even though the students seem have to generally disliked this type of SRL support. It is important to further investigate students' dissatisfaction with the SRL support. Support seems to work better when students are satisfied and recognize the added value of it, and dissatisfaction may even prevent them from engaging in SRL activities (cf. Schuitema et al., 2012).

We suggest two possible explanations for the results in our study that have important implication for future SRL research in secondary education. First, some direction can be found in students' answers that state that the SRL support should be reduced, because they already knew why and how to apply SRL strategies. Previous research

suggested that students in primary education benefit more from SRL interventions as they lack knowledge and experience with SRL, in contrast to students in secondary education (Donker et al., 2014; Lai & Hwang, 2016). It could be that the secondary education students had already acquired a certain level of SRL and that therefore explicit SRL instruction did not enhance their SRL (cf. the expertise reversal effect; Kalyuga et al., 2003). However, our results indicate that the added value of SRL instruction to SRL prompts enhances learning outcomes.

Second, it could be that students disliked the SRL support because applying SRL strategies is more difficult than not using them as it requires a certain effort. While students may think SRL support distracts them from learning, we know from research that students tend to misjudge their own learning based on the perceived ease of processing (Bjork et al., 2013). The large variation between students in how they valued the SRL support, and also in their performance of online SRL activity (e.g., *rewind actions*), indicates that some students were very positive about the SRL support and some were very negative.

Limitations and Directions for Future Research

The way we operationalized and measured prior knowledge did not explain the variation in effects of SRL support on SRL and learning outcomes. In future research, it would be helpful to include students from different educational tracks. We included students from only the pre-university track, to diminish pre-existing differences between conditions and to accurately measure the effects of the SRL support using comparable samples. However, this track represents only approximately 20% of the student population in the Netherlands. Including students from other tracks would increase the differences in students' prior knowledge and SRL skills. Previous research into the effectivity of FL in secondary education suggested that FL seems to work well only with students who are good at SRL and are sufficiently motivated to benefit from the increased autonomy (Heyma et al., 2015). As the relation between prior knowledge and SRL behavior has been demonstrated in previous research (Schwonke et al., 2013), we recommend that future research investigates how the individual learning processes of a more diverse student population develop during FL, including their SRL activities, satisfaction with the SRL support, how this is related to prior knowledge and other explanatory variables.

It is also possible that we did not find differences in SRL between conditions due to methodological reasons. For example, it is possible that the design of our FL environment also improved the SRL of students in the no-SRL-support condition to such an extent that it was hard to detect any differences between conditions. Meta-analytic research has found that students applying FL improve in terms of SRL in comparison with students in a traditional classroom (Tan et al., 2017). Other research

has also shown that students in an authentic context may spontaneously take advantage of the interactive possibilities in terms of SRL activities by manipulating instructional videos (e.g., by pausing or rewinding) without being prompted or instructed (Henderson, Selwyn et al., 2017; Merkt et al., 2011; Schwan & Riempp, 2004). This is in line with data from Table 3, in which some students stated that they especially valued FL because they could manipulate the instruction presented in the video. The current sample size was large enough to detect a medium effect on learning outcomes, but too small to detect small effects on SRL. Denying the no-SRL-support condition access to these interactive design features or the quiz questions in their workbook, which potentially increase the SRL activities of this group, could perhaps have helped in detecting the effects of the SRL support. However, this creates a conflict for researchers between maintaining the different conditions as equally as possible, manipulating the independent variables, the ecological validity of the study, and the associated ethical aspects (Bétrancourt & Benetos, 2018).

In addition, we added a bonus point for students as a reward for watching the instructional videos in time before class. Students stated that this was an important incentive to complete the homework videos (Figure 1). This greatly reduced the removal of students from the analysis due to non-compliance (cf. van Alten et al., 2020a). However, a negative effect could have been that this bonus made students in the control group somewhat aware of SRL and perhaps obstructed the detection of an effect of the SRL support on the video completion rate because there was little natural variance due to this external motive. Our suggestion for future research is to keep all these factors in mind while designing SRL support interventions and to achieve a balance between detecting effects and obtaining a large enough sample size.

Implications for Practice

Our results show that it is important for teachers and curriculum developers to consider how the SRL of students is supported in FL environments. As FL increases students' autonomy over their learning, it depends on their SRL skills in order to benefit from the possibilities that FL can offer. We demonstrated that combining SRL prompts embedded in instructional videos with SRL instruction enhanced students' learning outcomes. For instance, we prompted students to rewind parts of the instructional videos when they did not understand the content to stimulate the quantity of SRL activities. We recommend the use of open-ended prompt questions, to provoke deeper reflection by the students (e.g., Panadero et al., 2016).

With the SRL instruction, we aimed to improve the quality of students' SRL activities by providing clear instruction about how to perform effective SRL activities and why it supports their learning. As shown in Table A.1, our SRL instruction was short,

presented to students at the beginning and/or ending of the video, and focused on different aspects of SRL (e.g., planning, monitoring, setting goals, and environmental structuring). Moreover, as we also found that the students generally disliked the explicit SRL instruction, we advise careful consideration of the frequency and mode of the SRL support, as it could make students dissatisfied when they see it as a distraction or burden. We suggest presenting SRL support in a scaffolding strategy, in which the SRL instruction tips build on each other and are gradually reduced to encourage the students to internalize the SRL strategies instructed (Kollar & Fischer, 2006). In addition, more research is needed on the question of whether SRL support should be tailored according to the different levels of students' SRL skills (e.g., Vanslambrouck et al., 2019). Caution should be taken by presenting SRL support as optional (e.g., for students with better SRL skills), because students do not seem to comply with SRL support if it is not embedded in the learning environment (Clarebout et al., 2010; Jansen et al., 2020).

Conclusion

Our study shows that in the context of an eight-week flipped History class in secondary education, supporting students' SRL by explicit instruction and prompts in the instructional videos leads to improved learning outcomes. However, this effect cannot be explained directly by an improvement in SRL skills, as we did not find effects of SRL support on SRL self-reports or online SRL activities. Although the SRL support helped students to be more aware of their own learning, some students clearly disliked the SRL instruction. We conclude that SRL support is beneficial for students' learning, but that it should be carefully designed to prevent this effect from being hampered by students' dissatisfaction.

Appendix A

Table A.1
Overview of all the video embedded SRL prompts and the SRL instruction the SRL-support condition. Hints (in italics) were provided after students answered a that particular SRL prompt. Video length is provided for the no-SRL-support condition and the SRL-support condition

Video	SRL instruction before	Forethought prompts (start of video)	
0 – Introduction (05:29/05:29)	-	-	
1 - Agricultural and demographic revolution (08:45/10:40)	Planning + monitoring by making use of the cognitive questions	(1) Make a planning for the next knowledge clip in order to be well prepared and prevent running out of time.	(2) Which advices to learn with the KC are useful for you?
2 – Rise of factories (07:09/09:23)	Setting goals	What are your goals when learning from this homework video? Which learning behavior fits your goal best?	
3 – The power of steam (06:01/07:55)	Rationale behind metacognition + learning strategies (e.g., spaced practice, how to make notes)	Name a reason why it is important to think about the way you are learning from the videos.	
4 – Factory workers (06:08/06:51)	Environmental structuring (e.g., prevent distractions)	How do you deal with distractions while learning from the videos?	

Performance prompts (middle of video)	SRL instruction after	Self-reflection prompts (end of video)
-	-	-
(1) You have now learned about the demographic revolution. Answer quiz question 1 in your workbook to test whether you can explain in your own words.	(2) Do you understand what enclosure acts are? In that case, answer quiz question 2 in your workbook. What could you do now if you don't understand it yet?	Learning strategy (summarize) + help seeking
Is there any information so far that you did not understand? You can check yourself by answering the quiz questions in your workbook.	Learning strategies (e.g., monitoring understanding) + environmental structuring + time management.	Describe briefly what your approach has been to learn with this video. Was it a good approach, or are you going to change your learning behavior next time?
(1) Answer quiz question 1 in your workbook about the benefits of steam power to check your understanding.	(2) Quiz question 3 is about two important developments in transport. Answer this question to know if you understand them, or if you have to watch the previous instruction again.	Did you accomplish the goals you have set at the beginning of this knowledge clip? If not, what can you still do to reach them (for example focus on the quiz questions, rewatch, read the textbook)?
Answer quiz question 1 in your workbook and recall the causes for the poor working conditions and child labor. <i>Hint: The correct answers to the quiz questions make a nice little summary of what is important learning content. In addition, it helps you monitoring your understanding of the video.</i>	Rationale behind quiz questions as an effective strategy + follow up strategies of a student does not know an answer to the monitoring questions (e.g., rewatch video, read textbook, write down question for peers or teacher in class).	Do you understand all the learning content from this video? If you have any questions, write them down in your workbook and bring them to class. <i>Hint: You can write down questions about the learning content, or about your study approach.</i>
	-	Rate your own concentration while learning from this video. <i>Hint: If you gave yourself a low rating, what are you going to do next time to improve it?</i>

Video	SRL instruction before	Forethought prompts (start of video)	
5 - Urbanization (04:03/04:40)	Activate prior knowledge	<p>(1) Which two goals to you want to achieve while learning from the coming video?</p> <p><i>Hints: for example, you could set goals like understanding the most important learning content, or spent a maximum of 10 minutes to my homework.</i></p>	<p>(2) The previous video was about the poor working conditions in the factories. Look at the title of the current video, and the quiz questions in your workbook; what will be the current video about? How does it connect to the previous video?</p> <p><i>Hint: previous video about poor working conditions in the factories, current video about poor living conditions outside the factories.</i></p>
6 - The social issue (08:19/09:09)	-	<p>Have you changed your way of learning from the videos in the past weeks?</p> <p><i>Hint: Have you become more aware of your study approach?</i></p>	
7 - Modern imperialism (07:47/07:47)	-	<p>This video lasts 07:47 minutes and you will also need some time to answer the questions and make notes. Have you planned enough time for this? Are you working in a quiet place where you can concentrate well?</p>	

Note: The length of the video is provided in this column for subsequently the no-SRL-support condition and the SRL-support condition.

Performance prompts (middle of video)		SRL instruction after	Self-reflection prompts (end of video)
<p>Answer quiz question 2 about why urbanization happened in the 19th century.</p> <p><i>Hint: Do you ever rewind the video if you have doubts about your answer, or to check your understanding?</i></p>		-	<p>Think back to the two goals that you have set at the beginning of the video. Briefly explain whether you have achieved them, and how you accomplished that.</p>
<p>(1) How can you check whether you understand the explanation of trade unions?</p> <p><i>Hint: make quiz question 1 / make notes / rewind / ask your teacher or a classmate for help</i></p>	<p>(2) Answering quiz question 2 is actually a kind of note taking about the differences between the 3 concepts. Is this something you will do yourself in the next video?</p>	Recapture several SRL instructions (e.g., monitoring strategies and environmental structuring)	<p>Rate your own concentration while learning from this video.</p> <p><i>Hint: compare your rating with the previous rating in video 4.</i></p>
<p>(1) Answer the quiz questions that belong to modern imperialism ask yourself whether you now understand the difference between imperialism and modern imperialism.</p>	<p>(2) You now know that in the following part an explanation of three major causes of modern imperialism will be provided. How will you ensure that you understand and remember this?</p> <p><i>Hint: Pause the video after each cause and summarize the cause in your workbook. Rewind if it goes too fast. Explain in your own words. In this way, you remember it better for the test.</i></p>	-	<p>Think about some advantages and disadvantages of learning from the videos and how you have experienced this during the past few weeks. Do you think that you are better prepared for the test this way?</p>

Appendix B

Table B.1
Internal-Consistency Reliabilities of the Contextualized SOL-Q-R Scales

Scale	pretest			posttest		
	<i>n</i>	Items	α	<i>n</i>	Items	α
Metacognitive activities before	111	4	.80	110	5	.75
Metacognitive activities during	111	6	.65	113	6	.71
Metacognitive activities after	113	4	.83	112	4	.86
Persistence	113	5	.82	111	5	.82
Help seeking	109	4	.77	112	4	.83
Environmental structuring	112	3	.51	113	3	.36
Time management	112	4	.72	113	4	.59

Table B.2
Internal-Consistency Reliabilities of the Contextualized MSLQ Scales

Scale	<i>n</i>	Items	α
Intrinsic goal orientation	112	4	.69
Extrinsic goal orientation	111	4	.73
Task value	113	4	.90
Expectancy: self-efficacy for learning	112	4	.82

Table B.3
Goodness-of-Fit Indicators of Satisfaction Questionnaire Model

Model	<i>n</i>	χ^2	<i>p</i> -value	<i>df</i>	χ^2 / df	RMSEA	CFI	TLI
Satisfaction (one factor)	148	15.37	.081	9	1.71	0.07	.98	.97

Note: a non-significant χ^2 suggests model acceptance; *df* = degrees of freedom; χ^2 / df should be lower than 3; RMSEA = Root Mean Square Error of Approximation and should be < .08; CFI = Comparative Fit Index and should be > .95; TLI = Tucker Lewis Index and should be > .95.

Table B.4

Pearson Correlation Matrix for all the Dependent Variables (N between 113 and 115)

Dependent variable	MAB	MAD	MAa	PER	HS	TM	VCR	VTI	RA	LO
Metacognitive activities before (MAB)	-									
Metacognitive activities during (MAD)	.633**	-								
Metacognitive activities after (MAa)	.746**	.714**	-							
Persistence (PER)	.142	.467**	.268**	-						
Help seeking (HS)	.428**	.494**	.419**	.439**	-					
Time management (TM)	.336*	.526**	.466*	.496**	.481**	-				
Video completion rate (VCR)	-.002	.032	-.106	.116	.152	.249**	-			
Video timing (VTI)	.030	.081	.081	.100	.129	.153	.789**	-		
Rewind actions (RA)	.171	.051	.134	-.005	.001	-.007	.111	.153	-	
Learning outcome test (LO)	-.080	.170	.045	.289**	.107	.226*	.117	.181	.173	-
Satisfaction (SAT)	.289**	.450**	.340**	.504**	.393**	.541**	.319**	.160	.080	.183

* $p < 0.05$ (1-tailed).

** $p < 0.01$ (1-tailed).

Table B.5*Mean and Standard Deviation for the Dependent Variables per Condition and P-values of the Analyses.*

Pretest variables	SRL-support condition			No-SRL-support condition		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
<i>SRL self-report scales (p = .619)</i>						
Metacognitive activities before	50	2.83	1.20	65	2.89	1.18
Metacognitive activities during	50	4.00	0.79	65	3.74	1.09
Metacognitive activities after	50	2.95	1.13	65	2.89	1.30
Persistence	50	4.34	1.27	65	4.29	1.19
Help seeking	50	4.77	1.31	65	4.58	1.22
Time management	50	4.06	1.36	65	3.89	1.29
<i>Motivation self-report scales (p = .028)</i>						
Intrinsic goal orientation	50	4.03	1.09	65	3.63	1.18
Extrinsic goal orientation	50	4.62	1.45	65	4.69	1.19
Task value*	50	4.27	1.52	65	3.53	1.44
Expectancy: self-efficacy for learning	50	4.71	1.25	65	4.67	1.06
<i>Cognitive abilities (p = .337)</i>						
Pre-knowledge test (range: 0-14.5)	43	9.71	1.56	64	9.56	1.76
History performance level (range: 0-10)	50	7.12	1.04	65	6.81	1.12

* $p < 0.05$.

5

Chapter 5

Secondary Students' Online Self-Regulated Learning During Flipped Learning: A Latent Profile Analysis

This chapter is based on:

van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2021). Secondary Students' Online Self-Regulated Learning During Flipped Learning: A Latent Profile Analysis. *Computers in Human Behavior*, 118, 106676. <https://doi.org/10.1016/j.chb.2020.106676>

Acknowledgement of author contributions:

DvA, CP, JJ, and LK designed the study; DvA collected the data; DvA and JJ planned the data analysis; DvA analyzed the data; DvA drafted the manuscript; DvA, CP, JJ, and LK contributed to critical revision of the manuscript; CP, JJ, and LK supervised the study.

Supplementary material to this chapter can be found on:

<https://doi.org/10.34894/WRUQOG>

Abstract

Flipped learning (FL) makes greater use of students' self-regulated learning (SRL) skills when they regulate their online learning behavior. Previous research has shown the value of SRL support during FL to enhance students' SRL and learning outcomes. However, as previous studies have indicated that SRL behavior varies, should SRL support be tailored to these differences in SRL? We applied latent profile analysis to identify subgroups in 150 eighth-graders during FL. We used practically relevant online behavioral data to represent students' online SRL activities, which we gathered unobtrusively in an ecologically valid secondary educational classroom setting. We found five distinct SRL profiles from low completion and no activity to full completion and very high activity. In addition, students in the profile who showed low SRL activity achieved significantly worse learning outcomes than students in the three profiles with higher SRL activity. Finally, we explored whether SRL activity profile membership can be explained by student characteristics (i.e., self-reported SRL, motivation, and prior knowledge). None of the student-level variables predicted profile membership, but our approach offers leads for future research to further investigate the potential of tailored SRL support.

Introduction

The most basic definition of flipped learning (FL) is that it is a teaching method by which students learn instructional material before class (e.g., by watching videos) and apply the content of the instructional material during class (van Alten et al., 2019). Several meta-analyses have confirmed that FL enhances learning outcomes (e.g., van Alten et al., 2019; Cheng et al., 2019; Låg & Sæle, 2019). However, the role of students' self-regulated learning (SRL) skills in regulating their learning behavior in FL is less well validated in previous research. Self-regulated learning skills are utilized when students need to regulate where, when, and how they study instructional material outside of the classroom, especially if pre-class learning takes place in an online learning environment (He et al., 2016; Kim et al., 2014; Lee & Tsai, 2011; Shih & Huang, 2019; Sletten, 2015; Tan et al., 2017). As the increase in students' autonomy in pre-class online learning puts a higher demand on their SRL skills, this could lead to ineffective study behavior, such as inadequate time management due to the freedom of navigation or sequencing of instruction (Butzler, 2016; Lee et al., 2010; Sletten, 2015).

Some researchers have been exploring the question of whether a personalized learning approach, which aims to tailor the pace of learning and the instructional approach to the needs of individual learners (Xie et al., 2019), could offer a solution for the higher demand on SRL skills in online learning. Personalized learning has gained popularity amongst educators and researchers due to the increased possibilities that online education could offer, such as implementing intelligent learning systems and analyzing individual learners' data (Xie et al., 2019). Therefore, educational researchers question if and how online learning environments can be adapted to the individual learner (e.g., Vandewaetere et al., 2011). This personalization can be achieved in different forms. For example, what kind of feedback about learning behavior should be provided to the learner and the teacher to enhance learning (Ali et al., 2012; Sedrakyan et al., 2020)? Personalization of online learning environments also uses different personalization parameters (e.g., difficulty, sequence and pace of the materials, and student variables such as cognitive, SRL, and motivational abilities; Xie et al., 2019).

Previous research has shown the value of SRL support to the whole group of students during online pre-class FL. This *variable-centered approach* demonstrated that SRL support on average enhances students' SRL and learning outcomes (e.g., van Alten et al., 2020b; Lai & Hwang, 2016; Moos & Bonde, 2016; Yilmaz et al., 2018). However, given the personalized learning perspective, previous research on students' SRL in online learning environments showed that students' SRL skills vary greatly (Barnard-Brak et al., 2010; Broadbent & Fuller-Tyszkiewicz, 2018; Jovanović et al., 2017; Maldonado-Mahauad et al., 2018; Ning & Downing, 2015; Vanslambrouck et

al., 2019). The question that prompts the current study is whether the differences in students' SRL skills should form the basis for tailoring SRL support in online learning environments (Kizilcec et al., 2017; Lehmann et al., 2014; Vanslambrouck et al., 2019). Therefore, we take a *person-centered approach* to narrow this knowledge gap by identifying if subgroups exist amongst secondary education students corresponding to their online SRL behavior in an FL setting. The novelty of our contribution to the scientific debate is twofold. First, we used practically relevant and available online behavioral data that represented students' online SRL activities. Second, we unobtrusively gathered these data in an ecologically valid secondary educational classroom setting. This yields important insights, as no studies have been conducted to date in secondary education taking a *person-centered approach* towards online SRL activities. In addition, previous research suggested that focus on the development of SRL skills is of particular importance for younger students (Dignath & Büttner, 2008, 2018; Muijs & Bokhove, 2020; Ramdass & Zimmerman, 2011; Veenman et al., 2006; Wigfield et al., 2011).

The Demand on Younger Students' Self-Regulated Learning Skills in Online Learning Environments

Self-regulated learning has been conceptualized in different ways (Panadero, 2017). Most researchers agree that self-regulated learners regulate and monitor their own learning by making use of cognitive and metacognitive strategies and simultaneously regulate their motivation to perform these strategies. A common definition of SRL is that it is a cyclical process during which students sequentially move from the forethought phase via the performance phase to the self-reflection phase (Zimmerman, 2000; Zimmerman & Moylan, 2009). In the forethought phase, students analyze the task (e.g., goal setting and planning). Motivational beliefs, such as self-efficacy, task value, and goal orientation, affect students' task analysis and the activation of learning strategies (Panadero, 2017). In the performance phase, students work on the learning tasks, while actively metacognitively monitoring their progress. They perform self-control strategies such as time management and help seeking. Metacognition, which can be defined as knowledge, awareness, and regulation of one's learning, is one of the important skills for students to guide them through the SRL phases. In the self-reflection phase, students evaluate their performance and could adapt their approach to future learning.

Previous research has shown that students' SRL skills predict their learning outcomes during online FL pre-classroom preparations (Jovanović et al., 2019; Sun et al., 2018; Yilmaz & Baydas, 2017). For example, students with better SRL skills achieve improved learning outcomes during FL compared with those with poorer skills (Lee & Choi, 2019; Shibukawa & Taguchi, 2019). However, acquiring and performing SRL is difficult,

and in particular for younger students age plays an important role in developing SRL skills. First, although SRL skills begin to develop during pre-school and early-school years, they tend to strengthen and continuously develop during primary and secondary education (Muijs & Bokhove, 2020; Veenman et al., 2006). In addition, it seems that secondary school students have less confidence in their SRL skills in comparison with primary school pupils (Wigfield et al., 2006; Wigfield et al., 2011). This could mean that students in secondary education, with less confidence in their ability to regulate their learning than primary education pupils, set easier goals which could lead to a negative influence on learning outcomes (Wigfield et al., 2011). Research has shown that SRL instruction can help young students to develop SRL skills faster and in a more structured manner (Dignath & Büttner, 2008, 2018; Muijs & Bokhove, 2020). For example, SRL support has an impact on learning outcomes in secondary education if the SRL support includes elaboration of the application of SRL strategies that students have already acquired (Dent & Koenka, 2016; Dignath & Büttner, 2008).

Second, students' SRL skills also develop through maturation, interaction, and imitation of older peers, regardless of SRL instruction (Muijs & Bokhove, 2020). This implies that children develop SRL skills differently, according to the opportunities they receive at school but also at home, in which social background potentially also plays a role (Muijs & Bokhove, 2020; Veenman et al., 2006). Therefore, we argue that it is important to investigate the differences in SRL behavior in the secondary school context. We already know from previous research that SRL support can enhance students' SRL development and learning outcomes (Dignath & Büttner, 2008, 2018; Muijs & Bokhove, 2020). What we do not know yet is whether there are significant differences amongst secondary education students' online SRL behavior. If this is the case, SRL support in school environments can play an important role in recognizing these differences in SRL development amongst children to further enhance their confidence and skills in performing SRL activities. We argue that this is especially important for students with a social background that offers fewer opportunities to develop SRL at home (Muijs & Bokhove, 2020; Veenman et al., 2006).

Supporting Self-Regulated Learning During Flipped Learning

Previous research has found positive effects of SRL support on FL outcomes in primary education (e.g., Lai & Hwang, 2016), secondary education (e.g., van Alten et al., 2020b), and higher education (e.g., Moos & Bonde, 2016; Yilmaz et al., 2018). In these studies, students' SRL was supported during learning in (online) pre-class activities, for instance by SRL prompts in the instructional videos, a self-regulated monitoring system to provide students feedback on their SRL activities, or SRL instructions about effective SRL behavior. These studies showed that a one-size-fits-all approach to SRL support enhanced the learning outcomes of students who received the SRL support

in contrast to those without SRL support. This type of research can be characterized as a *variable-centered approach*: it describes associations among variables and assumes that the students are homogeneous with respect to how the predictors operate on the outcomes (Laursen & Hoff, 2006). This approach clarifies the relative importance of predictor variables in explaining variance amongst the included students with regards to outcome variables (Laursen & Hoff, 2006) and is helpful in demonstrating the main effects of SRL support on the average SRL behavior and learning outcomes of the student population.

In contrast, however, a *person-centered approach* distinguishes different subgroups amongst this population and assumes that the population is heterogeneous with respect to how the predictors operate on the outcomes (Laursen & Hoff, 2006). This approach describes differences among individuals in how variables are related to each other and identifies groups of individuals who share particular attributes or relationships among attributes (Laursen & Hoff, 2006). In SRL research, such an approach adds valuable insights, as it investigates the possibility that some subgroups benefit more from SRL support than others (Barnard-Brak et al., 2011; Broadbent & Fuller-Tyszkiewicz, 2018; Kizilcec et al., 2017; Laursen & Hoff, 2006; Ning & Downing, 2015).

Previous research that took a *person-centered approach* towards students' SRL in online learning environments (e.g., blended learning types such as FL or massive open online courses) showed the existence of student subgroups with divergent SRL profiles and learning patterns (Barnard-Brak et al., 2010; Broadbent & Fuller-Tyszkiewicz, 2018; Jovanović et al., 2017; Maldonado-Mahauad et al., 2018; Ning & Downing, 2015; Vanslambrouck et al., 2019). These studies used SRL self-report questionnaires for latent profile analyses (LPA) to identify and analyze the subgroups within the student population. However, previous research suggested the supplementation of self-report data with observational data (Panadero et al., 2016; Rovers et al., 2019). For example, because students' SRL behavior changes throughout a course, which cannot be captured in a time-fixed questionnaire (Wang, 2019). Self-reports also stimulate students to think about their own learning and therefore possibly also affect their SRL behavior (Greene & Azevedo, 2010). In addition, students can have difficulty in accurately assessing their own SRL activity (e.g., due to overconfidence) or are simply subject to response sets and memory biases (Li et al., 2020; Pekrun, 2020; Wang, 2019). Therefore, our aim in the current study was to complement self-report data by log data that were unobtrusively gathered in online learning environments. Recent studies used online traces of SRL behavior in LPA to identify patterns by which students can be grouped (Jovanović et al., 2017; Kovanović et al., 2015; Maldonado-Mahauad et al., 2018). In fact, other research has shown that online traces of SRL better predict SRL behavior that enhances learning outcomes compared with self-reports alone (Li et al., 2020; Rovers

et al., 2019; Wang, 2019). Several online SRL behavioral patterns have been identified, such as *comprehensive* learners, *targeting* learners, and *sampling* learners in the context of massive open online courses (Maldonado-Mahauad et al., 2018) and *intensive* learners, *strategic* learners, *highly strategic* learners, *selective* learners, and *highly selective* learners in the context of an FL setting (Jovanović et al., 2017). Students who were more active in regulating their learning (e.g., students showed more online SRL activities) also achieved improved learning outcomes compared with less active regulating students (Jovanović et al., 2017).

Current Study

By taking a *person-centered approach* toward online SRL behavior in secondary education, the current study addresses several existing knowledge gaps. First, there is a need to study the relationship between SRL and FL in secondary education, as previous research has shown the importance of supporting students' SRL at younger ages when they acquire and further develop their SRL skills (Dent & Koenka, 2016; Dignath & Büttner, 2008, 2018; Muijs & Bokhove, 2020). Moreover, previous SRL profile research has almost exclusively focused on blended learning contexts in higher education. One exception is the study of Abar and Loken (2010), who included a sample of 11th and 12th graders in high school, but the difference between this research and the current study is twofold. The intervention of Abar and Loken (2010) was mainly classroom based, and an online web-based study tool was presented to students as supplementary material. In our study, we examine eighth-grade secondary school students' SRL in an ecologically valid FL setting, where they watched instructional videos in the online learning environment before every class. Also, Abar and Loken (2010) based the clustering of students on students' self-reported SRL. In our study, we take online SRL activity trace data to cluster students. Accordingly, our first research question is: *Which SRL profiles can be identified for secondary education students during FL according to their online SRL activities?*

Second, empirical research has shown that students with better SRL skills revisit previously studied learning content more often than those with poorer SRL skills (Kizilcec et al., 2017). In addition, You (2016) found that regularity as indicator of SRL behavior (i.e., accessing and watching instructional videos) was the strongest predictor of learning outcomes. Therefore, it can be hypothesized that students with better SRL skills engage in online SRL activities more often than those with poorer SRL skills, which in turn will predict their learning outcomes. However, Vanslambrouck et al. (2019) suggested that a quantitative interpretation of SRL activities (i.e., more is better) could fall short, as students with fewer SRL skills may undertake more but inadequate SRL activities, and students with better SRL skills may undertake fewer but more efficient SRL activities (Wormington et al., 2012). To address this knowledge

gap, we relate the latent SRL profiles and their distinct SRL behaviors to their learning outcomes. Therefore, our second research question is: *To what extent do SRL activity profiles of secondary education students in FL differ in learning outcomes?*

Third, previous research has shown that students' self-reported motivation, self-reported SRL skills, and prior knowledge play a critical role in the motives of students to perform SRL activities (Abar & Loken, 2010; Kizilcec et al., 2017; Ning & Downing, 2015; Sun et al., 2018; Vanslambrouck et al., 2019; Yang et al., 2018). Therefore, the last question we address in the current study is to what extent the SRL activity profiles can be explained by student characteristics. If these variables on the student level predict SRL activity profile membership, this information can be used to guide further research into tailoring SRL support to other SRL-related individual differences (Kizilcec et al., 2017; Vanslambrouck et al., 2019; Wang, 2019). Therefore, our third research question is: *To what extent do self-reported motivation, self-reported SRL, and prior knowledge in secondary education students in FL predict SRL profile membership?*

Method

Participants

We asked 156 second-year students (grade 8, 13–14 years old) and their parents for their consent to participate. We excluded five students because consent was not obtained, and one student because of missing data. The final sample of 150 second-year students consisted of six pre-existing classes with three teachers (including the first author) who each taught two classes. The Netherlands has a tracked secondary education system (from grade 7 onwards), and the participating students were enrolled in the two highest tracks: senior general ($n = 21$) and pre-university level, consisting of *atheneum* level ($n = 119$) and *gymnasium* ($n = 10$) level (which includes Latin and Greek as compulsory courses). The study took place in the school's regular History curriculum in one secondary school in the Netherlands. This school is a large urban school (2200 students) that is generally representative of the majority of secondary schools in the Netherlands. Ethical approval was obtained from the Faculty Ethics Review Board, and students and parents gave permission for the authors to obtain students' data for study purposes.

Instruments

Online Self-Regulated Learning Activity Log Data

In online learning environments, trace data that indicates SRL behavior can be

gathered automatically and unobtrusively. Previous research has shown the potential of data-mining processes (e.g., sequential pattern mining) to gather these traces of students' SRL from online learning environments (Jovanović et al., 2017; Kovanović et al., 2015; Maldonado-Mahauad et al., 2018). In contrast to these previous studies, which used complicated clustering and pattern-mining techniques to extract SRL activity indications from large amounts of clickstream data, we used SRL activity indications from online learning environments in secondary education. These variables, such as video viewing time or *rewind actions*, are commonly provided by online learning environments and can be easily interpreted by teachers, which makes them practically relevant. We used Edpuzzle (www.edpuzzle.com) as a freely available online learning environment and collected the following online SRL activity variables: *completion rate*, *watch time*, *on time rate*, *rewind actions*, and *revisit actions*.

We obtained log data for each video that the students watched in the online learning environment over the course of eight weeks (eight videos in total). We aggregated the SRL activity variables from the extracted log data variables in the following way. *Completion rate* indicated to what extent the student watched each individual instructional video (on a scale from 0 to 100%). We calculated the total percentage of all the video completions for each individual student. *Watch time* was the total time a student spent on a particular instructional video. Edpuzzle did not take into account the paused time. This was unfortunate, as deliberate pausing actions could indicate SRL behavior (e.g., monitoring), but also fortunate, because this prevented a long *watch time* from being measured if students just ran the videos but were not watching them. We added the individual scores of each student to create a total sum of *video watch time*. *On time rate* indicated if a student completed the instructional video on time (i.e., before the deadline, which was the beginning of class), too late, or never completed the assignment. We coded this variable for each video (i.e., 0 for not complete, 1 for too late, 2 for on time) and created a sum score that indicates a student's overall degree of completing video instructions before the deadline. *Rewind actions* referred to occasions when a student watched a portion of a video more than once. Every video was automatically divided into 10 segments, and log data were available about how often a student watched a particular segment. We created a sum score for all the deliberate *rewind actions*, which we calculated by taking the total sum of portions viewed and subtracting the number of segments that indicated a first-time viewing action. *Revisit actions* indicated the number of occasions when students watched (a portion of) an instructional video on another part of the day (i.e., spacing principle). Edpuzzle provides trace data of the time when a student completed a video for the first time and the time when a student watched that particular video for the last time. For each video, we calculated if these dates were similar (i.e., no revisit action, 0) or different (i.e., at least one revisit action, 1) and created a sum score for all the videos. In comparing the first-time completed and last-watched dates, we set

a minimum interval of six hours, to ensure that this variable is an indication of a deliberate revisit action after one part of a day (e.g., not a difference of a few minutes due to technical issues).

Table 1
Relationship Between the Online SRL Activity Measurements and the Three-Phase Model of SRL by Zimmerman and Moylan (2009)

Online SRL activity measurement	Forethought phase	Performance phase	Self-reflection phase
On time rate	X		
Watch time		X	
Completion rate		X	
Rewind actions			X
Revisit actions			X

In Table 1, we indicate how these practically relevant trace-data variables relate to each phase of the SRL model (Zimmerman & Moylan, 2009). Each SRL phase is covered by at least one SRL activity measurement that best assesses that SRL phase. For example, student learning in the forethought phase requires planning and regulation of motivation to complete the video instructions before the deadline. Previous research has shown that late submissions from students in an online learning environment predict learning outcomes (You, 2016). This makes the *on time rate* a suitable indication of SRL activities related to planning skills in the forethought phase. In the performance phase, students control and regulate their learning behavior to complete learning tasks. Studies have shown that these trace data variables suitably represent SRL behavior; for example, it was found that low-achieving students often fail to successfully complete assignments (Jovanović et al., 2017). This makes *watch time* and *completion rate* suitable indications of SRL activities in the performance phase. In the self-reflection phase, students monitor and evaluate their understanding of the video and can decide to watch (a part of) the instructional video again (i.e., *rewind* and *revisit actions*). For instance, the variable *revisit actions* indicated a deliberate SRL action of a student to revisit already-completed videos at least six hours after completion. Table 2 presents the descriptive statistics for the online SRL activity variables.

Table 2

Descriptive Statistics for the Online SRL Activity Variables: Mean, Standard Deviation, Median, and Two-Tailed Pearson Correlation Coefficients (1–5)

Online SRL activity variable	<i>M</i>	<i>SD</i>	Median	Min. - Max.	2.	3.	4.	5.
1. Completion rate	94.33	16.39	100	0 – 100	.46*	.88*	.22*	.24*
2. Watch time	78.37	36.83	67	0 – 209	-	.45*	.96*	.81*
3. On time rate	14.35	3.20	16	0 – 16		-	.24*	.27*
4. Rewind actions	38.02	45.21	20	0 – 228			-	.82*
5. Revisit actions	2.57	3.01	1	0 – 8				-

Note. Interpretational guidance. 1 = mean percentage of watched videos (0 to 100); 2 = total minutes watching videos (total length without repeating portions was 54 minutes); 3 = high score (16) means that a student completed every video on time; 4 = all deliberate rewind actions; 5 = high score (8) means that a student revisited already-completed videos after more than six hours at least once.

* = $p < .01$.

Self-Regulated Learning and Motivation Questionnaire

To measure how students themselves assessed their SRL skills and motivation before the flipped classroom started, we developed a pre-test with a focus on learning tasks that they had to complete at home (i.e., homework for History). The Likert-scale range for each question was 1 (disagree) to 7 (agree). For each individual scale, we calculated means for the relevant items, and then we calculated two separate mean values for the SRL and motivation scores.

The first part of the questionnaire was largely based on the revised self-regulated online learning questionnaire (SOL-Q-R; Jansen et al., 2018). We maintained the questionnaire's seven constructs of SRL: *metacognition* (i.e., activities) *before learning* (e.g., "I ask myself questions about what I am to study before I begin"), *during learning* (e.g., "I am aware of what learning strategies I use"), and *after learning* (e.g., "After studying History, I reflect on what I have learned"), *time management* (e.g., "I make good use of my study time for History homework"), *environmental structuring* (e.g., "I find a comfortable place to complete my History homework"), *persistence* (e.g., "When History homework is difficult, I continue to keep working"), and *help seeking* (e.g., "When I have trouble learning, I ask for help"). We did not include *environmental structuring*, due to the low Cronbach's α (.55). We contextualized the questions to the student's age and environment context (e.g., we changed "I set specific goals before I begin a task in this online course" to "I set specific goals before I begin working on my History homework").

The second part of the questionnaire was largely based on the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al., 1991). The four constructs *intrinsic goal orientation*, *extrinsic goal orientation*, *task value*, and *self-efficacy for learning* were

used to measure how students valued and perceived these motivational aspects while completing homework for History. Table 3 presents the descriptive statistics of the pre-test self-report questionnaires' scales of SRL and motivation.

Prior Knowledge Test

We developed a prior knowledge test, which assessed how much students already knew about the content of the FL lessons. The test contained seven multiple-choice questions on the recall level (e.g., "Which concept does not belong to the Industrial Revolution?") and comprehension level (e.g., "Explain if there was a population growth or decline after the new agricultural developments in the 18th century"). Two additional open questions on the transfer level tested their knowledge about the Industrial Revolution by analyzing a historical source (e.g., "Name two differences between Robert Owen and other factory bosses in his day"). One item was deleted due to a negative distinctive value. The first author scored all the tests and created a sum score. As the items together measured an underlying construct that was difficult to measure (e.g., students with less prior knowledge tend to guess more), we accounted for a low reliability (8 items; $\alpha = .44$) in terms of Cronbach's α (Taber, 2018; Taub et al., 2014). Table 4 presents the descriptive statistics of the prior knowledge test.

Learning Outcomes Test

The learning outcomes test consisted of 24 items (four multiple-choice questions) and was similar to other tests that students were used to completing to assess their learning about different History topics. The test contained recall questions (e.g., "Define urbanization"), comprehension questions (e.g., "Explain if Karl Marx would support the idea of labor unions"), and transfer questions (e.g. "Analyze the following historical source and argue if this is a reliable source for the working conditions in the 19th century"). As each teacher scored their own students, the inter-rater reliability of the open question scores was assessed. The first author independently scored a random sample of five unscored tests from the other classes (20 tests and 400 items in total). These scores were compared with the other teachers' scores. One item was deleted due to a low Cohen's κ ($\kappa = .30$, 60% agreement rate). The inter-rater reliability for the remaining items was substantial: an average Cohen's κ of .80 (SD = 0.14, range $\kappa = .53$ to 1, $p < .001$) and an average total agreement of 85%. In general, the learning outcomes test was found to be reliable (23 items; $\alpha = .82$). Finally, we created a sum score for the learning outcomes test and used this in further analyses. Table 4 presents the descriptive statistics of the learning outcomes test.

Procedure, Design, and Course Materials

The students completed the questionnaire and the prior knowledge test during regular class time, before the flipped lesson series started. The final learning outcomes test was given to the students in a regular test week at the end of the school year, one week after the last lesson. Before the lesson series started, students were briefed about the workings of FL and how to access the videos in the online learning environment Edpuzzle.

The design and data of the current study originated from a previously conducted quasi-experiment, in which the researchers asked the *variable-centered* question as to what was the effect of SRL support in FL pre-class videos on students' SRL and learning outcomes (van Alten et al., 2020b). The lesson series was about the Industrial Revolution and consisted of 10 lessons, with eight instructional videos, over the course of eight weeks. The learning materials (both online and offline) were exactly the same for all the students, except for one difference in the online instructional videos. The content of the video was similar, but a total of 75 students received three SRL prompts at the start, middle, and end of the video (e.g., "Did you accomplish the goals you have set at the beginning of this knowledge clip? If not, what can you still do to reach them?"). In addition, they received a short (<60-second) spoken SRL instruction about effective SRL activities (e.g., instruction on why it is effective to think about setting goals before learning) during some of the videos. The other 75 students received no SRL prompts or SRL instruction.

The three participating teachers each taught two classes, in which both types of support occurred once. Teachers were instructed not to change their usual teaching interaction with their students. They received training on how to apply the learning materials for each lesson and were also asked to give feedback on all the learning materials before the intervention started. We provide a detailed analysis of the impact of these two types of support in the Appendix, which shows that it did not significantly affect SRL behavior in the current study.

The workbooks with additional content questions (e.g., "What are the advantages of building steam factories in the city?") and all the face-to-face lessons were the same for all the students. The videos had an average length of seven minutes and were developed and recorded by the first author. Fast-forwarding was disabled for students when they watched the video for the first time. Students were rewarded with a little extra credit on their final mark if they completed the video instructions on time.

Table 3

Descriptive Statistics for Each Questionnaire Scale (1–7) Of SRL And Motivation: Mean and Standard Deviation, Cronbach's α , and Pearson Correlation Coefficients (1–10)

Scale of the pre-test	Items	n	M (SD)	α	2.	3.	4.	5.	6.	7.	8.	9.	10.
<i>Self-regulated learning</i>													
1. Metacognition before	4	145	3.04 (1.23)	.78	.65**	.69**	.42**	.28**	.52**	.47**	.49**	.41**	.14
2. Metacognition during	6	145	3.95 (0.97)	.66	-	.66**	.52**	.44**	.59**	.47**	.36**	.46**	.25**
3. Metacognition after	4	147	3.02 (1.23)	.84		-	.47**	.32**	.54**	.47**	.41**	.46**	.18*
4. Persistence	5	147	4.44 (1.23)	.84			-	.26**	.60**	.60**	.47**	.60**	.48**
5. Help seeking	4	143	4.63 (1.21)	.77				-	.34**	.12	.28**	.26**	.04
6. Time management	4	146	4.14 (1.27)	.71					-	.57**	.42**	.58**	.31**
<i>Motivation</i>													
7. Intrinsic goal orientation	4	146	3.96 (1.15)	.70							.40**	.73**	.48**
8. Extrinsic goal orientation	4	144	4.77 (1.29)	.74							-	.45**	.27**
9. Task value	4	146	4.01 (1.52)	.91								-	.50**
10. Expectancy: self-efficacy	4	146	4.75 (1.13)	.81									-

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

Table 4

Descriptive Statistics for the Prior Knowledge and Learning Outcomes Tests

Cognitive variable	n	M	SD	Median	Min.-Max.	Correlation 1–2
1. Prior knowledge pre-test	141	9.62	1.89	9.50	4–15.50	.30**
2. Learning outcomes post-test	149	18.41	5.73	18.50	3–29.50	-

Note. ** = $p < .001$

During the 65-minute class time, the teachers started with retrieval practice activities about the content of the videos, and students could ask questions. This was usually followed by a complementary micro lecture (approximately 10 minutes) to address students' misconceptions that came forward during the retrieval practice activities and to provide more in-depth aspects of the learning content from the video (e.g., appealing examples and primary source material). For the greatest part of class time, students engaged in learning activities that applied the learning materials (e.g., analyzing historical sources or creating advertisement posters about industrial machines).

Data Analysis

To answer the first research question as to which SRL profiles can be identified according to the online SRL activity variables, we performed LPA in Mplus 8.3 (Muthén & Muthén, 2012) with maximum likelihood estimation. As our trace data also provides values for students who did not engage in SRL activities online, we had no missing data to address. Latent profile analysis consists of evaluating a series of iterative LPA models, in which each model adds the possibility of one more profile (Ferguson et al., 2019). We explored the hypotheses of one to seven different profiles, and the following model fit criteria were used to evaluate the best model fit and interpretability (Ferguson et al., 2019; Vanslambrouck et al., 2019). Model retention in LPA is evaluated by the Bayesian Information Criterion (BIC), Sample-Adjusted BIC (SABIC), and Akaike's Information Criterion (AIC), for which lower values indicate a better fit (Ferguson et al., 2019). Log-likelihood values can be used to determine where the difference between models (i.e., lower indicates better fit) starts to become relatively small (Ferguson et al., 2019; Masyn, 2013). Entropy was used as an indication of classification uncertainty. Higher entropy values indicate a model that better divides the data into profiles, with values of $>.80$ supporting the minimal uncertainty of the profile classification of students (Celeux & Soromenho, 1996; Ferguson et al., 2019; Masyn, 2013). The Lo, Mendell, and Rubin (LMR) test was used to test the likelihood ratio of one model to the model with $k - 1$ profiles (Lo et al., 2001). The LMR test assesses significance across differences in degrees of freedom, and significance indicates that the model with fewer profiles fits the data significantly worse than the model with more profiles (Ferguson et al., 2019; Y. Lo et al., 2001). We considered the sample size of the smallest latent profile, as profiles with less than 5% of the total sample should be carefully considered for interpretability (Ferguson et al., 2019; Masyn, 2013). Finally, we further tested with a one-way MANOVA the statistical significance between the latent SRL profiles identified, with SRL profile membership as an independent variable and the online SRL activity variables as dependent variables (e.g., Kovanović et al., 2015; Vanslambrouck et al., 2019). For completeness, we also performed a one-way MANOVA with SRL profile membership as the independent variable and the student level pre-test variables as dependent variables to further ascertain differences between SRL profiles.

To answer the second research question of whether there is a difference in learning outcomes between the latent SRL profiles, we conducted a one-way ANOVA in SPSS with the learning outcomes test as the dependent variable and SRL profile membership as the independent variable. *Post hoc* comparisons between the SRL profiles were used to further explore differences in learning outcomes between SRL profiles.

To answer the third research question of whether certain student-level variables predict SRL profile membership, we conducted a multinomial logistic regression with profile membership as the dependent variable and the individual scales of the self-reported motivation and SRL questionnaire and also the prior knowledge test as predictor variables in one model. We centered the predicting variables around the mean.

Results

Self-Regulated Learning Activity Profiles

Selection of the Number of Clusters

As presented in Table 2, the standard deviations for the online SRL activity variables are rather high (especially for *watch time*, *rewind actions* and *revisit actions*). This confirms our hypothesis that there is large variance in students' SRL behavior.

Table 5 contains seven different model fit statistics that were used to evaluate which LPA model fits the data best. We concluded that model 5 was the best model fit to the data based on the lower log-likelihood, AIC, BIC, and SABIC values, the higher entropy value, and the nonsignificant LMR test. The log-likelihood, AIC, BIC, and SABIC values showed relatively large decreases in the comparison between model 1 and 2, 2 and 3, until the difference between models 5 and 6. Entropy was best (i.e., $>.80$) in models 3, 4, and 5. The results of the LMR test indicate that model 5 was a better fit than model 4, and that model 4 was not a better fit than model 3. The smallest profile was $<5\%$ in models 3–7, which could indicate a statistically spurious profile. However, closer examination of this profile revealed that this small profile was aligned with theory and previous LPA in education, as this was the small group of students who engaged very little in the online learning activities. As a means of control, we also performed an LPA analysis in which we excluded these non-active students and concluded that the clustering of the other profiles remained stable. Therefore, we argue that it is justifiable to consider models 3–6 including the non-active students' distinguishable cluster. Overall, model 5 indicating five different SRL profiles fits the data best.

Description and Analysis of the Self-Regulated Learning Profiles

Table 6 contains descriptive statistics of the five SRL profiles we found. We labeled each SRL profile after careful analysis and interpretation of the differences (cf. Ferguson et al., 2019). In these interpretative labels, we contrast the two main differences between profiles: the *completion rate* of the instructional videos and the indication of the relative intensity of SRL activities that involved watching the videos such as rewinding, revisiting, and completion on time. We regarded 12% *completion rate* as low, 73% as medium, 99% as high, and 100% as full. We regarded virtually no *rewind* and *revisit actions* and a very low *on time rate* as no SRL activity, in comparison with low SRL activity (50% *on time rate*, few *rewind* and *revisit actions*), medium SRL activity (94% *on time*, few *rewind* and *revisit actions*), high SRL activity (99% *on time*, more than 4 times more *rewind* and *revisit actions* than medium), and very high SRL activity (99% *on time*, approximately 2 times more *rewind* and *revisit actions* than high). Figure 1 visualizes the differences between the profiles based on standardized z-scores for each SRL activity variable. The zero line represents the mean score for all the students for that particular variable.

Table 5
Latent Profile Analysis Model Fit Summary for Different Models

Model	# of free parameters	AIC	BIC	SABIC	Log-likelihood	Entropy	Smallest profile %	LMR <i>p</i> and meaning
1	10	5887.09	5917.20	5885.55	-2933.55	-	-	-
2	16	5465.74	5513.91	5463.27	-2716.87	0.978	27	.010 2>1
3	22	5175.46	5241.70	5172.07	-2565.73	0.989	4	.071 3<2
4	28	4960.75	5045.05	4956.43	-2452.38	0.988	3	.090 4<3
5	34	4812.93	4915.29	4807.68	-2372.46	0.988	3	.030 5>4
6	40	4725.90	4846.33	4719.73	-2322.95	0.978	3	.133 6<5
7	46	4670.86	4809.35	4663.77	-2289.43	0.977	1	.707 7<6

Note. *n* = 150; The LMR test compares the current model with a model with *k* - 1 profiles. AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; SABIC = Sample-Adjusted BIC; LMR = Lo-Mendell Ruben Likelihood Ratio Test.

Table 6

Descriptive Statistics (Mean and Standard Deviation) of the Five SRL Profiles for Online SRL Activity, and The Bonferroni Corrected Post Hoc Pairwise Comparisons Among the Profiles

Online SRL activity variable	Low completion no activity _a (n = 4, 3%)	Medium completion low activity _b (n = 13, 9%)	High completion medium activity _c (n = 92, 61%)	High completion high activity _d (n = 31, 21%)	Full completion very high activity _e (n = 10, 7%)
1. Completion rate	11.88 (15.02) _{bcd}	72.69 (12.76) _{acde}	98.61 (3.49) _{ab}	99.52 (2.25) _{ab}	100.00 (0.00) _{ab}
2. Watch time	3.00 (5.35) _{bcd}	50.00 (19.89) _{acde}	64.15 (11.17) _{abde}	112.61 (16.39) _{acde}	170.10 (24.82) _{abcd}
3. On time rate	1.00 (1.15) _{bcd}	8.69 (2.84) _{acde}	15.14 (1.26) _{ab}	15.65 (0.71) _{ab}	15.80 (0.42) _{ab}
4. Rewind actions	1.25 (1.89) _{de}	17.00 (18.23) _{de}	15.37 (13.46) _{de}	79.16 (20.78) _{acde}	160.90 (32.08) _{abcd}
5. Revisit actions	0 (0) _{de}	0.46 (1.13) _{de}	1.01 (1.43) _{de}	6.90 (1.04) _{abc}	7.30 (0.67) _{abc}

Note. Subscripts A-E indicate significant differences ($p < .05$) from other profiles according to the Bonferroni corrected *post hoc* tests. Interpretational guidance: 1 = mean percentage of watched videos (0 to 100); 2 = total minutes watching videos (total length without repeating portions was 54 minutes); 3 = high score (16) means that a student completed every video on time, 4 = all deliberate rewind actions; 5 = high score (8) means that a student revisited already completed videos after at least six hours at least once.

5

Figure 1

Comparison of SRL Profile (n) Online SRL Activity Variables in Z-Scores

We conducted a one-way MANOVA with SRL profile membership as the independent variable and the online SRL activity variables as dependent variables. This was performed to further ascertain statistically significant differences between the five SRL profiles. We followed Kovanović et al. (2015) using the Pillai's trace statistic in combination with the conservative Bonferroni correction for *post hoc* testing, as it is more robust to the assumption of homogeneity of covariances violation (Field, 2018).

We found a significant main effect of profile membership on the online SRL activity variables, Pillais' trace = 2.34, $F(20, 576) = 40.94$, $p < .001$; partial $\eta^2 = .59$. Results of the Bonferroni *post hoc* pair-wise tests are provided in Table 5, with subscripts indicating significant differences between profiles.

For a full description of the SRL profiles, we provide the profile means and standard deviations on the scales of the SRL and motivation pre-test, the prior knowledge pre-test, and the learning outcomes post-test in Table 7. Figure 2 visualizes the differences between the profiles based on standardized z-scores for these variables. The zero line represents the mean score for all the students for that particular variable. The included variables were used to answer research questions 2 and 3.

In addition, we assessed differences between students' pre-test SRL profiles with a one-way MANOVA, with SRL profile membership as the independent variable and the student-level pre-test variables (i.e., self-reported SRL, self-reported motivation, and prior knowledge) as dependent variables. After assessing the equality of covariances using the Box's M test and homogeneity of variances using Levene's test, we performed separate non-parametric Kruskal–Wallis tests for the student-level pre-test variables due to violation of assumptions. We found a significant different distribution across SRL profiles ($p = .011$) for *help seeking*, but not for the other variables (p ranges from .105 to .987). Only significant Kruskal–Wallis tests were followed up by Bonferroni corrected pairwise comparisons, and results are provided in subscripts in Table 7.

Relating Self-Regulated Learning Profiles to Learning Outcomes

To answer the second research question about differences amongst the SRL profiles for learning outcomes, we conducted a one-way ANOVA with the learning outcomes test as the dependent variable and SRL profile membership as a factor. First, homogeneity of variance amongst SRL profiles was checked using Levene's test. This showed that the variances for the learning outcomes test were equal, $F(4,144) = 1.48$, $p = .211$. We found a significant effect of SRL profile membership on learning outcomes, $F(4, 144) = 4.02$, $p = .004$; partial $\eta^2 = .10$. Hochberg's GT2 test was used for *post hoc* comparisons, as the sample sizes for the profiles varied greatly (Field, 2018). *Post hoc* comparisons indicated significantly ($p < .05$) lower means for students in the profile *medium completion–low activity* in contrast with students in the profiles *high completion–medium activity*, *high completion–high activity*, and *full completion–very high activity*. All other comparisons were not significant. Means and standard deviations of the learning outcomes test are provided in Table 7.

Table 7*Descriptive Statistics (Mean and Standard Deviation) of the Five SRL Profiles in the pre- and post-tests*

Variables	Low completion no activity ^a (n = 4, 3%)	Medium completion low activity ^b (n = 13, 9%)	High completion medium activity ^c (n = 92, 61%)	High completion high activity ^d (n = 31, 21%)	Full completion very high activity ^e (n = 10, 7%)
<i>Self-regulated learning pre-test</i>					
Metacognition before	2.56 (1.26)	3.88 (1.38)	3.03 (1.25)	2.74 (1.06)	3.25 (1.01)
Metacognition during	3.42 (1.09)	4.47 (0.99)	3.93 (0.95)	3.81 (1.05)	4.17 (0.64)
Metacognition after	2.44 (0.99)	3.69 (1.39)	2.92 (1.19)	2.96 (1.24)	3.55 (1.22)
Persistence	3.35 (1.54)	4.90 (1.03)	4.44 (1.25)	4.33 (1.26)	4.68 (0.87)
Help seeking	2.75 (0.89) _{be}	5.06 (0.98) _a	4.57 (1.17)	4.63 (1.28)	5.33 (0.96) _a
Time management	4.13 (1.05)	4.56 (1.14)	4.21 (1.31)	3.76 (1.28)	4.18 (0.94)
<i>Motivation pre-test</i>					
Intrinsic goal orientation	4.25 (0.46)	4.38 (1.18)	3.95 (1.18)	3.76 (1.20)	4.10 (0.84)
Extrinsic goal orientation	4.50 (1.67)	5.03 (1.36)	4.73 (1.31)	4.73 (1.26)	5.05 (1.24)
Task value	4.13 (1.81)	4.31 (1.90)	3.98 (1.46)	3.95 (1.59)	4.14 (1.52)
Expectancy: self-efficacy	5.00 (0.89)	4.58 (1.22)	4.74 (1.17)	4.77 (1.13)	4.86 (0.78)
<i>Cognition pre-test</i>					
Prior knowledge	7.67 (1.53)	8.35 (2.13)	9.23 (1.75)	9.26 (1.70)	9.72 (1.80)
<i>Cognition post-test</i>					
Learning outcomes	16.50 (3.58)	13.69 (5.25) _{cde}	18.54 (5.84) _b	18.97 (4.64) _b	22.78 (5.65) _b

Note. Subscripts A-E indicate significant differences ($p < .05$) from other profiles according to the Bonferroni corrected post hoc tests for the pre-tests, and Hochberg's GT2 correction for the post-test. Interpretational guidance: Likert-item range for the SRL and motivation tests was 1–7. Maximum score for the prior knowledge test was 14.50, and for the learning outcomes test 30.

Table 8*Results of the Multinomial Logistic Regression Predicting SRL Profile Membership*

Variables	Profile medium completion–low activity (n = 12) vs. Profile high completion–medium activity (n = 87)			Profile medium completion–low activity (n = 12) vs. high completion–high activity (n = 29)			Profile medium completion–low activity (n = 12) vs. full completion–very high activity (n = 9)		
	B	SE	p	B	SE	p	B	SE	p
<i>Self-regulated learning pre-test</i>									
Metacognition before	-.32	.40	.426	.78	.47	.097	.58	.60	.332
Metacognition during	-.41	.60	.494	.35	.65	.589	.72	.80	.364
Metacognition after	-.23	.40	.560	.24	.46	.608	.61	.58	.295
Persistence	-.50	.48	.300	.26	.52	.613	.25	.63	.691
Help seeking	-.22	.34	.518	.11	.37	.766	.50	.49	.314
Time management	.32	.43	.453	.21	.48	.663	.44	.66	.503
<i>Motivation pre-test</i>									
Intrinsic goal orientation	-.51	.51	.294	.69	.56	.217	.31	.72	.665
Extrinsic goal orientation	.26	.38	.408	.40	.41	.320	.57	.49	.245
Task value	.32	.36	.340	.42	.40	.295	.08	.52	.877
Expectancy: self-efficacy	.40	.37	.321	.44	.42	.293	.45	.55	.421
<i>Cognition</i>									
Prior knowledge test	.22	.20	.272	.23	.23	.308	.48	.30	.114

Note. The profile *medium completion–low activity* served as the reference group. Pseudo $R^2 = .19$ (Cox & Snell), .22 (Nagelkerke), .10 (McFadden). Model $\chi^2(33) = 248.19, p = .699$.

Figure 2Comparison of SRL Profile (*n*) pre-test and post-test Measurements in Z-Scores

Predicting self-regulated learning profile membership

To answer the third research question about which student-level variables predict SRL profile membership, we conducted a multinomial logistic regression with the scales of the SRL and motivation questionnaire pre-tests and the prior knowledge test as predictors and profile membership as the dependent variable (e.g., Vanslambrouck et al., 2019). We removed the SRL profile *low completion–no activity* from the analysis due to missing values that lead to an even smaller sample size than the original $n = 4$. We set the profile *medium completion–low activity* as the reference category, as this sample had the lowest *completion rate* and SRL activity that contrasted best with the other SRL profiles. Results of the multinomial logistic regression analysis are provided in Table 8. The nonsignificant model fit indicates that the full model does not predict profile membership better than the intercept-only model. In combination with nonsignificant likelihood tests in which the predictive value of the independent variables is tested for membership of the profile *medium completion–low activity* in comparison with membership of the other profiles, this leads us to conclude that none of the predictors in the model predict profile membership.

Discussion and Conclusions

By taking a *person-centered approach* to online SRL activity trace data from a secondary education FL course, we aimed to show differences in students' SRL that can be used to guide further research on tailored SRL support according to individual differences in SRL skills. We asked (1) which SRL profiles can be identified according to students' online SRL activities, (2) if these SRL profiles differ in learning outcomes, and (3) if self-reported motivation, self-reported SRL, and prior knowledge predict SRL profile membership. In line with previous research in higher education, We found five distinct online SRL profiles from low completion and no activity to full completion and very high activity. We also found that some of these SRL profiles significantly differed in learning outcomes. However, we found no student-level predictor variables to explain SRL profile membership. We interpret the results of the research questions through a detailed description of the similarities and differences between these SRL profiles.

Online Self-Regulated Learning Activity Profiles: Similarities and Differences

The profile *low completion-no activity* was the smallest group of students (3%), who did not follow instructions and engaged very little in watching the instructional videos. The students in this profile scored 6.0 (out of 10) points in the learning outcomes test. This is comparable to a C grade, and it was not statistically different to other profiles. As the sample size of the profile *low completion-no activity* was small, we cannot draw general conclusions about the relationship between their average learning outcomes and SRL skills. In contrast to previous research in which the majority of students were found to belong to a highly selective or even non-user group, this group was particularly small in our study (Broadbent & Fuller-Tyszkiewicz, 2018; Kovanović et al., 2015; Ning & Downing, 2015). This could be due to a difference in educational level (secondary vs. higher education), a more binding educational context (e.g., stricter secondary school culture of following homework instructions), or the incentive that was given to students to complete the instructional videos (e.g., a few extra credits on their final mark; see Radhakrishnan et al., 2009).

The profile *medium completion-low activity* comprised a small group of students (9%) who struggled to complete all the instructional videos (73%) and generally failed to complete half of the instructional videos before the deadline. This could be an indication of insufficient planning skills in these students, but further research with a broader view on SRL skills (e.g., planning) is needed to confirm this hypothesis. Their average watch time was slightly shorter than the total length of the instructional videos and was significantly different from all of the other profiles. On average, the students in this profile scored 5.1 (out of 10) points in the learning outcomes test, which is

comparable to an F grade. This score, which indicates a fail grade, was statistically different from the learning outcomes of students from the three profiles with high to full completion. This is in line with previous research that also found that students who were less active in regulating their learning achieve worse learning outcomes (Bannert et al., 2014; Jovanović et al., 2017).

The profile *high completion–medium activity* included the majority of students (61%), who generally completed all the instructional videos and completed them before the face-to-face class. The students in this profile generally scored 6.6 (out of 10) points in the learning outcomes test, which is comparable to a B grade. They also seem to be able to plan well and persist in their online learning activities (i.e., sufficient completion and *on time rate*). Similarly to students in the profile *medium completion–low activity*, they performed few monitoring activities (i.e., *rewind* and *revisit actions*). In this respect, this group shows behavior comparable to the selective SRL profiles in previous studies, in which students seem to do only the minimum to pass the course (cf. Barnard-Brak et al., 2010; Jovanović et al., 2017; Kovanović et al., 2015). Similarly, other studies found that the majority of the students were clustered in an SRL profile that can be described as average SRL behavior (Abar & Loken, 2010; Vanslambrouck et al., 2019).

The profile *high completion–high activity* was the second-largest subgroup of students (21%), who also generally completed all the instructional videos and before the deadline. The students in this profile generally scored 6.7 (out of 10) points in the learning outcomes test, which is comparable to a B grade. Only a few students did not complete a few portions of the instructional videos. In contrast to students in the *medium completion–low activity* profile, they performed significantly more *rewind* and *revisit actions*, and thus also spent significantly more time watching the instructional videos. These results are similar to previous research that also found a profile of students who performed more SRL activities, but achieved as well as competent regulators in terms of learning outcomes (Barnard-Brak et al., 2010).

The profile *full completion–very high activity* consisted of a small proportion of the students (7%), who fully completed each portion of the instructional videos. The average students' score of 7.8 (out of 10) points in the learning outcomes test, which is comparable to an A grade, was not statistically different from the learning outcomes of students in the profiles *high completion–medium activity* and *high completion–high activity*. Students from the profile *full completion–very high activity* performed significantly more *rewind actions* and thus spent more time watching than students from all other profiles. In addition, students from this group revisited the instructional videos significantly more than students in profiles with low to medium SRL activities. Therefore, the profile *full completion–very high activity* is comparable to the super regulators profile described in previous research (Broadbent & Fuller-Tyszkiewicz, 2018).

Theoretical Implications

One contribution of the current study is that it offers several nuances to the scientific debate about the quantitative interpretation of SRL activities. One of the supposed benefits of FL is that students in the online pre-class learning phase can revisit content and re-watch instructional videos without restrictions (Abeysekera & Dawson, 2015; Lo et al., 2017; Lundin et al., 2018). On the one hand, previous research found that more SRL behavior leads to better learning outcomes (e.g., You, 2016). In our study, we indeed found that students from the profile *medium completion–low activity* undertook significantly fewer SRL activities than students from the profiles *high completion–high activity* and *full completion–very high activity* and achieved significantly worse learning outcomes. In contrast, however, students from the profiles *high completion–high activity* and *full completion–very high activity* did not receive higher learning outcomes than students in the *high completion–medium activity* profile. On the other hand, it has been suggested in previous research that students with poorer SRL skills may undertake more but inadequate SRL activities and students with better SRL skills may undertake fewer but more efficient SRL activities (Vanslambrouck et al., 2019; Wormington et al., 2012). In our study, we found that students from the profile *full completion–very high activity* performed significantly more SRL activities than students from the profiles *high completion–high activity* and *full completion–very high activity*, but achieved similar learning outcomes. Although there is no conclusive evidence for the discussion, our results suggest that the first interpretation does exclude the second interpretation.

Our study expands the current knowledge about *person-centered approaches* to online SRL behavior to eighth-grade secondary education. Moreover, we used practically relevant and commonly provided variables from an authentic learning context and found results comparable to previous research in higher education (Jovanović et al., 2017; Kovanović et al., 2015; Maldonado-Mahauad et al., 2018). It is for future research to confirm that our proposed method of relating behavioral data to SRL activity was accurate by testing it on a larger scale and in different subjects and educational levels within secondary education.

Practical Implications

As we found no student-level variables that significantly predicted profile membership, our findings provide no concrete leads on how to tailor SRL support in FL environments. However, our results can still aid practitioners of FL in tailoring their SRL support in FL environments. For example, teachers and researchers may want to focus their SRL support on students in the profile *medium completion–low activity* in particular, as they generally failed to pass the course. This group of students failed to watch more than three-quarters of the instructional videos, and for approximately half of the time also not before the deadline. It seems that this group could benefit from specific SRL

support in the forethought phase, such as goal setting, strategic planning, and self-motivational beliefs. This can be achieved by instruction in the classroom (e.g., Ness & Middleton, 2012) or by personalized online prompts focused on the forethought phase (e.g., Lehmann et al., 2014).

Another example relates to the students in the profile *low completion-no activity*. Previous research indicates that insufficient SRL skills can lead to non-compliance (Eriksson et al., 2017). Although we did not measure the complete set of SRL skills of the students (e.g., how they regulate their own learning outside the online learning environment), their online SRL behavior could be used as a starting inference. The SRL of this group of students could, for example, be supported by enhancing their intrinsic motivation to watch and interact with the videos. Due to the small sample size of this subgroup, it was not possible in the current study to clarify if, for example, low (self-reported) motivation or low (self-reported) SRL skills are underlying factors that could explain this behavior. However, external motivators such as extra credits seemed insufficient in persuading this group to engage in the homework activities. Supporting the intrinsic motivation of these students seems a requirement for them to engage in SRL activities (cf. Lehmann et al., 2014). Furthermore, this small sample of students was also quite negative about seeking help from their teacher or peers and thus seems to prefer learning on their own. Further research could reveal whether it helps to specifically teach the students from this profile the value and usefulness of help-seeking activities (cf. Karabenick & Gonida, 2018)."abstract": "No longer viewed as evidence of dependency, seeking help when needed is now considered an important self-regulated learning (SRL. However, as this group generally scored a sufficient grade, it is also possible that these students' SRL skills are adequate and that they learned just enough by attending classes, or that they studied other offline available learning materials (e.g., their textbook).

Moreover, students from the profile *high completion-high activity* scored almost the same as those from the profile *high completion-medium activity* in the learning outcome and prior knowledge tests and demonstrated comparable *on time* and *completion rates*. However, what stands out is that the students from the former profile performed significantly more *rewind* and *revisit actions* and thus spent more time on the online learning activities, but achieved as highly as the majority of students in the latter profile who performed fewer of these monitoring activities (cf. Barnard-Brak et al., 2010). On the one hand, one could argue that the students from the profile *high completion-high activity* need all these SRL activities to score as highly as the students from the other profile. On the other hand, one could argue that, while their behavior does not lead to higher learning outcomes, students who seem to perform SRL activities excessively often could benefit if their self-efficacy, efficiency, confidence, or merely the quality of their SRL activities is supported.

Limitations and Directions for Future Research

Previous research has found that students' self-reported motivation, self-reported SRL skills, and prior knowledge are related to differences in students' SRL (Abar & Loken, 2010; Kizilcec et al., 2017; Ning & Downing, 2015; Sun et al., 2018; Vanslambrouck et al., 2019; Yang et al., 2018). However, none of the self-reported SRL scales, self-reported motivation scales, and prior knowledge variables predicted profile membership.

First, it is possible that other predictive variables that we did not measure may play a role. For example, there is modest evidence that students' SRL is related to their socio-economic background, which in turn could affect students' online SRL activities (Muijs & Bokhove, 2020). In addition, students' anxiety levels could also predict students' online SRL activities (Broadbent & Fuller-Tyszkiewicz, 2018). More research in secondary education is needed to determine how we can predict SRL behavior on the basis of student-level variables, which in turn can guide the design of SRL support.

Second, the lack of prediction could be due to the higher accuracy of predicting SRL behavior by SRL behavior measurements than by self-reports (Li et al., 2020; Rovers et al., 2019; Wang, 2019). While we used online behavioral data to find SRL profiles, we used self-reports to measure self-reported SRL and motivation. It could therefore also be possible that the Dunning-Kruger effect plays a role, such that students with poor SRL skills tend to overestimate their SRL ability (Kruger & Dunning, 1999). This is especially likely given the relatively young student population in the current study.

Third, the measurement of prior knowledge had low reliability, which could mean that these differences may not have been optimally measured. We therefore advise treating this variable with caution. Although our measurement of prior knowledge was of a similar nature to that in other studies (e.g., Taub et al., 2014; Yang et al., 2018), we advise careful consideration of the reliability of measuring prior knowledge and suggest exploring different methods of including this variable in multinomial logistic regression analyses to predict profile membership of secondary education students. For example, Ning and Downing (2015) used prior academic performance (i.e., GPA) as an indication of cognitive ability and prior knowledge, while Sun et al. (2018) asked students to report on their highest level of similar course content prior to the course, in addition to the measurement of pre-class homework grades as an indicator of prior domain knowledge.

Finally, we acknowledge that the following limitations may have an impact on the generalizability of our results. Although the participating school is generally representative of urban secondary schools in the Netherlands, more research is needed to reduce the possible bias of differences in schools that may affect students' differences in online SRL behavior (for example caused by school-wide SRL training

in the school's curriculum). Moreover, as the subsample sizes of the profiles *low completion-no activity* and *full completion-very high activity* are quite small, we have to be cautious with our interpretations of these profiles. More research with larger subsample sizes is needed to determine, for example, whether the extraordinarily high SRL activity of students in the profile *full completion-very high activity* is related to prior knowledge (Taub et al., 2014; Yang et al., 2018) or improved learning outcomes (Jovanović et al., 2017, 2019).

In addition, the ratio of student numbers in the profiles could possibly be affected. First, it may be affected by a novelty effect, as the FL method and the use of the online learning environment was new for the students. Although this does not explain the differences in SRL behavior between students, as it was new for all the participants, it might have caused more students than would otherwise have been expected to be included in SRL profiles with higher *completion rates* just because it was new and interesting for them. Second, we provided a small number of extra credits to students for watching the instructional videos to reduce non-compliance that would lead to a decrease in our sample size. This incentive was the same for every student and is common in this natural educational context. It is likely that the same SRL profiles would have been found without this external incentive. However, it could be an explanation for the very small sample size of the *low completion-no activity* profile.

Conclusions

Broad academic consensus has been reached on the importance of sufficient SRL support in the development of SRL skills (Muijs & Bokhove, 2020; Quigley et al., 2018). Previous research has shown that one effective way to implement this is by supporting students' SRL in an FL context (e.g., van Alten et al., 2020b; Lai & Hwang, 2016; Moos & Bonde, 2016; Yilmaz et al., 2018). However, less is known about the extent to which SRL support should be personalized to students' differences in SRL skills. By taking a *person-centered approach*, the current study revealed clear differences in SRL behavior amongst secondary education students in an ecologically valid FL context. These differences are also related to differences in learning outcomes, as students in the profile *medium completion-low activity* failed in the learning outcomes test and scored significantly worse than those in the profiles with higher SRL activity. The differences in online SRL behavior, which are related to differences in learning outcomes, suggest that, compared with one-size-fits-all SRL support, personalized SRL support could be an important improvement for learning environments which aim to support students' SRL.

Appendix: Analyses on the impact of two types of self-regulated learning support

The data from this study originate from a previously conducted quasi-experiment. The learning materials (both online and offline) were exactly the same for all the students, except for one difference in the online instructional videos. The content of the video was similar, but a total of 75 students received video-embedded SRL support, while the other 75 students received no SRL support. Here, we provide a detailed analysis to determine whether these two types of support affected SRL behavior in the current study.

First, we checked whether there was a difference in prior knowledge for educational levels, which would have required us to take this into account in later analyses. Participating students were enrolled in two educational levels: senior general ($n = 21$) and pre-university level, which consists of *atheneum* level ($n = 119$) and *gymnasium* ($n = 10$) level (which includes Latin and Greek as compulsory courses). We found a significant main effect of class level (independent variable) on prior knowledge as the dependent variable in an ANOVA, $F(2, 138) = 3.74, p = .026$, partial $\eta^2 = .05$. Subsequently, we found no significant main effect of SRL support variant (independent variable) on prior knowledge as the dependent variable in an ANOVA, $F(1, 139) = 2.43, p = .121$, partial $\eta^2 = .02$. This indicates that the students with different prior knowledge were equally distributed over both types of SRL support.

Second, we checked *a priori* differences in the SRL support variant in self-reported SRL and motivation. We found no significant difference between the two SRL support variants (independent variable) in a MANOVA including the 11 SRL and motivation self-reported pre-test scales as dependent variables, $F(10, 138) = 1.33, p = .221$; Wilk's $\Lambda = 0.912$, partial $\eta^2 = .09$. Therefore, the students from both SRL support variants did not differ in relevant *a priori* measurements.

Third, we analyzed the possible impact of SRL support on students' actual online SRL behavior. We performed a MANOVA with *completion rate*, *on time rate*, *rewind actions*, and *revisit actions* as dependent variables, SRL support variant as the independent variable, and class level as the controlling factor given the differences in prior knowledge for class level. *Watch time* was omitted from the analysis because of a very high correlation ($R > 0.95, p < .01$) with *rewind actions* and the problem of multicollinearity. We excluded nine students due to non-compliance, as they did not watch at least 70% of the instructional videos (cf. van Alten et al., 2020b). For the students in the SRL support group, non-compliance indicates that they were not sufficiently exposed to our intervention, which hinders a fair comparison of the impact of the SRL support variant. We found no significant difference between the

two SRL support variants for the online SRL activity variables, $F(4, 134) = 2.26$, $p = .066$; Wilk's $\lambda = 0.937$, partial $\eta^2 = .06$. Overall, we conclude that the difference in SRL support in the offering of the course material did not significantly affect SRL behavior.

6

Chapter 6

General Discussion

In flipped learning (FL), students prepare learning materials before class and apply what they have learned during class. FL is supposed to benefit the delivery of information and instruction as students are able to control the viewing frequency and viewing pace of instructional videos before class to reduce cognitive load (Abeysekera & Dawson, 2015). In addition, students in a flipped classroom devote class time to learning activities focused on applying the instructional material with the guidance of their teacher (Roehl et al., 2013). Students' learning processes benefit from a teacher that guides and facilitates active learning (i.e., a class in which students engage in learning activities instead of passively listening) and provides direct feedback (Van den Bergh et al., 2014). However, in previous empirical research, we noticed that there was ambiguity in discussions about the effects of FL (Bishop & Verleger, 2013; O'Flaherty & Phillips, 2015). Therefore, the first part of this dissertation aimed to synthesize the current scientific literature about the effects of FL to be able to draw more general conclusions.

Summary of the Results

The main research questions in the study in **Chapter 2** were: 1) What is the effect of FL on assessed and perceived learning outcomes and student satisfaction, compared with traditional teaching methods?; 2) Does the effect of FL on assessed and perceived learning outcomes and student satisfaction depend on instructional design characteristics, educational context characteristics, or study quality characteristics? We quantitatively synthesized the effect sizes of 114 studies on the effects of FL on learning outcomes and student satisfaction in secondary and higher education. The studies included in the meta-analyses compared students in an FL setting to students in a more traditional setting. We found that, on average, applying FL leads to higher assessed learning outcomes. The results for perceived learning outcomes and student satisfaction were not significant. For all meta-analyses, we found significant heterogeneity in effect sizes between studies, which was caused by more than random sampling alone. Therefore, we investigated if instructional design, sample, and study quality characteristics moderated the effectiveness of FL. We found that sustaining face-to-face time and adding quizzes significantly moderated the effect of FL on assessed learning outcomes.

The second part of this dissertation focused on the role of supporting SRL in FL. During FL, students have more autonomy over their learning process. This puts a higher demand on students' SRL skills (He et al., 2016; Lee & Tsai, 2011; Shih & Huang, 2019; Tan et al., 2017). As students may struggle to apply SRL to fully benefit from FL, supporting students' SRL in an FL environment seems worthwhile. Previous research has shown that supporting SRL by means of SRL interventions enhances students' SRL and in turn (at least partly) enhances learning outcomes (Dent & Koenka, 2016;

Dignath & Büttner, 2008; Donker et al., 2014; Jansen et al., 2019; Zheng, 2016). However, because most studies on supporting students' SRL were conducted in higher education, we noticed an absence of studies about how students' SRL should be supported in FL environments in secondary education. Therefore, in the second part of this dissertation, we investigated the effects of SRL support embedded in flipped video instructions on SRL and learning outcomes.

In the study in **Chapter 3**, we investigated the effects of video embedded SRL prompts on secondary education students' (a) SRL self-reports, (b) online SRL activity, and (c) learning outcomes. We conducted our experiment in an ecologically valid setting over the course of six weeks. We found that students provided with SRL prompts did not significantly improve in terms of their SRL self-reports and learning outcomes in comparison with students who were not provided with the SRL prompts. The results for students' online SRL activity only showed a significant difference in *video completion rate*. This means that students in the SRL-prompts condition watched more videos than students in the no prompts condition. This effect could indicate that the SRL prompts made students more aware of their learning process and that they therefore completed their homework more consistently than students without the SRL prompts. This could, for instance, be caused by SRL prompts that explicitly connected the coherence across the homework videos. However, we also expected to find differences in the other SRL measures on the basis of SRL interventions in other educational contexts (e.g., Bannert et al., 2015; Lai & Hwang, 2016; Moos & Bonde, 2016; Shyr & Chen, 2018), but this expectation was not confirmed.

Therefore, in the study in **Chapter 4**, we added explicit SRL instruction to the design of embedded SRL support to explain to students why SRL activities are beneficial, and how they can be performed. In a similar ecologically valid setting of eight weeks, we again investigated what the effects of video embedded SRL support are on students' (a) SRL self-reports, (b) online SRL activity, and (c) learning outcomes. In addition, we investigated if SRL support affected (d) student satisfaction, and (e) how students in the SRL-support condition valued this SRL support. We found that students in the SRL-support condition achieved higher learning outcomes, but, in contrast to our hypothesis, we found no difference in students' SRL activities as indicated by their SRL self-reports and their SRL online activity variables. Although the SRL support helped students to be more aware of their own learning, some students clearly disliked the SRL-instruction.

As we also observed large variance in students' online SRL activities, we hypothesized that the one-size-fits-all SRL support studies from **Chapters 3** and **4**, although improving learning outcomes, might explain why we did not find an effect on students' SRL activities. For instance, if variation in online SRL activities of students in

this context appears to be large, the results can be used to guide future research on personalized SRL support based on individual differences in students' SRL skills (cf. Kizilcec et al., 2017; Vanslambrouck et al., 2019). Therefore, in the study in **Chapter 5**, we investigated the large variance in students' online SRL activities. We asked (a) which SRL profiles can be identified according to students' online SRL activities, (b) if these SRL profiles differ in learning outcomes, and (c) if self-reported motivation, self-reported SRL, and prior knowledge predict SRL profile membership. In line with previous research in higher education, we found five significantly distinct SRL profiles from low completion and no activity to full completion and very high activity. In addition, students in the profile which showed low SRL activity achieved significantly worse learning outcomes than students in the profiles with higher SRL activity. However, we did not find any student-level predictor variable (e.g., prior knowledge) that explained SRL profile membership.

Scientific Implications

Here, we address how our results contribute to the current scientific debates and how these could guide future research. First, FL seems to be evaluated based on the results of single studies (e.g., Setren et al., 2019) while a synthesis of research findings is necessary to come to robust conclusions. With our study in **Chapter 2**, we contributed a statistical synthesis of data on the effects of FL. With this synthesis, we are now able to draw the robust conclusion that FL in general leads to an improvement in learning outcomes. The generalizability of this conclusion has been reinforced, as comparable results from other meta-analyses on FL have been found in the past five years (Chen et al., 2018; Cheng et al., 2019; Hew et al., 2020; Hew & Lo, 2018; Hu et al., 2018; Jang & Kim, 2020; Låg & Sæle, 2019; Strelan et al., 2020; Tan et al., 2017; Wagner et al., 2020; Zheng et al., 2020). Therefore, we urge the research field to move beyond the question if FL can be an effective way to enhance student learning, but rather take the insights (e.g., quizzing, not reducing class time) to improve interventions and ask new questions about what is still unknown. For example, about the role of SRL, motivation, or teacher level variables such as workload and satisfaction with work (e.g., Wanner & Palmer, 2015).

Second, we showed that, in contrast to widespread views, students do not necessarily like FL in comparison to traditional methods. This might be explained by students' perception of active learning during class. For example, one of the explanations participants of FL give for their dissatisfaction, is that is they perceive a higher workload because of the active learning activities in class (Låg & Sæle, 2019; Strelan et al., 2020). Previous research showed that when students experience an increased cognitive effort that is often associated with active learning, they initially take that effort to signify poorer learning (Deslauriers et al., 2019). In other words, FL requires more

effort to engage in active learning activities in class, but students might associate this increased effort with lower learning outcomes. The dissatisfaction with the learning environment based on this misjudgment could have negative effects on students' motivation, engagement, and ability to self-regulate their own learning (Deslauriers et al., 2019). This connects to what we found in our SRL support experiments. Some students reported that they did not like the SRL support because it required more effort and distracted them from learning from the videos, while they actually scored significantly better on the learning outcome test than students who did not receive SRL support. Because of this finding, we advocate to critically consider the validity of student satisfaction as outcome measure. On the other hand, we recommend both FL practitioners as future researchers to address this misconception among students in FL settings and explain the value of active learning and SRL support for their own learning.

Third, the results of our studies in **Chapters 3 and 4** are the first steps to examine how SRL can best be supported in an FL setting in secondary education. We found that positive results from previous SRL interventions involving video embedded SRL prompts in other educational contexts (e.g., higher education) on students' SRL activities do not automatically yield similar results in secondary education (Moos & Bonde, 2016). We found that for students in their second year of secondary education, SRL prompts need to be supplemented with SRL instruction to enhance students' learning outcomes. In this way, students' SRL activities can be supported during FL both qualitatively by SRL instruction, and quantitatively by SRL prompts (Panadero et al., 2016). This is in contrast with insights from previous research suggesting that students in primary education benefit more from SRL interventions that focus on instruction, as they lack knowledge and experience with SRL (Donker et al., 2014; Lai & Hwang, 2016), compared with students in secondary education who already acquired a certain level of SRL which could make explicit SRL instruction ineffective (cf. the expertise reversal effect; Kalyuga et al., 2003).

Fourth, the large variance in students' online SRL activities we found could be interesting to investigate in further research. In **Chapter 5**, we explored this variance to see if this could be an explanation for the absence of effects of SRL support on students' online SRL activities in **Chapters 3 and 4**. The differences in online SRL behavior, which were related to differences in learning outcomes, suggest that personalized SRL support could be an important improvement as compared to one-size-fits-all SRL support. This can be further explored, for example, in learning environments where students' online SRL activity and/or SRL skills are measured, and SRL support is adapted to these results.

Practical Recommendations

First, we conclude that the general effect of FL on assessed learning outcomes is meaningful in the context of education. It means that, as compared to the mean learning outcomes of students in a more traditional classroom, 64% of the students in FL achieve higher learning outcomes. However, the significant and large heterogeneity in the effect sizes of the included studies in our meta-analyses makes clear that FL is not a magic bullet; it matters how FL is designed and implemented (cf. Liu et al., 2016). Based on our results, we recommend to sustain face-to-face time, and add quizzes as both elements are critical features for a successful implementation of FL. The results of our meta-analyses illustrate that the working principle of FL might be that it enables teachers to combine proven teaching methods in the classroom that enhance learning outcomes. For instance, teachers can test students' knowledge of the pre-class instructional videos with learning activities that are beneficial due to the testing effect (e.g., DeLozier & Rhodes, 2017; McDaniel et al., 2007). The FL method is particularly suitable for applying repeated retrieval practice activities. These activities require student's effort, and allow for teacher's feedback. This improves the long-term retention of the learning content (Kang et al., 2007).

Second, one of the novelties of our research approach is that we tested SRL interventions in an ecologically valid setting over the course of eight weeks. An appropriate number of students participated in the studies during their regular History lessons. In this way, we tried to yield practically relevant results for educational practitioners in secondary education. We conclude that it is necessary to teach secondary education students how they could benefit from FL, as the increased autonomy during FL can be challenging. We showed that an effective way to do this is to add SRL prompts and SRL instruction to the online instructional videos. We recommend to present SRL support to students in three phases: the forethought phase at the beginning of the learning task (e.g., task analysis and planning), the performance phase during the learning task (e.g., self-control and monitoring), and the self-reflection phase at the end of the learning task (e.g., self-judgment and change behavior accordingly; Zimmerman & Moylan, 2009). In addition, we recommend to gradually reduce the SRL support (i.e., fading or scaffolding) to let students internalize the instructed SRL strategies (Kollar & Fischer, 2006).

Third, as we found that students' online SRL activities varied greatly, we advise teachers to recognize these differences between students, and adapt SRL support accordingly. For example, we found a student SRL profile that generally failed to pass the course. This group of students only watched about 70% of the instructional videos, and about half of the times after the deadline (i.e., not before class). It seems that this group could benefit from specific SRL support in the forethought phase, such as goal setting, strategic planning, and self-motivational beliefs. In contrast, students

who rewind videos often without a payoff in higher learning outcomes, could benefit if their self-efficacy, efficiency, confidence, or merely the quality of their SRL activities is supported. This can be done by instruction in the classroom (e.g., Ness & Middleton, 2012), or by personalized online prompts focused on the forethought phase (e.g., Lehmann et al., 2014).

Limitations and Future Directions

In the remaining part of the general discussion, we will discuss the overarching limitations of this dissertation. In contrast to our expectations, we did not find large effects of our SRL support interventions on students' self-reported SRL and their online SRL activities. A meta-analysis showed that SRL activity only partially mediates the effects of SRL interventions on learning outcomes (Jansen et al., 2019). This suggests that there are other factors leading to enhanced learning outcomes resulting from SRL support, besides improving students' SRL self-reports and activities. On the one hand, we suggest that this could be due to an increase in time-on-task for students whose SRL was supported. On the other hand, the extent to which we measured SRL (i.e., log data and self-reports) mainly provides quantitative SRL indications. While we did not find that the effect of SRL support on learning outcomes was due to an increase in the quantity of students' SRL activities, it is possible that our intervention mostly improved the quality of students' SRL activities. One way to better understand the quality and underlying reasons of students' SRL in future research is to supplement SRL trace data with retrospective interview data in which students reflect on their online SRL activities (e.g., Jansen et al., 2020). While the trace data leaves some interpretations open about the quality of students' SRL, interviews will provide this additional information.

Besides, it is a limitation of our SRL intervention studies that we did not measure the impact of our SRL support on the cognitive load of students. We know from research that complex learning activities that require simultaneous processing of complex information and tasks can lead to an overload of students' working memory (Paas et al., 2003). In the qualitative interviews and the satisfaction questionnaire from **Chapters 3 and 4**, we saw that some students indeed reported that the SRL support during the instructional videos distracted them from the learning tasks. We can not rule out the possibility that adding SRL support increased cognitive load of students and in turn might have an impact on our results (cf. Seufert, 2020).

In addition, one alternative explanation for the lack of effects on students' SRL could be the way we designed our FL environment. In this dissertation, we argued that FL leads to an increase in autonomy over students' learning process, and that this puts a higher demand on their SRL skills (He et al., 2016; Shih & Huang, 2019; Tan et al.,

2017). However, the extent to which students are given autonomy differs according to the implementation of FL (Van Laer & Elen, 2017). Because the participating students' mean age was 14 years, we chose to implement a structure in our SRL interventions in which students had relatively little control and freedom over the learning tasks that they had to complete before class as part of their homework. They still had autonomy over where and when they learned, but for every lesson there was a strict planning which instructional video they had to prepare, including a workbook with quiz questions about the content in both the intervention and control group. This intervention setup could have had the effect that students had little opportunities to apply a wide range of SRL skills (such as planning and task selection). In addition, we provided a small extra credits to students for watching the instructional videos, to reduce students' non-compliance that would lead to a reduction of our sample size. This may have partly taken over students' SRL, which in turn could explain that the SRL support we provided could not fully enhance their SRL skills. Future research could further investigate the ideal ratio between student and teacher control to enhance SRL skills in FL settings in secondary education (cf. Corbalan et al., 2009; and for higher education: Vandewaetere & Clarebout, 2011).

General Conclusion

With the studies in this dissertation, we contributed to the scientific knowledge about the effects of FL on learning outcomes, satisfaction, and the role of SRL support to enhance the effectiveness of FL. We conclude that applying FL is a beneficial method to increase students' learning outcomes. We recommend to sustain face-to-face time and add quizzes to the FL environment, as this significantly moderated the effect of FL on assessed learning outcomes. In contrast to what often is supposed, we did not find any evidence that FL leads to enhanced student satisfaction. Furthermore, we strongly advise to consider SRL support when designing FL environments, as FL puts a higher demand on students' SRL skills. We found that in secondary education, video embedded SRL support that includes both SRL prompts to improve the quantity of students' SRL and SRL instruction to improve the quality of students' SRL enhance learning outcomes. Yet, we also found large variance in students' online SRL activities profiles. This provides opportunities for future research to further improve SRL support in secondary education, for example by personalizing SRL support to students' SRL skills.

References

References marked with an asterisk indicate studies included in the meta-analyses in Chapter 2. A complete reference list of studies included in the meta-analyses is available as appendix B (online only) at <https://ars.els-cdn.com/content/image/1-s2.0-S1747938X18305694-mmc5.docx>.

- Abar, B., & Loken, E. (2010). Self-regulated learning and self-directed study in a pre-college sample. *Learning and Individual Differences*, 20(1), 25–29. <https://doi.org/10.1016/j.lindif.2009.09.002>
- Abeysekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research. *Higher Education Research & Development*, 34(1), 1–14. <https://doi.org/10.1080/07294360.2014.934336>
- *Adams, A. E. M., Garcia, J., & Traustadóttir, T. (2016). A Quasi Experiment to Determine the Effectiveness of a “Partially Flipped” versus “Fully Flipped” Undergraduate Class in Genetics and Evolution. *CBE-Life Sciences Education*, 15(2), 1–9. <https://doi.org/10.1187/cbe.15-07-0157>
- Adèr, H. J., Mellenbergh, G. J., & Hand, D. J. (2008). *Advising on research methods: A consultant's companion*. Johannes van Kessel Publishing.
- Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers and Education*, 126, 334–345. <https://doi.org/10.1016/j.compedu.2018.07.021>
- Ali, L., Hatala, M., Gašević, D., & Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers and Education*, 58(1), 470–489. <https://doi.org/10.1016/j.compedu.2011.08.030>
- van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: a meta-analysis. *Educational Research Review*, 28, 1–18. <https://doi.org/10.1016/j.edurev.2019.05.003>
- van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2020a). Effects of self-regulated learning prompts in a flipped history classroom. *Computers in Human Behavior*, 108, 106318. <https://doi.org/10.1016/j.chb.2020.106318>
- van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2020b). Self-regulated learning support in flipped learning videos enhances learning outcomes. *Computers and Education*, 158, 104000. <https://doi.org/10.1016/j.compedu.2020.104000>
- Armijo-Olivo, S., Warren, S., & Magee, D. (2009). Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: a review. *Physical Therapy Reviews*, 14(1), 36–49. <https://doi.org/10.1179/174328809X405928>
- *Asef-Vaziri, A. (2015). The Flipped Classroom of Operations Management : A Not-For-Cost-Reduction Platform. *Decision Sciences Journal of Innovative Education*, 13(1), 71–89. <https://doi.org/10.1111/dsji.12054>
- *Asiksoy, G., & Özdamli, F. (2016). Flipped classroom adapted to the ARCS model of motivation and applied to a physics course. *Eurasia Journal of Mathematics, Science and Technology Education*, 12, 1589–1603. <https://doi.org/10.12973/eurasia.2016.1251a>
- Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology*, 96(3), 523–535. <https://doi.org/10.1037/0022-0663.96.3.523>

- Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition - Implications for the design of computer-based scaffolds. *Instructional Science*, 33(5-6), 367-379. <https://doi.org/10.1007/s11251-005-1272-9>
- *Baepler, P., Walker, J. D., & Driessen, M. (2014). It's not about seat time: Blending, flipping, and efficiency in active learning classrooms. *Computers and Education*, 78, 227-236. <https://doi.org/10.1016/j.compedu.2014.06.006>
- Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C.-L. C. (1991). Effects of Frequent Classroom Testing. *The Journal of Educational Research*, 85(2), 89-99. <https://doi.org/10.1080/00220671.1991.10702818>
- Banks, G. C., Kepes, S., & Banks, K. P. (2012). Publication Bias: The Antagonist of Meta-Analytic Reviews and Effective Policymaking. *Educational Evaluation and Policy Analysis*, 34, 259-277. <https://doi.org/10.3102/0162373712446144>
- Bannert, M., Hildebrand, M., & Mengelkamp, C. (2009). Effects of a metacognitive support device in learning environments. *Computers in Human Behavior*, 25(4), 829-835. <https://doi.org/10.1016/j.chb.2008.07.002>
- Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, 9(2), 161-185. <https://doi.org/10.1007/s11409-013-9107-6>
- Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior*, 52, 293-306. <https://doi.org/10.1016/j.chb.2015.05.038>
- Barnard-Brak, L., Lan, W. Y., & Paton, V. O. (2010). Profiles in Self-Regulated Learning in the Online Learning Environment. *International Review of Research in Open and Distance Learning*, 11(1), 61-79. <https://doi.org/10.19173/irrodl.v11i1.769>
- Barnard-Brak, L., Lan, W. Y., & Paton, V. O. (2011). Measuring and profiling self-regulated learning in the online environment. In G. Dettori & D. Persico (Eds.), *Fostering Self-Regulated Learning through ICT* (pp. 27-38). IGI global. <https://doi.org/10.4018/978-1-61692-901-5.ch002>
- Belski, R., & Belski, I. (2014). Cultivating student skills in self-regulated learning through evaluation of task complexity. *Teaching in Higher Education*, 19(5), 459-469. <https://doi.org/10.1080/13562517.2014.880685>
- Bétrancourt, M., & Benetos, K. (2018). Why and when does instructional video facilitate learning? A commentary to the special issue "developments and trends in learning with instructional video." *Computers in Human Behavior*, 89, 471-475. <https://doi.org/10.1016/j.chb.2018.08.035>
- van den Bergh, L., Ros, A., & Beijard, D. (2014). Improving Teacher Feedback During Active Learning: Effects of a Professional Development Program. *American Educational Research Journal*, 51(4), 772-809. <https://doi.org/10.3102/0002831214531322>
- Bergmann, J., & Sams, A. (2012). *Flip Your Classroom: Reach Every Student in Every Class Every Day*. Alexandria: International Society for Technology in Education.
- Bergmann, J., & Sams, A. (2014). Flipped Learning: Maximizing Face Time. *T&D*, 68(2), 28-31.
- Bernard, J. S. (2015). The flipped classroom: Fertile ground for nursing education research. *International Journal of Nursing Education Scholarship*, 12(1), 1-11. <https://doi.org/10.1515/ijnes-2015-0005>
- Bethiavas, V., Bridgman, H., Kornhaber, R., & Cross, M. (2016). The evidence for "flipping out": A systematic review of the flipped classroom in nursing education. *Nurse Education Today*, 38,

- 15–21. <https://doi.org/10.1016/j.nedt.2015.12.010>
- Bishop, J. L., & Verleger, M. (2013). The Flipped Classroom: A Survey of the Research. *Proceedings of the Annual Conference of the American Society for Engineering Education*, 30(9), 1–18.
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology*, 64(1), 417–444. <https://doi.org/10.1146/annurev-psych-113011-143823>
- Blasiman, R. N., Dunlosky, J., & Rawson, K. A. (2017). The what, how much, and when of study strategies: comparing intended versus actual study behaviour. *Memory*, 25(6), 784–792. <https://doi.org/10.1080/09658211.2016.1221974>
- Boevé, A. J., Meijer, R. R., Bosker, R. J., Vugteveen, J., Hoekstra, R., & Albers, C. J. (2017). Implementing the flipped classroom: an exploration of study behaviour and student performance. *Higher Education*, 74(6), 1015–1032. <https://doi.org/10.1007/s10734-016-0104-y>
- Bol, L., Campbell, K. D. Y., Perez, T., & Yen, C. J. (2016). The effects of self-regulated learning training on community college students' metacognition and achievement in developmental math courses. *Community College Journal of Research and Practice*, 40(6), 480–495. <https://doi.org/10.1080/10668926.2015.1068718>
- Bond, M. (2020). Facilitating student engagement through the flipped learning approach in K-12: A systematic review. *Computers and Education*, 151, 103819. <https://doi.org/10.1016/j.compedu.2020.103819>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd. <http://doi.org/10.1002/9780470743386>
- *Bossaer, J. B., Panus, P., Stewart, D. W., & Hagemeyer, N. E. (2016). Student Performance in a Pharmacotherapy Oncology Module Before and After Flipping the Classroom. *American Journal of Pharmaceutical Education*, 80(2), 3–8. <https://doi.org/10.5688/ajpe80231>
- Bouwmeester, R. A. M., de Kleijn, R. A. M., van den Berg, I. E. T., ten Cate, O. T. J., van Rijen, H. V. M., & Westerveld, H. E. (2019). Flipping the medical classroom: Effect on workload, interactivity, motivation and retention of knowledge. *Computers & Education*, 139, 118–128. <https://doi.org/10.1016/j.compedu.2019.05.002>
- Broadbent, J., & Fuller-Tyszkiewicz, M. (2018). Profiles in self-regulated learning and their correlates for online and blended learning students. *Educational Technology Research and Development*, 66(6), 1435–1455. <https://doi.org/10.1007/s11423-018-9595-9>
- Bryan, T., Burstein, K., & Bryan, J. (2001). Students With Learning Disabilities: Homework Problems and Promising Practices. *Educational Psychologist*, 36(3), 167–180. https://doi.org/10.1207/S15326985EP3603_3
- Butzler, K. B. (2016). The Synergistic Effects of Self-Regulation Tools and the Flipped Classroom. *Computers in the Schools*, 33(1), 11–23. <https://doi.org/10.1080/07380569.2016.1137179>
- *Calimeris, L., & Sauer, K. M. (2015). Flipping out about the flip: All hype or is there hope? *International Review of Economics Education*, 20, 13–28. <https://doi.org/10.1016/j.iree.2015.08.001>
- Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13(2), 195–212. <https://doi.org/10.1007/BF01246098>
- Chandra, V., & Lloyd, M. (2008). The methodological nettle: ICT and student achievement. *British Journal of Educational Technology*, 39(6), 1087–1098. <https://doi.org/10.1111/j.1467-8535.2007.00790.x>

- Chen, K. S., Monrouxe, L., Lu, Y. H., Jenq, C. C., Chang, Y. J., Chang, Y. C., & Chai, P. Y. C. (2018). Academic outcomes of flipped classroom learning: a meta-analysis. *Medical Education*, 52(9), 910–924. <https://doi.org/10.1111/medu.13616>
- Cheng, L., Ritzhaupt, A. D., & Antonenko, P. (2019). Effects of the flipped classroom instructional strategy on students' learning outcomes: a meta-analysis. *Educational Technology Research and Development*, 67(4), 793–824. <https://doi.org/10.1007/s11423-018-9633-7>
- Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292. <https://doi.org/10.3102/0013189X16656615>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Cicchinelli, A., Veas, E., Pardo, A., Pammer-Schindler, V., Fessler, A., Barreiros, C., & Lindstädt, S. (2018). Finding traces of self-regulated learning in activity streams. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*, 191–200. <https://doi.org/10.1145/3170358.3170381>
- Clarebout, G., Horz, H., Schnotz, W., & Elen, J. (2010). The relation between self-regulation and the embedding of support in learning environments. *Educational Technology Research and Development*, 58(5), 573–587. <https://doi.org/10.1007/s11423-009-9147-4>
- Clark, R. C., Nguyen, F., & Sweller, J. (2005). *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. Pfeiffer.
- Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development*, 42(2), 21–29. <https://doi.org/10.1007/BF02299088>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.
- Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2009). Combining shared control with variability over surface features: Effects on transfer test performance and task involvement. *Computers in Human Behavior*, 25(2), 290–298. <https://doi.org/10.1016/j.chb.2008.12.009>
- DeLozier, S. J., & Rhodes, M. G. (2017). Flipped classrooms: A review of key ideas and recommendations for practice. *Educational Psychology Review*, 29(1), 141–151. <http://dx.doi.org/10.1007/s10648-015-9356-9>
- Dent, A. L., & Koenka, A. C. (2016). The Relation Between Self-Regulated Learning and Academic Achievement Across Childhood and Adolescence: A Meta-Analysis. *Educational Psychology Review*, 28(3), 425–474. <https://doi.org/10.1007/s10648-015-9320-8>
- Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences of the United States of America*, 116(39), 19251–19257. <https://doi.org/10.1073/pnas.1821936116>
- 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, C., & Büttner, G. (2018). Teachers' direct and indirect promotion of self-regulated

- learning in primary and secondary school mathematics classes – insights from video-based classroom observations and teacher interviews. *Metacognition and Learning*, 13(2), 127–157. <https://doi.org/10.1007/s11409-018-9181-x>
- Dirkx, K. J. H., Camp, G., Kester, L., & Kirschner, P. A. (2019). Do secondary school students make use of effective study strategies when they study on their own? *Applied Cognitive Psychology*, 33(5), 952–957. <https://doi.org/10.1002/acp.3584>
- Dirkx, K. J. H., Kester, L., & Kirschner, P. A. (2014). The testing effect for learning principles and procedures from texts. *The Journal of Educational Research*, 107(5), 357–364. <https://doi.org/10.1080/00220671.2013.823370>
- 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>
- Duncan, T. G., & McKeachie, W. J. (2005). The Making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist*, 40(2), 117–128.
- *Eichler, J. F., & Peebles, J. (2016). Flipped classroom modules for large enrollment general chemistry courses: a low barrier approach to increase active learning and improve student grades. *Chemistry Education Research and Practice*, 17, 197–208. <https://doi.org/10.1039/C5RP00159E>
- Eriksson, T., Adawi, T., & Stöhr, C. (2017). “Time is the bottleneck”: a qualitative study exploring why learners drop out of MOOCs. *Journal of Computing in Higher Education*, 29(1), 133–146. <https://doi.org/10.1007/s12528-016-9127-8>
- Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of clinical epidemiology*, 43(6), 543–549. [https://doi.org/10.1016/0895-4356\(90\)90158-L](https://doi.org/10.1016/0895-4356(90)90158-L)
- Ferguson, S. L., Moore, E. W. G., & Hull, D. M. (2019). Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. *International Journal of Behavioral Development*. <https://doi.org/10.1177/0165025419881721>
- Fergusson, D., Aaron, S. D., Guyatt, G., & Hébert, P. (2002). Post-randomisation exclusions: The intention to treat principle and excluding patients from analysis. *British Medical Journal*, 325(7365), 652–654.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (Fifth ed.). Sage.
- Field, A. P., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy*, 98, 19–38. <https://doi.org/10.1016/j.brat.2017.05.013>
- Flipped Learning Network. (2014). *The Four Pillars of F-L-I-P™*. In: [flippedlearning.org](http://www.flippedlearning.org/definition). <http://www.flippedlearning.org/definition>
- Graham, C. R. (2006). Blended learning systems: Definition, Current Trends, and Future Directions. In C. J. Bonk & C. R. Graham (Eds.), *The handbook of blended learning: Global perspectives, local designs* (pp. 3–21). Pfeiffer.
- Greene, J. A., & Azevedo, R. (2010). The measurement of learners’ self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist*, 45(4), 203–209. <https://doi.org/10.1080/00461520.2010.515935>
- Greene, J. A., Bolick, C. M., Caprino, A. M., Deekens, V. M., McVea, M., Yu, S., & Jackson, W. P. (2015). Fostering High-School Students’ Self-Regulated Learning Online and Across Academic Domains. *The High School Journal*, 99(1), 88–106. <https://doi.org/10.1353/hsj.2015.0019>

- 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 and Education*, 54(1), 230–243. <https://doi.org/10.1016/j.compedu.2009.08.006>
- *Gross, D., Pietri, E. S., Anderson, G., Moyano-Camihort, K., & Graham, M. J. (2015). Increased preclass preparation underlies student outcome improvement in the flipped classroom. *CBE Life Sciences Education*, 14(4), 1–8. <https://doi.org/10.1187/cbe.15-02-0040>
- *Hagen, E. J., & Fratta, D. (2014). Hybrid learning in geological engineering: Why, how, and to what end? Preliminary results. *Geo-Congress Technical Papers: Geo-Characterization and Modeling for Sustainability, USA*, 3920–3929. <https://doi.org/10.1061/9780784413272.380>
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.
- He, W., Holton, A., Farkas, G., & Warschauer, M. (2016). The effects of flipped instruction on out-of-class study time, exam performance, and student perceptions. *Learning and Instruction*, 45, 61–71. <https://doi.org/10.1016/j.learninstruc.2016.07.001>
- Hedges, L. V. (1981). Distribution Theory for Glass's Estimator of Effect size and Related Estimators. *Journal of Educational Statistics*, 6(2), 107–128. <https://doi.org/10.3102/10769986006002107>
- Hedges, L. V., & Pigott, T. D. (2004). The Power of Statistical Tests for Moderators in Meta-Analysis. *Psychological Methods*, 9(4), 426–445. <http://dx.doi.org/10.1037/1082-989X.9.4.426>
- Hempel, S., Miles, J. N., Booth, M. J., Wang, Z., Morton, S. C., & Shekelle, P. G. (2013). Risk of bias: a simulation study of power to detect study-level moderator effects in meta-analysis. *Systematic reviews*, 2(1), 107 <https://doi.org/10.1186/2046-4053-2-107>
- Henderson, M., Selwyn, N., & Aston, R. (2017). What works and why? Student perceptions of 'useful' digital technology in university teaching and learning. *Studies in Higher Education*, 42(8), 1567–1579. <https://doi.org/10.1080/03075079.2015.1007946>
- Hew, K. F., & Lo, C. K. (2018). Flipped classroom improves student learning in health professions education: A meta-analysis. *BMC Medical Education*, 18(1), 1–12. <https://doi.org/10.1186/s12909-018-1144-z>
- Heyma, A., Bisschop, P., van den Berg, E., Wartenbergh-Cras, F., Kurver, B., Muskens, M., & Spanjers, I. (2015). *Effectmeting Innovatielimpuls Onderwijs: eindrapport*. (SEO-rapport No. 2015-28). <https://www.rijksoverheid.nl/documenten/rapporten/2015/06/29/effectmeting-innovatieimpuls-onderwijs>
- *Hibbard, L., Sung, S., & Wells, B. (2016). Examining the Effectiveness of a Semi Self Paced Format in a College General Chemistry Sequence. *Journal of Chemical Education*, 93(1), 24–30. <https://doi.org/10.1021/acs.jchemed.5b00592>
- Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3), 172–177. <https://doi.org/10.1111/j.1750-8606.2008.00061.x>
- *Hotle, S. L., & Garrow, L. A. (2015). Effects of the Traditional and Flipped Classrooms on Undergraduate Student Opinions and Success. *Journal of Professional Issues in Engineering Education and Practice*, 142(1), 5015005. [https://doi.org/10.1061/\(ASCE\)EI.1943-5541.0000259](https://doi.org/10.1061/(ASCE)EI.1943-5541.0000259)
- Hsu, H. C. K., Wang, C. V., & Levesque-Bristol, C. (2019). Reexamining the impact of self-determination theory on learning outcomes in the online learning environment. *Education and Information Technologies*, 24(3), 2159–2174. <https://doi.org/10.1007/s10639-019-09863-w>
- Hu, R., Gao, H., Ye, Y., Ni, Z., Jiang, N., & Jiang, X. (2018). Effectiveness of flipped classrooms in Chinese baccalaureate nursing education: A meta-analysis of randomized controlled trials. *International*

- Journal of Nursing Studies*, 79, 94–103. <https://doi.org/10.1016/j.jijnurstu.2017.11.012>
- *Hudson, D. L., Whisenhunt, B. L., Shoptaugh, C. F., Visio, M. E., Cathey, C., & Rost, A. D. (2015). Change takes time: Understanding and responding to culture change in course redesign. *Scholarship of Teaching and Learning in Psychology*, 1(4), 255–268. <https://doi.org/10.1037/stl0000043>
- Jackson, L. A., Zhao, Y., Kolenic, A., Fitzgerald, H. E., Harold, R., & Von Eye, A. (2008). Race, gender, and information technology use: The new digital divide. *CyberPsychology & Behavior*, 11(4), 437–442. <https://doi.org/10.1089/cpb.2007.0157>
- Jang, H. Y., & Kim, H. J. (2020). A meta-analysis of the cognitive, affective, and interpersonal outcomes of flipped classrooms in higher education. *Education Sciences*, 10(4). <https://doi.org/10.3390/educsci10040115>
- Jansen, R. S., van Leeuwen, A., Janssen, J., Conijn, R., & Kester, L. (2020). Supporting learners' self-regulated learning in Massive Open Online Courses. *Computers and Education*, 146. <https://doi.org/10.1016/j.compedu.2019.103771>
- Jansen, R. S., van Leeuwen, A., Janssen, J., Jak, S., & Kester, L. (2019). Self-regulated learning partially mediates the effect of self-regulated learning interventions on achievement in higher education : A meta-analysis. *Educational Research Review*, 28. <https://doi.org/10.1016/j.edurev.2019.100292>
- Jansen, R. S., van Leeuwen, A., Janssen, J., & Kester, L. (2020). A mixed method approach to studying self-regulated learning in MOOCs. *Frontline Learning Research*, 8(2), 35–64. <https://doi.org/10.14786/flr.v8i2.539>
- Jansen, R. S., van Leeuwen, A., Janssen, J., & Kester, L. (2018). Validation of the Revised Self-regulated Online Learning Questionnaire. In V. Pammer-Schindler, M. Pérez-Sanagustín, H. Drachler, R. Elferink, & M. Scheffel (Eds.), *Lifelong Technology-Enhanced Learning* (pp. 116–121). Springer International Publishing.
- Jansen, R. S., van Leeuwen, A., Janssen, J., Kester, L., & Kalz, M. (2017). Validation of the self-regulated online learning questionnaire. *Journal of Computing in Higher Education*, 29(1), 6–27. <https://doi.org/10.1007/s12528-016-9125-x>
- Jo, B. (2002). Statistical power in randomized intervention studies with noncompliance. *Psychological Methods*, 7(2), 178–193. <https://doi.org/10.1037/1082-989X.7.2.178>
- Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *Internet and Higher Education*, 33, 74–85. <https://doi.org/10.1016/j.iheduc.2017.02.001>
- Jovanović, J., Mirriahi, N., Gašević, D., Dawson, S., & Pardo, A. (2019). Predictive power of regularity of pre-class activities in a flipped classroom. *Computers and Education*, 134, 156–168. <https://doi.org/10.1016/j.compedu.2019.02.011>
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist*, 38(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19(4–5), 528–558. <https://doi.org/10.1080/09541440601056620>
- Karabenick, S. A., & Gonida, E. N. (2018). Academic help seeking as a self-regulated learning strategy: Current issues, future directions. In *Handbook of self-regulation of learning and performance*, 2nd ed. (pp. 421–433). Routledge/Taylor & Francis Group.

- Kauffman, D. F. (2004). Self-Regulated Learning in Web-Based Environments: Instructional Tools Designed to Facilitate Cognitive Strategy Use, Metacognitive Processing, and Motivational Beliefs. *Journal of Educational Computing Research*, 30(1–2), 139–161. <https://doi.org/10.2190/AX2D-Y9VM-V7PX-OTAD>
- Kim, M. K., Kim, S. M., Khera, O., & Getman, J. (2014). The experience of three flipped classrooms in an urban university: An exploration of design principles. *Internet and Higher Education*, 22, 37–50. <https://doi.org/10.1016/j.iheduc.2014.04.003>
- King, A. (1993). From Sage on the Stage to Guide on the Side. *College Teaching*, 41(1), 30–35.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work. *Educational Psychologist*, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1
- Kistner, S., Rakoczy, K., Otto, B., Dignath-van Ewijk, C., Büttner, G., & Klieme, E. (2010). Promotion of self-regulated learning in classrooms: Investigating frequency, quality, and consequences for student performance. *Metacognition and Learning*, 5(2), 157–171. <https://doi.org/10.1007/s11409-010-9055-3>
- Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers and Education*, 104, 18–33. <https://doi.org/10.1016/j.compedu.2016.10.001>
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. (T. D. Kenny & D. A. Little, Eds.). Guilford.
- Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22, 2693–2710. <https://doi.org/10.1002/sim.1482>
- Kollar, I., & Fischer, F. (2006). Supporting self-regulated learners for a while and what computers can contribute. *Journal of Educational Computing Research*, 35(4), 425–435. <https://doi.org/10.2190/RK02-7384-2723-G744>
- *Koo, C. L., Demps, E. L., Farris, C., Bowman, J. D., Panahi, L., & Boyle, P. (2016). Impact of flipped classroom design on student performance and perceptions in a pharmacotherapy course. *American Journal of Pharmaceutical Education*, 80(2), 1–9. <https://doi.org/10.5688/ajpe80233>
- Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction*, 22(2), 121–132. <https://doi.org/10.1016/j.learninstruc.2011.08.004>
- Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., & Adesope, O. (2015). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *Internet and Higher Education*, 27, 74–89. <https://doi.org/10.1016/j.iheduc.2015.06.002>
- Kramarski, B., Weiss, I., & Sharon, S. (2013). Generic Versus Context-Specific Prompts for Supporting Self-Regulation in Mathematical Problem Solving Among Students With Low or High Prior Knowledge. *Journal of Cognitive Education and Psychology*, 12(2), 197–214.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educational Research Review*, 10, 133–149. <https://doi.org/10.1016/j.edurev.2013.02.002>
- Kyriazos, T. A. (2018). *Applied Psychometrics: Writing-Up a Factor Analysis Construct Validation Study*

- with Examples. *Psychology*, 09(11), 2503–2530. <https://doi.org/10.4236/psych.2018.911144>
- Låg, T., & Sæle, R. G. (2019). Does the Flipped Classroom Improve Student Learning and Satisfaction? A Systematic Review and Meta-Analysis. *AERA Open*, 5(3). <https://doi.org/10.1177/2332858419870489>
- Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the Classroom: A Gateway to Creating an Inclusive Learning Environment. *The Journal of Economic Education*, 31(1), 30–43. <https://doi.org/10.1080/00220480009596759>
- Lai, C.-L., & Hwang, G.-J. (2016). A self-regulated flipped classroom approach to improving students' learning performance in a mathematics course. *Computers & Education*, 100, 126–140. <https://doi.org/10.1016/j.compedu.2016.05.006>
- *Lancellotti, M., Thomas, S., & Kohli, C. (2016). Online video modules for improvement in student learning. *Journal of Education for Business*, 91(1), 19–22. <https://doi.org/10.1080/08832323.2015.1108281>
- *Lape, N. K., Levy, R., Yong, D. H., Haushalter, K. A., Eddy, R., & Hankel, N. (2014). Probing the inverted classroom: A controlled study of teaching and learning outcomes in undergraduate engineering and mathematics. *ASCE Annual Conference and Exposition, Conference Proceedings, USA*, 121, 9475. <https://www.asee.org/public/conferences/32/papers/9475/download>
- Laursen, B., & Hoff, E. (2006). Person-Centered and Variable-Centered Approaches to Longitudinal Data. *Merrill-Palmer Quarterly (Special Issue: Person-Centered and Variable-Centered Approaches to Longitudinal Data)*, 52(3), 377–389.
- Lazonder, A. W., & Harmsen, R. (2016). Meta-Analysis of Inquiry-Based Learning: Effects of Guidance. *Review of Educational Research*, 86(3), 681–718. <https://doi.org/10.3102/0034654315627366>
- Lee, H. W., Lim, K. Y., & Grabowski, B. L. (2010). Improving self-regulation, learning strategy use, and achievement with metacognitive feedback. *Educational Technology Research and Development*, 58(6), 629–648. <https://doi.org/10.1007/s11423-010-9153-6>
- Lee, J., & Choi, H. (2019). Rethinking the flipped learning pre-class: Its influence on the success of flipped learning and related factors. *British Journal of Educational Technology*, 50(2), 934–945. <https://doi.org/10.1111/bjett.12618>
- Lee, S. W. Y., & Tsai, C. C. (2011). Students' perceptions of collaboration, self-regulated learning, and information seeking in the context of Internet-based learning and traditional learning. *Computers in Human Behavior*, 27(2), 905–914. <https://doi.org/10.1016/j.chb.2010.11.016>
- Lehmann, T., Hähnlein, I., & Ifenthaler, D. (2014). Cognitive, metacognitive and motivational perspectives on prelection in self-regulated online learning. *Computers in Human Behavior*, 32, 313–323. <https://doi.org/10.1016/j.chb.2013.07.051>
- Levine, T. R., Asada, K. J., & Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlated in meta-analyses: Evidence and implications of a publication bias against nonsignificant findings. *Communication Monographs*, 76(3), 286–302. <https://doi.org/10.1080/03637750903074685>
- Li, Q., Baker, R., & Warschauer, M. (2020). Using clickstream data to measure, understand, and support self-regulated learning in online courses. *Internet and Higher Education*, 45(January), 100727. <https://doi.org/10.1016/j.iheduc.2020.100727>
- Lim, D. H., & Morris, M. L. (2009). Learner and instructional factors influencing learning outcomes within a blended learning environment. *Educational Technology and Society*, 12(4), 282–293. <https://doi.org/10.2307/jeductechsoci.12.4.282>

- Lim, K. Y., Lee, H. W., & Grabowski, B. (2009). Does concept-mapping strategy work for everyone? the levels of generativity and learners' self-regulated learning skills. *British Journal of Educational Technology*, 40(4), 606–618. <https://doi.org/10.1111/j.1467-8535.2008.00872.x>
- Liu, L., Ripley, D., & Lee, A. (2016). Flipped Learning and Influential Factors: Case Analysis. *Journal of Educational Technology Development and Exchange*, 9(2). <https://doi.org/10.18785/jetde.0902.05>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications.
- Lo, C. K., & Hew, K. F. (2017). A critical review of flipped classroom challenges in K-12 education: possible solutions and recommendations for future research. *Research and Practice in Technology Enhanced Learning*, 12(4), 1-22. <https://doi.org/10.1186/s41039-016-0044-2>
- Lo, C. K., Hew, K. F., & Chen, G. (2017). Toward a set of design principles for mathematics flipped classrooms: A synthesis of research in mathematics education. *Educational Research Review*, 22, 50–73. <https://doi.org/10.1016/j.edurev.2017.08.002>
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767–778. <https://doi.org/10.1093/biomet/88.3.767>
- López-López, J. A., Marín-Martínez, F., Sánchez-Meca, J., Noortgate, W., & Viechtbauer, W. (2014). Estimation of the predictive power of the model in mixed-effects meta-regression: A simulation study. *British Journal of Mathematical and Statistical Psychology*, 67(1), 30-48. <https://doi.org/10.1111/bmsp.12002>
- Lundin, M., Bergviken Rensfeldt, A., Hillman, T., Lantz-Andersson, A., & Peterson, L. (2018). Higher education dominance and siloed knowledge: a systematic review of flipped classroom research. *International Journal of Educational Technology in Higher Education*, 15(20), 1-30. <https://doi.org/10.1186/s41239-018-0101-6>
- Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Munoz-Gama, J. (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Computers in Human Behavior*, 80, 179–196. <https://doi.org/10.1016/j.chb.2017.11.011>
- Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In *The Oxford handbook of quantitative methods: Statistical analysis, Vol. 2* (pp. 551–611). Oxford University Press.
- Mazur, E. (1991). Can We Teach Computers to Teach? *Computers in Physics*, 5(1), 31. <https://doi.org/10.1063/1.4822968>
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19(4–5), 494–513. <https://doi.org/10.1080/09541440701326154>
- *McLaughlin, J. E., Griffin, L. M., Esserman, D. A., Davidson, C. A., Glatt, D. M., Roth, M. T., & Mumper, R. J. (2013). Pharmacy student engagement, performance, and perception in a flipped satellite classroom. *American Journal of Pharmaceutical Education*, 77(9), 1-8 (No. 196). <https://doi.org/10.5688/ajpe779196>
- Merkt, M., Weigand, S., Heier, A., & Schwan, S. (2011). Learning with videos vs. learning with print: The role of interactive features. *Learning and Instruction*, 21(6), 687–704. <https://doi.org/10.1016/j.learninstruc.2011.03.004>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine*, 6(7), 1–6. <https://doi.org/10.1371/journal.pmed.1000097>

References

- Moos, D. C., & Azevedo, R. (2009). Self-efficacy and prior domain knowledge: To what extent does monitoring mediate their relationship with hypermedia learning? *Metacognition and Learning*, 4(3), 197–216. <https://doi.org/10.1007/s11409-009-9045-5>
- Moos, D. C., & Bonde, C. (2016). Flipping the Classroom: Embedding Self-Regulated Learning Prompts in Videos. *Technology, Knowledge and Learning*, 21(2), 1–18. <https://doi.org/10.1007/s10758-015-9269-1>
- Moos, D. C., & Ringdal, A. (2012). Self-Regulated Learning in the Classroom: A Literature Review on the Teacher's Role. *Education Research International*, 2012, 1–15. <https://doi.org/10.1155/2012/423284>
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2), 364–386. <https://doi.org/10.1177/1094428106291059>
- Muijs, D., & Bokhove, C. (2020). *Metacognition and Self-Regulation: Evidence Review*. London. <https://educationendowmentfoundation.org.uk/evidence-summaries/evidence-reviews/metacognition-and-self-regulation-review/>
- *Mullen, J. S., & Sullivan Jr., J. M. (2015). Student-Perceived Effectiveness of Online Content Delivery Modes. *Proceedings of Frontiers in Education Conference, USA*, 1906–1909. <https://doi.org/10.1109/FIE.2015.7344335>
- Müller, N. M., & Seufert, T. (2018). Effects of self-regulation prompts in hypermedia learning on learning performance and self-efficacy. *Learning and Instruction*, 58(April), 1–11. <https://doi.org/10.1016/j.learninstruc.2018.04.011>
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide* (Seventh Ed). Muthén & Muthén.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation*, 26, 125–141.
- Ness, B. M., & Middleton, M. J. (2012). A Framework for Implementing Individualized Self-Regulated Learning Strategies in the Classroom. *Intervention in School and Clinic*, 47(5), 267–275. <https://doi.org/10.1177/1053451211430120>
- Niemi, H. (2002). Active learning—a cultural change needed in teacher education and schools. *Teaching and Teacher Education*, 18(7), 763–780. [https://doi.org/10.1016/S0742-051X\(02\)00042-2](https://doi.org/10.1016/S0742-051X(02)00042-2)
- Ning, H. K., & Downing, K. (2015). A latent profile analysis of university students' self-regulated learning strategies. *Studies in Higher Education*, 40(7), 1328–1346. <https://doi.org/10.1080/03075079.2014.880832>
- Nye, B., Spyros, K., & Hedges, L. V. (2004). How Large Are Teacher Effects? *Educational Evaluation and Policy Analysis*, 26(3), 237–257. <https://doi.org/10.3102/01623737026003237>
- O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *The Internet and Higher Education*, 25, 85–95. <https://doi.org/10.1016/j.iheduc.2015.02.002>
- *Olitsky, N. H., & Cosgrove, S. B. (2016). The better blend? Flipping the principles of microeconomics classroom. *International Review of Economics Education*, 21, 1–11. <https://doi.org/10.1016/j.iiree.2015.10.004>
- *Ossman, K. A., & Warren, G. (2014). Effect of Flipping the Classroom on Student Performance in First Year Engineering Courses. *ASEE Annual Conference and Exposition, Conference Proceedings, USA*, 121, 8754. <https://peer.asee.org/20342>

- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist*, 38(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, 1–28. <https://doi.org/10.3389/fpsyg.2017.00422>
- Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: when measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research*, 60(6), 723–735. <https://doi.org/10.1080/00313831.2015.1066436>
- Paris, S. G., & Paris, A. H. (2001). Classroom Applications of Research on Self-Regulated Learning. *Educational Psychologist*, 36(2), 89–101. <https://doi.org/10.1207/S15326985EP3602>
- Pekrun, R. (2020). Self-Report is Indispensable to Assess Students' Learning. *Frontline Learning Research*, 8(3), 185–193. <https://doi.org/10.14786/flr.v8i3.637>
- *Peterson, D. J. (2015). The Flipped Classroom Improves Student Achievement and Course Satisfaction in a Statistics Course: A Quasi-Experimental Study. *Teaching of Psychology*, 43(1), 10–15. <https://doi.org/10.1177/0098628315620063>
- 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.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. MI. <https://doi.org/10.5901/mjss.2015.v6n1p156>
- Polanin, J. R., & Pigott, T. D. (2015). The use of meta-analytic statistical significance testing. *Research synthesis methods*, 6(1), 63–73. <https://doi.org/10.1002/jrsm.1124>
- Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the Difference Between Published and Unpublished Effect Sizes: A Meta-Review. *Review of Educational Research*, 86(1), 207–236. <https://doi.org/10.3102/0034654315582067>
- Puustinen, M., & Pulkkinen, L. (2001). Models of Self-regulated Learning: A review. *Scandinavian Journal of Educational Research*, 45(3), 269–286. <https://doi.org/10.1080/00313830120074206>
- Quigley, A., Muijs, D., & Stringer, E. (2018). *Metacognition and Self-regulated Learning: Guidance Report. The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit*. London. <https://educationendowmentfoundation.org.uk/tools/guidance-reports/metacognition-and-self-regulated-learning/>
- R Development Core Team. (2008). R: A Language and Environment for Statistical Computing. Vienna, Austria. <http://www.r-project.org>
- Radhakrishnan, P., Lam, D., & Ho, G. (2009). Giving university students incentives to do homework improves their performance. *Journal of Instructional Psychology*, 36(3), 219.
- Ramdass, D., & Zimmerman, B. J. (2011). Developing Self-Regulation Skills: The Important Role of Homework. *Journal of Advanced Academics*, 22(2), 194–218. <https://doi.org/10.1177/1932202X1102200202>
- *Reza, S., & Ijaz Baig, M. (2015). A study of inverted classroom pedagogy in computer science teaching. *International Journal of Research Studies in Educational Technology*, 4(2), 19–30. <https://doi.org/10.5861/ijrset.2015.1091>
- Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>

- Roehl, A., Reddy, S. L., & Shannon, G. J. (2013). The flipped classroom: An opportunity to engage millennial students through active learning strategies. *Journal of Family & Consumer Sciences*, 105(2), 44–49.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Vol. 6). Sage Publications.
- Rovers, S. F. E., Clarebout, G., Savelberg, H. H. C. M., de Bruin, A. B. H., & van Merriënboer, J. J. G. (2019). Granularity matters: comparing different ways of measuring self-regulated learning. *Metacognition and Learning*, 14(1), 1–19. <https://doi.org/10.1007/s11409-019-09188-6>
- Ryan, R., & Deci, E. (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>
- Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R. M., Abrami, P. C., Surkes, M. A., Wade, C. A., & Woods, J. (2014). The effects of technology use in postsecondary education: A meta-analysis of classroom applications. *Computers and Education*, 72, 271–291. <https://doi.org/10.1016/j.compedu.2013.11.002>
- Schmidt, H. G., Wagener, S. L., Smeets, G. A. C. M., Keemink, L. M., & Van Der Molen, H. T. (2015). On the Use and Misuse of Lectures in Higher Education. *Health Professions Education*, 1(1), 12–18. <https://doi.org/10.1016/j.hpe.2015.11.010>
- Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological bulletin*, 143(6), 565. <https://doi.org/10.1037/bul0000098>
- Schuitema, J., Peetsma, T., & van der Veen, I. (2012). Self-regulated learning and students' perceptions of innovative and traditional learning environments: A longitudinal study in secondary education. *Educational Studies*, 38(4), 397–413. <https://doi.org/10.1080/03055669.2011.643105>
- Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. *Learning and Instruction*, 14(3), 293–305. <https://doi.org/10.1016/j.learninstruc.2004.06.005>
- Schwonke, R., Ertelt, A., Otieno, C., Renkl, A., Aleven, V., & Salden, R. J. C. M. (2013). Metacognitive support promotes an effective use of instructional resources in intelligent tutoring. *Learning and Instruction*, 23(1), 136–150. <https://doi.org/10.1016/j.learninstruc.2012.08.003>
- *Scott, C. E., Green, L. E., & Etheridge, D. L. (2016). A comparison between flipped and lecture-based instruction in the calculus classroom. *Journal of Applied Research in Higher Education*, 8(2), 252–264. <https://doi.org/10.1108/JARHE-04-2015-0024>
- Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2020). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*, 107. <https://doi.org/10.1016/j.chb.2018.05.004>
- Seery, M. K. (2015). Flipped learning in higher education chemistry: emerging trends and potential directions. *Chemistry Education Research and Practice*, 16(4), 758–768. <https://doi.org/10.1039/C5RP00136F>
- Sergis, S., Sampson, D. G., & Pelliccione, L. (2018). Investigating the impact of Flipped Classroom on students' learning experiences: A Self-Determination Theory approach. *Computers in Human Behavior*, 78, 368–378. <https://doi.org/10.1016/j.chb.2017.08.011>
- Setren, E., Greenberg, K., Moore, O., & Yankovich, M. (2019). Effects of the Flipped Classroom:

- Evidence from a Randomized Trial. *School effectiveness & inequality initiative*. <https://seii.mit.edu/wp-content/uploads/2019/08/SEII-Discussion-Paper-2019.07-Setren.pdf>
- Seufert, T. (2020). Building Bridges Between Self-Regulation and Cognitive Load—an Invitation for a Broad and Differentiated Attempt. *Educational Psychology Review*, 32, 1151–1162. <https://doi.org/10.1007/s10648-020-09574-6>
- Shibukawa, S., & Taguchi, M. (2019). Exploring the difficulty on students' preparation and the effective instruction in the flipped classroom. *Journal of Computing in Higher Education*, 31(2), 311–339. <https://doi.org/10.1007/s12528-019-09220-3>
- Shih, H. J., & Huang, S. C. (2019). College students' metacognitive strategy use in an EFL flipped classroom. *Computer Assisted Language Learning*, 1–30. <https://doi.org/10.1080/09588221.2019.1590420>
- Shih, M., Liang, J. C., & Tsai, C. C. (2018). Exploring the role of university students' online self-regulated learning in the flipped classroom: a structural equation model. *Interactive Learning Environments*, 27(8), 1192–1206. <https://doi.org/10.1080/10494820.2018.1541909>
- Shyr, W. J., & Chen, C. H. (2018). Designing a technology-enhanced flipped learning system to facilitate students' self-regulation and performance. *Journal of Computer Assisted Learning*, 34(1), 53–62. <https://doi.org/10.1111/jcal.12213>
- Sitzmann, T., Bell, B. S., Kraiger, K., & Kanar, A. M. (2009). A Multilevel Analysis Of The Effect Of Prompting Self-Regulation In Technology-Delivered Instruction. *Personnel Psychology*, 62, 697–734.
- Sitzmann, T., & Ely, K. (2010). Sometimes You Need a Reminder: The Effects of Prompting Self-Regulation on Regulatory Processes, Learning, and Attrition. *Journal of Applied Psychology*, 95(1), 132–144. <https://doi.org/10.1037/a0018080>
- Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction. *Personnel Psychology*, 59(3), 623–664. <https://doi.org/10.1111/j.1744-6570.2006.00049.x>
- Slavin, R. E. (1991). Synthesis of research of cooperative learning. *Educational Leadership*, 48(5), 71–82.
- Sletten, S. R. (2015). Investigating Self-Regulated Learning Strategies in the Flipped Classroom. In *Society for Information Technology & Teacher Education International Conference* (Vol. 1, pp. 497–501).
- Sletten, S. R. (2017). Investigating Flipped Learning: Student Self-Regulated Learning, Perceptions, and Achievement in an Introductory Biology Course. *Journal of Science Education and Technology*, 26(3), 347–358. <https://doi.org/10.1007/s10956-016-9683-8>
- Spanjers, I. A. E., Könings, K. D., Leppink, J., Verstegen, D. M. L., de Jong, N., Czabanowska, K., & van Merriënboer, J. J. G. (2015). The promised land of blended learning: Quizzes as a moderator. *Educational Research Review*, 15, 59–74. <https://doi.org/10.1016/j.edurev.2015.05.001>
- Steiner, P. M., Wroblewski, A., & Cook, T. D. (2009). Randomized Experiments and Quasi-Experimental Designs in Educational Research. In K. E. Ryan & J. Bradley Cousins (Eds.), *The SAGE International Handbook of Educational Evaluation*. SAGE Publications. <https://doi.org/10.4135/9781452226606>
- Strayer, J. F. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning Environments Research*, 15(2), 171–193. <https://doi.org/10.1007/s10984-012-9108-4>

References

- Strelan, P., Osborn, A., & Palmer, E. (2020). Student satisfaction with courses and instructors in a flipped classroom: A meta-analysis. *Journal of Computer Assisted Learning*, 36(3), 295–314. <https://doi.org/10.1111/jcal.12421>
- Sun, J. C., Wu, Y., & Lee, W. (2016). The effect of the flipped classroom approach to OpenCourseWare instruction on students' self-regulation. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.12444>
- Sun, Z., Xie, K., & Anderman, L. H. (2018). The role of self-regulated learning in students' success in flipped undergraduate math courses. *Internet and Higher Education*, 36, 41–53. <https://doi.org/10.1016/j.iheduc.2017.09.003>
- Surma, T., Vanhoyweghen, K., Camp, G., & Kirschner, P. A. (2018). The coverage of distributed practice and retrieval practice in Flemish and Dutch teacher education textbooks. *Teaching and Teacher Education*, 74, 229–237. <https://doi.org/10.1016/j.tate.2018.05.007>
- Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Tan, C., Yue, W.-G., & Fu, Y. (2017). Effectiveness of flipped classrooms in nursing education: Systematic review and meta-analysis. *Chinese Nursing Research*, 4(4), 192–200. <https://doi.org/10.1016/j.CNRE.2017.10.006>
- Talbert, R. (2015). Inverting the Transition-to-Proof Classroom. *PRIMUS*, 25(8), 614–626. <https://doi.org/10.1080/10511970.2015.1050616>
- Taub, M., Azevedo, R., Bouchet, F., & Khosravifar, B. (2014). Can the use of cognitive and metacognitive self-regulated learning strategies be predicted by learners' levels of prior knowledge in hypermedia-learning environments? *Computers in Human Behavior*, 39, 356–367. <https://doi.org/10.1016/j.chb.2014.07.018>
- Torgerson, C. J. (2006). Publication bias: The achilles' heel of systematic reviews? *British Journal of Educational Studies*, 54(1), 89–102. <https://doi.org/10.1111/j.1467-8527.2006.00332.x>
- *Touchton, M. (2015). Flipping the Classroom and Student Performance in Advanced Statistics: Evidence from a Quasi-Experiment. *Journal of Political Science Education*, 11(1), 28–44. <https://doi.org/10.1080/15512169.2014.985105>
- *Tune, J. D., Sturek, M., & Basile, D. P. (2013). Flipped classroom model improves graduate student performance in cardiovascular, respiratory, and renal physiology. *Advances in Physiology Education*, 37(4), 316–320. <https://doi.org/10.1152/advan.00091.2013>
- Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). "How many studies do you need? A primer on statistical power for meta-analysis": Erratum. *Journal of Educational and Behavioral Statistics*, 35(3), 375. <http://dx.doi.org/10.3102/1076998610376621>
- Van Laer, S., & Elen, J. (2017). In search of attributes that support self-regulation in blended learning environments. *Education and Information Technologies*, 22(4), 1395–1454. <https://doi.org/10.1007/s10639-016-9505-x>
- Vandewaetere, M., & Clarebout, G. (2011). Can instruction as such affect learning? the case of learner control. *Computers and Education*, 57(4), 2322–2332. <https://doi.org/10.1016/j.compedu.2011.05.020>
- Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, 27(1), 118–130. <https://doi.org/10.1016/j.chb.2010.07.038>

- 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>
- Veenman, M. V. J. (2016). Learning to self-monitor and self-regulate. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 233–257). Routledge.
- Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, 1(1), 3–14. <https://doi.org/10.1007/s11409-006-6893-0>
- Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal Of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125. <https://doi.org/10.1002/jrsm.11>
- Vukman, K. B., & Licardo, M. (2010). How cognitive, metacognitive, motivational and emotional self-regulation influence school performance in adolescence and early adulthood. *Educational Studies*, 36(3), 259–268. <https://doi.org/10.1080/03055690903180376>
- Wagner, M., Gegenfurtner, A., & Urhahne, D. (2020). Effectiveness of the Flipped Classroom on Student Achievement in Secondary Education: A Meta-Analysis. *Zeitschrift Für Pädagogische Psychologie*.
- Walker, J. D., Cotner, S. H., Baepler, P. M., & Decker, M. D. (2008). A Delicate Balance: Integrating Active Learning into a Large Lecture Course. *CBE Life Sciences Education*, 7, 361–367. <https://doi.org/10.1187/cbe.08-02-0004>
- Wang, F. H. (2019). On prediction of online behaviors and achievement using self-regulated learning awareness in flipped classrooms. *International Journal of Information and Education Technology*, 9(12), 874–879. <https://doi.org/10.18178/ijiet.2019.9.12.1320>
- Wanner, T., & Palmer, E. (2015). Personalising learning: exploring student and teacher perceptions about flexible learning and assessment in a flipped university course. *Computers & Education*, 88, 354–369. <https://doi.org/10.1016/j.compedu.2015.07.008>
- *Wasserman, N. H., Quint, C., Norris, S. A., & Carr, T. (2015). Exploring Flipped Classroom Instruction in Calculus III. *International Journal of Science and Mathematics Education*, 15, 545–568. <https://doi.org/10.1007/s10763-015-9704-8>
- Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R. W., & Davis-Kean, P. (2006). Development of Achievement Motivation. In *Handbook of child psychology: Social, emotional, and personality development*, Vol. 3, 6th ed. (6th ed., pp. 933–1002). John Wiley & Sons.
- Wigfield, A., Klauda, S. L., & Cambria, J. (2011). Influences on the Development of Academic Self-Regulatory Processes. In D. H. S. Barry J. Zimmerman (Ed.), *Handbook of Self-Regulation of Learning and Performance* (pp. 33–48). New York: Routledge. <https://doi.org/10.4324/9780203839010.ch3>
- Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning*, 9(2), 229–237. <https://doi.org/10.1007/s11409-014-9113-3>
- Wittrock, M. C. (1992). Generative Learning Processes of the Brain. *Educational Psychologist*, 27, 531–541. https://doi.org/10.1207/s15326985ep2704_8
- Wolters, C. A., Pintrich, P. R., & Karabenick, S. A. (2005). *Assessing academic self-regulated learning*. In K. A. Moore & L. H. Lippman (Eds.), *What do children need to flourish: conceptualizing and measuring indicators of positive development*. Springer Science Business Media.

- <https://doi.org/10.1007/0-387-23823-9>
- Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G. J., & Paas, F. (2019). Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review. *International Journal of Human-Computer Interaction*, 35(4–5), 356–373. <https://doi.org/10.1080/10447318.2018.1543084>
- Wormington, S. V., Corpus, J. H., & Anderson, K. G. (2012). A person-centered investigation of academic motivation and its correlates in high school A person-centered investigation of academic motivation and its correlates in high school. *Learning and Individual Differences*, 22(4), 429–438. <https://doi.org/10.1016/j.lindif.2012.03.004>
- Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers and Education*, 140. <https://doi.org/10.1016/j.compedu.2019.103599>
- Yang, T.-C., Chen, M. C., & Chen, S. Y. (2018). The influences of self-regulated learning support and prior knowledge on improving learning performance. *Computers & Education*, 126, 37–52. <https://doi.org/10.1016/j.compedu.2018.06.025>
- Yeh, Y. F., Chen, M. C., Hung, P. H., & Hwang, G. J. (2010). Optimal self-explanation prompt design in dynamic multi-representational learning environments. *Computers and Education*, 54(4), 1089–1100. <https://doi.org/10.1016/j.compedu.2009.10.013>
- Yilmaz, R. M., & Baydas, O. (2017). An examination of undergraduates' metacognitive strategies in pre-class asynchronous activity in a flipped classroom. *Educational Technology Research and Development*, 65(6), 1547–1567. <https://doi.org/10.1007/s11423-017-9534-1>
- Yilmaz, F. G. K., Olpak, Y. Z., & Yilmaz, R. (2018). The Effect of the Metacognitive Support via Pedagogical Agent on Self-Regulation Skills. *Journal of Educational Computing Research*, 56(2), 159–180. <https://doi.org/10.1177/0735633117707696>
- You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *Internet and Higher Education*, 29, 23–30. <https://doi.org/10.1016/j.iheduc.2015.11.003>
- 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(2), 187–202. <https://doi.org/10.1007/s12564-016-9426-9>
- Zheng, L., Bhagat, K. K., Zhen, Y., & Zhang, X. (2020). The Effectiveness of the Flipped Classroom on Students' Learning Achievement and Learning Motivation: A Meta-Analysis. *Educational Technology & Society*, 23(1), 1–15. <https://www.jstor.org/stable/26915403>
- Zimmerman, B. J. (2000). Attaining Self-Regulation a Social Cognitive Perspective. In *Handbook of Self-Regulation* (pp. 13–39). New York: University of New York. <https://doi.org/10.1016/B978-012109890-2/50031-7>
- Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, 82(1), 51–59. <https://doi.org/10.1037/0022-0663.82.1.51>
- Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In A. C. Hacker D. J., Dunlosky J., Graesser (Ed.), *Handbook of Metacognition in Education* (pp. 299–315). Routledge.

Ref

List of Supplementary Materials

Chapter 2: Effects of Flipped Learning on Learning Outcomes and Satisfaction: A Meta-Analysis

Dataverse: <https://doi.org/10.34894/NSXMET>

- Table S1: Description of the studies that were excluded because insufficient effect size information was provided
- Table S2: Descriptive information for each included study, effect size(s), and sample sizes
- Appendix A: References of studies reported in Table S1 which were excluded due to insufficient effect size information
- Appendix B: References included studies in the meta-analyses
- Appendix C: Calculation formulas 8, 9, 10 and 25 from Morris (2008) were used to calculate effect size and variance from studies with a pretest-posttest-Control group design
- Figure S1: Funnel plot of assessed learning outcome
- Figure S2: Funnel plot of perceived learning outcome
- Figure S3: Funnel plot of satisfaction
- Figure S4: Cumulative meta-analysis assessed learning outcome
- Figure S5: Cumulative meta-analysis perceived learning outcome
- Figure S6: Cumulative meta-analysis satisfaction
- Figure S7: Forest plot assessed learning outcome sorted on effect size
- Figure S8: Forest plot perceived learning outcome sorted on effect size
- Figure S9: Forest plot satisfaction sorted on effect size
- Coding schemes study level and effect size level
- Research data, search process, data syntax, and data analysis

Chapter 3: Effects of Self-Regulated Learning Prompts In a Flipped History Classroom

Dataverse: <https://doi.org/10.34894/C5SRAZ>

- Learning materials in class (in Dutch)
- Learning materials out class (instructional videos and SRL prompts; in Dutch)
- Data collection materials (in Dutch)
- Raw data and syntax

Chapter 4: Self-Regulated Learning Support in Flipped Learning Videos Enhances Learning Outcomes

Dataverse: <https://doi.org/10.34894/YIOF09>

- Learning materials in class (in Dutch)
- Learning materials out class (instructional videos and SRL prompts; in Dutch)
- Data collection materials (in Dutch)
- Research data and syntax

Chapter 5: Secondary Students' Online Self-Regulated Learning During Flipped Learning: A Latent Profile Analysis

Dataverse: <https://doi.org/10.34894/WRUQOG>

- Learning materials in class and out class (in Dutch)
- Data collection materials (in Dutch)
- Research data and syntax

Sup

Nederlandse Samenvatting (Summary)

In dit proefschrift worden vier onderzoeken beschreven naar *flipped learning* (FL) en de rol die zelfregulerend leren (ZRL) daarin speelt. In **Hoofdstuk 1** geven we een overkoepelende inleiding met daarin de theoretische achtergrond van onze onderzoeksvragen. Onder FL verstaan we het didactische principe dat leerlingen (of studenten) de lessen thuis voorbereiden. Dat doen ze bijvoorbeeld door de leerstof via een (online) instructievideo te bestuderen. In de les passen de leerlingen vervolgens deze kennis toe door verschillende actieve verwerkingsopdrachten uit te voeren. Deze methode, ook wel bekend als *flipping the classroom*, is in de afgelopen tien jaar populair geworden in het voortgezet- en hoger onderwijs. In de praktijk wordt FL vaak toegepast met behulp van (online) videotechnologie. Wij beargumenteren in dit inleidende hoofdstuk dat wij voor een brede definitie van FL kiezen, omdat FL vormen waarbij geen gebruik gemaakt wordt van technologie ook mogelijk zijn. Verder schetsen wij in de theoretische achtergrond dat FL vaak als effectieve onderwijsmethode wordt gezien, vanwege een aantal theoretische veronderstellingen. Er zijn onder andere bepaalde beperkingen aan het geven van lange groepsinstructie in traditionelere vormen van onderwijs, zoals de beperkte mate van interactie en actieve verwerking (Chi & Wylie, 2014; Schmidt et al., 2015). FL kan wat dit laatste betreft voordelen bieden. Als de instructie in een *flipped class* door de leerling individueel voor de les wordt bestudeerd, geeft dit de leerling meer ruimte naar eigen inzicht te interacteren met de instructie bijvoorbeeld door de frequentie en het tempo van de instructie zelf te bepalen (Abeysekera & Dawson, 2015). Deze controle over de instructie door de leerling zou cognitieve overbelasting kunnen verminderen, bijvoorbeeld omdat leerlingen de instructie kunnen pauzeren en terugspoelen (Abeysekera & Dawson, 2015). Daarnaast kunnen leerlingen tijdens FL de moeilijke leerstof beter toepassen in de les (een actieve verwerking van de leerstof), omdat de docent dan meer tijd en ruimte heeft om dit leerproces te sturen (Roehl et al., 2013; Van den Bergh et al., 2014).

Veel van deze aannames zijn theoretisch goed onderbouwd. Echter was er bij de start van dit promotieonderzoek een grote behoefte aan empirisch onderzoek naar de effecten van FL ten opzichte van een gedegen controlegroep zonder FL (Bishop & Verleger, 2013; O'Flaherty & Phillips, 2015). Onze eerste onderzoeksvraag werd dan ook: wat zijn de effecten van FL op leerprestaties en tevredenheid van leerlingen?

In **Hoofdstuk 2** beschrijven we de resultaten van een systematische review met drie meta-analyses naar de effecten van FL op (a) door een docent vastgestelde leerprestaties (bijvoorbeeld een toets), (b) door leerlingen gerapporteerde leerprestaties (bijvoorbeeld een vragenlijst), en (c) de tevredenheid van leerlingen over de leeromgeving. Na onze zorgvuldige zoektocht en kritische beoordeling van de

gevonden studies bleven 114 studies (met in totaal 20.318 deelnemende leerlingen) over. We vonden dat de door een docent vastgestelde leerprestaties tijdens FL gemiddeld beter waren dan in de controlegroep waar een traditionele manier van lesgeven werd gebruikt. Over het gemiddelde effect van FL op leerling gerapporteerde leerprestaties kunnen we minder duidelijke conclusies trekken door het kleine aantal studies die dit gemeten hebben. Het effect voor de door een leerling gerapporteerde leerprestaties was even groot als de door een docent vastgestelde leerprestaties. Verder vonden we geen effect van FL op de tevredenheid van leerlingen over de leeromgeving. In tegenstelling tot wijdverspreide veronderstellingen, leidt FL dus niet automatisch tot meer tevredenheid over de leeromgeving onder leerlingen. Omdat we in alle analyses een significante spreiding vonden tussen de effectgroottes, hebben we met moderator meta-regressie analyses onderzocht of bepaalde designkenmerken van FL, namelijk, studieomgevingskenmerken of studiekwaliteitskenmerken deze spreiding konden verklaren. We vonden dat het gelijk houden van de lestijd (dus niet verkorten omdat de leerlingen nu meer thuis voorbereiden) en het toepassen van quizjes over de leerstof (tijdens de instructievideo of tijdens de verwerking in de les) leidden tot effectievere vormen van FL in termen van hogere leerprestaties.

In het tweede deel van dit proefschrift onderzoeken we de relatie tussen zelfregulatie van het leren (ZRL) en FL. Onder ZRL verstaan we cognitieve, metacognitieve en motivationele strategieën die een leerling inzet om het eigen leerproces te reguleren (Zimmerman & Moylan, 2009). Tijdens FL wordt er een groter beroep gedaan op de autonomie van leerlingen, wanneer zij de leerstof voor de les zelfstandig voorbereiden (He et al., 2016; Lee & Tsai, 2011; Shih & Huang, 2019; Tan et al., 2017). Uit eerdere onderzoeken blijkt dat leerlingen hier moeite mee kunnen hebben, en dat ondersteuning van ZRL kan leiden tot het verbeteren van deze zelfregulatievaardigheden en de leerprestaties (Dent & Koenka, 2016; Dignath & Büttner, 2008; Donker et al., 2014; Jansen et al., 2019; Zheng, 2016). In het primair en hoger onderwijs zijn deze vormen van ZRL-ondersteuning in de context van FL onderzocht en zijn er positieve resultaten gevonden (Lai & Hwang, 2016; Moos & Bonde, 2016; Shyr & Chen, 2018; Yilmaz et al., 2018). In het voortgezet onderwijs is hier nog geen onderzoek naar gedaan.

Hoofdstuk 3 beschrijft de studie die we hebben gedaan naar de effectiviteit van ZRL prompts tijdens FL instructievideo's onder 154 tweedeklas leerlingen tijdens een zesweekse lessenserie in een ecologisch valide schoolsituatie. Door de helft van deze leerlingen te voorzien van ZRL prompts tijdens de video's en de andere helft niet, onderzochten we het effect van ZRL prompts op (a) een zelf gerapporteerde ZRL (vragenlijst), (b) het online ZRL gedrag (gemeten in de elektronische leeromgeving, bijvoorbeeld het terugspoelen van video's en het op tijd bekijken van de video's), en (c) de leerprestaties (eindtoets). De ZRL prompts waren korte onderbrekingen

waarin leerlingen gestimuleerd en geholpen werden om op een effectieve manier over hun eigen leerproces na te denken. Aan het begin van een video vroegen we leerlingen bijvoorbeeld om te reflecteren op hun studieaanpak, en na te denken over hoe ze ervoor gingen zorgen dat ze alle belangrijke leerstof zouden gaan onthouden. In tegenstelling tot eerder onderzoek in primair en hoger onderwijs, lieten de resultaten van onze kwantitatieve analyses zien dat de ZRL prompts geen significant effect hadden op zelf gerapporteerde ZRL van leerlingen en de leerprestaties op de afsluitende toets. Voor het online ZRL gedrag vonden we een significant verschil voor het video voltooiingspercentage. Dat betekent dat leerlingen die de ZRL ondersteuning kregen gemiddeld meer video's hebben (af)gekeken dan de leerlingen die geen ZRL ondersteuning ontvingen.

In de studie beschreven in **Hoofdstuk 4** hebben we de studieopzet uit **Hoofdstuk 3** nog een keer toegepast, maar dan met een verbeterd ontwerp van de ZRL ondersteuning. We hebben de ZRL ondersteuning aangepast op basis van de kwalitatieve feedback van leerlingen uit de studie uit **Hoofdstuk 3**. Zo hebben we bijvoorbeeld expliciete ZRL instructies toegevoegd aan de ondersteuning. Het bleek voor leerlingen niet vanzelfsprekend waarom het regelmatig nadenken over het eigen leerproces nuttig is. Met ZRL instructies kunnen we leerlingen expliciet uitleggen waarom de ZRL prompts en het regelmatig reflecteren op het eigen leerproces nuttig is, maar ook begeleiden bij hoe ze dit het beste kunnen aanpakken (Bannert et al., 2009; Jansen et al., 2020; Kistner et al., 2010). Met dezelfde meetinstrumenten en lesmaterialen als in het onderzoek uit **Hoofdstuk 3** hebben we de vervolgstudie uit **Hoofdstuk 4** in een nieuwe groep van 156 tweedeklasleerlingen uitgevoerd. We onderzochten de effecten van ZRL ondersteuning op (a) de zelf gerapporteerde ZRL, (b) het online ZRL gedrag, (c) de leerprestaties, en (d) de tevredenheid van leerlingen over de leeromgeving. Daarnaast hebben we kwalitatieve data verzameld over hoe de leerlingen die ZRL ondersteuning kregen dit ervoeren. De resultaten van de kwantitatieve analyses gaven aan dat de leerlingen die ZRL ondersteuning kregen een significant hoger cijfer behaalden op de eindtoets. Echter vonden we ook in deze studie geen effect op de zelf gerapporteerde ZRL vaardigheden en het online ZRL gedrag van leerlingen. En hoewel de ZRL ondersteuning de leerlingen hielp zich meer bewust te worden van hun eigen leerproces en het leidde tot betere leerprestaties, gaven sommige leerlingen duidelijk aan dat ze deze ondersteuning negatief waardeerden. Leerlingen gaven bijvoorbeeld aan dat ze het gevoel hadden dat de prompts om na te denken over hun leerproces voor afleiding zorgden bij het bestuderen van de instructievideo. Toch haalden ze door de ZRL ondersteuning een beter cijfer op de toets dan de leerlingen die dat niet kregen.

Tijdens onze empirische onderzoeken viel het op dat er een hele grote spreiding was in het online ZRL gedrag van leerlingen. Veel onderzoek in het hoger onderwijs heeft

zich gericht op het onderzoeken van ZRL profielschetsen op basis van zelfrapportages (Maldonado-Mahauad, Pérez-Sanagustín, Kizilcec, Morales, & Munoz-Gama, 2018; Ning & Downing, 2015; Vanslambrouck et al., 2019). In **Hoofdstuk 5** hebben wij een dergelijke latente profielanalyse voor het eerst toegepast op data van *flipped* klassen uit het voortgezet onderwijs. Daarnaast hebben wij voor de onderwijspraktijk relevante en vrij standaard verzamelde online ZRL gedragsvariabelen gebruikt, die we hebben gemeten in een online leeromgeving (zoals hoe vaak een video werd teruggespoeld, of hoe vaak de video op tijd werd bekeken). De onderzoeksvragen zijn (a) welke ZRL profielschetsen van leerlingen kunnen we maken op basis van hun waargenomen online ZRL gedrag?, (b) zijn deze ZRL profielen gerelateerd aan verschillen op de eindtoets die leerprestaties meet?, en (c) kunnen zelf gerapporteerde ZRL en motivatie, en voorkennis van leerlingen het ZRL profiel lidmaatschap voorspellen?

In lijn met eerder onderzoek in andere onderwijscontexten vonden wij vijf significant te onderscheiden ZRL profielen. De ZRL profielen lopen uiteen van een groep leerlingen met vrijwel geen gemeten ZRL activiteiten (zoals video's niet terugspoelen, video's na de deadline pas bekijken, en niet op een later moment terugkijken) tot een groep leerlingen die heel veel ZRL activiteiten liet zien. Daarnaast vonden we dat leerlingen in het ZRL profiel waarin minder video's zijn afgekeken en minder ZRL gedrag werd vertoond, lager scoorden op de leeruitkomsten van de eindtoets dan leerlingen in de drie profielen met een hoger videovoltooingspercentage die meer ZRL gedrag lieten zien. Tot slot vonden we geen bewijs dat zelf gerapporteerde ZRL en motivatie, en voorkennis van leerlingen het ZRL profiel lidmaatschap konden voorspellen.

Tot slot bevat **Hoofdstuk 6** een overkoepelende discussie waarin we de methoden en resultaten uit alle individuele hoofdstukken aan elkaar relateren. Wij gaan eerst in op de wetenschappelijke implicaties en praktische aanbevelingen, en daarna op limitaties van ons onderzoek en aanbevelingen voor vervolgonderzoek. Onze meta-analyse over de effecten van FL op leerprestaties heeft allereerst een belangrijke bijdrage geleverd aan de wetenschappelijke kennisbasis. Zowel in maatschappelijke als wetenschappelijke discussies over de effecten van FL worden vaak enkele negatieve of positieve studies aangehaald. Door het grote aantal studies dat wij hebben kunnen includeren in de statistische synthese is er steviger bewijs gekomen voor de generaliseerbare conclusie dat FL een positief effect heeft op leerprestaties. Echter concluderen wij ook dat het toepassen van FL geen wondermiddel is, gezien de grote spreiding van de effecten tussen de studies. Op basis van onze moderatoranalyses bevelen we dan ook aan om de lestijd gelijk te houden en quizjes toe te passen over de leerstof (tijdens de instructievideo of tijdens de verwerking in de les), omdat dit de effectiviteit van FL versterkt.

Daarnaast hebben we aangetoond dat FL, in tegenstelling tot wat vaak wordt

verondersteld, niet automatisch leidt tot meer tevredenheid van leerlingen. Een mogelijke verklaring hiervoor is dat leerlingen ervaren dat ze tijdens de lessen meer inspanning moeten leveren bij de verwerkingsactiviteiten en dit ten onrechte associëren met slechtere prestaties op de toets (Deslauriers et al., 2019). Deze misvatting kan mogelijk negatieve gevolgen hebben voor hun motivatie en betrokkenheid, maar ook de bereidheid om ZRL toe te passen (Deslauriers et al., 2019). Dit geldt mogelijk ook voor de leerlingen die in onze experimenten rapporteerden dat ze de ZRL ondersteuning als afleidend ervaren, terwijl ze betere leerprestaties haalden dan leerlingen die deze ondersteuning niet kregen. We raden daarom aan om aan de ene kant kritisch te staan tegenover de validiteit van student tevredenheid als uitkomstmaat. Aan de andere kant raden we aan om deze misvatting onder studenten bij het toepassen van FL danwel ZRL ondersteuning tijdig te adresseren.

Ten derde hebben we met onze ZRL interventiestudies bijgedragen aan belangrijke nieuwe kennis over de rol van het ondersteunen van ZRL bij het toepassen van FL in het voortgezet onderwijs. Sterke punten van ons onderzoek hierbij zijn de relatief lange duur van onze interventie, de ecologisch valide context in het nog weinig onderzochte voortgezet onderwijs, het aantal leerlingen dat heeft meegedaan, en de toepassing van verschillende en elkaar versterkende meetinstrumenten. Op basis van onze bevindingen is ons advies om ZRL prompts in instructievideo's in te zetten om leerlingen regelmatig te stimuleren over hun leerproces te reflecteren, en ZRL instructies daar aan toe te voegen om ook de kwaliteit van deze activiteiten te versterken. De ZRL instructies kunnen mogelijk ook helpen om leerlingen het nut in te laten zien van regelmatige zelfregulatieactiviteiten tijdens het leren, omdat wij aanwijzingen vonden dat sommige leerlingen dit bijvoorbeeld als afleiding ervaren bij het leren van de video.

Ten vierde hebben we met onze studie in **Hoofdstuk 5** aangetoond dat ook in het voortgezet onderwijs veel verschil in online ZRL gedrag is tussen leerlingen in een klas. Hoewel we geen directe verklaringen vonden voor deze grote verschillen, zagen we wel dat deze verschillen gerelateerd zijn aan verschillen op leerresultaten bij de eindtoets. We adviseren docenten die FL toepassen om daar waar het mogelijk is de ZRL ondersteuning aan te laten sluiten bij de specifieke ZRL vaardigheden van de leerling. Leerlingen die bijvoorbeeld moeite hebben om de video's af te kijken zouden meer voordeel van ZRL ondersteuning kunnen ervaren bij het plannen. Leerlingen die overmatig vaak terugspoelen, terwijl uit ons onderzoek blijkt dat ze daar geen extra profijt van hebben bij de toets, zouden meer voordeel kunnen hebben bij het ondersteunen van hun *self efficacy*. Hiermee kan bijvoorbeeld de studie effectiviteit en het zelfvertrouwen van de leerling worden bevorderd.

We hebben in onze studies weinig tot geen bewijs gevonden voor onze hypotheses

dat ZRL ondersteuning leidt tot meer online ZRL gedrag en hogere ZRL zelfrapportage, in vergelijking met leerlingen die geen ZRL ondersteuning krijgen. Wel vonden we dat leerlingen die ZRL ondersteuning kregen betere leerprestaties behaalden dan de leerlingen die dit niet kregen. Onze hypothese was dat de ondersteuning van ZRL zou leiden tot meer ZRL gedrag, wat dan zou leiden tot betere leeruitkomsten. Een mogelijke verklaring voor deze discrepantie is dat we de effectieve leertijd van leerlingen mogelijk hebben vergroot door ze meer te laten reflecteren op hun leerproces. Een andere verklaring is dat de wijze waarop wij ZRL gemeten hebben vooral iets zegt over de hoeveelheid van ZRL activiteiten, maar niet zozeer over de kwaliteit. Een mogelijke manier om dit in vervolgonderzoek ook te kunnen onderzoeken is het toevoegen van retrospectieve interviewdata waarin leerlingen expliciet op de kwaliteit van hun ZRL gedrag worden bevraagd (Jansen et al., 2020). Daarnaast is het mogelijk dat onze onderzoeksopzet de verschillen tussen de leerlingen die wel en geen ZRL ondersteuning hebben gekregen zijn verkleind, omdat wij de relatief jonge leerlingen niet in één keer alle autonomie gaven over hun leerproces: per les was het duidelijk welke instructievideo moest worden bestudeerd. Het is interessant om in vervolgonderzoek te onderzoeken of de effecten van ZRL ondersteuning wel gevonden worden als de leerlingen in deze context meer autonomie krijgen in de onderzoeksopzet.

Op basis van alle hoofdstukken van dit proefschrift concluderen we in de eerste plaats dat het toepassen van FL een effectieve methode kan zijn om de leerprestaties van leerlingen te vergroten, mits de effectieve lestijd niet wordt verkort en quizjes worden toegepast om de kennis van de instructievideo's op te halen en te activeren. Ook concluderen we dat FL niet automatisch leidt tot meer tevredenheid onder leerlingen. Bovendien raden we sterk aan om ZRL te ondersteunen bij het toepassen van FL, aangezien FL hogere eisen stelt aan de ZRL vaardigheden van leerlingen. We hebben aangetoond dat ZRL ondersteuning in FL video's de leerresultaten verbetert als (a) leerlingen regelmatig gestimuleerd worden om te reflecteren op het eigen leerproces, en (b) ZRL instructies worden toegevoegd die leerlingen duidelijk uitleggen waarom ZRL activiteiten nuttig zijn, en hoe ze deze activiteiten effectief kunnen doen. We vonden echter ook grote verschillen in de online ZRL profielen van leerlingen, die mogelijkheden bieden voor toekomstig onderzoek om de ZRL ondersteuning in het voortgezet onderwijs verder te verbeteren.

Curriculum Vitae

David van Alten was born in 1990 in Rotterdam, The Netherlands. He received a Bachelor's degree in History at Utrecht University in 2012 and completed the Researchmaster Ancient studies in 2014 (*cum laude*). During this time, he received a six week scholarship at the Royal Netherlands Institute in Rome (KNIR), and participated in the Archaeological fieldwork project "Thorikos" (Greece) in July 2013 and July 2014, led by the Universities of Ghent and Utrecht. To further explore his interest in teaching History, he completed a second Master to become a History teacher in secondary education in 2015 (*cum laude*). Then, he got the chance to become a PromoDoc and combine his passions for educational research and practice. For five years, he worked on his PhD project about flipped learning for three days a week. His research is about the effects of flipped learning, as well as the role of self-regulated learning behavior of the students. During his PhD, he presented his research findings at several (inter)national conferences and educational institutions. He was also asked to contribute to research projects on the subject of teaching sensitive history subjects and participated in research conferences in Rome and Groningen. His scholarly work is published in journals such as: Educational Research Review, Computers in Human Behavior, and Computers & Education. The other two days of the week he worked as a secondary school History Teacher at Cals College Nieuwegein. In 2019 he also became Information manager, to assist with constructing business plans and the development of ICT in (the organization of) education. Currently, David is combining the roles of Information manager and History teacher at Cals College.

List of Publications and Professional Contributions

Peer Reviewed Publications

- van Alten, D. C. D. (2017). Glocalization and Religious Communication in the Roman Empire: Two Case Studies to Reconsider the Local and the Global in Religious Material Culture. *Religions*, 8(8), 140. <https://doi.org/10.3390/rel8080140>
- van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of Flipping the Classroom on Learning Outcomes and Satisfaction: a Meta-Analysis. *Educational Research Review*, 28, 1–18. <https://doi.org/10.1016/j.edurev.2019.05.003>
- van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2020). Effects of self-regulated learning prompts in a flipped history classroom. *Computers in Human Behavior*, 108, 106318. <https://doi.org/10.1016/j.chb.2020.106318>
- van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2020). Self-regulated learning support in flipped learning videos enhances learning outcomes. *Computers & Education*, 158, 104000. <https://doi.org/10.1016/j.compedu.2020.104000>
- van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2021). Secondary Students' Online Self-Regulated Learning During Flipped Learning: A Latent Profile Analysis. *Computers in Human Behavior*, 118, 106676. <https://doi.org/10.1016/j.chb.2020.106676>

Conference Presentations and Proceedings

- Van Alten, D. C. D. (2017, June). De effecten van flipping the classroom op leerprestaties en tevredenheid: een systematische review en meta-analyse. In J. Janssen (Chair), *Onderwijskundige meta-analyses met een twist*. Symposium conducted at the Onderwijs Research Dagen (ORD), Antwerpen, Belgium.
- Van Alten, D. C. D. (2018, August). *Effects of Flipping the Classroom on Learning Outcomes and Satisfaction: a Meta-Analysis*. Paper presented at the EARLI SIG 6-7: Instructional Design/Technology-Enhanced Learning and Instruction, Bonn, Germany.
- Van Alten, D. C. D. (2018, August). *Self-Regulated Learning prompts in a Flipped History Classroom: effects on Learning Outcomes and Self-Regulated Learning*. Paper presented at the EARLI SIG 6-7: Instructional Design/Technology-Enhanced Learning and Instruction, Bonn, Germany.
- Van Alten, D. C. D. (2018, September). Invited speaker, *Workshop Flipping the Classroom*. *Conferentie opleiden en onderzoek in de school 2018*, Amersfoort.
- Van Alten, D. C. D. & Lozano Parra, S. (2019, January). *Teaching the sensitive past, experiences, challenges and practical examples: a Dutch Approach*. Paper presented at the Symposium *The Sensitive Past*, Royal Netherlands Institute in Rome (KNIR).
- Van Alten, D. C. D. & Lozano Parra, S. (2019, August). Meerstemmige geschiedenis: inclusiviteit in onderwijs over het verleden? Lesgeven over zwarte piet: ervaringen uit de praktijk. *Historicidagen 2019 – Inclusieve Geschiedenis* Groningen.
- van Alten, D. C. D. & Janssen, J. (2020, January). *Best-evidence in onderwijsonderzoek: meta-analyses? Flipping the classroom als casus* Invited speaker ResearchED 2020 Nijkerk.

Van Alten, D. C. D. (2020, January). *The effects of flipping the classroom*. Keynote speaker at Utrecht University, Strengthen your education week.

Professional Contributions

van Alten, D. C. D. (2019, September). *Betere leerprestaties als je flipping the classroom toepast?*, in: Centre for Academic Teaching: Educational Database. https://cat-database.sites.uu.nl/knowledge_item/betere-leerprestaties-als-je-flipping-the-classroom-toepast/

Van De Meent, Y. (2020, January). *Flipping the classroom is geen wondermiddel*, in: Onderwijsblad. <https://www.aob.nl/nieuws/flipping-the-classroom-is-geen-wondermiddel/>

van Alten, D. C. D. & Janssen, J. (2020, January). *Flipping the classroom: wat werkt?*, in: Didactief. <https://didactiefonline.nl/artikel/flipping-the-classroom-wat-werkt>

Leraar24 (2020, April). *Als je goed flipt, kan dit de prestaties van leerlingen verbeteren*. <https://www.leraar24.nl/2619500/als-je-goed-flipt-kan-dit-de-prestaties-van-leerlingen-verbeteren/>

Rous, R. (2019, September). *Succesvol 'flipping': "Begin eenvoudig en klein en maak gebruik van quizjes"*, in: ICTheek/Breens. <https://ictheek.nl/ictrends/succesvol-flippen-begin-eenvoudig-en-klein-en-maak-gebruik-van-quizjes/>

van Alten, D. C. D. (2016, August). *Is het wenselijk om tabletgebruik door jonge kinderen af te stemmen op hun lichamelijke kenmerken of ontwikkeling?* <https://www.nro.nl/wp-content/uploads/2016/10/086-Antwoordformulier-ergonomie.pdf>

van Alten, D. C. D. (2016, September). *Wat weten we over het effect van zittenblijven of versnellen op leerprestaties en de sociaal-emotionele ontwikkeling?* <https://www.nro.nl/wp-content/uploads/2016/10/095-Antwoordformulier-Zittenblijven-en-Versnellen.pdf>

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 (the list will be updated every year in January)

362. Tuithof, J.I.G.M. (23-06-2017). *The characteristics of Dutch experienced history teachers' PCK in the context of a curriculum innovation*. Utrecht: Utrecht University.
363. Van Waes, S. (23-06-2017). *The ties that teach: Teaching networks in higher education*. Antwerp: University of Antwerp.
364. Evens, M. (30-06-2017). *Pedagogical content knowledge of French as a foreign language: Unravelling its development*. Leuven: KU Leuven.
365. Moses, I. (07-09-2017). *Student-teachers' commitment to teaching*. Leiden: Leiden University.
366. Wansink, B.G.J. (15-09-2017). *Between fact and interpretation. Teachers' beliefs and practices in interpretational history teaching*. Utrecht: Utrecht University.
367. Binkhorst, F. (20-10-2017). *Connecting the dots. Supporting the implementation of Teacher Design Teams*. Enschede: University of Twente.
368. Stoel, G.L. (14-11-2017). *Teaching towards historical expertise. Developing students' ability to reason causally in history*. Amsterdam: University of Amsterdam.
369. Van der Veen, M. (28-11-2017). *Dialogic classroom talk in early childhood education*. Amsterdam: Vrije Universiteit Amsterdam.
370. Frèrejean, J. (08-12-2017). *Instruction for information problem solving*. Heerlen: Open University of the Netherlands.
371. Rezende Da Cunha Junior, F. (19-12-2017). *Online groups in secondary education*. Amsterdam: Vrije Universiteit Amsterdam.
372. Van Dijk, A.M. (22-12-2017). *Learning together in mixed-ability elementary classrooms*. Enschede: University of Twente.
373. Bouwmans, M.H.C.F. (12-01-2018) *The role of VET colleges in stimulating teachers' engagement in team learning*. Wageningen: Wageningen University.
374. Jansma, D.J. (25-01-2018) *This is wrong, right? The role of moral components in anti- and prosocial behaviour in primary education*. Groningen: University of Groningen.
375. Okkinga, M. (02-02-2018) *Teaching reading strategies in classrooms- does it work?* Enschede: University of Twente.
376. Thomsen, M. (09-02-2018) *Teachers Trust. Measurement, sources and consequences of teacher's interpersonal trust within schools for vocational education and training*. Amsterdam: University of Amsterdam.
377. Van der Wurff, I.S.M. (09-02-2018) *Fatty acids, Cognition, School Performance and Mental Well-Being in Children and Adolescents*. Heerlen: Open University of the Netherlands.

378. Raaijmakers, S.F. (16-02-2018) *Improving self-regulated learning: Effects of training and feedback on self-assessment and task-selection accuracy*. Utrecht: Utrecht University.
379. Zhao, X. (07-03-2018) *Classroom assessment in Chinese primary school mathematics education*. Utrecht: Utrecht University.
380. Van Rooij, E.C.M. (15-03-2018) *Secondary school students' university readiness and their transition to university*. Groningen: University of Groningen.
381. Vanlommel, K. (26-03-2018) *Opening the black box of teacher judgement: the interplay of rational and intuitive processes*. Antwerp: University of Antwerp.
382. Boevé, A.J. (14-05-2018), *Implementing Assessment Innovations in Higher Education*. Groningen: University of Groningen.
383. Wijsman, L.A. (30-05-2018) *Enhancing Performance and Motivation in Lower Secondary Education*. Leiden: Leiden University.
384. Vereijken, M.W.C. (22-05-2018) *Student engagement in research in medical education*. Leiden: Leiden University.
385. Stollman, S.H.M. (23-05-2018) *Differentiated instruction in practice: A teacher perspective*. Leiden: Leiden University.
386. Faddar, J. (11-06-2018) *School self-evaluation: self-perception or self-deception? Studies on the validity of school self-evaluation results*. Antwerp: University of Antwerp.
387. Geeraerts, K. (25-06-2018) *Dood hout of onaangeboorde expertise? Intergenerationale kennisstromen in schoolteams*. Antwerp: University of Antwerp.
388. Day, I.N.Z. (28-06-2018), *Intermediate assessment in higher education* (Leiden: Leiden University)
389. Huisman, B.A. (12-09-2018) *Peer feedback on academic writing*. Leiden: Leiden University.
390. Van Berg, M. (17-09-2018) *Classroom Formative Assessment. A quest for a practice that enhances students' mathematics performance*. Groningen: University of Groningen.
391. Tran, T.T.Q. (19-09-2018) *Cultural differences in Vietnam : differences in work-related values between Western and Vietnamese culture and cultural awareness at higher education*. Leiden: Leiden University
392. Boelens, R. (27-09-2018) *Studying blended learning designs for hands-on adult learners*. Ghent: Ghent University.
393. Van Laer, S. (4-10-2018) *Supporting learners in control: investigating self-regulated learning in blended learning environments*. Leuven: KU Leuven.
394. Van der Wilt, F.M. (08-10-18) *Being rejected*. Amsterdam: Vrije Universiteit Amsterdam.
395. Van Riesen, S.A.N. (26-10-2018) *Inquiring the effect of the experiment design tool: whose boat does it float?* Enschede: University of Twente.
396. Walhout, J.H. (26-10-2018) *Learning to organize digital information* Heerlen: Open University of the Netherlands.
397. Gresnigt, R. (08-11-2018) *Integrated curricula: An approach to strengthen Science & Technology in primary education*. Eindhoven: Eindhoven University of Technology.
398. De Vetten, A.J. (21-11-2018) *From sample to population*. Amsterdam: Vrije Universiteit Amsterdam.
399. Nederhand M.L. (22-11-2018) *Improving Calibration Accuracy Through Performance Feedback*. Rotterdam: Erasmus University Rotterdam.
400. Kippers, W.B. (28-11-2018) *Formative data use in schools. Unraveling the process*. Enschede: University of Twente.
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 preparation 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

Dankwoord (Acknowledgements)

*I've seen the bridge, and the bridge is long
And they built it high, and they built it strong
Strong enough to hold the weight of time
Long enough to leave some of us behind*

*And every one of us has to face that day
Do you cross the bridge or do you fade away?
And every one of us that ever came to play
Has to cross the bridge or fade away*

Elton John, *The Bridge, The Captain and The Kid* (2006)

De kloof en de brug

Op 30 april 2015 vond de officiële aftrap plaats van het PromoDoc (promoverend docent) pilottraject plaats. Op bezoek bij het ministerie van Onderwijs, Cultuur en Wetenschap vertelde Jan van Tartwijk aan de toenmalige minister over het doel van het PromoDoc project. Er bestaat een diepe kloof tussen onderwijsonderzoek en de praktijk, tussen de universiteit en de school. Het ambitieuze doel is om daar een brug tussen te slaan. En wat heb ik daar de afgelopen vijf jaren enorm van genoten. Ik dank daarbij met name Artur Bakker, Gijsbert Erkens, Hanneke Tuithof, en Jan van Tartwijk die mij tijdens de lerarenopleiding het laatste zetje gaven om het avontuur aan te gaan.

En hoewel de kloof tussen onderzoek en praktijk soms inderdaad diep kan zijn, hoop ik dat ik mijn eigen bescheiden bijdrage heb kunnen leveren. Balancerend over de kloof, elke werkdag op zoek naar de juiste balans en loyaliteit. Soms waren er dagen dat de brug echt even heel zichtbaar was. Bijvoorbeeld op 20 mei 2016 in een taxibusje met Jan van Tartwijk en Michel Rog (2^e kamerlid onderwijs) op weg naar het Cals. Of op 19 april 2017 toen een ambtenaar van OCW, twee afgevaardigden van het ISO, en Jan van Tartwijk op bezoek kwamen in Nieuwegein. Ik vergeet nooit meer dat het de traditionele 'stuntdag' was van de examenleerlingen, dus er volgde een hilarische slalom tussen de met linten afgezette gangen en schuimbaden door. Joke vertelde ook nog eens dat ik door mijn 2 havo klas was genomineerd voor de leraar van het jaar verkiezing van het Rijksmuseum. Misschien ben ik vanaf dat moment pas echt gaan geloven dat ik het combineren van promoveren en lesgeven aan zou kunnen.

UU

Wie bruggen bouwt, moet ze ook over durven te steken. Mede door de leuke collega's aan de Universiteit heb ik met zoveel plezier gewerkt aan dit proefschrift. Als je dan het laatste jaar van je PhD plotsklaps vanuit huis moet werken, maakt dat het afscheid nemen van hen alleen nog maar moeilijker. Om te beginnen van mijn mede promo.docx^(© Casper Hulshof). Jullie weten als geen ander hoe het voelt om voor de klas te staan en te promoveren. Bij de onregelmatig gehouden intervisiesessies vond ik altijd wel een fijn luisterend oor.

Ik dank alle collega's van LV3 die mij toch het gevoel gaven 'dat ik bij de club hoorde', ondanks dat ik niet altijd overal bij kon zijn. Jos en Caroline, bedankt dat ik altijd even bij jullie binnen kon lopen met technische of andersoortige vragen. Dank ook aan iedereen uit de pubgroup sessies van SIG1 waar ik zoveel heb geleerd.

Bjorn, bedankt voor alle keren dat je mijn kamer binnenkwam en me met één rake opmerking weer met beide benen op de grond zette. Met een relativerende grap en bulderende lach liet je mij vaak weer vragend achter wat nu eigenlijk het hele nut van mijn onderzoek nou toch was. Met gevleugelde uitspraken als "publicatie is alleen maar zielenrust" kon je mij gelijktijdig geruststellen en doen afvragen waar ik het eigenlijk allemaal voor deed. Maar, begrijp me niet verkeerd, ik kijk hier echt met heel veel plezier op terug. Ook bedank ik je voor de nodige en nuttige afleiding als we samen gingen sparren projecten als virtual reality en gevoelige geschiedenis bespreken in de klas. Het leidde zelfs tot een fantastisch en leerzaam weekend op het KNIR in Rome waar ik na enkele jaren weer even terug mocht keren als onderzoeker.

Saro, ik kijk met plezier terug op alle mooi dingen die we hebben meegemaakt, zoals de onderzoektripjes naar Antwerpen en Rome en alle interessante nevenprojecten. Het is altijd fijn om met je te sparren, of het nou over ons onderzoek, schoolzaken, of het leven gaat. Zo jammer dat we geen koffie-en-bijpraat momentjes meer hadden het laatste jaar, maar gelukkig hebben we ze vervangen door bel-en-loop-sessies.

Renée, het was heel fijn om met je samen te werken en ervaringen te delen. Ik was altijd onder de indruk van hoe snel en consciëntieus jij je lastige theorie of onderzoeksmethodes eigen maakte. Hoewel ik vaak grapte dat ik alleen maar jouw ideeën en werk hoefde te kopiëren door MOOC te veranderen door flipped learning, was het altijd een geruststelling dat mijn onderzoek zich in een vergelijkbare richting bewoog als iemand die hele hoge eisen stelt aan eigen onderzoek. Heel leuk dat we zo af en toe nog eens bijpraten nu we geen collega's meer zijn. Misschien gaat dat gezamenlijke artikel er ooit nog eens van komen...

F3.01

I will switch to English as I would like to thank all my colleagues from room F3.01. I look back with great pleasure to all the nice conversations and support in the room and the traditional Fruit Breaks at 15:00hrs. We created a lot of nice memories with yearly trips to the Efteling, the museum-and-exploring-the-Netherlands-groups, dinners, trips to conferences in Germany, and a beautiful holiday in the Ardennes. We shared great wisdoms about our lives as researchers, such as "It's all about the money", "No money, no shit", and "No audience, no fame". As a lot of people came in and left F3.01, it would take too many pages to mention all the wonderful colleagues personally. However, I would like to thank a few colleagues with whom I was often working on the same days. My fellow Promodocs in the room Anne, Mare, Marloes, and Sophie for being part of this awesome experience from the beginning. Jonne and Katrijn for all the fun and your unmatched dance moves. Xiaojing for always being so kind and helpful. Christa, thanks for all the last minute submission tips and sharing the last moments of stress together. Anouk, Sophia, Michaela and YuanYuan, thanks for brighten up my last years as 'newcomers' in the room; I wish you all the best in your own further PhD journeys. Mei, I treasure our close bond that grew in the past years. We shared so many happy memories and saying goodbye for your return to China was very difficult. It was an honor to be part of so many of your mean jokes, but most of all to be your paranimph. Sadly, the reunion trip to China with Anne and Marloes was cancelled due to an unforeseen pandemic. I wish we can still visit you and Xiaojing in the near future...

Begeleidingsteam

Zonder mijn begeleidingsteam was het me nooit gelukt om zonder grote kleerscheuren de eindstreep van dit promotietraject te halen. Het is voor promovendi enorm belangrijk om regelmatig contactmomenten te hebben met je team, en al helemaal als je weinig achtergrond in een vakgebied hebt. Ik prijs mij dan ook gelukkig met jullie als begeleiders en de regelmaat waarmee jullie tijd voor mij vrijmaakten in jullie drukke agenda's.

Chris, vanaf ons eerste gesprek stelde je je open, gaf je mij het gevoel dat ik alles met je kon bespreken wanneer het ook maar nodig was. Jij was altijd waakzaam hoe het met me ging door na de vergaderingen nog even te blijven kletsen, lette er op dat ik mezelf niet voorbij liep, en gaf me altijd het gevoel dat ik het kon en dat het goed zou komen. En hoewel je een nieuwe positie binnen de UU ging bekleden, ben ik heel blij dat je onderdeel wilde blijven van het team. Ik had me geen betere mentor kunnen wensen. En hoe toevallig is het dat ik enkele jaren later "de Chris van het Cals College" (© Liesbeth) zou worden? Daarnaast heb je een bijzonder voorspellende gave. Toen ik ergens in mei 2019 aan je vroeg of ik het allemaal wel op tijd zou halen, weet ik nog

dat je zei: “Ach David, je hebt nu al je data binnen, dan hoef je je dat laatste half jaar alleen maar op de sluiten in huis en dan kun je lekker doorschrijven.” Wie had toen gedacht dat dit door de COVID-19 uitbraak ook letterlijk zo de laatste maanden van mijn PhD zouden zijn...

Jeroen, je hebt er een hele grote bijdrage aan geleverd dat ik mij nu na vijf jaar echt een onderwijskundig onderzoeker durf te noemen. Zeker nadat ik jouw statistiekcolleges had gevolgd voelde ik mij nog wel eens bezwaard om vrij basale statistiekwedsties met je te bespreken. Maar je stond daar altijd voor open en hielp me altijd weer op de juiste weg. Waar ik in het begin van het grote meta-analyse avontuur al trots was op het zelf uitrekenen van de cohen's κ , waren we maanden later bezig om enorm ingewikkelde meta-regressie poweranalyses te repliceren in Excel. De belangrijkste statistiek les gaf je me op 3 september 2019, toen je me de volgende wijsheid meegaf: “Tja, statistiek houdt helaas geen rekening met gevoelens.” Tijdens ons wekelijks overlegmoment zorgde je er altijd voor dat ik weer verder kon. Ik vond het een eer om met je samen te presenteren op de ORD en de ResearchED.

Liesbeth, bedankt dat je me altijd het vertrouwen gaf die ik nodig had om er in te geloven dat ik het kon. Jouw idee om met Jeroen en Chris een team om mij heen te vormen heeft in mijn ogen fantastisch uitgepakt. Tijdens ons eerste B&O gesprek zei je: “Denk niet dat het normaal is, dat je je zo snel in een compleet ander vakgebied inwerkt en je het je eigen maakt.” Eigenlijk ben ik er vanaf dat moment pas écht in gaan geloven dat ik het wel zou gaan redden. Ik was in mijn begeleiding ook heel erg op zoek naar kritische feedback als ik iets had bedacht of opgeschreven wat niet volgens de basisregels in de onderwijskunde was en ik kon daarin altijd rekenen op jouw eerlijkheid. Ook wist je op de juiste momenten het project de goede richting in te helpen bij de ‘knopen-doorhak-sessies.’ En toen we daar even een keer minder goed in slaagden, belde je mij een week later heel attent op om te vragen wat er nog nodig was om weer verder te kunnen. En natuurlijk zal ik de avontuurlijke reis in het busje naar Bad Schussenried nooit vergeten...

Cals College

Promodoc zijn werkt alleen als de school waarop je werkt je daarin voor de volle 100% steunt. Harrie, Joke, Guus, Nico, en Gabriëlle, bedankt voor de steun en betrokkenheid die ik bij jullie als mijn leidinggevendenden altijd gevoeld heb. Jullie losten onhandige roosterzaken voor mij op en gaven mij autonomie en vertrouwen om promotie- en schoolzaken te combineren. Ook dank ik Peter en Robert voor de kans die ze me hebben geboden door mij aan te nemen als PromoDoc, in een tijd dat er een groot tekort is aan docenten behalve bij het vak Geschiedenis. Bij de vaksectie vond ik al vrij snel een diverse groep van leuke en inspirerende vakidioten waar ik veel van heb

geleerd. En hoe fijn was het dat Rivka, Vera, Lana, Rutger, en Jan-Willem met hun klassen mee wilden doen met mijn onderzoeken.

Ik vond op school de nodige afleiding van alle promotiezaken en een groep fijne collega's om mee samen te werken. Bijvoorbeeld de ICT-werkgroep collega's, de Volleybal-op-donderdagdag-groep, en natuurlijk ook de leden van het *Pentagoon*, en iedereen van Cals Centraal met wie ik met veel plezier samenwerk om alle ambities van het Cals College te verwezenlijken. De vele vergaderingen (en ja ook de lachmomentjes, Henk) zijn een fijne afleiding gebleken op de momenten dat ik weer even vast zat met het schrijven van mijn proefschrift.

Paul, we begonnen samen op het Cals en ik koester de vriendschap die is ontstaan. Ik waardeer je interesse die je toonde voor mijn onderzoek en je immers (zelf)kritische houding waarmee je jezelf en anderen inspireert om het beste in onszelf naar boven te halen. In het gamen en sporten vonden we gedeelde hobby's wat fijn dat we elkaar nog steeds blijven zien en spreken!

Rivka, jij hebt de fases die ik doorlopen heb in dit proefschrift van zoveel kanten meegemaakt. Als deelnemende docent, als geïnteresseerde in het publiek van mijn workshops, als dierbare vriend en toehoorder tijdens onze talrijke fietsbelletjes en lunchloopjes, maar bovenal als paranimf. We vinden elkaar voor een luisterend oor, de beste adviezen, voor het beste vermaak, en natuurlijk in CHBL. Door de hechte band die is ontstaan ken jij mij als geen ander. Op het moment van de verdediging weet ik me gesteund met jou als paranimf.

Leven naast werk?

Er werd mij nog wel eens gevraagd of er nog wel een leven naast het werk van een PromoDoc mogelijk was. Gelukkig had ik veel lieve mensen om mij heen die er voor zorgden dat ik voldoende kon ontspannen en opladen om doordeweeks weer vol aan het werk te kunnen. Zoals Ilona, Jarno, Luuk, en Diandra en de bijzondere vriendschap die we al sinds het Emmauscollege met elkaar delen. Heel fijn dat we elkaar nog zo (on) regelmatig zien, op plaatsen van Maastricht tot Friesland. De zaalvoetballers van Kok Makelaars, jammer dat het er tegenwoordig niet meer in zit mannen. Joris, bedankt voor de wijze levenslessen die je met je *padawan* wilde delen. Ewoud, Luke, en Tristan met wie ik vele game & chill uurtjes doorbracht en de leuke tripjes die we samen maken. Jennifer, bedankt voor je steun en de ruimte die je me gaf om mijn ambities te verwezenlijken. De leuke collega's van Ticketmaster met wie ik tijdens mijn promotie nog altijd de nodige concerten en evenementen beleefde. De grote groep badmintonners van BC Schollevaar waarmee ik samen op de baan mijn energie eruit kon slaan, en ook met een hele fijne groep vrienden buiten de baan van het leven mag genieten.

Ik dank natuurlijk ook mijn familie. Voor de gekke humor die we delen met elkaar, en het op elkaar kunnen rekenen op de momenten dat het echt nodig is. Oma's Truus en Tiny die altijd zo graag wilden weten hoe het er voor stond met het promoveren. Nu is het eindelijk echt af! Joke en Ria voor alle leuke avonden met nieuwe invalshoeken en fanatieke spelletjes. En in het bijzonder natuurlijk mijn liefdevolle gezin. Met zijn vieren delen we lief en leed. Hanno, Tike, Hester, jullie staan altijd voor me klaar en dat veilige vertrouwde gevoel is het waardevolste dat er bestaat. Hoewel ik jullie nooit de oren van jullie hoofd heb afgekletst over mijn werk, maakte deze fijne hechte basis in mijn leven het wel mogelijk om dit avontuur aan te gaan en vol te houden. En natuurlijk ons bijzondere gevoel voor humor en zelfspot. Hester, onze broer-zusband is heel sterk en dat koester ik. Of we nu samen op vakantie gaan of hele jeugdfilms uit ons hoofd na al die jaren nog kunnen meepraten. Ik ben heel blij dat jij me bijstaat als paranimf bij de laatste stap richting de finish. Dat we maar altijd blijven lachen, wat er ook gebeurt, want na regen komt zonneschijn!

Roos, ik ben heel blij met je liefde, je steun, ons samenzijn. Waar ik altijd enorm heb opgekeken tegen de laatste fase van het promoveren, ben ik dat mede dankzij jou zonder kleerscheuren doorgekomen. Ik hoop dat het ons nog vele jaren gegund is samen te zijn.

Aan allen die tot het einde van dit lange dankwoord gekomen zijn: proficiat! Ik steek nu weer een brug over naar het volgende avontuur. Als ik over mijn schouder terugkijk doe ik dat met veel dankbaarheid en weemoed. Naar voren kijk ik met veel enthousiasme naar wat er nog mag komen, klaar om weer nieuwe bruggen te bouwen.