A Prediction-Based Framework to Reduce Procrastination in Adaptive Learning Systems

Inaugural Dissertation

submitted to the Faculty of Human Sciences of the University of Bern for the Degree of

Doctor of Philosophy

Christof Imhof

Binn, VS

Bern, March 2022

Under the supervision of

Prof. Dr. Fred W. Mast

Prof. Dr. Per B. Bergamin

Accepted by the Faculty of Human Sciences of the University of Bern

in the request of Prof. Dr. Fred W. Mast

and Prof. Dr. Per B. Bergamin

Bern, The dean: Prof. Dr. Stefan Troche

This dissertation comprises the following publications and manuscripts:

Study 1

Imhof, C., Bergamin, P. B., Moser, I., & Holthaus, M. (2018). Implementation of an adaptive instructional design for a physics module in a learning management system. In *Proceedings of the 15th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2018)*.

Study 2

Imhof, C., Bergamin, P., & McGarrity, S. (2020). Implementation of Adaptive Learning Systems: Current State and Potential. In *Online teaching and learning in higher education* (pp. 93-115). Springer, Cham. https://doi.org/10.1007/978-3-030-48190-2_6

Study 3

Imhof, C., Bergamin, P., & McGarrity, S. (2021). Prediction of dilatory behaviour in online assignments. *Learning and Individual Differences*, *88*. https://doi.org/10.1016/j.lindif.2021.102014

Study 4

Imhof, C., Comsa, I.-S., Hlosta, M., Parsaeifard, B., Moser, I., & Bergamin, P. B. (2022). *Prediction of Dilatory Behavior in eLearning: A Comparison of Multiple Machine Learning Models*. arXiv preprint arXiv:2206.15079

Abstract

Procrastination and other types of dilatory behaviour are common in online learning, especially in higher education. While procrastination is associated with worse performance and discomfort, positive forms of delay can be used as a deliberate strategy without any such consequences. Although dilatory behaviour has received attention in research, it has to my knowledge never been included as an integral part of an adaptive learning system. Differentiating between different types of delay within such a system would allow for tailored interventions to be provided in the future without alienating students who use delay as a successful strategy. In this thesis, I present four studies that provide the basis for such an endeavour. I first discuss the results of two studies that focussed on the prediction of the extent of dilatory behaviour in online assignments. The results of both studies revealed an advantage of objective predictors based on log data over subjective variables based on questionnaires. The predictive performance slightly improved when both sets of predictors were combined. In one of these studies, we implemented Bayesian multilevel models while the other aimed at comparing various machine learning algorithms to determine the best candidates for a future inclusion in real-time predictive models. The results reveal that the most suitable algorithm depended on the type of predictor, implying that multiple models should be implemented in the field, rather than selecting just one. I then present a framework for an adaptive learning system based on the other two studies, where I highlight how dilatory behaviour can be incorporated into such a system, in light of the previously discussed results. I conclude this thesis by providing an outlook into the necessary next steps before an adaptive learning system focussing on delay can be established.

Keywords: procrastination, strategic delay, adaptive learning, log data, predictions, machine learning

Table of Contents

Table of Figures

1 Introduction

Ever since its inception, the Internet has continued to expand its reach into almost every aspect of life, including learning. The ongoing digitalisation has not only given rise to forms of learning such as distance learning or technology-based learning but has also delivered the necessary tools such as learning management systems to enhance pre-existing approaches including adaptive, personalised, self-regulated, and self-directed learning. These approaches can now benefit from the collection and subsequent analysis of learner- and context-specific data, allowing for learning to be optimised in real-time; a concept better known as *learning analytics* (see Mavroudi, Giannakos, & Krogstie, 2018; Ferguson, 2012).

Besides these new possibilities, the advent of online learning has brought with it an illustrious array of other advantages, including but not limited to: access to higher education for groups that previously did not have any due to temporal or spatial constraints (e.g., full-time employees, parents, students living in remote locations), efficient and convenient ways to achieve learning goals (Chen, Lambert, & Guidry, 2010), and higher levels of student engagement (see Dumford & Miller, 2018 for an overview). However, it also casts a shadow and within that shadow, something sinister lurks that most of us are already more than familiar with: procrastination. Despite not appearing to be threatening at first glance, it is widespread, persistent, and associated with various negative consequences, including increased risk of drop-out and feelings of guilt (see Höcker, Engberding, & Rist, 2022).

The overall aim of this thesis is to present a framework of how to address procrastination and other types of dilatory behaviour within an adaptive learning system, which could provide tailored interventions to reduce procrastination while not affecting positive types of dilatory behaviour. The goal of the first study, published as a conference paper titled *Implementation of* *an Adaptive Instructional Design for a Physics Module in a Learning Management System* (Imhof, Bergamin, Moser, & Holthaus, 2018), was to demonstrate how an adaptive instructional design, in this case rule-based task recommendations embedded in feedback based on students' prior knowledge and current performance, can be implemented in a learning management system (LMS). This paper then served as the basis for a book chapter called *Implementation of Adaptive Learning Systems: Current State and Potential* (Imhof, Bergamin, & McGarrity, 2020), which marks the second entry for this thesis. The chapter details the basic mechanisms of adaptive learning, what the current research has revealed about its effectiveness, and where the future potential lies. In the third and fourth studies, titled *Prediction of Dilatory Behaviour in Online Assignments* (Imhof, Bergamin, & McGarrity, 2021) and *Prediction of Dilatory Behaviour in eLearning: A Comparison of Multiple Machine Learning Models* (Imhof et al., 2022) respectively, me and my co-authors investigated the prediction of dilatory behaviour when handing in assignments in an LMS, focussing on a comparison between questionnaire- and log data - based predictors. The underlying data was the same for both studies, what differed was the prediction approach behind it (Bayesian multilevel models in study 3, expanded by multiple Machine Learning algorithms in study 4). For the remainder of the thesis, these four studies will be referred to as studies 1, 2, 3, and 4, respectively.

Within the framework, study 2 provides the theoretical basis for the proposed adaptive learning system. Studies 3 and 4 present solutions for how to predict the extent of dilatory behaviour, which is crucial since it allows for interventions to be made, be it by lecturers or teachers or by an adaptive system. In future, such an adaptive learning system could implement machine learning decision-making, providing adaptive instructional feedback akin to study 1, yet in a more sophisticated manner and tailored specifically towards maladaptive forms of dilatory behaviour based on dilatory profiles.

I will first give an overview over the basics of procrastination and other types of dilatory behaviour, followed by a discussion about how it can be measured, where the contributions of studies 3 and 4 will also be highlighted. In the next chapter, I will then present a framework of how the models detailed in studies 3 and 4 could be incorporated in an adaptive learning system based on the contributions of all four studies, followed by a discussion about what implications for interventions this framework offers, before providing a conclusion.

2 Procrastination and other types of dilatory behaviour in online learning

In this chapter, I present an overview over conceptualisations of procrastination in general, followed by a definition of academic procrastination, the reasons behind it, the consequences thereof, and how it differs from other types of dilatory behaviour.

Based on its literal meaning, derived from the Latin expression *procrastinus*, composed of the prefix *pro*, meaning *forward*, and *crastinus*, meaning *of tomorrow*, one could assume that procrastination simply refers to moving something from one day to the other, or in other words, the act of delaying something. This is exemplified by Alcoholics Anonymous founder Bill Wilson, who once wrote that procrastination means “really sloth in five syllables”. However, this does not capture the true essence of procrastination. After all, attributing laziness to a student cleaning their entire flat top to bottom just to avoid working on their essay seems misguided. What is missing from Wilson’s interpretation is highlighted in many definitions found in the literature: some focus on the inability to convert an intention into an action (Blunt & Pychyl, 2005), others stress the delay of specific parts of a task (e.g., the beginning or the end, Lavoie & Pychyl, 2001; or the aversive nature of the task, e.g., Alexander & Onwuegbuzie, 2007), or focus on the lack of motivation (Senécal, Koestner, & Vallerand, 1995) or self-regulation (Rebetez, Rochat, Barsics, & Van der Linden, 2016; Tuckman & Sexton, 1989). Similarly, Sirois and Pychyl (2013) define procrastination as a "self-regulatory failure of not exerting self-control necessary for task engagement" (p. 116). Steel (2007) adds another element in his often-cited definition, namely the irrational and dysfunctional nature of procrastination. He thus defines it as a "voluntary delay of an intended course of action despite expecting to be worse off for the delay (p. 66)."

Importantly, this awareness highlights that procrastination is not caused by intellectual differences between procrastinators and non-procrastinators (see Ferrari, 1991), but rather differences in the level of self-regulation, volition, and organisational skills (Rabin, Fogel, & Nutter-Upham, 2011). Procrastinators have trouble setting goals (Lee, 2005), correctly estimating how long it takes to reach, prioritise, and follow through with them, even if they were set (Ferrari, 2001; Howell & Watson, 2007), which are all time and effort management issues (see Hailikari, Katajavuori, & Asikainen, 2021). When viewed from a self-regulation perspective, a lack of inhibition capacities may play a large role, since they are necessary for the regulation of thoughts, emotions, impulses, and behaviour (Tice, Bratslavsky, & Baumeister, 2001), which could explain the low conscientiousness, high impulsivity, thought control issues, and mood regulation deficits that are often associated with procrastination (see Dewitte & Schouwenburg, 2002; Gustavson, Miyake, Hewitt, & Friedman, 2014; Harriott, Ferrari, & Dovidio, 1996; Rebetez, Rochat, & Van der Linden, 2015; Sirois & Pychyl, 2013). Another approach, the Temporal Motivational Theory (Steel & König, 2006), conceptualises procrastination as the absence of positive reinforcers (i.e., a task lacking personal value), low self-efficacy (i.e., confidence in one's ability to complete tasks), unclear or distant deadlines, and an inability to delay gratification. Links between procrastination and various facets of motivation have also been reported (e.g., intrinsic motivation, extrinsic motivation, and a lack of motivation, Rakes & Dunn, 2010; Grunschel et al., 2013a; Lee, 2005; Rebetez et al., 2015; Tice & Baumeister, 1997). Other authors conceptualise procrastination as a failure-avoidance strategy (e.g., Helmke & Aken, 1995).

For the purposes of this thesis, I draw from multiple of the definitions above and thus view procrastination as an irrational, voluntary delay of important tasks characterised by a lack of self-regulation and time management skills.

2.1 Academic procrastination

While not being exclusive to online learning - after all, procrastination can be encountered in basically any given circumstance in life that involve tasks that need to be completed eventually, such as housekeeping, making appointments, paying bills, oiling that squeaky door, answering that important e-mail at work, or doing taxes - *academic procrastination* is thought to be a particularly prevalent type of procrastination, as the reported numbers suggest. While the prevalence rates are around 10-25% in the general population (Ferrari, Díaz-Morales, O'Callaghan, Díaz, & Argumedo, 2007; Svartdal, 2017), the self-reported numbers among university students are as high as 70% (Rahimi & Hall, 2021; Schouwenburg, Lay, Pychyl, & Ferrari, 2004). Students in distance education are particularly susceptible to procrastination, due to the higher demands this autonomous learning environment puts on students' persistence, responsibility, attention, self-efficacy, self-regulation, and self-directedness (e.g., Deimann & Bastiaens, 2010; You, 2015). In this context, self-regulation refers to efforts to monitor, manipulate and improve one's own learning, e.g., by implementing strategies (Corno & Mandinach, 1983). Learning in a self-regulated manner is a demanding task for many students, especially since it not only requires knowledge about how to learn, but also actual implementation of learning strategies, be they cognitive, metacognitive, motivational, or behavioural (see e.g., Zimmerman, 2008). Studies have shown that procrastinating is positively correlated with reduced use of cognitive and metacognitive learning strategies (Howell & Watson, 2007; Klingsieck, Fries, Horz, & Hofer, 2012; Wolters, 2003) and that self-regulation strategies serve as a mediator of the effects of procrastination on learning behaviours, meaning successful implementations of such strategies are able to counter personal tendencies to procrastinate (see Pintrich, 2004).

One self-regulation strategy that is particularly relevant in online learning is effort regulation (also known as volition, see Corno, 1993; Pintrich, 2000). It refers to a learner's ability to control their attention and resist distractions, which is a challenge, given the saliency of distractors in online learning. While distractions such as daydreaming have been an issue before (see Harriott et al., 1996; Pychyl, Lee, Thibodeau, & Blunt, 2000), online learning has pried Pandora's Box of distractions wide open, as every distraction imaginable pines for students' limited attention. Media multitasking has become particularly tempting (Dontre, 2021) and students may find it increasingly difficult to stay on task without getting distracted due to a growing use of personal digital devices (Gazzaley & Rosen, 2016). All it takes is a simple mouse click or touchscreen tap. Checking Facebook (Meier, Reinecke, & Meltzer, 2016), watching cat videos (Myrick, 2015), or being on the Internet in general (Thatcher, Wretschko, & Fridjhon, 2008) have thus become frequent sources of distraction from typical tasks in online learning such as studying learning materials, participating in single or group activities, or completing online assignments.

2.2 Reasons for procrastination

The reasons for procrastination suggested by the literature are linked to its postulated nature: approaches that view procrastination as a personality trait focus on its relation to other personality traits, with mixed results (e.g., extraversion, conscientiousness, and external locus of control, Senécal, Lavoie, & Koestner, 1997; Steel, Brothen, & Wambach, 2001; the Dark Triad, Lyons & Rice, 2014; responsibility, Çelikkaleli & Akbay, 2013; perfectionism, Renn, Allen, Fedor, & Davis, 2005), whereas approaches that stress the situational factors underlying procrastination note that it is more likely to occur when the task at hand is perceived as aversive (Blunt & Pychyl, 2005). This aversiveness can take many forms: a task can be perceived as boring, too difficult or challenging, anxiety-inducing (fear of failure, see Renn et al., 2005; Abdi Zarrin, Gracia, & Paixão, 2020; Rahimi & Hall, 2021), or not granting the student enough control or autonomy, leading to it being avoided (Senécal et al., 1997). This is in line with self-determination theory, according to which psychological needs have to be fulfilled as a basis for autonomous motivation (Deci & Ryan, 2008). Procrastination could also be viewed as a self-handicapping (or self-defeating) strategy: by waiting until the last possible minute to start with or finish a task, a procrastinator could then attribute the poor outcome, should it occur, to their low effort, rather than low ability, thus protecting their self-worth (Garcia & Pintrich, 1994). Psychological reactance, meaning the protection of freedom of choice by favouring suddenly unavailable options over the remaining alternatives, has also been linked to procrastination and could serve as an additional reason (Malatincová, 2015).

2.3 Consequences of procrastination

Regardless of the underlying reasons, procrastination is traditionally associated with a list of negative consequences, which includes unsatisfactory academic performance (Klassen, Krawchuk, & Rajani, 2008), higher levels of stress and anxiety, (Ferrari, O’Callaghan, & Newbegin, 2005; Sirois, 2004), higher number of task-related errors (Ferrari, 2001), reduced mood, and feelings of guilt (Pychyl et al., 2000). As You (2015) notes, this is particularly damaging in online learning settings, since procrastination affects drop-out rates (Doherty, 2006), and there is evidence to suggest that its impact on achievement could be more severe compared to traditional classroom settings (see Tuckman, 2005). Similarly, Yilmaz (2017) found a stronger negative link between procrastination and assignment scores in distance learning compared to face-to-face environments.

2.4 Strategic delay and other positive forms of dilatory behaviour

Nonetheless, a rather recent strand of research has noted that not all acts of delay are necessarily ill-fated and that in fact, a functional counterpart to procrastination may exist. Initially coined *active procrastination* by Chu and Choi (2005) to differentiate the concept from the regular, "passive" version, other terms have since been introduced to avoid a linguistic conundrum that Pychyl (2008) pointed out. He argued that procrastination is negative and maladaptive by definition; and a positive form would thus be an oxymoron. These other terms include *active delay*, *purposeful delay* (Corkin, Yu, & Lindt, 2011; Grunschel et al., 2013a), and *strategic delay* (Klingsieck, 2013). Despite referring to the term *purposeful delay* in studies 3 and 4, I will instead use the term *strategic delay* during the remainder of this thesis, the reason for which will be addressed in chapter 4.

Klingsieck (2013) identifies four aspects that procrastination and its functional counterpart have in common and three in which they differ, based on constituent parts of their definitions found throughout the literature. The shared aspects are the involvement of an act of delay, an intention to start or to complete a task, the necessity or personal importance of the task, and the voluntary nature of the delay. Voluntary in this sense means it is not imposed on the individual purely by external factors, e.g., a student missing a train, preventing them from handing in an assignment on time. The three aspects that are unique to procrastination are the irrationality of the delay, delaying a task despite being aware of the potential consequences, and the subjective discomfort (or other negative consequences) it entails. Aspects that are unique to strategic delay include a preference to work under pressure to be properly motivated and the ability to meet deadlines, which along with a higher level of satisfaction and the intentionality of the decision mark the four dimensions of a strategic delay - specific questionnaire by Choi and Moran (2009) (even though it uses the original term *active procrastination*). Another key difference between procrastination and strategic delay is the aforementioned level of self-efficacy. Self-efficacy, particularly in the context of self-regulated learning, is thought to be crucial since the confidence in using appropriate strategies is a key factor when beginning and completing tasks (Klassen et al., 2008). Wäschle, Allgaier, Lachner, Fink, and Nückles (2014) for example found that self-efficacy is at the centre of both a vicious and a virtuous circle. Students that experience low self-efficacy and/or perceive tasks as aversive are less likely to engage with them, tend to then delay their completion and thus probably fail to achieve their goals, which might cause feelings of disappointment and an increased likelihood of avoiding similar tasks in future. This vicious circle could be counteracted by higher self-efficacy: perceptions of high self-efficacy positively affect the motivation to engage with a task, which increases the likelihood of implementing learning strategies, resulting in higher chances of success and the achievement of goals, in turn increasing self-efficacy for the next task.

3 Measurement of dilatory behaviour

As many authors note (e.g., Wäschle et al., 2014; Malatincová, 2015; Stainton, Lay, & Flett, 2000), a distinction should be made between trait and state procrastination since actual dilatory actions are influenced by personal tendencies (i.e., personality traits), situational factors, and self-regulation strategies (see Steel, 2007; Pintrich, 2004). This is evidenced by a medium to high correlation between trait and state of .51 (Stainton et al., 2000), implying that state and trait are related, but not identical. In general, there is little consensus whether procrastination is primarily a personality trait or the result of situational factors (see Thatcher et al., 2008). Despite a meta-analysis by Steel (2007) showing that procrastination is stable across time and across different situations, favouring a trait perspective, there is also empirical evidence to highlight the role of contextual factors, e.g., task characteristics (Blunt & Pychyl, 2005). Another reason not to conflate trait and state procrastination is provided by Malatincová (2015), who proposes that procrastination may actually be a multi-dimensional construct or even a set of constructs. Subjective experience and actual delay should thus clearly be distinguished from one another, since they might relate to different underlying, potentially opposed psychological variables. Anxiety for instance may increase self-reported questionnaire scores while reducing actual delay at the same time. Conflating questionnaire scores and temporal indicators of delay would thus be misleading.

Defining procrastination is not the only point of contention in the research on the topic, how to measure it has also been subject to debate (Svartdal et al., 2016). By far the most common approach for measuring trait procrastination has been the implementation of questionnaires, usually self-report instruments with varied theoretical underpinnings (see Steel, 2010 or Rozental et al., 2014, for an overview). In contrast, dilatory behaviour in online learning scenarios is commonly assessed with log data collected from LMS, e.g., Levy and Ramim (2012), who measured the difference between the due times and submission times of tasks, del Puerto Paule-Ruiz, Riestra-González, Sánchez-Santillán, and Pérez-Pérez (2015) and Cerezo, Esteban, Sánchez-Santillán, and Núñez (2017), who both used a variety of log data variables to form association rules, Akram et al. (2019), Yang et al. (2020b), and Yang et al. (2020a), who all used homework submission data to classify students as procrastinators or non-procrastinators, or Abdi

Zarrin et al. (2020), Cao (2012a; 2012b), and Motie, Heidari, and Sadeghi (2012), who all used trait procrastination as the outcome variable.

3.1 Prediction of dilatory behaviour

Dilatory behaviour has thus served as a predictor in multiple studies, yet it has rarely been the outcome variable itself. While the effect that dilatory behaviour exerts on achievement or subjective well-being is important, predicting whether a student is going to delay a task in the first place - and by how much - is also valuable information. It allows lecturers or teachers to intervene in time or gives an adaptive system the information it needs to provide instructional guidance. These potential benefits make the lack of research in this area all the more surprising. A notable exception is a study by Hlosta, Zdrahal, and Zendulka (2018), where a predictive approach that did not rely on data from previous courses (“Self-Learning”) was implemented to predict whether students were going to meet the first submission deadline. However, the focus of the study lied on dealing with the underlying issue of unbalanced data, meaning an uneven distribution of observations between levels of a class (in this case delayed and punctually-submitted tasks). Moreover, no theoretical links between the predictors and the outcome were established. Another exception is a study by Zuber et al. (2020), who successfully predicted dilatory behaviour with the Pure Procrastination Scale (PPS) and its subdimensions but included no other predictors in their regression models.

The goal of study 3 (Imhof et al., 2021) was thus to address this research gap by calculating models to determine the best predictors of delay and to compare objective and subjective predictors. We differentiated between trait and state delay by measuring dilatory behaviour - the outcome variable - based on log data and trait procrastination as well as trait strategic delay - both serving as predictors - with one questionnaire each. In total, we used seven predictors, four subjective ones based on questionnaire scores and three objective indicators derived from log data from our LMS (i.e., Moodle) and other internal systems.

The first objective predictor was a temporal variable, namely the interval between the start of a block (i.e. learning unit) and the first click on an assignment, indicating how long it takes until a student first reads the task description. Our courses are structured in consecutive thematic blocks, which is why the official start of the block the assignments were located in was used as the starting point for this calculation. The other two objective predictors were click-based activity indicators, similar to the variable used in study 1. The sum of clicks on what we deemed learning-relevant course activities (forums, learning videos, Moodle books, wikis, glossaries etc.) indicated engagement with learning materials across the course. The other click-based measure, the sum of all clicks on the assignment itself, indicated how often students re-read the instructions.

The subjective predictors were derived from the literature outlined in chapter 2. Trait procrastination and trait strategic delay were selected due to the expected links between them and dilatory behaviour, based on the reported correlation between trait and state (Stainton et al., 2000). We chose a self-efficacy score as the third indicator due to the role self-efficacy plays in dilatory behaviour, having strong ties to motivation in general, self-regulation, self-directedness, appropriate goal-setting, and the successful implementation of effective learning strategies (see e.g., Wäschle et al., 2014; Klassen et al., 2008). Moreover, it is a relevant factor in multiple types of delay and may help differentiate between them. The fourth and final questionnaire-based indicator was a self-directed learning score. As highlighted in chapter 2, self-regulation is a key concept in procrastination, even being at the core of some of the conceptualisations (e.g., the one by Sirois & Pychyl, 2013 who view procrastination as a self-regulation failure), which is why a related measure is crucial to include. Our choice was between a measure of self-directed learning (SDL) or a scale assessing of self-regulated learning (SRL), which are similar concepts that often get used interchangeably. While there are differences between these two concepts (see Saks & Leijen, 2014; Linkous, 2021), they are both strongly linked to self-efficacy, procrastination, and strategic delay (see Sundaramoorthy, 2018; Corkin et al., 2011; Grunschel, Patrzek, Klingsieck, & Fries, 2018; Onji & Kikuchi, 2011; Saeid & Eslaminejad, 2017; Schommer-Aikins & Easter, 2018; Wolters, 2003). Since we collaborated with the Research Unit Self-Directed Learning at the North-West University in South Africa for a follow-up project, we settled for an instrument assessing SDL, which was also advantageous due to its brevity (10 items).

In general, we decided to implement short questionnaires whenever possible in order not to alienate the participants, who were busy students in higher education, and usually fully-employed as well. We then calculated multiple Bayesian multilevel models to determine which individual predictors worked best, which category of predictors would perform better, and whether a combination of the two categories would yield more favourable results.

3.2 Empirical findings on the prediction of dilatory behaviour

Our results in study 3 demonstrate on the one hand that dilatory behaviour can indeed be predicted, and on the other hand, that objective factors (log data) are far more successful in doing so than subjective predictors (questionnaire scores). Despite the close links between self-efficacy, self-directed learning, trait procrastination and trait strategic delay in the literature, the respective questionnaire scores were mostly negligible as predictors. The missing effect of self-efficacy lines up with Cao (2012a) and a recent study by Hailikari et al. (2021), who also did not find the expected link between self-efficacy and delay. The weak relation between trait procrastination and dilatory behaviour was also confirmed in a recent study by Wieland, Ebner-Priemer, Limberger, and Nett (2021) that was published shortly after study 3. By using an event-based experience sampling method involving appraisals, the authors found that general procrastination tendencies did not predict task-specific delays. This contrasts with a study by Zuber et al. (2020), who found that trait procrastination did predict actual behaviour in a behavioural task (sending in a signed attendance sheet before a deadline). However, none of these studies (including study 3) implemented the same scales to assess trait procrastination, which could partially explain the contrasting results, as can the nature of the tasks. While we investigated mandatory assignments, the tasks in Wieland et al. (2021) were selected by the students themselves and the task in Zuber et al. (2020) was, while mandatory, purely administrative.

Among the log data predictors, the interval between the start of a block and the first click turned out to be the most successful, aligning with the findings of del Puerto Paule-Ruiz et al. (2015) and Cerezo et al. (2017). Surprisingly, the purely click-based log data variables produced non-intuitive effects, e.g., clicks on relevant activities being a positive predictor, which could be explained by students using other learning content as way to delay working on the actual assignment. The click-based variable in study 1 however worked as intended and had the expected effect on learning progress. The major difference was the granularity of the measure: in study 1, it served as a general measure of online activity (i.e., a sum across a longer period of time) whereas in study 3, the measures were more specific (on the course level for the clicks on course activities and on the more fine-grained assignment level for the assignment-related clicks).

Performance-wise, there was little worth in combining both subjective (trait) and objective (state) measures in study 3 as the objective models barely improved when the subjective predictors were added. On the one hand, this is hardly surprising since at least some overlap can be expected due to the moderate correlation reported in the literature, yet on the other hand, they are not supposed to measure the exact same construct as pointed out by Malatincová (2015) and can thus not serve as replacements or proxies for one another, implying a potential for additional variance to be explained. Importantly, our results do not imply that subjective variables should be discarded, as they can serve as predictors in the absence of objective variables. Besides not measuring the same construct, subjective and objective variables are not interchangeable for a practical reason: They are not equally intrusive. While log data can be extracted without the student ever noticing (giving consent at some point notwithstanding), questionnaires need to be strategically selected and distributed, as students tend to find them tedious and disruptive, if they outstay their welcome.

Our results have wider implications for measurement of delay: First of all, we discovered an effect of the type of deadline, meaning that clear deadlines were associated with less delay. Unclear deadlines are a potential issue since they may appear more distant than they are, thus reducing the sense of urgency and increasing the likelihood of dilatory behaviour (see Huang, Zhang, Burtch, Li, & Chen, 2021). There is also a limitation in that the calculation of delay requires a deadline, which is not always appropriate, either because they are not available on a technical level (meaning it cannot simply be extracted from an LMS alongside the other data), the task does not have any, or because the deadline does not exist outside of the students' mind. Self-imposed deadlines are important, yet inaccessible unless specifically asked for. Another implication is the point in time our indicators allow a prediction to be made. The issue with our log data variables is that they require a minimum of clicks in order to actually be able to make a prediction about potential delay. For instance, if the student has not clicked on an assignment yet, the interval variable cannot be calculated. This was not of much relevancy in study 3 itself since all data was extracted after the fact and our predictions were actually post-dictions. However, this would be an issue for interventions based on these variables, which cannot be provided if key variables are missing. After all, the long-term goal of this line of research is to implement prediction models into an LMS, allowing lecturers and/or systems to intervene once potential maladaptive dilatory behaviour is detected. This process will be detailed in chapter 4.

3.3 Comparison between machine learning models

Before this kind of prediction model may be implemented in a live LMS, appropriate machine learning (ML) algorithms need to be identified. The purpose of study 4 (Imhof et al., submitted) was thus to determine potential candidates by comparing the predictive performance of multiple ML algorithms, trained and tested on the data set obtained in study 3.

In study 4, we chose the Bayesian multilevel models from study 3 as our baseline and compared the predictive performance of the following ML algorithms: Naive Bayes (NB), K-Nearest neighbours (KNN), Radial Basis Function Networks (RBFN), Feed-Forward Neural Networks (FFNN), Regression Trees (RT), Gradient Boosting Machines (GBM), Random Forests (RF), and several types of Support Vector Regression (SVR). The goal was to determine the algorithms that minimised the error between the predicted and actual delay values and obtained the best classification performance of two classes (delayed and timely submission, i.e., dichotomised delay values) for each of the three types of predictors (subjective, objective, and combined). Since our data was unbalanced (66% of assignments having been submitted in time), common performance indicators such as accuracy or the Matthews Correlation Coefficient (MCC) were not easily interpretable (see Hlosta et al., 2018). For this reason, we opted to calculate a score that balances the measures for both classes in complex ML modelling, the novel *G-score*. The G-score takes the trade-off between the performance of predicting delay, the performance of predicting timely submission, and the error between the predicted and ground truth value into account. The objective of each ML algorithm was then to maximise the G-score when training and validating the solutions.

3.4 Empirical findings on the performance of machine learning models

The results show that the algorithms with the best G-scores were not always the same depending on the type of predictor selected: GBM was the best approach for subjective predictors, the Bayesian multilevel models with random slopes from study 3 for the objective predictors, and RF when all predictors were combined in a single model. This implies that choosing one single particular ML algorithm may not be the best solution when applying these models for real-time predictions.

The results also allow us to draw more conclusions about the predictor types, since the advantage of the objective predictors over the subjective variables was consistent across the various ML algorithms, thus supporting the findings from study 3. In contrast to that study however, the superiority of the combined predictors was slightly more pronounced. While the ranking was unclear before, the ML models revealed an advantage of combining all seven of the predictors in terms of predictive performance, further underlining the role that subjective predictors play despite their lower predictive performance.

4 Adaptive framework to address dilatory behaviour

Now that suitable predictor variables and appropriate models have been determined, the next step is to actually implement these models into an LMS to allow for interventions aiming at the reduction of procrastination. Adaptive learning systems provide the necessary conditions for such an endeavour, as they are able to deliver instructional countermeasures (e.g., automated feedback, warnings, or task variations) specifically tailored towards certain learner characteristics, for example the type of delayer. In the following, I will first discuss the implications of study 1, before presenting a framework of how an adaptive learning system may address dilatory behaviour based on the models we calculated in studies 3 and 4, combined with conclusions drawn from studies 1 and 2.

4.1 Basis of the framework

Study 1 did not address dilatory behaviour directly, but rather aimed at demonstrating how a simple rule-based adaptive learning system may be implemented by analysing the effects of detailed and non-detailed tasks, which were recommended based on the learners' level of knowledge. Nonetheless, the study has two important implications for the proposed framework: First off, it further highlights that log data can successfully be used to indicate activity (or lack thereof), even just in the form of simple click sums. Secondly, we encountered an effect that at least superficially resembled the Expertise Reversal Effect (Kalyuga, Ayres, Chandler, & Sweller, 2003). This effect means that interventions may actually be weaker for experts or even detrimental to their performance in comparison to novices, since the additional information may interfere with prior knowledge, increasing their cognitive load. This is relevant in the context of the thesis in two ways: On the one hand, as previous studies (e.g., Ferrari, 2001) have shown, procrastinators perform worse under conditions that require high levels of self-regulation, which is the case when cognitive load is increased. Not accounting for dilatory behaviour could thus be considered a shortcoming of study 1 that was not addressed in the paper itself. On the other hand, interventions aimed at reducing procrastination could have a detrimental impact in the same vein as the Expertise Reversal Effect, which is why the learner characteristics that instructions and/or content are adapted to need to be assessed carefully and as accurately as possible. In that sense, providing non-procrastinators or students that delay tasks in a strategic manner with unneeded interventions may have adverse effects, such as distracting them or increasing their cognitive load with information they may not need.

As highlighted in studies 3 and 4, this requires a differentiation between strategic delay and procrastination as state variables, meaning the behaviour must be categorised to allow for accurate predictions to be made, which then enable adaptive interventions. Not intervening is also an option in case students do not delay their tasks in the first place, or do so as part of a deliberate strategy. In both studies, the outcome was dilatory behaviour in general, without the possibility to determine its type. Ways to differentiate between different types without relying on questionnaires include gaining contextual data or analysing click-based and/or temporal log patterns. Lindblom-Ylänne, Saariaho, Inkinen, Haarala-Muhonen, and Hailikari (2015) investigated the former by conducting interviews. Based on the seven core aspects of procrastination and purposeful delay as identified by Klingsieck (2013), they distinguished between five dilatory profiles. Each of their participants was then assigned to one of the five categories: students that showed no delay, strategic delayers, unnecessary delayers, procrastinators without discomfort, and procrastinators with discomfort. This distinction is crucial, since interventions may not always be welcome. They may even be counterproductive, as discussed above. Whether help is appreciated depends on multiple factors, e.g., self-efficacy, self-regulation, motivation, whether fear of failure is involved, or if the delay is motivated by self-handicapping. While they cannot account for all the intricacies of dilatory behaviour and the multitude of reasons behind it, these profiles could provide the means to deliver more specific interventions compared to a one-size-fits-all approach. Moreover, they capture subjective components of dilatory behaviour (Malatincová, 2015) on top of the actual delay.

4.2 The framework

I thus present a framework under the umbrella of design-based research that demonstrates how procrastination and other types of delay may be addressed within an adaptive learning system. Since this framework involves adaptive learning, I first briefly address the basics of adaptive learning systems, which are explained in more detail in study 2. In our view, adaptive learning refers to technologies that adapt the teaching process to the behaviours and needs of individual learners and monitor their learning process (Imhof et al., 2020, see Adams Becker et al., 2018). Adaptive learning systems commonly involve three core components, the domain model, the learner model, and the adaptive model (Vagale & Niedrite, 2012).

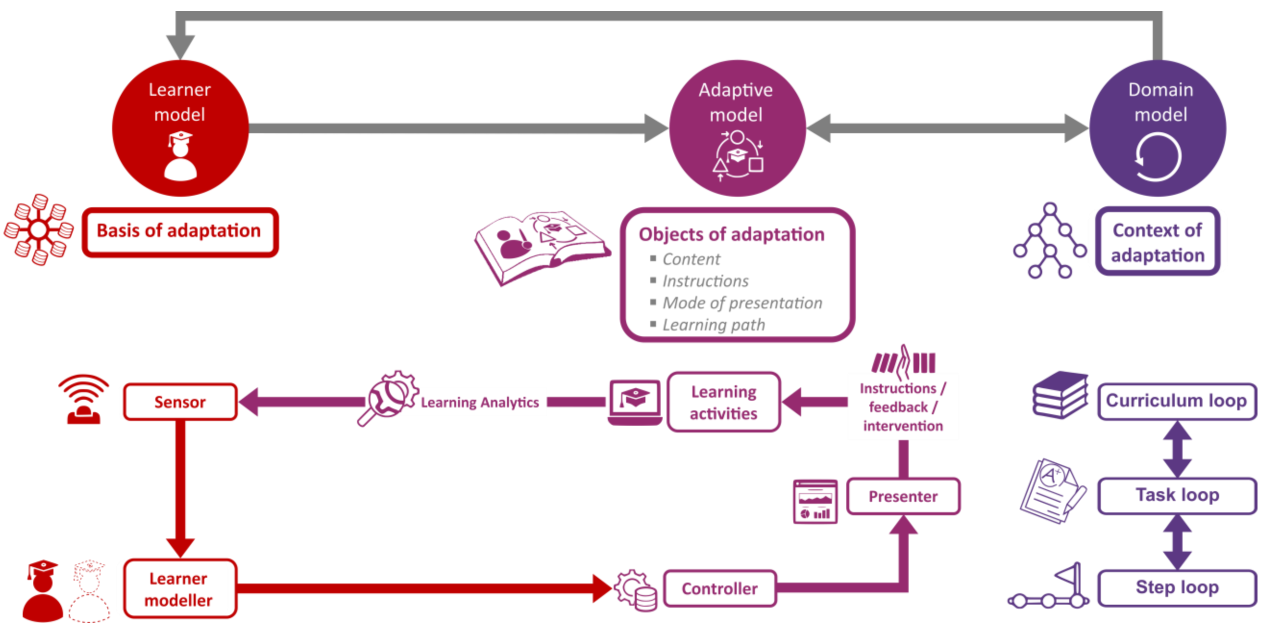

Figure 1. Illustration of the three core components of an adaptive learning system

Figure 1 shows a graphic representation of what elements these models comprise and how they interact. The basis of adaptation is provided by the learner model, which is a representation of the learner via specific learner characteristics that are detected by sensors and further processed by the learner modeller (e.g., abilities, knowledge, personality traits, and preferences).

This information is then passed to the adaptive model, which adapts the objects of adaptation (content, instructions, mode of presentation, learning paths) via the presenter by checking if certain conditions are met via the controller, e.g., if a students' prior knowledge is low enough to justify an intervention by providing more guidance. This decision is also influenced by the domain model, which provides the context of adaptation and involves the topic to be taught (more specifically its contents and structure) as well as the intended learning outcome. It also consists of multiple loops that define the granularity of the adaptive mechanism (step loop, task loop, curriculum loop). The impact of the adapted content (instructions, feedback, intervention) on the student's performance can then be assessed with learning analytics, i.e., analysing collected learner-specific data. These then inform the sensor, thus updating the learner model and restarting the process. Of particular relevance to the framework are the learner and adaptive models.

So far, to my knowledge, no adaptive learning system has focussed directly on procrastination as part of its learner or adaptive models. Even related concepts such as self-regulated learning (Harati, Sujo-montes, Tu, Armfield, & Yen, 2021), motivation, and general meta-cognitive abilities have rarely been used to inform learner models (see Nakić, Granić, & Glavinić, 2015 for an overview over commonly used learner characteristics). Moreover, the assessment of relevant learner characteristics is still mostly rooted in rule-based approaches, rarely employing ML (see Afini Normadhi et al., 2019 for an overview). This highlights the two most crucial novelties of the framework: providing adaptive interventions based on dilatory profiles and doing so with an ML approach, allowing for flexible real-time predictions and interventions.

The goal of this framework is thus to implement the models created in studies 3 and 4 into an LMS to provide real-time predictions of delay, which then serve as the basis for adaptive interventions to reduce or prevent procrastination. However, the predictions will be expanded: instead of solely predicting delay, the models will also predict which dilatory profiles students belong to, inspired by the profiles presented in Lindblom-Ylänne et al. (2015). These profiles will serve as a means to differentiate between different types of delayers in order to provide appropriate automated interventions, as discussed above. This is also the reason why I favour the term *strategic delay* in the context of this thesis over *purposeful delay*, as used in studies 3 and 4.

Since the subject trait variables can be assessed at any time during a semester, they can serve as early predictors to inform the learner model, thus providing the means to overcome the cold-start problem. While their predictive performance is not fully satisfactory as study 4 particularly highlights, they can still serve as initial informants of the learner model before the objective, log data - based predictors become active (e.g., by students first clicking on an assignment). Naturally, this raises a concern since it appears to counteract the usual fading process, where support and guidance decrease over time rather than increasing, as is the case here. The solution I provide in this framework is to split the process into two phases: in phase 1, no historical data (i.e., student data from previous semesters) is available, meaning there is no fading and the amount of data entering the learner model increases over time. In the second phase, historical data can be used to assign students to the dilatory profiles, which can then be updated over time as new information enters the learner model. This process works in the form of a loop.

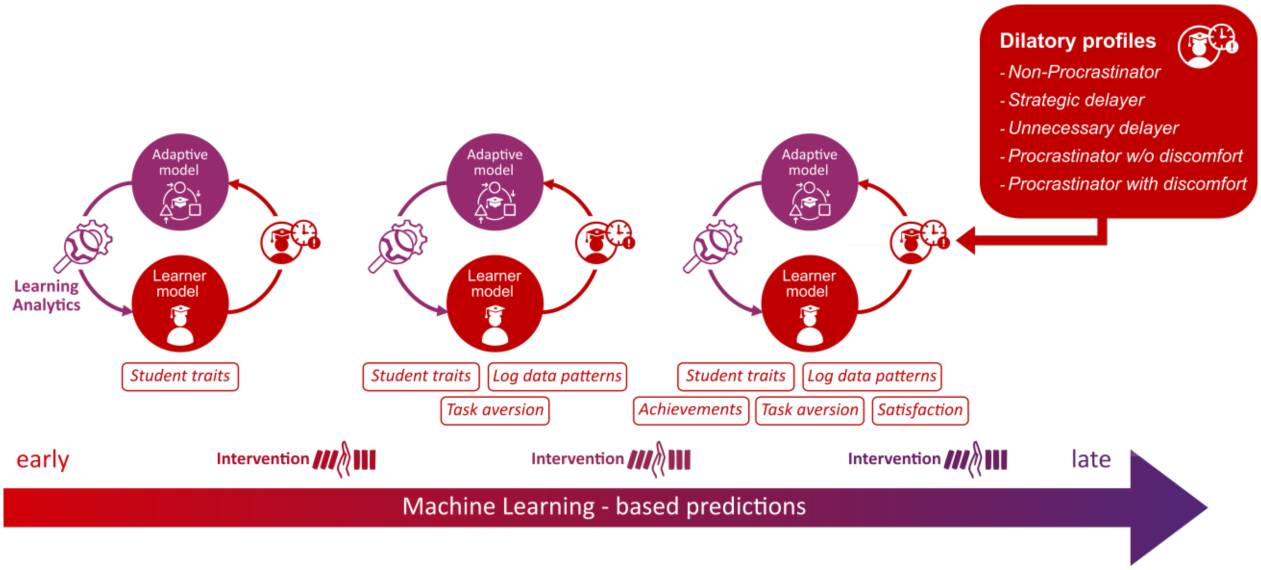

Figure 2. Illustration of the progress of the adaptive mechanism over the span of a semester

Figure 2 shows how phase one of the process is supposed to work under the framework: The learner model is first created based on trait variables, resulting in a dilatory profile which then informs the adaptive model. A student could thus be categorised as a potential procrastinator early in the semester, even though no intervention is necessary yet. As the semester progresses, more and more bits of information enter the learner model via learning analytics (e.g., log data patterns), updating and refining the profile, perhaps changing the initial categorisation. This is a crucial step since early-assessed profiles could introduce biases that get perpetuated in later predictions and ultimately lead to unsuitable interventions. Once enough data is collected, the adaptive model can then provide an intervention, e.g., through a cue, a recommendation, or psychoeducational measures to increase self-efficacy and reduce anxiety. Importantly, the point in time an intervention is made does not have to be fixed and can instead be adapted as well, and so can the deployed algorithm. After all, as study 4 reveals, there is no clear preferable algorithm that works equally well for all types of predictors, so this choice should also be incorporated into the adaptive loop, based on the data available at any given time. After the semester has ended, the final profile can be kept as historical data (with students' consent, of course) and serve as the basis for predictions in the following semester, i.e., phase two. However, this does not necessarily preclude a reassessment of the subject variables, which should be updated from time to time despite their trait-like nature, as discussed in study 4.

4.3 Outlook & potential intervention strategies

However, before the framework can actually be incorporated into an LMS, more research is required. In the next steps, more indicators should be involved in order to increase the predictive performance, which is not as high as it could be (see study 4). While self-efficacy is important and strongly linked to motivation, other motivational factors should also be considered. Fine-grained, task-specific measures to discover task aversiveness to name an example (see Wieland et al., 2021), environmental factors such as institutional conditions (Nordby, Klingsieck, & Svartdal, 2017), and the reasons behind dilatory behaviour (Grunschel et al., 2013b) also need to be involved in upcoming studies to gain a more differentiated, comprehensive picture. Another avenue is to further analyse clickstream patterns, which can serve as an indicator of time management skills (Park et al., 2018) or self-regulation (Li, Baker, & Warschauer, 2020). Increasing the data set used for training the ML models is another suitable way to make sure predictions are performing better once they are part of an LMS. Another concern that still needs to be addressed is the accuracy of the dilatory profiles, which requires its own separate study. After all, we do not know yet how well our models are suited to predict the affiliation with dilatory profiles as opposed to predicting delay values. Sensitivity and the positive predictive value are also of importance in this context: should false positives or false negatives be prioritised, i.e., is it preferable to potentially alienate strategic delayers or to miss actual procrastinators?

Careful consideration also needs to be given to the interventions themselves, which were not part of any of the four studies presented in this thesis. A vital component to any procrastination-related intervention strategy are time management skills, as the literature suggests (see e.g., Wolters & Brady, 2020), along with effort management and psychological flexibility (see Hailikari et al., 2021). The importance of time management was also evident in study 3 in regard to the effect of the deadline type we discovered, as previously discussed. Potential interventions could thus include reminder systems similar to the ones implemented by Onji and Kikuchi (2011) or Antunes et al. (2016).

Nonetheless, improving time management skills should not be the sole strategy, as it fails to directly address emotional issues (Eerde & Klingsieck, 2018). Fear of failure for instance is an important driving factor of procrastination, especially for students (whether they are undergraduates or and graduates, see Rahimi & Hall, 2021). Another emotional key factor is task aversiveness (Afzal & Jami, 2018). As a meta-analysis by Eerde and Klingsieck (2018) reveals, the most effective intervention strategies are based on cognitive behavioural therapy (CBT), which reduced procrastination to a larger degree than the other investigated types of interventions (self-regulation, therapeutic interventions outside of CBT, and interventions focussing on resources and strengths). However, as the authors note, most interventions lack a coherent theoretical basis. They thus suggest future studies should have theoretical underpinnings rooted in goal-setting or self-regulation principles, be they cognitive-motivational or emotional. Moreover, the authors stress that the studies that were part of this meta-analysis did not account for self-determination theory. An intervention study specifically based on that theory was conducted recently by Oram (2021).

In their review, focussing on academic interventions, Zacks and Hen (2018) concluded that self-regulation skills and study habits should be addressed and that instructors ought to discuss career goals, anxieties, and learning experiences in a supportive manner with their students to reduce the risk of procrastination. How these principles can successfully be translated into automated interventions from the perspective of instructional design remains to be seen in follow-up studies.

5 Conclusion

In conclusion, the four studies that form the bulk of this thesis highlight the value of log data - based approaches in the context of student online activity and dilatory behaviour and provide a viable basis for an adaptive framework allowing maladaptive forms of delay such as procrastination to be reduced or prevented in learning management systems. While subjective variables did not fare as well as the objective predictors in terms of predictive performance in any of our delay models, they should not be dismissed out of hand as they may serve an important role as early predictors in real-time adaptive scenarios. As our results imply, the choice of model should not be rigid and may involve multiple different algorithms, depending on the data available.

The proposed framework, which aims to reduce procrastination via adaptive interventions based on dilatory profiles informed by a combination of log data - based and questionnaire - based variables, needs to be evaluated in future studies before it can be implemented in the field. The next steps will include an expansion of the models to allow for affiliation to the profiles to be predicted, the addition of new fine-grained and context-specific predictors to refine the models and increase their predictive performance, as well as determining which interventions are the most suitable in an online learning context.

Acknowledgments

I would like to thank both Prof. Dr. Fred Mast and Prof. Dr. Per Bergamin for granting me the opportunity to write a dissertation in the first place, and for giving me valuable feedback and advice. Thanks also go to Dr. Egon Werlen and Per Bergamin again for many fruitful and insightful discussions over the years and for serving as a real-life reminder system to keep me from ~~procrastinating~~ delaying strategically too much, with varying degrees of success. Many thanks also go to my co-authors Dr. Ivan Moser, Matthias Holthaus, Dr. Stéphanie McGarrity, Dr. Ioan-Sorin Comsa, Dr. Martin Hlosta, and Dr. Behnam Parsaeifard for their input and contributions to our papers, their feedback, and support. Additional thanks go to Ivan for introducing me to R, making my life a whole lot easier in the process, Ioan for his helpful input in regard to the illustrations in this thesis, Dr. Tansu Pancar for providing the log data vital to all of my studies; and of course my other current and former colleagues at the Institute for Research in Open, Distance, and eLearning (IFeL): Victoria Mirata, Dr. Sukanya Nath, Nicole Bergamin, and Franziska Hirt. And special thanks to Ivan, Ioan, and Egon for proofreading this thesis!

A final round of thanks goes out to my girlfriend Rebecca, plus my friends and family, who provided moral support for which I'm eternally grateful for (despite the latter committing the ultimate sin of enquiring about the status of a PhD student's thesis in regular intervals).

References

Abdi Zarrin, S., Gracia, E., & Paixão, M. P. (2020). Prediction of academic procrastination by fear of failure and self-regulation. *Educational Sciences: Theory and Practice*, *20*(3), 34–43. https://doi.org/10.12738/jestp.2020.3.003

Adams Becker, S., Brown, M., Dahlstrom, E., Davis, A., DePaul, K., Diaz, V., & Pomerantz, J. (2018). *NMC Horizon Report 2018: Higher Education Edition* (p. 60). Retrieved from https://library.educause.edu/$\sim$/media/files/library/2018/8/2018horizonreport.pdf

Afini Normadhi, N. B., Shuib, L., Md Nasir, H. N., Bimba, A., Idris, N., & Balakrishnan, V. (2019). Identification of personal traits in adaptive learning environment: Systematic literature review. *Computers and Education*, *130*, 168–190. https://doi.org/10.1016/j.compedu.2018.11.005

Afzal, S., & Jami, H. (2018). Prevalence of Academic Procrastination and Reasons for Academic Procrastination in University Students. *Journal of Behavioural Sciences*, *28*(1), 51–69. Retrieved from http://pu.edu.pk/images/journal/doap/PDF-FILES/04_v28_1_18.pdf

Akram, A., Fu, C., Li, Y., Javed, M. Y., Lin, R., Jiang, Y., & Tang, Y. (2019). Predicting Students' Academic Procrastination in Blended Learning Course Using Homework Submission Data. *IEEE Access*, *7*, 102487–102498. https://doi.org/10.1109/access.2019.2930867

Alexander, E. S., & Onwuegbuzie, A. J. (2007). Academic procrastination and the role of hope as a coping strategy. *Personality and Individual Differences*, *42*(7), 1301–1310. https://doi.org/10.1016/J.PAID.2006.10.008

Antunes, O., Lucena, K. K. T., Oliveira, E. H. T., Gusmão, T. H. Y., Nozawa, E. H., & Brodbeck, W. (2016). LMS monitor: an application for academic and performance monitoring of students in LMS. In *XXI congreso internacional de informática educativa (tise)* (Vol. 12, pp. 45–52).

Blunt, A., & Pychyl, T. A. (2005). Project systems of procrastinators: A personal project-analytic and action control perspective. *Personality and Individual Differences*, *38*(8), 1771–1780. https://doi.org/10.1016/J.PAID.2004.11.019

Cao, L. (2012a). Differences in procrastination and motivation between undergraduate and graduate students. *Journal of the Scholarship of Teaching and Learning*, *12*(2), 39–64.

Cao, L. (2012b). Examining 'active' procrastination from a self-regulated learning perspective. *Educational Psychology*, *32*(4), 515–545. https://doi.org/10.1080/01443410.2012.663722

Cerezo, R., Esteban, M., Sánchez-Santillán, M., & Núñez, J. C. (2017). Procrastinating behavior in computer-based learning environments to predict performance: A case study in Moodle. *Frontiers in Psychology*, *8*, 1–11. https://doi.org/10.3389/fpsyg.2017.01403

Chen, P. S. D., Lambert, A. D., & Guidry, K. R. (2010). Engaging online learners: The impact of Web-based learning technology on college student engagement. *Computers and Education*, *54*(4), 1222–1232. https://doi.org/10.1016/j.compedu.2009.11.008

Choi, J. N., & Moran, S. V. (2009). Why not procrastinate? Development and validation of a new active procrastination scale. *The Journal of Social Psychology*, *149*(2), 195–211. https://doi.org/10.3200/SOCP.149.2.195-212

Chu, A. H. C., & Choi, J. N. (2005). Rethinking procrastination: positive effects of "active" procrastination behavior on attitudes and performance. *The Journal of Social Psychology*, *145*(3), 245–264. https://doi.org/10.3200/SOCP.145.3.245-264

Corkin, D. M., Yu, S. L., & Lindt, S. F. (2011). Comparing active delay and procrastination from a self-regulated learning perspective. *Learning and Individual Differences*, *21*(5), 602–606. https://doi.org/10.1016/j.lindif.2011.07.005

Corno, L. (1993). The Best-Laid Plans: Modern Conceptions of Volition and Educational Research. *Educational Researcher*, *22*(2), 14–22. https://doi.org/10.3102/0013189X022002014

Corno, L., & Mandinach, E. B. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist*, *18*(2), 88–108. https://doi.org/10.1080/00461528309529266

Çelikkaleli, Ö., & Akbay, S. E. (2013). The Investigation of University Students' Academic Procrastination, General Self-Efficacy Belief and Responsibility, 237–255.

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology*, *49*(3), 182–185. https://doi.org/10.1037/a0012801

Deimann, M., & Bastiaens, T. (2010). The role of volition in distance education: An exploration of its capacities. *The International Review of Research in Open and Distributed Learning*, *11*(1), 1–16. https://doi.org/10.19173/IRRODL.V11I1.778

del Puerto Paule-Ruiz, M., Riestra-González, M., Sánchez-Santillán, M., & Pérez-Pérez, J. R. (2015). The procrastination related indicators in e-learning platforms. *Journal of Universal Computer Science*, *21*(1), 7–22. https://doi.org/10.3217/jucs-021-01-0007

Dewitte, S., & Schouwenburg, H. C. (2002). Procrastination, temptations, and incentives: the struggle between the present and the future in procrastinators and the punctual: *European Journal of Personality*, *16*(6), 469–489. https://doi.org/10.1002/PER.461

Doherty, W. (2006). An analysis of multiple factors affecting retention in Web-based community college courses. *The Internet and Higher Education*, *9*(4), 245–255. https://doi.org/10.1016/J.IHEDUC.2006.08.004

Dontre, A. J. (2021). The influence of technology on academic distraction: A review. *Human Behavior and Emerging Technologies*, *3*(3), 379–390. https://doi.org/10.1002/HBE2.229

Dumford, A. D., & Miller, A. L. (2018). Online learning in higher education: exploring advantages and disadvantages for engagement. *Journal of Computing in Higher Education*, *30*(3), 452–465. https://doi.org/10.1007/s12528-018-9179-z

Eerde, W. van, & Klingsieck, K. B. (2018). Overcoming procrastination? A meta-analysis of intervention studies. *Educational Research Review*, *25*, 73–85. https://doi.org/10.1016/j.edurev.2018.09.002

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5-6), 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Ferrari, J. R. (1991). Compulsive procrastination: some self-reported characteristics. *Psychological Reports*, *68*(2), 455–458. https://doi.org/10.2466/pr0.1991.68.2.455

Ferrari, J. R. (2001). Procrastination as self-regulation failure of performance: effects of cognitive load, self-awareness, and time limits on 'working best under pressure': *European Journal of Personality*, *15*(5), 391–406. https://doi.org/10.1002/PER.413

Ferrari, J. R., Díaz-Morales, J. F., O'Callaghan, J., Díaz, K., & Argumedo, D. (2007). Frequent behavioral delay tendencies by adults: International prevalence rates of chronic procrastination. *Journal of Cross-Cultural Psychology*, *38*(4), 458–464. https://doi.org/10.1177/0022022107302314

Ferrari, J. R., O'Callaghan, J., & Newbegin, I. (2005). Prevalence of procrastination in the United States, United Kingdom, and Australia: Arousal and avoidance delays among adults. *North American Journal of Psychology*, *7*(1), 1–6.

Garcia, T., & Pintrich, P. R. (1994). Regulating motivation and cognition in the classroom: The role of self-schemas and self-regulatory strategies. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 127–153). Hillsdale, NJ: Lawrence Erlbaum Associates. Retrieved from http://arxiv.org/abs/arXiv:1011.1669v3

Gazzaley, A., & Rosen, L. D. (2016). *The distracted mind: Ancient brains in a high-tech world*. MIT Press.

Grunschel, C., Patrzek, J., & Fries, S. (2013a). Exploring different types of academic delayers: A latent profile analysis. *Learning and Individual Differences*, *23*(1), 225–233. https://doi.org/10.1016/j.lindif.2012.09.014

Grunschel, C., Patrzek, J., & Fries, S. (2013b). Exploring reasons and consequences of academic procrastination: An interview study. *European Journal of Psychology of Education*, *28*(3), 841–861. https://doi.org/10.1007/s10212-012-0143-4

Grunschel, C., Patrzek, J., Klingsieck, K. B., & Fries, S. (2018). "Jetzt höre ich auf zu prokrastinieren!" Förderung spezifischer Prozesse selbstregulierten Lernens zur Verringerung von akademischer Prokrastination. *Journal of Prevention and Intervention in the Community*, *46*(2), 143–157. https://doi.org/10.1080/10852352.2016.1198166

Gustavson, D. E., Miyake, A., Hewitt, J. K., & Friedman, N. P. (2014). Genetic Relations Among Procrastination, Impulsivity, and Goal-Management Ability: Implications for the Evolutionary Origin of Procrastination. *Psychological Science*, *25*(6), 1178. https://doi.org/10.1177/0956797614526260

Hailikari, T., Katajavuori, N., & Asikainen, H. (2021). Understanding procrastination: A case of a study skills course. *Social Psychology of Education*, *24*(2), 589–606. https://doi.org/10.1007/s11218-021-09621-2

Harati, H., Sujo-montes, L., Tu, C.-h., Armfield, S. J. W., & Yen, C.-j. (2021). Assessment and Learning in Knowledge Spaces (ALEKS) Adaptive System Impact on Students' Perception and Self-Regulated Learning Skills. *Education Sciences, 11*(10), 603. https://doi.org/10.3390/educsci11100603

Harriott, J. S., Ferrari, J. R., & Dovidio, J. F. (1996). Distractibility, daydreaming and self-critical cognitions as determinants of indecision. *Journal of Social Behavior & Personality*, *11*(2), 337–344. Retrieved from https://psycnet.apa.org/record/1996-04808-010

Helmke, A., & Aken, M. A. G. van. (1995). The Causal Ordering of Academic Achievement and Self-Concept of Ability During Elementary School: A Longitudinal Study. *Journal of Educational Psychology*, *87*(4), 624–637. https://doi.org/10.1037/0022-0663.87.4.624

Hlosta, M., Zdrahal, Z., & Zendulka, J. (2018). Are we meeting a deadline? classification goal achievement in time in the presence of imbalanced data. *Knowledge-Based Systems*, *160*, 278–295. https://doi.org/10.1016/j.knosys.2018.07.021

Howell, A. J., & Watson, D. C. (2007). Procrastination: Associations with achievement goal orientation and learning strategies. *Personality and Individual Differences*, *43*(1), 167–178. https://doi.org/10.1016/j.paid.2006.11.017

Höcker, A., Engberding, M., & Rist, F. (2022). *Prokrastination - Extremes Aufschieben*. Hogrefe Verlag. https://doi.org/10.1026/03081-000

Huang, N., Zhang, J., Burtch, G., Li, X., & Chen, P. (2021). Combating Procrastination on Massive Online Open Courses via Optimal Calls to Action. *Information Systems Research*, *32*(2), 301–317. https://doi.org/10.1287/isre.2020.0974

Imhof, C., Bergamin, P. B., Moser, I., & Holthaus, M. (2018). Implementation of an adaptive instructional design for a physics module in a learning management system. In *Proceedings of the 15th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2018)*.

Imhof, C., Bergamin, P., & McGarrity, S. (2020). Implementation of Adaptive Learning Systems: Current State and Potential. In *Online teaching and learning in higher education* (pp. 93-115). Springer, Cham. https://doi.org/10.1007/978-3-030-48190-2_6

Imhof, C., Bergamin, P., & McGarrity, S. (2021). Prediction of dilatory behaviour in online assignments. *Learning and Individual Differences*, *88*. https://doi.org/10.1016/j.lindif.2021.102014

Imhof, C., Comsa, I.-S., Hlosta, M., Parsaeifard, B., Moser, I., & Bergamin, P. B. (2022). *Prediction of Dilatory Behavior in eLearning: A Comparison of Multiple Machine Learning Models*. arXiv preprint arXiv:2206.15079

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

Klassen, R. M., Krawchuk, L. L., & Rajani, S. (2008). Academic procrastination of undergraduates: Low self-efficacy to self-regulate predicts higher levels of procrastination. *Contemporary Educational Psychology*, *33*(4), 915–931. https://doi.org/10.1016/j.cedpsych.2007.07.001

Klingsieck, K. B. (2013). Procrastination when good things don't come to those who wait. *European Psychologist*, *18*(1), 24–34. https://doi.org/10.1027/1016-9040/a000138

Klingsieck, K. B., Fries, S., Horz, C., & Hofer, M. (2012). Procrastination in a distance university setting. *Distance Education*, *33*(3), 295–310. https://doi.org/10.1080/01587919.2012.723165

Lavoie, J. A. A., & Pychyl, T. A. (2001). Cyberslacking and the Procrastination Superhighway: A Web-Based Survey of Online Procrastination, Attitudes, and Emotion. *Social Science Computer Review*, *19*(4), 431–444. https://doi.org/10.1177/089443930101900403

Lee, E. (2005). The relationship of motivation and flow experience to academic procrastination in university students. *Journal of Genetic Psychology*, *166*(1), 5–15. https://doi.org/10.3200/GNTP.166.1.5-15

Levy, Y., & Ramim, M. M. (2012). A Study of Online Exams Procrastination Using Data Analytics Techniques. *Interdisciplinary Journal of E-Skills and Lifelong Learning*, *8*, 097–113. https://doi.org/10.28945/1730

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*, 100727. https://doi.org/10.1016/j.iheduc.2020.100727

Lindblom-Ylänne, S., Saariaho, E., Inkinen, M., Haarala-Muhonen, A., & Hailikari, T. (2015). Academic procrastinators, strategic delayers and something betwixt and between: An interview study. *Frontline Learning Research*, *3*(2), 47–62. https://doi.org/10.14786/flr.v3i2.154

Linkous, H. M. (2021). Self-Directed Learning and Self-Regulated Learning: What's the Difference? A Literature Analysis. *American Association for Adult and Continuing Education*.

Lyons, M., & Rice, H. (2014). Thieves of time? Procrastination and the Dark Triad of personality. *Personality and Individual Differences*, *61-62*(April), 34–37. https://doi.org/10.1016/j.paid.2014.01.002

Malatincová, T. (2015). The mystery of "should": Procrastination, delay, and reactance in academic settings. *Personality and Individual Differences*, *72*, 52–58. https://doi.org/10.1016/j.paid.2014.08.015

Mavroudi, A., Giannakos, M., & Krogstie, J. (2018). Supporting adaptive learning pathways through the use of learning analytics: developments, challenges and future opportunities. *Interactive Learning Environments*, *26*(2), 206–220. https://doi.org/10.1080/10494820.2017.1292531

Meier, A., Reinecke, L., & Meltzer, C. E. (2016). Facebocrastination? Predictors of using Facebook for procrastination and its effects on students' well-being. *Computers in Human Behavior*, *64*, 65–76. https://doi.org/10.1016/j.chb.2016.06.011

Motie, H., Heidari, M., & Sadeghi, M. A. (2012). Predicting Academic Procrastination during Self-Regulated Learning in Iranian first Grade High School Students. *Procedia - Social and Behavioral Sciences*, *69*, 2299–2308. https://doi.org/10.1016/j.sbspro.2013.02.023

Myrick, J. G. (2015). Emotion regulation, procrastination, and watching cat videos online: Who watches Internet cats, why, and to what effect? *Computers in Human Behavior*, *52*, 168–176. https://doi.org/10.1016/j.chb.2015.06.001

Nakić, J., Granić, A., & Glavinić, V. (2015). Anatomy of student models in adaptive learning systems: A systematic literature review of individual differences from 2001 to 2013.

Journal of Educational Computing Research, *51*(4), 459–489. https://doi.org/10.2190/EC.51.4.e

Nordby, K., Klingsieck, K. B., & Svartdal, F. (2017). Do procrastination-friendly environments make students delay unnecessarily? *Social Psychology of Education*, *20*(3), 491–512. https://doi.org/10.1007/s11218-017-9386-x

Onji, K., & Kikuchi, R. (2011). Procrastination, prompts, and preferences: Evidence from daily records of self-directed learning activities. *Journal of Socio-Economics*, *40*(6), 929–941. https://doi.org/10.1016/j.socec.2011.08.019

Oram, R. A. J. (2021). *Stop Procrastinating on Procrastination: An Exploration of Academic Procrastination Through the Lens of Self-Determination Theory* (PhD thesis). Université d'Ottawa/University of Ottawa.

Park, J., Yu, R., Rodriguez, F., Baker, R., Smyth, P., & Warschauer, M. (2018). *Understanding Student Procrastination via Mixture Models*. International Educational Data Mining Society.

Pintrich, P. R. (2000). An Achievement Goal Theory Perspective on Issues in Motivation Terminology, Theory, and Research. *Contemporary Educational Psychology*, *25*(1), 92–104. https://doi.org/10.1006/ceps.1999.1017

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Pychyl, T. A. (2008). Savouring the Flavours of Delay. *English Studies in Canada*, *34*(2-3), 25. Retrieved from https://www.questia.com/read/1G1-214398915/savouring-the-flavours-of-delay

Pychyl, T. A., Lee, J. M., Thibodeau, R., & Blunt, A. (2000). Five days of emotion: An experience sampling study of undergraduate student procrastination. *Journal of Social Behavior & Personality*, *15*(5), 239–254.

Rabin, L. A., Fogel, J., & Nutter-Upham, K. E. (2011). Academic procrastination in college students: The role of self-reported executive function. *Journal of Clinical and Experimental Neuropsychology*, *33*(3), 344–357. https://doi.org/10.1080/13803395.2010.518597

Rahimi, S., & Hall, N. C. (2021). Why Are You Waiting? Procrastination on Academic Tasks Among Undergraduate and Graduate Students. *Innovative Higher Education*, *46*, 759–776. https://doi.org/10.1007/s10755-021-09563-9

Rakes, G. C., & Dunn, K. E. (2010). The impact of online graduate students' motivation and self-regulation on academic procrastination. *Journal of Interactive Online Learning*, *9*(1), 78–93.

Rebetez, M. M. L., Rochat, L., Barsics, C., & Van der Linden, M. (2016). Procrastination as a self-regulation failure: The role of inhibition, negative affect, and gender. *Personality and Individual Differences*, *101*, 435–439. https://doi.org/10.1016/j.paid.2016.06.049

Rebetez, M. M. L., Rochat, L., & Van der Linden, M. (2015). Cognitive, emotional, and motivational factors related to procrastination: A cluster analytic approach. *Personality and Individual Differences*, *76*, 1–6. https://doi.org/10.1016/j.paid.2014.11.044

Renn, R. W., Allen, D. G., Fedor, D. B., & Davis, W. D. (2005). The roles of personality and self-defeating behaviors in self-management failure. *Journal of Management*, *31*(5), 659–679. https://doi.org/10.1177/0149206305279053

Rozental, A., Forsell, E., Svensson, A., Forsström, D., Andersson, G., & Carlbring, P. (2014). Psychometric evaluation of the Swedish version of the pure procrastination scale, the irrational procrastination scale, and the susceptibility to temptation scale in a clinical population. *BMC Psychology*, *2*(1), 1–12. https://doi.org/10.1186/s40359-014-0054-z

Saeid, N., & Eslaminejad, T. (2017). Relationship between Student's Self-Directed-Learning Readiness and Academic Self-Efficacy and Achievement Motivation in Students. *International Education Studies*, *10*(1). https://doi.org/10.5539/ies.v10n1p225

Saks, K., & Leijen, Ä. (2014). Distinguishing Self-directed and Self-regulated Learning and Measuring them in the E-learning Context. *Procedia - Social and Behavioral Sciences*, *112*, 190–198. https://doi.org/10.1016/j.sbspro.2014.01.1155

Schommer-Aikins, M., & Easter, M. (2018). Cognitive flexibility, procrastination, and need for closure linked to online self-directed learning among students taking online courses. *Journal of Business and Educational Leadership*, *8*(1), 112–121.

Schouwenburg, H. C., Lay, C. H., Pychyl, T. A., & Ferrari, J. R. (2004). *Counseling the procrastinator in academic settings*. American Psychological Association. https://doi.org/10.1037/10808-000

Senécal, C., Koestner, R., & Vallerand, R. J. (1995). Self-regulation and academic procrastination. *Journal of Social Psychology*, *135*(5), 607–619. https://doi.org/10.1080/00224545.1995.9712234

Senécal, C., Lavoie, K., & Koestner, R. (1997). Trait and Situational Factors in Procrastination: An Interactional Model. *Journal of Social Behaviour and Personality*, *12*(4), 889–903.

Sirois, F. M. (2004). Procrastination and intentions to perform health behaviors: The role of self-efficacy and the consideration of future consequences. *Personality and Individual Differences*, *37*(1), 115–128. https://doi.org/10.1016/J.PAID.2003.08.005

Sirois, F., & Pychyl, T. (2013). Procrastination and the Priority of Short-Term Mood Regulation: Consequences for Future Self. *Social and Personality Psychology Compass*, *7*(2), 115–127. https://doi.org/10.1111/SPC3.12011

Stainton, M., Lay, C. H., & Flett, G. L. (2000). Trait Procrastinators and Behavior/Trait-Specific Cognitions. *Journal of Social Behavior and Personality*, *15*(5).

Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, *133*(1), 65–94. https://doi.org/10.1037/0033-2909.133.1.65

Steel, P. (2010). Arousal, avoidant and decisional procrastinators: Do they exist? *Personality and Individual Differences*, *48*(8), 926–934. https://doi.org/10.1016/j.paid.2010.02.025

Steel, P., Brothen, T., & Wambach, C. (2001). Procrastination and personality, performance, and mood. *Personality and Individual Differences*, *30*(1), 95–106. https://doi.org/10.1016/S0191-8869(00)00013-1

Steel, P., & König, C. J. (2006). Integrating Theories of Motivation. *Academy of Management Review*, *31*(4), 889–913.

Sundaramoorthy, J. (2018). Self-regulated learning strategies of active procrastinating pre-university students. *Electronic Journal of Research in Educational Psychology*, *16*(45), 325–343. https://doi.org/10.25115/ejrep.v16i45.2096

Svartdal, F. (2017). Measuring procrastination: Psychometric properties of the Norwegian versions of the Irrational Procrastination Scale (IPS) and the Pure Procrastination Scale (PPS). *Scandinavian Journal of Educational Research*, *61*(1), 18–30. https://doi.org/10.1080/00313831.2015.1066439

Svartdal, F., Pfuhl, G., Nordby, K., Foschi, G., Klingsieck, K. B., Rozental, A., … Rebkowska, K. (2016). On the measurement of procrastination: Comparing two scales in six European countries. *Frontiers in Psychology*, *7*, 1307. https://doi.org/10.3389/fpsyg.2016.01307

Thatcher, A., Wretschko, G., & Fridjhon, P. (2008). Online flow experiences, problematic Internet use and Internet procrastination. *Computers in Human Behavior*, *24*(5), 2236–2254. https://doi.org/10.1016/j.chb.2007.10.008

Tice, D. M., & Baumeister, R. F. (1997). Longitudinal Study of Procrastination, Performance, Stress, and Health: The Costs and Benefits of Dawdling. *Psychological Science*, *8*(6), 454–458. https://doi.org/10.1111/j.1467-9280.1997.tb00460.x

Tice, D. M., Bratslavsky, E., & Baumeister, R. F. (2001). Emotional Distress Regulation Takes Precedence Over Impulse Control: If You Feel Bad, Do It! *Journal of Personality and Social Psychology*, *80*(1), 53–67. https://doi.org/10.1037/0022-3514.80.1.53

Tuckman, B., & Sexton, T. (1989). Effects of relative feedback in overcoming procrastination on academic tasks. In *Meeting of the american psychological association, new orleans, la*.

Tuckman, B. W. (2005). Relations of Academic Procrastination, Rationalizations, and Performance in a Web Course with Deadlines. *Psychological Reports*, *96*, 1015–1021. https://doi.org/10.2466/pr0.96.3c.1015-1021

Vagale, V., & Niedrite, L. (2012). Learner Model's Utilization in the e-Learning Environments. In *Tenth international baltic conference on databases and information systems* (pp. 162–174). Vilnius, Lithuania.

Wäschle, K., Allgaier, A., Lachner, A., Fink, S., & Nückles, M. (2014). Procrastination and self-efficacy: Tracing vicious and virtuous circles in self-regulated learning. *Learning and Instruction*, *29*, 103–114. https://doi.org/10.1016/j.learninstruc.2013.09.005

Wieland, L. M., Ebner-Priemer, U. W., Limberger, M. F., & Nett, U. E. (2021). Predicting Delay in Goal-Directed Action: An Experience Sampling Approach Uncovering Within-Person Determinants Involved in the Onset of Academic Procrastination Behavior. *Frontiers in Psychology*, *12*(July), 1–21. https://doi.org/10.3389/fpsyg.2021.695927

Wolters, C. A. (2003). Understanding procrastination from a self-regulated learning perspective. *Journal of Educational Psychology*, *95*(1), 179–187. https://doi.org/10.1037/0022-0663.95.1.179

Wolters, C. A., & Brady, A. C. (2020). College Students' Time Management: a Self-Regulated Learning Perspective. *Educational Psychology Review*. https://doi.org/10.1007/s10648-020-09519-z

Yang, Y., Hooshyar, D., Pedaste, M., Wang, M., Huang, Y. M., & Lim, H. (2020a). Predicting course achievement of university students based on their procrastination behaviour on Moodle. *Soft Computing*, 1–17. https://doi.org/10.1007/s00500-020-05110-4

Yang, Y., Hooshyar, D., Pedaste, M., Wang, M., Huang, Y. M., & Lim, H. (2020b). Prediction of students' procrastination behaviour through their submission behavioural pattern in online learning. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-020-02041-8

Yilmaz, M. B. (2017). The Relation between Academic Procrastination of University Students and Their Assignment and Exam Performances: The Situation in Distance and Face-to-Face Learning Environments. *Journal of Education and Training Studies*, *5*(9), 146. https://doi.org/10.11114/jets.v5i9.2545

You, J. W. (2015). Examining the Effect of Academic Procrastination on Achievement Using LMS in e-Learning. *Educational Technology & Society*, *18*(3), 64–74.

Zacks, S., & Hen, M. (2018). Academic interventions for academic procrastination: A review of the literature. *Journal of Prevention and Intervention in the Community*, *46*(2), 117–130. https://doi.org/10.1080/10852352.2016.1198154

Zimmerman, B. J. (2008). Goal setting: a key proactive source of academic self-regulation. (D. H. Schunk, Ed.). New York: Lawrence Erlbaum.

Zuber, S., Cauvin, S., Haas, M., Daviet, A. S., Da Silva Coelho, C., & Kliegel, M. (2020). Do self-reports of procrastination predict actual behavior? *International Journal of Methods in Psychiatric Research*, *29*(4), 1–6. https://doi.org/10.1002/mpr.1843

IMPLEMENTATION OF AN ADAPTIVE INSTRUCTIONAL DESIGN FOR A PHYSICS MODULE IN A LEARNING MANAGEMENT SYSTEM

Christof Imhof, Per Bergamin, Ivan Moser and Matthias Holthaus
UNESCO chair on personalized and adaptive distance education, Swiss Distance University of Applied Sciences (FFHS)
Überlandstrasse 12, Brig, Switzerland

ABSTRACT

This article demonstrates how an adaptive instructional design for a physics module can be realized in a standard learning management system. We implemented a didactic design with physics-specific online exercises that were accompanied by either detailed or non-detailed instructions, depending on the results of the previous task (or a prior knowledge test for the very first exercise). This was realized by use of simple technological tools within the framework of a straightforward recommender system with four components. Consequently, students with less prior knowledge and/or lower learning achievements received more and different teaching assistance than those with high levels of prior knowledge or performance. This was done in the form of recommendations embedded within task feedback, suggesting which task to tackle next. We present first results which show that prior knowledge and online activity contribute to the learning progress in different ways depending on the type of task that was chosen. The detailed versions of the tasks were beneficial only to the students with lower or medium prior knowledge test scores while the students with higher levels of prior knowledge had less learning progress. In the future, our simple recommender system may serve as the basis for a more complex adaptive system, further closing the gap between research and practice in the field of technology-based adaptive learning.

KEYWORDS

Technology-based Learning, Adaptive Learning, Cognitive Load, Expertise Reversal Effect, Learning Management System, Log Files

1. INTRODUCTION

As countless aspects of our lives have become more and more digitalized in the past few decades, so has learning. Accordingly, new forms of learning such as distance learning or technology-based learning have emerged and have been gaining in importance ever since (Bergamin et al. 2012). Due to their flexible nature, these new forms of learning allow learners independence and autonomy, offer freedom of choice and break space-time barriers, thus granting many people the opportunity to pursue academic studies in circumstances that do not commonly allow for that to occur (e.g. full-/part-time employment or parenthood). In addition, such flexibility allows for individual requirements and contexts to be taken into account, which can vary substantially between learners and which can then determine the appropriate instructional design. In university education, learners are usually expected to develop the same competences over the course of their studies despite differing in key characteristics such as different prior knowledge or experience in relation to certain topics or learning skills. One way to achieve the goal of comparable learning success despite heterogeneous preconditions is through adaptation of the learning process, replacing the classic “one-size-fits-all” approach.

The importance of adapting the learning process to the needs of the learner can be demonstrated by the finding that instructional techniques (e.g. guidance by a tutor or detailed instructions) that prove beneficial for novices in a given field can lose their effectiveness or even be counterproductive when applied to experts; a phenomenon known as the Expertise Reversal Effect (Kalyuga et al. 2003). On a technical level, adaptive learning environments may be provided through Learning Management Systems (LMS), which are increasingly accessible due to the raise of technology-based learning.

Despite the popularity of studies on adaptive educational hypermedia, actual practical implementations of adaptive technology-based learning are still scarce (Somyürek 2015). Scanlon et al. (2013) found "a surprising failure" (p. 4) to translate research results in the field of technology-based learning (e.g. prototypes) into commercial applications, due to the so-called "valley of death" (i.e. failure due to lack of funding amongst other factors). According to Price et al. (2017), this gap between research and practice in technology-based learning appears to be systemic in nature, requiring change on multiple levels, including institutional change. Murray and Pérez (2015) claim that intelligent technology-based learning environments are still "years away", in spite of the advances that have been made, and appeal for pedagogy rather than technology to "drive the evolution of advanced learning systems" (p. 124). Oxman and Wong (2014) also identify the challenges of (further) implementation of technology-based learning systems as structural (e.g. term length) and operational rather than technological. Therefore, bridging the gap between research and practice requires an interdisciplinary approach that involves large-scale field studies in appropriate contexts as well as well-founded instructional designs (cf. Scanlon et al. 2015). This study seeks to narrow the gap between experimental research and its practical application by addressing the question whether an adaptive learning system can be implemented in a traditional learning environment without the use of high-end technology (such as deep neural networks). In this paper, we demonstrate how adaptive learning can be implemented in practice on a university level by applying it to a physics module within the learning management system "Moodle". Our approach utilises a fairly simple rule-based instructional design. We further explore to what extent adaptive instruction design, online activity and prior knowledge are related and how much they contribute to the learning progress. Finally, we discuss the potential and limitations of rule-based adaptive learning systems within the boundaries of a standard LMS.

2. THEORETICAL BACKGROUND

As explained above, the Expertise Reversal Effect – the phenomenon that teaching support which is beneficial for novices can turn out to be superfluous or even detrimental to experts and vice versa – may obstruct learning success in many learning scenarios, especially in classic settings where all learners receive the same instructions or guidance (Kalyuga 2007b). "Reversal" refers to the circumstance that the effectiveness of didactical aspects may be reversed for different levels of learners' expertise (Lee & Kalyuga 2014). The most common explanation for this effect is the Cognitive Load Theory (Sweller 1988), which focusses on the interactions between long-term and working memory. According to this theory, the former is used to store knowledge and has an unlimited storing capacity while the latter is involved in consciously processing novel information, but is limited in its capacity of storing it, both in terms of amount and duration (van Merriënboer & Sweller 2005). Recent accounts of the theory differentiate between (at least) two kinds of cognitive load, the intrinsic load and the extraneous load. Intrinsic load concerns cognitive processes that are required for processing learning materials and may be affected by the (perceived) complexity or difficulty of the material. In contrast, extraneous load refers to cognitive processes caused by factors that are not directly related to the learning task but are nevertheless crucial for the learning process, for instance convoluted instructional design or unfavourable presentation of the learning material (Kalyuga 2009a).

The basis of the Cognitive Load Theory is the notion that the intrinsic and extraneous loads combined cannot exceed the limitations of the working memory (Paas et al. 2003). If high extrinsic load results from an unnecessary processing of design or presentation aspects, less capacity can be made available for the processing of the actual learning tasks (i.e. intrinsic load). Consequently, learning is impaired if the learning activities require too much cognitive capacity, resulting in overload.

Importantly, the current cognitive load of a learner is not only determined by objective factors (such as difficulty of the learning content and instructional design), but also by characteristics of the learner. In parts, human expertise results from cognitive schemata with varying degrees of complexity and automation housed by the long-term memory (van Merriënboer & Sweller 2005). Knowledge is stored in and organized by these schemata, which may become automated through training, thus freeing space in the working memory, thereby reducing the intrinsic load and leaving more cognitive capacity for the processing of new content (Kalyuga 2009a). Put differently, the level of available knowledge exerts considerable influence on the cognitive load (Kalyuga 2007b). This implies that optimal teaching of complex learning content needs to take learners' cognitive load into account and actively manage it though instructional interventions (Somyürek, 2015; Rey & Buchwald, 2011). In the case of the Expertise Reversal Effect, this concerns the degree of instructional guidance: On the one hand, a lack of sufficient guidance during a complex task may result in the application of poor problem-solving strategies or arbitrary trial-and-error behaviour. On the other hand, vast amounts of instructional guidance may lead learners to squandering their resources by comparing and contrasting their prior knowledge with the incoming information, thus inflating their intrinsic load (Kalyuga 2007b).

Consequently, at the start of the learning process, novices should be provided with instructional guidance (e.g. step-by-step instruction) in order to help them with their tasks and optimise their cognitive load. As the learners gain more expertise over time, this guidance can then gradually be reduced (cf. the concept of fading scaffolds; van Merriënboer & Sluijsmans 2009). The educational implications of the Cognitive Load Theory in general and the Expertise Reversal Effect in particular have been confirmed by numerous studies (e.g. Rey & Buchwald 2011). However, it should also be noted that the cognitive load approach is limited to the acquisition of subject-specific knowledge as the learning goal (Kalyuga & Singh, 2016) and is less applicable to other learning objectives such as the acquisition of self-regulated learning skills or the increase of learning motivation. Moreover, learners may feel overburdened with the necessary monitoring and adaptation of the learning process (Kirschner & van Merriënboer 2013). Technology-based adaptive learning can assist these processes, thus reducing the extrinsic load and increasing the effectiveness of learning.

In contrast to traditional technology-based approaches, adaptive learning allows learning aspects (contents, navigation, support) to be presented in a dynamic environment that continually changes in response to information collected in the course of learning. This raises the question, which sources of a learning scenario are most suitable as a basis for adaptation processes in a course module (cf. Nakić et al. 2015, for an overview). Principally, three main groups of characteristics can be identified: (1) stable or situational personal characteristics of the learners such as gender, culture, learning style, prior knowledge or emotional state, (2) specific characteristics of the content such as topics or task difficulty and (3) characteristics of the context such as learning time or self-regulation (Wauters et al. 2010). The learning process itself can then be adapted by means of altering the instructional design regarding the relevant learning objects in accordance with the needs of an individual or a group of learners.

3. INSTRUCTIONAL DESIGN AND SYSTEM IMPLEMENTATION

In our instructional design, we focus on task difficulty as the adaptive factor, similar to Brunstein et al. (2009); Hsu et al. (2015); and van Der Kleij et al. (2015). Distance students in general and the students at our university in particular tend to have significantly different levels of background knowledge and learning strategies (mostly due to different educational and/or professional careers). Therefore, we implemented an adaptive instruction design in the learning management system used by our university (Moodle), which recommended tasks with step-by-step detailed or non-detailed instructions and thus varied the task difficulty accordingly. In the context of self-regulated learning, students could either follow the recommendation or choose an alternative learning path. Aiming to reduce the cognitive load, less proficient students were given instructions that offered more support and assistance while more proficient students received less support, thus increasing the task difficulty. The recommendations and interventions were each embedded in the feedback of the previously processed task. Our framework for processing and linking learning data with adaptive learning instructions was inspired by a model by Zimmermann et al. (2005). Conceptually, the system was based on four components that together formed the adaptation mechanism. The first component was the sensors, which were linked to the task data (specifically if a task has been solved correctly or not). The second component was the analyser, which collected the data measured by the sensors. This information was then transferred to the third component, the controller, which determined whether a certain threshold had been met. Depending on the outcome, the controller determined if the learning object (for example a task) was to be adjusted. The last component was the presenter, which then displayed objects of learning support (such as recommendations). As an entry point, we used the data from a prior knowledge test that was administered at the beginning of the course and assessed the level of expertise with which the students started the course. In the course of the semester, further tests and assessments were then fed to the sensor component data base. Consequently, students received instructions and learning support adapted to their learning performance and behaviour. The learning support focused on three different elements: the initial sensor, the step loop and the task loop.

The initial sensor was an assessment of prior knowledge in physics in the form of a set of standard exercises solved at the very start of the course. Based on their performance, the students were then divided into two groups, "novices" (less than 50% correct answers) and "experts" (more than 50% of the answers were correct). Depending on the score, the first proper task in the module appeared in a detailed (high learning support) or a standard form (low learning support). The second element, the step loop, measured the current level of knowledge within a task and accordingly determined the appropriate learning support. Based on the correctness of the answer, the students received feedback which was provided after each step in the task and changed depending on how often the same question was answered incorrectly. This served the purpose of clarifying possible misunderstandings the students might have had as quickly as possible, e.g. by reminding them of forgotten information (Durlach & Ray 2011). The third element, the task loop, consisted of two kinds of tasks, namely standard tasks and transfer tasks. The former assessed the ability to solve a particular physics problem, while the latter had two objectives: On the one hand, the transfer task checked whether a student had understood a particular problem and was thus able to solve the task in its standard form (see "vertical transfer", van Eck & Dempsey 2002), and on the other hand, it evaluated whether the student was able to apply the now acquired problem-solving knowledge to a similar task in a different topic (see "horizontal transfer", van Eck & Dempsey 2002). Within each set of tasks, the system recommended which task the student should tackle next and in what form (detailed or standard). The detailed version featured numerous small solution steps, while the standard version was composed of few solving steps. We chose a rule-based adaptive learning system with a fixed set of rules. The reason for this decision was two-fold: On the one hand, the sensors in our learning scenario do not generate enough data for a complex self-learning system and on the other hand, we wanted to keep the adaptation mechanisms transparent for our students in the sense of an "open learner model".

Starting in the autumn semester of 2015/16, we carried out a two-year field study that implemented our adaptive learning approach into one of our university's study modules. The chosen bachelor module was part of the course of studies in industrial engineering and featured three main physics topics (thermodynamics, optics and microphysics). The module was organized in a blended learning format, which means there was a mixture of face-to-face sessions (20% of the overall expected effort) and both on- and off-line self-study (80%). In order to promote the acceptance of our system among students, we chose a mixed control approach between the system (adaptation) and the learner (adaptability). In their feedback, the students thus received recommendations as to which tasks they should ideally complete and in which form they should choose it. As previously stated, compliance with these recommendations was always optional. The sensor data (i.e. the current state of knowledge) was made available to students in a transparent and concise way to foster their self-assessment skills as well as their acceptance of the recommendations (see open learning models, e.g. Long & Aleven 2017; Suleman et al. 2016).

4. ANALYSIS

4.1 Object of Investigation, Subjects, Procedure and Hypotheses

As previously stated, the students attending a physics module in the semesters 2015/16 and 2016/17 served as the participants in our investigation. The module is offered each autumn semester and was chosen for its reputation as a "problem module", due to its above-average failure rate. Each semester, the students are divided into seven or eight classes, split among different lecturers. Considering the high degrees of employment of our students, the class division was based on the students' preferences in terms of optimal time and location for the five face-to-face events (with four options to choose from in terms of location), which had no bearing on the module content or the online part of the course. In the autumn semester 2015/16, 105 students were enlisted in the course while 106 students participated in the course the following year (2016/17). In both years, the data of several students had to be excluded for our main analyses since they either didn't complete the prior knowledge test (11 in the 2015/16 semester) or the final exam at the end of the semester (7 in the first and 16 in the second year). There were no missing prior knowledge test scores in the second year since the test was not optional anymore. The course offered 43 adaptive tasks not all of which had to be completed by the students. More proficient students for example may only need to solve a fraction of the task array after having successfully completed the first assessment (initial sensor) in order to succeed at the final test at the end of the semester while less proficient students may have to complete a larger portion of the task set in order to achieve the same goal (following the recommendations).

As for the procedure, we first evaluated the distributions of the prior knowledge test scores and three different indicators of online activity (sum of daily clicks, sum of completed standard tasks and sum of completed detailed tasks) based on the students' log files. The sum of daily clicks was used as a general online activity measure, while the other two were specific for engagement with either detailed or non-detailed tasks. In order to investigate how our instruction design impacted the learning progress of our participants, we then explored the relationship between online activity, prior knowledge and the learning progress by calculating several regression models. We formulated four hypotheses for this exploratory study: H_1: prior knowledge is negatively related with learning progress (the less one knows, the more one can learn, i.e. a ceiling effect); H_2: general online activity (in the form of the sum of daily clicks) and learning progress are positively related; H_3: engagement with the tasks (be it detailed or standard tasks) is positively related with learning progress;

H_4: There is a negative interaction between online activity (all three forms thereof) and prior knowledge (more online activity benefits less knowledgeable students more than it does those with high levels of prior knowledge). The whole procedure was done separately for both semesters since the second semester served as a replication and expansion of the first. All statistical analyses were performed with R (R Core Team 2013).

4.2 Distributions of the Prior Knowledge Test Scores and Online Activity Measures

Before examining the distribution of the prior knowledge test scores, the scores in the separate three topics (microphysics, thermodynamics and optics) were added and standardised to an index of 100. Using the R psych package (Revelle 2017), we calculated the mean scores of the prior knowledge test and the mean sums for the online activity measures, as well as the skewness and kurtosis of the distributions (see *Table 1* for an overview). The normality of the distributions was then tested using the Shapiro-Wilk normality test. In the 2015/16 semester, the null hypothesis that the data was normally distributed wasn't rejected (W = 0.99, p = .59) while it clearly was in the 2016/17 semester (W = 0.94, p < .001). Thus, for the semester of 2015/16, the test scores were found to be approximately normally distributed while in the 2016/17 semester, the data was not normally distributed but instead skewed to the left due to overall low test scores (not a single person reached a score higher than 45 out of 100) and a high proportion of test scores of 0.

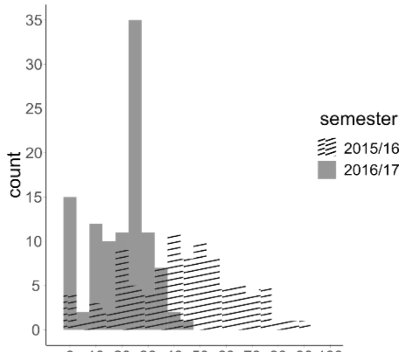

Figure 1. Two overlapping histograms showing the distributions of the prior knowledge test scores in both semesters. The frequencies of the scores did not differ from a normal distribution in year 1 (2015/16) but were left-skewed in year 2 (2016/17)

As for the online activity measures, only completed tasks were accounted for since the recommendation in the task loop was only given after having completed a task; aborted tasks were not counted. All six online measure distributions were clearly left-skewed due to the high frequencies of low levels of activity.

Table 1. Distribution of the prior knowledge test scores and three online activity measures in the two semesters, including the number of participants, means, standard deviations, the Shapiro-Wilk test statistic W and the according p-value

Measure	Mean (*SD*)	Skewness	Kurtosis	W (*p*)
	Semester 2015/16			
Prior knowledge test score (n=94)	42.88.0 (20.93)	-0.06	-0.53	0.99 (.59)
Sum of daily clicks (n=87)	831.56 (693.19)	1.25	1.77	0.89 (<.001)
Sum of completed standard tasks (n=87)	7.03 (8.72)	1.40	1.26	0.80 (<.001)
Sum of completed detailed tasks (n=87)	9.25 (8.30)	0.65	-0.72	0.90 (<.001)
	Semester 2016/17			
Prior knowledge test score (n=106)	19.71 (11.02)	-0.36	-0.67	0.94 (<.001)
Sum of daily clicks (n=90)	860.75 (626.45)	0.61	-0.07	0.95 (<.001)
Sum of completed standard tasks (n=90)	11.10 (9.55)	0.75	0.11	0.92 (<.001)
Sum of completed detailed tasks (n=90)	7.85 (7.32)	0.76	-0.10	0.90 (<.001)

4.3 Prior Knowledge, Online Activity and Learning Progress

In the following section, we report the relationship between prior knowledge, online activity and the learning progress within the adaptive module. For this purpose, the learning progress for each student was defined as the difference between the results of the prior knowledge test and the final test, both standardized to 100. We explored the relationship between the prior knowledge, the students' online activity and their learning progress by calculating three regression models for each of the three online activity measures ("OA" models 1 to 3) as well as a regression model solely containing prior knowledge (the "PK" model), all of which was done separately for both semesters. We used the R package ggplot2 (Wickham 2009) for the regression plots. The first OA model included only the online activity measure as a predictor. The second OA model added a second predictor in the form of the prior knowledge test score and the third and final OA model added the interaction between the online activity measure and the prior knowledge test score (see *Figure 2*). In all instances, the three models were supposed to predict the learning progress.

4.3.1 Prior Knowledge

For both semesters, the PK model yielded a significant regression with an adjusted R^2 of 0.29 for the 2015/16 semester (F(1, 85) = 36.01, *p* <.001) and an adjusted R^2 of 0.03 for the 2016/17 semester (F(1,85) = 4.02, *p* = .047). In line with our hypothesis H_1, there was a significant negative relationship between prior knowledge test score and predicted learning progress (b = -0.61, *p* < .001) for the 2015/16 semester, i.e. students with higher prior knowledge showed less learning progress. A similar relationship was found for the 2016/17 semester (b = -0.31, *p* = .048).

4.3.2 Sum of Daily Clicks and Prior Knowledge

For the 2015/16 data, two significant regression models were found, namely OA models 2 (F(2, 84) = 21.8, *p* < .001) with an adjusted R^2 of 0.33 and model 3 (F(3, 83) = 15.82, p < .001) with an adjusted R^2 of 0.34. In model 2, there was a positive relationship between sums of daily clicks and predicted learning progress (b = 0.01, *p* = .02) and a negative relation between prior knowledge and learning progress (b = -0.67, *p* < .001), supporting H_2 and H_1 respectively. Contrary to H_4, the interaction between the two predictors (model 3, see *Figure 2*) was not significant (b = -0.0003, *p* = .09). Model 2 significantly improved model fit compared to model 1 (F(1, 84) = 43.96, *p* < .001), while model 3 did not did not improve model fit compared to model 2 (F(1, 83) = 2.88, *p* = .09). In contrast, none of the regression models were significant for the 2016/17 data.

4.3.3 Sum of Completed Standard Tasks and Prior Knowledge

We again found two significant regression models in the 2015/16 semester, namely for OA models 2 (F(2, 84) = 29.07, *p* < .001) and 3 (F(3, 83) = 19.18, *p* < .001), both with an R^2 of 0.39. In model 2, both factors were significant predictors of learning progress. As before, participants' predicted learning progress was positively related to the sum of completed standard tasks (b = 1.08, $p < .001$) and negatively related to the prior knowledge test score (b = -0.91, $p < .001$). In model 3 (see *Figure 2*), the interaction between the two predictors again did not reach significance (b = -0.003, $p = .80$). Thus, the results supported hypotheses H_1 and H_3 but not H_4. As before, model 2 provided a better model fit (F(1, 84) = 56.51, $p < .001$) compared to model 1. The same could not be said about model 3 (F(1, 83) = 0.06, $p = .80$). Again, none of the regression equations were significant in case of the 2016/17 data.

4.3.4 Sum of Completed Detailed Tasks and Prior Knowledge

In the 2015/16 semester, all three models were significant, with OA model 1 (F(1, 85) = 6.54, $p = .01$) having an R^2 of 0.06, model 2 (F(2, 84) = 21.2 , $p < .001$) an R^2 of 0.32 and model 3 (F(3, 83) = 17.67 , $p < .001$) an R^2 of 0.37. In model 1, the predicted learning progress was positively related to the sum of completed detailed tasks (b = 0.75, $p = .01$). This was also in the case in model 2 (b = 0.55, $p = .03$). Like before, the relationship between predicted learning progress and prior knowledge test was negative (b = -0.58, $p < .001$). Unlike before, the interaction between online activity and prior knowledge (see *Figure 2*) was significant (b = -0.04, $p = .01$). Hence, hypotheses H_1, H_3 and H_4 were all supported by the data. A simple slope analysis (using the R package jtools, Long 2018) revealed that the slope of the sum of completed detailed tasks was only significantly different from zero when the prior knowledge test score was either medium (b = 0.52, $p = .04$) or low (b = 1.28, $p < .001$). As before, model 2 (F(1, 84) = 35.92, $p < .001$) was an improvement over model 1 in terms of model fit. Unlike before, model 3 further improved upon model 2 in this case (F(1, 83) = 7.39, p =.01). However, none of the regression models were significant for the 2016/17 data.

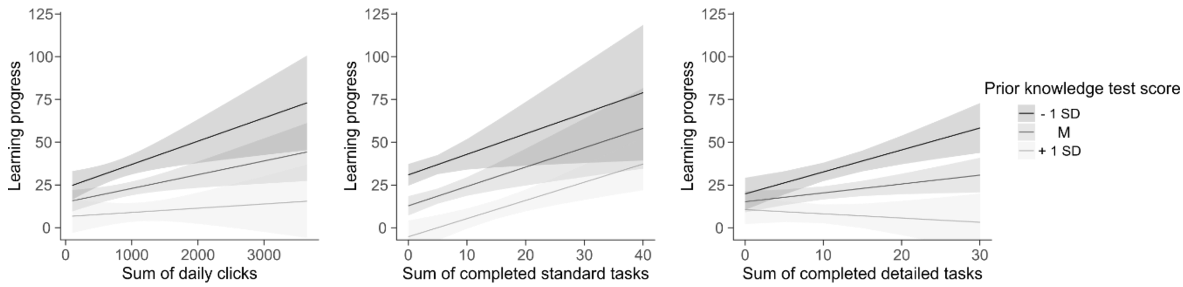

Figure 2. Regression model 3 for all three measures of online activity separately. All models are based on 2015/2016 data exclusively. In model 3, learning progress is predicted by *prior knowledge test score*, *online activity* and the interaction between the two. That interaction was significant only when using sum of completed detailed task as the measure of online activity. In order to illustrate the interaction, we plotted the predicted effect of online activity on learning progress using simple slopes for students that scored one standard deviation below the mean (-1 SD), students that scored one SD above the mean (+1 SD) and students with mean prior knowledge test scores (M)

5. DISCUSSION

The purpose of this article was to demonstrate how an adaptive instruction design could be implemented in a standard learning management system. On a theoretical level, the instruction design was based on the Cognitive Load Theory in general and the Expertise Reversal Effect in particular. We developed an adaptive task set which was then combined with a rule-based recommendation system within the learning management system Moodle. Our adaptive system was applied to a physics module during two semesters (2015/16 and 2016/17) with the respective enlisted students of our university serving as our participants.

The results for the 2015/16 semester reveal that the prior knowledge test score and the level of online activity are both important predictors of learning progress. As hypothesized, the relationship between online activity (be it the sum of daily clicks or the amount of standard tasks solved) and learning progress was positive (the higher the level of online activity, the larger the learning progress), while the relationship between prior knowledge test scores and learning gain was negative (the higher the initial score, the smaller the learning gain). Contrary to our hypotheses, this effect was independent from the level of online activity in most cases with the detailed tasks being the exception, where we found the expected negative interaction: the amount of tasks solved was positively related with learning gains when the prior knowledge test score was low and negatively related when the score was high. This means the less proficient students (i.e. the "novices") benefitted from the detailed tasks more than the advanced students (i.e. the "experts"). Even though this result suggests a presence of the Expertise Reversal Effect (Kalyuga et al. 2003), an according conclusion cannot be drawn unless we know whether the students followed the recommendations. Moreover, the ceiling effect we found could also account for that result, especially when considering that classic exams always have a maximum score that cannot be exceeded.

Surprisingly, neither the prior knowledge test scores nor the levels of online activity predicted the learning progress in the 2016/17 semester, neither separately nor in interaction. We argue that this result was found due to the peculiar distribution of the prior knowledge test scores. As stated before, the scores in that semester were very low with not even a single student reaching 50 points out of a 100. The test was exactly the same as the year before, where it showed no floor or ceiling effects and yielded a mean slightly below what one would expect for a test of medium difficulty (in this case, 50). Since the test itself was identical, the difference between the semesters must have a different explanation. Even though the distributions of the online activity levels were also left-skewed, this could not explain the difference since those distributions were similar in both semesters. As previously stated, what differed between the semesters was the circumstance that the prior knowledge test was mandatory in the 2016/17 while it was not the year before. Motivational aspects thus may factor into a possible explanation or a potential homogeneity within the particular group of students, which has yet to be investigated (e.g. similar previous education that was light on physics).

A major limitation of the study is the lack of a control group. Control groups are hard to achieve in contexts like the one this field study was conducted in since the grades the students receive at the end of the semester are real, thus posing ethical problems comparable to withholding of treatment in clinical studies. In this particular case, there was no non-adaptive past version of the course available to compare to the adaptive one, and even if there was, comparisons might not be all that conclusive given the difference in key factors we witnessed between the 2015/16 and 2016/17 semesters. As of the time of this writing, the adaptive system is still being used in the 2017/18 semester, the results of which will be available soon. Another limitation is the lack of information concerning whether the students actually complied with the recommendations (rather than making their choice either at random or as they saw fit). However, the students' compliance with the recommendations will be analysed soon as well.

6. CONCLUSION

We demonstrated that a simple rule-based adaptive system can be implemented in a common learning management system, getting one step closer to bridging the gap between theory and practice of implementing adaptive systems. As our results show, the implementation of the task-difficulty based adaption process was successful (to an extent), even though we (expectedly) encountered a ceiling effect concerning the learning progress of the students with higher levels of prior knowledge. Our results also demonstrate the importance of reliable assessments serving as the sources of adaption (e.g. well-balanced tests). Future studies in this line of research could grant insight into the transfer tasks, which were very rarely solved by the students and thus could not be analysed properly. Despite the limitations of this study, this line of research holds potential since the successful implementation of simple rules can serve as the basis for a more complex adaptive system, for instance by expanding the sources for adaption (e.g. mood) or by refining the instruction design (e.g. by adding more levels of difficulty). In the near future, research in the field of learning analytics (see e.g. Bannert et al. 2017) may result in the development of more holistic sensors that measure several metrics relevant to learning at once, providing multiple sources for adaption, potentially even allowing implementations within common LMS. Such improvements could result in adaptive systems with more intelligent and complex algorithms at their core, which could then lead to more accurate prognoses, higher efficiency of the adaptive process and more suitable recommendations for the students, which in turn may enhance the acceptance of such systems among students and lecturers.

ACKNOWLEDGEMENT

We would like to thank the many students who participated in the course as well as Dr. Stéphanie McGarrity for proofreading this manuscript and giving us invaluable advice.

REFERENCES

Bannert, M. et al., 2017. Relevance of Learning Analytics to Measure and Support Students' Learning in Adaptive Educational Technologies. In *LAK'17 Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. Vancouver, BC, Canada: ACM New York, NY, USA, pp. 568–569.

Bergamin, P.B. et al., 2012. The Relationship between Flexible and Self-Regulated Learning in Open and Distance Universities. *International Review of Research in Open and Distance Learning*, 13(2), pp.101–123.

Brunstein, A., Betts, S. & Anderson, J.R., 2009. Practice Enables Successful Learning Under Minimal Guidance. *Journal of Educational Psychology*, 101(4), pp.790–802.

Durlach, P.J. & Ray, J.M., 2011. *Designing adaptive instructional environments: Insights from empirical evidence (Technical Report 1297)*, Arlington, VA: U. S. Army Research Institute for the Behavioral Social Sciences.

van Eck, R. & Dempsey, J., 2002. The Effect of Competition and Contextualized Advisement on the Transfer of Mathematics Skills in a Computer-Based Instructional Simulation Game. *Educational Technology Research & Development*, 50(3), pp.23–41.

Hsu, Y. et al., 2015. Interactions Between Levels of Instructional Detail and Expertise When Learning with Computer Simulations. *Journal of Educational Technology & Society*, 18, pp.113–127.

Kalyuga, S., 2007b. Expertise Reversal Effect and Its Implications for Learner-Tailored Instruction. *Educational Psychology Review*, 19(4), pp.509–539.

Kalyuga, S., 2009. Managing cognitive load in adaptive ICT-based learning. *Journal of Systemics, Cybernetics and Informatics*, 7(5), pp.16–21.

Kalyuga, S. et al., 2003. The Expertise Reversal Effect. *Educational Psychologist*, 38(1), pp.23–31.

Kalyuga, S. & Singh, A.-M., 2016. Rethinking the Boundaries of Cognitive Load Theory in Complex Learning. *Educational Psychology Review*, 28(4), pp.831–852.

van der Kleij, F.M., Feskens, R.C.W. & Eggen, T.J.H.M., 2015. Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes: A Meta-Analysis. *Review of Educational Research*, 85(4), pp.475–511.

Lee, C.H. & Kalyuga, S., 2014. Expertise reversal effect and its instructional implications. In *Applying science of learning in education: Infusing psychological science into the curriculum*. Washington, DC: Society for the Teaching of Psychology, pp. 32–44.

Long, J.A., 2018. *jtools: Analysis and Presentation of Social Scientific Data*, R package version 1.0.0, https://cran.r-project.org/package=jtools.

Long, Y. & Aleven, V., 2017. Enhancing learning outcomes through self-regulated learning support with an Open Learner Model. *User Modeling and User-Adapted Interaction*, 27(1), pp.55–88.

van Merriënboer, J.J.G. & Sluijsmans, D.M.A., 2009. Toward a Synthesis of Cognitive Load Theory, Four-Component Instructional Design, and Self-Directed Learning. *Educational Psychology Review*, 21(1), pp.55–66.

van Merriënboer, J.J.G. & Sweller, J., 2005. Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review*, 17(2), pp.147–177.

Murray, M.C. & Pérez, J., 2015. Informing and performing: A study comparing adaptive learning to traditional learning. *Informing Science*, 18(1), pp.111–125.

Nakić, J., Granić, A. & Glavinić, V., 2015. Anatomy of student models in adaptive learning systems: A systematic literature review of individual differences from 2001 to 2013. *Journal of Educational Computing Research*, 51(4), pp.459–489.

Oxman, S. & Wong, W., 2014. White Paper: Adaptive Learning Systems. *DV X Innovations DeVry Education Group*.

Paas, F., Renkl, A. & Sweller, J., 2003. Cognitive Load Theory and Instructional Design: Recent Developments. *EDUCATIONAL PSYCHOLOGIST*, 38(1), pp.1–4.

Price, L., Casanova, D. & Orwell, S., 2017. Modeling an institutional approach to developing Technology Enabled Learning: Closing the gap between research and practice. *INTED2017 Proceedings*, pp.5009–5018.

R Core Team, 2013. R: a language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.*

Revelle, W.R., 2017. psych: Procedures for Personality and Psychological Research.

Rey, G.D. & Buchwald, F., 2011. The expertise reversal effect: Cognitive load and motivational explanations. *Journal of Experimental Psychology: Applied*, 17(1), pp.33–48.

Scanlon, E. et al., 2013. *Beyond prototypes: Enabling innovation in technology-enhanced learning*, Open University, Milton Keynes.

Scanlon, E., O'Shea, T. & McAndrew, P., 2015. Technology-Enhanced Learning : Evidence-based Improvement. In *L@S '15 Proceedings of the Second (2015) ACM Conference on Learning @ Scale*. Vancouver, BC, Canada: ACM New York, NY, USA, pp. 229–232.

Somyürek, S., 2015. The new trends in adaptive educational hypermedia systems. *International Review of Research in Open and Distributed Learning*, 16(1), pp.221–241.

Suleman, R.M., Mizoguchi, R. & Ikeda, M., 2016. A New Perspective of Negotiation-Based Dialog to Enhance Metacognitive Skills in the Context of Open Learner Models. *International Journal of Artificial Intelligence in Education*, 26(4), pp.1069–1115.

Sweller, J., 1988. Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, 12, pp.257–285.

Wauters, K., Desmet, P. & Van den Noortgate, W., 2010. Adaptive item-based learning environments based on the item response theory: possibilities and challenges. *Journal of Computer Assisted Learning*, 26, pp.549–562.

Wickham, H., 2009. *Ggplot2 : elegant graphics for data analysis*, Springer.

Chapter 6
Implementation of Adaptive Learning Systems: Current State and Potential

Christof Imhof, Per Bergamin, and Stéphanie McGarrity

6.1 Introduction

Countless aspects of our lives have become increasingly digitalized in the past few decades, learning being no exception. In the wake of digitalization, new forms of learning have emerged such as distance learning or technology-based learning, which are increasingly gaining importance today (Bergamin et al. 2012). Due to their flexible nature, these new forms of learning allow learners more independence and autonomy than ever before. Moreover, they overcome space-time barriers, thus granting many people the opportunity to pursue academic studies in circumstances that usually prevent or at least hinder such ambitions, e.g. full- or part-time employment or parenthood. Such flexibility allows for the inclusion of personal needs and contexts, which can differ considerably between individual learners. In higher education, such characteristics might be prior knowledge, learning skills, experience in regard to certain topics, use of strategies or affective states. Even with these differences, learners are usually expected to develop the same competences throughout their studies.

One way to achieve these comparable learning outcomes despite heterogeneous preconditions is to continuously adapt the learning process to the needs of the learners. This and related concepts can be covered under the umbrella term *adaptive learning*. In contrast to other technology-based learning approaches, adaptive learning enables the presentation of learning resources (e.g. content, support or navigation) in a dynamic form. This mostly occurs as a reaction to collected and evaluated data which can change during the learning processes, e.g. due to learning progress. In essence, adaptive learning systems continuously identify what a learner does or does not understand and provide help accordingly until a certain learning goal is

C. Imhof (✉) · P. Bergamin · S. McGarrity
Swiss Distance University of Applied Sciences, Brig, Switzerland
e-mail: christof.imhof@ffhs.ch; per.bergamin@ffhs.ch; stephanie.mcgarrity@ffhs.ch

P. Isaias et al. (eds.), *Online Teaching and Learning in Higher Education*,
Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-48190-2_6

met. This help can take different forms. One described by Oxman and Wong (2014) is the presentation of content situated just above the learner's current level in order to balance challenge and frustration. On this basis, adaptive learning has the potential to reduce dropout rates, lead to better learning outcomes and help students to achieve their learning goals faster. The notion of providing learners with assistance tailored towards their specific needs has a long history in pedagogy (e.g. in the form of one-to-one teacher support). However, technology-based adaptive learning systems provide forms of adaptivity beyond what can realistically be accomplished in traditional classroom settings in terms of resources or scale (cf. Koedinger et al. 2013).

The overall research problem addressed in this chapter is how the theoretical and conceptual foundation of an adaptive system needs to be specified in order for such a system to be implemented successfully in a university setting. This chapter aims to contribute the following to the discussion: We will first determine what it entails for a learning system to operate adaptively. In order to characterise the research in this area, we will then explore six basic questions in the design process of adaptive learning systems: why, what, what to, when, where and how a system can or should adapt (Brusilovsky 1996, 2001; Knutov 2012). We will also address the features and functions that are central to adaptive systems, followed by an overview over the current state of research in the area of adaptive learning. Practical implications and future potential of the research will also be discussed.

6.2 Definition of Adaptive Learning

Adaptive learning may be viewed from different theoretical and disciplinary perspectives, which is reflected in the definitions found in the literature. Depending on the perspective, the definitions may thus emphasise different elements. Jameson (2003), for example, approaches adaptivity from a computer science perspective and highlights the system's interactivity and its adaptation to different users based on user models (see below) as its core functionalities. He therefore defines a user-adaptive system as "an interactive system that adapts its behaviour to individual users on the basis of processes of user model acquisition and application that involve some form of learning, inference, or decision making" (p. 2). Interactivity and a focus on individual learners are elements also present in a more recent conceptualization by Aleven et al. (2017). In contrast to Jameson (2003), the authors argue from an educational point of view and further specify which kind of measure a system should base its adaptation upon. The authors identify three conditions a learning environment must meet in order to be considered adaptive. First off, its design needs to reflect topic-related challenges that learners often encounter. Secondly, the environment's pedagogical decision-making has to be based on psychological measures of individual learners (such as current knowledge, skills or affective states). Lastly, it is required to respond interactively to learner actions. All three of these aspects require data about learners, which are either pre-existing (condition 1) or collected and processed by the system (conditions 2 and 3).

In our view, these two definitions, although emphasising important learning-related components of adaptivity, do not explicitly address instructional aspects of adaptive learning. One element we deem crucial in this context is the monitoring of changes regarding the learners' progress. In our understanding, adaptive learning thus refers to technologies that monitor learning progress and repeatedly or continuously adapt the teaching process to the behaviours and needs of individual learners (see Adams Becker et al. 2018).

6.3 Core Components of Adaptive Learning Systems and Their Implementation

As indicated by the definitions of adaptive learning systems, there are certain elements that need to be accounted for when implementing such systems. Three core elements commonly found in adaptive learning systems, regardless of their degree of sophistication, are the domain model, the learner model and the adaptive model (cf. Vagale and Niedrite 2012). The *domain model* (also known as *content model* or *expert model*) refers to the content and structure of the topic to be taught, i.e. the relationships between the domain elements, and can address the intended learning outcomes as well as their sequence. The *learner model* (also known as *user model* or *student model*) is – as the name implies – a representation of the learner. The model consists of sensors and the learner modeller. The sensors capture and measure specific learner characteristics and pass the information to the learner modeller which then either uses the information as is (e.g. age, gender, prior knowledge) or further processes it (e.g. current knowledge, abilities, learning styles, motivational or emotional state). Depending on what characteristics the sensors measure, learner models can be either static or dynamic. While static models assess learner characteristics once, dynamic variants repeatedly measure and update them. In order for the learner model to be sound, the assessment of the learner characteristics (and the ensuing inferences) needs to be reliable and valid (see Shute and Towle 2003). The information from the sensors is in turn processed by the learner model and then further relayed to the *adaptive model* (also known as *adaptation model*, *instructional model*, *pedagogical model* or *tutoring model*). This model combines the processed information from the learner model with information from the domain model. The adaptive model can proceed to adapt content, instruction, or recommendations accordingly to support the learner in their progress. The model encompasses an instructional strategy that determines not only what can be adapted but also the context in which the adaptive process will occur.

Another way to look at adaptive learning systems is to focus on the design process. One way to characterise this process and its facets is by considering the six dimensions of the classic adaptive hypermedia approach (cf. Brusilovsky 1996): the

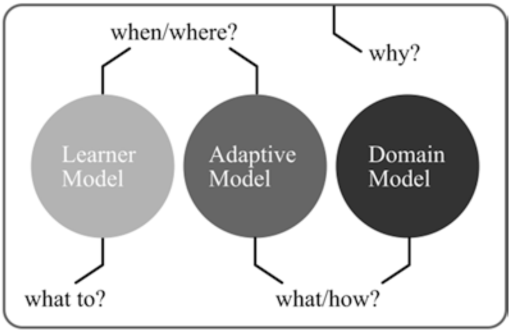

Fig. 6.1 Core components of adaptive learning systems and facets of the design process

goals, targets, sources, temporal contexts, situational contexts and methods/techniques of adaptation. These dimensions can be rephrased as the following six questions: *Why* is adaptation wanted? *What* can or should a system adapt? *What* can or should it adapt *to*? *Where* and *when* can it be applied? And *how* does the system adapt? These questions will be elaborated on in the following sections, starting with the *why* question. Due to similarities between them, some of the subsequent questions will be bundled, specifically the *when* and *where* questions that both concern the context of adaptation and the *what* and *how* questions which both address the adaptive model. The relation between the three core components and the six questions is illustrated in Fig. 6.1.

6.3.1 Why *Is Adaptation Wanted? The Reasons for and Goals of Adaptation*

The first didactic question for the development of adaptive learning objects or entire systems is why adaptation of learning to particular needs is even desired (Knutov 2012). On the one hand, it relates to the identification and fulfilment of user-related needs that require such methods and techniques in the first place (i.e. the goals of adaptation). Through adaptive learning, personal learning paths, assistance and advice, a variety of learning requirements can be met, which is difficult to achieve in traditional learning settings. For instance, uneven levels of prior knowledge between learners, which could lead to adverse effects (e.g. overwhelming inexperienced learners while simultaneously boring advanced learners), can be mitigated through adaptive instructional design. Another example is adaptive learning systems can support novices that require navigational help, e.g. by limiting the amount of alternatives or recommending relevant links (Brusilovsky 1996). On the other hand, this question concerns the course designers' motivation behind applying different adaptive methods and techniques (i.e. the reasons for adaptation). In principle, the *why* question thus concerns the pedagogical rationale underlying the implementation of adaptive systems (cf. Mavroudi et al. 2018). The pedagogical rationale itself can be derived from a variety of different basic theories, such as aptitude-treatment interactions, the zone of proximal development, fading scaffolds, the expertise reversal paradigm and self-regulated learning.

The concept of *aptitude-treatment interactions* (see Cronbach and Snow 1977) refers to the circumstance that instructional strategies (Cronbach and Snow refer to these as "treatments") are not equally successful for each individual learner and may instead depend on specific abilities of the learners that forecast their potential success – in other words, their *aptitude*. From this point of view, adaptive learning provides options to find optimal treatments to match individual learners' aptitudes. Another concept which adaptive learning can build on is the *zone of proximal development* (see Vygotsky 1978). The core idea of this concept is to give the learners tasks they are able to complete with guidance, as opposed to tasks they are able to do unaided or task they cannot complete even with guidance. As the learner progresses, this guidance can gradually be reduced (cf. the concept of *fading scaffolds*; Collins et al. 1988; van Merriënboer and Sluijsmans 2009). The importance of adapting the learning process to characteristics of the learner is further supported by the finding that instructional techniques (e.g. guidance by a tutor or detailed instructions) that benefit novices can lose their effectiveness or even be counterproductive to experts, a phenomenon known as the *expertise reversal effect* (Kalyuga et al. 2003).

In this context, "reversal" refers to the idea that the effectiveness of instructional techniques may be reversed for different levels of expertise, e.g. that instructions may help novices yet hinder experts (Lee and Kalyuga 2014). The expertise reversal effect is usually explained by the *Cognitive Load Theory* (Sweller 1988). The basis of the theory is the notion that the cognitive load, i.e. information that is currently stored and processed in the working memory, cannot exceed its limitations. While the long-term memory holds cognitive schemata with varying degrees of complexity within an unlimited storing capacity, the working memory is thought to be quite limited in its capacity to store information, both in terms of amount and duration (van Merriënboer and Sweller 2005). Classic accounts of the Cognitive Load Theory differentiate between two kinds of cognitive load, the intrinsic load and the extraneous load. *Intrinsic load* refers to cognitive processes involved in processing novel learning materials, which may be affected by the (perceived) complexity of the material. *Extraneous load* concerns factors that affect cognitive processes despite not being directly related to the task at hand, such as convoluted instructional design or unfavourable presentation of the learning material (Kalyuga 2009).

The two forms of cognitive load interact with one another so that an abundance of extraneous load (e.g. by giving learners too much unnecessary information or by having a cluttered visual design) reduces the capacity left for proper processing of the learning material due to the working memory's limitations. Importantly, the current cognitive load of a learner also depends on learner characteristics such as expertise. In parts, expertise is represented by cognitive schemata with varying degrees of complexity and automation housed by the long-term memory (van Merriënboer and Sweller 2005). When schemata become automated through training, space in the working memory is freed, which then reduces the intrinsic load, leaving more cognitive capacity for the processing of new content (Kalyuga 2009). This implies that instructional interventions should be adjusted (adapted) to the learners' cognitive load when teaching complex content (Rey and Buchwald 2011; Somyürek 2015). This may be achieved through instructional guidance: low levels of guidance or instructional scarcity can affect novices negatively as they might lack the expertise to compensate for the missing or incomplete information, which can lead to poor problem-solving strategies or mere guess work. Experts on the other hand are not affected as much since they can rely on their prior knowledge. When the amount of guidance is overabundant, the inverse effect may occur: novices benefit from the detailed instructions while experts' cognitive load is increased since they need to compare and contrast the flux of incoming information with their prior knowledge, inflating their intrinsic load (cf. Kalyuga 2007). Consequently, at the start of the learning process, novices should be provided with instructional guidance (e.g. step-by-step instruction) in order to guide them through their tasks and reach an optimal level of cognitive load. The concept of fading scaffolds applies here again (Collins et al. 1988; van Merriënboer and Sluijsmans 2009).

The educational implications of the Cognitive Load Theory and its role in explaining the expertise reversal effect have been explored and confirmed in numerous studies (e.g. Rey and Buchwald 2011). However, the cognitive load approach is limited to a specific learning goal in its application, namely, the acquisition of subject-specific knowledge (Kalyuga and Singh 2016). Other learning goals such as enhancing self-regulated learning are beyond the scope of the approach and may best be addressed by other theoretical perspectives within adaptive learning. *Self-regulated learning* refers to self-directive processes and motivational self-beliefs that learners use to proactively acquire academic skills (Zimmerman 2008). These skills include the setting of challenging goals, the employment of appropriate strategies to achieve these goals and the self-monitoring of one's activities and effectiveness until said goals are met. Adaptive learning environments can support self-regulated learning, e.g. by facilitating monitoring via continuous self-assessments and improving regulation of learning processes via instructional guidance (Scheiter et al. 2017).

These theories all provide guidelines for pedagogical decision-making. Despite representing vastly different perspectives, they are not mutually exclusive. The pedagogical strategies of adaptive learning systems can draw from multiple theoretical sources at once, e.g. by combining self-regulated learning with fading scaffolds.

6.3.2 What Can or Should Be Adapted and How? The Objects, Methods and Techniques of Adaptation

The next questions concern what can be adapted within a system to meet the guidelines illustrated above and how this may be accomplished. On one hand, the *what* question depends on the domain model since that model provides a structure of the topic also entailing which aspects can be adapted (see Knutov 2012). Brusilovsky (2001) suggests two aspects that can be adapted, namely, presentation and navigation support. *Adaptive presentation* focusses – as the name implies – on the presentation of the content in accordance with various learner characteristics (which will be discussed later). For example, a more experienced learner may be provided with less detailed instructions for a task, while novices may receive additional explanations to support their understanding of the topic. *Adaptive navigation support* is based on personalised learning paths that are supposed to guide the learner to appropriate learning content. Knutov (2012) adds a third approach in the form of *content adaptation support*, which addresses the presence or absence of specific bits of information, thus regulating their accessibility. This kind of support may also vary the emphasis that is put on the information. Other parts of the instructional design that can be adapted include hints, prompts and recommendations.

On the other hand, the *what* question also revolves around the adaptive model, as does the *how* question. How the adaptive process works can be described on two levels, either on a conceptual/design level or on an implementation level. The adaptive process involves *techniques*, which are usually applied at the implementation level of a system and adhere to specific approaches or algorithms, as well as *methods*, which are generalisations of techniques (Knutov 2012). Examples for techniques in content adaptation support include inserting, removing or modifying information, which change the accessibility of information, thus altering the content itself. Other techniques, which are also shared by adaptive presentation support, do not change the content but rather lead the learner to focus only on parts of the content. These include dimming, sorting, zooming or stretchtext (Knutov 2012). The latter two are also useful techniques when presenting information that only needs to be seen by a subset of learners. Techniques applied in the context of adaptive navigation support can either be enforced or recommended. These techniques include guidance (e.g. by recommending links, which can also be classified as an adaptive presentation support technique), link generation or link hiding (Knutov 2012).

The decision between enforced or recommended paths taps into the self-regulation dilemma, which concerns the amount of control that is given to the system versus the control given to its user (see Bergamin and Hirt 2018; Kobsa et al. 2001). On one end of the spectrum, learners are given complete control over their learning process (i.e. choice of topics, resources and support). Such systems are also called *adaptable* systems. The learner-control approach might entail positive consequences since freedom can be a motivating factor and learners may enjoy being in control. However, this level of freedom may also overwhelm and thus demotivate learners, especially at the beginning of the learning process, when learners lack self-regulation skills, or when a complex topic is concerned. On the other end of the spectrum, *adaptive* systems choose and present learning content, which may lead to decisions that are more sound than decisions that novices would make, but the lack of control on the learner's part may frustrate them, especially when the decisions by the system are faulty or not what the learner anticipates. This may be the case when the learner model is not accurate enough or when the learner's view is skewed. One way to bypass the dilemma is by allowing the control to be shared between the system and the learner, which is often achieved by implementing *recommender systems*. These systems offer learners recommendations or advice on how to adapt their learning process (e.g. by recommending tasks, supplementary material and so on) instead of forcing a system-made decision upon them. The learner is thus free to follow the recommendation or ignore it.

Since instructional interventions in this type of system are dependent on the learner's initiative, they are referred to as *non-embedded* (Clarebout and Elen 2006). A more *embedded* alternative exists in the form of the *two-step approach* (cf. Bergamin and Hirt 2018). In the first step, the system selects a set of appropriate learning objects (e.g. tasks), which the learner is then able to choose from. The main advantage of this approach is that learners can be prevented from being overwhelmed by countless options or from selecting counterproductive tasks while still being allowed to be in control, at least to a degree. Chou et al. (2015) present another option that allows simultaneous shared control between the system and the learner, the *negotiation-based adaptation mechanism*. This mechanism compares the system's learner model with the student's self-assessment, and if they do not match, modifications to the learner model will be "negotiated" between the learner and the system. It supports learners with low meta-cognitive skills while allowing learners to correct inaccurate learner models.

Moreover, methods and techniques applied in adaptive learning systems can vary substantially in terms of complexity and level of detail. A common distinction is made between *rule-based* and *algorithm-based systems* (Murray and Pérez 2015; cf. Oxman and Wong 2014). The former usually relies on a series of if-then functions with varying degrees of complexity (e.g. through different branching paths). If learners get answers right, the system directs them to the next task, and if they do not, it provides assistance in the form of a hint, repeated content or different explanations of the same content. Rule-based adaptive systems are transparent in their functionalities, which makes them easier to use; however, they do not tap into the computational potential that more sophisticated systems do. Algorithm-based approaches are far more complex and often involve methods related to machine learning, such as item-response theory (e.g. Wauters et al. 2010; Pliakos et al. 2019), Bayesian Knowledge Tracing (Corbett and Anderson 1995), fuzzy-logic (Ennouamani and Mahani 2019) or deep learning (Goodfellow et al. 2016). Additionally, they may involve elaborated techniques such as (big) data mining (e.g. Yuan 2019) or learning analytics in order to continuously predict the success of an individual learner based on specific bits of information. As Ge et al. (2019) note in their literature review, there is a tendency for adaptive systems to rely on established algorithms, rather than implementing game engines or developing their own algorithms.

A noteworthy example for algorithm-based approaches are *micro-adaptive systems* (Vandewaetere et al. 2011). Micro-adaptive systems are learning systems that employ micro-adaptive instructions that dynamically decide which instructional treatments are the most appropriate at any given time (e.g. intelligent tutoring systems). They accordingly provide tailored on-time instructions based on within-task measures. The fine-grained and precise measures this approach requires are thought to warrant the implementation of artificial intelligence techniques. However, this alleged necessity has attracted controversy since some authors, e.g. Essa (2016), argue that domain-specific micro-adaptivity should be regarded as “the primary realm of the instructor” (p. 11). The authors speculate that for the foreseeable future, machine learning will not surpass the instructor’s knowledge and experience, at least as far as providing feedback and correcting errors is concerned. We would like to emphasise that machine learning and the instructor’s experience are not mutually exclusive and may complement one another. Examples for this are supervised machine learning and co-creation strategies (see Dollinger and Lodge 2018).

6.3.3 What Can or Should Be Adapted to? The Basis of Adaptation

The fourth question concerns which characteristics of the learner should be captured by the sensor part of the learner model. As these characteristics form the basis for adaptive processes, they need to be selected carefully. What characteristics are most valuable in the context of a learning task, a course or even degree programmes to be adapted in regard to a particular goal is not a trivial question and has led to some disagreement in the literature (see Granić and Nakić 2010). In order to provide a potential answer, Nakić et al. (2015) conducted one of the most encompassing literature reviews regarding adaptation to learner characteristics. The authors explored 22 different learner characteristics over 98 studies released between 2001 and 2013, which include age, gender, working memory capacity, (meta-)cognitive abilities, anxiety and so on.

Given how wide the variety of characteristics to choose from is, several attempts have been made to categorise them. Vandewaetere et al. (2011) differentiate between three categories, which they derive from the combination of empirical research with theoretical propositions. These three categories are (1) cognition (working memory capacity, intelligence, prior knowledge, cognitive and learning styles), (2) affect (frustration, confusion, delight, mood and self-efficacy) and (3) behaviour (need for learner control, help and/or feedback, self-regulated learning, number of tries per task and grades). Although these categories seem to differ clearly, the boundaries between them are often blurred. The category *affect* includes states that are blends between affect and cognition (e.g. confusion and self-efficacy), while the characteristics in the *behaviour* category can be viewed as consequences of cognitive and affective states. Another classification stems from Aleven et al. (2017) who identify five groups of learner characteristics: prior knowledge and knowledge growth; strategies and errors; affect and motivation; self-regulated learning strategies, metacognition and effort; and learning styles. As they note, determining which characteristics are worth adapting to the most is ultimately an empirical question. Based on the results of the studies that Nakić et al. (2015) examined, the authors conclude that adapting to one or more of the following characteristics proves to be the most successful: learning styles, prior knowledge, cognitive styles, preferences for particular types of learning materials and motivation. The latter is noted to have been subject to increasing attention in research, along with characteristics such as emotions and metacognitive abilities (Nakić et al. 2015). Adapting to cognitive abilities and personality is also deemed promising, although those characteristics have been explored to a lesser degree (see, e.g. Afini Normadhi et al. 2019). Further details will be provided in the section discussing the current state of the research.

6.3.4 When and Where Can Adaptation Be Applied? The Context of Adaptation

Knowing on which pedagogical basis we want to adapt what aspects to which characteristics with which techniques, the final questions are when and where adaptation takes place. One way to answer both of these questions at once is by addressing loop levels, which determine when and where instructions can be varied within the adaptive model. According to Bergamin and Hirt (2018), there are three levels on which adaptation can occur: the curriculum loop, the task loop and the step loop. In the *curriculum loop*, the adaptive system recommends (or enforces) learning domains (curricula) based on the learners' needs and preconditions. This can be illustrated with an example: A learner succeeds in a particular course and may thus be recommended an advanced course on the same topic. Since it concerns in-between-course adaptation, the curriculum loop only occasionally adapts to the learner model.

In the *task loop* (also known as *outer loop*), the system makes decision regarding the instructional support, complexity of the content or sequencing (i.e. task selection) depending on the individual learner's current conditions. An adaptive system may thus recommend (or enforce) more challenging tasks to successful learners while presenting tasks that involve more assistance to less proficient learners. Since it concerns tasks, the task loop adapts to the learner model more frequently than the curriculum, but less frequently than the step loop. In the *step loop* (also known as *inner loop*), the system provides hints, feedback and prompts regarding the current learning activity within a learning object (e.g. a task). This adaptation depends on the individual learner's most recent learning behaviour. Aleven et al. (2017) also differentiate between three loop levels; but instead of the curriculum loop, they include a *design loop* in their conceptualisation. Design-loop adaptivity refers to data-driven changes between different iterations of the same course on the basis of similarities between learners. For example, a course designer may receive the feedback that a high percentage of students displayed the same misconception in a physics task, which leads to them accounting for that misconception in the next version of the course. In contrast to the other loops, this loop does not concern the individual learner and takes on a different perspective (namely, that of a course designer charged with redesigning an existing course).

The *when* and *where* questions can further be addressed by considering another aspect of adaptive systems, namely, their application area. While e-learning remains the main application area of adaptive learning, its range has expanded significantly over the years. Adaptive learning systems are applied in various educational institutions (primary school, secondary school, senior school, university, etc.) as well as organisations, e.g. for training purposes. Moreover, there has been an increase in context-aware adaptive systems that try to incorporate context characteristics in addition to learner characteristics, e.g. the time and place of a learning activity or the device used by the learner. This can be achieved by either expanding the learner model or adding a fourth model to the three core components (for instance, a *context model*; see Knutov 2012).

6.4 Current State of the Research

In this next part, we will concentrate on three aspects of current application-oriented research: the evaluation of the effectiveness and efficiency of adaptive learning systems, the satisfaction of learners with such systems and their actual implementation. We highlight application-oriented research over theoretical literature to emphasise the practical implementation of adaptive learning systems.

6.4.1 Learner Performance: Effectiveness and Efficiency of Adaptive Learning Systems

Instructional effectiveness and efficiency are key aspects of adaptive learning since optimising learning is one of the central objectives of this approach (Sottilare and Goodwin 2017). *Instructional effectiveness* refers to enhancing learning capacity to acquire knowledge or skill. Importantly, the time in which this learning gain is supposed to transpire is fixed and the learning content is varied, so that at the end of the course, learners may be below, at or above their expected level (Sottilare and Goodwin 2017). In contrast, *instructional efficiency* refers to the acceleration of learning, which means a reduction of the time learners need to reach a desired level of knowledge or skill. By providing learners with instruction tailored to their needs (e.g. based on their current level of knowledge), the amount of information they are presented with can be reduced. However, allowing learners to skip information is not always recommended since learning materials may need to be revisited from time to time to retain proficiency (Sottilare and Goodwin 2017). Adaptive learning reveals its potential addressing both of these points, as it permits a large variety of learning materials and instructional strategies to be tailored to the needs of individual learners. Effectiveness and efficiency depend, among other things, on the context of the deployment of adaptive learning, higher education being by far the most common context (see Xie et al. 2019, for an overview).

One part of the literature concerns the effectiveness and efficiency of adaptive learning systems. This line of research is concerned with the research question how effective and efficient adaptive learning systems are, usually in comparison to either non-adaptive alternatives or other adaptive systems with diverging features. Accordingly, most researchers hypothesise that adaptive learning systems are more effective and efficient than their non-adaptive counterparts. While some studies have assessed both effectiveness and efficiency of adaptive learning systems, others have focussed on one of these two performance measures. Verdú et al. (2008), for example, examined the evidence for the effectiveness of adaptive learning by comparing studies that analysed adaptive systems in various institutional contexts. They found that with varying levels of statistical significance and effect sizes, all 18 of the studies in their pool reported positive results, i.e. students improved in their academic achievement when using adaptive systems in comparison to control groups. The variation between effect sizes indicates a vast range of effects. One study yielded an effect size of 0.1, which indicates a small, statistically not significant learning gain. Large effects (i.e. effect sizes of at least 0.66) were found in ten of the studies, with the remainder yielding medium to small effects. Further studies show that the results concerning the effectiveness and efficiency of adaptive learning are rather mixed: while there is evidence to suggest that the implementation of adaptive learning can lead to improved achievements, higher self-perceived learning gains and reduced cognitive load (e.g. Yang et al. 2013), other studies were only able to detect positive effects on learning outcomes under specific conditions. In their evaluation of an adaptive online learning system, Griff and Matter (2013) only found positive effects in two out of the six participating institutions. Similarly, Murray and Pérez (2015), who implemented a micro-level adaptive approach, only found a negligible impact of adaptive learning on learning outcomes when compared to a traditional non-adaptive approach. In a recent experimental classroom study, Eau et al. (2019) did not find any significant impact of adaptive learning on exam scores, course grades or progress. In contrast, Ghergulescu et al. (2016), who conducted a field study with a total sample size of 10,000 students across 1700 mathematics sessions, report significant improvements across ability levels (i.e. ranging from low to high achievers). Low achievers improved more than high achievers, thus reducing the achievement gap.

Another part of the literature addresses effectiveness and efficiency in relation to the temporal context the systems operate in as well as the learner characteristics their learner model is based on. Here we will illustrate this based on the findings by Aleven et al. (2017), who evaluated the effectiveness of adapting to various learner characteristics by systematically reviewing studies that either addressed design-loop, task-loop or step-loop adaptations to learner characteristics stemming from their previously presented five categories (prior knowledge, strategies and errors, affect and motivation, self-regulation of learning and learning styles). Since we do not consider design-loop adaptivity to be on the same dimension as the task and step loops as explained above, we will only include the latter two in our overview.

First off, Aleven et al. (2017) present evidence to support the effectiveness of adapting to prior knowledge. Evidence on the task-loop adaptivity suggests that adapting the task selection to the learners' prior knowledge improves both effectiveness and efficiency of learning. Corbett et al. (2000), for instance, observed that students scored twice as high in the assessment of an algebra problem and 10% higher in a standard test when using the Cognitive Tutor Algebra I in comparison to traditional courses. Cognitive Tutors are intelligent tutoring systems that present tasks which train aspects students are unlikely to have mastered yet. Comparable results have been achieved by promoting learning by analogue problem-solving, where students solve problems by transferring knowledge from an analogue, adaptively selected example (cf. Muldner and Conati 2007). Increased learning gains were also observed when examining step-loop adaptivity, even though the evidence is not quite as abundant in this context. Conati (2013), for example, reported larger learning gains after implementing a self-explanation coach for physics problem-solving (i.e. a system that adaptively selected steps of worked examples and provided a structure template as well as feedback). This effect was larger for students with low levels of prior knowledge, which is also what Albacete and VanLehn (2000) observed. The opposite was found by Own (2006): in his study, the difference in learning progress was only significant for students that had more prior knowledge. E. Verdú et al. (2008) identified differences in contexts, systems and analyses between the studies as the most likely cause for this discrepancy.

Overall, Aleven et al. (2017) note that the evidence supporting the value of adapting to prior knowledge is consistent with the widespread notion that learners' prior knowledge is a key factor in learning. In fact, the authors assert that adapting to prior knowledge within the task-loop yielded the largest effects out of all the possible combinations between the learner characteristics and loops they examined.

Adapting to learners' affect was also found to improve effectiveness and efficiency. An example concerning task-loop adaptivity is a study by Walkington (2013), who implemented interest in her tutoring system by adapting the cover stories of algebra problems to students' interests. This resulted in higher accuracy and increased learning efficiency in the course and led to accelerated learning later on. Regarding the step loop, affect-aware tutoring systems were found to enhance learning. Examples include studies by D'Mello et al. (2010), who used *AutoTutor*, a system capable of detecting boredom, confusion, frustration and neutral affective states, or D'Mello et al. (2012), who implemented eye-trackers in their tutoring system in order to detect and adaptively counteract disengagement. Some systems even feature hybrid adaptivity, i.e. algorithms that combine affective with cognitive factors (e.g. Mazziotti et al. 2015). In contrast, Aleven et al. (2017) note that research focussed on adapting to learners' motivation has been comparatively scarce with only the groundwork being laid, e.g. in the form of self-efficacy-detecting algorithms using machine-learning models (McQuiggan et al. 2008).

Task-loop adaptivity to self-regulation can be effective as well, even though the evidence seems to be mixed. The most promising approach appears to be a combination between open learner models (i.e. a representation of the learner characteristics used by the system, often presented to the learner in a visual form) and self-assessment support (cf. Arroyo et al. 2014; Long and Aleven 2013). There is also evidence to suggest that adapting to self-regulated learning yields positive results in the step loop by improving learners' self-regulated learning processes (e.g. help-seeking, Tai et al. 2013).

In contrast, the evidence for the effectiveness and efficiency of adapting to learners' learning strategies and error patterns is mixed (Aleven et al. 2017). While step-loop adaptivity to strategies and errors is also deemed effective, particularly when applied in the form of step-level feedback (see Koedinger and Aleven 2007), the evidence presented by Aleven et al. (2017) does not support any clear advantage of task-loop adaptivity over non-adaptive tutoring. Adapting to learning styles also yielded little conclusive evidence, despite the popularity of the concept in past and present research (e.g. Kolekar et al. 2019). Many researchers argue that learning styles lack a firm theoretical basis (e.g. Aleven et al. 2017; Kirschner and van Merriënboer 2013; Lu et al. 2003), an issue that is further compounded by other controversies surrounding the topic, with some researchers even dismissing them as a "myth" (see Kirschner 2017).

A learner characteristic not present in the overview presented by Aleven et al. (2017) that was recently investigated was aptitude (Eldenfria and Al-Samarraie 2019). In their study, Eldenfria and Al-Samarraie (2019) found their aptitude-based adaptive mechanism to be effective, which was supported by EEG data.

Current research thus shows that adaptive learning can be both effective and efficient, be it in general or addressing specific temporal contexts (i.e. loops) or learner characteristics. The effects found in the literature may vary in their size from no effect to large effects, but all reported effects are positive, supporting the potential for future research.

6.4.2 Satisfaction Among Learners

Effectiveness and efficiency are not the only measures to indicate the success of a learning system. No matter how effective a system is, the prospects of success are jeopardised if students and/or teachers reject it. Assessing student satisfaction is therefore key when judging the quality of a system. Moreover, studies have shown positive links between student satisfaction and motivation, student retention and recruitment (see Schertzer and Schertzer 2004). Levy (2007) additionally shows that dropouts occur at substantially higher rates in e-learning as compared to offline courses, stressing the importance of student satisfaction for student retention. The research question that guides this strand of research is thus how satisfied students are with adaptive learning systems. Usually, students are hypothesised to feel satisfied with adaptive learning systems. Verdú et al. (2008) compared the results of 11 studies that assessed the level of students' satisfaction with adaptive learning systems via questionnaires. Since the results were based on questionnaires with different scales, the values were normalised before the comparison. One study reported medium (0.5) and the others high learner satisfaction (0.66–0.81) with adaptive learning systems. They conclude that most learners thought that the adaptive systems supported their learning progress and met their requirements.

In a more recent study, Dziuban et al. (2017) investigated how students from two contextually different universities reflected on the adaptive learning platform *Realizeit*. Despite differing in demographic and educational backgrounds, most students reacted positively to the adaptive system by giving it high marks regarding its perceived educational effectiveness and were able to make a near-seamless transition from non-adaptive systems. However, there are certain conditions that have to be met in order for learners not to reject adaptive systems. If systems are unstable, unreliable, too cumbersome in their use or plagued by usability problems, the risk of students (and teachers) abandoning it rises. Lack of transparency is an additional risk factor that can lead to trust issues (e.g. when the system is perceived as a "black box" without any comprehensible rationale behind its decisions; see Khosravi et al. 2020).

Assessing the usability of adaptive systems is therefore worthwhile (cf. Khosravi et al. 2020). Alshammari et al. (2015), for example, compared an adaptive learning system with a non-adaptive version in an experimental setting and found that the adaptive learning system yielded higher ratings regarding its perceived usability than its non-adaptive counterpart. Similarly, Vesin et al. (2018) examined the usability of the adaptive learning system *ProTuS* using the *System Usability Scale* (*SUS*). The resulting score was 67.2 out of 100, indicating a marginally acceptable usability, i.e. on the verge of being acceptable (with a score of 70 being the threshold). More recently, a German translation of the SUS was used to assess the usability of adaptive courses in the learning management system *Moodle* (Pancar et al. 2019). In contrast to previous results, the adaptive courses yielded lower usability scores (55.08 and 57.8) than their non-adaptive counterparts (62.87 and 67.51), meaning their usability was "ok".

As the research above illustrates, adaptive learning systems tend to be satisfying to learners, which is an important condition for the success of such systems. However, research on their usability opened up a clear gap which needs to be further addressed. Given how crucial usability is to the acceptance of adaptive learning systems, improving it is a key challenge.

6.4.3 Implementation of Adaptive Learning Systems

Another avenue of research within adaptive learning concerns the actual implementation of adaptive systems in practice, providing potential answers to the *how* and *when*/*where* dimensions. The research questions in this area are thus if adaptive learning systems can be successfully implemented in educational practice and under what conditions. Despite the wealth of studies on adaptive learning systems, there has been a notable lack of successfully implemented adaptive technology-based learning systems in practice (Cavanagh et al. 2020; Somyürek 2015), with a few exceptions, e.g. the previously mentioned study by Ghergulescu et al. (2016). Scanlon et al. (2013) found what they called a "surprising failure" (p. 4) to translate research results in the field of technology-enhanced learning, including prototypes, into commercial products. This gap between research and successful application is the so-called valley of death, which can be caused by a lack of funding, weaknesses in the didactic concept, scalability-related issues, inaccuracies in the core components or lack of sustainability.

Moreover, as Lerís et al. (2017) point out, technological issues are a contributing factor as well since one of the main reasons why some adaptive systems have failed is a lack of easy-to-use technology for the teachers meant to design adaptive tasks and instructions. Instructors that produce and follow sound instructional designs are essential to adaptive learning, which is why it is key to involve them from the very beginning (cf. Shelle et al. 2018). One potential solution is to implement adaptive learning within environments that teachers are already familiar with, such as learning management systems (e.g. *Moodle*). In one of our own studies, we demonstrate how a simple rule-based adaptive design based on a recommender system can be implemented in a physics course on Moodle (see Imhof et al. 2018). Our system recommended tasks with either detailed or non-detailed instructions to our students, depending on their current level of knowledge (i.e. a prior knowledge test score for the first task and task performance for the remainder of the task set). We deemed the implementation successful enough to serve as a good basis for future, more complex adaptive instructional designs in the same or similar contexts.

6.5 Practical Implications and Future Potential of Adaptive Learning Systems

The results presented above have practical implications for designing and implementing adaptive learning systems. In this discussion, we will refer to the six questions introduced in the beginning of this chapter again. *Why* is adaptation wanted? Research reveals arguments for the implementation of adaptive learning systems by demonstrating effectiveness and efficiency. *Where* and *when* can adaptation be applied? Adaptive learning systems have yielded positive effects in a variety of different contexts, be it in terms of institutions, the topics to be learned (despite the noticeable focus on STEM topics, especially in the realm of micro-adaptivity; cf. Essa 2016), the target audience or the loop levels within the adaptive model. *What* can or should it adapt *to*? Not all options are equally recommendable in regard to learner characteristics. For instance, the evidence for adapting to learning styles is mixed at best (cf. Aleven et al. 2017), despite their popularity. Importantly, no matter which learner characteristics are chosen, they need to be assessed reliably and validly in order for the system to adapt to the learners' needs accurately. *What* can or should it adapt and *how*? In contrast to the other questions, these two are difficult to answer on the basis of the literature we considered. To our knowledge, systems usually follow one specific approach in terms of methods and techniques and stick to them. This renders direct, unbiased comparisons with other approaches nigh impossible, since the list of potential confounding variables is vast (e.g. learning support, learning topics, educational contexts, outcome variables, learning devices, differences between learners and so on; cf. Xie et al. 2019).

Moreover, adaptivity on its own is no guarantee for success. In our view, the success of an adaptive system is instead linked to three crucial elements of its adaptive design, each addressing multiple of the six basic questions:

1. The concept behind an adaptive learning system needs to be specific and sound. Adaptive learning has unique requirements and is, as Freda (2016) states, not a "magic bullet". The quality of the adaptive design (and thus the recommendations a system makes) depends on the monitoring and diagnosis of changing learning requirements, which could result in insufficient adaptation rules if neglected (cf. Dounas et al. 2019).
2. The loop level the system operates on has to be specified. As Essa (2016) notes, a considerable amount of research has been dedicated to the inner loop (i.e. step loop or micro-adaptivity), whereas research on the outer loop (i.e. task loop or macro-adaptivity) has been described as "modest" (Rus et al. 2013).
3. Special care ought to be given to the algorithms behind the adaptive learning system. Most systems rely on existing algorithms (cf. Ge et al. 2019) which are not necessarily the ideal solution in every individual case.

In summary, adaptive systems need a concise concept behind them as well as a suitable adaptive mechanism supported by the proper algorithms. Differences in these three design elements could explain why some studies found adaptive learning systems to be effective (Eldenfria and Al-Samarraie 2019; Ghergulescu et al. 2016) while others did not (Eau et al. 2019) or had mixed results (Griff and Matter 2013). This is especially important when estimating the effectiveness of adaptive learning systems in practice.

Furthermore, the results illustrated above highlight that usability should be a major focal point when designing and implementing such systems (Khosravi et al. 2020). Systems burdened with usability problems satisfy neither learners nor teachers, increasing the risk of systems being swiftly abandoned.

As our overview depicts, the processes of designing and implementing adaptive learning systems are very complex since there are countless options one could choose when designing adaptive systems. Not only are these processes non-linear since the questions inform and influence each other; there is also a notable lack of guidance for them, at least currently (Hou and Fidopiastis 2017).

All in all, the practical implications of adaptive learning are somewhat limited at the moment since there are still various challenges that adaptive learning systems have to overcome in order to truly bridge the gap between research prototypes and application tools. In their Delphi study, Mirata and Bergamin (2019) identified three dimensions of the challenges for adaptive learning: technology; teaching and learning; and organisation. In the dimension *technology*, the challenges are *infrastructure and hard- and software*, which include the usability of adaptive learning systems, and *perceptions and beliefs about adaptive technology*, e.g. acceptance and attitude towards technology, both from the lecturers' and students' points of view. In the context of the dimension *teaching and learning*, the identified challenges are *instructional and curriculum elements* (e.g. the need to redesign courses) as well as *lecturer and learner characteristics* (e.g. their motivation and commitment). The final dimension, *organisation*, contains *institutional strategies* (including commitment on the part of the management), *management* (e.g. support for lecturers and learners) and *resources* (e.g. the hiring of instructional designers). Further challenges are identified by Zliobaite et al. (2012), who present additional technological challenges, and Freda (2016), who highlights the organisational challenges. Zliobaite et al. (2012) add scalability and having to deal with "realistic data" as additional challenges for technology. In order to improve usability, trust and acceptance, they state that the practical application of adaptive learning systems might have to be broken down into adaptive tools that non-experts are also able to use. This latter point is also stressed by Cavanagh et al. (2020), who include understanding of the mechanism behind adaptive learning systems as one of the items on their list of pedagogical best practices.

Similar to Mirata and Bergamin (2019), Freda (2016) stresses securing monetary resources and convincing parties other than students and teachers of the value of adaptive learning (e.g. project managers and instructional technologists) as two important obstacles when transitioning from traditional to adaptive learning systems.

Future research has the potential to address most if not all of these issues, thereby getting closer to bridging the gap between research and application, potentially leading to widespread successful implementations of adaptive learning systems. As the research presented in this chapter shows, adaptive learning systems hold considerable potential to improve scalability (i.e. reaching more learners with less effort) and learners' performance. This complex development is still ongoing, but the current state of the research indicates great promise for the future.

References

Adams Becker, S., Brown, M., Dahlstrom, E., Davis, A., DePaul, K., Diaz, V., & Pomerantz, J. (2018). *NMC horizon report 2018: higher education edition* (p. 60). Retrieved from https://library.educause.edu/{~}/media/files/library/2018/8/2018horizonreport.pdf

Afini Normadhi, N. B., Shuib, L., Md Nasir, H. N., Bimba, A., Idris, N., & Balakrishnan, V. (2019). Identification of personal traits in adaptive learning environment: Systematic literature review. *Computers and Education, 130*, 168–190. https://doi.org/10.1016/j.compedu.2018.11.005.

Albacete, P., & VanLehn, K. (2000). Evaluating the effectiveness of a cognitive tutor for fundamental physics concept. *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society* (pp. 25–30). Mahwah, NJ: Erlbaum.

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2017). Instruction based on adaptive learning technologies. In R. E. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 522–560). New York: Routledge.

Alshammari, M., Anane, R., & Hendley, R. J. (2015). Design and usability evaluation of adaptive e-learning systems based on learner knowledge and learning style. In *Human-computer interaction – Interact 2015* (Vol. 9297, pp. 584–591). https://doi.org/10.1007/978-3-319-22668-2.

Arroyo, I., Woolf, B.P., Burelson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education, 24*(4), 387–426.

Bergamin, P. B., & Hirt, F. S. (2018). Who's in charge? – Dealing with the self-regulation dilemma in digital learning environments. In K. North, R. Maier, & O. Haas (Eds.), *Knowledge management in digital change – New findings and practical cases*. Cham: Springer.

Bergamin, P. B., Werlen, E., Siegenthaler, E., & Ziska, S. (2012). The relationship between flexible and self-regulated learning in open and distance universities. *International Review of Research in Open and Distance Learning, 13*(2), 101–123. https://doi.org/10.19173/irrodl.v13i2.1124.

Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction, 6*(2–3), 87–129. https://doi.org/10.1007/BF00143964.

Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction, 11*(1/2), 87–110. https://doi.org/10.1023/A:1011143116306.

Cavanagh, T., Chen, B., Lahcen, R. A. M., & Paradiso, J. (2020). Constructing a design framework and pedagogical approach for adaptive learning in higher education: A practitioner's perspective. *The International Review of Research in Open and Distributed Learning, 21*(1), 172–196. https://doi.org/10.19173/irrodl.v21i1.4557.

Chou, C.-Y., Lai, K. R., Chao, P.-Y., Lan, C. H., & Chen, T.-H. (2015). Negotiation based adaptive learning sequences: Combining adaptivity and adaptability. *Computers & Education, 88*, 215–226. https://doi.org/10.1016/J.COMPEDU.2015.05.007.

Clarebout, G., & Elen, J. (2006). Tool use in computer-based learning environments: Towards a research framework. *Computers in Human Behavior, 22*(3), 389–411. https://doi.org/10.1016/J.CHB.2004.09.007.

Collins, A., Brown, J. S., & Newman, S. E. (1988). Cognitive apprenticeship. *Thinking: The Journal of Philosophy for Children, 8*(1), 2–10. https://doi.org/10.5840/thinking19888129.

Conati, C. (2013). Modeling and scaffolding self-explanation across domains and activities. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 367–383). New York: Springer. https://doi.org/10.1007/978-1-4419-5546-3_24.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing. *User Modeling and User-Adapted Interaction, 4*(4), 253–278.

Corbett, A. T., McLaughlin, M., & Scarpinatto, K. C. (2000). Modeling student knowledge: Cognitive tutors in high school and college. *User Modeling and User-Adapted Interaction, 10*(2/3), 81–108. https://doi.org/10.1023/A:1026505626690.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.

D'Mello, S. K., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., et al. (2010). A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. https://doi.org/10.1007/978-3-540-69132-7.

D'Mello, S. K., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human Computer Studies, 70*(5), 377–398. https://doi.org/10.1016/j.ijhcs.2012.01.004.

Dollinger, M., & Lodge, J. M. (2018). Co-creation strategies for learning analytics. In *Proceedings of the 8th international conference on learning analytics and knowledge – LAK'18* (pp. 97–101). New York: ACM Press. https://doi.org/10.1145/3170358.3170372.

Dounas, L., Salinesi, C., & El Beqqali, O. (2019). Requirements monitoring and diagnosis for improving adaptive e-learning systems design. *Journal of Information Technology Education: Research, 18*, 161–184. https://doi.org/10.28945/4270.

Dziuban, C., Moskal, P., Johnson, C., & Evans, D. (2017). Adaptive learning: A tale of two contexts. *Current Issues in Emerging eLearning, 4*(1). Retrieved from https://scholarworks.umb.edu/ciee/vol4/iss1/3/

Eau, G., Judah, K., & Shahid, H. (2019). How can adaptive platforms improve student learning outcomes? A case study of open educational resources and adaptive learning platforms. http://dx.doi.org/10.2139/ssrn.3478134

Eldenfria, A., & Al-Samarraie, H. (2019). The effectiveness of an online learning system based on aptitude scores: An effort to improve students' brain activation. *Education and Information Technologies, 24*(5), 2763–2777. https://doi.org/10.1007/s10639-019-09895-2.

Ennouamani, S., & Mahani, Z. (2019). Towards adaptive learning systems based on fuzzy-logic. In *Advances in intelligent systems and computing* (Vol. 997, pp. 625–640). Cham: Springer. https://doi.org/10.1007/978-3-030-22871-2_42.

Essa, A. (2016). A possible future for next generation adaptive learning systems. *Smart Learning Environments, 3*. https://doi.org/10.1186/s40561-016-0038-y.

Freda, B. (2016). *Clearing the hurdles to adaptive learning*. Retrieved from https://universitybusiness.com/clearing-the-hurdles-to-adaptive-learning/

Ge, Z., Xi, M., & Li, Y. (2019). A literature review of the adaptive algorithms adopted in adaptive learning systems. In *2019 IEEE 4th international conference on signal and image processing, ICSIP 2019* (pp. 254–258). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/SIPROCESS.2019.8868893.

Ghergulescu, I., Flynn, C., & O'Sullivan, C. (2016). Learning effectiveness of adaptive learning in real world context. In *EdMedia+ innovate learning* (pp. 1391–1396). Waynesville: Association for the Advancement of Computing in Education (AACE).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.

Granić, A., & Nakić, J. (2010). Enhancing the learning experience: Preliminary framework for user individual differences. In (pp. 384–399). Berlin/Heidelberg: Springer. https://doi.org/10.1007/978-3-642-16607-5_26.

Griff, E. R., & Matter, S. F. (2013). Evaluation of an adaptive online learning system. *British Journal of Educational Technology, 44*(1), 170–176. https://doi.org/10.1111/j.1467-8535.2012.01300.x.

Hou, M., & Fidopiastis, C. (2017). A generic framework of intelligent adaptive learning systems: From learning effectiveness to training transfer. *Theoretical Issues in Ergonomics Science, 18*(2), 1–30.

Imhof, C., Bergamin, P. B., Moser, I., & Holthaus, M. (2018). Implementation of an adaptive instructional design for a physics module in a learning management system. In *Proceedings of the 15th international conference on cognition and exploratory learning in the digital age, CELDA 2018*.

Jameson, A. (2003). Adaptive interfaces and agents. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction* handbook (pp. 305–330). Hillsdale: Lawrence Erlbaum. Retrieved from https://chusable.com/archives/homepage-2017/pdf/hci-handbook.jameson.pdf

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*(4), 509–539. https://doi.org/10.1007/s10648-007-9054-3.

Kalyuga, S. (2009). Instructional designs for the development of transferable knowledge and skills: A cognitive load perspective. *Computers in Human Behavior, 25*(2), 332–338. https://doi.org/10.1016/j.chb.2008.12.019.

Kalyuga, S., & Singh, A.-M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 28*(4), 831–852. https://doi.org/10.1007/s10648-015-9352-0.

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.

Khosravi, H., Gasevic, D., & Sadiq, S. (2020). Development and adoption of an adaptive learning system. https://doi.org/10.1145/3328778.3366900.

Kirschner, P. A. (2017). Stop propagating the learning styles myth. *Computers & Education, 106*, 166–171. https://doi.org/10.1016/J.COMPEDU.2016.12.006.

Kirschner, P. A., & van Merriënboer, J. J. G. (2013). Do learners really know best? Urban legends in education. *Educational Psychologist, 48*(3), 169–183. https://doi.org/10.1080/00461520.2013.804395.

Knutov, E. (2012). *Generic adaptation framework for unifying adaptive web-based systems*. Eindhoven: Technische Universiteit Eindhoven. https://doi.org/10.6100/IR732111

Kobsa, A., Koenemann, J., & Pohl, W. (2001). Personalised hypermedia presentation techniques for improving online customer relationships. *The Knowledge Engineering Review, 16*(2), 111–155. https://doi.org/10.1017/s0269888901000108.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*(3), 239–264. https://doi.org/10.1007/s10648-007-9049-0.

Koedinger, K. R., Brunskill, E., Baker, R. S. J., McLaughlin, E. A., & Stamper, J. C. (2013). New potentials for data-driven intelligent tutoring system development and optimization. *AI Magazine, 34*(3), 27. https://doi.org/10.1609/aimag.v34i3.2484.

Kolekar, S. V., Pai, R. M., & Manohara Pai, M. M. (2019). Rule based adaptive user interface for adaptive e-learning system. *Education and Information Technologies, 24*(1), 613–641. https://doi.org/10.1007/s10639-018-9788-1.

Lee, C. H., & Kalyuga, S. (2014). Expertise reversal effect and its instructional implications. In *Applying science of learning in education: Infusing psychological science into the curriculum* (pp. 32–44). Washington, DC: Society for the Teaching of Psychology. Retrieved from http://psycnet.apa.org/record/2013-44868-003

Lerís, D., Sein-Echaluce, M. L., Hernández, M., & Bueno, C. (2017). Validation of indicators for implementing an adaptive platform for MOOCs. *Computers in Human Behavior, 72*, 783–795. https://doi.org/10.1016/j.chb.2016.07.054.

Levy, Y. (2007). Comparing dropouts and persistence in e-learning courses. *Computers & Education, 48*(2), 185–204. https://doi.org/10.1016/J.COMPEDU.2004.12.004.

Long, Y., & Aleven, V. (2013). Skill diaries: Improve student learning in an intelligent tutoring system with periodic self-assessment. In *International Conference on Artificial Intelligence in Education* (pp. 249–258). Springer, Berlin, Heidelberg.

Lu, J., Yu, C.-S., & Liu, C. (2003). Learning style, learning patterns, and learning performance in a WebCT-based MIS course. *Information & Management, 40*(6), 497–507. https://doi.org/10.1016/S0378-7206(02)00064-2.

Mavroudi, A., Giannakos, M., & Krogstie, J. (2018). Supporting adaptive learning pathways through the use of learning analytics: Developments, challenges and future opportunities. *Interactive Learning Environments, 26*(2), 206–220. https://doi.org/10.1080/10494820.2017.1292531.

Mazziotti, C., Holmes, W., Wiedmann, M., Loibp, K., Rummel, N., Mavrikis, M., et al. (2015). Robust student knowledge: Adapting to individual student needs as they explore the concepts and practice the procedures of fractions. *CEUR Workshop Proceedings, 1432*, 32–40.

McQuiggan, S. W., Mott, B. W., & Lester, J. C. (2008). Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *User Modeling and User-Adapted Interaction, 18*(1–2), 81–123. https://doi.org/10.1007/s11257-007-9040-y.

Mirata, V., & Bergamin, P. B. (2019). Developing an implementation framework for adaptive learning: A case study approach. In *Proceedings of the 18th European conference on e-learning*. Copenhagen, Denmark, 7–8 November 2019.

Muldner, K., & Conati, C. (2007). Evaluating a decision-theoretic approach to tailored example selection. In *IJCAI 2007, proceedings of the 20th international joint conference on artificial intelligence*. Hyderabad, India. Retrieved from https://www.ijcai.org/Proceedings/07/Papers/076.pdf

Murray, M. C., & Pérez, J. (2015). Informing and performing: A study comparing adaptive learning to traditional learning. *Informing Science, 18*(1), 111–125. https://doi.org/10.28945/2165.

Nakić, J., Granić, A., & Glavinić, V. (2015). Anatomy of student models in adaptive learning systems: A systematic literature review of individual differences from 2001 to 2013. *Journal of Educational Computing Research, 51*(4), 459–489. https://doi.org/10.2190/EC.51.4.e.

Own, Z. (2006). The application of an adaptive web-based learning environment on oxidation–reduction reactions. *International Journal of Science and Mathematics Education, 4*(1), 73–96. https://doi.org/10.1007/s10763-004-8129-6.

Oxman, S., & Wong, W. (2014). White Paper: Adaptive learning systems. Retrieved from http://kenanaonline.com/files/0100/100321/DVx{_}Adaptive{_}Learning{_}White{_}Paper.pdf

Pancar, T., Holthaus, M., & Bergamin, P. B. (2019). Enhanced system usability scale for adaptive courses. In *The twelfth international conference on advances in human-oriented and* *personalized mechanisms, technologies, and services, centric 2019,* Valencia, Spain, 24–28 November 2019.

Pliakos, K., Joo, S. H., Park, J. Y., Cornillie, F., Vens, C., & Van den Noortgate, W. (2019). Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems. *Computers and Education, 137*, 91–103. https://doi.org/10.1016/j.compedu.2019.04.009.

Rey, G. D., & Buchwald, F. (2011). The expertise reversal effect: Cognitive load and motivational explanations. *Journal of Experimental Psychology: Applied, 17*(1), 33–48. https://doi.org/10.1037/a0022243.

Rus, V., Baggett, W., Gire, E., Franceschetti, D., Conley, M., & Graesser, A. (2013). *Towards learner models based on learning progressions in DeepTutor*. Retrieved from www.semantic-similarity.org

Scanlon, E., Sharples, M., Fenton-O'Creevy, M., Fleck, J., Cooban, C., Ferguson, R., et al. (2013). *Beyond prototypes: Enabling innovation in technology-enhanced learning*. Milton Keynes: Open University. Retrieved from http://beyondprototypes.com/

Scheiter, K., Fillisch, B., Krebs, M.-C., Leber, J., Ploetzner, R., Renkl, A., et al. (2017). How to design adaptive information environments to support self-regulated learning with multimedia. In *Informational environments* (pp. 203–223). Cham: Springer. https://doi.org/10.1007/978-3-319-64274-1_9.

Schertzer, C. B., & Schertzer, S. M. B. (2004). Student satisfaction and retention: A conceptual model. *Journal of Marketing for Higher Education, 14*(1), 79–91. https://doi.org/10.1300/J050v14n01.

Shelle, G., Earnesty, D., Pilkenton, A., & Powell, E. (2018). Adaptive learning: An innovative method for online teaching and learning. *Journal of Extension, 56*(5). Retrieved from https://joe.org/joe/2018september/pdf/JOE{_}v56{_}5a5.pdf

Shute, V., & Towle, B. (2003). Adaptive e-learning. *Educational Psychologist, 38*(2), 105–114. https://doi.org/10.1207/S15326985EP3802_5.

Somyürek, S. (2015). The new trends in adaptive educational hypermedia systems. *The International Review of Research in Open and Distance Learning, 16*(1), 221–241. https://doi.org/10.19173/irrodl.v16i1.1946.

Sottilare, R. A., & Goodwin, G. A. (2017). *Adaptive instructional methods to accelerate learning and enhance learning capacity*. Retrieved from https://gifttutoring.org/attachments/download/2427/Sottilare and Goodwin DHSS2017.pdf

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.

Tai, M., Arroyo, I., & Woolf, B. P. (2013). Teammate relationships improve help-seeking behavior in an intelligent tutoring system. *International Conference on Artificial Intelligence in Education* (pp. 239–248). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39112-5_25 .

van Merriënboer, J. J. G., & Sluijsmans, D. M. A. (2009). Toward a synthesis of cognitive load theory, four-component instructional design, and self-directed learning. *Educational Psychology Review, 21*(1), 55–66. https://doi.org/10.1007/s10648-008-9092-5.

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147–177. https://doi.org/10.1007/s10648-005-3951-0.

Vagale, V., & Niedrite, L. (2012). Learner model' s utilization in the e-learning environments. In *Tenth international Baltic conference on databases and information systems*. Vilnius, Lithuania, pp. 162–174.

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*, 118–130. https://doi.org/10.1016/j.chb.2010.07.038.

Verdú, E., Regueras, L. M., Verdú, M. J., De Castro, J. P., & Pérez, M. Á. (2008). An analysis of the research on adaptive learning: The next generation of e-learning. *WSEAS Transactions on Information Science and Applications, 5*(6), 859–868.

Vesin, B., Mangaroska, K., & Giannakos, M. (2018). Learning in smart environments: User-centered design and analytics of an adaptive learning system. *Smart Learning Environments, 5*(1). https://doi.org/10.1186/s40561-018-0071-0.

Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes* (p. 159). Cambridge, MA: Harvard University Press.

Walkington, C. A. (2013). Using adaptive learning technologies to personalize instruction to student interests: The impact of relevant contexts on performance and learning outcomes. *Journal of Educational Psychology, 105*(4), 932–945. https://doi.org/10.1037/a0031882.

Wauters, K., Desmet, P., & Van Den Noortgate, W. (2010). Adaptive item-based learning environments based on the item response theory: Possibilities and challenges. *Journal of Computer Assisted Learning, 26*(6), 549–562. https://doi.org/10.1111/j.1365-2729.2010.00368.x.

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 & Education, 140*, 103599. https://doi.org/10.1016/J.COMPEDU.2019.103599.

Yang, T. C., Hwang, G.-J., & Yang, S. J. H. (2013). Development of an adaptive learning system with multiple perspectives based on students' learning styles and cognitive styles. *Educational Technology and Society, 16*(4), 185–200.

Yuan, X. (2019). Model and implementation of personalized adaptive learning and analysis technology based on large data. *International Conference on Artificial Intelligence and Advanced Manufacturing* (pp. 202–205), Dublin, Ireland. https://doi.org/10.1109/AIAM48774.2019.00048.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166–183. https://doi.org/10.3102/0002831207312909.

Zliobaite, I., Bifet, A., Gaber, M., Gabrys, B., Gama, J., Minku, L., & Musial, K. (2012). Next challenges for adaptive learning systems. *ACM SIGKDD Explorations Newsletter, 14*(1), 48. https://doi.org/10.1145/2408736.2408746.

Learning and Individual Differences 88 (2021) 102014

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Prediction of dilatory behaviour in online assignments

Christof Imhof[a,*], Per Bergamin[a,b], Stéphanie McGarrity[a]

[a] Fernfachhochschule Schweiz, UNESCO Chair on Personalized and Adaptive Distance Education, Ueberlandstrasse 12, 3900 Brig, Switzerland
[b] North-West University, Potchefstroom Campus, 11 Hoffman Street, Potchefstroom 2531, South Africa

ARTICLE INFO

Keywords:
Procrastination
Purposeful delay
Self-efficacy
Self-directed learning
Log data

ABSTRACT

Procrastination has been increasing since the proliferation of online learning. While traditionally assessed with self-report instruments, online learning offers the possibility to measure objective indicators (log data). In the present study, we aim to find out whether the combination of short scales on procrastination-related traits and log data predict the extent of dilatory behaviour in online tasks and performance (assignment scores). The log data models (which include the number of clicks on the assignment, the interval between thematic block start and first click, and the number of clicks on course activities as predictors) have a better fit and explain more variance than the questionnaire models when predicting delay; and the predictions barely improve when combined. The prediction of performance did not yield any noteworthy effects. Future studies need to diversify predictors by incorporating contextual factors to improve early and/or late predictions and allow classification of dilatory behaviour (e.g. procrastination vs purposeful delay).

1. Introduction

Procrastination is a widespread phenomenon chronically affecting the general population (Blunt & Pychyl, 1998; Harriott & Ferrari, 1996) in a variety of situations, be it by putting off housekeeping, doing taxes or seeing a dentist. A remarkable frequency of such behaviour has been reported in learning settings on an academic level, where the self-assessed prevalence is approximately 70% (Ellis & Knaus, 1977; Steel, 2007). This particular type of procrastination – academic procrastination – is characterised by favouring short-term (or immediate) gratifications such as sleeping, playing, watching TV (Pychyl, Lee, Thibodeau, & Blunt, 2000), checking Facebook (Meier, Reinecke, & Meltzer, 2016), or watching cat videos (Myrick, 2015) over long-term assignments, such as studying for an upcoming exam or writing papers, that only provide distal rewards. It is therefore often described as a failure to complete academic tasks within a given timeframe or delaying academic work in general (Schraw, Wadkins, & Olafson, 2007; Senécal, Koestner, & Vallerand, 1995). This fits with the common concept of procrastination as a failure of self-regulation (Klingsieck, Fries, Horz, & Hofer, 2012; Steel, 2007). However, not all dilatory behaviour is necessarily classified as procrastination since other types of delay have been identified, such as purposeful delay (see Corkin, Yu, & Lindt, 2011; Grunschel, Patrzek, & Fries, 2013). In the following sections, we define procrastination and purposeful delay, show their connections to other relevant learning concepts that can serve as indicators for them, followed by an overview over findings concerning prediction models involving dilatory behaviour.

Given how prevalent and potentially damaging procrastination can be (e.g. in regards to performance, see Steel, 2007), researchers have identified several indicators and measures of dilatory behaviour, with the goal of detecting such tendencies early-on, especially in online courses. These indicators are either subjective (i.e. self-report questionnaires) or objective (e.g. log data). The aim of this study is to explore which factors predict the extent of dilatory behaviour (i.e. how long students delay the submission of their assignments) the best and to determine whether a combination of both kinds of indicators improves the predictability in the context of online assignments.

2. Theoretical background, prior research, and current study

2.1. Academic procrastination

Steel (2007) describes procrastination as a "voluntary delay of an intended course of action despite expecting to be worse off for the delay" (p. 66). The second half of the quote highlights that procrastinators are aware of the ill-fated nature of their behaviour. Thus, what sets procrastinators apart are not intellectual differences (see Ferrari, 1991), but rather differences in the level of self-regulation (Rabin, Fogel, & Nutter-

* Corresponding author.
E-mail address: christof.imhof@ffhs.ch (C. Imhof).

https://doi.org/10.1016/j.lindif.2021.102014
Received 16 November 2020; Received in revised form 14 April 2021; Accepted 16 April 2021
Available online 7 May 2021

Upham, 2011; Wolters, 2003). Procrastination tendencies in academia are thought to have been increasing ever since the advent of online learning due to the self-directedness such learning environments demand (see You, 2015). Online learning environments require students to be more intrinsically motivated to prevent procrastinating (Wighting, Liu, & Rovai, 2008). This poses challenges to students' responsibility and time management, which increase the risk of procrastinating. In such environments, procrastination is often considered to be more damaging compared to traditional classroom settings (Tuckman, 2005) and thought to lead to higher drop-put rates (Doherty, 2006). Other negative outcomes procrastination is associated with include lower performance, reduced mood, increased stress, feelings of guilt, poor self-rated health, and lower self-efficacy (Pychyl, Lee, Thibodeau, & Blunt, 2000; Sirois, Melia-Gordon, & Pychyl, 2003; Steel, 2007; Tice & Baumeister, 1997).

A further characteristic of procrastination is its stability across situations and over longer time spans as evidenced by Steel's meta-analysis (Steel, 2007). However, there is also evidence to suggest that contextual and situational aspects, e.g. task characteristics or mood, may influence procrastination. The moderate correlation of 0.51 between state and trait procrastination (see Steel, 2007) indicates that they are related, but not one and the same. Hence, procrastination as a trait does not necessarily determine behaviour, which leaves room for other learning-related factors.

2.2. *Purposeful delay and connections to other learning-related factors*

Since procrastination is strongly linked to self-regulation (i.e. efforts by students to monitor, manipulate and improve their own learning, Corno & Mandinach, 1983), learning concepts related to or including self-regulation may be useful when it comes to predicting procrastination. One such concept is self-directed learning (SDL). SDL is defined as a process in which learners take charge of their own learning process by identifying their learning needs, goals, and resources, deploying appropriate strategies and evaluating the outcomes (Knowles, 1975). SDL has several crucial cornerstones that are relevant in this context, namely self-monitoring, self-management as well as evaluation and regulation of students' own learning (see Bolhuis, 1996; Garrison, 1997). Self-regulation is also one of the primary dimensions of SDL, along with motivation and metacognition (cf. Mentz & Van Zyl, 2018). Studies have shown that SDL has negative links to procrastination (e.g. Onji & Kikuchi, 2011; Schommer-Aikins & Easter, 2018) and positive links to self-efficacy (e.g. Saeid & Eslaminejad, 2017).

Self-efficacy is another learning concept related to self-regulation with strong ties to procrastination and is usually defined as the belief in one's ability to succeed, or more specifically "people's judgements of their capability to organize and execute courses of action required in attaining designated types of performances" (Bandura, 1986, p. 391). In their study, Wäschle, Allgaier, Lachner, Fink, and Nückles (2014) found a virtuous circle of self-efficacy to counteract the vicious circle of procrastination, the latter of which students with low levels of self-efficacy are prone to fall into. High self-efficacy is also associated with the functional counterpart of procrastination, a concept initially known as *active procrastination* (Chu & Choi, 2005). Active procrastination denotes behaviour that closely resembles "regular" procrastination on its surface, with four key differences: People who display active procrastination still achieve satisfactory outcomes, prefer to work under pressure, do it intentionally and are able to meet deadlines. These differences also serve as the four subscales of the Active Procrastination Scale (APS, Choi & Moran, 2009).

The naming of this concept received criticism on a linguistic level for being an oxymoron (see Pychyl, 2008) since procrastination is classically defined as irrational, ill-fated behaviour which directly contradicts the benefits that active procrastination is supposed to have. For this reason, the term *active delay* was introduced to avoid that linguistic conundrum. However, as Hensley (2014) notes, the dichotomy between active and passive is an oversimplification, which led to a third term being introduced to refer to the same concept, namely *purposeful delay* (see Corkin, Yu, & Lindt, 2011; Grunschel, Patrzek, & Fries, 2013). This is also the expression we use. A similar term has been coined by Klingsieck (2013), namely *strategic delay*, which can be considered synonymous with purposeful delay. This type of delay is characterised as a rational and intentional strategy that is positively related to self-efficacy. Reasons for delaying tasks in such a manner include prioritising other tasks, acquiring more information before executing a task, enhancing motivation (e.g. if one prefers to work under pressure), reaching a state of cognitive flow and maximising learning in a minimal amount of time (Ferrari, 2011; Schraw, Wadkins, & Olafson, 2007).

In their study, Corkin, Yu, and Lindt (2011) found that purposeful delay and procrastination are indeed inversely related. Moreover, procrastination was found to be negatively related to self-regulation whereas purposeful delay is positively linked with aspects of self-regulation such as the aforementioned self-efficacy (see Chowdhury & Pychyl, 2018) or self-regulated learning (Sundaramoorthy, 2018). Self-regulated learning (SRL) is a concept similar to self-directed learning (SDL) with subtle differences, mainly regarding its origins, with SRL being rooted in psychology and SDL stemming from pedagogy. Furthermore, purposeful delay was found to be linked to higher grades. Combined, these findings support the notion that purposeful delay is a functional adaptive form of delay, in stark contrast to procrastination. However, this notion has recently been challenged by Pinxten, De Laet, Van Soom, Peeters, and Langie (2019), who attempted to replicate the four factor-structure of the APS, as reported by Choi and Moran (2009). The authors instead found a three-factor structure (intentional decision, ability to meet deadlines, and a combined factor consisting of preference for pressure and outcome satisfaction) and the findings did not support the expected positive effect of purposeful delay on achievements.

2.3. *Measurement of dilatory behaviour*

Traditionally, procrastination and other types of delay are assessed with self-report instruments. However, as Malatincová (2015) points out, the subjective component of procrastination that these instruments assess needs to be distinguished from the actual delay component. This is a crucial distinction since hindsight bias or a situation-specific subjective sense of procrastination may influence people's perception of their own behaviour. A reliable measurement of the delay component of procrastination therefore requires objective methods less susceptible to this type of bias. In the context of academic procrastination, there are tools that have emerged since the advent of online learning which provide exactly that. One prominent example are log data, which can be extracted from most learning management systems (LMS). Such data include the number of clicks within the course, the time spent on certain learning activities, deadlines, submission times and so on, thus providing the means to calculate delay or find indicators thereof. Several studies have analysed dilatory patterns in this manner. Morris, Finnegan, and Wu (2005) for instance found significant differences in participation (i.e. number of views for content and hours spent on certain activities) between what they called "withdrawers" and "completers". The amount of viewed discussion posts and content pages not only predicted student participation, but also their finale grades.

Levy and Ramim (2012) investigated the relationship between procrastination and academic performance by using data analytic techniques and measured procrastination as the difference between the due time and the submission time in hours. Results indicate that students who procrastinated performed worse in online exams, with additional trends being the time of day the exams were taken (exams taken in the morning yielded better results), the students' academic level (sophomores procrastinated the most), and gender (women procrastinated more often). Similarly, You (2015) measured procrastination by observing late or absent submission. Instead of implementing data analytics, the author conducted a multiple regression analysis and found that procrastination negatively predicted course achievement, with the predictability increasing over time.

A different approach was followed by del Puerto Paule-Ruiz, Riestra-González, Sánchez-Santillán, and Pérez-Pérez (2015), who identified indicators of procrastination based on additional log data extracted from the LMS *Moodle*, such as the number of clicks in mandatory resources, time spent on quizzes and the time it takes to first click on or submit various tasks. These indicators were then discretised (low/medium/high) and formed into association rules. These rules were reformulated in the form of conditionals, e.g. if the number of clicks in a mandatory is high and the time it takes until the first submission of an assignment is low, then the performance is good. The time-related indicators turned out to be the most important ones for students' performance.

A similar study was conducted by Cerezo, Esteban, Sánchez-Santillán, and Núñez (2017), who also implemented association rules based on indicators on Moodle, albeit in a blended learning context. The results were comparable, as the time-related indicators (e.g. the time that passes between the unlocking of a block on Moodle and the submission of an assignment within that block) were strongly associated with performance. In their study, Yamada, Oi, and Konomi (2017) assessed self-efficacy in addition to procrastination and learning performance, albeit in the context of self-regulated learning (SRL) rather than self-directed learning (SDL). They found that self-efficacy was related to the frequency of out-of-class activities, submission times, and performance. Akram et al. (2019) predicted academic procrastination based on homework submissions by using ten different classification algorithms. Homework submission data were also used in two recent studies by Yang and colleagues: in Yang et al. (2020b), they classified students as procrastinators, procrastination candidates and non-procrastinators with a clustering algorithm based on homework submission data and in Yang et al. (2020a) the authors implemented educational data mining in order to predict course achievement with students' homework-based procrastination behaviour.

2.4. The present study

These studies demonstrate that dilatory behaviour can be measured by various instruments independent from questionnaires. However, most of these studies use state procrastination as a predictor for achievement rather than trying to predict dilatory behaviour itself, specifically its extent. Trait procrastination has been predicted before (e.g. Cao, 2012, who predicted both procrastination and purposeful delay based on APS scores with self-efficacy, learning strategies, and motivational factors), state procrastination less so. To our knowledge, which predictors comparatively work best for predicting the extent of dilatory behaviour is hardly explored, since the studies presented above rely on algorithms based on patterns in the submission dates alone, rather than assessing the value of other indicators such as click data or learning concepts related to delay (like self-directed learning and self-efficacy). Despite our focus on the predictability of dilatory behaviour, achievement was still of interest to us since procrastination and achievement are closely linked in most of the research in this area (poor grades or dropouts being commonly researched consequences of this behaviour). Achievement is usually operationalised with final exam or cumulative course scores, but we are interested to see whether a more direct approach works as well, in this case the raw scores of individual online assignments.

Our research questions are thus as follows: 1. Which predictor types and individual predictors are the most successful in predicting the extent of dilatory behaviour? 2. Does a combination of subjective trait predictors (questionnaires) and objective state predictors (log data) improve the predictability of dilatory behaviour? 3. Which predictors types and individual predictors are able to predict achievement in online assignments (scores) the best? 4. Does a combination of subjective and objective predictors improve the predictability of scores?

As far as we are aware, no other studies have combined log data analysis with questionnaires in this context, specifically the combination of procrastination, purposeful delay, and the two learning-relevant concepts of self-efficacy and self-directed learning. We expect all predictors to contribute to the prediction of delay as well as achievement, in accordance with the relations between the involved concepts in the literature. Self-efficacy and self-directed learning are thus hypothesised to be negative predictors of delay and positive predictors of achievement. Academic procrastination and purposeful delay are expected to be positive predictors of delay (both being types of trait delay) while mixed for achievement (positive for purposeful delay and negative for procrastination). The number of clicks per assignment and the number of clicks on relevant activities, both indicating activity, are expected to negatively predict delay and positively predict achievement. The final predictor, the interval between the start of a block and the first click, is hypothesised to be a positive predictor of delay and a negative predictor of achievement. Moreover, we hypothesise that the combination of the subjective and objective predictors improves predictability.

3. Methods

3.1. Participants

The participants ($N = 243$, 115 women, 128 men, mean age 33.53, $SD = 9.03$) were students at a Central-European distance university who partook in the study voluntarily after being invited via e-mail. Each student was enrolled in between one and four courses the university offered during the autumn semester of 2019. The courses all followed a blended learning approach with 80% online (or self-paced) studies and 20% traditional face-to-face classes. As a reward for their participation, the students had a chance of winning one of 3 sets of cinema vouchers. The invitation to our study was sent to every student that was registered at our institution at that point (around 2000), resulting in a response rate of about 12%. The questionnaires were filled in online at the very end of the semester, preceded by a briefing about the duration and purpose of the study. The participating students were also informed about the planned use of their log data on Moodle and consented by entering their e-mail address after reading an informed consent form, which was in turn used to match the datasets. Since the ethics committee of our institution was in its early development at the time of data collection, we could not obtain their approval for this study. However, we conducted it in accordance with the Declaration of Helsinki, assuring consent by providing an informed consent form, highlighting the voluntary nature of the participation, and disclosing the purpose of the study as well as our data protection measures.

Out of 243 initial participants, 134 remained in the final dataset due to multiple issues. First, 24 e-mail addresses were unidentifiable, which meant the questionnaire scores of these participants could not be connected to the students' log data. Then, 31 students were not enrolled in courses that actually had any online assignments, which meant there was no delay to be predicted. 51 students were additionally lost since the deadlines for the assignments in their courses could not be deduced by log data alone, meaning the lecturers communicated the deadlines on a different channel (e.g. via e-mail, course files or in a face-to-face class). Finally, 3 students had to be excluded due to errors in the automatically generated log protocols. A total of 126 courses were involved, each including one to eight students from our sample.

3.2. Instruments & procedure

The variables of interest were trait variables assessed by online questionnaires and state variables in the form of log data. The four trait variables *self-efficacy*, *self-directed learning*, *procrastination* and *purposeful delay* were measured with one questionnaire each. Self-efficacy was measured with the General Self-efficacy Scale (GASE), which consists of 5 items, all related to beliefs about one's academic abilities (e.g. conviction to pass the exam if putting in enough effort) and is aimed at measuring academic self-efficacy regardless of academic discipline

(Nielsen, Dammeyer, Vang, & Makransky, 2018). In order to measure self-directed learning, we used the Self-directed Learning Scale (SDLS), which consists of a single factor comprising 10 items (Lounsbury, Levy, Park, Gibson, & Smith, 2009). The items all involve the concept of the capacity to learn without relying on others (e.g. finding answers on one's own for content not explained in class). We measured academic procrastination with the Academic Procrastination Scale — Short form (APS-S) that consists of 5 items, all of which include avoiding or delaying academic tasks (McCloskey, 2012). This short form contains the items that the author deemed the most promising from his full-length version, based on their psychometric properties. The validity of the suggested short form was then confirmed in a study by Yockey (2016).

We chose the Active Procrastination Scale (APS) to assess purposeful delay, which consists of 4 subscales with 4 items each, i.e. a total of 16 items (Choi & Moran, 2009). The subscales are *intentional decision*, *ability to meet deadlines*, *preference for pressure*, and *outcome satisfaction*. Three of the questionnaires utilise a five-point Likert scale (APS-S, GASE, SDLS) while the APS implements a seven-point Likert scale. All questionnaires exclusively implement the agree/disagree type of question. We chose the GASE, SDLS, and APS-S for their short length, which is beneficial for online surveys with busy students. To our knowledge, no shorter alternatives to the APS are available, which is why we implemented the original questionnaire. All of the questionnaires were translated into German in a multi-step process. We first translated the original questions from English into German ourselves, then used translation software to translate the German versions back into English, compared the two versions to detect specific expressions that may differ between them, and finally asked a native speaker for a final verification.

The internal consistency, item difficulty, item discrimination, and example items of all four of the questionnaires are shown in Table 1 below. Item difficulty usually refers to the likelihood of a correct answer in a test, but can also be used to determine the likelihood of an affirmative answer in a questionnaire. The internal consistency, expressed by Cronbach's alpha, is 0.68 for the GASE, 0.81 for the SDLS, 0.76 for the APS, and 0.85 for the APSS.

The log data was extracted from the database of the university's instance of Moodle. Since variables related to durations (e.g. the duration of solving a quiz or reading an interactive book) were not available to us at the time, we used click-based variables included in the raw log files, signifying activity. The first such variable was *the number of clicks on an assignment*, which is the sum of all individual clicks on a particular assignment section in the course before the assignment was submitted. Clicks that occurred after the submission (e.g. to check whether the lecturer provided their feedback) were ignored because they were irrelevant for the prediction. Next, we considered the overall *number of clicks on relevant activities* as a predictor. We consider all activities on Moodle to be relevant that are provided as optional tasks that involve engagement with the learning material. The variable was thus calculated as the sum of all clicks students made on the following course activities: quizzes, wikis, forums, chats, glossaries, embedded learning videos, and interactive books. Moodle activities we did not consider were stat reports, grade reports, and schedules. The courses at our institution generally consist of multiple blocks (between five and ten), each with its own assignments, optional tasks, and face-to-face class. The third and final log variable was the *block-click-interval*, which was the difference between the point in time the first click on an assignment was registered and the day the block the assignment was a part of started (generally two weeks after the previous face-to-face class). We included this variable due to its face validity: the longer a student waits to start working on an assignment (indicated by the first click since the task description is not visible in the main course and requires at least one click to be read), the less time they have to complete it before the deadline.

Table 1
Questionnaire descriptives.

Scale	Cronbach's alpha	Item difficulty	Item discrimination	Example item
GASE	0.68	0.71–0.86	0.2–0.66	I generally manage to solve difficult academic problems if I try hard enough.
SDLS	0.81	0.62–0.86	0.32–0.62	I like to be in charge of what I learn and when I learn it.
APS	0.76	0.45–0.86	−0.01–0.68	I frequently find myself putting important deadlines off.
APSS	0.85	0.42–0.52	0.62–0.73	In order to make better use of my time, I intentionally put off some tasks.

Note. Cronbach's Alpha, item difficulty, item discrimination values, and example items for all four questionnaires

Since the outcome variable *delay* depends on the existence of a deadline for its calculation, we focus solely on the submission of mandatory assignments instead of involving additional learning tasks that are self-paced and/or optional. Thus, we defined delay as the temporal distance between the deadline of an assignment and its submission. The other outcome variable of interest, *score*, was calculated based on raw assignment scores extracted from the log files, which were then divided by the respective maximum score. This was done in order to turn all of the scores into percentage scores, rendering them comparable.

On top of these predictors, we considered five covariates for our models, including *age* and *gender*. Since learning behaviour could change over time, we considered *study experience* as an additional covariate, which refers to the number of years the students have been studying at our institution up to the point of data collection. *Department* is another covariate we included since the effects could vary between study subjects. Moreover, we considered the *type of deadline* as a final relevant covariate. Since our institution grants lecturers many liberties in regards to course-specific details, assignment deadlines are specified in a diverse range of ways. We categorised the deadlines into two groups for this study: relative dates (i.e. in relation to a certain face-to-face class, e.g. two weeks after the next class) and absolute dates (e.g. on the 31st of August, 23:59).

3.3. Analyses

The analyses were conducted with R (Version 4.0.2; R Core Team, 2020) and the R-packages *brms* (Version 2.13.5; Bürkner, 2017, 2018), *chron* (Version 2.3.56; James & Hornik, 2020), *fitdistrplus* (Version 1.1.1; Delignette-Muller & Dutang, 2015), *flextable* (Version 0.5.11; Gohel, 2020), *ggplot2* (Version 3.3.2; Wickham, 2016), *ggpubr* (Version 0.4.0; Kassambara, 2020), *knitr* (Version 1.30; Xie, 2015), *lubridate* (Version 1.7.9; Grolemund & Wickham, 2011), *moments* (Version 0.14; Komsta & Novomestky, 2015), *papaja* (Version 0.1.0.9997; Aust & Barth, 2020), *performance* (Version 0.5.0; Lüdecke, Makowski, Waggoner, & Patil, 2020), *psych* (Version 2.0.8; Revelle, 2020), *sjPlot* (Version 2.8.5; Lüdecke, 2020), *stringr* (Version 1.4.0; Wickham, 2019), and *tidyverse* (Version 1.3.0; Wickham et al., 2019).

We followed Bayesian approaches for our analyses since we intend to use the results of this study to inform the priors of a follow-up study.

4. Results

4.1. Descriptives

The mean delay was −1.66 days, which translates to an early submission of assignments roughly one and a half days before the deadline ($SD = 18.26$, $min = -113.51$, $max = 132.43$). Positive values indicate delay and negative values mean early submissions. The skewness of the distribution was low (1.20), the kurtosis however was excessively high

(18.91) due to a vast majority of the submissions having been made shortly before or after the deadline (see Fig. 1). The total amount of assignments was 1107, equating an average of 8.26 assignments per student. The mean score is surprisingly high at 89.14% ($SD = 14.58$). The median revealed that half of the scores are above 93.81%. The mean values of the predictors were as follows: self-efficacy ($mean = 19.71$; $SD = 3.10$, $min = 8$, $max = 25$), self-directed learning ($mean = 38.74$, $SD = 5.29$, $min = 18$, $max = 50$), academic procrastination ($mean = 11.04$, $SD = 4.16$, $min = 5$, $max = 21$), purposeful delay ($mean = 72.51$, $SD = 11.87$, $min = 43$, $max = 104$), number of clicks on the assignment ($mean = 6.58$, $SD = 4.91$, $min = 1$, $max = 34$), the number of clicks on relevant activities ($mean = 173.83$, $SD = 168.44$, $min = 0$, $max = 1237$), and the block-click-interval ($mean = -7.13$, $SD = 32.40$, $min = -150.30$, $max = 98.72$).

The values for our covariates were as follows: age ($mean = 31.61$, $SD = 7.97$, $min = 19$, $max = 61$), gender ($N_{men} = 65$, $N_{women} = 69$), study experience in years ($mean = 2.63$, $SD = 1.24$, $min = 1$, $max = 5$), the department ($N_{assignEconomics} = 366$, $N_{assignComputerScience} = 418$, $N_{assignHealth} = 323$), and the type of deadline ($N_{relative} = 944$, $N_{absolute} = 163$). The predictors and covariates were all standardised (z-transformed) for the multilevel models due to their different scales (with the exception of the non-binary categorial variable *department*).

4.2. Bayesian multilevel models

Due to the way our data was structured, we decided to calculate multilevel models. The basic structure was as follows: Every student was enrolled in at least one course and handed in at least one assignment for each of their courses. The assignments were therefore located on level 1 and were nested in two grouping factors, namely the *course* and the *student* (level 2), resulting in repeated measures for some of the variables. In our dataset, the variables student and course were interlaced, i.e. students could be part of multiple courses and courses could contain multiple students. Our dataset was not large enough to inform highly complex multilevel models with nested grouping factors. Therefore, we settled on 2-levels model with course and student as level-2 variables.

Our predictors and covariates were either associated with the *assignment*, the *course*, or the *student*. The variables associated with *assignment* (level 1) were the *number of clicks on the assignment* and the *block-click-interval* as predictors and the covariate *type of deadline*. The predictors associated with *student* (level 2) were the four questionnaire scores as well as the covariates *age*, *gender*, *study experience*, and *stress*, since these variables concerned the participants rather than the assignments. One predictor (*number of clicks on relevant activities*) and one covariate (*department*) were associated with *Course* (also level 2).

We then calculated six Bayesian multilevel models with *brms* in order to address the first and second research questions. The first model was an intercept-only model ($model_{IO}$). The second one ($model_{quest}$) then added the four trait predictors (i.e. GASE, SDLS, APSS, and APS). The third and fourth models ($model_{logRI}$ and $model_{logRS}$) contained state predictors in the form of the three log data variables described above. The trait predictors were not included in these two models. Model three included random intercepts and model four added random slopes. The grouping factor for the random parts of these two models was either *course* or *student*, depending on the predictor. The fifth and sixth models ($model_{allRI}$ and $model_{allRS}$) included all seven predictors, again with the same random parts as in models three and four. The five covariates were featured in all of the models except for the intercept-only model.

All models shared the same outcome variable, namely *delay*. Despite the unusual distribution seen in the left-hand plot in Fig. 1, we opted for a Gaussian distribution in the multilevel models due to the low skewness of the distribution. As an example for how we specified the models in R, we chose $model_{allRS}$ (see R code chunk below). Since we did not have any informative priors at our disposal, we used the default weakly-informative priors provided by *brms*.

```
delay_model_all_RS <- brm(delay ~ GASE_z + SDLS_z + APSS_z + APS_z + n_clicks_assign_z + interval_block_click_z + n_clicks_relevant_activity_z + age_z + gender_z + study_exp_z + deadline_type_z + depart + (n_clicks_assign_z + interval_block_click_z + deadline_type_z + depart|CourseID) + (n_clicks_assign_z + interval_block_click_z + n_clicks_relevant_activity_z + deadline_type_z|StudentID), data=delay_df, family='gaussian', chains=4)
```

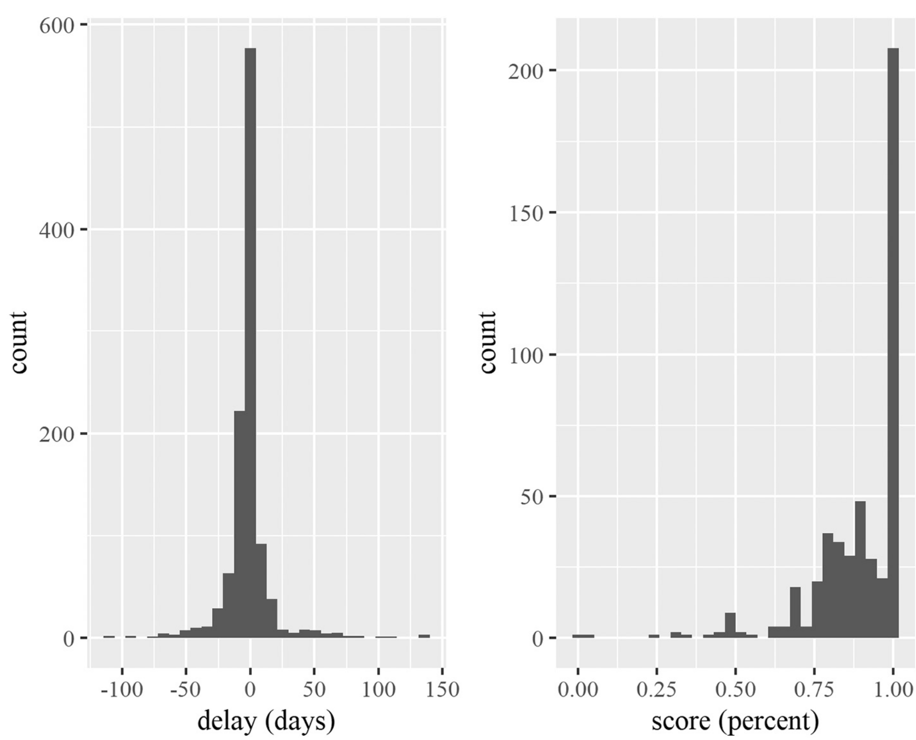

Fig. 1. Distribution of delay (in days) and assignment scores (in percent).

First, we report the magnitude of the fixed effects in our models. According to Lorah (2018), the effect size of fixed effects in multilevel models can either be represented by the standardised regression coefficients or by Cohen's f^2. We decided to report both since the high-kurtosis distribution may lead to effects based on regression coefficients appearing larger than they actually are. Cohen's f^2 effect sizes for all predictors and covariates are displayed in Table 2. We found two fixed effects that were deemed at least "small" judging by Cohen's f^2 (see Fig. 2). The larger of the two was produced by the predictor *block-click-interval*, which was present in all of the models it was included in (E_{logRI} = 6.20, [5.20, 7.23]; E_{logRS} = 6.48, [4.57, 8.42]; E_{allRI} = 6.10, [5.10, 7.15]; E_{allRS} = 6.43, [4.48, 8.39]). The effect size was medium to small (f^2_{logRI} = 0.15, f^2_{logRS} = 0.15, f^2_{allRI} = 0.14,f^2_{allRS} = 0.14). The other small effect was produced by the predictor *number of clicks on the assignment*, which also appeared in all four respective models (E_{logRI} = 4.65, [3.75, 5.54]; E_{logRS} = 3.82, [2.18, 5.44]; E_{allRI} = 4.68, [3.81, 5.58]; E_{allRS} = 3.92, [2.34, 5.54]).

The remaining fixed effects did not quite pass the threshold set by Cohen's f^2 and were not as consistent across models as the two main effects, but are still worth reporting due to the credible interval of their regression coefficients not including zero. These included four predictors and two covariates. A fixed effect of *self-directed learning* emerged in two models (E_{quest} = −1.81, [−3.25, −0.35]; E_{allRI} = −2.06, [−3.59, −0.45]), as did an effect of *number of clicks on relevant activities* (E_{allRI} = 2.12, [0.10, 4.15]; E_{allRS} = 2.12, [0.01, 3.02]). Moreover, we found effects for *self-efficacy* (E_{quest} = 2.86, [1.44, 4.27]) and *academic procrastination* (E_{allRS} = 0.99, [0.06, 1.93]) in one model each. The covariate *study experience* produced fixed effects in three models (E_{quest} = 1.57, [0.33, 2.80]; E_{logRI} = 3.23, [0.39, 6.03]; E_{allRI} = 3.27, [0.37, 6.10]), as did *deadline type* (E_{quest} = −1.92, [−3.07, −0.80]; E_{logRS} = −2.03, [−3.93, −0.15]; E_{allRS} = −1.99, [−3.98, −0.10]).

The credible intervals of the remaining predictors and covariates all included 0, rendering them negligible. Random effects can be found in all RS models and for both grouping factors. The fixed and random effects of $model_{allRS}$ are displayed in Table 3 (see Annex A for tables concerning the other models). Two of the random (i.e. group level) effects were correlated, namely the *block-click-interval* and the *number of clicks on the assignment*. This effect existed for both grouping factors (*course*: R_{allRS} = 0.66, [0.36, 0.88]; *student*: R_{allRS} = 0.68, [0.33, 0.92]). The correlation between the variables themselves was outside the realm of multicollinearity (R = −0.27). The other correlated effects did not involve the predictors.

We then conducted a model fit comparison and calculated the effect size for the models, represented by R^2 (see Table 4). $Model_{allRS}$ had the best fit in comparison to the other five models; however, the difference to the second best model ($model_{logRS}$) was lower than the respective standard error, rendering the exact ranking unclear (see Table 4). The percentage of explained variance in $model_{allRS}$, calculated with the *r2_bayes* function from the package *performance*, was R^2 = 0.74 when random effects were included (conditional) and 0.18 without random effects (marginal). The loo-adjusted R^2, which is the closest equivalent to an adjusted R^2 measure, was R^2 = 0.62.

In order to investigate the third and fourth research questions, we calculated the same models a second time, this time with *score* as the outcome variable and *delay* as an additional eighth predictor. Due to the peculiar distribution of the outcome variable, we first inspected a Cullen and Frey plot (see Fig. 3) with the *fitdistrplus* package to determine the most appropriate distribution based on skewness and kurtosis, in this case a gamma distribution. Since the scores included 0, we chose the hurdle gamma family for our models.

The resulting models included very few fixed effects for any of the predictors. Cohen's f^2 was not applicable due to the underlying distribution, which is why we focused on the standardised regression coefficients. Two predictors produced fixed effects, namely *self-directed learning* (E_{quest} = 0.03, [0.01, 0.05]; E_{allRI} = 0.03, [0.01, 0.05]; E_{allRS} = 0.04, [0.01, 0.06]) and *academic procrastination* (E_{quest} = 0.03, [0.01, 0.05]). Two covariates also had effects: *study experience* (E_{logRS} = −0.08, [−0.13, −0.01]; E_{allRS} = −0.08, [−0.14, −0.02]) and the *department*, specifically computer science when compared to health (E_{quest} = 0.09, [0.03, 0.15]). See Annex B for the model summaries. The model fit comparison (Table 5) revealed an unclear ranking (again due to large standard errors) and small R^2s.

5. Discussion

The main goal of this study was to find out whether the combination of subjective questionnaire data (trait) and objective log data extracted from an LMS (state) improves the predictability of two outcome variables, dilatory behaviour and achievement in online assignments. Moreover, we intended to determine the best individual predictors for these two outcome variables. We calculated multiple Bayesian multilevel models to answer these questions.

Even though the high-kurtosis distribution in the delay data and the peculiar distribution of the scores were not ideal, the models all converged with satisfying Rhats, indicating a good fit between the data and our models. In regards to delay, the two strongest and most consistent effects were both based on log data predictors, namely the *block-click-interval* and the *number of clicks on an assignment* in that order. A higher interval was indicative of a longer delay, which can be explained by the circumstance that students who start their work assignment later have less time to complete it and therefore likely hand it in later. This result is in line with the findings of del Puerto Paule-Ruiz, Riestra-González, Sánchez-Santillán, and Pérez-Pérez (2015) and Cerezo, Esteban, Sánchez-Santillán, and Núñez (2017). The latter effect was

Table 2
Effect sizes.

	Cohen's f^2 quest	Cohen's f^2 log_RI	Cohen's f^2 log_RS	Cohen's f^2 all_RI	Cohen's f^2 all_RS
Predictors					
GASE_z	0.01	–	–	0.01	0.01
SDLS_z	0.01	–	–	0.01	0.01
APSS_z	0	–	–	0.00	0.00
APS_z	0	–	–	0.00	0.00
n_clicks_assign_z	–	0.05	0.04	0.05	0.05
interval_block_click_z	–	0.15	0.15	0.14	0.14
n_clicks_relevant_activity_z	–	0	0	0.00	0.00
Covariates					
age_z	0	0	0	0.00	0.00
gender_z	0.01	0	0	0.00	0.00
study_exp_z	0.01	0.01	0.01	0.01	0.01
deadline_type_z	0.01	0	0	0.00	0.00
depart	0.01	0	0	0.00	0.00

Note. Effect sizes (Cohen's f^2) of predictors and covariates for all models expect the intercept-only model

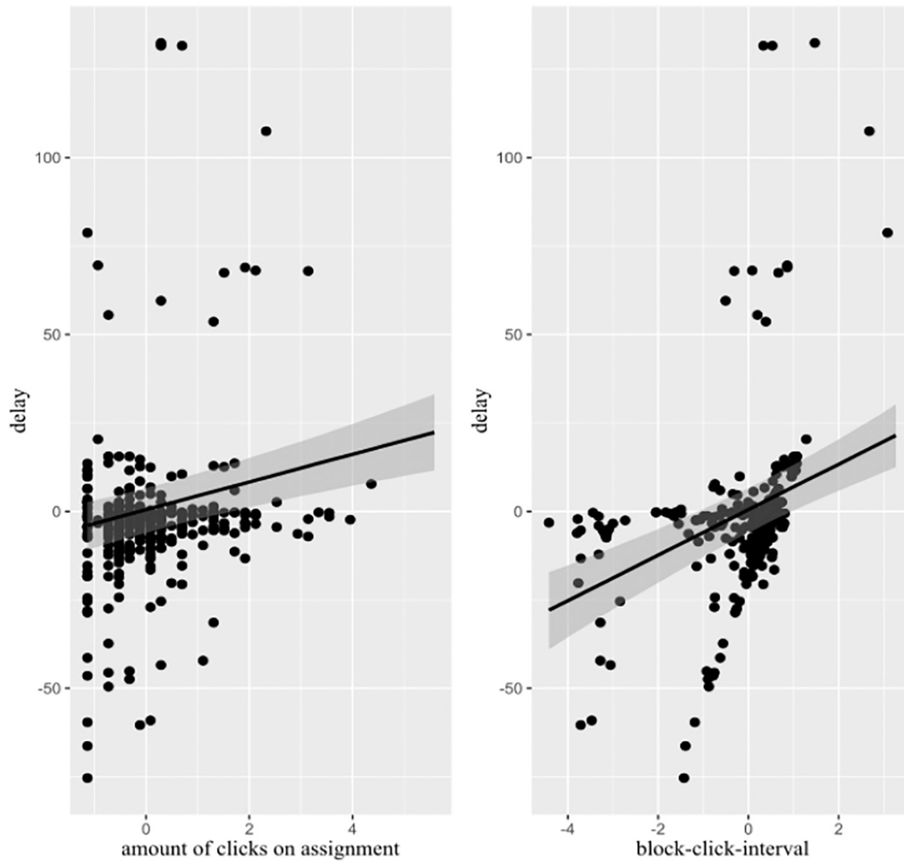

Fig. 2. Effects of number of clicks on assignment ($model_{allRS}$), and block-click-interval ($model_{allRS}$).

also positive, meaning a higher number of clicks indicates longer delay. More activity would usually imply less procrastination (see Cerezo, Esteban, Sánchez-Santillán, & Núñez, 2017), in this case however we look at the activity specific to the page containing the description of the assignment and the according submission form itself, rather than meaningful activity within the course (which is covered by the third log variable). Less clicks may be associated with swift submissions of assignments considering there is little benefit to be gained from opening this particular page within the course multiple times on different occasions. Also, we only considered the overall number of clicks rather than inspecting its temporal distribution across the semester, so increased activity can emerge through random clicking or repeated visits to this part if the students needed to refresh their memories (e.g. assignment instructions).

The third log variable, *clicks on relevant activities*, only producing a tiny effect is surprising, given the abovementioned connection between activity and procrastination. Moreover, the effect is positive, which may appear counterintuitive, but not implausible when considering that even relevant learning tasks can be used as a means of procrastinating, similar to how people may perform tasks they do not enjoy on a regular basis to distract themselves from a task they enjoy even less (e.g. house-cleaning instead of learning for an upcoming exam). The models solely containing objective predictors not only included the most noteworthy effects, they had a much better model fit than their subjective counterparts and, given the unclear model fit ranking, may even surpass the combined models.

These results suggest that paying attention to students' log behaviour is worthwhile when trying to predict (and ultimately prevent) dilatory behaviour. One potential intervention that could be developed based on these findings would be a time management tool that provides prompts on an LMS, reminding students of upcoming deadlines, similar to the system presented by Onji and Kikuchi (2011). This idea is also supported by the consistent effect of the covariate *deadline type*: the more absolute the deadline was, the less delay there was, implying a benefit of clearly communicated deadlines. The timing of such prompts would be key, since deadlines that appear distant may reduce the sense of urgency, particularly with students that have a low course load (Huang, Zhang, Burtch, Li, & Chen, 2021). Another possibility would be a tool that informs lecturers about dilatory tendencies of their students, allowing them to intervene in order to minimise the risk of drop-outs, similarly to the application presented by Antunes et al. (2016). Attention also needs to be paid to another covariate, the students' *study experience*. Interestingly, experienced students tended to delay the submission of their assignments more, which we suspect may be the result of an attitude of increased relaxation towards deadlines as the years pass by.

Given how little the log data models improve in their level of explained variance when the questionnaire predictors were added, it appears to be more efficient to only use log-data-based predictors rather than combining them with questionnaire-based predictors. However, these results do not imply that models with subjective predictors should be discarded altogether. Both of our successful predictors rely on at least one click having been made on the assignment, which often occurs rather close to the deadline. This could be an issue in practical applications of these models. If one wishes to make early predictions, prediction models devoid of log data may prove useful when no such data is yet available. Appropriate data for early predictions can be provided by subjective measures, which may theoretically be assessed via questionnaires at any given time during a semester. In our case, we found three effects of subjective predictors across the delay models, namely those of *self-efficacy*, *self-directed learning*, and *academic procrastination*. Despite these effects being considerably weaker than their objective counterparts and not passing the threshold for one of our two effect size measures, we still considered them worth discussing.

The higher the self-directed learning score was, the shorter the delay,

Table 3
Fixed and random effects in *model$_{allRS}$.*

	Estimate	Estimated error	Lower 95% CI	Upper 95% CI	Rhat
Fixed effects					
Intercept	0.44	3.24	−6.08	6.62	1.00
GASE_z	0.84	0.66	−0.45	2.13	1.00
SDLS_z	−1.07	0.58	−2.24	0.08	1.00
APSS_z	0.99	0.47	0.06	1.93	1.00
APS_z	−0.16	0.45	−1.03	0.71	1.00
n_clicks_assign_z	3.92	0.82	2.34	5.54	1.00
interval_block_click_z	6.43	1.00	4.48	8.39	1.00
n_clicks_relevant_activity_z	1.50	0.76	0.01	3.02	1.00
age_z	−0.45	0.46	−1.34	0.45	1.00
gender_z	−0.52	0.56	−1.58	0.58	1.00
study_exp_z	0.83	0.71	−0.55	2.26	1.00
deadline_type_z	−1.99	0.97	−3.98	−0.10	1.00
department comp.sci	−2.86	3.67	−9.85	4.44	1.00
department economics	0.57	3.82	−6.70	8.30	1.00
Random effects (course)					
sd(Intercept)	14.40	1.73	11.32	17.96	1.00
sd(n_clicks_assign_z)	4.91	0.73	3.60	6.41	1.00
sd(interval_block_click_z)	6.77	0.86	5.19	8.52	1.00
sd(deadline_type_z)	5.46	0.93	3.77	7.46	1.00
sd(department comp.sci)	15.98	2.10	12.11	20.46	1.00
sd(department economics)	3.66	2.77	0.17	10.45	1.02
cor(n_clicks_assign_z,interval_block_click_z)	0.66	0.14	0.36	0.88	1.01
cor(Intercept,deadline_type_z)	−0.84	0.10	−0.97	−0.57	1.00
cor(Intercept,department comp.sci)	−0.97	0.02	−0.99	−0.93	1.00
cor(deadline_type_z,department comp.sci)	0.88	0.09	0.66	0.98	1.00
Random effects (student)					
sd(Intercept)	1.01	0.62	0.06	2.33	1.00
sd(n_clicks_assign_z)	4.46	0.76	3.01	5.99	1.00
sd(interval_block_click_z)	3.62	0.74	2.23	5.13	1.00
sd(n_clicks_relevant_activity_z)	0.80	0.60	0.03	2.25	1.01
sd(deadline_type_z)	0.81	0.57	0.03	2.06	1.00
cor(n_clicks_assign_z,interval_block_click_z)	0.68	0.15	0.33	0.92	1.00

Note. Estimates, standard errors, lower and upper ends of credible intervals, and Rhats for fixed effects and random effects (both grouping factors). For the random effects, the estimates which include 0 in their credible intervals are not shown. “z” indicates z-transformed variables.

Table 4
Model fit comparison (delay).

Model	elpd diff	SE diff	R^2 Bayes	R^2 Bayes marginal	R^2 loo-adjusted
model$_{logRS}$	0.00	0.00	0.74	0.18	0.62
model$_{allRS}$	−3.11	3.24	0.74	0.17	0.62
model$_{allRI}$	−200.84	39.84	0.54	0.19	0.46
model$_{logRI}$	−201.30	39.68	0.54	0.18	0.47
model$_{quest}$	−538.19	65.79	0.05	–	0.02
model$_{IO}$	−551.21	65.34	0.00	–	0.00

Note. Comparison of the model fit by leave-one-out cross validation (LOO). Elpd diff indicates the difference between the expected log pointwise predictive density for a new dataset and SE diff is the standard error of elpd diff. Negative elpd diffs favour the first model, in this case model logRS. The marginal R^2 indicates the percentage of explained variance without random effects.

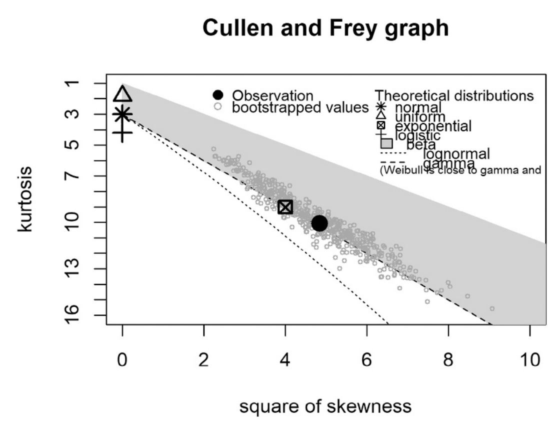

Fig. 3. Cullen and Frey graph for the distribution of scores.

which was consistent with our hypothesis. Given that self-directed learning involves teachable skills, this result implies that fostering it may be a fruitful intervention strategy. In contrast, the effect of self-efficacy was positive, meaning that higher scores indicated longer delay, negating our corresponding hypothesis. One possible explanation could be the positive relation between self-efficacy and purposeful delay (trait), meaning that students who delay on purpose may feel more confident in their decision compared to procrastinators. However, since we cannot distinguish between types of dilatory behaviour in this context (i.e. whether participants procrastinated or delayed the submission on purpose for valid reasons), this needs to be explored further. Another explanation could be the comparatively low internal consistency of the GASE. This raises the question if a different questionnaire may have been more appropriate (at the cost of efficiency). In contrast, the weak link between academic procrastination and delay is consistent with the literature, given the often-reported medium correlation between trait and state procrastination and the fact our delay measure conceptually encompasses more than one type of delay.

In contrast to the delay models, the score models only produced a few weak effects. The most consistent effect across models was *self-directed learning*, further stressing the potential benefits of fostering it. Additional effects were provided by *academic procrastination*, which was positive, and the covariates *study experience* (experienced students having slightly lower scores) and *department* (computer science students scoring slightly higher than students in the health department). The weakness of these effects prevents them from being interpreted in any meaningful way.

Table 5
Model fit comparison (score).

Model	elpd diff	SE diff	R^2 Bayes	R^2 Bayes marginal	R^2 loo-adjusted
$\text{model}_{\text{logRS}}$	0.00	0.00	0.46	0.19	0.22
$\text{model}_{\text{allRS}}$	−1.04	2.80	0.46	0.21	0.22
$\text{model}_{\text{allRI}}$	−55.62	50.80	0.31	0.12	0.23
$\text{model}_{\text{logRI}}$	−56.85	52.40	0.33	0.15	0.21
$\text{model}_{\text{quest}}$	−79.69	47.22	0.14	–	0.05
model_{IO}	−91.82	44.20	0.00	–	0.00

Note. Comparison of the model fit by leave-one-out cross validation (LOO). Elpd diff indicates the difference between the expected log pointwise predictive density for a new dataset and SE diff is the standard error of elpd diff. Negative elpd diffs favour the first model, in this case model allRS. The marginal R^2 indicates the percentage of explained variance without random effects.

Overall, the models produced several effects and the amount of explained variance is respectable. The presence of a sizable amount of random effects for both grouping factors, combined with the difference in explained variance between models that do or do not include them shows that individual and course-specific differences play a large role. Despite the importance of individual and course-specific factors this result implies, our predictors related to these two aspects produced either few or no effects at all, meaning other, contextual factors need to be taken into consideration, e.g. motivational ones (cf. Howell & Watson, 2007). Additional predictors could for example be located on the course level, e.g. interest in the topic, instructional clarity, or perceived difficulty (cf. Ackerman & Gross, 2005; Steel, 2007). Self-report measures, which were confined to the subject-level in this study, could also be applied on the course level or, even more fine-grained, on the assignment level. While more fine-grained subjective state measures could provide valuable insights in addition to the objective state and subjective trait measures already employed, they would require frequent data collection throughout the semester, which may be perceived as disruptive from the students' perspective. If this sort of measure is to be involved, short questionnaires would be crucial to prevent participants from dropping out of the study. The high percentage of variance explained by individual as well as course differences implies potential in assessing more contextual factors, such as specific reasons for delaying a submission or the procrastination-friendliness of the environment (Nordby, Klingsieck, & Svartdal, 2017). This could for instance be achieved by interviews, which allow for more nuanced information to be collected. Data gathered from interviews could also provide the necessary information to help categorise dilatory behaviour.

In addition, the log data predictors may benefit from a classification into distinguishable patterns. In the current models, click-based and temporal log data are treated as individual predictors, but merging them (e.g. by factoring in their distribution of clicks across the semester) may yield patterns that are able to give more insight into dilatory behaviour.

6. Limitations

The first limitation in our study concerns a potential sample bias, which is a two-fold issue. First, we lost a considerable amount of participants during data preparation due to missing data. These losses can mostly be explained by course characteristics, which is not uncommon in field studies. Second, self-selection biases could be involved since students who decided to participate in the study may procrastinate less than non-participants, rendering the sample less representative. Both of these possibilities could affect sampling and lead to a biased sample. However, the second possibility presents an additional advantage of predicting dilatory behaviour based on log data: unlike subjective predictors, which require participants to fill in a potentially lengthy questionnaire, thus opening the possibility of self-selection biases, objective predictors can be extracted from the LMS more directly and without interrupting the learning process. Participation would only require a permission by the student to extract the data. Biases may still apply since students could refuse to grant access to their data, but they at least would not be directly related to dilatory behaviour.

An additional limitation is the focus on mandatory assignments. Dilatory behaviour concerns learning as a whole, rather than the timing of assignments hand-ins alone. However, analysing delay requires clear deadlines, which is often lacking in non-mandatory tasks. Technically, optional tasks (e.g. quizzes) and studying for final exams also involve deadlines, but they are often implicit and not bound to a specific date. Moreover, they may involve blind spots from a data collection point of view due to a lack of direct, objective access to offline learning activities. How delay can be explored in regards to learning activities other than mandatory assignments should be addressed in future studies.

7. Conclusions

In conclusion, Bayesian multilevel models that include predictors based on questionnaire and log data can be implemented in order to predict the extent of dilatory behaviour in a learning management system, despite not all the involved predictors being equally consistent. Log data turned out to be the more relevant type of predictor, indicated by the negligible improvement of models that include both types of predictors over ones that only included log data predictors and the strength of the individual log data predictors. Predicting achievement yielded much weaker results. The results have implications for the timing of predictions in practical applications of such models, meaning the ability to make early predictions is not guaranteed for all of the predictors. Future studies need to take more (fine-grained) context factors into consideration, especially on the individual and course levels, in order to improve predictions and allow for differentiation between procrastination and other types of dilatory behaviour.

Funding

This work was supported by a research grant from the Canton of Valais.

Declaration of competing interest

There are no competing financial interests.

Acknowledgements

We would like to thank all students who participated in this study as well as Tansu Pancar for painstakingly extracting the log data from our learning management system.

Appendix A. Remaining tables for models predicting delay

Table 6
Fixed effects in $model_{quest}$.

	Estimate	Estimated error	Lower 95% CI	Upper 95% CI	Rhat
Fixed effects					
Intercept	−1.34	1.13	−3.56	0.92	1
GASE_z	2.86	0.72	1.44	4.27	1
SDLS_z	−1.81	0.74	−3.25	−0.35	1
APSS_z	0.60	0.59	−0.53	1.76	1
APS_z	−0.64	0.59	−1.77	0.54	1
age_z	−0.56	0.55	−1.65	0.54	1
gender_z	−1.10	0.60	−2.31	0.09	1
study_exp_z	1.57	0.63	0.33	2.80	1
deadline_type_z	−1.92	0.59	−3.07	−0.80	1
department comp.sci	−1.43	1.59	−4.50	1.66	1
department economics	0.67	1.68	−2.58	3.93	1

Note. Estimates, standard errors, lower and upper ends of credible intervals, and Rhats for fixed effects.

Table 7
Fixed effects in $model_{logRI}$.

	Estimate	Estimated error	Lower 95% CI	Upper 95% CI	Rhat
Fixed effects					
Intercept	0.65	3.31	−5.62	7.22	1.00
n_clicks_assign_z	4.65	0.46	3.75	5.54	1.00
interval_block_click_z	6.20	0.52	5.20	7.23	1.00
n_clicks_relevant_activity_z	1.64	1.06	−0.43	3.70	1.00
age_z	−0.73	0.72	−2.14	0.69	1.00
gender_z	−1.13	0.80	−2.72	0.39	1.00
study_exp_z	3.23	1.45	0.39	6.03	1.01
deadline_type_z	−0.30	0.52	−1.33	0.69	1.00
department comp.sci	−3.27	4.14	−11.37	4.65	1.00
department economics	2.49	4.23	−5.95	10.62	1.00

Note. Estimates, standard errors, lower and upper ends of credible intervals, and Rhats for fixed effects.

Table 8
Fixed and random effects in $model_{logRS}$.

	Estimate	Estimated error	Lower 95% CI	Upper 95% CI	Rhat
Fixed effects					
Intercept	1.24	3.11	−5.06	7.15	1.00
n_clicks_assign_z	3.82	0.84	2.18	5.44	1.00
interval_block_click_z	6.48	0.98	4.57	8.42	1.00
n_clicks_relevant_activity_z	1.27	0.75	−0.21	2.71	1.00
age_z	−0.26	0.48	−1.20	0.67	1.00
gender_z	−0.58	0.57	−1.69	0.54	1.00
study_exp_z	0.82	0.71	−0.59	2.25	1.00
deadline_type_z	−2.03	0.96	−3.93	−0.15	1.00
department comp.sci	−3.51	3.49	−10.19	3.39	1.00
department economics	−0.62	3.66	−7.55	6.63	1.00
Random effects (course)					
sd(Intercept)	14.34	1.66	11.34	17.95	1.00
sd(n_clicks_assign_z)	4.88	0.70	3.60	6.32	1.00
sd(interval_block_click_z)	6.73	0.89	5.03	8.52	1.00
sd(deadline_type_z)	5.47	0.92	3.75	7.40	1.01
sd(department comp.sci)	15.87	2.00	12.16	19.96	1.00
sd(department economics)	3.83	2.74	0.16	9.87	1.01
cor(n_clicks_assign_z,interval_block_click_z)	0.66	0.15	0.31	0.90	1.02
cor(Intercept,deadline_type_z)	−0.84	0.11	−0.97	−0.57	1.00
cor(Intercept,department comp.sci)	−0.97	0.02	−0.99	−0.93	1.01
cor(deadline_type_z,department comp.sci)	0.88	0.09	0.65	0.98	1.00
Random effects (student)					
sd(Intercept)	1.35	0.67	0.09	2.67	1.01
sd(n_clicks_assign_z)	4.65	0.74	3.27	6.14	1.00
sd(interval_block_click_z)	3.74	0.73	2.33	5.25	1.00
sd(n_clicks_relevant_activity_z)	0.82	0.61	0.04	2.28	1.00
sd(deadline_type_z)	0.83	0.57	0.04	2.12	1.01
cor(n_clicks_assign_z,interval_block_click_z)	0.69	0.14	0.37	0.91	1.00

Note. Estimates, standard errors, lower and upper ends of credible intervals, and Rhats for fixed effects and random effects (both grouping factors). For the random effects, the estimates which include 0 in their credible intervals are not shown.

Table 9
Fixed effects in $model_{allRI}$.

	Estimate	Estimated error	Lower 95% CI	Upper 95% CI	Rhat
Fixed effects					
Intercept	0.03	3.47	−7.03	6.58	1.01
GASE_z	0.87	0.94	−0.98	2.69	1.00
SDLS_z	−2.06	0.79	−3.59	−0.45	1.00
APSS_z	0.86	0.64	−0.36	2.12	1.00
APS_z	0.16	0.62	−1.06	1.36	1.00
n_clicks_assign_z	4.68	0.45	3.81	5.58	1.00
interval_block_click_z	6.10	0.52	5.10	7.15	1.00
n_clicks_relevant_activity_z	2.12	1.03	0.10	4.15	1.00
age_z	−0.99	0.67	−2.30	0.34	1.00
gender_z	−1.03	0.71	−2.47	0.38	1.00
study_exp_z	3.27	1.45	0.37	6.10	1.00
deadline_type_z	−0.34	0.53	−1.35	0.68	1.00
department comp.sci	−2.61	4.39	−10.90	6.22	1.01
department economics	3.94	4.43	−4.46	13.06	1.00

Note. Estimates, standard errors, lower and upper ends of credible intervals, and Rhats for fixed effects.

Appendix B. Tables for models predicting score

Table 10
Fixed effects in $model_{quest}$.

	Estimate	Estimated error	Lower 95% CI	Upper 95% CI	Rhat
Fixed effects					
Intercept	−0.17	0.03	−0.22	−0.12	1
delay_z	−0.01	0.01	−0.03	0.01	1
GASE_z	0.00	0.01	−0.03	0.02	1
SDLS_z	0.03	0.01	0.01	0.05	1
APSS_z	0.03	0.01	0.01	0.05	1
APS_z	−0.01	0.01	−0.03	0.01	1
age_z	0.01	0.01	−0.01	0.03	1
gender_z	−0.01	0.01	−0.03	0.01	1
study_exp_z	−0.02	0.01	−0.04	0.00	1
deadline_type_z	0.01	0.01	−0.01	0.02	1
department comp.sci	0.09	0.03	0.03	0.15	1
department economics	0.01	0.04	−0.06	0.08	1

Note. Estimates, standard errors, lower and upper ends of credible intervals, and Rhats for fixed effects.

Table 11
Fixed effects in $model_{logRI}$.

	Estimate	Estimated error	Lower 95% CI	Upper 95% CI	Rhat
Fixed effects					
Intercept	−0.17	0.04	−0.25	−0.10	1
delay_z	0.00	0.01	−0.02	0.02	1
n_clicks_assign_z	0.00	0.01	−0.02	0.02	1
interval_block_click_z	0.00	0.01	−0.02	0.02	1
n_clicks_relevant_activity_z	0.00	0.02	−0.03	0.03	1
age_z	0.02	0.01	−0.01	0.04	1
gender_z	−0.03	0.01	−0.05	0.00	1
study_exp_z	−0.03	0.02	−0.06	0.01	1
deadline_type_z	0.01	0.01	−0.01	0.04	1
department comp.sci	0.08	0.05	−0.01	0.17	1
department economics	0.00	0.06	−0.11	0.12	1

Note. Estimates, standard errors, lower and upper ends of credible intervals, and Rhats for fixed effects.

Table 12
Fixed and random effects in $model_{logRS}$

	Estimate	Estimated error	Lower 95% CI	Upper 95% CI	Rhat
Fixed effects					
Intercept	−0.16	0.04	−0.23	−0.09	1.00
delay_z	−0.01	0.01	−0.03	0.01	1.00
n_clicks_assign_z	0.00	0.01	−0.02	0.02	1.00
interval_block_click_z	0.00	0.01	−0.02	0.03	1.00
n_clicks_relevant_activity_z	−0.01	0.02	−0.05	0.03	1.00
age_z	0.02	0.01	−0.01	0.04	1.00
gender_z	−0.02	0.01	−0.05	0.00	1.00
study_exp_z	−0.08	0.03	−0.13	−0.01	1.00

(*continued on next page*)

Table 12 (*continued*)

	Estimate	Estimated error	Lower 95% CI	Upper 95% CI	Rhat
deadline_type_z	0.01	0.01	−0.01	0.04	1.00
department comp.sci	0.04	0.08	−0.11	0.19	1.00
department economics	−0.07	0.08	−0.22	0.08	1.00
Random effects (course)					
sd(department comp.sci)	0.37	0.08	0.22	0.53	1.01
sd(department economics)	0.12	0.06	0.02	0.26	1.01

Note. Estimates, standard errors, lower and upper ends of credible intervals, and Rhats for fixed effects and random effects (course only). For the random effects, the estimates which include 0 in their credible intervals are not shown.

Table 13
Fixed effects in $model_{allRI}$.

	Estimate	Estimated error	Lower 95% CI	Upper 95% CI	Rhat
Fixed effects					
Intercept	−0.17	0.04	−0.25	−0.10	1
delay_z	0.00	0.01	−0.02	0.02	1
GASE_z	−0.02	0.01	−0.05	0.01	1
SDLS_z	0.03	0.01	0.01	0.05	1
APSS_z	0.02	0.01	0.00	0.04	1
APS_z	0.00	0.01	−0.02	0.02	1
n_clicks_assign_z	0.00	0.01	−0.02	0.01	1
interval_block_click_z	0.00	0.01	−0.02	0.02	1
n_clicks_relevant_activity_z	0.00	0.02	−0.03	0.03	1
age_z	0.01	0.01	−0.01	0.03	1
gender_z	−0.02	0.01	−0.04	0.01	1
study_exp_z	−0.03	0.02	−0.06	0.01	1
deadline_type_z	0.02	0.01	0.00	0.04	1
department comp.sci	0.09	0.05	−0.01	0.18	1
department economics	0.00	0.07	−0.13	0.13	1

Note. Estimates, standard errors, lower and upper ends of credible intervals, and Rhats for fixed effects.

Table 14
Fixed and random effects in $model_{allRS}$.

	Estimate	Estimated error	Lower 95% CI	Upper 95% CI	Rhat
Fixed effects					
Intercept	−0.17	0.03	−0.23	−0.10	1.00
delay_z	0.00	0.01	−0.02	0.02	1.00
GASE_z	−0.02	0.02	−0.05	0.01	1.00
SDLS_z	0.04	0.01	0.01	0.06	1.00
APSS_z	0.02	0.01	−0.01	0.04	1.00
APS_z	0.00	0.01	−0.02	0.02	1.00
n_clicks_assign_z	0.00	0.01	−0.02	0.02	1.00
interval_block_click_z	0.01	0.01	−0.02	0.03	1.00
n_clicks_relevant_activity_z	−0.01	0.02	−0.05	0.03	1.00
age_z	0.01	0.01	−0.01	0.04	1.00
gender_z	−0.01	0.01	−0.04	0.02	1.00
study_exp_z	−0.08	0.03	−0.14	−0.02	1.00
deadline_type_z	0.02	0.01	−0.01	0.04	1.00
department comp.sci	0.05	0.08	−0.11	0.19	1.00
department economics	−0.07	0.08	−0.23	0.09	1.00
Random effects (course)					
sd(department comp.sci)	0.38	0.08	0.22	0.53	1.01
sd(department economics)	0.12	0.06	0.01	0.27	1.00

Note. Estimates, standard errors, lower and upper ends of credible intervals, and Rhats for fixed effects and random effects (course only). For the random effects, the estimates which include 0 in their credible intervals are not shown.

References

Ackerman, D. S., & Gross, B. L. (2005). My instructor made me do it: Task characteristics of procrastination. *Journal of Marketing Education, 27*(1), 5–13. https://doi.org/10.1177/0273475304273842.

Akram, A., Fu, C., Li, Y., Javed, M. Y., Lin, R., Jiang, Y., & Tang, Y. (2019). Predicting students' academic procrastination in blended learning course using homework submission data. *IEEE Access, 7*, 102487–102498. https://doi.org/10.1109/access.2019.2930867.

Antunes, O., Lucena, K. K. T., Oliveira, E. H. T., Gusmão, T. H. Y., Nozawa, E. H., & Brodbeck, W. (2016). LMS monitor: An application for academic and performance monitoring of students in LMS. In *, Vol. 12. XXI congreso internacional de informática educativa (TISE)* (pp. 45–52).

Aust, F., & Barth, M. (2020). papaja: Create APA manuscripts with R markdown. Retrieved from https://github.com/crsh/papaja.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory* (pp. xiii, 617–xiii, 617). Englewood Cliffs, NJ, US: Prentice-Hall, Inc..

Blunt, A., & Pychyl, T. A. (1998). Volitional action and inaction in the lives of undergraduate students: State orientation, procrastination and proneness to boredom. *Personality and Individual Differences, 24*(6), 837–846. https://doi.org/10.1016/S0191-8869(98)00018-X.

Bolhuis, S. (1996). Towards active and self-directed learning. Preparing for lifelong learning, with reference to Dutch secondary education. Retrieved from https://eric.ed.gov/?id=ED396141.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01.

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal, 10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017.
Cao, L. (2012). Examining "active" procrastination from a self-regulated learning perspective. *Educational Psychology, 32*(4), 515–545. https://doi.org/10.1080/01443410.2012.663722.
Cerezo, R., Esteban, M., Sánchez-Santillán, M., & Núñez, J. C. (2017). Procrastinating behavior in computer-based learning environments to predict performance: A case study in Moodle. *Frontiers in Psychology, 8*(AUG), 1–11. https://doi.org/10.3389/fpsyg.2017.01403.
Choi, J. N., & Moran, S. V. (2009). Why not procrastinate? Development and validation of a new active procrastination scale. *The Journal of Social Psychology, 149*(2), 195–211. https://doi.org/10.3200/SOCP.149.2.195-212.
Chowdhury, S. F., & Pychyl, T. A. (2018). A critique of the construct validity of active procrastination. *Personality and Individual Differences, 120*, 7–12. https://doi.org/10.1016/j.paid.2017.08.016.
Chu, A. H. C., & Choi, J. N. (2005). Rethinking procrastination: Positive effects of "active" procrastination behavior on attitudes and performance. *The Journal of Social Psychology, 145*(3), 245–264. https://doi.org/10.3200/SOCP.145.3.245-264.
Corkin, D. M., Yu, S. L., & Lindt, S. F. (2011). Comparing active delay and procrastination from a self-regulated learning perspective. *Learning and Individual Differences, 21*(5), 602–606. https://doi.org/10.1016/j.lindif.2011.07.005.
Corno, L., & Mandinach, E. B. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist, 18*(2), 88–108. https://doi.org/10.1080/00461528309529266.
Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software, 64*(4), 1–34. Retrieved from http://www.jstatsoft.org/v64/i04/.
Doherty, W. (2006). An analysis of multiple factors affecting retention in web-based community college courses. *The Internet and Higher Education, 9*(4), 245–255. https://doi.org/10.1016/J.IHEDUC.2006.08.004.
Ellis, A., & Knaus, W. J. (1977). Overcoming procrastination: Or how to think and act rationally in spite of life's inevitable hassles (p. 180). Institute for Rational Living. Retrieved from https://books.google.ch/books/about/Overcoming_Procrastination.html?id=ZWpynQAACAAJ&redir_esc=y.
Ferrari, J. R. (1991). Compulsive procrastination: Some self-reported characteristics. *Psychological Reports, 68*(2), 455–458. https://doi.org/10.2466/pr0.1991.68.2.455.
Ferrari, J. R. (2011). *Still procrastinating?: The no-regrets guide to getting it done*. John Wiley & Sons.
Garrison, D. R. (1997). Self-directed learning: Toward a comprehensive model. *Adult Education Quarterly, 48*(1), 18–33. https://doi.org/10.1177/074171369704800103.
Gohel, D. (2020). Flextable: Functions for tabular reporting. Retrieved from https://CRAN.R-project.org/package=flextable.
Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software, 40*(3), 1–25. Retrieved from http://www.jstatsoft.org/v40/i03/.
Grunschel, C., Patrzek, J., & Fries, S. (2013). Exploring different types of academic delayers: A latent profile analysis. *Learning and Individual Differences, 23*(1), 225–233. https://doi.org/10.1016/j.lindif.2012.09.014.
Harriott, J., & Ferrari, J. R. (1996). Prevalence of procrastination among samples of adults. *Psychological Reports, 78*(2), 611–616. https://doi.org/10.2466/pr0.1996.78.2.611.
Hensley, L. C. (2014). Reconsidering active procrastination: Relations to motivation and achievement in college anatomy. *Learning and Individual Differences, 36*, 157–164. https://doi.org/10.1016/j.lindif.2014.10.012.
Howell, A. J., & Watson, D. C. (2007). Procrastination: Associations with achievement goal orientation and learning strategies. *Personality and Individual Differences, 43*(1), 167–178. https://doi.org/10.1016/j.paid.2006.11.017.
Huang, N., Zhang, J., Burtch, G., Li, X., & Chen, P. (2021). Combating procrastination on massive online open courses via optimal calls to action. *Information Systems Research.*, Article isre.2020.0974. https://doi.org/10.1287/isre.2020.0974.
James, D., & Hornik, K. (2020). Chron: Chronological objects which can handle dates and times. Retrieved from https://CRAN.R-project.org/package=chron.
Kassambara, A. (2020). Ggpubr: 'Ggplot2' based publication ready plots. Retrieved from https://CRAN.R-project.org/package=ggpubr.
Klingsieck, K. B. (2013). Procrastination when good things don't come to those who wait. *European Psychologist, 18*(1), 24–34. https://doi.org/10.1027/1016-9040/a000138.
Klingsieck, K. B., Fries, S., Horz, C., & Hofer, M. (2012). Procrastination in a distance university setting. *Distance Education, 33*(3), 295–310. https://doi.org/10.1080/01587919.2012.723165.
Knowles, M. S. (1975). Self-directed learning: A guide for learners and teachers (p. 135). Chicago: Association Press. Retrieved from https://search.library.wisc.edu/catalog/999499643502121.
Komsta, L., & Novomestky, F. (2015). Moments: Moments, cumulants, skewness, kurtosis and related tests. Retrieved from https://CRAN.R-project.org/package=moments.
Levy, Y., & Ramim, M. M. (2012). A study of online exams procrastination using data analytics techniques. *Interdisciplinary Journal of E-Skills and Lifelong Learning, 8*, 097–113. https://doi.org/10.28945/1730.
Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education, 6*(1), 8. https://doi.org/10.1186/s40536-018-0061-2.
Lounsbury, J. W., Levy, J. J., Park, S. H., Gibson, L. W., & Smith, R. (2009). An investigation of the construct validity of the personality trait of self-directed learning. *Learning and Individual Differences, 19*(4), 411–418. https://doi.org/10.1016/j.lindif.2009.03.001.
Lüdecke, D. (2020). SjPlot: Data visualization for statistics in social science. Retrieved from https://CRAN.R-project.org/package=sjPlot.
Lüdecke, D., Makowski, D., Waggoner, P., & Patil, I. (2020). Performance: Assessment of regression models performance. *CRAN.*. https://doi.org/10.5281/zenodo.3952174.
Malatincová, T. (2015). The mystery of "should": Procrastination, delay, and reactance in academic settings. *Personality and Individual Differences, 72*, 52–58. https://doi.org/10.1016/j.paid.2014.08.015.
McCloskey, J. D. (2012). *Finally, my thesis on academic procrastination*. PhD thesis.
Meier, A., Reinecke, L., & Meltzer, C. E. (2016). Facebocrastination? Predictors of using Facebook for procrastination and its effects on students' well-being. *Computers in Human Behavior, 64*, 65–76. https://doi.org/10.1016/j.chb.2016.06.011.
Mentz, E., & Van Zyl, S. (2018). The impact of cooperative learning on self-directed learning abilities in the computer applications technology class. *International Journal of Lifelong Education, 37*(4), 482–494. https://doi.org/10.1080/02601370.2018.1513426.
Morris, L. V., Finnegan, C., & Wu, S. S. (2005). Tracking student behavior, persistence, and achievement in online courses. *Internet and Higher Education, 8*(3), 221–231. https://doi.org/10.1016/j.iheduc.2005.06.009.
Myrick, J. G. (2015). Emotion regulation, procrastination, and watching cat videos online: Who watches internet cats, why, and to what effect? *Computers in Human Behavior, 52*, 168–176. https://doi.org/10.1016/j.chb.2015.06.001.
Nielsen, T., Dammeyer, J., Vang, M. L., & Makransky, G. (2018). Gender fairness in self-efficacy? A Rasch-based validity study of the general academic self-efficacy scale (GASE). *Scandinavian Journal of Educational Research, 62*(5), 664–681. https://doi.org/10.1080/00313831.2017.1306796.
Nordby, K., Klingsieck, K. B., & Svartdal, F. (2017). Do procrastination-friendly environments make students delay unnecessarily? *Social Psychology of Education, 20*(3), 491–512. https://doi.org/10.1007/s11218-017-9386-x.
Onji, K., & Kikuchi, R. (2011). Procrastination, prompts, and preferences: Evidence from daily records of self-directed learning activities. *Journal of Socio-Economics, 40*(6), 929–941. https://doi.org/10.1016/j.socec.2011.08.019.
Pinxten, M., De Laet, T., Van Soom, C., Peeters, C., & Langie, G. (2019). Purposeful delay and academic achievement. A critical review of the Active Procrastination Scale. *Learning and Individual Differences, 73*, 42–51. https://doi.org/10.1016/j.lindif.2019.04.010.
del Puerto Paule-Ruiz, M., Riestra-González, M., Sánchez-Santillán, M., & Pérez-Pérez, J. R. (2015). The procrastination related indicators in e-learning platforms. *Journal of Universal Computer Science, 21*(1), 7–22. https://doi.org/10.3217/jucs-021-01-0007.
Pychyl, T. A. (2008). Savouring the flavours of delay. *English Studies in Canada, 34*(2–3), 25. Retrieved from https://www.questia.com/read/1G1-214398915/savouring-the-flavours-of-delay.
Pychyl, T. A., Lee, J. M., Thibodeau, R., & Blunt, A. (2000). Five days of emotion: An experience sampling study of undergraduate student procrastination. *Journal of Social Behavior & Personality, 15*(5), 239–254.
R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.
Rabin, L. A., Fogel, J., & Nutter-Upham, K. E. (2011). Academic procrastination in college students: The role of self-reported executive function. *Journal of Clinical and Experimental Neuropsychology, 33*(3), 344–357. https://doi.org/10.1080/13803395.2010.518597.
Revelle, W. (2020). *Psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych.
Saeid, N., & Eslaminejad, T. (2017). Relationship between student's self-directed-learning readiness and academic self-efficacy and achievement motivation in students. *International Education Studies, 10*(1). https://doi.org/10.5539/ies.v10n1p225.
Schommer-Aikins, M., & Easter, M. (2018). Cognitive flexibility, procrastination, and need for closure linked to online self-directed learning among students taking online courses. *Journal of Business and Educational Leadership, 8*(1), 112–121.
Schraw, G., Wadkins, T., & Olafson, L. (2007). Doing the things we do: A grounded theory of academic procrastination. *Journal of Educational Psychology, 99*(1), 12–25. https://doi.org/10.1037/0022-0663.99.1.12.
Senécal, C., Koestner, R., & Vallerand, R. J. (1995). Self-regulation and academic procrastination. *Journal of Social Psychology, 135*(5), 607–619. https://doi.org/10.1080/00224545.1995.9712234.
Sirois, F. M., Melia-Gordon, M. L., & Pychyl, T. A. (2003). "I'll look after my health, later": An investigation of procrastination and health. *Personality and Individual Differences, 35*(5), 1167–1184. https://doi.org/10.1016/S0191-8869(02)00326-4.
Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin, 133*(1), 65–94. https://doi.org/10.1037/0033-2909.133.1.65.
Sundaramoorthy, J. (2018). Self-regulated learning strategies of active procrastinating pre-university students. *Electronic Journal of Research in Educational Psychology, 16*(45), 325–343. https://doi.org/10.25115/ejrep.v16i45.2096.
Tice, D. M., & Baumeister, R. F. (1997). Longitudinal study of procrastination, performance, stress, and health: The costs and benefits of dawdling. *Psychological Science, 8*(6), 454–458. https://doi.org/10.1111/j.1467-9280.1997.tb00460.x.
Tuckman, B. W. (2005). Relations of academic procrastination, rationalizations, and performance in a web course with deadlines. *Psychological Reports, 96*(3_suppl), 1015–1021. https://doi.org/10.2466/pr0.96.3c.1015-1021.
Wäschle, K., Allgaier, A., Lachner, A., Fink, S., & Nückles, M. (2014). Procrastination and self-efficacy: Tracing vicious and virtuous circles in self-regulated learning. *Learning and Instruction, 29*, 103–114. https://doi.org/10.1016/j.learninstruc.2013.09.005.
Wickham, H. (2016). Ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org.

Wickham, H. (2019). Stringr: Simple, consistent wrappers for common string operations. Retrieved from https://CRAN.R-project.org/package=stringr.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686.

Wighting, M. J., Liu, J., & Rovai, A. P. (2008). Distinguishing sense of community and motivation characteristics between online and traditional college students. *Quarterly Review of Distance Education, 9*(3), 285–295.

Wolters, C. A. (2003). Understanding procrastination from a self-regulated learning perspective. *Journal of Educational Psychology, 95*(1), 179–187. https://doi.org/10.1037/0022-0663.95.1.179.

Xie, Y. (2015). Dynamic documents with R and knitr (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://yihui.org/knitr/.

Yamada, M., Oi, M., & Konomi, S. (2017). Are learning logs related to procrastination? From the viewpoint of self-regulated learning. In *14th international conference on cognition and exploratory learning in the digital age, celda 2017 — Vilamoura, Algarve, Portugal* (pp. 3–10).

Yang, Y., Hooshyar, D., Pedaste, M., Wang, M., Huang, Y. M., & Lim, H. (2020a). Predicting course achievement of university students based on their procrastination behaviour on Moodle. *Soft Computing*, 1–17. https://doi.org/10.1007/s00500-020-05110-4.

Yang, Y., Hooshyar, D., Pedaste, M., Wang, M., Huang, Y. M., & Lim, H. (2020b). Prediction of students' procrastination behaviour through their submission behavioural pattern in online learning. *Journal of Ambient Intelligence and Humanized Computing.*. https://doi.org/10.1007/s12652-020-02041-8.

Yockey, R. D. (2016). Validation of the short form of the academic procrastination scale. *Psychological Reports, 118*(1), 171–179. https://doi.org/10.1177/0033294115626825.

You, J. W. (2015). Examining the effect of academic procrastination on achievement using LMS in e-learning. *Educational Technology & Society, 18*(3), 64–74. Retrieved from http://eds.a.ebscohost.com/eds/detail/detail?vid=2&sid=c055a577-bc0c-4d81-9954-f30273da2472%40sessionmgr4001&hid=4202&bdata=Jmxhbmc9cHQtYnImc2l0ZT1lZHMtbGl2ZSZzY29wZT1zaXRl#AN=109155621&db=a9h.

Prediction of Dilatory Behavior in eLearning: A Comparison of Multiple Machine Learning Models

A Preprint

Christof Imhof, Ioan-Sorin Comsa, Martin Hlosta, Behnam Parsaeifard, Ivan Moser, and Per Bergamin

June 29, 2022

Abstract

Procrastination, the irrational delay of tasks, is a common occurrence in online learning. Potential negative consequences include higher risk of drop-outs, increased stress, and reduced mood. Due to the rise of learning management systems and learning analytics, indicators of such behavior can be detected, enabling predictions of future procrastination and other dilatory behavior. However, research focusing on such predictions is scarce. Moreover, studies involving different types of predictors and comparisons between the predictive performance of various methods are virtually non-existent. In this study, we aim to fill these research gaps by analyzing the performance of multiple machine learning algorithms when predicting the delayed or timely submission of online assignments in a higher education setting with two categories of predictors: subjective, questionnaire-based variables and objective, log-data based indicators extracted from a learning management system. The results show that models with objective predictors consistently outperform models with subjective predictors, and a combination of both variable types perform slightly better. For each of these three options, a different approach prevailed (Gradient Boosting Machines for the subjective, Bayesian multilevel models for the objective, and Random Forest for the combined predictors). We conclude that careful attention should be paid to the selection of predictors and algorithms before implementing such models in learning management systems.

Keywords Procrastination, dilatory behavior, machine learning, learning analytics, predictive performance

C. Imhof, I.-S. Comsa, M. Hlosta, B. Parsaeifard, I. Moser, and P. Bergamin are with the Institute for Research in Open-, Distance- and eLearning, Swiss Distance University of Applied Sciences, Brig, CH-3900, Switzerland (e-mail addresses: christof.imhof@ffhs.ch, ioan-sorin.comsa@ffhs.ch, martin.hlosta@ffhs.ch, behnam.parsaeifard@ffhs.ch, ivan.moser@ffhs.ch, per.bergamin@ffhs.ch). P. Bergamin is also affiliated with the North-West University, Potchefstroom, 2531, South Africa.

1 Introduction

The ubiquity of digitization continues to pervade many aspects of life, including learning. While this ongoing trend entails numerous advantages for students, such as more flexible schedules, more engaging distance education or a greater variety of interactive learning tasks Dumford and Miller [2018], drawbacks also need to be considered. One potential drawback is the increased likelihood of academic procrastination due to the many distractions our digital world offers. *Academic procrastination* is a type of dilatory behavior that is commonly defined as the irrational delay of academic tasks such as writing assignments or studying course literature Steel [2007]. While seemingly harmless, procrastination may lead to various negative consequences, which include increased stress, lower performance, and reduced mood Tice and Baumeister [1997]. Moreover, this behavior is linked to increased risk of drop-out Doherty [2006] and is reportedly highly prevalent among students in higher education Ellis and Knaus [1977], Steel [2007].

However, not all acts of delay are necessarily maladaptive since they can also be used as a deliberate strategy. The positive, productive counterpart to procrastination is known as *purposeful delay* Corkin et al. [2011a]. Alternative names for this concept include active delay, strategic delay or active procrastination, the latter of which is considered to be an oxymoron Pychyl [2008]. The term *dilatory behavior* thus serves as an umbrella term that refers to both positive and negative forms of delay.

A promising approach to measuring indicators of dilatory behavior comes in the form of *Learning Analytics* (LA). LA is often implemented as a conceptual framework to analyse course characteristics, such as prediction of students' learning performance, educational data analysis, data collection and measurement, and early intervention Hwang et al. [2017]. Given the potential consequences of negative dilatory behavior, the ability to detect it early utilizing an LA framework could prove very useful to teachers, lecturers, and students alike. This especially concerns learning management systems (LMS), which are able to provide objective indicators for such behavior. The earlier the detection, the more time remains to intervene and to ultimately prevent potential drop-outs.

Predicting dilatory behavior in academic tasks is not trivial for multiple reasons, especially in the field: various courses must be incorporated to allow for a generalization of such predictions and their timing also needs to be considered. In this context, data collection and the type of variables (be they predictors or outcome variables) play a crucial role. Despite the variety of types of *Machine Learning* (ML) models that could be employed, their predictive performance may vary as a result of the size of the collected data set, the number of predictor variables, and the point in time predictions are made. Therefore, the selection of an appropriate prediction model must involve such considerations.

Previous studies indicate the great potential of ML-based prediction models in learning analytics Azimi et al. [2020], Pardo et al. [2017], Conijn et al. [2017], even when predicting dilatory behavior Akram et al. [2019]. To our knowledge, however, comparisons between multiple ML algorithms are rather rare in this context (see Akram et al. [2019] and Cerezo et al. [2017]). Moreover, delay is usually treated as a predictor for other outcome variables rather than being the outcome variable itself. To fill this gap, we employ eight different ML algorithms to predict dilatory behavior in online assignments. For our analyses, we used the same data set as in a prior study, where the prediction of dilatory behavior was investigated, comparing models with subjective and objective predictors using Bayesian multilevel regression Imhof et al. [2021]. We provide a comparative analysis to determine which algorithms deliver the best predictions and compare the performance with the Bayesian models as a baseline.

2 Related Work

In this section, we present the related work on dilatory behavior and the machine learning algorithms employed to predict such tendencies in various tasks. The analytical framework, research questions, and contributions of this paper are also detailed.

2.1 Theoretical Background of Dilatory Behavior

As stated in the introduction, dilatory behavior is an umbrella term that encompasses at least two types of delay, procrastination and purposeful delay. However, another important distinction should be made when investigating delay, namely the one between trait and state delay. *State* variables are situational, whereas *trait* variables are stable with less situation- or time-specific variance. In the case of procrastination, the correlation between trait and state (or in other words, attitude and actual behavior) is reportedly medium to large at .51 Steel [2007]. This not only implies that situational factors are important when trying to predict delay, but also that trait procrastination can still serve as a predictor for state procrastination (or vice versa), albeit not a particularly strong one.

This train of thought leads to the question, which additional predictors besides trait procrastination should then be considered when predicting procrastination and other types of dilatory behavior. Two broad categories present themselves: subjective and objective predictors. The former are commonly assessed with questionnaires, while the latter can be gained by observation, or in the case of LMS, collection of log data.

When predicting delay with subjective predictors, motivational aspects and learning-related factors suggest themselves. One option are factors in the orbit of self-regulation, given that procrastination is often being considered a "self-regulation failure" Steel [2007] whereas purposeful delay appears as its successful counterpart Corkin et al. [2011b]. Learning concepts related to self-regulation include academic self-efficacy and self-directed learning. *Academic self-efficacy* refers to the belief in one's ability to succeed at academic tasks such as writing exams Bandura [1986] and is negatively linked to procrastination Wäschle et al. [2014], forming a vicious circle, but positively connected to purposeful delay, counteracting that circle Chu and Choi [2005].

Self-directed learning (SDL) is a process in which students take responsibility for their own learning, which involves self-monitoring, self-management, and self-evaluation of the learning process Bolhuis [1996], Knowles [1975], Garrison [1997]. Research reveals positive links between SDL and self-efficacy Saeid and Eslaminejad [2017] and negative links with procrastination Schommer-Aikins and Easter [2018].

In the context of LMS, it is self-evident to consider objective predictors based on data extracted from such systems. Popular approaches in learning analytics are to analyze click-based data (e.g., Cirigliano et al. [2020], Knight et al. [2017]), which provide insights into student activity and engagement with learning content on the platform.

2.2 Delay-related Prediction Models in Learning Analytics

Prediction models that involve delay as a predictor are quite common in the literature. A variety of methods have been implemented to investigate the relationship between procrastination (as the most commonly researched type of delay) and other variables, usually achievement. A first example is You [2015], who used multiple regression analysis to investigate the effects of procrastination on course achievement. Their regression model predicted achievement at each point in time, with the predictability increasing as time passed. The authors in Levy and Ramim [2012] instead implemented data analytics techniques to detect anomalies in their data, corroborating the finding that procrastination negatively predicts performance (grades). The same conclusion was reached by the authors in Gareau et al. [2019], who used Structure Equation Modeling and discovered a mediating role of task-oriented and disengagement-oriented coping. Another technique, path analysis, was applied in Yamada et al. [2016], where the authors found that positive time management promoted early submissions of paper reports. A different method, association rule mining, was implemented in del Puerto Paule-Ruiz et al. [2015] and Cerezo et al. [2017]. Both research groups identified indicators of procrastination on Moodle, which were used to create association rules in the form of conditionals by implementing either Apriori and/or Predictive Apriori algorithms. In both studies, time-related indicators were closely related to students' performance. In contrast, models that explicitly try to predict delay as the outcome variable are rather scarce. Exceptions include studies conducted in Akram et al. [2019], Hooshyar et al. [2020], Abidi et al. [2020], and Yang et al. [2020], who all intended to classify students based on homework submission data. Ten different ML algorithms (ZeroR, OneR, ID3, J48, Random Forest, decision stump, JRip, PART, NBTree, and Prism) were implemented in Akram et al. [2019] to classify students as procrastinators or non-procrastinators based on feature vectors. The optimal amount of clusters was determined to be three (one non-procrastinating group and two procrastinating classes that differ in submission scores).

Multiple clustering algorithms were also used in Hooshyar et al. [2020] and Yang et al. [2020] to detect the optimal amount of clusters, followed by a classification of students into the three resulting clusters (procrastinators, non-procrastinators, and procrastinator candidates). In both studies, the authors compared eight methods of classification (linear and radial basis function kernel support vector machine, Gaussian Processes, decision tree, Random Forest, neural network, AdaBoost, and Naive Bayes). The former implemented a feature vector algorithm involving categorical and continuous features based on spare time (i.e., interval between submission and deadline) and inactive time (i.e., time before the first click on an assignment), and found that neural networks worked best with categorical features and that linear support vector machines outperformed the others in the case of continuous features. The latter reported that the linear support vector machine approach delivered the best predictive performance for their clusters.

Binary classifiers were employed in Abidi et al. [2020], using four supervised-learning algorithms (logistic regression, decision tree, gradient boosting, and Random Forest) to classify students as procrastinators or non-procrastinators based on data that was extracted from an intelligent tutoring system (ITS). Among the ML algorithms, gradient boosting had the best performance in terms of classification precision. However, these examples involve a classification of students into clusters, rather than predictions regarding the extent of the delay itself. This research gap was one of the reasons we conducted a previous study to determine if dilatory behavior could be predicted based on a mixture of questionnaire scores and log data Imhof et al. [2021].

In that study, seven predictors were implemented across six different Bayesian multilevel models to determine which type of predictor (objective vs. subjective) and which individual predictors were able to predict dilatory behavior the best. The four subjective predictors were questionnaire scores and the three objective predictors were based on log data. The model fit comparison favored the models that included all seven predictors, but their advantage over the models that only included objective predictors was minimal at best. This implies that the models with objective predictors barely improve when subjective predictors are added. However, this result does not imply that subjective predictors should be discarded altogether. First off, the results may not be the same when following other approaches for prediction models, for instance models based on ML. Secondly, we focused on individual predictors and comparisons between types of predictors rather than analyzing the actual performance of the predictions in terms of accuracy and other measures.

2.3 Research Questions

Therefore, the goal of this study was to extend the findings reported in Imhof et al. [2021] by comparing and contrasting the performance of different approaches to prediction models, determining which ML algorithm delivers the best predictions for delay and for which type of predictor.

Before the predictive performance of these models can be assessed, it needs to be clarified how the hyperparameters for each ML algorithm and type of predictor must be determined in order for them to be optimized. We present a novel cross-validation approach to address this and compare the results of the cross-validation with the rest results. We then intend to identify the ML model with the best predictive performance for the following three predictor types: subjective (`subj`), objective (`obj`), and a combination between the two (`comb`). Finally, we determine whether models with objective predictors still outperform models with subjective predictors when calculated with various ML algorithms and whether there is still an advantage of models that combine both types of predictors for all of our algorithms. The research questions are thus as follows:

RQ1: How well do the results of the cross-validation with the proposed measure compare to the test results?

RQ2: Which machine learning algorithm delivers an improvement of the predictive performance compared to the baseline models based on the subjective, objective, and combined sets of predictors (intra-model comparison)?

RQ3: Which type of predictors (subjective, objective, and combined) allow for the highest predictive performance (inter-model comparison)?

2.4 Paper Organization

The rest of the paper is organized as follows: In Section III, we present the methods and instruments used to collect the data. Then, we introduce the prediction framework and briefly illustrate each ML algorithm. A novel cross-validation procedure is also presented to determine the best set of hyperparameters for all of the models. Next, we present our results in Section IV, which includes the results of the cross-validations and the test results of each ML model, which are compared with the performance of the Bayesian multilevel models. This is followed by a discussion of our results in Section V, including implications and limitations of our findings. Finally, the paper concludes with Section VI.

3 Methods and Prediction Models

3.1 Participants and Courses

We used the same data set as in a prior study, which included 134 students from a distance university located in Central Europe Imhof et al. [2021]. The sample included 65 male and 69 female participants (mean age 31.61 years, sd = 7.97, min = 19, max = 61). The students were enrolled in at least one course each during the autumn semester of 2019. The number of involved online courses was 126, each belonging to one of three departments (computer science, economics, and health). Every course consisted of several blocks (usually between five and ten), each with their own assignment(s). The total number of assignments across all students and courses was 1107. The students participated voluntarily and consented to have their log data extracted from the institution's LMS by filling in an online survey. As compensation, all participants automatically entered a raffle with a chance of winning cinema vouchers.

3.2 Instruments, Procedures, and Variables

As described in Imhof et al. [2021], the procedure started with an e-mail invitation that was sent to all students enrolled at our institution at the end of the semester. The volunteers then followed a link to an online survey, consisting of four questionnaires, whose scores formed the subjective predictors for our models: the General Academic Self-Efficacy (GASE) Scale Nielsen et al. [2018], consisting of five items, the Self-Directed Learning Scale (SDLS) Lounsbury et al. [2009] with ten items, the Academic Procrastination Scale - short form (APSS) McCloskey [2012] with five items, and the Active Procrastination Scale (APS) Choi and Moran [2009] with four subscales and a total of 16 items.

The three objective predictors, which the authors selected based on log data variables used in other studies del Puerto Paule-Ruiz et al. [2015], Cerezo et al. [2017], were the number of clicks on an assignment, the interval between the start of a block and the first click on an assignment, and the number of clicks on relevant activities in the course. The *number of clicks on an assignment* reflects the sum of all clicks on the assignment section of the course made before the deadline had passed. The second predictor, the *interval between the start of a block and the first click on an assignment*, indicates the time (in days) that passed between when a block had started and when the first click on the assignment was made (i.e., when the task description was first read). The final objective predictor was the *number of clicks on relevant activities in the course*, which reflected the overall engagement with the course material on the platform (i.e., the sum of all clicks on learning videos, forums, interactive books, etc.). The interval turned out to be the strongest and most consistent individual predictor in the previous study, followed by the number of clicks on the assignment.

The to-be-predicted outcome variable was *delay*, which can assume positive and negative values, with positive values indicating a delayed submission of an assignment (in hours) and negative values meaning a timely submission. In this study, *delay* was used for regression and classification alike, the former to determine the error between predicted and

Table 1: Description of Subjective and Objective Predictors and Outcome variable

Variable	Role	Mean	SD	Min	Max	Description
GASE	subjective predictor	19.71	3.10	8	25	General Academic Self-Efficacy Scale, measures belief in one's academic abilities
SDLS	subjective predictor	38.74	5.29	18	50	Self-Directed Learning Scale, measures how much one feels in charge of one's learning process
APSS	subjective predictor	11.04	4.16	5	21	Academic Procrastination Scale – Short Form, measures tendency to procrastinate in academic tasks
APS	subjective predictor	72.51	11.87	43	104	Active Procrastination Scale, measures purposeful delay, i.e., tendency to delay tasks strategically
Number of clicks on assignment	objective predictor	6.58	4.91	1	34	Sum of clicks on an assignment before submission
Interval between start of block and first click on assignment (days)	objective predictor	-7.13	32.40	-150.30	98.72	Difference between start of a block (learning unit on Moodle) and the first click on an assignment (days)
Number of clicks on relevant activities in the course	objective predictor	173.83	168.44	0	1237	Sum of all clicks on activities (quizzes, videos, books, etc.), indicates general activity in the course
Delay (days)	outcome variable	-1.66	18.26	-113.51	132.43	Difference between deadline of an assignment and time of submission, positive values indicating delay and negative values meaning early submissions

actual values, and the latter to determine whether the two classes (delay vs. timely submission) would be correctly predicted.

In Table I, we provide an overview of the characteristics of the outcome variable, the four subjective, and the three objective predictors (i.e., their role in the models, their mean, standard deviation, minimum and maximum values, and a brief description). These characteristics are very useful in adopting different normalization techniques (e.g., standard score or min-max feature scaling). In this paper, we adopt the max-absolute normalization technique to have the block-click-interval and delay features in the interval of [-1, 1].

3.3 Machine Learning Models

Before training the models, we first removed all missing values (e.g., assignments that were not handed in at all) and then normalized the data. The processed data set (1107 rows, each representing one individual assignment) was then split into training and testing sets, the former including 80% of the data (885 rows) and the latter 20% (222 rows). Since the original data set is rather small, the subsets could differ in their distributions depending on the way the data happens to be split. We addressed this concern by repeating the randomized splitting process ten times, thus creating ten pairs of training and testing sets. When individually compared, the testing subsets shared 44.7 rows on average (sd = 5.69, min = 31, max = 57), ruling out potential split-related biases. For further processing, we then split all training and testing sets into three subsets each based on the type of predictor: a subset with subjective predictors, one with objective predictors, and a subset with both types of predictors combined. All sets and subsets were of equal length, the only difference being the included predictors.

For every training set, we then determined the hyperparameters of each ML algorithm based on cross-validation. Once the models' hyperparameters were fixed, we trained each ML algorithm by individually exposing them to all ten training data sets to learn predicting the outcome variable *delay* based on the subjective, objective or both sets of predictors combined. Our objective was twofold: predicting real values in terms of delay and assessing how accurate the classifications were. Therefore, the ML algorithms to be analyzed in this paper all needed to be able to produce both regression and classification models. We thus chose the following eight algorithms: Naive Bayes (NB), K-Nearest neighbors (KNN), Radial Basis Function Networks (RBFN), Feed-Forward Neural Networks (FFNN), Regression Trees (RT), Gradient Boosting Machines (GBM), Random Forests (RF), and Support Vector Regression (SVR). Out of this selection, we aimed to determine the algorithms that *a)* minimized the error between the predicted and actual delay values from the testing data sets; and *b)* obtained the best classification performance when predicting the delay and timely-submission classes. In the following, we introduce each of these techniques, starting with the Bayesian multilevel models that serve as the baseline for our comparisons.

3.3.1 Bayesian Multilevel Models

We selected Bayesian multilevel regression models as the baseline for our comparisons since they were already implemented in Imhof et al. [2021], whose data this study is based on. In that study, Bayesian multilevel models were chosen to match the nested data structure (students being enrolled in multiple courses, each with their own assignments). The authors favored a Bayesian approach over frequentist regression models as they intended to use the results of the study to inform the priors of the models in a planned follow-up study. In total, six models were calculated: one without any predictors to serve as their baseline (which we do not include in this study), one with questionnaire predictors only, two with log data predictors only (one with random intercepts and the other with additional random slopes), and two with all seven predictors (again with random intercepts and additional random slopes respectively). Multilevel models allow the relationship between predictors and outcome variable to vary depending on a grouping factor, which is recommended when there is no valid (e.g., theoretical) reason to assume the relationship remains the same for all values of that factor. For this reason, random slope models were calculated. The grouping factors were the student or the course, depending on the predictors, with the questionnaire scores being associated with the level 2 - grouping factor *student* and the number of clicks on relevant activities being associated with *course*. The remaining predictors were associated with the assignment and thus located on level 1.

3.3.2 Naive Bayes

Naive Bayes (NB) utilizes Bayes theorem and assumes that features are independent. In this paper, we use NB for both classification and regression. When applying NB to regression, we first divide the entire range of the outcome variable into N parts and consider each part as a new class. The algorithm then classifies them based on the features and finally reassigns the target values to the classes. The only hyperparameter of NB model is the number of splits, which is determined through cross-validation.

3.3.3 K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a supervised learning algorithm used for both classification and regression. In order to predict the class of a new query point in KNN, we first find its k-nearest neighbors in the training set and then assign the class of the majority of its neighbors to it. The most common metric for measuring the distances and closeness of the points is the Euclidean distance. We can also have nearer neighbors contributing more by assigning a weight proportional to the inverse of distance to the neighbors. When applying KNN to regression problems, the (weighted) average of the neighbors' target values is calculated and assigned to the test point. The hyperparameter k is determined based on cross-validation.

3.3.4 Radial Basis Function Network

Conceptually, a neural network is a non-linear function in which the weights must be tuned through training data samples to fit the testing data or unseen observations. RBFN is a simple type of a neural network that has the weights structured in only two layers, a hidden and an output layer. RBFN differ from other neural networks in the type of activation function used in the hidden layer. Each hidden neuron makes use of a radial activation function that calculates the Euclidian distance between each data point and some data centers which are a priori computed through clustering algorithms over the training set. For regression tasks, the activation function of the output neuron is a linear one. When training the RBFN, the weights of hidden and output layers are updated in iterations based on the reinforced error between the predicted and real values. In this sense, the gradient descent algorithm is used to train the weights, where the error rate is an important hyperparameter to be optimized together with the number of hidden nodes and Gaussian parameter.

3.3.5 Feed-Forward Neural Network

Feed-forward neural networks (FFNN) differ from RBFN in the sense that the number of hidden layers can be greater than one, the activation function at the level of each hidden node takes simple non-linear representations (e.g., tangent hyperbolic) and at the output layer, the activation function is linear. The weights are trained similarly to RBFN in iterations, but the error is back-propagated each time through a greater number of layers, and the weights are updated by using the same gradient descent algorithm. Alongside the error rate, the number of hidden layers and nodes must be decided through cross-validation before training and testing the FFNN structure.

Compared to RBFN, training an FFNN is less complex since a clustering approach is no longer needed. Generally, the most notable disadvantage when training any type of neural network is the lack of criteria when to abort the learning process to prevent over-fitting. In this paper, we propose the following approach to deal with the issue of setting a proper stopping criterion: Out of the training data, a number of validation samples are carefully selected to monitor the performance of the training process. Instead of imposing a fixed stopping criterion, we conducted the training over a large number of iterations and a multi-objective function (to be detailed in Section III.D). The function aimed at balancing the classification performance and minimizing the regression error, and was computed over the validation data. The weights of the neural networks are saved each time a new maximum value is found during the training process.

3.3.6 Regression Trees

The training process of Regression Trees (RT) is known as a binary recursive partitioning, which splits training data into partitions or branches. The algorithms work iteratively and continue splitting each partition into smaller sub-partitions as the training process moves up to each branch. The process of building the tree until each node reaches a specific minimum node size and becomes a terminal node. Once this node is reached, all the responses from all data points are averaged. When testing the RT, each new point follows the split values and variables given by the RT, which was built based on the train data and the predictions are given as a response of the terminal node each new sample ends on. When the minimum node size is 1, the RT can over-fit the training data. In this paper, we determine the minimum node size for each training set based on cross-validation. A very important aspect that concerns the performance of RT is the splitting rule. Standard split criteria (e.g., linear rank statistics, log rank statistics) cannot detect non-linear effects in the outcome variable. To overcome this potential drawback, we use the maximally selected rank statistics for selecting the split point in which splitting variables are compared on the p-value scale Wright et al. [2017]. Alongside minimum node size, the lower quantile of the co-variate distribution and the significance level for splitting are all determined through cross-validation.

3.3.7 Gradient Boosting Machines

Compared to RT, Gradient Boosting machines (GBM) greedily construct several trees. A new tree is constructed in each round, minimizing the errors given the previously constructed trees. Hence, in each round, the model focuses on the errors made by the previous trees Friedman [2001]. The commonly tuned parameters include the number of trees, maximum depth, minimum samples to split the data, learning rate, and the type of loss function. Moreover, by manipulating the sub-sample parameter, a stochastic version of GBM Friedman [2002] uses bootstrapping averaging similar to RF models, when each iteration is trained only on a fraction of the data.

GBM and its variant XGBoost are used in many existing deployed LA systems Hlosta et al. [2021], Ruipérez-Valiente et al. [2017] and were used in winning solutions for predicting student drop-out in the KDD15 competition 2015 [2015]. Their high performance is due to their ability to learn from previous mistakes without requiring normalization. As a drawback, the model might suffer from over-fitting, which needs to be overcome by cross-validation and exploring a relatively large number of hyperparameters.

3.3.8 Random Forest

In regression tasks, Random Forest (RF) collects the efforts of multiple decision trees and predicts each new data point based on the average responses of all the trees. To decrease the variance of the model without increasing the bias, the trees must be uncorrelated. In this sense, a bootstrapping procedure is applied, and the training subset of each tree is randomly sampled with replacement in the standard proportion of 63.21% from the overall training set Wright and Ziegleri [2017]. Moreover, a random subset of features will be considered at each candidate split when training the regression trees. This overcomes a high degree of correlation between trees when some variables are very strong predictors for the delay variable. As a splitting rule, in this paper we employ the principle of ExtraTrees in which random cut-points are selected in the top-down splitting process Wright and Ziegleri [2017]. For each possible splitting variable, a number of random cut-points are generated and the one with the highest decrease of impurity is selected to split the node. Therefore, the RF algorithm adjusts four hyperparameters based on cross-validation: the number of regression trees, the minimum node size for all trees, the number of possible variable splits, and the number of cut values that are randomly generated in the range of min/max values for each possible split variable.

RF is less prone to over-fitting compared to simple RT or GBM. However, the RF algorithm may change considerably by a small change in the training data. Moreover, the RF training time increases considerably in larger data sets, especially when computing the optimal cut-point locally for each feature (i.e., based on log-rank statistics, maximal rank statistics, etc.).

3.3.9 Support Vector Regression

Being originally proposed for binary classification problems, support vector machines aim to find the best hyperplane that separates a given set of data points. Training points are mapped in space so that the width of the gap or the soft margin between the two categories would be maximized. Test points are mapped in this space and predicted as being part of one of such categories depending on which side of the gap they belong. In practice, the soft margin that separates the two categories of data points can be controlled based on parameter C. When C is relatively small, a larger separation margin between the classes will be used at the price of higher rate of misclassifications. When C is higher, the separation gap gets smoother and the rate of misclassifications is lower. A lower misclassification rate increases the risk of over-fitting the training data that can lead to a much higher rate of misclassification for new data points. An optimal value of soft-margin hyperparameters must be found on training data to fit the testing data.

The accuracy of classification depends on the shape of the separation hyperplane that can be set through the kernel function. In this paper, we test different types of kernel functions (linear, polynomial, radial basis function, radial basis function with vector subtraction, hyperbolic tangent), that can be used to best fit the collected data. Moreover, some of these kernel functions must be parameterized before properly training and testing the machine. In regression tasks, the goal of Support Vector Regression (SVR) is comparable to classification tasks with the amendment that a certain error is tolerated. In this paper, we optimize the soft margin parameter, the kernel function and error tolerance through cross-validation.

3.4 Proposed Cross-Validation Measure

In order to optimize the hyperparameters of the implemented ML algorithms (NB, KNN, RBFN, FFNN, RT, RF, GBM, SVR), we employed a cross-validation procedure based on the training sets. Each ML model defines its own grid of hyperparameter configuration that falls in predefined ranges. Each possible configuration of hyperparameters from the grid is then evaluated based on the training data. Finally, the trained ML model with the highest evaluation outcome is employed to predict new data examples from the testing set.

The algorithm splits the training set in K number of sub-partitions. Then, the K-fold cross-validation trains the ML model iteratively for each hyperparameter configuration band on $K-1$ sub-partitions and the results are validated based on the remaining sub-partition every time. For each configuration, the validation results are averaged over K number of partitions. Once all hyperparameter configurations are evaluated, the K-fold cross-validation algorithm selects the scheme with the highest validation outcome. The same procedure is repeated for each ML algorithm with its predefined grid search of hyperparameter configurations. The validation outcome can be measured by monitoring the regression error (e.g., mean absolute error). However, in our prediction task, the selected hyperparameters should also improve the classification performance.

Moreover, the ratio of early and late submissions in the training data is often unbalanced. Therefore, we propose a novel multi-objective function that measures the validation performance of each hyperparameter configuration on each split of data set. Based on this proposed function, the cross-validation selects the hyperparameter configuration that minimizes the regression error and maximizes the classification performance between the two classes (delay and timely submission).

Given the function f that must be trained by each ML model to best predict delay, we denote by $\hat{y} = f(\mathbf{x})$ the current estimate of prediction function f, where $\mathbf{x}$ is a test data point. In order to evaluate the prediction performance, the ground truth values of delay y should be known in advance. In this study, delay y is measured as the difference between the submission of an assignment and its deadline in days. The deadline is met when $y \leq 0$ (timely submission), and the tasks are submitted after the deadline when $y > 0$ (delay). When a new data sample from the test set is predicted, the decision can take one of the following forms: *a)* true positive (TP): $\hat{y} > 0$ and $y > 0$; *b)* false positive (FP): $\hat{y} > 0$ and $y \leq 0$; *c)* true negative (TN): $\hat{y} \leq 0$ and $y \leq 0$; *d)* false negative (FN): $\hat{y} \leq 0$ and $y > 0$. Once all test samples are predicted, the classification performance is evaluated based on TP, FP, TN and FN indicators which are summed over the number of test data examples. To measure the ML model performance when dilatory behavior is correctly detected, the F1-score function can be employed as follows:

$$F_{tp} = \frac{2TP}{2TP + FP + FN} \quad (1)$$

where a higher F_{tp} value implies a better performance when predicting delay. In unbalanced data sets, F_{tp} does not give any measure of how well the prediction performance is for the timely submission class. In this case, a similar version of F1-score measure can be computed to monitor the ratio of TN over the number of false predictions:

$$F_{tn} = \frac{2TN}{2TN + FP + FN} \quad (2)$$

By maximizing both F_{tp} and F_{tn} in cross-validation, the hyperparameters will be selected according to the best prediction balance between the two classes. To measure how close the predicted value $\hat{y}$ is from its delay pattern $\hat{y}$, we measure the Mean Absolute Error (MAE) E given by:

$$E = \frac{1}{T}\sum_{i=1}^{T}\frac{|\hat{y}_i - y_i|}{y_{max}} \quad (3)$$

where T is the number of samples in each data set split and y_{max} is the maximum absolute value of delay. Considering (1), (2) and (3), the multi-objective function to be maximized during the cross-validation process becomes:

$$G = \frac{(1-E) + F_{tp} + F_{tn}}{3} \quad (4)$$

The configuration of hyperparameters able to maximize (4) will be selected to train the ML model for the entire training set and the prediction performance is analyzed in Section IV. Each ML algorithm follows the same cross-validation principle when tuning the corresponding hyperparameters.

3.5 Analyses

The baseline models (i.e., the Bayesian multilevel models) were calculated with the R package *brms* Bürkner [2017] and the NB, KNN and GBM algorithms were employed using the Python *scikit-learn* package Pedregosa et al. [2011]. The remaining ML approaches (FFNN, RBFN, RT, RF, SVR) were deployed in C/C++ using dedicated functions for cross-validation, training and testing. For the `obj` and `comb` predictors, the Bayesian models included random intercepts (BA-RI) and random slopes (BA-RS) due to the nested structure of the data, which was not the case for the `subj` predictors. The implementation of RT and RF was based on RANGER C++ packages which are publicly available Wright and Ziegleri [2017]. The proposed SVR ML algorithm followed the regression model from Parrella [2007] with five different types of kernel functions: linear (SVR-LIN), polynomial (SVR-POL), tangent hyperbolic (SVR-TAH), radial-basis function (SVR-RBF) and with the vector subtraction function (SVR-VS). All these variants of SVR model are cross-validated, trained, tested and compared to the other ML algorithms. The FFNN model we used was previously employed for the optimization of video quality in remote education Comsa et al. [2021], while the RBFN model was also used to classify high dimensional vectors in radio communications systems Comsa [2014].

The complexity of cross-validation processes depends on the grid size, which differs from one ML algorithm to another. A special case in cross-validation is represented by RBFN, where the number of hidden nodes is equivalent to the number of data centers computed for each data set separately based on the clustering analysis. The clustering process is conducted before cross-validation by employing a heuristic algorithm that iteratively combines a classical k-means algorithm for a more precise calculation of centers with an algorithm that uses the random swapping of data centers from the available data set to enhance the searching time of globally optimal solutions Comsa [2014]. Based on the clustering algorithm, we determined the number of optimal clusters that characterizes each training set by employing an additional algorithm to calculate the Silhouettes Index (SI) Rousseeuw [1987]. The SI index interprets and validates the consistency of the clusters for each training set, and provides a measure of how well each data point is matched to its own cluster compared to the neighboring clusters. Higher SI values (max value of 1) denote that the data points are very well suited to their clusters, while lower values (min value of -1) indicate that the clustering configuration may have too many or too few clusters. In general, when a higher number of clusters is obtained through the computation of SI index, then the data set is much better represented when training the RBFN model and a higher prediction performance is expected.

We evaluated the performance of the ML algorithms for each type of predictor based on classification metrics and regression errors (G-Scores for the former and MAE for the latter). As classification measures, we considered the F_{tp} and F_{tn} scores in the computation of the multi-objective G function. When training data is unbalanced, we recommend the use of both scores to find the best configuration of hyperparameters that can balance precision and robustness in both directions. This is the case for our original data set, which is unbalanced in favor of the timely-submission class, meaning that 67% of assignment data points were classified as timely submissions ($y < 0$), and the remaining 33% as

Table 2: Comparison of G-scores between different predictor types and ML approaches obtained in cross-validation and prediction

	Cross-validation Mean G-Score (SD)			Test Mean G-Score (SD)		
ML Alg.	`subj`	`obj`	`comb`	`subj`	`obj`	`comb`
NB	0.6457 (0.0084)	0.6845 (0.013)	0.6961 (0.0048)	0.6375 (0.021)	0.6628 (0.0267)	0.681 (0.019)
KNN	0.7071 (0.0068)	0.7482 (0.0039)	0.7525 (0.0069)	0.6758 (0.0222)	0.7279 (0.0132)	0.7319 (0.019)
RBFN	0.7401 (0.0051)	0.7707 (0.0072)	0.7688 (0.0054)	0.6778 (0.0179)	0.701 (0.0218)	0.7214 (0.0204)
FFNN	0.7143 (0.012)	0.7631 (0.0102)	0.7673 (0.008)	0.6695 (0.0368)	0.6957 (0.0316)	0.71884 (0.0223)
RT	0.6772 (0.0106)	0.7491 (0.0046)	0.7478 (0.0047)	0.6811 (0.0251)	0.7126 (0.025)	0.7181 (0.035)
RF	0.7081 (0.008)	0.755 (0.0057)	0.7685 (0.0054)	0.7064 (0.0153)	0.7457 (0.018)	0.7626 (0.015)
GBM	0.70686 (0.01143)	0.76012 (0.00919)	0.7713 (0.0071)	0.71624 (0.025)	0.73166 (0.021)	0.7442 (0.0271)
SVR-LIN	0.6785 (0.0071)	0.7253 (0.0083)	0.7275 (0.0069)	0.6591 (0.0422)	0.6888 (0.0264)	0.7134 (0.019)
SVR-POL	0.6975 (0.0056)	0.7348 (0.0041)	0.7413 (0.0055)	0.6835 (0.0312)	0.6934 (0.0273)	0.715 (0.023)
SVR-TAH	0.6833 (0.01)	0.7305 (0.0072)	0.7331 (0.0062)	0.6544 (0.0318)	0.6916 (0.0334)	0.708 (0.024)
SVR-RBF	0.7208 (0.0094)	0.7406 (0.0059)	0.7536 (0.0077)	0.7129 (0.0189)	0.7288 (0.026)	0.7538 (0.018)
SVR-VS	0.7208 (0.0084)	0.7608 (0.0045)	0.7631 (0.0048)	0.7145 (0.0175)	0.7364 (0.0257)	0.7589 (0.018)

delay ($y > 0$). Therefore, we included the following metrics: Positive Predicted Value ($PPV = TP/(TP + FP)$), which is the precision of detecting delay; and the True Positive Rate ($TPR = TP/(TP + FN)$) that indicates how well delay can be predicted without negatively affecting the prediction of timely submissions (sensitivity or recall). However, when data is perfectly balanced, the use of F_{tp} and F_{tn} is not needed. In this case, other classification metrics could be used instead, including accuracy (ACC) as the ratio of the correct and total predictions, and the Matthews Correlation Coefficient (MCC) as a metric of difference between correct and wrong predictions. Generally, the MCC is reportedly more informative compared to other measures Chicco and Jurman [2020], Chicco [2017]. We included ACC and the MCC for the sake of completion.

4 Results

The aim of this section is to evaluate the performance of ML algorithms in regard to cross-validation and prediction. First, we measured and evaluated the multi-objective function G-Score for each algorithm and type of predictor in the cross-validation process. This was followed by a calculation of the same G-Scores for the test data sets. These results were then compared to those of the cross-validation (RQ1). RQ2 is then addressed by comparing the ML approaches for each type of predictor in more detail while considering the classification performance indicators and the mean absolute error. Finally, we address RQ3 by evaluating how high the predictive performance is when considering the `obj` and `comb` predictors compared to predictions based solely on `subj` variables.

As explained above, the ML algorithms were cross-validated, trained and tested separately for ten randomized subsets to account for the discovered unbalance. The reported results include the mean and SD values over the ten subsets for each of the three types of predictors.

4.1 Evaluation in Cross-Validation

In this paper, we employed a 4-fold cross-validation scheme for all of the ML algorithms, in which all training sets were divided in four subsets with 222 elements each. The same cross-validation procedure was conducted for the three types of predictors and each ML algorithm, resulting in a total of 360 different cross-validation processes.

4.1.1 Mean G-Score in Cross-Validation

To evaluate the performance of the ML algorithms during cross-validation, we computed the mean G-score and SD over ten data sets for each of the three predictor types. The values exposed in Table II for each ML algorithm are the maximum G-scores of the best configuration of hyperparameters obtained through the cross-validation averaged over ten training data sets.

We do not report hyperparameters for the Bayesian multilevel models since we selected non-informative priors, as was the case in the previous study Imhof et al. [2021]. Looking at the `subj` and `obj` predictors, RBFN obtained the highest scores of 0.74 and 0.771, respectively. The performance regarding the validation sets was higher compared to the other ML algorithms since the data centers obtained through the proposed clustering approach cover the entire training set. For the `comb` predictors, RBFN, RF and GBM obtained comparable G-scores. When using SVR, the LIN and TAH

Table 3: Intra-model comparison between ML approaches for `subj` predictors

ML Alg.	PPV (SD) [%]	TPR (SD) [%]	F_{tp} (SD) [%]	F_{tn} (SD) [%]	MCC (SD) [%]	ACC (SD) [%]	MAE (SD)
BA	36.19 (3.61)	32.59 (8.16)	33.66 (5.59)	71.33 (2.69)	5.91 (4.22)	60.14 (2.39)	9.19 (0.99)
NB	41.1 (4.4)	41.6 (8.5)	40.9 (5.1)	73.3 (2.9)	14.8 (5.4)	63.4 (2.6)	30.38 (2.47)
KNN	39.44 (5.32)	44.84 (8.62)	41.68 (5.84)	70.09 (2.78)	12.4 (7.81)	60.55 (3.27)	11.94 (2.57)
RBFN	40.18 (3.28)	42.66 (7.65)	40.9 (3.37)	71.06 (3.039)	12.73 (3.02)	61.3 (2.4)	11.39 (1.35)
FFNN	40.46 (5.45)	37.1 (14.06)	37.04 (5.33)	71.69 (6.49)	11.21 (5.88)	61.62 (5.09)	10.44 (1.03)
RT	39.35 (4.03)	46.01 (8.97)	42.19 (5.61)	69.76 (2.81)	12.68 (6.57)	60.45 (6.48)	10.09 (1.15)
RF	45.82 (4.15)	45.86 (5.83)	45.43 (2.68)	74.28 (2.74)	20.22 (3.09)	65.18 (2.42)	10.31 (1.1)
GBM	47.69 (4.75)	43.94 (7.18)	45.56 (5.4)	76.2 (1.77)	22.07 (7.04)	66.89 (2.67)	9.11 (0.95)
SVR-LIN	42.46 (7.88)	25.96 (12.69)	29.50 (9.81)	75.42 (3.52)	9.66 (4.35)	64.05 (2.92)	9.54 (0.89)
SVR-POL	42.02 (2.79)	36.47 (10.14)	38.31 (5.59)	73.80 (4.29)	13.30 (3.72)	63.65 (3.19)	9.34 (0.96)
SVR-TAH	39.84 (7.95)	22.39 (8.85)	27.52 (7.66)	75.89 (2.62)	7.28 (6.33)	64.05 (2.78)	9.37 (0.95)
SVR-RBF	50.76 (4.45)	36.56 (5.99)	42.24 (4.88)	78.30 (1.62)	22.14 (5.11)	68.51 (2.02)	8.82 (1.09)
SVR-VS	51.22 (4.29)	36.57 (5.46)	42.51 (4.7)	78.54 (1.31)	22.63 (5.05)	68.78 (1.83)	8.84 (1.03)

Table 4: Intra-model comparison between ML approaches for `obj` predictors

ML Alg.	PPV (SD) [%]	TPR (SD) [%]	F_{tp} (SD) [%]	F_{tn} (SD) [%]	MCC (SD) [%]	ACC (SD) [%]	MAE (SD)
BA-RI	47.97 (5.11)	57.53 (4.78)	52.23 (4.51)	74.17 (3.02)	27.13 (6.51)	66.53 (3.32)	8.57 (0.68)
BA-RS	51.02 (3.52)	60.88 (6.75)	55.41 (4.37)	76.14 (2.32)	32.29 (5.4)	69.01 (2.38)	7.57 (0.63)
NB	40.6 (6.1)	29.2 (7.8)	33.7 (7.2)	76.3 (1.7)	11.5 (6.5)	65.2 (2.0)	14.75 (1.19)
KNN	49.1 (1.77)	48.6 (5.24)	48.69 (2.94)	76.26 (1.68)	25.15 (3.27)	67.62 (1.69)	8.72 (0.88)
RBFN	43.49 (4.34)	52.59 (9.99)	46.79 (3.16)	70.72 (4.36)	19.45 (3.08)	62.52 (3.25)	9.64 (1.26)
FFNN	43.55 (3.8)	45.29 (11.01)	43.43 (5.74)	72.54 (4.38)	17.22 (4.44)	63.47 (3.46)	9.57 (0.87)
RT	48.39 (4.25)	40.76 (8.06)	43.8 (5.73)	76.71 (2.62)	21.39 (5.29)	67.25 (2.62)	8.92 (1.09)
RF	50.36 (4.31)	57.12 (5.09)	53.48 (4.4)	76.16 (1.68)	29.98 (5.7)	68.51 (2.25)	7.88 (0.79)
GBM	50.7 (5.04)	45.13 (7.04)	47.51 (5.46)	77.53 (1.74)	25.52 (6.12)	68.61 (2.24)	7.34 (0.86)
SVR-LIN	43.78 (5.21)	35.72 (9.61)	38.48 (6.38)	75.08 (2.27)	14.99 (5.99)	64.69 (2.47)	9.13 (0.93)
SVR-POL	43.13 (6.54)	38.39 (8.43)	40.23 (6.24)	74.32 (2.62)	15.21 (7.79)	64.19 (3.15)	8.66 (0.89)
SVR-TAH	47.18 (5.35)	35.44 (13.82)	38.63 (8.16)	76.26 (3.55)	17.77 (4.93)	66.26 (2.78)	9.81 (2.22)
SVR-RBF	46.47 (5.54)	41.09 (8.14)	43.23 (5.96)	75.76 (2.46)	19.64 (6.31)	66.17 (2.58)	8.39 (0.85)
SVR-VS	49.73 (4.57)	51.81 (9.62)	50.52 (6.82)	76.50 (1.59)	27.35 (7.06)	68.29 (2.07)	8.09 (0.91)

kernel functions are not suitable options to predict delay when employing cross-validation over the training sets. The non-linear kernels (POL, RBF, VS) can fit the validation sets better in the training $K-1$ sub-partitions. In the next step, we verified if the same prediction trend in cross-validation is followed when predicting new examples from the test sets.

4.1.2 Mean G-Score in Test Data Sets

The hyperparameters obtained through cross-validation were used to train each ML algorithm by using the entire training sets (K folds) for each type of predictor. Their performance is then evaluated in the test sets for all ML algorithms based on G-Scores. In Table II (right side), we present the mean G-scores and SD values calculated over the ten test data sets for each type of predictor. In case of the `subj` predictors, the GBM and SVR-VS models outperformed the others with mean G-scores higher than 0.71. When analyzing the `obj` predictors, the best performance of $G = 0.75$ was achieved by the baseline Bayesian models with random slopes. When combining both types of predictors, RF and SVR-VS remained the best options with a mean G-score higher than 0.76.

4.1.3 Mean G-Score Comparison

In order to validate the configuration of hyperparameters selected for each ML algorithm, the G-scores were compared between cross-validation and testing. The most successful ML approach would have the highest G-score among all candidates and all types of predictors with the smallest performance deprecation between cross-validation and testing. NB and KNN both had a rather small difference in G-scores in the validation and testing sets, but the performance level is much lower compared to the other ML approaches. RT and GBM are well known for over-fitting the training data, which can also be observed when looking at their G-scores. SVR is sensitive to the training data, and the performance of delay prediction depends very much on the selected configuration of hyperparameters. This explains the 3% deprecation between validation and test sets for all types of predictors, especially for the case of kernels that use the linear and tangent hyperbolic functions. When analyzing the performance of RBFN and FFNN in the validation and test data sets, we observed the highest degradation in performance. This aspect is explained by the fact that both RBFN and FFNN were trained on 60% of the data, while the remaining 20% were used for the stopping criteria. Also, the G-scores

Table 5: Intra-model comparison between ML approaches for `comb` predictors

ML Alg.	PPV (SD) [%]	TPR (SD) [%]	F_{tp} (SD) [%]	F_{tn} (SD) [%]	MCC (SD) [%]	ACC (SD) [%]	MAE (SD)
BA-RI	49.72 (5.17)	59.94 (3.82)	54.25 (4.14)	75.14 (2.82)	30.20 (5.85)	67.84 (3.14)	8.56 (0.68)
BA-RS	51.04 (2.65)	61.08 (5.43)	55.55 (3.49)	76.19 (1.6)	32.43 (4.09)	69.05 (1.62)	7.61 (0.66)
NB	42 (5.3)	40.9 (5.5)	41.2 (4.1)	74.4 (3.)	15.9 (6.1)	64.5 (3.3)	15.12 (1.74)
KNN	49.34 (5.22)	51.8 (4.72)	50.41 (4.19)	67.67 (2.19)	26.6 (5.92)	67.67 (2.64)	9.01 (0.89)
RBFN	46.28 (5.02)	59.69 (6.79)	51.69 (3.37)	71.95 (3.69)	25.58 (4.68)	64.64 (3.35)	9.56 (1.22)
FFNN	47.72 (4.45)	47.28 (7.54)	47.21 (5.34)	75.59 (2.16)	23.15 (5.36)	66.76 (2.17)	9.46 (1.07)
RT	48.87 (7.61)	49.42 (10.77)	47.92 (5.09)	74.41 (6.28)	24.02 (6.63)	66.26 (5.31)	9.13 (1.24)
RF	52.75 (3.99)	62.22 (3.8)	57.05 (3.58)	77.23 (1.61)	34.86 (4.6)	70.27 (1.98)	7.27 (0.82)
GBM	51.60 (5.47)	51.35 (8.96)	51.15 (6.58)	77.45 (2.19)	28.98 (7.71)	69.23 (2.94)	7.05 (0.78)
SVR-LIN	47.28 (5.28)	43.79 (4.78)	45.24 (3.85)	75.66 (2.43)	21.26 (5.71)	66.35 (2.84)	9.1 (0.73)
SVR-POL	47.03 (3.59)	44.86 (8.40)	45.55 (5.39)	75.59 (2.22)	21.63 (5.94)	66.39 (2.45)	8.82 (0.91)
SVR-TAH	47.84 (3.28)	38.94 (10.18)	42.27 (6.43)	76.87 (1.48)	20.46 (5.28)	67.12 (1.51)	8.93 (0.87)
SVR-RBF	53.83 (5.27)	54.67 (7.06)	53.92 (4.31)	78.23 (1.88)	32.57 (5.44)	70.5 (2.2)	7.97 (0.89)
SVR-VS	53.12 (4.73)	58.82 (4.54)	55.72 (3.99)	77.66 (2.12)	33.69 (5.86)	70.32 (2.66)	7.57 (0.99)

reported in cross-validation are the best values that could be found while training in each split. RF was the most stable ML algorithm for all three types of predictors, since the levels of test G-scores were very close to those scores obtained in cross-validation.

4.2 Intra-Model Comparisons

In this section, we provide a comprehensive report to compare the employed ML algorithms for each type of predictor (intra-model comparison). Thus, we analyse the mean and SD of the PPV, TPR, F_{tp}, F_{tn}, MCC, ACC, and MAE performance indicators, where the best values of these indicators are highlighted in green. To answer RQ2, the goal of this section is to determine the most successful ML algorithm for each type of predictor that would maximize the F_{tp} measure and minimize the loss in F_{tn} and regression error (MAE).

4.2.1 Subjective Predictors

In Table III we present the classification and regression performance metrics (mean and SD) for each ML algorithm and calculated over the data sets with `subj` predictors. When looking at the PPV metric, we obtained the highest amount of correct predictions for the delay class when using SVR with RBF and VS kernels. However, the robustness of these models in predicting the timely submission class is deprecated more than 15% when compared to PPV. The best trade-off between PPV and TPR was observed with the RF and GBM models. When computing the F_{tp} score, GBM performed slightly better than RF with 45.5%. However, the SD of the RF model is reduced by half when compared to GBM, meaning the RF classifier is less sensitive to the type of data set which is used to train the model. The F_{tn} score was also calculated in Table III to measure the trade-off between the precision for timely submission and the robustness to delay. Since the false predictions of delay have a greater impact than the false predictions of timely submissions due to the unbalance of the data sets, the SVR model with the RBF and VS kernels achieved the highest F_{tn} scores. By monitoring the MCC and ACC classification metrics, it can be concluded that the absolute difference between the false predictions of delay and timely submission was higher for the SVR model (with RBF and VS kernel functions) than for the RF and GBM models. Based on the results collected in Table III, a precision to detect delay of nearly 48% and a classification accuracy of about 67% can be obtained with the GBM model when exclusively considering the `subj` predictors. When computing the regression error, SVR model with RBF and VS kernels performed slightly better than GBM. However, we recommend the use of the GBM model when predicting the `subj` predictors due to the best trade-off between regression error, F_{tp} and F_{tn}.

4.2.2 Objective Predictors

In Table IV, we present the classification and regression performance metrics (mean and SD) obtained when predicting delay with `obj` predictors. When evaluating the predictive performance of the delay class, the Bayesian model with random slopes (BA-RS) provided the best results among all models with the following metrics: a) the precision (PPV = 51.02%) when predicting delay; b) robustness (TPR = 60.88%) to predict timely submission; c) trade-off between PPV and TPR ($F_{tp} = 55.41\%$). Also, BA-RS was the best option when measuring the MCC (32.29%) and accuracy (ACC = 69.01%). On average, the GBM model achieved the highest F_{tn} score of 77.53% due to higher robustness to predict delay. Also, its error was lower compared to the other models. When looking at the trade-off between F_{tp}, F_{tn} and mean absolute error, BA-RS was the best option to predict delay with `obj` predictors.

Table 6: Inter-model comparison between the highest-performing ML approach for each type of predictor: GBM (subj), BA-RS (obj) and RF (comb)

Pred type	PPV (SD) [%]	TPR (SD) [%]	F_{tp} (SD) [%]	F_{tn} (SD) [%]	MCC (SD) [%]	ACC (SD) [%]	MAE (SD)
subj	47.69 (4.75)	43.94 (7.18)	45.56 (5.4)	76.2 (1.77)	22.07 (7.04)	66.89 (2.67)	9.11 (0.95)
obj	51.02 (3.52)	60.88 (6.75)	55.41 (4.37)	76.14 (2.32)	32.29 (5.4)	69.01 (2.38)	7.57 (0.63)
comb	52.75 (3.99)	62.22 (3.8)	57.05 (3.58)	77.23 (1.61)	34.86 (4.6)	70.27 (1.98)	7.27 (0.82)

4.2.3 Subjective and Objective Predictors combined

When looking at the performance of the ML algorithms for the `comb` predictors in Table V, we observed that the SVR-RBF performed better in precision (with about 2%) but with lower robustness value (with more than 7%) when compared to the RF model. This explains the larger gain obtained by the RF algorithm when measuring the F_{tp} score of about 57%. The same trend can be observed when comparing the SVR-VS and RF models. When measuring the F_{tn} scores, SVR with RBF kernel provided the best results due to better robustness of the model to predict delay. By comparing the RF, SVR-VS, and SVR-RBF models, we noted that higher F_{tp} scores involve higher MCC levels, while higher F_{tn} scores align with higher accuracy values. When calculating the average regression error, GBM and RF models outperformed the other ML models. Although SVR-RBF and SVR-VS provided better precision (PPV), the RF model remained the best option to be employed for the `comb` predictors due to the best trade-off between precision and robustness to delay and timely submission classes, and a low regression error.

4.3 Inter-Model Comparison

In this section, we compare the predictive performance between each type of predictor to answer RQ3 by taking into account the results displayed in Tables II-V.

When analyzing the performance of the mean G-scores in both the cross-validation and testing stages (Table II), we concluded that the `obj` predictors are more informative than the `subj` predictors when predicting delay in assignments. However, a performance gain can be achieved by combining both types of predictors. When analyzing the best G-scores in Table II (on the right side), we observe a prediction gain of 4% when comparing the prediction with the `obj` and `subj` predictors, and a gain of 1% when predicting with the `comb` predictors compared to the `obj` predictors alone. By combining both `subj` and `obj` predictors, almost all ML candidates benefit from the perspective of both classification and regression metrics as shown in Table V. For example, the SVR model with all types of kernel functions obtained higher precision (PPV) and robustness (TPR) when involving the `comb` predictors compared to the `subj` or `obj` predictors alone. The same performance gain can be observed when measuring MCC, ACC, or MAE. Another concluding example in this sense is the RF approach that enhanced its F_{tp} score from 45.43% for `subj` predictors to 53.48% for `obj` data, and when combining both, the performance was higher than 57%. An explanation of this gain is given by the importance of the `subj` predictors that changes when the `obj` predictors are added. In this sense, we computed the Gini importance index or the mean decrease in impurity that gives the feature importance when regression trees are employed T. Daniya and Kumar [2020]. When training the RF model, we make the following observations based on the importance index: *a)* in case of `obj` predictors, the interval between the start of a block and the first click on an assignment is the most important predictor, followed by the number of clicks on an assignment and the number of clicks on relevant activities; *b)* when training with `subj` predictors, APS is the most informative, followed by APSS, SDLS, and GASE; *c)* by combining both types of predictors, the most important variables are the `obj` predictors in the same order as above in *a)*, followed by the `subj` predictors in the order as before (*b*), with the only difference that the GASE variable gained a higher importance than APS. While APS and APSS are more predictive in general, when combined with `obj` predictors, they do not add any new information. Therefore, the GASE variable could enhance the precision and accuracy above 52% and 70% respectively when combining both sets of predictors.

As observed in Tables III-IV, the highest performing ML approach differs from one type of predictor to another when balancing the performance between F_{tp} and F_{tn} scores and the regression error. GBM is recommended to be used for `subj` data due to a better trade-off between the indicators above. When using the `obj` predictors only in the form exposed in Table I, the BA-RS approach was the best option to be employed. However, for combined predictors, the RF method outperformed other approaches due to a higher importance of GASE feature among other `subj` ones. In Table VI, we summarize the results from Tables III-IV and present the inter-model comparison between the highest performing ML approach for each type of predictor. The combined predictors brought a performance gain of more than 1% when compared to `obj` data sets and more than 11% when compared to `subj` predictors. The regression error was the lowest when predicting with the `comb` data sets while the deprecation in F_{tn} was negligible between different types of predictors. The highest accuracy value of 70.23% was achieved when combining the predictors, with a slight depreciation of about 1% for `obj` and 3% for `subj` predictors. The same trend was observed when monitoring the MCC indicator with a much larger performance gap of 10%.

5 Discussion

The aim of this study was to expand the findings in Imhof et al. [2021] by employing multiple ML algorithms to determine which one delivered the best predictions of delay based on objective and subjective variables. First, due to the unbalanced nature of the data (meaning there were twice as many timely submissions as there was delay), we needed to account for it in our approach to optimize each of the ML algorithms. We achieved this by a cross-validation procedure we conducted based on a novel multi-objective function, the G-score, which involves a trade-off between positive and negative $F1$-scores and regression error. We measured the performance degradation of G-score between test and cross-validation to verify the authenticity of the selected hyper-parameters for each ML algorithm. Except for RBFN and FFNN, all other approaches provided a low degradation in G-score. We concluded that RBFN and FFNN need more data to cope with such a performance gap. Among the ML algorithms in our pool, RF obtained the lowest degradation in performance, meaning that cross-validation was highly efficient (RQ1).

We then identified the best prediction performances for each of the types of predictors (`subj`, `obj`, `comb`) by measuring the trade-off between F_{tp}, F_{tn}, and MAE (RQ2). GBM turned out to be the best approach for subjective predictors, BA-RS for the objective predictors, and RF for the combined sets.

The highest predictive performance of RF is congruent with the existing literature in Learning Analytics, where RF consistently ranks among the best models Hlosta et al. [2018], Alcaraz et al. [2021]. To the best of our knowledge, our second best approach, however, Bayesian multilevel models with random slopes, is not mentioned in any LA studies, due to its nature as a statistical approach. Since most of the studies in LA are focused on predictions from objective data and BA-RS was the best algorithm in our case, it might be useful for other researchers to consider multilevel models in their work. After all, repeated measures are a common occurrence when assessing student data and multilevel models provide the means to reflect nested data structures. Nevertheless, these comparisons should be taken cautiously since the outcomes differ in both studies.

Next, we compared the different types of predictors (RQ3). Predicting delay with objective predictors yielded better results than relying on the subjective variables alone. By mixing both types of predictors, we achieved a slightly higher predictive performance, meaning that some of the subjective predictors increase their relative predictive power when combining them with the objective predictors. These results are unsurprisingly in line with the previous study our data stems from Imhof et al. [2021]. The superiority of the objective predictors was evident in all the ML models regardless of the underlying algorithm, strengthening the finding of the previous study. Similarly, Pardo et al. [2017] reported that more objective variables correlated with academic outcomes than subjective variables, and Yu et al. [2020] also found a higher accuracy of objective features to predict course and next semester outcomes. However, in contrast with the previous study, the advantage of combining the two sets of predictors was more apparent here since the exact ranking of the models was rather unclear before.

As expected, the best results across all the performance metrics were achieved when the subjective and objective variables were combined. This is also in line with Pardo et al. [2017], Tempelaar et al. [2021], despite the different levels of granularity between our study and theirs. While both of these studies focused on the final course outcome, we were interested in a much more fine-grained measure (the delay of individual assignments). In both cases, the authors reported an increase in explained variance when objective variables were added to the subjective variables. Our results demonstrate that this effect remains consistent across different ML algorithms when predicting delay. However, it remains unclear if combined data has a long-term predictive advantage. In Yu et al. [2020] for instance, the authors found that combined data only had higher predictive accuracy for short-term predictions. For long-term predictions (e.g., the outcome of the following semester), the authors determined that single objective predictors yielded better results. We suspect this may be due to fluctuations in students' learning-related dispositions, which cannot be fully captured by a single measurement. Despite their trait-like nature, it would thus be advisable to repeatedly assess subjective factors instead of exclusively relying on once-collected historical student data. This comes with its own risks, e.g., alienating students by a high frequency of reassessments.

Compared to Pardo et al. [2017], our results show that using ML algorithms might shuffle the order of importance of the subjective variables and the magnitude of their contribution. While APSS is the most predictive feature for subjective data, when combined with objective predictors, it is superseded by the GASE factor, which was the least important factor for subjective-only predictions. This makes it more challenging to determine which variables to favor in a more parsimonious model. In the work of Pardo et al. [2017], the order of importance remained consistent. We assume that this discrepancy could either be traced back to the covariance structure of the predictors and/or the complex nature of some of the implemented algorithms, particularly the underlying decision trees that empower both RF and GBM. In decision trees, more complex dependencies of the variables are considered when building the model.

5.1 Limitations

This study has two major limitations that need to be addressed. First off, the comparatively low predictive accuracy of the models (some even performing barely above chance) raises the question of why it is not as high as one might expect based on models in other studies. One issue is our rather diminutive sample size. This could be remedied in a number of ways, for example by collecting more data across multiple semesters with the same students. The data accumulated over time can then serve as historical data, serving as training data for future predictions. The other contributing factor is the unbalance of our data. Two thirds of the assignments being handed in on time is not necessarily a representative finding. Given the reported prevalence of procrastination in higher education Steel [2007], one could assume that there is more of an equilibrium in other populations. Another plausible possibility is a stricter enforcement of deadlines, resulting in an even higher proportion of timely submissions and thus more unbalanced data. The ML approaches involved in this type of prediction thus need to be stable enough to account for such differences in the studied populations.

The second issue concerns the models themselves. Models with objective predictors outperforming models with subjective predictors is not that surprising, considering that the outcome variable delay and our objective predictors are all state variables. When interpreting the results, it is also important to note that the two sets of variables are not directly comparable, mainly due to their different granularity (the subjective predictors being based on data assessed on the student level and the objective variables being located on the assignment level in two cases and the course level in one case). This is also reflected in the outcome variable, which shared its granularity only with a few predictors, which were also revealed to be the most important variables, both in this study and the original one Imhof et al. [2021].

5.2 Implications & future work

The results imply that future research in procrastination and other types of dilatory behavior should include ML algorithms such as RF, rather than relying on traditional statistical approaches alone. This holds true even for cases with low numbers of variables. In order to select the best variables to enhance predictions, we thus recommend not to refrain from implementing ML algorithms and evaluating data sets that include both subjective and objective predictors. This is particularly crucial when applying such models in the field, e.g., for for real-time predictions, considering these two types of predictors may not be available simultaneously. While our data was all collected at the end of the semester, the same set of predictors could be split into early and late predictors if they were to be assessed in an LMS with real-time predictions in the future.

For instance, our subjective predictors are trait variables, meaning they are supposedly stable across certain periods, and could thus already be assessed at the start of the semester, making them viable as early predictors. Despite not being as strong or consistent as their objective counterparts in our models, the subjective predictors could still potentially have value since they may provide some clues long before the objective predictors become active. Even though a lot of the algorithms performed below chance when operating with subjective variables alone (meaning the performance was lower than simply assuming all assignments would be submitted on time), SVR-VS still managed to perform above chance.

As discussed in Imhof et al. [2021], the objective predictors we used have the drawback of requiring information about students' activity across the semester, meaning predictions based on them cannot be made unless that data is accumulated (at least partially). This again highlights the utility of combining both types of predictors. Moreover, some of these variables change throughout the course (e.g., the number of clicks on an assignment), and so would the predictive effect of that variable when doing real-time predictions. From this perspective, the optimal points in time should be determined during the semester to achieve the best possible predictive performance that allows for instructional feedback and other interventions to have a positive impact on dilatory behavior and other performance metrics.

These results also imply that selecting a single algorithm and sticking with it for all models may not be the ideal path to take. Considering that a different approach was favored for each of the three categories, it may be advisable to employ multiple ML algorithms for real-time predictions, depending on the current stage of the semester (e.g., starting with SVR to work with early predictors, which can then be accompanied by a multilevel model for the late predictors or get replaced entirely by RF once all predictors become available).

Combined, our results show promise that these types of models could be used for real-time predictions of delay in the future, which would be a necessary step if the end goal is to provide timely interventions to reduce maladaptive forms of dilatory behavior (e.g., procrastination). However, the sample size is currently not large enough to provide accurate real-time predictions, meaning more data needs to be collected first, involving an expanded array of predictors. This could include task-specific factors (e.g., motivational aspects such as students' interest in a topic) or predictors related to students' time management skills. The latter could be achieved by analyzing behavioral patterns, which was successfully incorporated in a mixture models study by Park and colleagues Park et al. [2018]. Another promising avenue is to assess some of the learning-related factors objectively rather than subjectively, e.g., by replacing the self-directed learning questionnaire with clickstream-based indicators of self-regulated learning (see Li et al. [2020]).

6 Conclusion

In conclusion, when applying ML prediction models in the field, comparisons between various algorithms are needed to determine which ones deliver the highest performance. The suitability of a given ML algorithm when predicting dilatory behavior depends on the type of predictor. While objective predictors work best with a statistical approach (Bayesian multilevel models), subjective predictors are better served with Gradient Boosting Machines. When both types of variables are combined, Random Forests were the preferable ML algorithm in our case. Future studies need to increase the predictive performance (e.g., by expanding the roster of predictors) to allow such models to be implemented in learning management systems, ultimately enabling real-time predictions during a semester. This can then serve as the basis for interventions aiming at reducing procrastination and promoting timely submissions.

References

Amber D. Dumford and Angie L. Miller. Online learning in higher education: exploring advantages and disadvantages for engagement. *Journal of Computing in Higher Education*, 30(3):452–465, dec 2018. ISSN 18671233. doi:10.1007/s12528-018-9179-z.

Piers Steel. The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, 133(1):65–94, 2007. ISSN 1939-1455. doi:10.1037/0033-2909.133.1.65.

Dianne M Tice and Roy F. Baumeister. Longitudinal Study of Procrastination, Performance, Stress, and Health: The Costs and Benefits of Dawdling. *Psychological Science*, 8(6):454–458, nov 1997. ISSN 0956-7976. doi:10.1111/j.1467-9280.1997.tb00460.x.

William Doherty. *The Internet and Higher Education*, 9(4):245–255, oct 2006. ISSN 1096-7516. doi:10.1016/J.IHEDUC.2006.08.004.

Albert Ellis and William J. Knaus. *Overcoming procrastination : or how to think and act rationally in spite of life's inevitable hassles*. Institute for Rational Living, 1977. ISBN 0917476042.

Danya M Corkin, L Yu Shirley, and Suzanne F Lindt. Comparing active delay and procrastination from a self-regulated learning perspective. *Learning and Individual Differences*, 21(5):602–606, 2011a.

Timothy A. Pychyl. Savouring the Flavours of Delay. *English Studies in Canada*, 34(2-3):25, jun 2008. ISSN 0317-0802.

Gwo-Jen Hwang, Hui-Chun Chuz, and Chengjiu Yin. Objectives, Methodologies and Research Issues of Learning Analytics. *in Interactive Learning Environments*, 25(2):143 – 146, March 2017. ISSN 1744-5191. doi:https://doi.org/10.1080/10494820.2017.1287338.

Sepinoud Azimi, Carmen-Gabriela Popa, and Tatjana Cucić. Improving Students Performance in Small-Scale Online Courses - A Machine Learning-Based Intervention. *in International Journal of Learning Analytics and Artificial Intelligence for Education (iJAI)*, 2(2):80 – 95, November 2020. doi:10.3991/ijai.v2i2.19371.

Abelardo Pardo, Feifei Han, and Robert A. Ellis. Combining University Student Self-Regulated Learning Indicators and Engagement with Online Learning Events to Predict Academic Performance. *in IEEE Transactions on Learning Technologies*, 10(1):82 – 92, March 2017. ISSN 1939-1382. doi:10.1109/TLT.2016.2639508.

Rianne Conijn, Chris Snijders, Ad Kleingeld, and Uwe Matzat. Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS. *in IEEE Transactions on Learning Technologies*, 10(1):17 – 29, March 2017. ISSN 1939-1382. doi:10.1109/TLT.2016.2616312.

Aftab Akram, Chengzhou Fu, Yuyao Li, Muhammad Yaqoob Javed, Ronghua Lin, Yuncheng Jiang, and Yong Tang. Predicting Students' Academic Procrastination in Blended Learning Course Using Homework Submission Data. *IEEE Access*, 7:102487–102498, 2019. ISSN 2169-3536. doi:10.1109/access.2019.2930867.

Rebeca Cerezo, María Esteban, Miguel Sánchez-Santillán, and José C. Núñez. Procrastinating behavior in computer-based learning environments to predict performance: A case study in Moodle. *Frontiers in Psychology*, 8(AUG): 1–11, 2017. ISSN 16641078. doi:10.3389/fpsyg.2017.01403.

Christof Imhof, Per Bergamin, and Stéphanie McGarrity. Prediction of dilatory behaviour in online assignments. *Learning and Individual Differences*, 88, May 2021. doi:10.1016/j.lindif.2021.102014.

Danya M. Corkin, Shirley L. Yu, and Suzanne F. Lindt. Comparing active delay and procrastination from a self-regulated learning perspective. *Learning and Individual Differences*, 21(5):602–606, 2011b. ISSN 10416080. doi:10.1016/j.lindif.2011.07.005.

Albert Bandura. *Social foundations of thought and action: A social cognitive theory.* Prentice-Hall series in social learning theory. Prentice-Hall, Inc, Englewood Cliffs, NJ, US, 1986. ISBN 0-13-815614-X (Hardcover).

Kristin Wäschle, Anne Allgaier, Andreas Lachner, Siegfried Fink, and Matthias Nückles. Procrastination and self-efficacy: Tracing vicious and virtuous circles in self-regulated learning. *Learning and Instruction*, 29:103–114, 2014. ISSN 09594752. doi:10.1016/j.learninstruc.2013.09.005.

Angela Hsin Chun Chu and Jin Nam Choi. Rethinking procrastination: positive effects of "active" procrastination behavior on attitudes and performance. *The Journal of Social Psychology*, 145(3):245–64, 2005. ISSN 0022-4545. doi:10.3200/SOCP.145.3.245-264.

Sanneke Bolhuis, apr 1996.

Malcolm Shepherd Knowles. *Self-directed learning : a guide for learners and teachers*. Chicago : Association Press, 1975.

D. R. Garrison. Self-Directed Learning: Toward a Comprehensive Model. *Adult Education Quarterly*, 48(1):18–33, nov 1997. ISSN 0741-7136. doi:10.1177/074171369704800103.

Nasim Saeid and Tahere Eslaminejad. Relationship between Student's Self-Directed-Learning Readiness and Academic Self-Efficacy and Achievement Motivation in Students. *International Education Studies*, 10(1), 2017. ISSN 1913-9039. doi:10.5539/ies.v10n1p225.

Marlene Schommer-Aikins and Marilyn Easter. Cognitive flexibility, procrastination, and need for closure linked to online self-directed learning among students taking online courses. *Journal of Business and Educational Leadership*, 8(1):112–121, 2018.

Matthew M. Cirigliano, Charles D. Guthrie, and Martin V. Pusic. Click-level Learning Analytics in an Online Medical Education Learning Platform. *Teaching and Learning in Medicine*, 32(4):410–421, aug 2020. ISSN 1040-1334. doi:10.1080/10401334.2020.1754216.

Simon Knight, Alyssa Friend Wise, and Bodong Chen. Time for Change: Why Learning Analytics Needs Temporal Analysis. *Journal of Learning Analytics*, 4(3):7–17, dec 2017. doi:10.18608/jla.2017.43.2.

Ji Won You. Examining the Effect of Academic Procrastination on Achievement Using LMS in e-Learning. *Educational Technology & Society*, 18(3):64–74, 2015. ISSN 1436-4522.

Yair Levy and Michelle M. Ramim. A Study of Online Exams Procrastination Using Data Analytics Techniques. *Interdisciplinary Journal of e-Skills and Lifelong Learning*, 8:097–113, 2012. ISSN 2375-2084. doi:10.28945/1730.

Alexandre Gareau, Mélodie Chamandy, Kristina Kljajic, and Patrick Gaudreau. The detrimental effect of academic procrastination on subsequent grades: the mediating role of coping over and above past achievement and working memory capacity. *Anxiety, Stress, & Coping*, 32(2):141–154, 2019.

Masanori Yamada, Yoshiko Goda, Takeshi Matsuda, Yutaka Saito, Hiroshi Kato, and Hiroyuki Miyagawa. How does self-regulated learning relate to active procrastination and other learning behaviors? *Journal of Computing in Higher Education*, 28(3):326–343, 2016. ISSN 18671233. doi:10.1007/s12528-016-9118-9.

María del Puerto Paule-Ruiz, Moises Riestra-González, Miguel Sánchez-Santillán, and Juan Ramón Pérez-Pérez. The procrastination related indicators in e-learning platforms. *Journal of Universal Computer Science*, 21(1):7–22, 2015. ISSN 09486968.

Danial Hooshyar, Margus Pedaste, and Yeongwook Yang. Mining educational data to predict students' performance through procrastination behavior. *Entropy*, 22(1):12, 2020.

Syed Muhammad Raza Abidi, Wu Zhang, Saqib Ali Haidery, Sanam Shahla Rizvi, Rabia Riaz, Hu Ding, and Se Jin Kwon. Educational sustainability through big data assimilation to quantify academic procrastination using ensemble classifiers. *Sustainability*, 12(15):6074, 2020.

Yeongwook Yang, Danial Hooshyar, Margus Pedaste, Minhong Wang, Yueh Min Huang, and Heuiseok Lim. Prediction of students' procrastination behaviour through their submission behavioural pattern in online learning. *Journal of Ambient Intelligence and Humanized Computing*, 2020. ISSN 18685145. doi:10.1007/s12652-020-02041-8. URL `https://doi.org/10.1007/s12652-020-02041-8`.

T. Nielsen, J. Dammeyer, M. L. Vang, and G. Makransky. Gender Fairness in Self-Efficacy? A Rasch-Based Validity Study of the General Academic Self-Efficacy Scale (GASE). *Scandinavian Journal of Educational Research*, 62(5): 664–681, 2018. ISSN 14701170. doi:10.1080/00313831.2017.1306796.

John W. Lounsbury, Jacob J. Levy, Soo Hee Park, Lucy W. Gibson, and Ryan Smith. An investigation of the construct validity of the personality trait of self-directed learning. *Learning and Individual Differences*, 19(4):411–418, 2009. ISSN 10416080. doi:10.1016/j.lindif.2009.03.001.

Justin D. McCloskey. *Finally, My Thesis On Academic Procrastination*. PhD thesis, 2012.

Jin Nam Choi and Sarah V Moran. Why not procrastinate? Development and validation of a new active procrastination scale. *The Journal of social psychology*, 149(2):195–211, 2009. ISSN 0022-4545. doi:10.3200/SOCP.149.2.195-212.

Marvin N Wright, Theresa Dankowski, and Andreas Ziegler. Unbiased split variable selection for random survival forests using maximally selected rank statistics. *Statistics in Medicine*, 36(8):1272–1284, 2017.

Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.

Martin Hlosta, Christothea Herodotou, Vaclav Bayer, and Miriam Fernandez. Impact of predictive learning analytics on course awarding gap of disadvantaged students in stem. In *International Conference on Artificial Intelligence in Education*, pages 190–195. Springer, 2021.

José A Ruipérez-Valiente, Ruth Cobos, Pedro J Muñoz-Merino, Álvaro Andujar, and Carlos Delgado Kloos. Early prediction and variable importance of certificate accomplishment in a mooc. In *European Conference on Massive Open Online Courses*, pages 263–272. Springer, 2017.

KDD Cup 2015. Kdd cup 2015: Predicting dropouts in mooc, 2015.

M. N. Wright and A. Ziegleri. ranger: A fast implementation of random forests for high dimensional data in c++ and r. *Journal of Statistical Software*, 77(1):1 – 17, 2017.

Paul-Christian Bürkner. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28, 2017. doi:10.18637/jss.v080.i01.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

Francesco Parrella. *Online support vector regression [master's thesis]*. PhD thesis, University of Genoa, Italy, 2007.

I.-S. Comsa, A. Molnar, I. Tal, P. Bergamin, C. Hava Muntean, G.-M. Muntean, and R. Trestian. A Machine Learning Resource Allocation Solution to Improve Video Quality in Remote Education. *in IEEE Transactions on Broadcasting*, pages 1 – 21, April 2021. ISSN 1557-9611. doi:10.1109/TBC.2021.3068872.

Ioan-Sorin Comsa. *Sustainable Scheduling Policies for Radio Access Networks Based on LTE Technology*. University of Bedfordshire, U.K., 2014. ISBN PhD Thesis.

Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53 – 65, 1987.

Davide Chicco and Giuseppe Jurman. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1):1–13, 2020.

Davide Chicco. Ten quick tips for machine learning in computational biology. *BioData mining*, 10(1):1–17, 2017.

M. Geetha T. Daniya and K. Suresh Kumar. Classification and Regression Trees with Gini Index. *in Advances in Mathematics: Scientific Journal*, 9(10):8237 – 8247, 2020. ISSN 1857-8438. doi:https://doi.org/10.37418/amsj.9.10.53.

Martin Hlosta, Zdenek Zdrahal, and Jaroslav Zendulka. Are we meeting a deadline? classification goal achievement in time in the presence of imbalanced data. *Knowledge-Based Systems*, 160:278–295, 2018.

Raul Alcaraz, Arturo Martinez-Rodrigo, Roberto Zangroniz, and Jose Joaquin Rieta. Early prediction of students at risk of failing a face-to-face course in power electronic systems. *IEEE Transactions on Learning Technologies*, 2021.

Renzhe Yu, Qiujie Li, Christian Fischer, Shayan Doroudi, and Di Xu. Towards accurate and fair prediction of college success: Evaluating different sources of student data. *International Educational Data Mining Society*, 2020.

Dirk Tempelaar, Bart Rienties, and Quan Nguyen. The contribution of dispositional learning analytics to precision education. *Journal of Educational Technology and Society*, 24(1):109–122, January 2021.

Jihyun Park, Renzhe Yu, Fernando Rodriguez, Rachel Baker, Padhraic Smyth, and Mark Warschauer. Understanding student procrastination via mixture models. *International Educational Data Mining Society*, 2018.

Qiujie Li, Rachel Baker, and Mark Warschauer. Using clickstream data to measure, understand, and support self-regulated learning in online courses. *The Internet and Higher Education*, 45:100727, 2020.

Christof Imhof has been employed at the Institute for Research in Open, Distance, and e-Learning (IFeL) since 2016 and is currently a doctoral student at the University of Bern. His research focus lies primarily on procrastination and other types of dilatory behavior in the context of adaptive learning, which also serves as the topic of his doctoral thesis. Other research interests include the detection of emotions with objective measures such as eye-tracking combined with emotional word lists.

Ioan-Sorin Comsa is a Data Scientist with the Institute for Open, Distance and eLearning (IFeL). He received a joint PhD from the University of Bedfordshire, U.K. and the University of Applied Sciences of Western Switzerland, in 2015. He worked as a Senior Research Engineer with CEA-LETI, France. Since 2017, he has been a Research Assistant with Brunel University London being actively involved in EU NEWTON project. His research interests include personalised and adaptive learning, machine learning and wireless communications.

Martin Hlosta is a research fellow at the Institut for Distance Learning and eLearning Research (IFeL) working on projects for adaptive learning in education. Before, he led OUAnalyse at The Open University, a project improving student retention via machine learning, selected by UNESCO among four of the best projects using AI in Education in 2020. His research mainly focuses on Predictive Learning Analytics and its impact, in particular to target existing educational inequalities.

Behnam Parsaeifard has joined the Institute for Open-, Distance- and eLearning (IFeL) as a researcher in November 2021. Prior to that, he received his PhD in computational physics from the University of Basel and was a research intern at the Visual Intelligence for Transportation (VITA) lab of EPFL. He is interested in applying data mining and machine learning techniques to improve adaptive learning and learning analytics.

Ivan Moser is a Researcher and Lecturer at the Institute for Research in Open, Distance and eLearning (IFeL). Previously, he studied Psychology at the University of Bern, where he earned his PhD at the Institute of Cognitive Psychology. In the course of his research career, Ivan Moser developed an increasing fascination for the numerous new possibilities that virtual reality (VR) opens as a learning technology. As leader of several VR-related research projects at FFHS, he investigates the potential and limits of virtual learning environments.

Per Bergamin is Professor for Didactics in Distance Education and E-Learning at the Swiss Distance University of Applied Sciences (FFHS). Since 2006 he acts as the Director of the Institute for Research in Open-, Distance- and eLearning (IFeL) and from 2016 on he holds also the UNESCO Chair on personalised and adaptive Distance Education. His research activities focus on self-regulated and technology-based personalized and adaptive learning. Central aspects are instructional design, usability and application implementation.

Selbständigkeitserklärung

„Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe r des Gesetzes über die Universität vom 5. September 1996 und Artikel 69 des Universitätsstatuts vom 7. Juni 2011 zum Entzug des Doktortitels berechtigt ist.

Für die Zwecke der Begutachtung und der Überprüfung der Einhaltung der Selbständigkeitserklärung bzw. der Reglemente betreffend Plagiate erteile ich der Universität Bern das Recht, die dazu erforderlichen Personendaten zu bearbeiten und Nutzungshandlungen vorzunehmen, insbesondere die Dissertation zu vervielfältigen und dauerhaft in einer Datenbank zu speichern sowie diese zur Überprüfung von Arbeiten Dritter zu verwenden oder hierzu zur Verfügung zu stellen.“