IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 May 2022, accepted 21 June 2022, date of publication 1 July 2022, date of current version 13 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3187732

RESEARCH ARTICLE

Profiling Students' Self-Regulation With Learning Analytics: A Proof of Concept

MARTÍN LIZ-DOMÍNGUEZ, (Student Member, IEEE), MARTÍN LLAMAS-NISTAL, (Senior Member, IEEE), MANUEL CAEIRO-RODRÍGUEZ, (Senior Member, IEEE), AND FERNANDO A. MIKIC-FONTE, (Member, IEEE)
atlanTTic Research Center, University of Vigo, 36310 Vigo, Spain

Corresponding author: Martín Liz-Domínguez (mliz@gist.uvigo.es)

This work was supported in part by the Xunta de Galicia (Galicia's Singular Research Center, accreditation 2019–2022); in part by the European Union through the European Regional Development Fund (ERDF); in part by the Galician Regional Government under Project ED431B 2020/33; and in part by the Galician Regional Government through the "Programa de Axudas á Etapa Predoutoral da Xunta de Galicia—Consellería de Educación, Universidade e Formación Profesional," with the purpose of supporting research activities carried out by Ph.D. students.

ABSTRACT The ability to regulate one's own learning processes is a key factor in educational scenarios. Self-regulation skills notably affect students' efficacy when studying and academic performance, for better or worse. However, neither students or instructors generally have proper understanding of what self-regulated learning is, the impact that it has or how to assess it. This paper has the purpose of showing how learning analytics can be used in order to generate simple metrics related to several areas of students' self-regulation, in the context of a first-year university course. These metrics are based on data obtained from a learning management system, complemented by more specific assessment-related data and direct answers to self-regulated learning questionnaires. As the end result, simple self-regulation profiles are obtained for each student, which can be used to identify strengths and weaknesses and, potentially, help struggling students to improve their learning habits.

INDEX TERMS Data analysis, data processing, engineering education, learning management systems, linear regression, self-regulated learning.

I. INTRODUCTION

Historically, one of the most important issues faced in educational institutions in general, and universities in particular, is student underperformance, leading to failure in courses and possible dropout from the educational program altogether. This has negative implications not only for the student as an individual, but also for learning institutions, which may acquire negative reputations due to high failure and dropout rates. It could even be considered detrimental to society as a whole, as resources allocated into the education of an important amount of students ultimately go to waste [1].

One of the focuses of this paper is learning analytics (LA). In recent years, the advent of this discipline has provided new possibilities for the implementation of tools to support and improve learning. The discipline of LA started to gain mainstream popularity at the start of the decade of 2010, and was defined then as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" [2]. Since then, the interest garnered by LA as a research discipline grew drastically [3], being a core part of many types of applications, such as analytics dashboards [4], resource recommenders [5], or early warning systems [6].

The other focus of this document is self-regulated learning (SRL). The concept of SRL and the first relevant self-regulation models surfaced in the 1980s, with Barry J. Zimmerman as the main contributor at the time, although other researchers had previously worked on elements which are considered to be SRL components to this day [7]. There are several features that characterize a

The associate editor coordinating the review of this manuscript and approving it for publication was Dongxiao Yu.

self-regulated student: they are able to control their behavior, motivation, and cognition; they pursue specific learning goals; and they can change their own actions and behavior for what they consider their best interest towards optimizing learning and performance [8].

Zimmerman's original SRL model pictured interactions between three levels of self-regulation: personal, behavioral and environmental [9]. Since then, many other authors have proposed different and more complex SRL models of their own [10].

Self-regulated learning capabilities are widely seen as key aspects that are directly related to academic success [11]–[13]. And, despite this, SRL is an often overlooked factor by teachers and students alike, as typically neither collective has had specific formation in order to understand what self-regulation is, its importance regarding any learning activity and its impact over academic performance. Naturally, poor understanding of SRL leads to the impossibility of assessing individual students' self-regulation capabilities. This may be seen as a wasted opportunity for improving students' learning processes, as SRL techniques can be taught and adopted by students of any age with proper training methods [14].

The purpose of this paper is showing how learning analytics techniques can be used to generate simple metrics in order to assess university students' proficiency in self-regulation. These metrics, focusing on different SRL categories, may be useful for teachers and students in order to have a basic understanding of how well individual students regulate several aspects of their own learning processes. Additionally, this paper shows how SRL metrics relates to student performance, testing whether they can be used to identify students at risk of failure.

The main contributions of this study are the following:

- The definition of a SRL framework made up of a limited set of categories, aiming to be practical and simple to understand. The selection and scope of the categories themselves are rooted in previous literature within the SRL field of knowledge.
- An experiment that shows an application of the aforementioned framework in a university-level course. Both observational and self-reported data from the course were gathered and transformed into metrics associated with the different categories in the framework. Correlation tests were performed in order to assess the influence of these metrics on each other and on students' course performance.

The rest of this document will follow this structure: Section II presents the SRL framework that has been defined for this study, covering the choice of SRL categories and summaries of relevant previous works for each of them. In Section III, contextual information regarding this study is provided, including the types of data that we worked with and the characteristics of the university course these data were obtained from. Section IV shows the resulting metrics computed for each of the defined SRL categories, which are then tested in order to assess their correlation with student performance. Finally, Section V provides some closing thoughts and lines for future work.

II. SELF-REGULATED LEARNING FRAMEWORK

A. DATA AS A WAY TO MEASURE SELF-REGULATED LEARNING

As should be expected, measures of self-regulated learning are all based on learner-related data which are collected in one way or another. SRL researchers may use several types of data sources, the choice of which may be based on factors such as the specific component that they want to measure, or simply working with the data that they have available.

Many pieces of research have focused on measuring SRL using self-reported instruments. Roth *et al.* [15] performed a systematic literature review regarding this topic, identifying the following instrument types: questionnaires, interviews, think-aloud protocols and learning diaries. Regardless of the instrument, the procedure involves learners being requested to provide key information about their attitude towards specific aspects of self-regulation. However, these procedures have some limitations. First, provided information may often be biased and inaccurate as it is affected by learners' perception and point of view. Second, these measurements may not accurately show how learners dynamically adapt and modify their behavior during learning, as they are focused on a more static and generic view. And finally, they involve an extra effort for students that, in many cases, they have no desire to perform. Despite these problems, the use of self-reported instruments is still common practice today due to their ease of implementation and the ability to obtain information regarding some SRL aspects which would be very difficult to obtain through different means.

A different approach to measure SRL involves obtaining data regarding learners' behavior in online learning tools and the subsequent application of data mining and machine learning algorithms. Measurements using this method have a key difference compared to the use of self-regulated learning instruments. While the previous case focuses on attitudes based on learners' perceptions [16], here the focus is on real learner behaviors, represented in traces that are logged by learning management systems and other tools to support learning. Thus, these works aim to assess SRL as an event (time-based and task-related of known start and end), rather than an aptitude (a steady feature of the learner.) [17]

Currently, SRL measurements using learning analytics techniques usually follow one of two approaches: either based solely on observational data (typically obtained from online tools such as early warning systems), or complemented by information from self-reported instruments [17].

B. SELF-REGULATED LEARNING CATEGORIES

Researchers that work with SRL typically define a set of categories, each one representing a specific aspect of self-regulation. Originally, Zimmerman and Pons established a set of fifteen categories of self-regulated learning strategies, including but not limited to: self-evaluation, goal-setting and planning, information seeking, rehearsing and memorizing, help seeking and record reviewing [18].

However, there is no clear standard on which SRL categories should be considered by research. Instead, authors define the number and type of categories which best suit their particular experiments. For instance, Perels *et al.* [19] work with six categories: motivation, learning strategies, self-reflection, self-efficacy, problem solving and goal setting. Other examples include Cheng [12], who works with just four categories: goal setting, action control, learning motivation, and learning strategies; or Fabriz *et al.* [20], who define nineteen much more specific categories in their study, including aspects such as procrastination or reflection.

In a previous study, we performed a systematic literature review on the use of data analytics techniques for the assessment of SRL features and behaviors of students [21]. A total of 109 studies were found that fulfilled the eligibility criteria for the period comprised between 2011 and 2019. Most of the studies relied on logged data from learning management systems and other learning supporting tools, such as assessment tools, chats, video recordings and annotation software. In some cases, multimodal trace data was also used, including heart rate, eye-tracking data, step count or weather conditions.

The techniques used were mainly cluster analysis, classification, and temporal data mining techniques like sequence mining and process mining.

The most frequently measured SRL components were:

1) Goal-setting, planning, and time management (measured in 57 studies): the learner is able to set precise goals for him or herself.
2) Keeping, reviewing records, and monitoring (measured in 36 studies): this refers to the learners' monitoring of their progression towards achieving goals.
3) Emotion regulation (measured in 21 studies.)
4) Learning strategies (measured in 19 studies): this includes behaviors such as reading, repeating, elaboration, judgment of relevance, taking notes, summarizing, coordinating different information sources, activating prior knowledge, peer-learning, processing, questioning, and problem-solving.
5) Self-evaluation (measured in 18 studies): this refers to "comparisons of self-observed performances against some standard, such as one's prior performance, another person's performance, or an absolute standard of performance."
6) Seeking information and social help (measured in 16 studies): this pertains learners looking for assistance and information from others.
7) Organizing and transforming (measured in 6 studies.)
8) Environmental structuring (measured in 4 studies): this involves learners' use of e-learning platforms to alter the digital area for achieving their learning objectives [22].
9) Rehearsing and memorizing (measured in 4 studies.)

Our priority in this study is to define a reduced set of categories that could be easy to understand for teachers and students without any specific SRL knowledge. Additionally, we want our categories to be feasibly associated with the data at our disposal. Considering the information presented in this section, we defined the following five categories:

- Learning strategies.
- Time management.
- Resource management.
- Self-monitoring and self-assessment.
- Motivation and self-confidence.

The following subsections provide more in-depth descriptions of each category, as well as summaries of relevant pieces of work.

C. LEARNING STRATEGIES

Learning strategies encompass the variety of ways in which students interact with course resources and tasks. Some learning strategies can be more effective than others depending on the context of each course, which is defined by factors such as the assessment method or type of contents. Thus, an important aspect of self-regulation is the fact that the student is aware of the learning strategies that they use and can adapt and change these strategies depending on the learning context [23].

Researchers Jelena Jovanovic, Dragan Gasevic and Abelardo Pardo have led a series of studies within the context of a flipped classroom and focused on first-year engineering university courses.

In [24], Gasevic *et al.* make a distinction between self-reported learning strategies — that is, those deducted from students' answers to a questionnaire — and strategies that are discovered because of applying learning analytics techniques. For self-reported strategies, the authors make use of the Study Process Questionnaire [25] and on the other hand, the authors use LMS activity logs (reading learning material, watching videos or performing formative or summative exercises.) Thus, using either self-reported or log-based data sources, the authors obtain two separate classifications of students regarding their use of superficial or deep learning strategies, which were observed to be weakly correlated. The authors consider that self-reported strategies are the ones perceived by students, and that they do not necessarily correspond to what analysis of learning sequences tell — the actual strategies. Thus, log-based data is usually more reliable than self-reports: classification using learning sequences ended up being more tightly correlated to exam scores. Either way, it was observed that students using deep strategies obtained significantly better scores.

In [16], Jovanovic *et al.* classified students regarding their learning strategies, determined via learning analytics. This study considers not only the types of activities that students prioritize, but also the total number of learning sequences. As a result, students are classified into five different groups: intensive, strategic, highly strategic, selective and highly selective, in decreasing level of activity. Students in the intensive group are the ones more closely associated with deep learning strategies, while the ones in the selective groups tend to use superficial strategies. This study also shows important performance differences between students using deep and superficial approaches to learning.

Also involving Jovanovic, Gasevic and Pardo, Matcha *et al.* published a study in the same line of work [26], distinguishing between tactics and strategies and implementing a personalized feedback system for students [27]. Tactics are defined as a classification for learning sequences depending on the predominant type of activity. The combination of tactics that a student normally utilizes define their placement into one of the three groups of strategies (selective, strategic and intensive.)

It is worth mentioning that studies [16] and [26] present analysis results on a weekly basis, regularly highlighting predominant tactics for each learning strategy.

Highlighting the work of a different research group, Uguina-Gadella *et al.* are the authors of a study that is smaller in scale [28], involving a self-regulation questionnaire [29], and thus, working with self-reported strategies. This specific study focuses on the ability or inability of students in keeping a regular study pattern, summarized in a single question: Which of the following better represents your study habits?: a) I normally distribute my study sessions throughout multiple days or weeks; and b) I usually study in a single session before the exam. Students were classified into one of two groups depending on their answer to this question. Throughout the course, it was observed that the study pattern for each student roughly matched their answer in the questionnaire, suggesting a certain level of self-consciousness regarding their own habits.

We can consider learning sequences as the main basis in order to detect strategies. Additionally, questionnaires can be used in order to assess whether students are aware or not of the strategies they use, by comparing their answers with the observed learning sequences.

D. TIME MANAGEMENT

The topic of time management by students has been widely studied by researchers. A student's capability to appropriately manage their time plays an important role in their academic performance and learning efficiency. In order to achieve an adequate use of their time, a student must control not only the amount of time they allocate to do academic tasks, but also the time frames they choose in order to do so. Procrastination is a problem that has been approached in many ways, due to its virtual omnipresence: estimations say that between 80% and 95% of students procrastinate, and 50% of them do so in a consistent and problematic manner [30].

Wolters *et al.* [31] are the authors of a study in which procrastination is predicted using a questionnaire. This work includes a detailed preamble with explanations on what time management and procrastination are, as well as a solid literature review. This study makes a distinction between two types of procrastination:

1) Traditional or passive procrastination corresponds to the typical definition of the term: a delay in actions or decisions necessary to complete academic tasks.
2) Active procrastination involves an intentional delay in the start and completion of tasks, with a strategic justification. Active procrastination implies a calculated decision by a student, who must be sure they can complete the delayed tasks on time.

The experiment results confirmed the hypothesis that delays in starting and finishing academic tasks tend to seriously jeopardize students' capacity to meet deadlines. It is important to note that this is true for the cases of both passive and active procrastination. A limitation of this study is that it exclusively uses self-reported data.

Asarta and Schmidt [32] focus on a blended learning context, with slides and recorded speech replacing traditional lectures, while using in-classroom time to solve practical exercises, answer questions and, at the end of each course module, doing an exam. This study considered the volume, intensity, regularity and moments in which students accessed online contents. Out of these factors, it was observed that volume and intensity of online accesses were not significantly related to the final grade. However, a correlation was detected between both regularity and moments and the students' final grade. The authors highlight three aspects that are related to regularity and help avoid procrastination:

- Pace, which involves the student's capacity to keep up-to-date with watching video lectures.
- Non-accumulation, which implies that the student avoids watching the bulk of the lessons during the two days prior to the exam.
- Consistency, which a student achieves if they dedicate a similar amount of time every week to watching lectures.

Finally, Steel [33] analyzes several questionnaires that aim to predict procrastination, and combines aspects from many of them in order to create his own: the Pure Procrastination Scale. As an important conclusion of this study, and similar to what Wolters *et al.* wrote [31], Steel asserts that the distinction between several types of procrastination, such as passive or active, is not actually important.

Most studies dealing with student time management focus on the moments in which students do academic tasks, rather than the amount of time spent. This may be because the choice of time frames in which to perform academic tasks, and more specifically, procrastination, can be generalized more easily than the amount of time to invest, which could significantly vary depending on the student. This is why the detection of procrastination is probably the best way to assess students' time management, in terms of self-regulated learning.

E. RESOURCE MANAGEMENT

In order to complete tasks and carry out a successful learning process, students must take advantage of all types of resources they have available. We define as a resource any kind of material or entity which the student can interact with in order to favor their learning interests (learning materials, attendance to lectures, interactions with teachers and other students, and the use of libraries and other study spaces).

All these factors may have an important influence over the study habits of a student, and can be difficult for them to optimally manage, as well as for a teacher or supervisor to monitor.

Most of the studies that analyze student resource management and its impact over academic results focus on the use of online academic material (slides, notes or videos, etc.) due to them being easier to monitor. However, some researchers had the opportunity to access data relative to other types of resources, and have performed analysis tasks using them.

A study by Jovanovic *et al.* in which students were classified depending on their use of online learning material within the context of a flipped classroom [16] is a great example of analysis of resource use by students in order to obtain performance predictions.

A study by a Chinese research group led by Wang [34] included information relative to grades and credits enrolled, as well as usage logs of university buildings, such as the library and dorms. Principal component analysis confirmed the existence of a correlation between book borrow patterns from the library and academic achievement. This study is particularly interesting because it is able to utilize types of data that are rarely seen in similar works.

Finally, a Spanish research group formed by Díaz-Lázaro, Solano-Fernández and Sánchez-Vera performed a study in order to assess how students learn and collaborate through social networks, under the perspective of social learning analytics (SLA) [35]. The study analyzes messages and content posted in a private Facebook group, as well as reactions and comments by group members. The authors note that reactions (''likes'') are the most common activity type by students, while comments on post are much rarer. Additionally, student participation is very lopsided: about 20% of students are responsible for half the content posted on the platform. Lastly, the presence of the professor is instrumental in order to trigger student activity (the teacher is, by far, the member who poster the highest amount of publications in the group). On the other hand, an important amount of students never uploaded an original post to the platform.

There exists a great amount of literature relative to resource use in education. However, documents that analyze this topic under the viewpoint of SRL are scarcer. Most of them focus on several applications related to learning analytics, such as predictive analysis and production of indicators.

F. SELF-MONITORING AND SELF-ASSESSMENT

Self-monitoring is defined as the student's capability to realize that they are making progress towards their academic objectives as they study or perform tasks. Meanwhile, self-assessment skills involve reflection on a previous task or study session, making sure that all goals established for said session were accomplished. In both cases, the student must be able to detect deficiencies in their work methods and apply solutions to improve them.

Fabriz *et al.* aimed to improve students' self-monitoring by making them keep a study diary [20]. In order to fill said diary, students were asked to write down their reflections regarding multiple SRL categories, such as study planning, motivation, help seeking or self-monitoring [19]. This way, students make a specific effort in order to supervise their own learning. However, the authors specify that benefits were only observed when coupling this activity with specific formation regarding self-regulation.

From the perspective of data visualization, there are dashboards that were developed with the main purpose of foster student reflection on their own activity. Santos *et al.* are the creators of StepUp! [36], a tool that registers student activity in the different online tools that may be used within a course, allowing students to compare the intensity of their activity with their peers. According to the students that used this tool, the possibility of visualizing their activity allowed them to understand their own effort management in different tasks. However, the provided information was generally insufficient in order to understand how other students allocated their time and effort.

As we could observe, there are many approaches to fostering self-monitoring and self-assessment of learning: questionnaires, performing specific activities or creating data visualizations. There is, however, a common problem: it is difficult to control how well a student exercises self-monitoring and self-assessment using only LMS activity data. Therefore, the inclusion of self-monitoring and self-assessment questions seems necessary. A clear example is the one performed by Fabriz *et al.* [20].

G. MOTIVATION AND SELF-CONFIDENCE

This category involves several types of emotional factors experienced by students, which directly affect their learning, performance, and self-regulation capabilities. These factors can be reflected in aspects such as setting and pursuing learning goals, the milestones which the student consider as reachable or unreachable, the estimated value of tasks and subjects, or the mental strength to overcome the difficulties that the course poses. Moreover, proper management of motivation and self-confidence can help avoid negative emotions that can appear during the learning process, such as fear or anxiety.

Paul R. Pintrich is probably the most influential researcher on the subject of motivation as a learning component, with a series of papers published in the 90s and 2000s that are still relevant to this day. Pintrich is also the main author of the Multiple Strategies for Learning Questionnaire (MSLQ) [37], one of the most well-known and reference self-regulation questionnaire, which has a particular focus on learning strategies and motivational aspects.

In [38], Pintrich defines three categories for different types of motivational factors in learning:

- Self-sufficiency: the way students judge their own capacity to reach certain learning goals or complete academic tasks.
- Task value estimation: the perception that students have regarding the importance of specific tasks, determined by both the personal interest of the student in the task and the utility that said task can provide towards the future. In this context, the "future" comprises many types of short-, mid-, and long-term events, such as the next task in the course, following courses in the degree or required knowledge to perform a job once studies are finished, respectively.
- Goal orientation, which is itself divided into three subcategories:
  - Mastery goals: the student prioritizes mastering the contents of the subject, according to criteria that the very own student sets.
  - Extrinsic goals: the student focuses on obtaining good grades and meeting the standards of other people, such as teachers or family.
  - Relative goals: the student aims to outperform their peers.

In [39], Pintrich presents a conceptual framework for assessing self-regulation and motivation in university students. According to the model presented in this article, regulation of motivation by a student is considered as a self-regulation aspect, and includes the previously listed factors: self-sufficiency, task value estimation and goal orientation. The author states that the MSLQ questionnaire contains questions that assess several motivational factors of students, but it does not include questions regarding students' active efforts to monitor and regulate their own motivation and confidence.

Self-reported data is generally essential in order to assess students' motivation and self-confidence. Lonn *et al.* [40] performed two questionnaires in order to find out details regarding students' motivation. Similarly, Mega *et al.* [41] studied the effects of motivation, emotions and level of self-regulation by students over academic performance, using self-reported data, with questions focused on self-regulation, emotions and motivation. Using these data, after an exploratory analysis, the authors developed a structural equation model (SEM) in order to observe correlations between the data, and additionally, what factors influenced academic achievement.

Regarding non-self-reported data, participation measures could be used, especially if gamification activities are introduced. However, it is hard to identify the reasons for high or low levels of participation without any extra data. Participation data can be a good complement, but questionnaires will be the main source of data for this SRL component.

III. EXPERIMENT DESCRIPTION

A. EDUCATIONAL CONTEXT

This experiment targets a higher education context, more specific, a first-year Computer Architecture course from the Telecommunications Engineering degree at University of Vigo. The scope of activities in this course includes theory lectures covering the fundamentals of computer architecture and information representation, as well as lab assignments regarding assembly programming.

The theory part of this subject follows a flipped classroom methodology: lectures are provided to students in video format via the institutional LMS, which should be watched at home, and on-site class sessions are instead used for solving practical problems and answering questions. Additionally, the course features a continuous assessment model in which students undertake a series of short exams every two to three weeks, each covering the contents of the subject that were taught in the respective period. In total, students undertake 6 of these exams throughout the course. Alternatively, students may choose to be assessed by means of a single final exam covering the entire subject. However, almost every student chooses to follow the continuous assessment system, especially those who are taking the subject for the first time. [42].

Besides the university's own LMS, which is based on Moodle, the subject uses the online platform BeA (Blended e-Assessment) for exam management: design, assessment and reviews are all performed using this tool [43]. From the viewpoint of the teacher, the main advantage of BeA is speeding up the exam assessment process, as for every question, each possible mistake students may make only needs to be defined once, and then assigned as many times as needed. Meanwhile, students can use BeA to see their exam assessments and error explanations without the need to make an in-person appointment with the instructor.

This study will be analyzing data from the 2020/2021 academic year. The target course took place between February and May of 2021.

In recent academic years, the overall success rate in this subject ranged between 30% and 40% [44], and thus, students that retake the subject make up a significant fraction of each year's enrollments. Nevertheless, and for the sake of analyzing a subset of students with similar backgrounds and characteristics, we will only be considering students who were enrolled in the subject for the first time during the 2020/2021 academic year. According to our observations, the use patterns of online learning tools by retaking students are different from those of first-takers, and would thus warrant separate analysis.

During the studied academic year, there were 115 students who were enrolled for the first time in the Computer Architecture course.

B. AVAILABLE DATA

The data which this study works with comes from a combination of observational and self-reported data sources. As was mentioned in Section II-A, this approach is commonly used by researchers in SRL-measuring experiments. Observational data consists of log entries from the two online learning-supporting platforms that are used in the course:

Moodle and BeA. On the other hand, self-reported data is gathered using SRL questionnaires handed out to students.

Moodle logs register any activity that users — including students, teachers and course administrators — perform in the platform. If we focus specifically on students, we can observe that all actions performed on Moodle are indeed logged, together with a timestamp. These actions include, but are not limited to: accessing the main course page, viewing or downloading a document, watching a video, viewing or participating in a forum, perform self-assessment questionnaires, checking course grades, or viewing their own profile information or another user's public profile. When put together, these pieces of information can be used to reconstruct each student's sessions using the LMS, including the activities that they performed and the time lapse between them. This is akin to the concept of *learning sequences*, described by Khan and Pardo in [45] and then used in subsequent studies by Gasevic, Jovanovic *et al.* [16], [24].

An important parameter which defines the way learning sequences are interpreted is the maximum amount of time between two consecutive logged events by a student for them to be considered within the same sequence. In their definition of learning sequences, Khan *et al.* set a 30-minute maximum time within consecutive activities. However, we did not have the certainty that this value would properly fit our data. If the maximum interval between consecutive activities is too short, many events could be considered part of different learning sequences even though it would make sense to fit them within the same one. On the other hand, should this interval be too large, events belonging to clearly different sessions might be grouped into the same learning sequence.

In order to pick an appropriate time interval between events, we computed the learning sequences from our Moodle logs trying values ranging from 10 minutes to 6 hours, and counted the total number of sequences found for each one of them. Fig. 1 shows the results of these computations. It can

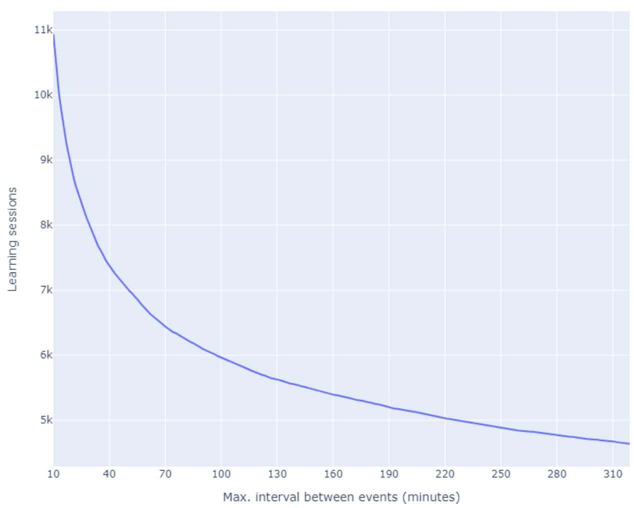


FIGURE 1. Number of observed learning sequences in moodle logs depending on the maximum time interval allowed between consecutive events.

be observed that the resulting curve follows a trend similar to exponential decay. In order to avoid the aforementioned problems derived from choosing inappropriately big or small interval values, we settled on a 50-minute maximum time between events, a value close to the knee of the represented curve.

Additionally, BeA logs are also available to utilize in this study. Thanks to this platform, it is possible to gather in-depth data regarding exam assessment, such as the specific errors that students make in each exam question. Additionally, it is possible to track student activity in the platform, obtaining data on the specific information that students check while using BeA.

On the other end of the spectrum, self-reported data was obtained directly from students via SRL questionnaires as a complement to the observational data. These questionnaires consisted of statements directly related to one of the five SRL categories of interest, to which students stated their level of agreement using a 1 to 5 Likert-like scale. The questionnaire items were inspired by existing SRL instruments in the literature, such as the Study Process Questionnaire [46] and the Multiple Strategies for Learning Questionnaire [37], although the questions were adapted to better fit the characteristics of the course and the learning context. More details regarding how the questionnaires were built can be found in [47].

Questionnaire distribution was performed as follows: at the beginning of the semester, students were asked to perform a 20-item SRL questionnaire on paper during one of the in-classroom sessions. Afterwards, three subsequent 7-item questionnaires were enabled on BeA for students to submit at home. These smaller questionnaires were spread throughout the duration of the course, with a new one being accessible for students every four weeks.

The non-mandatory nature of these SRL questionnaires led to an issue of low student participation. The initial 20-item questionnaire was answered by virtually every class-attending student, with a total of 113 submissions. When considering only the newly-enrolled students, 80 out of the total 115 answered the initial, on-site questionnaire. However, the following questionnaires in BeA only registered between 17 and 22 answers, less than 20% of the ones who performed the first questionnaire. Consequently, observational data is overall considered much more reliable in this particular study.

IV. RESULTS

As aforementioned, the goal of this data analysis experiment is obtaining metrics for simple SRL profiling of students. Said metric serve as ordering criteria that allow us to compare the proficiency of students in each of the five studied SRL categories.

This section presents the metrics that were chosen to represent each of the five SRL categories in our framework, which will form a data set together with students' final grades in the course. Afterwards, a measure of the correlation between these indicators and students' performance in the course, represented by the final grades, is included as a relevance assessment of the obtained metrics.

A. SRL METRICS

A summary of all generated metrics can be seen in Table 1, including the SRL category that each one is associated to and the minimum and maximum values observed in our data set. The following subsections provide more in-depth information regarding each category.

TABLE 1. Self-regulated learning metrics.

#	Category	Metric	[min, max]
1: LS1	Learning strategies	Total number of learning sessions throughout the course.	[0, 199]
2: LS2	Learning strategies	Average length (number of activities) in learning sessions.	[0, 14.4]
3: TM1	Time management	Average time spent in online learning sessions (minutes per week.)	[0, 262.75]
4: TM2	Time management	Number of weeks in which the student spent at least 48 minutes in learning sessions.	[0, 16]
5: RM	Resource management	Number of individual course resources that were accessed at least once.	[0, 100]
6: SMA1	Self-monitoring and self-assessment	Time taken by the student to see a exam correction since becoming available (minutes after first student.)	[0, 14911]
7: SMA2	Self-monitoring and self-assessment	Average score obtained in self-assessment tests.	[0, 10]
8: SMA3	Self-monitoring and self-assessment	Total number of attempts in self-assessment tests.	[0, 28]
9: MSC	Motivation and self-confidence	Average score in "Motivation and self-confidence" questionnaire items.	[2, 4.5]

1) LEARNING STRATEGIES

Metrics related to students' learning strategies are based on characteristics of their observed learning sequences. More specifically, we will be looking at the total number of learning sessions in which the student engages during the course, and also their average length (that is, the number of different activities, such as watching videos or performing self-assessment tests, which are included in each learning sequence.) It is important to note that, at least on paper, the *number* and *length* of learning sequences are not necessarily related to their *duration*, which refers to the amount of time the student invests in their learning sessions. This aspect will be studied more in-depth in Section IV-A2.

The justification for considering the number of learning sessions as a metric is similar to what is seen in Jovanovic *et al.* [16], as was mentioned in Section II-C: in their work, these authors observe the number of learning sequences in order to gauge the "intensity" of students' study habits. Fig. 2 shows the distribution of students regarding their total number of learning sessions.

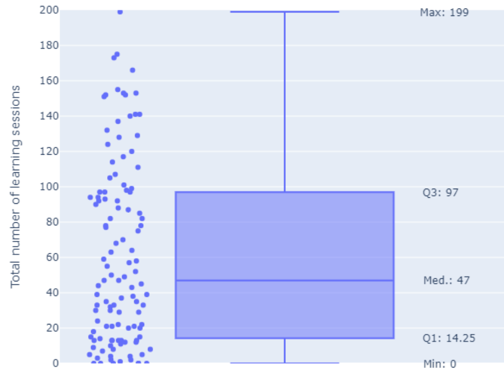


FIGURE 2. Distribution of the total number of learning sessions that each student engaged in during the course.

The length of students' learning sessions is also being considered as a complementary metric. Observing the available data, the average length of newly-enrolled students' learning sequences ranged between 1.25 and 7.5, with the exception of those who did not interact with the platform at any point, and a single outlier whose average learning sequence length was 14.4.

2) TIME MANAGEMENT

In order to obtain a simple metric of how students manage their time in regards to the present course, the duration of their learning sequences was aggregated. This results in an estimation of the most basic notion of time management by students: whether the time they spend preparing this subject is sufficient or not.

A representation of how much time students spend in online learning sessions, grouped by weeks, is shown in Fig. 3. In this graph, week 0 corresponds to the first week of February 2021, the week right before the course started. It is included for completeness, since course resources were made available for students during this week, but it is ignored in terms of calculating the total average. On the other hand, week 16 is the last week of May 2021, and the last week of the course overall.

As it may have been expected, the weeks with highest activity values are those that precede exam weeks — since during the 2020/2021 academic year, exams always took place at the start of their respective weeks. Therefore, since there were exams scheduled in weeks 4, 6, 9, 11, 13 and 16, the weeks preceding these presented relatively higher online activity. It is also worth highlighting the sharp decrease in overall activity from week 7 onward, as the median weekly activity values are close to 0 for the second half of the course. This suggests that a significant portion of students give up on the subject after the second exam.

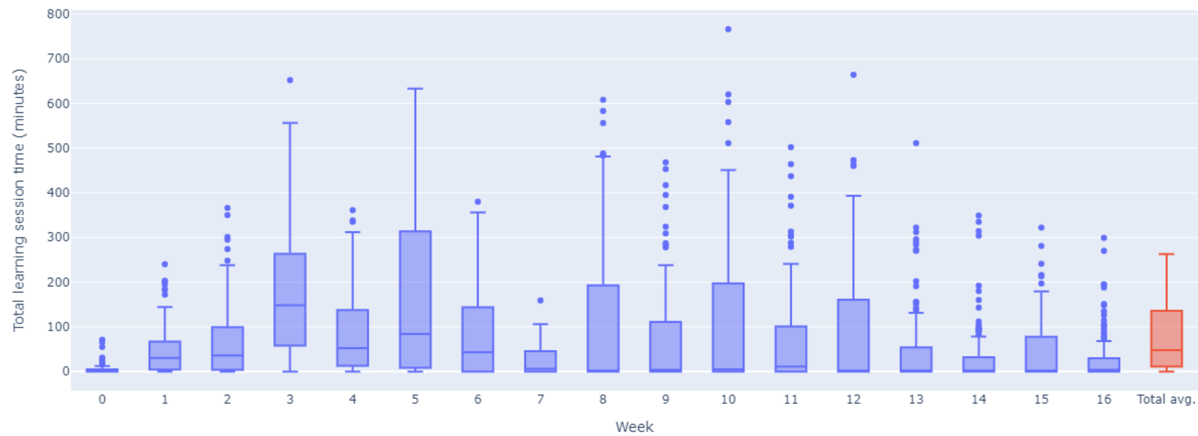


FIGURE 3. **Weekly activity by students in the course, as represented by the total time spent in online learning sessions. Notice that week 0 is the week right before the start of the course, during which course resources are made available for students.**

The primary metric that represents each student's time management is the weekly average time spent in online learning sessions, which is also included in Fig. 3. Differences observed in this regard can be significant, as the first quartile is located at 11 minutes per week, the median is 48 minutes and the third quartile reaches 135 minutes.

An additional metric was computed in order to account for regularity in the students' study schedules. As was seen in Section II-D, procrastination is one of the main issues that defines students' poor time management, which ultimately reflects in poor allocation of study time and leads to irregular schedules. Thus, the second time management metric that is being considered is the number of weeks in which the student spends at least the median average weekly time of 48 minutes in study sessions. Only 3 students surpassed this threshold in each of the 16 weeks the course lasted.

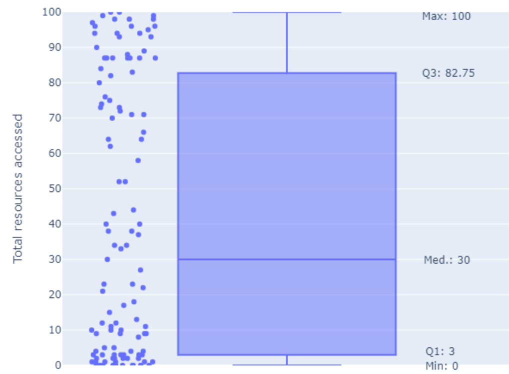


FIGURE 4. **Distribution of students regarding how many different resources available on Moodle they accessed at least once.**

3) RESOURCE MANAGEMENT

To assess resource management by the students, we tracked how many of the course material items that are made available to them via Moodle they actually end up using throughout the course.

As the target subject follows a flipped classroom system, most of the course material for theory classes are recorded video lectures: a total of 71 videos were uploaded to the learning management system for students to prepare the course with. Most videos were anywhere between 5 and 10 minutes long, with a few exceptions: some shorter videos only lasted for 2 to 3 minutes, and the longest ones were just under 15 minutes long.

On top of the videos, some complementary material was also made available for students:

- 13 collections of summaries for the provided videos, in the form of slide presentations.
- 16 documents with extra learning material, such as solved problems or data sheets.

These resources add up to 100 total items at the students' disposal related to theory contents. Fig. 4 summarizes the use of these resources by students. Only a single student accessed all available resources throughout the course. As in can be seen in the figure, the top 25% of students accessed 83% or more of the available resources at least once. On the other hand, the bottom 25% made negligible use of course resources, or no use at all. The median located at 30 implies that half of the students never used more than 70% of the course material items that were made available for them.

4) SELF-MONITORING AND SELF-ASSESSMENT

Metrics corresponding to the self-monitoring and self-assessment category are computed using two different sources. On the one hand, a total of 8 self-assessment tests were made available on Moodle for students to perform throughout the course. Out of these, only the first one was mandatory for students to be able to advance in the course. The main purpose of these tests was providing students a way to practice answering exam-like questions and problems.

There was no limit in the amount of times each student could attempt these tests, and the grade recorded in each one of them was the highest one obtained across all performed attempts.

Two aspects regarding the self-assessment tests were computed as metrics: the average grade obtained in them and the total amount of times that they were attempted.

The 115 newly-enrolled students in the course totaled up to 648 self-assessment test attempts. However, these attempts are not distributed evenly across the student population. 29 students never attempted any of the available tests. On the other hand, the highest number of attempts a single student made throughout the course was 28.

Fig. 5 summarizes the scores obtained by students in each of the self-assessment tests. It can be observed that there is great variance in the scores for every test, as not every student who performs poorly in a test decides to retake it in order to improve their score.

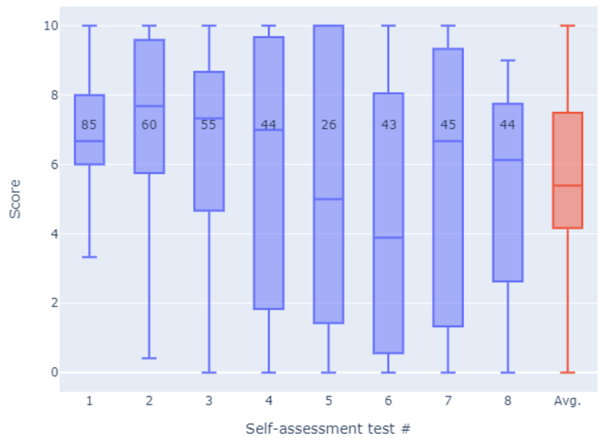


FIGURE 5. Distribution of student scores in the self-assessment tests. Labeled on each plot, number of students who attempted the corresponding self-assessment test at least once.

The other source that was taken into account for this category was student activity on the BeA platform. More specifically, we looked at how long it takes each student to check their exam assessments once they become available, under the hypothesis that it is beneficial for the student to do this as soon as possible so they can analyze their mistakes and contact the instructor for a double review if necessary.

Of course, this metric is only useful for students who take and review at least one exam during the course. Out of the 115 newly-enrolled students, 78 performed at least one exam, and 69 viewed their assessed exams at least once.

It was observed that students usually review their exams very soon after the assessments are published, typically within a couple hours. Cases in which students took more than 24 hours to see their exam assessments are very rare and, with the exception of a single outlier who first saw an exam assessment two months after it was posted, every student reviewed their exam results within a week of their publication.

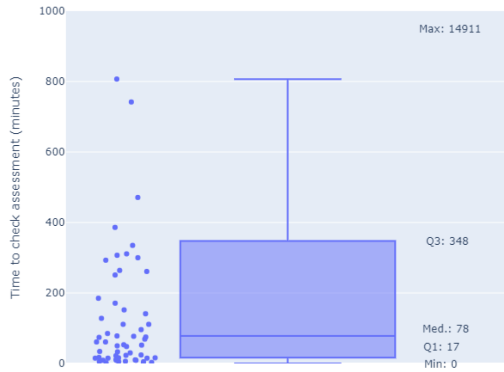


FIGURE 6. Distribution of the average time to check exam assessments on BeA (excluding 13 students with an average above 1000 minutes).

In order to obtain the metric, the times corresponding to each student were averaged. Fig. 6 shows how this metric is distributed.

5) MOTIVATION AND SELF-CONFIDENCE

In order to obtain a metric for the category of motivation and self-confidence, we relied on the students' own answers to the SRL questionnaires that were performed throughout the course, as outlined in Section III-B.

Appendix A at the end of this document lists the items related to motivation and self-confidence that were included in the questionnaires. The first 4 were among the initial questionnaire that was performed during an in-classroom session at the start of the course, and the rest were part of the subsequent smaller questionnaires that were made available on BeA.

Unfortunately, as was mentioned before, the reliability of this metric is limited due to insufficient student participation in these questionnaires. Most students only performed the initial questionnaire, which restricts the amount of answers related to motivation and self-confidence to just the 4 items that were included there.

The value of this metric was computed by averaging the scores corresponding to the answers of each student. As aforementioned, each questionnaire item provided a five-level Likert scale for stating the level of agreement or disagreement. It is important to note that by averaging the scores of all questions we are assuming that there is a constant difference between each consecutive pair of levels in the Likert scale, which constitutes a simplification of this model.

Computed averages for students who participated in these questionnaires ranged from 2 to 4.5 points.

B. CORRELATION WITH STUDENT PERFORMANCE

Once the SRL metrics have been defined, the next step is assessing their relevance by observing how they influence students' performance in the course, measured by the final grade obtained. Throughout this sections, data set features may be referred to by the number and acronym that was assigned to them in Table 1.

An important detail regarding the resulting data set is that there are a few features that may have undefined values for some students. More specifically:

- The average time to see an exam assessment (feature 6-SMA1) may be undefined if the student never reviews their assessments, or if they never attend an exam to begin with.
- The average score in self-assessment tests (feature 7-SMA2) may be undefined if the student does not perform any of these tests.
- The average score in ''motivation and self-confidence'' questionnaire items (feature 9-MSC) may be undefined if the student never performs any of the SRL questionnaires.

Out of the 115 newly-enrolled students, a total of 52 met at least one of the conditions listed above. When compared to the remaining 63, it was observed that:

- No students who had one or more undefined features passed the subject. 34 of them dropped out of the course altogether, and the top final score achieved by the other 22 was 1.8 points out of 10.
- Among the 63 students without observed undefined features, 28 passed the subject (44%), obtaining a final grade of at least 5 points out of 10.

It is clear that any of the condition that forces undefined data set values implies a significant risk of failure for the affected students. Considering these findings, only the latter group of students will be considered for the correlation tests in this section. While this reduced size is not ideal, keeping the students with undefined values would have compromised the balance of the data set.

Before proceeding with the correlation analysis, the following preparation procedure was executed:

- For feature 6-SMA1 specifically, a maximum value of 2000 was established, which was used for two outliers who greatly surpassed the values of the rest of the students.
- All features were subject to min-max normalization, fitting every value inside the range [0, 1], with the endpoints of this interval being the minimum and maximum value observed for the feature, respectively.

Having done this, Pearson correlation coefficients were computed between each pair of features in the data set. Fig. 7 depicts the result of this operation.

A few important insights can be extracted from these correlation values:

- Overall correlation values for many metrics are very high, which suggests that some are, indeed, redundant. This is particularly notable in the cases of metrics 1-LS1, 3-TM1, 4-TM2 and 5-RM, corresponding to the SRL categories ''learning strategies'', ''time management'' and ''resource management'', which all have Pearson correlation coefficients above 0.8 between each other.

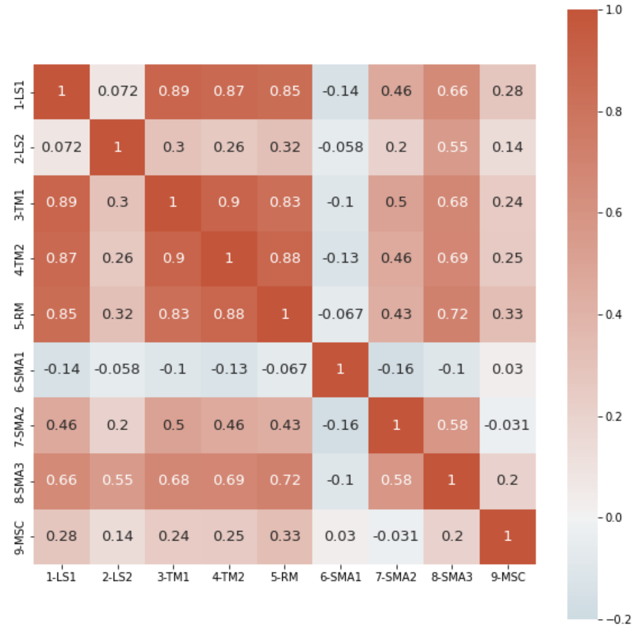


FIGURE 7. Pearson correlation matrix corresponding to the SRL metrics.

TABLE 2. Correlation between SRL metrics and final grade.

Metric #	Category	Corr. coefficient
1-LS1	Learning strategies	0.7290
2-LS2	Learning strategies	0.1571
3-TM1	Time management	0.6664
4-TM2	Time management	0.7149
5-RM	Resource management	0.7464
6-SMA1	Self-monitoring and self-assessment	−0.0666
7-SMA2	Self-monitoring and self-assessment	0.4345
8-SMA3	Self-monitoring and self-assessment	0.6847
9-MSC	Motivation and self-confidence	0.3600

- On the other end of the spectrum, feature 6 seems to not be correlated with any other feature in the data set. Metrics 2-LS2 and 9-MSC also show relatively low correlation coefficients. It is particularly interesting to see that features 1-LS1 and 2-LS2, corresponding to the total number of learning sequences and their average length respectively, are not correlated at all.

The existence of highly correlated metrics hints at the fact that students' self-regulation abilities in each of the categories which were proposed in this experiment are not independent from each other. While from the viewpoint of educational research it is still important to distinguish between different aspects of self-regulated learning, we can generally expect students to show proper self-regulation habits, or lack thereof, across the board.

A second correlation test was perform in order to find out the dependencies between the SRL metrics and the students' final grade in the course. Table 2 contains the outcome of this operation.

Metrics 1-LS1, 3-TM1, 4-TM2, 5-RM and 8-SMA3 are all significantly correlated with students' grades — and, as aforementioned, also between each other. On the other hand, grades are significantly less dependent on the remaining features, which presented lower coefficients overall in the Pearson correlation matrix.

C. EARLY WARNING POTENTIAL

As a final experiment, we will be assessing the usefulness of these SRL metrics as early indicators of student performance. This would be helpful in order to identify students at risk of failing the subject at an early stage in the course, potentially enabling interventions to improve struggling students' situations.

The metrics used for this experiment are the same as in the previous one, but they will be based exclusively on data generated during the weeks 1 through 4 of the course. This leads to some differences in the generation of metrics:

- The median average weekly time in study sessions is now 80 minutes instead of 48, and this is taken into account when computing metric 4-TM2.
- By week 4 in the course, only 31 resources were made available to students, which is considered for metric 5-RM.
- Only the first continuous assessment exam, performed in week 4, had been done. This limits the available data for generating metric 6-SMA1. For this experiment, undefined values in metric 6-SMA for students who performed the first exam were set to the maximum value of 2000, as some of them may first check their assessments after week 4.
- Only the first two self-assessment tests were available for students to perform, which affects metrics 7-SMA2 and 8-SMA3.
- For metric 9-MSC, only the initial SRL questionnaire had been performed at this point in the course.

With these limitations in mind, we can still make the same observation as in the last case regarding undefined values: 46 students presented undefined values in at least one of 7-SMA2 and 9-MSC, if they had not performed any self-assessment questionnaires or did not participate in the initial SRL questionnaire, and all of them failed or dropped out of the course, with the highest observed grade being 2.5 out of 10. This makes sense, as undefined values in these metrics signals that this group of students was very detached from the subject since the start. Additionally, there were 6 more students who did not present undefined values in 7-SMA2 or 9-MSC, but did not attend the first exam, dropping out of continuous assessment. Like last time, these records will be dropped from the data set, and we will be working with the 63 remaining students.

After normalizing all features like in the previous case, we built an ordinary least squares (OLS) linear regression model as a predictor of students' final grades based on the computed SRL metrics. The characteristics of this model are summarized in the first row of Table 3, and the parameters of the coefficients are listed in Table 4.

TABLE 3. Summary of OLS regression models.

Features	R^2	Adj. R^2	F-stat.	p-value	Res. std. err.
All	0.566	0.492	7.684	$4.04e-7$	0.2088
7-SMA2 + 9-MSC	0.505	0.489	30.67	$6.69e-10$	0.2095

TABLE 4. Regression coefficients for the complete OLS model.

Feature	Coef.	Std. Err.	t-stat.	p-value
(intercept)	−0.1717	0.114	−1.507	0.138
1-LS1	0.2080	0.309	0.672	0.504
2-LS2	−0.1113	0.213	−0.523	0.603
3-TM1	−0.3349	0.268	−1.249	0.217
4-TM2	0.0116	0.177	0.066	0.948
5-RM	0.2252	0.157	1.435	0.157
6-SMA1	0.0139	0.078	0.179	0.858
7-SMA2	0.5934	0.162	3.659	0.001
8-SMA3	0.0813	0.207	0.393	0.696
9-MSC	0.3518	0.156	2.254	0.028

However, as we already know, many of the features are either redundant or completely irrelevant. As a simplification, the two features with p-values smaller than 0.05, 7-SMA2 and 9-MSC, were selected to build a reduced OLS model including just these input variables. OLS regression results for this model can be found in the second row of Table 3, and the coefficient parameters are displayed in Table 5.

TABLE 5. Regression coefficients for the reduced OLS model.

Feature	Coef.	Std. Err.	t-stat.	p-value
(intercept)	−0.2129	0.084	−2.538	0.014
7-SMA2	0.6830	0.111	6.132	< 0.000
9-MSC	0.4194	0.138	3.035	0.04

Analyzing the results, it can be stated that the reduced model is of higher quality than the complete one, since it achieves virtually identical values for adjusted R^2 and residual standard error, while showing much better results in F-statistic and p-value. In either case, an R^2 value of around 0.5 and a residual standard error of 0.21 (it is important to remember than student grades are normalized between 0 and 1) show that the obtained regression models are fairly suboptimal, but they could be used in order to get a rough idea of which students are at risk of failing the course at very early stages.

V. CONCLUSION AND FUTURE WORK

Existing literature shows how relevant self-regulated learning features can be estimated based on available data by applying analysis techniques. We have attempted to apply some of these techniques to create simple student profiles, including features corresponding to 5 different SRL categories. This was an initial attempt, as new metrics could be obtained from the available data, but several lessons were learned from the performed work.

First, the difficulty to get significant amounts of reliable data. Most of the proposed metrics are obtained from LMS log data, mostly related to accesses of the students to the different available resources, but this does not provide significant information about the intensity of the students' effort. Basically, a student can access many videos or spend a lot of time in the platform, but this does not mean that they really pay attention or perform real cognitive effort to learn. We have also attempted to collect data from optional questionnaires including SRL items, but low student participation was an important issue. Questionnaires could be made mandatory, but in that case the reliability of the data would decrease as students could answer randomly or insincerely. In addition, we have collected data from the BeA e-assessment platform that provides data about student participation in exam review processes. Nevertheless, the use of this data to calculate metric 6-SMA1 provides the worst correlation with grades out of all the indicators (see Table 2). There is a large scarcity of relevant data about the students learning activities and processes.

Second, first year students have specific behaviors that need to be taken into account. As we have seen, around 30% of the students do not display any kind of participation in the subject from the start. Moreover, there are other features that are also indicative of dropout. In the paper we have identified three related features: 6-SMA1, related to the review of exams in BeA; 7-SMA1, related to the performance of self-assessment tests; and 9-MSC, related to the participation in optional SRL questionnaires. A total of 56 students did not engage in at least one of these activities, and none of them achieved a final score greater than 1.8 points out of 10. Therefore, these conditions are clear indicators of dropout or course failure. This makes sense, as these students have not shown any interest in the performance of complementary, yet optional activities.

Third, despite the fact that predictive potential of the SRL metrics has been observed to be limited, it could be enough for the development of early warning systems (EWS) that can detect students at risk of failing. Many of the computed metrics ended up being redundant, as demonstrated by the cross-correlation matrix in Fig. 7. However, they could still be used to build two OLS regression models of similar performance, one including all features and a second one excluding redundant and irrelevant metrics. In order to simulate an early state of the course, the models considered only data corresponding to the first few weeks. These models were able to explain about 50% of the variation in students' final grades, with an average standard error of 2 points if we consider grades to be in a 0 to 10 scale. Interestingly enough, feature 9-MSC, which was calculated using very limited information from SRL questionnaires, ended up being useful for building the reduced OLS regression models. This suggests that data obtained via self-reported instruments can still be useful as a complement to observational data.

The main aspect that differentiates this work from others in recent literature regarding SRL and learning analytics is the aim to simultaneously cover multiple dimensions of students' self-regulation, making use of both observational and self-reported data but prioritizing the former whenever possible. This endeavor required the acquisition, transformation and interpretation of multiple types of data available on the online platforms Moodle and BeA. As seen in Section II, most of the existing studies in the literature focus on a specific SRL aspect. While this work does not intend to provide as thorough of an analysis of each SRL category, it does show how student metrics obtained from course data can be interpreted through the lens of self-regulation in a broad sense.

Following the line of work presented in this paper, there is still room for improvements and future development. The predictive models can be improved through further training and testing using data from courses in the following academic years. Additionally, while OLS regression provides a simple framework for assessing feature relevance, other types of predictive models could be tested in order to find one that performs better. In many cases, it could be enough to label students depending on whether they are expected to pass or fail the course. In this regard, a classification or clustering algorithm may be better suited for the job.

Finally, the reliability of self-reported data could be improved by fostering student participation. While we still consider making mandatory SRL questionnaires to be counterproductive, performing activities in the classroom such as self-assessment quizzes mixed with some SRL questions could be a way to collect self-reported information in an engaging way.

APPENDIX A
MOTIVATION AND SELF-CONFIDENCE QUESTIONNAIRE ITEMS

These are the Likert-type questionnaire items related to motivation and self-confidence which were provided to students throughout the course, originally asked in Spanish. For questions labeled with (*), higher levels of agreement represent negative SRL attitudes, and their scores were reversed before computing the corresponding metric.

1) I am certain that I can learn even the toughest parts of the course.
2) I am able to make an effort to focus when I start getting distracted.
3) I consider really important to learn the contents of the course.
4) Whenever I do an exam, I always feel very nervous. (*)
5) I work hard to obtain good grades, even if I do not like the subject.
6) I think any topic can be interesting once I study it in depth.
7) When doing an exam, my mind focuses on the parts I do not know how to answer. (*)
8) I try to focus only in the easy parts of the subjects. (*)

9) When doing an exam, I worry about performing worse than my classmates. (*)
10) My main motivation for performing well is to please my family, friends or other people. (*)
11) I am certain that I can learn at least the basic concepts of all subjects.

REFERENCES

[1] E. M. Sosu and P. Pheunpha, "Trajectory of university dropout: Investigating the cumulative effect of academic vulnerability and proximity to family support," *Frontiers Educ.*, vol. 4, Feb. 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/feduc.2019.00006/full

[2] G. Siemens. (Jul. 2010). *1st International Conference on Learning Analytics and Knowledge 2011 | Connecting the Technical, Pedagogical, and Social Dimensions of Learning Analytics*. [Online]. Available: https://tekri.athabascau.ca/analytics/

[3] G. J. Hwang, D. Spikol, and K. C. Li, "Guest editorial: Trends and research issues of learning analytics and educational big data," *J. Educ. Technol. Soc.*, vol. 21, no. 2, pp. 134–136, 2018. [Online]. Available: http://www.jstor.org/stable/26388386

[4] S. Charleer, A. V. Moere, J. Klerkx, K. Verbert, and T. De Laet, "Learning analytics dashboards to support adviser-student dialogue," *IEEE Trans. Learn. Technol.*, vol. 11, no. 3, pp. 389–399, Jul. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/7959628/

[5] N. Manouselis, H. Drachsler, K. Verbert, and E. Duval, *Recommender Systems for Learning* (SpringerBriefs in Electrical and Computer Engineering). New York, NY, USA: Springer, 2013. [Online]. Available: http://link.springer.com/10.1007/978-1-4614-4361-2

[6] M. Liz-Domínguez, M. Caeiro-Rodríguez, M. Llamas-Nistal, and F. A. Mikic-Fonte, "Systematic literature review of predictive analysis tools in higher education," *Appl. Sci.*, vol. 9, no. 24, p. 5569, Dec. 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/24/5569

[7] B. J. Zimmerman, "Becoming a self-regulated learner: Which are the key subprocesses?" *Contemp. Educ. Psychol.*, vol. 11, no. 4, pp. 307–313, Oct. 1986. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0361476X86900275

[8] P. R. Pintrich, "Understanding self-regulated learning," *New Directions Teaching Learn.*, vol. 1995, no. 63, pp. 3–12, 1995, doi: 10.1002/tl.37219956304.

[9] B. J. Zimmerman, "A social cognitive view of self-regulated academic learning," *J. Educ. Psychol.*, vol. 81, no. 3, pp. 329–339, Sep. 1989. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-0663.81.3.329

[10] E. Panadero, "A review of self-regulated learning: Six models and four directions for research," *Frontiers Psychol.*, vol. 8, p. 422, Apr. 2017. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/28503157

[11] S. G. Paris and A. H. Paris, "Classroom applications of research on self-regulated learning," *Educ. Psychol.*, vol. 36, no. 2, pp. 89–101, Jun. 2001. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1207/S15326985EP3602_4

[12] E. C. K. Cheng, "The role of self-regulated learning in enhancing learning performance," *Int. J. Res. Rev.*, vol. 6, no. 1, pp. 1–16, 2011. [Online]. Available: https://tijrr.webs.com/A1_V6.1_TIJRR.pdf

[13] J. Li, H. Ye, Y. Tang, Z. Zhou, and X. Hu, "What are the effects of self-regulation phases and strategies for Chinese students? A meta-analysis of two decades research of the association between self-regulation and academic performance," *Frontiers Psychol.*, vol. 9, p. 2434, Dec. 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2018.02434/full

[14] L. Nilson, *Creating Self-Regulated Learners: Strategies to Strengthen Students' Self-Awareness and Learning Skills*. Sterling, VA, USA: Stylus Publishing, LLC, 2013. [Online]. Available: https://books.google.es/books?id=ZeBaAQAAQBAJ

[15] A. Roth, S. Ogrin, and B. Schmitz, "Assessing self-regulated learning in higher education: A systematic literature review of self-report instruments," *Educ. Assessment, Eval. Accountability*, vol. 28, no. 3, pp. 225–250, Aug. 2016. [Online]. Available: http://link.springer.com/10.1007/s11092-015-9229-2

[16] J. Jovanović, D. Gašević, S. Dawson, A. Pardo, and N. Mirriahi, "Learning analytics to unveil learning strategies in a flipped classroom," *Internet Higher Educ.*, vol. 33, pp. 74–85, Apr. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1096751617300684

[17] E. Panadero, J. Klug, and S. Järvelä, "Third wave of measurement in the self-regulated learning field: When measurement and intervention come hand in hand," *Scandin. J. Educ. Res.*, vol. 60, no. 6, pp. 723–735, Nov. 2016. [Online]. Available: https://www.tandfonline.comdoifull/10.1080/00313831.2015.1066436

[18] B. J. Zimmerman and M. M. Pons, "Development of a structured interview for assessing student use of self-regulated learning strategies," *Amer. Educ. Res. J.*, vol. 23, no. 4, p. 614, 1986. [Online]. Available: https://journals.sagepub.com/doi/10.3102/00028312023004614

[19] F. Perels, T. Gürtler, and B. Schmitz, "Training of self-regulatory and problem-solving competence," *Learn. Instruct.*, vol. 15, no. 2, pp. 123–139, Apr. 2005. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S095947520500023X

[20] S. Fabriz, C. Dignath-van Ewijk, G. Poarch, and G. Büttner, "Fostering self-monitoring of university students by means of a standardized learning journal—A longitudinal study with process analyses," *Eur. J. Psychol. Educ.*, vol. 29, no. 2, pp. 239–255, Jun. 2014. [Online]. Available: https://link.springer.com/article/10.1007/s10212-013-0196-z

[21] A. Elsayed, M. Caeiro Rodriguez, F. Mikic Fonte, and M. L. Nistal, "Research in learning analytics and educational data mining to measure self-regulated learning: A systematic review," in *Proc. 18th World Conf. Mobile Contextual Learn.*, Delft, The Netherlands, 2019, pp. 46–53. [Online]. Available: https://www.learntechlib.org/p/210600/

[22] R. Garcia, K. Falkner, and R. Vivian, "Systematic literature review: Self-regulated learning strategies using e-learning tools for computer science," *Comput. Educ.*, vol. 123, pp. 150–163, Aug. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0360131518301064

[23] P. R. Pintrich and E. V. de Groot, "Motivational and self-regulated learning components of classroom academic performance," *J. Educ. Psychol.*, vol. 82, no. 1, pp. 33–40, 1990. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-0663.82.1.33

[24] D. Gasevic, J. Jovanovic, A. Pardo, and S. Dawson, "Detecting learning strategies with analytics: Links with self-reported measures and academic performance," *J. Learn. Anal.*, vol. 4, no. 2, pp. 113–128, Jul. 2017. [Online]. Available: https://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/5085

[25] J. Biggs, D. Kember, and D. Y. P. Leung, "The revised two-factor study process questionnaire: R-SPQ-2F," *Brit. J. Educ. Psychol.*, vol. 71, no. 1, pp. 133–149, Mar. 2001, doi: 10.1348/000709901158433.

[26] W. Matcha, D. Gašević, N. A. Uzir, J. Jovanović, and A. Pardo, "Analytics of learning strategies: Associations with academic performance and feedback," in *Proc. 9th Int. Conf. Learn. Anal. Knowl.*, Tempe, AZ, USA, Mar. 2019, pp. 461–470, doi: 10.1145/3303772.3303787.

[27] A. Pardo, J. Jovanovic, S. Dawson, D. Gašević, and N. Mirriahi, "Using learning analytics to scale the provision of personalised feedback," *Brit. J. Educ. Technol.*, vol. 50, no. 1, pp. 128–138, Jan. 2019, doi: 10.1111/bjet.12592.

[28] L. U. Gadella, I. Estevez-Ayres, J. A. Fisteus, and C. Delgado-Kloos, "Application of learning analytics to study the accuracy of self-reported working patterns in self-regulated learning questionnaires," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Apr. 2020, pp. 1201–1205.

[29] M. K. Hartwig and J. Dunlosky, "Study strategies of college students: Are self-testing and scheduling related to achievement?" *Psychonomic Bull. Rev.*, vol. 19, no. 1, pp. 126–134, Feb. 2012, doi: 10.3758/s13423-011-0181-y.

[30] P. Steel, "The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure," *Psychol. Bull.*, vol. 133, no. 1, pp. 65–94, 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17201571

[31] C. A. Wolters, S. Won, and M. Hussain, "Examining the relations of time management and procrastination within a model of self-regulated learning," *Metacognition Learn.*, vol. 12, no. 3, pp. 381–399, Dec. 2017. [Online]. Available: http://link.springer.com/10.1007/s11409-017-9174-1

[32] C. J. Asarta and J. R. Schmidt, "Access patterns of online materials in a blended course," *Decis. Sci. J. Innov. Educ.*, vol. 11, no. 1, pp. 107–123, Jan. 2013, doi: 10.1111/j.1540-4609.2012.00366.x.

[33] P. Steel, "Arousal, avoidant and decisional procrastinators: Do they exist?" *Personality Individual Differences*, vol. 48, no. 8, pp. 926–934, Jun. 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0191886910000930

[34] Z. Wang, C. Zhu, Z. Ying, Y. Zhang, B. Wang, X. Jin, and H. Yang, "Design and implementation of early warning system based on educational big data," in *Proc. 5th Int. Conf. Syst. Informat. (ICSAI)*, Nov. 2018, pp. 549–553. [Online]. Available: https://ieeexplore.ieee.org/document/8599357/

[35] J. J. Díaz-Lázaro, I. M. Solano Fernández, and M. D. M. Sánchez-Vera, "Social learning analytics in higher Education. An experience at the primary education stage," *J. New Approaches Educ. Res.*, vol. 6, no. 2, pp. 119–126, Jul. 2017. [Online]. Available: https://naerjournal.ua.es/article/view/v6n2-5

[36] J. L. Santos, K. Verbert, and E. Duval, "Empowering students to reflect on their activity with StepUp!: Two case studies with engineering students," in *Proc. ARTEL12: 2nd Workshop Awareness Reflection*, Saarbrücken, Germany, vol. 931, Oct. 2012, pp. 73–86. [Online]. Available: https://lirias.kuleuven.be/handle/123456789/359546

[37] P. R. Pintrich, D. Smith, T. Garcia, and W. McKeachie, "A manual for the use of the motivated strategies for learning questionnaire (MSLQ)," Nat. Center Res. Improve Postsecondary Teach. Learn., Ann Arbor, MI, USA, Tech. Rep. NCRIPTAL-91-B-004, 1991. [Online]. Available: https://eric.ed.gov/?id=ED338122

[38] P. R. Pintrich, "The role of motivation in promoting and sustaining self-regulated learning," *Int. J. Educ. Res.*, vol. 31, no. 6, pp. 459–470, 1999. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0883035599000154

[39] P. R. Pintrich, "A conceptual framework for assessing motivation and self-regulated learning in college students," *Educ. Psychol. Rev.*, vol. 16, no. 4, pp. 385–407, Dec. 2004, doi: 10.1007/s10648-004-0006-x.

[40] S. Lonn, S. J. Aguilar, and S. D. Teasley, "Investigating Student motivation in the context of a learning analytics intervention during a summer bridge program," *Comput. Hum. Behav.*, vol. 47, pp. 90–97, Jun. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0747563214003793

[41] C. Mega, L. Ronconi, and R. De Beni, "What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic achievement," *J. Educ. Psychol.*, vol. 106, no. 1, pp. 121–131, Feb. 2014. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0033546

[42] M. Llamas-Nistal, F. Mikic-Fonte, M. Caeiro-Rodríguez, and M. Liz-Domínguez, "Supporting intensive continuous assessment with BeA in a flipped classroom experience," *IEEE Access*, vol. 7, pp. 150022–150036, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8865067/

[43] M. Llamas-Nistal, M. J. Fernández-Iglesias, J. González-Tato, and F. A. Mikic-Fonte, "Blended e-assessment: Migrating classical exams to the digital world," *Comput. Educ.*, vol. 62, pp. 72–87, Mar. 2013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0360131512002497

[44] M. Liz-Dominguez, "Contributions to learning analytics focused on assessment and self-regulated learning," in *Proc. Int. Symp. Comput. Educ. (SIIE)*, Malaga, Spain, Sep. 2021, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9583644/

[45] I. Khan and A. Pardo, "Data2U: Scalable real time Student feedback in active learning environments," in *Proc. 6th Int. Conf. Learn. Anal. Knowl. (LAK)*, New York, NY, USA, 2016, pp. 249–253, doi: 10.1145/2883851.2883911.

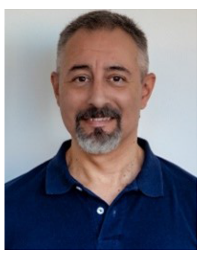
[46] J. Biggs, "The study process questionnaire (SPQ): Manual," Austral. Council Educ. Res., Hawthorn, QLD, Australia, Tech. Rep., 1987. [Online]. Available: https://eric.ed.gov/?id=ED308200

[47] M. Liz-Domínguez, M. Caeiro-Rodríguez, M. Llamas-Nistal, and F. Mikic-Fonte, "Monitoring students' self-regulation as a basis for an early warning system," in *Learning Analytics Summer Institute Spain 2021: Learning Analytics in Times of COVID-19: Opportunity from Crisis*, vol. 3029. Barcelona, Spain: CEUR-WS, 2021, pp. 38–51. [Online]. Available: http://ceur-ws.org/Vol-3029/paper04.pdf

MARTÍN LLAMAS-NISTAL (Senior Member, IEEE) received the Eng. and Ph.D. degrees in telecommunication from the Polytechnic University of Madrid, Spain, in 1986 and 1994, respectively. He has been a Faculty Member with the Higher Technical School of Telecommunication Engineers, University of Vigo, Spain, since March 1987. He is an author or coauthor of more than 300 papers in peer-reviewed international refereed journals and conference proceedings. He has directed several national and international research projects in telematics and technology enhanced learning fields. He has received several awards from the W3C and IEEE. He is very involved in the activities of the IEEE Education Society, serving in different positions: has been a member of the Board of Governors, since 2008, and the Strategic Planning Committee, since 2009; and was the Chair of the Publications Committee, from 2010 to 2018, the Vice-President for Publications, from 2011 to 2018, and the Vice-President for Member and Geographic Activity, from 2019 to 2020. Currently, he is a President Elect, from 2021 to 2022. Since April 2020, he has been the Director of atlanTTic Research Center for Telecommunication Technologies, University of Vigo.

MANUEL CAEIRO-RODRÍGUEZ (Senior Member, IEEE) received the Ph.D. degree in information and communication technologies about educational modeling languages, in 2007. He is an Associate Professor with the Department of Telematic Engineering, University of Vigo. He teaches computer programming, software engineering, and computer architectures with the School of Telecommunications Engineering. His research interests include e-learning technologies and standards, open educational resources, educational innovation, CSCL, and process-based systems. He has performed research stays at the University of Coimbra, Portugal; IRISA, Rennes, France; MTA-ZSTAKI, Budapest, Hungary; Kumamoto University, Japan; and ISEP, Porto, Portugal. He had received several awards from the W3C, NAE CASEE new faculty fellows, and the IEEE Spanish Chapter of the Education Society to the best Ph.D., in 2007. Currently, he is focused on learning analytics. He has served as the President for the Spanish Chapter of the IEEE Education Society, from 2018 to 2019.

MARTÍN LIZ-DOMÍNGUEZ (Student Member, IEEE) received the B.S. and M.S. degrees in telecommunication engineering from the University of Vigo, Spain, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree in information and communication technologies. He is also a member of the atlanTTic Research Center, Group of Telematics Systems Engineering (GIST), University of Vigo. In 2019, he received a scholarship from the Galician Regional Government to support his research activities during the Ph.D. stage. His current research interests include learning analytics and technology-enhanced learning.



FERNANDO A. MIKIC-FONTE (Member, IEEE) received the M.Sc. and Ph.D. degrees in telecommunication engineering from the University of Vigo, in 2001 and 2010, respectively. He is currently an Associate Professor with the Engineering School of Telecommunication, University of Vigo, where he is also with the Group of Telematics Systems Engineering (GIST), Telematics Engineering Department. His current research interests, carried out at the atlanTTic Research Center for Telecommunication Technologies, include e-learning, blended learning, self-regulated learning, serious games, and educational simulators.