# Using Process Mining to Analyze Time-Distribution of Self-Assessment and Formative Assessment Exercises on an Online Learning Tool

César Domínguez, Francisco J. García-Izquierdo, Arturo Jaime, Beatriz Pérez, Ángel L. Rubio, and María A. Zapata

***Abstract*—The study of the relationships between self-regulated learning and formative assessment is an active line of research in the educational community. A recent review of the literature highlights that the study of these connections has been mainly unidirectional, focusing on how formative assessment helps students to self-regulate their learning, being much less explored the effect of self-regulated learning strategies on formative assessment. In this context, analyzing automatically captured students' activities within online learning tools can provide us further insights on the interactions between these two topics. More specifically, this article examines the activity traces of 382 students who used an online tool to learn a programming language. The tool incorporates review exercises for promoting self-assessment (an important self-regulated learning strategy). Furthermore, the tool is used in supervised laboratories where students receive formative assessment. This study uses process mining techniques to analyze the temporal component of student behavior in both types of activities, their interaction, and how self-assessment relates to formative assessment. Some key lessons are learned: activities promoting self-assessment significantly improved students' involvement in formative assessment activities; increasing self-assessment cannot compensate for a lack of effort in formative assessment. We also underline that, to the best of our knowledge, to date no research has used process mining to consider the time component in the analysis of the relationships between formative assessment and self-assessment.**

***Index Terms*—Educational data mining, formative assessment, learning analytics, online learning, process mining, self-assessment technologies, self-regulated learning.**

## I. Introduction

SELF-REGULATED learning (SRL) and formative assessment (FA) are topics of great interest in the educational community [1]–[14]. In particular, the study of the relationships between SRL and FA is an active line of research [11], [12]. For example, a recent literature review [9] shows that instructors who use FA focus primarily on the effect of FA feedback on activities promoting SRL strategies. The results of this review suggest that FA, being a form of feedback, provides inputs to the SRL cycle, specifically assisting in the cognitive assessment of learning and in supporting goal and plan setting. However, the aforementioned review [9] states that the effect of specific SRL strategies on FA is much less explored, and highlights the potential benefits of analyzing how SRL strategies influence the way in which FA is used and its results. In this regard, when students strive to improve their SRL, the involved skills may boost their interest and motivation in FA activities and the benefits they reap from it [13].

This paper focuses on a specific SRL strategy, that of self-assessment (SA). This is the first strategy included in the list of strategies presented in [8], which is described as "student-initiated evaluations of the quality or progress of their work". Our work aims to contribute by considering the time component to examine patterns of student engagement with both SA and FA activities and the interaction on how SA relates to FA. The exploration of these patterns is especially relevant when online learning support tools [15] are used, as these tools provide lower levels of guidance than face-to-face classes [7].

In particular, our case study is based on the use of an online tool called *DELFOS* (that stands for Databases Exercises Laboratory For Online Study) in a course of a software engineering degree program. This tool has been used for several years with a two-fold purpose. The first is to perform exercises in FA supervised laboratory sessions. The second is to provide students with a set of review exercises that enable and promote SA. To this end, we study the moments throughout the course when certain events linked to students' interactions with the tool occur. In our work, the study of the timing of activities is addressed by means of Process Mining (PM) [16]. PM has been recognized as a powerful approach for identifying event

Manuscript received xxxx, 2020; revised xxxx, 2021 and xxxx, 2021; accepted xxxx, 2021. Date of publication xxxx, 2021; date of current version xxxx, 2021. This work was supported in part by MCIN/AEI/10.13039/501100011033, Spain, and "NextGenerationEU"/PRTR, European Union, under Grant PDC2021-121128-I00. (*Corresponding author: César Domínguez.*)

C. Domínguez, F. J. García-Izquierdo, A. Jaime, B. Pérez, and A. L. Rubio are with Departamento de Matemáticas y Computación, Universidad de La Rioja, 26006 Logroño, Spain (e-mail: cesar.dominguez@unirioja.es, francisco.garcia@unirioja.es, arturo.jaime@unirioja.es, beatriz.perez@unirioja.es, arubio@unirioja.es).

M. A. Zapata is with Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain (e-mail: mazapata@unizar.es).

patterns of SRL [17]. The analysis of traces of tool use has been admitted as a method of interest for a better understanding of the learning process [3], since what or how many real-time actions students made through the tool can be used for this purpose [18]. Nevertheless, to the best of our knowledge, to date no research has used PM to consider the time component in the analysis of the relationships between FA and SA.

## II. Background

### A. Review Exercises Promoting Self-Assessment

Pintrich defined *self-regulated learning* (SRL) as "an active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, motivation, and behavior, guided and constrained by their goals and the contextual features in the environment" [14]. SRL boasts an active line of research; and numerous models of SRL have been proposed (see, for instance, Zimmerman [19], Pintrich [14], or Winne and Hadwin [20] models). Most of these models share some characteristics [1]: a forethought or preparatory phase that includes task analysis, planning, and goal setting; a performance phase that involves the use of learning strategies and monitoring activities; and a reflection phrase, which includes evaluation of learning outcomes and reaction. SRL involves the use of specific strategies to achieve academic goals on the basis of self-efficacy perceptions [14]. SRL strategies are actions and processes directed at acquiring information or skills that involve agency, purpose, and instrumentality perceptions by learners [19]. In [8] fourteen categories of strategies are devised. The first of them is self-evaluation, described as "student-initiated evaluations of the quality or progress of their work". A recent bibliographic review of these types of strategies, applied in our field of interest (Computer Science), highlights self-evaluation over all other types of SRL activities [7]. In this field, self-evaluation usage is described as "student-initiated self-assessments to validate programming exercises". This type of activity can include optional exercises for self-assessment done through online tools. In this article, we call these optional exercises *review exercises promoting self-assessment*.

Goal orientation is an aspect that also appears in most SRL models (see, for instance, [14], [19]). Goal orientation reflects the motivation and purpose for doing a task, and can influence the different process of self-regulation and use of SRL strategies [14]. There exists different classification of goal orientations based on mastery or performance orientations [14]. An individual with a mastery goal orientation focuses on the development of knowledge, skills, and competences, and can use deeper processing strategies, whereas a work avoidant can use more surface ones [12], [21]. An individual with a performance orientation focuses on demonstrating competence by trying to outperform peers on academic tasks [14].

Some SRL models (such as Pintrich [14] or Winne and Hadwin [20] models) include the context as an area for student regulation. This involves individual perceptions of the task and the context. The student tries to actively monitor and regulate the context as well as to adapt to it [14]. Indeed, self-regulation activities can mediate between students personal and contextual characteristics and their academic performance. In this way, perception of the environment can be central in order to understand the student academic performance and achievement, and indeed some authors [14] claim that it is needed research on how different features of the context could shape, facilitate or constraint SRL.

### B. Formative Assessment Exercises

*Formative assessment* (FA) is defined by the United States Council of Chief State School Officers as "a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievements of intended instructional outcomes" [10]. This information can be used by teachers to shape instruction to meet students' needs, and by students to better understand and advance their learning [22]. FA achieves learning improvements through feedback received by students and teachers during the instruction process [23]. This type of assessment has some features such as learning progressions that articulate subgoals of the ultimate goal (which should be clearly identified and communicated to students), evidence of learning elicited during instruction, descriptive feedback of the instructional outcomes, self and peer assessments, and collaboration between teachers and students [10].

FA can take the format of a wide range of instruments or tasks with the purpose of assisting, strengthening, or shaping student learning during the educational process [22]. It may be formal or informal, spontaneous or planned, on an individual or group basis, oral or written, graded or ungraded [22]. Let us note an example: laboratory exercises built along the subgoals of a course, included in online tools and designed to be solved individually, but in a collaborative classroom where the instructor, the classmates, and/or the tool can offer FA feedback.

A controversial issue is whether deliverables produced by students in FA activities could be graded. Since the goal is to improve learning, requiring students to submit a deliverable which may be graded could distract them, and ultimately detract from their learning progress [24]. Meanwhile, some authors argue that "low stakes" assessments encourage students to pay more attention [23], [25], [26]. Nevertheless, a problematic situation arises when tasks are graded: some students may decide to copy other students' work, thus losing the opportunity to learn by completing the tasks on their own [23]. When the grade assigned to these activities is of minor importance and is clearly designed to motivate and recognize students' effort, students copying work may be considered "instructional disobedience", i.e., the student is engaging with the learning environment in a way different from the original intention [13].

### C. Relationships Between SRL and FA

The study of the interactions between SRL and FA is an active line of research [9], [11], [12]. A recently published literature review [9] discussed keystone publications on the relationship between FA and SRL. This review highlights that the study of these connections between assessment and SRL began in the early 1990s and have been primarily unidirectional: FA scholars have expanded their scope of assessment to include SRL [9]. FA helps students to conceptualize what and how they are learning. These processes activate students' cognitive and motivational capacities, focus students on their learning goals, and provide feedback and strategies they can use to help them to self-regulate their learning [9]. Practical methods that have been used to study the link between SRL and FA include self-assessment and formative feedback, for instance. Self-assessment involves reflecting on one's work which can serve to regulate one's own learning more effectively [27], [28]. Formative feedback provides information on how successfully a task has been done and on how and what can be done to perform it more effectively. This feedback equips students to self-regulate their learning [11], [12].

It is widely recognized by scholars in both the fields of FA and SRL that the interests of both strategies are reciprocal and that their combination is beneficial [29]. It seems clear that SRL skills are needed to take full advantage of student involvement in FA [9]. In this vein, researchers call for increased investment to explore this reciprocal relationship. It also seems logical to ask questions about how SRL strategies, as SA (or their absence), are related to the uses and results of FA [9].

### *D. Online Tools Supporting SRL and FA*

Online tools have the capability to support SRL and FA. On the one hand, online tools can be used to enhance SRL [3]. These tools provide learners with the flexibility and accessibility to be able to study anywhere, at any time. But this flexibility requires that students structure their own learning and manage their time efficiently. Broadbent *et al*. [3] described three types of these tools. Firstly, some educational technologies (e.g., online training or mobile-based apps) can provide direct instruction on how to acquire and develop SRL. Here the technology is used for the primary purpose of helping students learn how to regulate their learning [3]. Secondly, some tools (such as nStudy [30] or MetaTutor [31]) are embedded within online learning environments to support and promote SRL while students are completing course-specific content learning tasks. Finally, non-SRL tools can be used for SRL purposes. This is an alternative approach that uses already available tools to support and develop SRL (or build up on top of these tools). These types of tools can help learners to monitor their learning process while performing the activities, through SRL strategies that have proven to be beneficial, such as self-assessment, goal setting, planning or organization [30], [32]. Moreover, researchers seeking to produce a student reaction through a tool can use the tool to try to measure such reaction (this is called "third wave of SRL measurement" [4]). Specific studies on how SRL strategies affect learners' behavior in the use of Massive Open Online Courses [5], [33] and some literature reviews of online tools that support SRL have been published recently [3], [34].

On the other hand, technology can help instructors to create engaging FA. For example, an effective tool may provide immediate or personalized feedback to students [35]. Existing online learning tools enable FA to be offered in different formats: questionnaires, supervised or unsupervised [23]; peer reviews [36]; self-assessments [37]; or automatic evaluation and complementary feedback, in programming problems [38], [39]. Indeed, other recent literature reviews also show that computer-based tools can effectively enhance FA [40], [41]. Laboratory sessions that include programming exercises designed for FA are another example.

Clearly an online tool can be used for both FA and SRL purposes [9], [41]. For example, in [42], an FA tool is used to promote SRL skills. In addition, [43] presents an online game-based FA module that influences students' SRL-motivational strategies.

### *E. Educational Process Mining*

Both Broadbent *et al*. in [3] and Cicchinelli *et al*. in [6] included a definition of *learning analytics* 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". This discipline has great potential to improve the technologies that support and promote SRL, and several challenges have been identified, such as the need for data analysis methods capable of identifying patterns related to SRL [3]. In this line, in [44] learning analytics techniques are used to detect learning strategies. In particular, the authors present a comparison of three analytic approaches showing how different approaches influence results. More specifically, [44] identifies *process-oriented data analysis approaches* (which emphasize the timing of events), *sequence analysis* (which combines sequence techniques with unsupervised learning to detect learning strategies from trace data), and *network approaches* (where learning strategies are identified through networks based on the co-occurrence of learning states or actions). In order to focus specifically on the distribution of students' actions over time, it is logical to opt for a process-oriented data analysis.

A more specific field that falls within the scope of learning analytics is *educational data mining* (EDM). This field is concerned with the use of different data analysis techniques, backed by automated support, for the study of educational topics. It is a field still in development [45] that has been recognized to "play a signal role advancing research on motivation, metacognition, and self-regulated learning" [18]. A subset of the techniques used in EDM is process mining, whose use has given rise to the term *educational process mining* (EPM) [46], [47]. The most common uses of EPM include understanding educational processes, giving feedback for students and teachers, detecting challenges for students, or improving the management of learning objects [46]. At present, there is no well-established methodological approach for how to apply these techniques. Usually, data analyzed with EPM come from educational platforms, are collected in an event log, cleaned, and organized to eventually perform a process discovery [16]. Generally, the models obtained are complex and it is difficult to reach clear conclusions [33]. For this reason, clustering techniques (such as trace clustering [48]) are used to pre-process and segment the event log.

EPM has been recognized as "a powerful tool for identifying patterns of SRL" [17]. Some studies see SRL as a process and use EPM tools to analyze recorded data from student interactions with the online learning tool in order to identify different learning strategies [33], [49]. An important variable in this type of analysis is the moment during the course when activities occur, for example, SRL activities, or sessions during which students receive FA. For instance, Cicchinelli *et al.* [6] studied (albeit without using EPM) traces of interactions with a Learning Management System throughout a course, in order to analyze self-regulated behavior on the platform. As another example, a process-oriented analysis on a massive open online course (MOOC) is presented in [33]. The most frequent processes followed by students are identified, and then PM is applied to the entire group of those processes, but regardless of when they were conducted. The authors obtained three clusters called *Sampling*, *Comprehensive*, and *Targeting* learners. *Sampling* students explore some materials and do very few activities; *Comprehensive* students are those who make the most effort, delve deeper into content, and perform the most SRL; *Targeting* students focus their efforts on assessments and are less engaged. These clusters are also comparable to those obtained by [6], under the names *Probers*, *Continuously active*, and *Procrastinators*, respectively.

As far as we are aware, no study thus far has analyzed relationships among SRL and FA using EPM techniques considering the moment during a course in which students carry out both review exercises promoting SA and sessions with FA.

## III. Research Questions

When an online tool is used to perform both review exercises promoting SA and FA activities, the tool itself can track not only which exercises have been solved and the proposed solutions, but the time period over which each student has completed the activities. Traces of use of this tool could be used to analyze student engagement with review exercises promoting SA and FA activities. It is also interesting to examine the performance of students regarding their engagement. This performance will be measured through the academic results reflected in the course grades. In this context, and using tool traces to analyze the type and the timing of submitted exercises through an online tool, the research questions of this study are as follows:

RQ1. What are the patterns of student engagement with review exercises promoting SA over time?

RQ2. What are the patterns of student engagement with FA exercises over time?

RQ3. What are the patterns of student engagement when combining the SA and FA previous obtained patterns?

In the three questions above, we will also enquire how the identified patterns are associated with students' academic success (as measured by their course grades).

RQ4. How do review exercises promoting SA (or lack thereof) relate to students' engagement with FA exercises?

The implications of the results obtained in this study underscore the importance of taking into account the timing of SA and FA exercises when characterizing student behavior. In addition, this research represents progress in the bi-directional study of SA, as a SRL strategy, and FA. In this line, techniques such as process mining, that consider the temporal component of activities developed through an online tool, appear to be fundamental.

## IV. Material and Methods

### A. Academic Context and Online Tool

The present case study was conducted with students in a relational database course. It is a 60-hour course included in the first year of a software engineering degree program. Approximately 25% of the total class time was devoted to lectures on theoretical aspects of relational database systems and the SQL language [50], during which an instructor explains such concepts to the entire class. The remaining 75% of class time was reserved for hands-on learning of the SQL language, i.e., pen-and-paper SQL exercises, and scheduled SQL laboratories, which lasted 120 minutes. The SQL language is the predominant language used to define and manipulate relational databases. Although SQL syntax is relatively simple, it allows the construction of complex queries and learning the language requires considerable effort [51].

We developed an online tool, called DELFOS, that assists students in the process of learning the SQL language (see Supplementary Material for a description of the tool functionality and interfaces). DELFOS has two purposes. The first is to perform exercises in FA laboratory sessions. These sessions are a collaborative scenario in which the students, with the instructor help, try to progressively solve programming SQL queries provided by the tool. These labs are intended to articulate subgoals leading to the ultimate goal of learning SQL, which will be assessed through different written exams. DELFOS offers automatic assessment that informs students whether or not their answer matches the expected result. When the execution of the query is free of syntactical errors, the tool checks to see if the obtained query result and the expected query result match. This matching could overlook some students' misconceptions, such as the incorrect use of some SQL clauses or the inclusion of unnecessary tables or subqueries. Consequently, a basic help feature has been incorporated into DELFOS to call students' attention to these kinds of errors (Fig. 8 of the Supplementary Material shows a screenshot of the tool including a hint). Students also receive formative feedback from the session instructor. Depending on the circumstances, the instructor will address concerns individually, or explain issues to the whole group. In these labs, either another student or the instructor can help a student to solve a SQL query. Finally, each student individually submits the exercises developed using the tool. The work submitted throughout all the sessions receives a global grade intended to motivate and recognize students' effort in these activities. This is a "low stakes" grading that is automatically calculated through DELFOS, based on the number of exercises submitted by the students in these labs. However, as already noted by other authors [23], such relatively minor assessment can impel some students to falsely improve their deliverables by copying exercises from classmates. Nevertheless, this action has a negligible impact on the final grade and can be considered a type of instructional disobedience [13]. In the case of supervised FA laboratory sessions, students may have other motives for copying. For example, students may be running out of time, they are unable to solve a particular exercise, or they may have solved an exercise by collaborating with a classmate. In addition, well-intentioned students may collect solutions from peers to further study them, or ask classmates for help if they have problems with exercises [23].

The second purpose of DELFOS is for students to have access to a set of review exercises to promote their self-assessment. These review exercises are: optional, not scheduled, available at all times, and done by the students on their own. The tool assigns the review exercises to specific review sessions according to a set of options selected by the students. For instance, students may choose how many review exercises want to attempt in the session and the level of difficulty or kind of exercise. In this way, students can personalize each session according to their goals. The tool offers the automatic assessment and basic hints described above. Students can download a PDF document with the statements of the exercises and the solutions that they have proposed. This setting can be considered a use of a non-SRL tool for SRL purposes, in particular through a SA strategy [3].

The students' performance during the course was evaluated through three different formats (with grades ranging from 0 to 10 in all three cases):

1) *Laboratory sessions*: Seven supervised laboratory sessions with a total of 40 exercises were conducted using DELFOS. 3% of the course grade corresponds to students' efforts in laboratory sessions, which represents less than 0.05 points out of 10, per session.
2) *Interim tests and exams*: Three tests and three written exams were completed throughout the course. The tests included 6 to 10 questions about theoretical aspects of relational database systems. These tests constitute 3% of the final grade. The exams dealt with the SQL language, took place after the second, fourth, and seventh query laboratory sessions, and represented 24% of the final grade. The first exam includes simple queries and inner joins; the second: aggregations, outer joins, and set operations; and the third: subqueries.
3) *Final exam*: A written exam about relational databases and SQL exercises was given after all instruction had been completed and represented 70% of the final grade.

Students needed to earn at least 5 points to pass the course.

### *B. Participants*

After a pilot test of the tool was conducted during several laboratory sessions throughout the 2011 academic year, DELFOS was fully incorporated into laboratory sessions in the 2012 academic year. In 2014, the feature of review exercises promoting SA was introduced for the first time. However, during that year only the number of review sessions completed by each student was collected (but not the specific exercises solved in the session). With this in mind, data from two different periods were analyzed: (1) an initial period covering the 2012 and 2013 academic years, including 123 student participants (85% men), and during which only FA activities were conducted; and (2) a second period consisting of the academic years from 2015 through 2018, which includes 259 student participants (84% men), and during which both SA and FA activities were conducted. Both periods will be compared in order to analyze students' engagement and provide insight into the relationships between SA and FA. Let us note that the year 2014 has been discarded from the analysis for the reasons described above. Students who dropped out, i.e., students who did not take the final exam, have been excluded from the analysis. During all the years analyzed the instructors were the same and the difficulty of exercises, tests, and exams remained unchanged. The only variation between both periods was the inclusion of the review exercises promoting SA starting in 2014.

### *C. Log Events*

In this study, the interaction of the students with DELFOS was categorized as several series of events, grouped into two collections. Both collections were obtained from the tool log files. The first collected events are related to review exercises promoting SA. The kind of events considered here include the start and end-times of review sessions, and exercise submissions, noting which of the four written exams they preceded. A total of 13,407 events were collected in this log.

The second log collected the events of the laboratory sessions where students received FA. The kind of events considered include start and end-times of the laboratory session, and submission of original or copied exercise (hereinafter, original FA exercise or copied FA exercise). A total of 3,464 events of the 2012–2013 period and 8,288 events of the 2015–2018 period were collected in the second log. In this study, a submission is considered copied if it is a complete and exact copy (discarding word and line separators) of another solution submitted earlier by another student during the same session. Full copies have already been examined in other SQL learning studies [52]. Let us note that matches are not always due to copying, for example they could be the result of student collaboration. In addition, some submitted work that has some minor differences may be the result of copying. Obviously, there are more advanced mechanisms to detect student copying [53], but our experience in previous years indicates that the aforementioned measure of similarity is sufficient for our research purposes. We analyzed five of the seven laboratory sessions in which the tool was utilized. The first laboratory session was discarded from the study because it included very simple exercises. Such simple exercises resulted in many students proposing the same solution, even they were not necessarily copying each other's work (false positives). The last session was also discarded because it could not be held some years due to time constraints.

### *D. Instruments Used*

The Process Mining Project Methodology $PM^2$ was followed [16], which has also been used in other educational studies [33]. This methodology consists of six stages, and starts off by defining research questions (planning). Afterwards, the event data are extracted (extraction) and event logs are created structuring the obtained data in different views (data processing). Each event has information about the executed activity, as well as temporal information to sort the events. Fourthly, during the most important stage, process mining techniques are applied to event logs (mining and analysis). The discovered process models and other findings are evaluated (evaluation) and, finally, if the findings are satisfactory, they can be used to modify the process execution (process improvement and support).

In our case, once the research questions were established, the event data was extracted from the DELFOS database. Afterwards, following the research questions, the two event logs detailed in the previous subsection were created following a divide-and-conquer proposal [54], which allows each type of process to be analyzed separately with greater precision. The extraction and generation procedure for the two event logs was performed using the specialized tool XESame [55]. For the mining and analysis stage, we used the best-known process mining tool, called ProM [47]. In addition, we used Guide Tree Miner [56] as process diagnostic technique. Guide Tree Miner is a plugin for ProM that, working from event logs, uses agglomerative hierarchical clustering to automatically generate clusters, once it has been instructed about the number of clusters to create. For the evaluation stage, the clusters obtained were statistically analyzed using complementary data on academic performance (interim tests and exams, and final exam).

### *E. Measures*

The different grades obtained in the various assessment types are utilized to describe the academic performance of each cluster identified in terms of both SA and FA, as well as in the crossing of SA and FA clusters. In this analysis, ANOVA tests were conducted (including Bonferroni corrections to analyze each pair of data sets included in the ANOVA). Parametric conditions were verified prior to using these tests; and when parametric conditions were not fulfilled, the corresponding non-parametric tests (i.e., Kruskal–Wallis test) were used instead. Finally, a chi-square test was utilized to study the distribution independence for categorical data. The controlled variables included the instructors; the tool; the difficulty of the questions on the tests, exams, and labs; and the compulsory use of the tool in labs.

## V. Results

The following four subsections describe the results obtained for each research question posited by this study.

### *A. Clusters From the SA Event Log*

To answer the first research question, we analyzed the time information collected in the tool's trace to reveal students' engagement when doing review exercises promoting SA, and to compare the academic grades corresponding to each of them. Following the criteria and procedure indicated in the previous section, four clusters were obtained by applying PM to the log of SA events. We chose a four clusters solution, since adding more clusters did not provide relevant differences, and fewer clusters result in underspecification of students' traces (an analogue clustering approach is used in [6]). Fig. 1 contains a bubble chart representing the mean number of review exercises done by the students per week (within the semester) in each cluster. Table I includes the mean number of review exercises done by the students in the course and a comparison of the academic performance of each cluster. The labels chosen for these clusters were Strong, Initial, Final, and Weak SA.

By observing Fig. 1 one can appreciate that the cluster Strong SA contains a much greater amount of review exercises than the rest. In addition, the review exercises are distributed throughout the course in a more or less balanced manner, with intensity waning slightly as the final exam approaches. The Initial SA cluster contains fewer exercises than the Strong SA cluster and intensity decreases as the course progresses. Students started out investing a lot of work and then their effort decreased to a nominal level during the sessions associated with the final exam. The Final SA cluster is, in a way, the opposite of the Initial SA cluster because the pace of work increased over time and it was much higher in the sessions associated with the final exam. Last, the Weak SA cluster is the one with the fewest sessions and exercises throughout the course. The level of work was very low regardless of which exam the review sessions were associated with.

Considering the four clusters obtained, which reflect the individual self-assessment effort, comparative analyses were done with other data. Firstly, distribution does not differ among the different clusters according to gender: Weak (82 vs. 14), Final (65 vs. 12), Initial (33 vs. 5), and Strong (37 vs. 11); ($\chi^2$ = 2.059, $p$ = .560). On the other hand, drop-out students are concentrated in the Weak SA cluster (24), and very rare in the other clusters: Final (2), Initial (1), and Strong (2).

A comparative study was also done of the grades obtained by the clustered students on the different assessments (excluding drop-out students). Table I shows these comparisons, where significant differences between clusters can be observed in all assessments except for the exam 1. After comparing the groups by pairs (using Bonferroni correction), differences can be observed among exams 2, 3, and the final exam between the Strong SA cluster and the two with fewer review exercises (Final and Weak). There are also differences among exams 2 and 3 between the Initial and Final clusters, but not between Initial and Weak. Fig. 2 shows graphically how the Initial SA cluster got the best marks at the beginning, even surpassing the Strong SA cluster, but its performance declined as the course progressed. It can also be observed that the Weak SA cluster obtained better grades than the Final SA cluster on the first three exams, but is slightly surpassed by Final in the last exam. As we have seen, students in the Final SA cluster completed many review exercises associated with the last exam.

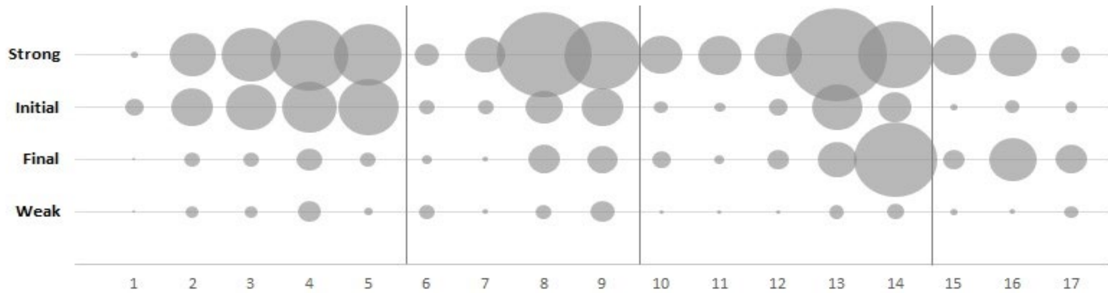


Fig. 1. Representation of the clusters (y-axis) obtained from the SA Event Log. The x-axis shows the number of the week (within the semester) when the students perform review exercises promoting SA. The size of the bubble represents the average number of exercises done by the students in that week. The vertical lines mark the moments when the three interim exams are scheduled.

TABLE I
MEAN (STANDARD DEVIATION) OF THE NUMBER OF EXERCISES SOLVED IN REVIEW SESSIONS PROMOTING SA AND OF THE GRADES FOR THE DIFFERENT ASSESSMENTS IN THE FOUR REVIEW GROUPS (GRADES ON A 10-POINT SCALE)

| | Groups regarding SA | | | | Statistic | After Bonferroni |
|---|---|---|---|---|---|---|
| | **S: Strong** | **I: Initial** | **F: Final** | **W: Weak** | | |
| **N** | 46 | 37 | 75 | 72 | | |
| **Reviews** | 41.69 (30.01) | 15.79 (11.49) | 12.13 (12.53) | 2.44 (4.17) | $\chi^2 = 160.304$*** | S > I, F > W |
| **Exam 1** | 5.41 (2.98) | 5.63 (2.92) | 4.54 (2.22) | 4.65 (2.91) | $\chi^2 = 1.656$; $p = .437$ | – |
| **Exam 2** | 5.71 (2.74) | 5.00 (2.59) | 3.47 (2.47) | 4.01 (2.70) | $F = 8.180$*** | S > F, W; I > F |
| **Exam 3** | 6.25 (3.33) | 5.14 (3.13) | 3.34 (2.91) | 4.20 (3.53) | $F = 8.345$*** | S > F, W; I > F |
| **Final Exam** | 6.06 (2.40) | 4.63 (2.36) | 4.16 (2.31) | 3.81 (2.77) | $F = 8.400$*** | S > F, W |
| **Labs** | 8.67 (1.96) | 8.73 (1.68) | 7.69 (2.33) | 7.58 (2.73) | $\chi^2 = 10.569$**; $p = .005$ | – |
| **Tests** | 5.18 (1.91) | 4.74 (1.78) | 4.29 (1.74) | 4.06 (2.16) | $F = 7.737$*** | S > W |

**$p < .01$, ***$p < .001$, Statistics: ANOVA $F$ test, Kruskal–Wallis $\chi^2$ test

### *B. Clusters From the FA Event Log*

To address the second research question, the time information and the submitted exercises (distinguishing between original and copied work) collected in the tool's trace are analyzed to identify student's engagement when completing FA exercises. Students' academic grades are also considered.

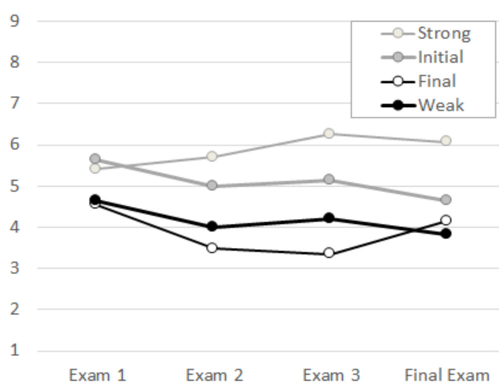


Fig. 2. Comparison of exams grades according to SA clusters.

Following the criteria and procedure indicated in the previous section, four clusters were obtained by applying PM to the laboratory session events log (FA). Fig. 3 contains a bubble chart representing the average number of exercises done by students (distinguishing between original and copied exercises) in each consecutive period of 20 minutes in the FA laboratory sessions. Table II includes the mean number of exercises done by the students (also distinguishing between original and copied exercises) and a comparison of the academic performance of each cluster. The labels chosen for these clusters were Very High, High, Medium, and Low FA.

One can observe in Fig. 3 how the Low FA cluster contains a much larger amount of copied exercises in all the sessions and that these are fairly evenly distributed throughout each session. In many cases, fewer exercises were submitted. The Medium FA cluster has fewer copied exercises than the Low FA cluster, but considerably more than the two remaining clusters. Copies are not concentrated at the beginning or at the end of the session. The High FA cluster has fewer copies than the Medium FA cluster. In addition, copies are concentrated at the end of some sessions. Finally, the Very High FA cluster is the one with the fewest copies and includes the largest number of original exercises.

Comparative analyses of these clusters were also performed. First of all, there are slight differences in the distribution of men and women: Very High (74 vs. 13), High (34 vs. 13), Medium (51 vs. 4), Low (58 vs. 12); ($\chi^2 = 7.947$, $p = .047$). A smaller proportion of women can be observed in the Medium cluster and a higher proportion in the High cluster, although due to the small total number of female students, conclusions cannot be drawn. In addition, drop-outs are concentrated in the Low cluster (19), and are very rare in the other clusters: Very High (1), High (1), and Medium (4).

A comparative study of the grades obtained by the different clusters of students on the various course evaluation assessments (again excluding drop-out students) was also conducted. Table II shows these comparisons, where significant differences can be observed between the clusters in all the assessments. After comparing the cluster pairs (applying the Bonferroni correction), differences are observed in all the assessments between the Very High and High clusters and the Low cluster. The Medium cluster obtained worse results on exams 2, 3, and the final exam as compared to the Very High

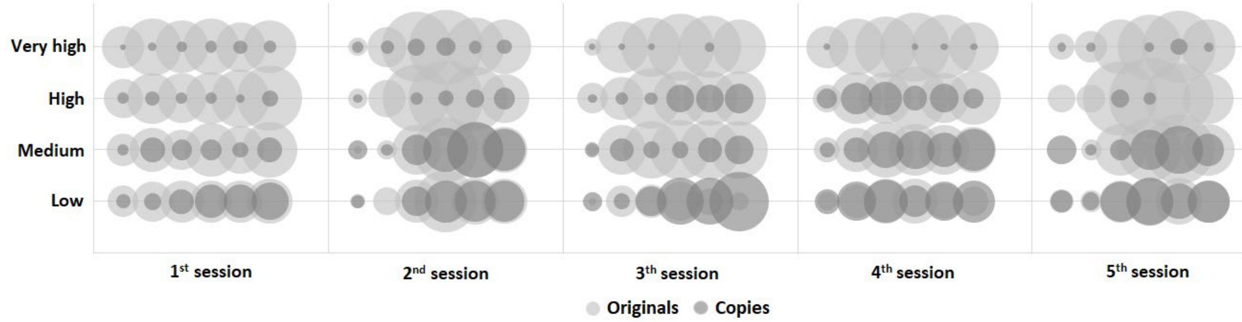


Fig. 3. Representation of the clusters (y-axis) obtained from the FA Event Log when review exercises promoting SA are included in DELFOS. The x-axis shows the five laboratory sessions. The size of the bubble represents the average number of exercises (original exercises, in light grey; copied exercises in dark grey) done by the students throughout the session (consecutive periods of 20 minutes has been considered in each laboratory session).

TABLE II
MEAN (STANDARD DEVIATION) OF THE GROUPS' GRADES FOR THE DIFFERENT ASSESSMENTS REGARDING FA (GRADES ON A 10-POINT SCALE)

| | Groups regarding FA | | | | | |
|---|---|---|---|---|---|---|
| | **V: Very High** | **H: High** | **M: Medium** | **L: Low** | **Statistic** | **After Bonferroni** |
| **N** | 86 | 46 | 51 | 51 | | |
| **Copies** | 0.76 (1) | 2.30 (1.47) | 5.78 (2.86) | 7.18 (6.27) | $\chi^2 = 117.870^{***}$ | V, H > M, L |
| **Originals** | 21.62 (4.37) | 20.57 (3.68) | 16.29 (4.86) | 10.26 (4.86) | $\chi^2 = 132.630^{***}$ | V, H > M > L |
| **Exam 1** | 5.52 (2.77) | 5.54 (2.56) | 4.53 (2.76) | 3.94 (2.61) | $\chi^2 = 14.335^{**}$; $p = .02$ | V, H > L |
| **Exam 2** | 5.32 (2.58) | 5.40 (2.77) | 3.74 (2.55) | 2.52 (1.97) | $F = 17.811^{***}$ | V, H > M, L |
| **Exam 3** | 5.42 (3.30) | 5.62 (3.37) | 3.67 (3.26) | 2.58 (2.63) | $F = 10.930^{***}$ | V, H > M, L |
| **Final Exam** | 5.35 (2.37) | 5.73 (2.23) | 3.58 (2.55) | 2.76 (2.18) | $F = 19.256^{***}$ | V, H > M, L |
| **Labs** | 8.49 (1.68) | 8.60 (1.80) | 8.53 (1.72) | 6.69 (3.10) | $\chi^2 = 16.279^{***}$ | V, H, M > L |
| **Tests** | 4.84 (1.97) | 5.12 (1.93) | 4.50 (1.52) | 3.63 (1.84) | $F = 7.737^{***}$ | V, H > L |

$^{**}p < .01$, $^{***}p < .001$, Statistics: ANOVA $F$ test, Kruskal–Wallis $\chi^2$ test

and High clusters. However, it achieved better results than the Low cluster in terms of submitting laboratory exercises (which contains students who failed to deliver some exercises). Fig. 4 shows graphically that the Very High and High clusters score similarly on the exams and that the other two clusters remain far below them, with the Low cluster getting the worst scores.

Applying PM to obtain 3 clusters, the Very High and High clusters are unified, and when applied to 2 clusters, Medium and Low are also unified. In the light of the results in Table II, the fusion of these two cluster pairs seems a natural simplification of the situation found in this study. In the following section, these two clusters are referred to as High and Low FA. This simplification does not alter the distribution of men and women in either of the two clusters: High (108 vs. 26), Low (105 vs. 16); ($\chi^2 = 1.765$, $p = .184$).

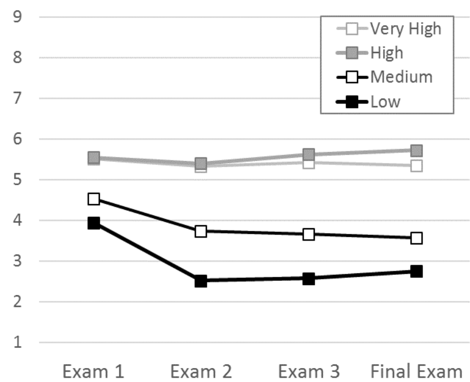


Fig. 4. Comparison of exams grades according to FA clusters.

### C. Clusters Combining SA and FA

To answer the third research question, an analysis that combines the engagement patterns obtained in the two previous sections is performed. Table III lists the eight clusters resulting from crossing the four on SA and the two on FA: Strong, Initial, Final, and Weak, each paired with High and Low. Drop-out students were excluded from these results (they were concentrated in the Weak–Low cluster (21) and very rare in the rest of clusters: Strong–Low (2), Initial–Low (1), Final–Low (1), Strong–High (2), Initial–High (0), Final–High (2), and Weak–High (2)). Table III shows differences in all assessments and the differences that remain after applying the Bonferroni correction. These results are displayed graphically in Fig. 5, where one can see that strategies with a high level of FA exercises (dashed lines), in most cases, obtained better results

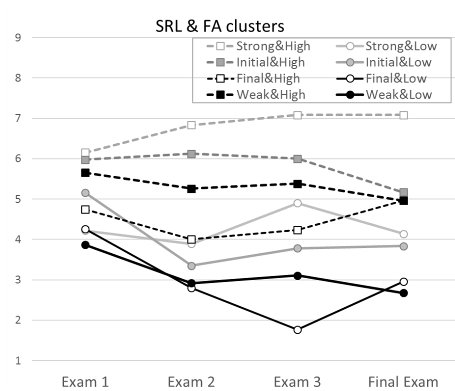


Fig. 5. Comparison of exams grades when combining SA clusters with High FA (dashed lines) and Low FA (straight lines) clusters.

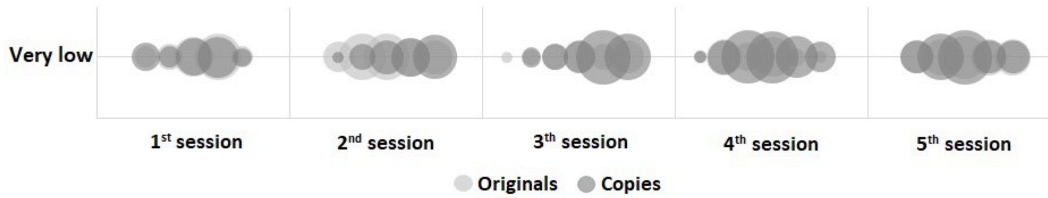


Fig. 6. Representation of the Very Low cluster obtained from the FA Event Log without review exercises. The x-axis shows the five laboratory sessions. The size of the bubble represents the average number of exercises (original exercises, in light grey; copied exercises in dark grey) done by the students throughout the session (consecutive periods of 20 minutes has been considered in each laboratory session).

TABLE III
MEAN (STANDARD DEVIATION) OF THE GROUPS' GRADES FOR THE DIFFERENT ASSESSMENTS REGARDING SA AND FA

| | Groups regarding SA and FA | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Strong High **SH** | Initial High **IH** | Final High **FH** | Weak High **WH** | Strong Low **SL** | Initial Low **IL** | Final Low **FL** | Weak Low **WL** | **Statistic** | **After Bonferroni** |
| **N** | 28 | 22 | 42 | 37 | 12 | 15 | 32 | 36 | | |
| **Copies** | 1.2 (1.3) | 1.5 (1.5) | 1.4 (1.5) | 1.2 (1.4) | 5.2 (2.8) | 6.3 (3) | 7 (6) | 6.8 (5.6) | $\chi^2 = 101.157$*** | H > L |
| **Originals** | 21.5 (4.8) | 21.8 (3.1) | 21.1 (4.5) | 21 (3.9) | 15.9 (4.8) | 14.9 (4.1) | 12.6 (5.9) | 11.8 (5.6) | $\chi^2 = 117.259$**** | L > H |
| **Reviews** | 50.4 (32.3) | 17.5 (13.4) | 15.2 (14.4) | 4.1 (5.5) | 27.1 (18.7) | 13.5(8) | 8.1 (8.5) | 1.4 (2.5) | $\chi^2 = 150.047$*** | SH > SL, IH, FH > WL, WH; SH > IL, FL; SL > FL |
| **Exam 1** | 6.2 (3.1) | 6 (2.9) | 4.8 (2.2) | 5.7 (2.7) | 4.2 (2.5) | 5.2 (3.1) | 4.3 (2.3) | 3.9 (2.9) | $\chi^2 = 21.311$** $p = .003$ | – |
| **Exam 2** | 6.8 (2.2) | 6.1 (2) | 4 (2.6) | 5.3 (2.6) | 3.9 (2.6) | 3.5 (2.5) | 2.8 (2.2) | 2.9 (2.3) | $F = 11.518$*** | (SI)H > FH, (SIFW)L; WH > (FW)L |
| **Exam 3** | 7.1 (3.3) | 6 (3) | 4.2 (2.7) | 5.4 (3.6) | 4.9 (3) | 3.8 (2.9) | 1.8 (2.4) | 3.1 (3.1) | $F = 8.733$*** | SH > FH, (IFW)H; IH > (FW)L; (FW)H, SL > FL |
| **Final Exam** | 7.1 (2) | 5.2 (1.8) | 5 (2) | 5 (2.7) | 4.1 (2) | 3.8 (2.9) | 3 (2.2) | 2.7 (2.4) | $F = 12.028$*** | SH > (FW)H, (SIFW)L; (IFW)H > (FW)L |
| **Labs** | 8.7 (1.9) | 9.1 (1.1) | 8.2 (1.8) | 8.4 (1.8) | 8.7 (2.9) | 8.2 (2.2) | 7.2 (2.5) | 7.1 (3.1) | $\chi^2 = 18.391$* $p = .01$ | – |
| **Tests** | 5.6 (1.9) | 5 (1.7) | 4.6 (1.7) | 4.8 (2.3) | 4.5 (1.7) | 4.4 (1.9) | 4.1 (1.6) | 3.7 (1.8) | $F = 3.514$** $p = .001$ | SH > (FW)L |

*$p < .05$, **$p < .01$, ***$p < .001$, Statistics: ANOVA $F$ test, Kruskal–Wallis $\chi^2$ test

than those with a low level of FA exercises (straight lines). Fig. 5 also shows that only the engagement pattern involving a high level of FA exercises and reviewing at the end (Final–High) was surpassed in some assessments by other patterns where students had a low level of FA exercises

### D. FA Clusters Without Review Exercises Promoting SA

In order to answer the fourth research question, we analyzed the log of events of laboratory exercises (FA) from the previous period (2012–2013) when there were no review exercises promoting SA. Four clusters were obtained using the same criteria than in the previous cases, and, after their visualization and comparison, similarities were found in three of them with the High, Medium, and Low clusters obtained with the log of the review-exercises period. The average number of copies and originals ("Copies" and "Originals" rows) of the clusters with the same name in Tables II and IV are similar, as are also similar the ratios between the average number of copies and the average number of originals: 0.11 and 0.16 for High, 0.35 and 0.34 for Medium, and 0.70 and 0.67 for Low. However, none of the clusters obtained in this new period is similar to the previous Very High. On the contrary, the Very Low cluster appears with a smaller number of originals and a larger number of copies (compare again rows "Copies" and "Originals" in Tables II and IV with ratios between the averages of copies and the averages of originals for Very High and Very Low of 0.03 and 1.54, which are very different from all the others). Fig. 6 contains a bubble chart representing the average number of exercises done by the students (distinguishing between original and copied exercises) each consecutive period of 20 minutes in the FA laboratory sessions for this Very Low cluster.

Distribution of gender does not differ in the various clusters: High (43 vs. 10), Medium (24 vs. 4), Low (20 vs. 1), Very Low (18 vs. 3); ($\chi^2 = 2.404$, $p = .493$). Drop-outs are concentrated in the Low (12) and Very Low (6) clusters, while less frequent in the High (3) and Medium (3) clusters. Drop-outs were excluded from the study of academic performance.

Table IV shows the comparative study of clustered students' grades on different course assessments. Significant differences between clusters are observed in all assessments, as is the case for the period that includes review exercises promoting SA. In the comparison between clusters, after performing the Bonferroni correction, differences are observed in all the assessments between the High and Medium clusters and the Very Low cluster except in the case of submission of laboratory exercise (labeled "Labs"). There are also differences among these two clusters and the Low cluster in all the assessments except the first two exams. In this case, the High and Medium clusters obtained similar results, while Low and Very Low also shared some similarities.

Finally, Fig. 7 graphically compares the evolution of exam scores in the FA clusters of the two periods studied. During the period without review exercises promoting SA (straight lines), there was a strong improvement in the final exam as compared to the previous exams (improvements were around two points). However, in the period with review exercises (dashed lines), the grades for exams 1, 2, and 3 were much better in general and similar to the final exam.

TABLE IV
MEAN (STANDARD DEVIATION) OF THE GROUPS' NUMBER OF EXERCISES SOLVED IN REVIEW SESSIONS AND OF THE GROUPS' GRADES FOR THE DIFFERENT ASSESSMENTS REGARDING FA WITHOUT REVIEW EXERCISES PROMOTING SA

| | Groups regarding FA without SA (Review Exercises) | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Very High** | **H: High** | **M: Medium** | **L: Low** | **V: Very Low** | **Statistic** | **After Bonferroni** |
| **N** | | 50 | 25 | 9 | 15 | | |
| **Copies** | | 2.74 (2.05) | 4.50 (2.68) | 5.57 (2.81) | 10.76 (4.24) | $\chi^2 = 48.410$*** | H < M, L < V |
| **Originals** | | 16.77 (5.32) | 13.39 (5.57) | 8.25 (4.33) | 7 (3.01) | $\chi^2 = 45.542$*** | H > M > L, V |
| **Exam 1** | | 4.31 (2.87) | 4.57 (2.94) | 3.07 (2.53) | 2.12 (2.81) | $\chi^2 = 16.902$** $p = .001$ | H, M > V |
| **Exam 2** | | 4.33 (2.67) | 3.73 (3.37) | 1.57 (2.26) | 1.21 (1.61) | $\chi^2 = 22.611$*** | H, M > V |
| **Exam 3** | | 3.44 (2.95) | 3.6 (3.37) | 0.44 (0.62) | 0.81 (1.61) | $\chi^2 = 19.666$*** | H, M > L, V |
| **Final Exam** | | 5.27 (2.73) | 5.31 (3.23) | 2.94 (2.12) | 2.35 (1.92) | $F = 6.084$** $p = .001$ | H, M > L, V |
| **Labs** | | 7.38 (1.92) | 7.16 (2.80) | 4.80 (2.19) | 6.77 (2.02) | $\chi^2 = 19.190$*** | H, M, V > L |
| **Tests** | | 5.45 (2.07) | 5.49 (2.51) | 3.43 (1.90) | 3.9 (1.60) | $F = 7.677$*** | H, M > L, V |

**$p < .01$, ***$p < .001$, Statistics: ANOVA $F$ test, Kruskal–Wallis $\chi^2$ test

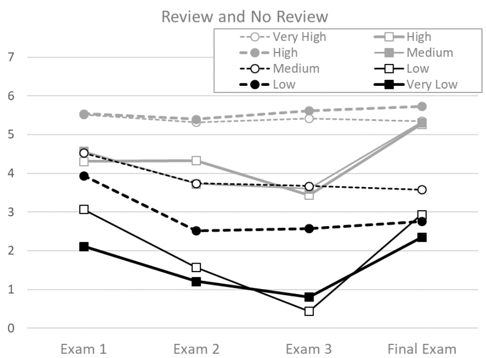


Fig. 7. Comparison of exam grades for periods with SA review exercises (dashed lines) and without them (straight lines).

## VI. DISCUSSION

The following four subsections discuss the results obtained for the four research questions set out in this study.

### *A. Discussion on SA Clusters*

The four clusters obtained from the review exercises log seems to indicate four engagement patterns regarding SA. Students in the Strong cluster did continuous review exercises throughout the course and achieved the best academic results on almost all assessments. Students in the Weak cluster made few review exercises and obtained poor results. These results are in line with what other studies have observed regarding academic performance and SRL [2], [7], [33].

Regarding the behaviors in the two intermediate clusters, students' efforts were concentrated into a few weeks, and the difference compared with the rest of weeks is very noticeable. In Initial, the greatest SA effort was condensed into the first weeks and in Final, in the last weeks. There are no differences between these two clusters in regards to average dedication; but Initial's efforts turned out to be more beneficial, as those students obtained better intermediate results, and even achieved the best results on the first exam. Then, as their SA investment waned, their results became worse. On the other hand, Final had poor results in general (the lowest grades on several assessments) and reacted too late, although this cluster's results were similar to the Initial cluster on the final exam. The Initial group's relaxation during the final stretch of the course could be the consequence of achieving sufficiently satisfactory results on initial assessments; while the Final cluster's reaction could result from not obtaining the desired grades. Some authors assert that SRL can be conceptualized as a dynamic series of behavioral and motivational events (among others) [4]. Furthermore, it is difficult to determine whether students' behavioral changes throughout the course derive from their SRL strategy, or involve other factors [33]. Based on the way they access to optional reviews, both intermediate groups could be classified as surface learning which is characterized by the student trying to avoid failure, and which differs greatly from the mastery learning orientation, adopted by the Strong group, which prioritizes real comprehension and is intrinsically motivated [14], [21].

Our clusters can be compared with those obtained in other studies regarding SRL [6], [33]. Clusters called *Sampling* [33] and *Probers* [6] could correspond to the Weak cluster from our study, or even to drop-out students. *Comprehensive* [33] and *Continuously active* [6] clusters would correspond to the Strong cluster. Initial and Final clusters would fit primarily to *Targeting* [33] and *Procrastinators* [6].

### *B. Discussion on FA Clusters With Review Exercises*

The clusters obtained from the log of laboratory sessions in the period with review exercises also seems to indicate four engagement patterns regarding FA. The trace of students in the group called Very High showed that these students did almost all the FA exercises and barely copied; while the ones in the Low group did not often do the FA exercises on their own, with a low "instructional obedience" according to the terminology of Elen [13]. These groups obtained the best and the worst academic results, respectively. This result was expected, and is noted in the literature on FA [23], [57]. Very High students are those who invested the most effort in finding their own solutions. These students learned much more than students who avoided tackling exercises, such as the Low students, for instance.

The intermediate patterns are: High, i.e., students whose trace in the tool reflects their engagement in almost all the FA exercises (as in the case of Very High) and with very few non-original exercises submitted at the end of a session (perhaps these students just needed more time); and Medium, whose trace indicates the submission of non-original exercises quite frequently throughout the session (perhaps they were overwhelmed by difficulties solving exercises). High's results were very similar to those of the Very High group. It seems that

High students tried their best, but when they ran out of time, they copied some exercises. Apparently, the Medium students avoided tackling challenging exercises and probably did not fully understand the solutions they copied. This last pattern obtained poor academic results in intermediate assessments, although copies helped them obtain a score comparable to Very High and High for a very small portion of the course grade (laboratories). The absence of this learning effort is evident in the supervised exams [23].

### *C. Discussion on Clusters That Combine SA and FA*

It is worth recalling that in this analysis the four SA clusters (Weak, Final, Initial, and Strong) have been crossed with the two FA clusters which gather the students that submitted the most (High) and least (Low) original FA exercises. When comparing the academic results of the eight resulting clusters, it is observed that all the clusters with a high level of involvement in FA laboratories (High) obtained better results on all the assessments than those with low level of involvement (Low), with only two exceptions located in the cluster with review exercises promoting SA at the end and many originals (Final–High). This last cluster does end up surpassing all of the Low ones in the final exam. Therefore, it can be inferred that, in general, students with better grades engaged more with FA exercises, regardless of how much they reviewed.

As could be expected, the group that submitted more FA original exercises and more review exercises promoting SA (Strong–High) obtained the best academic results (significantly better in the last two exams), and the group that reviewed only a little and had less originals (Weak–Low) is among the worst (on the last exam significantly). Another group that is among the worst is that which reviewed only at the end and submitted less original work (Final–Low). This third group, which began to review very late, obtained progressively worse results (and in some cases had the worst results). Although it did manage to slightly surpass Weak–Low on the final exam, it still got poor results. The first group (Strong–High) seems to fit into the mastery goal orientation, and the other two into work-avoidant [14], [21]. Mastery students focus on learning, finding problem-solving strategies, and strive consistently throughout the course. Work-avoidant students try to complete the course with the minimal effort necessary. The literature makes it clear that the first orientation tends to have satisfactory academic results and the second tends to obtain poor academic results [21].

Among the three intermediate groups that had more FA original exercises (Initial–High, Final–High, Weak–High), the one that reviewed the least (Weak–High) seemed to follow a performance-goal orientation [14], [21]. Performance students strive to prove their worth and are interested in obtaining strong academic results in comparison to other students. Therefore, they tend to present their own version of the exercises without copying others. However, they do not usually do self-study tasks [21], perhaps because this activity does not produce results that can be compared with other students. The engagement patterns of the other two intermediate groups that had more original FA exercises seem to be variations of the latter. The Initial–High group began like mastery students but, after obtaining satisfactory results, decided to perform fewer review exercises. The Final–High group, on the other hand, reacted to unsatisfactory results, applied themselves quite late in the course, and therefore failed to achieve the desired performance.

The remaining intermediate groups (Strong–Low and Initial–Low) had less original FA exercises. Strong–Low reviewed a lot, and seemed to follow a performance-avoidant goal orientation [14], [21]. Performance-avoidant students do not consider themselves to be very competent, and want to avoid revealing their lack of ability, so they are focused on avoiding failure [21]. Therefore, they try to copy exercises from other classmates and also do review exercises to compensate for their insecurities. Nevertheless, the effort avoided in FA sessions is not compensated by increased SA review sessions, since the obtained results are worse than the Weak–High students. The other group is Initial–Low, which looks like a variant of the previous orientation who, in view of its good initial results, decided to relax its SA effort.

Students with mastery goal orientation have been found to achieve better results than performance-orientation students [21]; and the latter, better than performance-avoidant [58]. This situation is also observed in the results presented herein, if we match our clusters with goal orientations.

### *D. Discussion on FA Clusters Without Review Exercises*

Data from the log of laboratory sessions, collected during the period when the tool did not include review exercises, allowed us to compare the student activity in laboratory sessions including FA before and after these review exercises promoting SA were available. Analyzing these data, four new clusters were obtained. Three of these clusters were labeled High, Medium, and Low, due to their great similarity with the clusters of the same name from the aforementioned period including review exercises. The fourth cluster was labeled Very Low, as students in that group had even less originals than those in the Low group. There is no cluster equivalent to the Very High cluster. Therefore, a change in the students' engagement in the laboratory sessions providing FA can be observed. The introduction of review exercises promoting SA could suppose a change in the contexts and the individual perceptions of the course activities, and could originate a student reaction [4], [14]. The new SA activities could mediate between students personal and contextual characteristics and formative labs engagement. In particular, students were impelled to position themselves in groups with greater instructional obedience, thereby enhancing their learning effort during FA sessions.

Comparing the academic results among the clusters of both periods also allows differences to be identified. In the period without review exercises promoting SA there was a marked improvement in the results of the final exam compared to previous assessments. However, in the period with the review exercises, the results of the final exam are very similar to the earlier assessments. In addition, intermediate exam scores are much better in the review-exercises period. This seems to indicate that promoting SA, by offering review exercises, enables students to distribute learning effort throughout the course, as other studies also suggest [59], [60]. On the contrary, when students did not have this possibility, they concentrated their learning effort into the weeks prior to the final exam. This reaction could also be a direct consequence of poor results on earlier exams. Some studies find that starting to study late has negative consequences on academic performance and can lead students to drop out of the course [59].

An improvement of results can also be observed in all the clusters of the review-exercises period as compared to the similar clusters from the other period. This difference is very evident in exams 1, 2, and 3; however, the results on the final exam are more similar. As noted above, in the review-exercises period, students put in more effort in FA laboratory sessions and performed review exercises. This additional effort was distributed throughout the course and explains the improvement on interim exams during this period. However, the students from this period did not feel inclined to make the same degree of final effort as their peers from the period without review exercises; and in the end, the final grades of the different groups are similar. Some research raises doubts about the effectiveness of interim exams in improving final assessment [10], [61]. Herein, we have found that improved interim exam results did not translate directly into better final assessment results.

### *E. Limitations of the Study*

Let us bear in mind that this study characterizes students' learning behavior solely based on their use of an online tool with review exercises promoting SA and FA exercises done in laboratories, without a prior assessment of SRL skills.

The work of Broadbent *et al.* [3] mentions that the technology could aid the learner with "distributed metacognition" that prompts and supports SA during the student's interaction with a tool. While this may improve learning outcomes, as appears to be the case in our study, more research is necessary to conclude whether it also enhances metacognitive knowledge and independent self-regulation outside the interaction with the technology.

Moreover, this study has been conducted on a concrete domain (relational databases and SQL programming), in a single course with specific methodological choices that could affect the study results. Nevertheless, we believe that the primary components of this research—review exercises promoting SA and FA exercises on a tool used in an undergraduate course—can be applied to other contexts, but more research is needed to generalize the obtained results.

### *F. Implications*

The knowledge provided by tracking the timing of tasks through process mining allows researchers to determine clusters of students, not only according to the amount of activity, but also the timing of their different types of activities. This information makes it possible to recognize, and even anticipate, each student's behavior during the course. On the other hand, it is beneficial to encourage activities that promote SA since it positively impacts FA activities by fostering strategies that make better use of FA laboratory time, probably because students come to laboratory sessions better prepared. Finally, empowering SA also has a strong influence on intermediate exams. This could be a consequence of less procrastination and, therefore, of a more balanced distribution of learning effort throughout the course.

## VII. Conclusion

Online computing tools support teaching in face-to-face degree programs in many different ways. For example, the two educational approaches studied herein—Self-Assessment (SA) and Formative Assessment (FA)—can benefit from these tools. In addition, researchers can track student activity with these tools, which allows them to examine student's engagement through techniques such as process mining. In our case, both SA and FA were integrated into the same tool, which enabled us to analyze the relationships between them.

Our study traces the students' interactions with a particular tool for both the diversity and quantity of activities completed by them, considering their distribution over time. This study allows us to explore whether it is possible to find different student engagement patterns when the tool is used alone in FA laboratories compared to when, through a modification of the learning context, the tool is used to additionally provide review exercises that promote SA. The main findings of the study are that activities designed to promote SA significantly improved students' involvement in FA tasks; on the other hand, it was also observed that a lack of effort in FA exercises cannot be compensated by increased SA activity.

## Acknowledgment

The authors would like to thank Diego Díez Ricondo and Juan-Francisco Diez Léglise for the development of the SQL learning tool Delfos as part of their capstone projects.

## References

[1] E. Panadero, "A review of self-regulated learning: Six models and four directions for research," *Frontiers Psychol.*, vol. 8, no. 422, Apr. 2017, doi: 10.3389/fpsyg.2017.00422.

[2] A. Pardo, F. Han, and R. A. Ellis, "Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance," *IEEE Trans. Learn. Technol.*, vol. 10, no. 1, pp. 82–92, Dec. 2016, doi: 10.1109/TLT.2016.2639508.

[3] J. Broadbent, E. Panadero, J. M. Lodge, and P. de Barba, "Technologies to enhance self-regulated learning in online and computer mediated learning environments," in *Handbook of Research in Educational Communications and Technology*, M.J. Bishop, J. Elen, E. Boling, and V. Svihla, Eds., New York, NY, USA: Springer-Verlag, 2020, pp. 37–52.

[4] 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," *Scand. J. Educational Res.*, vol. 60, no. 6, pp. 723–735, 2016, doi: 10.1080/00313831.2015.1066436.

[5] R. F. Kizilcec, M. Pérez-Sanagustín, and J. J. Maldonado, "Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses," *Comput. & Educ.*, vol 104, pp 18–33, Jan. 2017, doi: 10.1016/j.compedu.2016.10.001.

[6] A. Cicchinelli et al., "Finding traces of self-regulated learning in activity streams," in *Proc. 8th Int. Conf. Learning Analytics and Knowledge (LAK'18)*, Sydney, Australia, Mar. 7–9, 2018, pp. 191–200, doi: 10.1145/3170358.3170381.

[7] 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, doi: 10.1016/j.compedu.2018.05.006.

[8] B. J. Zimmerman and M. M. Pons, "Development of a structured interview for assessing student use of self-regulated learning strategies,"

*Amer. Educational Res. J.,* vol. 23, no. 4, pp. 614–628, Jan. 1986, doi: 10.3102/00028312023004614.

[9] E. Panadero, H. Andrade, and S. Brookhar, "Fusing self-regulated learning and formative assessment: A roadmap of where we are, how we got here, and where we are going," *Australian Educational Researcher*, vol. 45, no. 1, pp. 13–31, Mar. 2018, doi: 10.1007/s13384-018-0258-y.

[10] Formative Assessment for Students and Teachers (Fast) SCASS, "Distinguishing formative assessment from other educational assessment labels," Council of Chief State School Officers, Washington, DC, USA, 2012. [Online]. Available: https://www.michigan.gov/documents/mde/CCSSO_Assessment__Labels_Paper_ada_601108_7.pdf

[11] I. Clark, "Formative assessment: Assessment is for self-regulated learning," *Educational Psychol. Rev.*, vol. 24, no. 2, pp. 205–249, Jun. 2012, doi: 10.1007/s10648-011-9191-6.

[12] Y. Xiao and M. Yang, "Formative assessment and self-regulated learning: How formative assessment supports students' self-regulation in english language learning," *System*, vol. 81, pp. 39–49, Apr. 2019, doi: 10.1016/j.system.2019.01.004.

[13] J. Elen, "Instructional disobedience: A largely neglected phenomenon deserving more systematic research attention," *Educational Technol. Res. Develop.*, vol. 68, no.5, pp. 2021–2032, Apr. 2020, doi: 10.1007/s11423-020-09776-3.

[14] P. R. Pintrich, "The role of goal orientation in self-regulated learning," in *Handbook of Self-Regulation*, M. Boekaerts, P. R. Pintrich, and M. Zeidner, Eds., San Diego, CA, USA: Academic, 2000, pp. 452–502, doi: 10.1016/B978-012109890-2/50043-3.

[15] A. D. Alharthi, M. Spichkova, and M. Hamilton, "Sustainability requirements for eLearning systems: A systematic literature review and analysis," *Requirements Eng.*, vol. 24, no. 4, pp. 523–543, May 2019, doi: 10.1007/s00766-018-0299-9.

[16] M. L. van Eck, X. Lu, S. J. J. Leemans, and W. M. P. van der Aalst, "PM$^2$: A process mining project methodology," in *Proc. 27th Int. Conf. Advanced Information Systems Engineering (CAiSE'15)*, Stockholm, Sweden, Jun. 8–12, 2015, in Lecture Notes Comput. Sci., vol. 9097, May. 2015, pp. 297–313, doi: 10.1007/978-3-319-19069-3_19.

[17] P. H. Winne, "Issues in researching self-regulated learning as patterns of events," *Metacognition Learn.*, vol. 9, no. 2, pp. 229–237, Feb. 2014, doi: 10.1007/s11409-014-9113-3.

[18] P. H. Winne and R.S. Baker, "The potentials of educational data mining for researching metacognition, motivation and self-regulated learning," *J. Educational Data Mining*, vol. 5, no. 1, pp. 1–8, May 2013, doi: 10.5281/zenodo.3554619.

[19] B. J. Zimmerman, "Attaining self-regulation: A social cognitive perspective," in *Handbook of Self-Regulation*, M. Boekaerts, P. R. Pintrich, and M. Zeidner, Eds., San Diego, CA, USA: Academic, 2000, pp. 13–39, doi: 10.1016/b978-012109890-2/50031-7.

[20] P. H. Winne and A. F. Hadwin, "Studying as self-regulated learning," in *Metacognition in Educational Theory and Practice*, D. J. Hacker, J. Dunlosky, and A. C. Graesser, Eds., Mahwah, NJ, USA: Lawrence Erlbaum Associates Publishers, 1998, pp. 277–304.

[21] C. Was, "Academic achievement goal orientation: Taking another look," *Electron. J. Res. Educational Psychol.*, vol. 4, no. 3, pp. 529–550, Dec. 2006, doi: 10.25115/ejrep.v4i10.1209.

[22] E. Trumbull and A. Lash, "Understanding formative assessment. Insights from learning theory and measurement theory," WestEd.org, San Francisco, CA, USA, 2013, pp. 1–20, [Online]. Available: https://www2.wested.org/www-static/online_pubs/resource1307.pdf

[23] I. J. M. Arnold, "Cheating at online formative tests: Does it pay off?," *Internet Higher Educ.*, vol. 29, pp. 98–106, Apr. 2016, doi: 10.1016/j.iheduc.2016.02.001.

[24] T. Wolsey, "Efficacy of instructor feedback on written work in an online program," *Int. J. E-Learn.*, vol. 7, no. 2, pp. 311–329, Apr. 2008.

[25] G. Alexandron, M.E. Wiltrout, A. Berg, and J.A. Ruipérez-Valiente, "Assessment that matters: Balancing reliability and learner-centered pedagogy in MOOC assessment," in *Proc. 10th Int. Conf. Learning Analytics and Knowledge (LAK'20)*, Frankfurt, Germany (virtual), Mar. 23–27, 2020, pp. 512–517, doi: 10.1145/3375462.3375464.

[26] S. Schut, J. van Tartwijk, E. Driessen, C. van der Vleuten, and S. Heeneman, "Understanding the influence of teacher–learner relationships on learners' assessment perception," *Advances Health Sciences Educ.*, vol. 25, pp. 441–456, May 2020, doi: 10.1007/s10459-019-09935-z.

[27] S. G. Paris and A. H. Paris, "Classroom applications of research on self-regulated learning," *Educational Psychologist,* vol. 36, no. 2, pp. 89–101, 2001, doi: 10.1207/S15326985EP3602_4.

[28] E. Panadero, A. Jonsson, and J. Botella, "Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses," *Educational Res. Rev.*, vol. 22, pp. 74–98, Nov. 2017, doi: 10.1016/j.edurev.2017.08.004.

[29] E. Panadero and J. Alonso-Tapia, "Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students," *Electron. J. Res. Educational Psychol.*, vol. 11, no. 2, pp. 551–576, Jul. 2013, doi: 10.14204/ejrep.30.12200.

[30] P. H. Winne and A. F. Hadwin, "nStudy: Tracing and supporting self-regulated learning in the internet," in *International Handbook of Metacognition and Learning Technologies*, R. Azevedo and V. Aleven, Eds., New York, NY, USA: Springer-Verlag, 2013, pp. 293–308.

[31] R. Azevedo, A. Johnson, A. Chauncey, and C. Burkett, "Self-regulated learning with MetaTutor: Advancing the science of learning with MetaCognitive tools," in *New Science of Learning*, M. Khine and I. Saleh, Eds., New York, NY, USA: Springer-Verlag, 2010, ch. 11, pp. 225–247.

[32] R. Pérez-Álvarez, J. J. Maldonado-Mahauad, D. Sapunar-Opazo, and M. Pérez-Sanagustín, "NoteMyProgress: A tool to support learners' self-regulated learning strategies in MOOC environments," in *Proc. 12th European Conf. Technology Enhanced Learning (EC-TEL'17)*, Tallinn, Estonia, Sep. 12–15, 2017, in Lecture Notes Comput. Sci., vol. 10474, Sep. 2017, pp 460–466, doi: 10.1007/978-3-319-66610-5_43.

[33] J. Maldonado-Mahauad, M. Pérez-Sanagustín, R. F. Kizilcec, N. Morales, and J. Munoz-Gama, "Mining theory-based patterns from big data: Identifying self-regulated learning strategies in massive open online courses," *Comput. Human Behav.,* vol. 80, pp. 179–196, Mar. 2018, doi: 10.1016/j.chb.2017.11.011.

[34] R. Pérez-Álvarez, J. Maldonado-Mahauad, and M. Pérez-Sanagustín, "Tools to support self-regulated learning in online environments: Literature review," in *Proc. 13th European Conf. Technology Enhanced Learning (EC-TEL'18)*, Leeds, United Kingdom, Sep. 3–6, 2018, *in Lecture Notes Comput. Sci.*, vol. 11082, Sept. 2018, pp. 16–30, doi: 10.1007/978-3-319-98572-5_2.

[35] S. N. Robertson, S. M. Humphrey, and J. P. Steele, "Using technology tools for formative assessments," *J. Educators Online,* vol. 16, no. 2, Jul. 2019.

[36] S. E. M. Meek, L. Blakemore, and L. Marks, "Is peer review an appropriate form of assessment in a MOOC? Student participation and performance in formative peer review," *Assessment Eval. Higher Educ.*, vol. 42, no. 6, pp. 1000–1013, 2017, doi: 10.1080/02602938.2016.1221052.

[37] E. Jung, D. Kim, M. Yoon, S. Park, and B. Oakley, "The influence of instructional design on learner control, sense of achievement, and perceived effectiveness in a supersize MOOC course," *Comput. & Educ.*, vol. 128, pp. 377–388, Jan. 2019, doi: 10.1016/j.compedu.2018.10.001.

[38] L. Benotti, M. C. Martínez, and F. Schapachnik, "A tool for introducing computer science with automatic formative assessment," *IEEE Trans. Learn. Technol.*, vol. 11, no. 2, pp. 179–192, Apr.-Jun. 2018, doi: 10.1109/TLT.2017.2682084.

[39] M. Cvetanovic, Z. Radivojevic, V. Blagojevic, and M. Bojovic, "ADVICE—Educational system for teaching database courses," *IEEE Trans. Educ.*, vol. 54, no. 3, pp. 398–409, Aug. 2011, doi: 10.1109/TE.2010.2063431.

[40] R. Luckin, W. Clark, K. Avramides, J. Hunter, and M. Oliver, "Using teacher inquiry to support technology-enhanced formative assessment: A review of the literature to inform a new method," *Interactive Learn. Environments*, vol. 25, no. 1, pp. 85–97, 2017, doi: 10.1080/10494820.2015.1121152.

[41] T. McLaughlin and Z. Yan, "Diverse delivery methods and strong psychological benefits: A review of online formative assessment," *J. Comput. Assisted Learn.*, vol. 33, no. 6, pp. 562–574, Dec. 2017, doi: 10.1111/jcal.12200.

[42] N. H. Nadeem and H. A. Al Falig, "Kahoot! quizzes: A formative assessment tool to promote students' self-regulated learning skills," *J. Appl. Linguistics Lang. Res.*, vol. 7, no. 4, pp. 1–20, Jun. 2020. [Online]. Available: https://www.jallr.com/index.php/JALLR/article/view/1117/pdf1117

[43] Y. Manzhen and L. Li, "A longitudinal study of self-regulated learning in an online game-based EFL formative assessment blended learning environment," *Can. Social Sci.,* vol. 13, no. 7, pp. 35–41, Jul. 2017, doi: 10.3968/9749.

[44] W. Matcha et al., "Detection of learning strategies: A comparison of process, sequence and network analytic approaches," in *Proc. 14th European Conf. Technology Enhanced Learning (EC-TEL'19)*, Delft, The Netherlands, Sep. 14–18, 2019, in *Lecture Notes Comput. Sci.*, vol. 11722, Sep. 2019, pp. 525–540, doi: 10.1007/978-3-030-29736-7_39.

[45] A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining," *IEEE Access*, vol. 5, pp. 15991–16005, Jan. 2017, doi: 10.1109/ACCESS.2017.2654247.

[46] A. Bogarín, R. Cerezo, and C. Romero, "A survey on educational process mining," *WIREs: Data Mining Knowl. Discovery*, vol. 8, no. 1, Jan. 2018, doi: 10.1002/widm.1230.

[47] M. A. Ghazal, O. Ibrahim, and M. A. Salama, "Educational process mining: A systematic literature review," in *Proc. 2017 European Conf. Electrical Engineering and Computer Science (EECS'17),* Bern, Switzerland, Nov. 17–19, 2017, pp. 198–203, doi: 10.1109/eecs.2017.45.

[48] S. A. Priyambada, E. R. Mahendrawathi, and B. N. Yahya, "Curriculum assessment of higher educational institution using aggregate profile clustering," *Procedia Comput. Sci.*, vol. 124, pp. 264–273, 2017, doi: 10.1016/j.procs.2017.12.155.

[49] S. S. Beheshitha, D. Gašević, and M. Hatala, "A process mining approach to linking the study of aptitude and event facets of self-regulated learning," in *Proc. 5th Int. Conf. Learning Analytics and Knowledge (LAK'15),* Poughkeepsie, NY, USA, Mar. 16–20, 2015, pp. 265–269, doi: 10.1145/2723576.2723628.

[50] R. Elmasri and S. Navathe, *Fundamentals of Database Systems, 7th ed.* Pearson, 2016.

[51] M. de Raadt, S. Dekeyser, and T. Y. Lee, "A system employing peer review and enhanced computer assisted assessment of querying skills," *Inform. Educ.*, vol. 6, no. 1, pp. 163–178, Apr. 2007, doi: 10.15388/infedu.2007.11.

[52] S. Sadiq, M. Orlowska, W. Sadiq, and J. Lin, "SQLator: An online SQL learning workbench," *ACM SIGCSE Bull.*, vol. 36, no.3, p. 223–227, Sep. 2004, doi: 10.1145/1026487.1008055.

[53] M. Misic, Z. Siustran, and J. Protic, "A comparison of software tools for plagiarism detection in programming assignments," *Int. J. Eng. Educ*., vol. 32, no. 2, pp. 738–748, Jan. 2016.

[54] J. Carmona, J. Cortadella, and M. Kishinevsky, "Divide-and-conquer strategies for process mining," in *Proc. 7th Int. Conf. Business Process Management (BPM'09)*, Ulm, Germany, Sep. 8–10, 2009, in Lecture Notes Comput. Sci., vol. 5701, Sep. 2009, pp. 327–343, doi: 10.1007/978-3-642-03848-8_22.

[55] J. C. A. M. Buijs, "Mapping data sources to XES in a generic way," M.S. thesis, Dept. Math. Comput. Sci., Eindhoven Univ. of Technol., Eindhoven, The Netherlands, 2010. [Online]. Available: https://research.tue.nl/files/46977079/676586-1.pdf

[56] R. P. J. C. Bose, R. Jagadeesh, and W. M. P. van der Aalst, "Process diagnostics using trace alignment: Opportunities, issues, and challenges," *Inf. Syst.*, vol. 37, no. 2, pp. 117–141, Apr. 2012, doi: 10.1016/j.is.2011.08.003.

[57] J. W. Gikandi, D. Morrow, and N. E. Davis, "Online formative assessment in higher education: A review of the literature," *Comput. & Educ.*, vol. 57, no. 4, pp. 2333–2351, Dec. 2011, doi: 10.1016/j.compedu.2011.06.004.

[58] C. A. Was and T. L. R. Beziat, "Exploring the relationships between goal orientations, knowledge monitoring and academic achievement," *J. Educ. and Human Develop*., vol. 4, no. 3, pp. 67–77, Sep. 2015, doi: 10.15640/jehd.v4n3a8.

[59] H. Li, B. Flanagan, S. Konomi, and H. Ogata, "Measuring behaviors and identifying indicators of self-regulation in computer-assisted language learning courses," *Res. Pract. Technol. Enhanced Learn.*, vol. 13, Dec. 2018, pp. 1–12, Art. No. 19, doi: 10.1186/s41039-018-0087-7.

[60] C. Grunschel, J. Patrzek, K. B. Klingsieck, and S. Fries, "I'll stop procrastinating now! Fostering specific processes of self-regulated learning to reduce academic procrastination," *J. Prevention Intervention Community*, vol. 46, no. 2, pp. 143–157, Apr-Jun. 2018, doi: 10.1080/10852352.2016.1198166.

[61] J. A. Arter, "Interim benchmark assessments: Are we getting our eggs in the right basket?," presented at the Nat. Council Meas. Educ., Denver, CO, USA, May 1–3, 2010.


**César Domínguez** received the Ph.D. degree in mathematics from the University of La Rioja, Logroño, Spain, in 2003. He is currently an associate professor with the Department of Mathematics and Computing, University of La Rioja. His research interests include e-learning, innovative ways to teach engineering concepts, and learning technologies.

**Francisco J. García-Izquierdo** received the Ph.D. degree in telecommunications engineering from University of Zaragoza, Zaragoza, Spain, in 1999. He is currently an associate professor with the Department of Mathematics and Computing, University of La Rioja, Logroño, Spain. His research interests include modeling theories, and innovative ways to teach engineering concepts.

**Arturo Jaime** received the Ph.D. degree in computer science from the University of the Basque Country, Spain, in 1999. He is currently an associate professor with the Department of Mathematics and Computing, University of La Rioja, Logroño, Spain. His research interests include e-learning and innovative ways to teach engineering concepts.

**Beatriz Pérez** received the Ph.D. degree in computer science from the University of Zaragoza, Zaragoza, Spain, in 2011. She is currently an associate professor with the Department of Mathematics and Computing, University of La Rioja, Logroño, Spain. Her main research interests include modeling theories, process mining, and provenance.

**Angel Luis Rubio** received the Ph.D. degree in computer science from the University of La Rioja, Logroño, Spain, in 2002. He is currently an associate professor with the Department of Mathematics and Computing, University of La Rioja. His current research interests include metamodeling, model evolution, process mining, and performance assessment in the development of information systems.

**María A. Zapata** received the Ph.D. degree in computer science from the University of Zaragoza, Zaragoza, Spain, in 1994. She is currently an associate professor in computer science with the Faculty of Science, University of Zaragoza. Her main research interests are in metamodeling, system evolution, and process mining.