International Journal of Engineering Education Vol. 36, No. 1(B), pp. 470–482, 2020
Printed in Great Britain
0949-149X/91 $3.00+0.00

Teamwork Assessment in Collaborative Projects Through Process Mining Techniques*

JUAN ANTONIO CABALLERO-HERNÁNDEZ
EVAL for Research group, University of Cadiz, Av. República Árabe Saharaui s/n, 11519, Puerto Real, Spain.
E-mail: juanantonio.caballero@uca.es

ANTONIO BALDERAS, MANUEL PALOMO-DUARTE, PABLO DELATORRE
Department of Computer Science, University of Cadiz, Av. de la Universidad de Cádiz 10, 11519, Puerto Real, Spain.
E-mail: antonio.balderas@uca.es, manuel.palomo@uca.es, pablo.delatorre@uca.es

ANTONIO J. REINOSO
Department of ICT Engineering, Alfonso X El Sabio University, Av. de la Universidad 1, 28691, Vva. De la Cañada, Spain.
E-mail: areinpei@myuax.com

JUAN MANUEL DODERO
Department of Computer Science, University of Cadiz, Av. de la Universidad de Cádiz 10, 11519, Puerto Real, Spain.
E-mail: juanma.dodero@uca.es

Teamwork is one of the key issues in engineering projects success. Unfortunately, due to the high number of interactions, the assessment of collaborative tasks remains a challenge. Wikis are web-based systems that support collaborative work in enterprise engineering project documentation providing quantitative data from the members' contributions and interactions. While this objective data is interesting for the teamwork assessment, a qualitative assessment process can provide a complementary approach. We propose an architecture that combines information from both sources to conduct a scalable assessment of the teamwork in a wiki. It was implemented in a course of the degree on Computer Engineering using specific developed software tools and Process Mining techniques. Process Mining tools automatically apply artificial intelligence algorithms to extract knowledge from real processes and discover models. These models provided evidence of conducted behaviour. The actual dynamics of the teams in the wiki were automatically detected and could be analysed for assessment purposes. Finally, the followed mixed approach allowed a detailed and scalable teamwork skills assessment process.

Keywords: teamwork assessment; quantitative assessment; qualitative assessment; wikis; collaborative learning; process mining

1. Introduction

Engineering projects usually require different workers with varied profiles to effectively collaborate as a team to reach a common goal that meets stakeholders' expectations. As a consequence, teamwork is one of the generic skills that academia is focusing in recent years [1]. This focus promoted the inclusion of assignments that require teamwork competence in course syllabuses. Unfortunately, the assessment process can be more complex in collaborative assignments, due to the complex nature of the process with many interactions [2].

Wikis are web-based systems that support collaborative writing, providing a digital platform for asynchronous collaboration between peers. They can be enriched with additional features and integrated into larger ecosystems like Knowledge Management Systems [3]. As a result, they are widely used in enterprise engineering project documentation. A wiki does not only store the final document version of each page but also all the previous ones that resulted from every individual contribution made. This historical information is interesting not only for the contributors that can retrieve previously discarded information but also for assessment purposes by analysing the development process [4].

When it comes to assessing skills learned in a digital task, there are two complementary approaches, quantitative and qualitative [5]. On one hand, quantitative approaches, given proper access to the data stored in the system database, can automatically retrieve different objective metrics. The main advantage of quantitative techniques is that they are based on objective data that can scale to large groups. Unfortunately, they are rather coarse-grained, so different interesting aspects regarding the contributions to the wiki can be dismissed [6]. On the other hand, the complementary approaches are those based on qualitative assessment [7]. They can offer a more detailed assessment process, usually based on human intervention. Unfortunately, the human effort needed can lead to scalability problems when the size of the project or the group increases.

Collaborative projects based on wikis are struc-

* Accepted 19 November 2019.

tured as sequential contributions made by the different team members. We considered Sequence Analysis as a suitable alternative to alleviate scalability problems [8]. A set of Sequence Analysis techniques that could achieve our goals are Process Mining techniques because they automatically extract knowledge about a process from its event logs. In our case, we used model discovery to obtain the actual dynamic of the teams in the wiki without an a priori model.

Based on our previous experience, using both methods as complementary sources of information [6], we propose an architecture to conduct a sustainable assessment of wiki collaborative assignments by integrating data from them. The data retrieval and analysis process is supported by open-source software systems which present successful experiences in collaborative environments. First, StatMediaWiki was used to collect and aggregate quantitative information in a MediaWiki database, providing a general picture of the wiki [9]. Then, AssessMediaWiki assisted the qualitative assessment procedure implementing a scalable fine-grained qualitative assessment of wiki contributions using rubrics [6]. StatMediaWiki and AssessMediaWiki provide data sources to follow a mixed approach: quantitative and qualitative, respectively. Finally, we applied a Process Mining model discovery using ProM, an open source Process Mining platform [10]. ProM has been widely applied to real-life applications and presents multiple plugins (more than 600), providing a diverse set of techniques to conduct detailed data analysis.

We implemented our proposal in a case study developed in a course of the degree on Computer Engineering in the University of Cadiz (Spain). In the course, students had to work in groups to document their project in a wiki. Using our approach, the teamwork of the wiki could be easily modelled, showing a different dynamic when the quantitative information was enriched with that from a qualitative source.

The rest of the paper is organized as follows. Related works are analysed in section 2. Then, in section 3, we describe the experiment, including the research question, the methodology followed, the description of applied tools and the conducted case study. A discussion according to the results obtained is provided in section 4. Finally, some conclusions and future research lines are presented in section 5.

2. Related Works

As previously commented, wikis are widely used in learning processes [11], becoming a valuable tool for the improvement and enrichment of learning experiences. Their assessment use similar approaches than those in other collaborative assignments, like Version Control Systems [12]. We can find studies measuring different skills. First, an analysis is conducted about how the collaborative construction of an online project facilitates the acquisition of digital skills in [13]. This collaborative construction is carried out by a wiki as the main tool used to address the project's creation. An additional study on the development of digital skills using online tools like wikis is carried out in [14]. They conclude that the use of these tools contributes mainly to the development of two areas of digital competence: communication and content creation. Finally, the impact of using wikis to develop academic skills in critical thinking and collaborative work for higher education students is evaluated in [15]. According to this study, wikis facilitate the learning process, allowing for the discovery of new research-informational skills with the assistance of university libraries and enhancing critical thinking and collaborative work in the classroom.

Conversely, there are failed wiki experiences. In [16], a frustrating experience is presented involving students from the University of Girona that used wikis to develop collaborative research projects. They conducted a qualitative analysis of gathered data from online questionnaires and in-person meetings with students. Relevant conclusions were obtained. First, less than 5% of students had previously worked with wikis. Then, the majority of students preferred not to repeat the experience of working with wikis or to participate in collaborative projects. The main frustrations that appeared after working with wikis were the lack of time, lack of training and the scarce equality in the students' assessments. Therefore, a negative perception was concluded for collaborative works, specifically for those which include the mandatory use of wikis.

Wikis can be used as standalone systems or integrated in more complex systems, usually Learning Management Systems (LMS). In [17], the contributions to different activities (pages, comments and evaluations) in a wiki for teacher learning are assessed. In [18], a wiki was used for several years as the main support for an online higher education course. The case study analysed questionnaires filled by students about the developed skills and their expertise. Results showed that the collaboration between participants was less than expected. It was visible in the low amount of contribution to pages of other peers and scarce use of "Talk" pages for communication and coordination.

A case study where a skill assessment system in LMS is applied over a wiki is presented in [19]. This system is a group of decoupled web applications which interact between them. First, the system integrates a web service for designing assessment tools with a Restful web service used for managing skills and learning outcomes. Second, an extension provides additional functionality to mark evidence from web pages. Finally, all applications are integrated with a LMS. In the case study, this system is used to assess a collaborative project carried out by students using a wiki. Several skills related to wikis are defined and rated using an adapted assessment tool. Then, evidence of assessments were marked from the wiki. Finally, a report with grades and the wiki's evidence is displayed inside the LMS.

Additional teamwork assessment experiences with wikis have been detected in the literature. In [20, 21], indicators monitoring and learning analytics are used to apply CTMTC (Comprehensive Training Model of the Teamwork Competence). CTMTC is a teamwork assessment methodology focused on components such as leader behaviour, cooperation between peers, problems between team members and performance of each member. These studies are expanded in [22], where the learning analytics tool is applied in different contexts. In the conducted experiment, the presented tool is connected to Moodle LMS to obtain data from two different courses. Student used wikis to publish partial outcomes. Another case study of this tool is carried out in [23], including functionalities to label the student's learning evidence.

Wikis can present scalability problems when the number of students increases because of manual evaluations of the information stored in a wiki to retrieve objective metrics becomes a complex and time-consuming task [24]. Therefore, specific tools appear to overcome this kind of limitation and focus on retrieving quantitative information from wikis and supporting the assessment. In [25], EvalCourse, an open-source tool, is applied during two successive courses in order to obtain indicators for generating reports for generic skills. EvalCourse is an assessment support environment that executes queries written in SASQL, a Domain Specific Language (DSL) to tackle the complexity of customizing online learning assessments. It allows one to carry out simple queries to obtain information from the wiki. In [9], a study about the assessment of wiki contributions in a collaborative learning experience is conducted using StatMediaWiki. This tool is focused on quantitative analysis; it allows one to analyse users' performance and quantify the contributions carried out. StatMediaWiki provides the distribution of contributions between different users and an accurate description with information like number of pages, users, contributions or uploaded files. In addition, it provides wiki's temporal evolution through graphs with added bytes and carried out actions.

Finally, besides integrations with other systems or specific tools, other relevant experiences about wikis are described below. A learning-oriented collaborative assessment method supported by an open data framework is described in [26]. In addition, an architecture for the extraction of different indicators to facilitate the assessment process is presented. The assessment method and the open data framework are applied to a project-based course on web engineering. Projects were developed using an open-source forge which includes a task management tool for planning and monitoring, a repository of the version control system to store the software contributions and a wiki to store text contributions. The experience provided positive evidence because the grade measurement was backed up with assessment evidence and calculated with less effort.

Process Mining has been widely applied in educational environments [27, 28]. In [29], an analysis to discover bottlenecks in a Higher Education Degree is conducted. Frequencies of passed courses by students are analysed through Process Mining techniques to solve scalability problems of manual analysis. Another case of a study in Higher Education can be found in [30], where Process Mining techniques are used to periodically produce automated reports that relate the students' performance to their studying behaviour at the Eindhoven University of Technology. Then, in [31], a practical tutorial about how to apply clustering and Process Mining to Moodle data using open-source tools like ProM is carried out.

Successful studies about applying Process Mining to collaborative education projects are presented in the literature. In [32], Process Mining techniques are applied to discover and compare distinguished patterns of interaction and involvement between the student groups with high and low performance. Communications regarding textual and semantic contributions of the students in chat rooms during online distance activity are analysed as well. Then, in [33], three software processes are analysed through Process Mining techniques in the context of an undergraduate Software Engineering course: requirement engineering, development and maintenance, and issue tracking. In [34], Process Mining is used to explore sequences of social regulatory processes during a computer-supported collaborative learning task and their relationship to group performance. We found in [35] a quantitative survey is conducted to identify the most significant indicators affecting the collaboration process. In addition, several Process Mining techniques are applied with the purpose of increasing the instructor's knowledge about the collaborative dynamics in each group of students.

In essence, wikis have been mainly applied to collaborative learning processes as standalone services or integrated with additional systems, usually LMS. Although literature includes successful studies for measuring skills in wikis, failed experiences can also be found. Additionally, wikis can present scalability problems in the assessment process when the quantity of students increases and manual assessment becomes too complex. Specific tools focused on retrieving quantitative data appear to overcome this limitation and support the assessment. Process Mining has been used to solve scalability problems of manual analysis in education environments. Multiple studies can be found in the literature where Process Mining is successfully applied to collaborative education projects and diverse sources of data are used. Therefore, Process Mining techniques could be considered as a promising approach to address scalability problems of wikis using different sources of data.

3. Experiment Description

In this section, we define the research question of our study. Then, we describe the applied methodology, which includes the proposed architecture and the included tools. Finally, we explain the analysis for our case study.

3.1 Research Question

The Research Question (RQ) is: Can Process Mining techniques support a scalable teamwork assessment in a collaborative experience based on a wiki using different sources of data? Specifically, the data will be objective data from the wiki (i.e. the number of contributions to the wiki and the kind of wiki content that received those contributions) and grades resulting from a qualitative peer-assessment process.

3.2 Methodology

This work has been conducted by an Action-Research methodology. According to Argyris [36], "Action-Research takes its cues – its questions, puzzles and problems – from the perceptions of practitioners within particular, local practice contexts. It builds descriptions and theories within the practice context itself, and tests them through intervention experiments, that is, Action-Research through experiments that bear the double burden of testing hypotheses and effecting some (putatively) desirable change in the situation".

In essence, this methodology aims to improve an aspect of the research focus [37]. This goal of improvement is directed towards three areas: practice, the understanding of the practice by its practitioners, and the improvement of the situation where the practice takes place [38]. This goal can be met by examining actions carried out against the original hypotheses. The theory must solve a practical problem and generate knowledge within the context, in this case, the assessment process. To this end, different software tools have been combined to support the process: StatMediaWiki, AssessMediaWiki and ProM. As a result, this proposes an architecture to apply Process Mining techniques to quantitative and qualitative data sources.

3.3 Architecture

Implemented architecture is shown in Fig. 1.

StatMediaWiki is a tool to collect and aggregate quantitative information available in MediaWiki installation, providing user, page and category reports [39]. This tool allows one to analyse users' performance and quantify the contributions carried

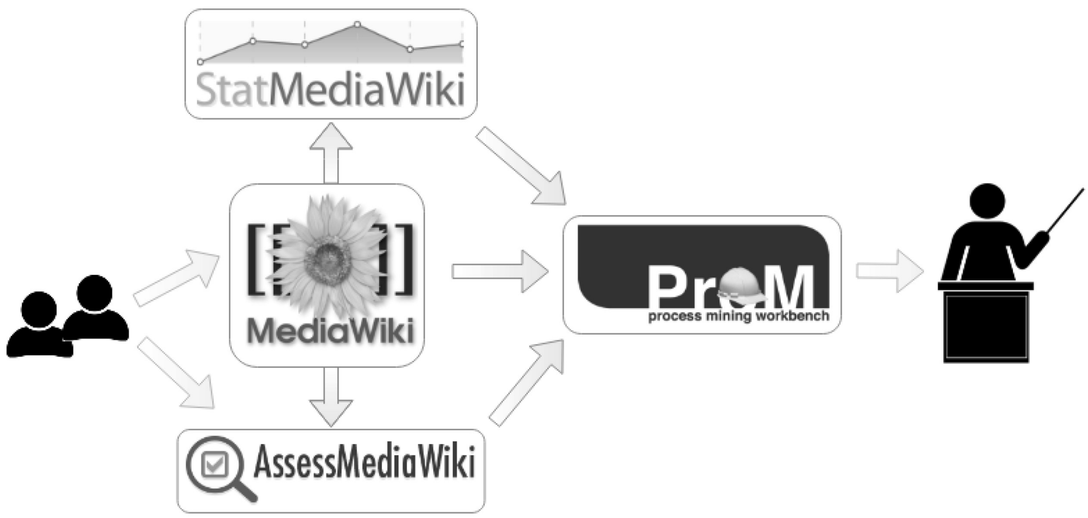

Fig. 1. Implemented Architecture (logo copyrights belong to their owners).

out according to different time periods. Weekly contribution reports correspond to the quantitative data applied in our analysis.

AssessMediaWiki is a web application that enables hetero-, self- and peer-to-peer assessment procedures, while keeping track of the compiled assessment data which works on wikis built on MediaWiki installations [40]. The assessing student uses a rubric defined by the supervisor to evaluate different aspects/skills developed in the specific wiki contribution. The rubric shows the wiki contribution and a form to assess the contribution with a numeric grade from 0 (minimum) to 10 (maximum) and a textual explanation. These grades correspond to the qualitative data applied in our analysis.

Additional information like timestamps and user-related data is taken directly from the MediaWiki database and used to process quantitative and qualitative data. This data defined a set of events to be imported in ProM, an open source framework for Process Mining algorithms [10]. ProM provides a platform which allows users to apply Model Discovery and others Process Mining techniques to event logs used as inputs. ProM also provides filtering tools that were used to refine models for each teamwork. These models were provided to the supervisor to assess the behaviour of the teams in the case study.

3.4 Case Study

We implemented our proposal in a case study developed in a course offered in the degree on Computer Engineering in the University of Cadiz (Spain) during the second semester. The course, with 22 students enrolled, aimed to introduce the benefits of functional programming to students in their fifth year (last) of the degree programme. In the course, students were grouped into teams of three members to carry out a collaborative project that was documented in a MediaWiki wiki.

Students were allowed to contribute to the wiki at any time during the semester. The project was mastering a library for the Haskell programming language and documenting the process. These libraries were scarcely documented, so they had to read different sources of information to learn about their installation, then learn to use it (usually integrating it with more abstract systems), create examples that demonstrate its capabilities and document the lessons learned. Each team had its own page where all the members made multiple contributions. Students were also encouraged to contribute to other team pages. Besides team pages, talk pages could be created by students to discuss the teamwork, set out questions, propose improvements, etc. At the end of the semester, when the project was finished, students peer-assessed the work done in the wiki using AssessMediaWiki.

3.4.1 Process Mining

We conducted a Process Mining analysis using the ProM tool. ProM implements an algorithm for the discovery of models that reflect the dynamic in an event log. Using Model Discovery, the wiki logs are analysed to provide a model for each team behaviour in terms of a Petri Net [41].

There are various metrics for measuring the quality of Process Mining results [42]. In this study, we focus on the analysis of Fitness and Precision metrics. First, a model with good Fitness allows for most of the behaviour seen in the event log, so the model is a faithful representative of the analysed behaviour. In addition, we can consider that a model has a perfect Fitness if all traces in the log can be replayed by the model from beginning to end. Second, a model is precise if it allows for the "proper" behaviour: same and similar sequences as the one presented in the log. A model with a low Precision is considered to be an "underfitting" model: an over-generalized model that allows for behaviours very different from what was seen in the log.

The Applied Model Discovery technique provided objective values for Fitness and Precision metrics, used in our analysis. In addition, it allowed us to control the ratio of paths used in the log to confirm that the logs with lower frequencies were not discarded. Therefore, we considered it to be a valid technique to measure the models' accuracy.

3.4.2 Data Processing

We processed the data from three different sources: StatMediaWiki, AssessMediaWiki and the MediaWiki database. First, StatMediaWiki provided the quantitative data: the number of weeks, type and quantity of the contributions carried out by the students in the wiki. Second, we obtained qualitative data from AssessMediaWiki: peer-assessed grades per contribution. Finally, we used Media-

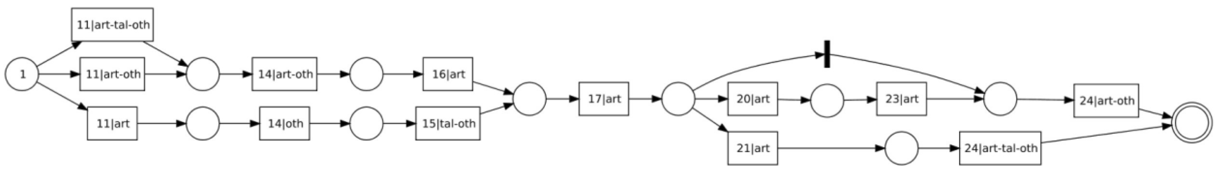

Fig. 2. Model for Crypto team (quantitative data).

Wiki to combine all these data into a single data set. The most relevant information for our case of study is explained below:

- week: number of weeks in the year,
- edits_type: type of contribution made by the student,
- edits_quantity: quantity of contributions made in the related week,
- edits_qualitative: grade obtained by the peer-assessment.

Weeks represent the number of weeks where the contribution was provided. In our data set, weeks belong to the second semester, from the 11th to 24th week. Types of contribution show the pages of the wiki where the contributions were made during a specific week: articles (own group pages), talks (discussion pages), others (other group pages) or a combination of these.

Contribution numeric data (edits_quantity and edits_qualitative) presented wide intervals of values, so we discretized them to group similar values and, in this way, avoid obtaining models with a large quantity of similar states and transitions. First, edits_quantity was processed to include the following values: min (minimum value of the data set), low (less than the average quantity), high (higher than the average quantity) and max (maximum value of the data set). These values were calculated according to the type of contribution. For instance, the average grade in article page contributions is different than average grades for talk page contributions. Second, edits_qualitative was discretized from lower grades to higher grades: C, B and A.

3.4.3 Analysis

After importing the data set in ProM, we aimed to carry out two different analyses to conduct a scalable teamwork assessment. First, we followed a quantitative approach to assess the weekly teamwork distribution of students considering their contributions. Second, we followed a mixed approach enriching quantitative data with qualitative assessment to assess how the teams contributed to the wiki. It should be noted that we reference two team's work by their alias in analysis and discussion: Crypto and Simple Latex.

Results for Fitness and Precision metrics show two well differentiated types of teams for the weekly teamwork distribution analysis. On one hand, three teams present a high Fitness (from 0.89 to 1.00) and a low Precision (from 0.08 to 0.12). Discovered models for these teams are over-generalized because they need to include a wide variety of paths. This can be taken as evidence that the members of those teams followed a heterogeneous distribution of the work. On the other hand, the other three teams present more intermediate values both for Fitness (from 0.61 to 0.65) and Precision (from 0.76 to 0.81). Models for those teams present a more linear structure and more defined paths. This can be taken as evidence that the members of those teams followed a more homogeneous distribution of the work than the first group of teams. It should be noted that obtained values for Fitness in all models (at least 0.61) are high enough to consider them as faithful representatives of the teamwork behaviour.

As in the weekly work distribution analysis, Fitness values for the mixed approach assessment are high enough (at least 0.84) to consider that the models are faithful representatives of the teamwork behaviour. Then, almost all teams present intermediate values (between 0.51 and 0.63) for the Precision metric except Simple Latex (0.29). These values can be taken as evidence that the majority of the teams involve a balanced heterogeneous/homogeneous level in the mixed assessments while only Simple Latex team members present more heterogeneous mixed assessments among them.

The most illustrative results are included and discussed in the section below. However, detailed anonymized results for all teams are openly available at figShare [43].

4. Discussion

In this section, we discuss the results obtained by applying Process Mining techniques to the data set. Firstly, we discuss the weekly teamwork distribution based on analysing two dimensions of the quantitative data: the week of the contributions and their types. Secondly, we followed a mixed approach to enrich the previous data. Results and differences are discussed.

4.1 Weekly Teamwork Distribution

Figs. 2 and 3 show the discovered models for two teams with illustrative results for weekly teamwork distribution. Labels of activities indicate the combined dimensions for this analysis: week number and, after a vertical bar, the type of contributions made during that week. First, each week is represented with the numeric value of the week in the year. Second, the types of contributions are presented with acronyms: art (articles), tal (talks) and oth (others). Weeks with more than one type of contribution were labeled joining those types by a hyphen character ('–'). Next, we discuss the obtained models.

4.1.1 Crypto Team

This model (Fig. 2) provided an acceptable value for the Precision metric (0.81) and a low quantity of

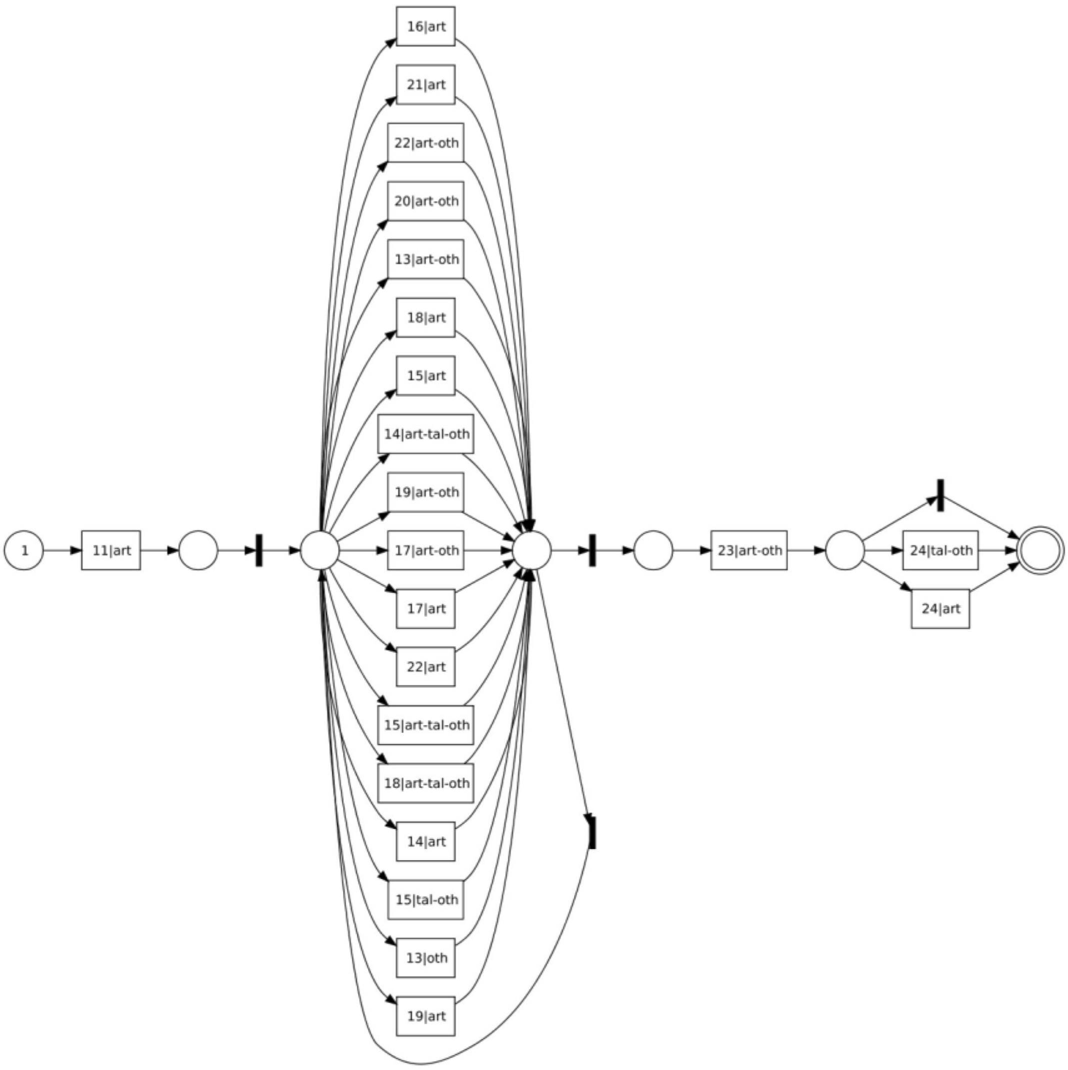

Fig. 3. Model for Simple-Latex team (quantitative data).

possible activities, 13 combinations for week + type of contribution. Therefore, we can consider that the members of this team showed quite homogeneous behaviour according to the work distribution.

The model presents a bifurcation at the beginning with similar states: in the first week, the three members contributed to their articles and two of them also contributed to articles of other groups. This can be taken as evidence that all team members started at the same week and made similar types of contributions, so there was no clear leader in the team. Later, two of the members contributed to an article page while the other one only added contributions to other teams and talks pages, following an independent path until the 17th week, while the rest of the team members continued with article contributions. Interestingly, that member only contributed to articles (not to talk of others) in the first week, so we could consider that this is a member with limited teamwork skills, not able to contribute. We can see that this member conducted different functions than the rest of the team members during those weeks.

In the second half of the model, we can see other bifurcation which reflects that one team member did not contribute for almost two months (from the 17th to 24th week). Therefore, teamwork was focused on article contributions by the other two members from the 17th to 23rd week. In fact, in the 17th week, all the members agreed to work on the teams' article. Two members alternated their contributions to the articles page from the 20th to 23rd week: one member contributed at the 20th week, the other one at the 21st, and then again the first one at the 23rd. This can be taken as evidence of how one of these members made contributions that the other needed to progress. As in the beginning, all the team members ended at the same week (24th) with similar contributions, complementing contributions in the article page with talk and others. It is likely that team members aimed to homogenize the distribution of the pending work.

Apart from these two bifurcations, we observe that the model is quite linear: it does not present any loops. In general, this linearity could also imply organized and properly planned teamwork.

4.1.2 Simple Latex Team

The obtained model (Fig. 3) provides different values for Fitness and Precision metrics than the previous one: 0.89 and 0.11 respectively. The low value for Precision of this over-generalized model can be considered as evidence of a very different weekly distribution among team members. The wide variety of sequences are reflected in the model through the central loop. Models with this structure are commonly known in the literature as "flower models".

Additional evidence on this heterogeneity are the amount of activities and the frequency of them. The model includes 22 different combinations of week and type of contribution, one of the highest values of the analysed teams. The highest frequency for an activity is 9.09%, which corresponds to only three occurrences. Grouping these activities by their week, we can see that at least one member provided contributions in almost all the weeks. This is evidence of the constant work carried out by this team during the semester, probably leading the teamwork from the beginning.

Besides article contributions, contributions in others (combined with articles and/or talks) are also made during most of the weeks. However, talk pages are gathered at some specific weeks: at the beginning (14 and 15), in the middle (18) and the last week (24). Therefore, this team regularly contributed in other team's pages but barely discussed in coordination talk pages. Team coordination was probably carried out by the member who led the work distribution.

Finally, we can observe in the model that all the team members started contributing into the article page of the team at the same week (11). However, the end of the model presents a high heterogeneity because it finishes with three different kind of contributions, article, talk and others, which reflects that the team members ended with different tasks. This implies three different roles at the end of the semester. One member probably ended his work in the previous weeks and the others distributed their pending work: one contributed in the article page and the other provided contributions in talk and others pages.

4.2 Mixed Approach Assessment

Figs. 4 and 5 show the discovered models for two illustrative teams with results for the mixed approach assessment. Labels of activities show the combined dimensions for this analysis: the quantity of the contribution and, after a vertical bar, its grade (qualitative assessment). First, quantities are classified using min, low, high and max. Second, grades are A, B and C.

4.2.1 Crypto Team

In this model (Fig. 4), we can appreciate an intermediate model between a linear model (without loops or a variety of paths) and an over-generalized "flower model" where all paths are possible. This structure probably reflects an intermediate level of heterogeneity between the quantitative and qualitative assessments in the contributions of the team members. However, the clear heterogeneity of the grades presented in this model should be noted because we can observe the four type of quantitative values (min, low, high, max) and the three grades (A, B, C).

First, we can observe two different paths at the beginning. One involves several different activities and iterations while the other one is totally linear. On one hand, the wider path probably reflects the behaviour of two team members because it includes different paths with states that can be omitted. We can observe that members who followed this path obtained higher grades (A and B) than those obtained by the other member who followed the linear path (C). They start with low or even minimum contributions and finish with a high|A activity: high contributions assessed with the highest grade. After that activity, there is a loop to the beginning, so again they provided a period combining low quantities followed by high quantities in their contributions. Analysing the frequencies of the event log, both low|B and high|A activities correspond to 22.22% (each one) of the paths followed by the team members. Therefore, this implies that multiple iterations starting with low contributions to increase the quantity later were carried out.

On the other hand, linear path only presents two contributions without loop. The first contribution provides the minimum quantity while the second one corresponds to the maximum quantity. Probably, this difference and order implies that one member started making a symbolic contribution and later made an effort to contribute as expected in the team, adding a big quantity of his work in one contribution. Both contributions were assessed with an average grade (C) and the last activity of the model also has a C grade. So, all the contributions of this member were assessed with the lowest grade. This behaviour evidences the worst performance from this team member while the most relevant teamwork and the leadership were carried out by the others members.

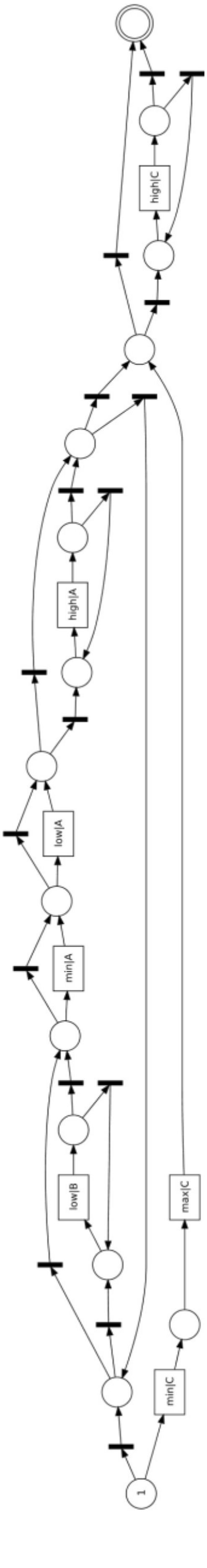

Fig. 4. Model for Crypto team (mixed data).

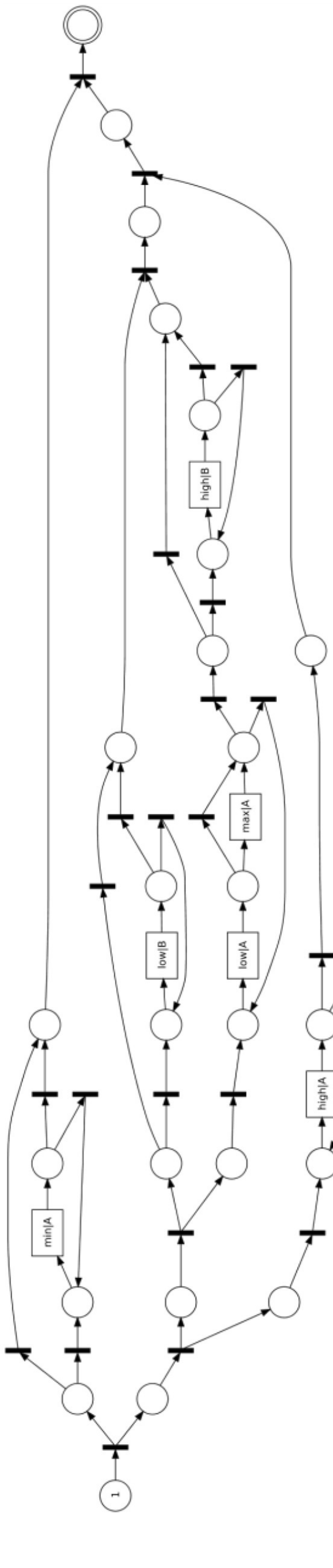

Fig. 5. Model for Simple-Latex team (mixed data).

The model ends with a high|C, which corresponds to 22.22% of the activities that occurred in the log. Therefore, it evidences a poor planning of the work: a large quantity of pending work had to be finished close to the deadline. This behaviour caused average grades in the qualitative assessments (C). Additionally, we can see that the first activities of the model contain minimum or low quantity of contributions and the rest present high and maximum. This can be taken as evidence that this team started working with less contributions than necessary ones and later they had to catch up.

If we compare this model with the previous one of weekly teamwork distribution (based on quantitative data) we can see some similarities and new findings. First, both models are quite linear, confirming an organized and properly planned teamwork. Additionally, both models share a bifurcation in the beginning: the team member who showed limited teamwork skills in the quantitative model corresponds to the behaviour of making a symbolic contribution and later made an effort to contribute as expected in the team. A new finding is that this student obtained average grades, confirming that the problems in teamwork also affected the quality of his/her work. At the same time, the other two members obtained high grades in their contributions. We can also see that at the end of the semester, grades are average, suggesting brilliant students had little to contribute to the project, while the other was unable to produce better work so close to the deadline.

4.2.2 Simple Latex Team

Although the obtained model (Fig. 5) is far away from an over-generalized "flower model", the low obtained Precision (0.29) implies that the model allows for other behaviours seen in the event log. Multiple loops are examples of this generalization. This structure contains a wide variety of sequences to model the data from the assessments. In addition, although there are few activities (only six), they present the four possible quantitative values (min, low, high and max) and two potential qualitative grades (A and B). Therefore, this general model could be taken as evidence of the heterogeneity in the teamwork of the members.

This model presents one key difference in comparison with the rest of the teams: multiple transitions produced more than one token. In this structure, the path of the team members is split and several paths are followed by each of them. It is symbolized as two (or more) arrows being created from a single transition (rectangle) and ending in different states (circles). An example of this can be seen at the beginning of the model, where all members parallelly follow both paths of the bifurcation. Therefore, four different paths were parallelly followed by all the members. The other aspect is the absence of C grades, so in general, the contributions of the members were sound.

The first path includes a loop with a min|A activity, a minimum quantitative value which obtained the maximum qualitative grade. Some team members probably provided a high amount of their work in a few contributions. The second path presents a similar structure, with a non-mandatory loop of one activity (low|B). This could imply unusual contributions in specific weeks provided by team members.

The third path includes a short loop with low and maximum. This can be taken as evidence that some team members alternated weeks with low contributions and others with a large quantity of work. All the work was highly assessed (A). The low|A activity corresponds to 33.33% of the contributions, so this iteration was constantly followed during several weeks. At the end of this path, there is a short loop with a high|B. This suggests that some team members likely had pending work to be finished and all was done in last weeks (high quantity) with a lower quality than before (B).

Finally, the fourth path presents a similar structure to the first and second paths: a short loop with a single activity (high|A). In this case, it corresponds to the 36.36% of the activities that occurred in the log, so this loop was commonly iterated. Therefore, team members carried out an alternating work pattern during these weeks: high quantitative/qualitative contributions and lower ones.

If we contrast this model with the previous one of weekly teamwork distribution (based on quantitative data) we can see similar structures with more detailed information. In general, the heterogeneity in the amount of activities and the frequency of them is now modelled with the multiple transitions that split the paths producing several tokens. The previously detected leader behaviour is modelled as the different contributions allowed the different loops of contributions. Checking quantitative information from StatMediaWiki, we can see that the leader made a significant contribution in the beginning and then maintained a continuous basis of contributing with an A grade. This is different to the profiles of the other two members: one significantly contributed in the middle (probably after the coordination in the 18th week) and at the end of the semester, and the other just contributed in the end. They are reflected in the loop of low, high and maximum amounts contributed that received good (B and A) grades. In this case, we can see that although some members were not constant, they made good contributions under their roles when required.

4.3 Findings

The discovered models for weekly work distribution found several collaborative dynamics about the students' teams. The models showed different levels of homogenization/heterogenization of the teamwork distribution. Leadership evidence was provided by the models by detecting how (type of contribution) and when (week) the team members started. The behaviour for specific team members was detected: the conduct of different roles and the type of contributions they focused on during particular periods of time. In addition, the discovering model techniques also showed evidence about teamwork skills like organization, planning, constant work and coordination.

Then, the discovered models in the mixed approach assessment also provided evidence. Contrasting them with the weekly teamwork distribution models, similarities and new findings are exposed. On one hand, levels of heterogeneity in the teamwork and other skills like leadership and planning were detected again. On the other hand, evidence on the consequences of teamwork problems in the quality of the work (average grades) were found. Additionally, poor work distributions and planning also affected the grades. Finally, findings were compared to previous and related studies.

The limitations of assessment processes in collaborative experiences based on wikis have been addressed in previous studies with related findings. First, StatMediaWiki supports the wiki contribution analysis providing a general picture of the wiki [9]. It groups data according to diverse periods of time, supporting the analysis of the distribution of effort in students' teams. Some transferable skills were observed, such as leadership or collaboration between students, however, these findings are only supported by quantitative data. Then, a more detailed assessment was carried out through multiple indicators applying both StatMediaWiki and AssessMediaWiki, following independent quantitative and qualitative approaches [6]. On one hand, these findings present detailed assessments supported by objective data. On the other hand, the current study follows a mixed approach, providing evidence on the collaborative dynamics of the students' teams.

Additional studies present related findings. In [22], the conducted experiment follows quantitative and qualitative approaches. A learning tool was implemented to obtain quantitative data from Moodle LMS in two different courses. These data were explored to provide a summative assessment according to students' contributions. Then, the qualitative approach was carried out by a manual review of questionnaires. On one hand, the learning tool provided indicators about the students' performances, reducing the time invested by teachers to assess students and showing the flexibility of the tool. On the other hand, the findings of the current study present a scalable assessment and illustrate the collaborative dynamics inside students' teams. Following a mixed approach, evidence of specific skills are detected and therefore relations between quantitative contributions and qualitative assessments are presented.

5. Conclusions

Teamwork is one of the key issues in the success of engineering projects. It involves several skills like leadership, planning, group distribution or continuous work. Unfortunately, due to the high number of interactions, the assessment of teamwork in collaborative projects is a challenging task. In addition, it is not feasible to assess some dynamics involved in teamwork, such as organization or work quality, with one single approach. In this study, we propose using Process Mining techniques to support a scalable teamwork assessment in a collaborative project based on a wiki using information from quantitative and qualitative sources.

Process Mining techniques allowed for the analysis of sequential processes from event logs. They proved to be a suitable solution for modelling the sequential contributions made by the different team members in the wiki. Our proposal is based on ProM, a Process Mining model discovery platform. It processed information from StatMediaWiki and AssessMediaWiki software packages. The implemented architecture automatically applied a model discovery technique to the mixed data to obtain models for each team's work according to different dimensions.

As a conclusion, we can affirm that the mixed approach allowed for a more detailed and scalable assessment. Applying the Process Mining model discovery proved to be a suitable technique to analyse the behaviour of team members in sequential processes like collaborative projects based on wikis. The tools used provided an automatic support for wiki contribution analysis in the collaborative processes. Therefore, we consider the previous evidence to answer the research question of this study.

The conducted experiment has been described in detail to ensure the clarity of the process and to enable replication of the procedure. However, the selected data set and executed processing limit the study in different ways. The data set belongs to one course in an academic year, so the experiment has been conducted considering this specific context. Then, some necessary operations were carried out in the data processing: quantitative data was grouped by weeks and qualitative data was discretized in three grade categories. Therefore, the discovered models are represented according to these abstractions.

Suggestions for future works include our proposal for the use of an LMS wiki such as Moodle wiki to enrich the model analysing the information provided by the learning platform. Additional sources of information could be applied, both in-the-wiki information, like the kind of text modified in a contribution that could be incorporated in the system, and out-of-the-wiki information, like task decomposition where the code is produced in a source control version system.

Acknowledgments – This work has been developed in the VISAIGLE project, funded by the Spanish National Research Agency (AEI) with ERDF funds under grant ref. TIN2017-85797-R.

References

1. K. Gibbard, A. Grocutt, A. Turner, T. O'Neill, R. Brennan and S. Li, Assessment of individual and teamwork attributes in undergraduate engineering students, *Proceedings of the Canadian Engineering Education Association (CEEA)*, Vancouver, BC, Canada, June 2018, pp. 1–4, 2018.
2. R. Lingard and S. Barkataki, Teaching teamwork in engineering and computer science, *2011 Frontiers in Education Conference (FIE)*, Rapid City, SD, October 2011, pp. F1C-1-F1C-5, 2011.
3. J. M. Dodero, M. Palomo-Duarte, I. Ruiz-Rube, I. Traverso, J. M. Mota and A. Balderas, Learning technologies and semantic integration of learning resources, *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, **10**(1), pp. 11–16, 2015.
4. S. D. Ravana, N. A. M. Amin and S. N. Raman, Effects of assessment on wiki activities during group work, *Proceedings of The Asian Conference on Society, Education, and Technology (ACSET)*, Osaka, Japan, October 2013, pp. 1–14, 2013.
5. G. P. Fairbrother, Quantitative and qualitative approaches to comparative education, in M. Bray, M. Bray, B. Adamson and M. Mason (eds), *Comparative Education Research. CERC Studies in Comparative Education*, vol. 19, 2nd edn, Springer Cham, Hong Kong, China, pp. 71–93, 2014.
6. A. Balderas, M. Palomo-Duarte, J. M. Dodero, M. S. Ibarra-Sáiz and G. Rodríguez-Gómez, Scalable authentic assessment of collaborative work assignments in wikis, *International Journal of Educational Technology in Higher Education (2018)*, **15**(1), pp. 40–61, 2018.
7. A. Balderas, M. Palomo-Duarte, J. M. Dodero and I. Ruiz-Rube, Qualitative assessment of wiki-based learning processes, *Proceedings of Ninth multidisciplinary symposium on the design and evaluation of digital content for education (SPDECE-2012)*, Alicante, Spain, June 2012, pp. 161–172, 2012.
8. A. Abbott, Sequence analysis: New methods for old ideas, *Annual Review of Sociology*, **21**(1), pp. 93–113, 1995.
9. M. Palomo-Duarte, J. M. Dodero, I. Medina-Bulo, E. J. Rodríguez-Posada and I. Ruiz-Rube, Assessment of collaborative learning experiences by graphical analysis of wiki contributions, *Interactive Learning Environments*, **22**(4), pp. 444–466, 2014.
10. H. M. W. Verbeek, J. C. A. M. Buijs, B. F. Van Dongenand and W. M. Van der Aalst, Prom 6: The process mining toolkit, *International Conference on Business Process Management Demonstration Track*, Hoboken, New Jersey, September 2010, pp. 34–39, 2010.
11. J. Ortega-Valiente, A. J. Reinoso and R. Muñoz-Mansill, Analysis of the implementation of wiki-based platforms in university education, *International Journal of Computer Information Systems and Industrial Management Applications*, **5**(1), pp. 41–49, 2013.
12. E. Kalliamvakou, D. Damian, L. Singer and D. M. German, The code-centric collaboration perspective: Evidence from github, *Technical Report DCS-352-IR, University of Victoria*, 2014.
13. M. Pérez-Mateo, La construcción colaborativa de proyectos como metodología para adquirir competencias digitales, *Comunicar* **21**(42), pp. 15–24, 2014.
14. M. D. V. Urías, S. A. Ortega and M. G. Navarro, Desarrollo de competencias digitales en docencia online: la asignatura Cimientos del curso de adaptación a grado en ingeniería de edificación, *Revista de Educación a Distancia*, **49**(8), pp. 1–29, 2016.
15. D. F. M. Rios and P. P. Robles, Desarrollo de competencias académicas utilizando la Wiki en la enseñanza de nivel superior: caso de dos universidades colombianas y sus bibliotecas, *VIRTUalis*, **6**(12), pp. 40–59, 2016.
16. A. R. González, Percepción de los estudiantes sobre el uso de wiki. Análisis de una experiencia de aprendizaje colaborativo en la Universitat de Girona, *Teoría de la Educación: Educación y Cultura en la Sociedad de la Información*, **14**(1), pp. 76–100, 2013.
17. S. Hadjerrouit, Wiki as a collaborative writing tool in teacher education: Evaluation and suggestions for effective use, *Computers in Human Behavior*, **32**(1), pp. 301–312, 2014.
18. L. V. Martínez and M. E. del Moral Pérez, Webquest y wikis: búsqueda de información en red y desarrollo de competencias en colaboración, *Teoría de la Educación: Educación y Cultura en la Sociedad de la Información*, **12**(1), pp. 190–208, 2011.
19. J. A. Caballero-Hernández, M. Palomo-Duarte, J. M. Dodero, G. Rodríguez and M. S. Ibarra-Sáiz, Integrating external evidences of skill assessment in Virtual Learning Environments, *Proceedings of the Fifth International Conference on e-Learning (eLearning-2014)*, Belgrade, Serbia, September 2014, pp. 65–70, 2014.
20. Á. Fidalgo-Blanco, D. Lerís, M. L. Sein-Echaluce and F. J. García-Peñalvo, Monitoring indicators for CTMTC: Comprehensive training model of the teamwork competence in engineering domain, *International Journal of Engineering Education (IJEE)*, **31**(3), pp. 829–838, 2015.
21. Á. Fidalgo-Blanco, M. L. Sein-Echaluce, F. J. García-Peñalvo and M. Á. Conde-González, Using Learning Analytics to improve teamwork assessment, *Computers in Human Behavior*, **47**(1), pp. 149–156, 2015.
22. M. Á. Conde-González, F. J. García-Peñalvo, Á. Fidalgo-Blanco and M. L. Sein-Echaluce, Study of the flexibility of a Learning Analytics tool to evaluate teamwork competence acquisition in different contexts, *Proceedings of the Learning Analytics Summer Institute Spain 2017: Advances in Learning Analytics (LASI-SPAIN 2017), Madrid, Spain, July 2017, no. 1925*, pp. 63–77, 2017.
23. M. Á. Conde-González, R. Colomo-Palacios, F. J. García-Peñalvo and X. Larrueca, Teamwork assessment in the educational web of data: A learning analytics approach towards ISO 10018, *Telematics and Informatics*, **35**(3), pp. 551–563, 2018.
24. M. Palomo-Duarte, J. M. Dodero, A. García-Domínguez, P. Neira-Ayuso, N. Sales-Montes, I. Medina-Bulo, F. Palomo-Lozano, C. Castro-Cabrera, E. J. Rodríguez-Posada and A. Balderas, Scalability of assessments of wiki-based learning experiences in higher education, *Computers in Human Behavior*, **31**(1), 2014.

25. A. Balderas, J. M. Dodero, M. Palomo-Duarte and I. Ruiz-Rube, A domain specific language for online learning competence assessments, *International Journal of Engineering Education*, **31**(3), pp. 851–862, 2015.
26. I. Traverso, A. Balderas, J. M. Dodero, I. Ruiz-Rube and M. Palomo-Duarte, Open Data Framework for Sustainable Assessment of Project-Based Learning Experiences, *Program Electronic Library and Information Systems*, **50**(4), pp. 380–398, 2016.
27. A. Bogarín, R. Cerezo and C. Romero, A survey on educational process mining, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **8**(1), pp. 1–17, 2018.
28. A. H. Cairns, B. Gueni, M. Fhima, A. Cairns, S. David and N. Khelifa, Process mining in the education domain, *International Journal on Advances in Intelligent Systems*, **8**(1), pp. 219–232, 2015.
29. J. A. Caballero-Hernández, J. M. Dodero, I. Ruiz-Rube, M. Palomo-Duarte, J. F. Argudo and J. J. Domínguez-Jiménez, Discovering bottlenecks in a computer science degree through process mining techniques, *2018 International Symposium on Computers in Education (SIIE)*, Cadiz, Spain, September 2018, pp. 1–6, 2018.
30. A. Bolt, M. De Leoni, W. M. Van Der Aalst and P. Gorissen, Exploiting process cubes, analytic workflows and process mining for business process reporting: a case study in education, *International Symposium on Data-driven Process Discovery and Analysis (SIMPDA)*, Vienna, Austria, December 2015, pp. 33–47, 2015.
31. C. Romero, R. Cerezo, A. Bogarín and M. Sánchez-Santillán, Educational process mining: A tutorial and case study using moodle data sets, in S. ElAltia, D. Ipperciel and O. R. Zaïane (eds), *Data Mining and Learning Analytics: Applications in Educational Research*, Wiley, New Jersey, p. 3, 2016.
32. W. Premchaiswadi and P. Porouhan, Process modeling and decision mining in a collaborative distance learning environment, *Decision Analytics*, **2**(1), pp. 6–40, 2015.
33. M. Mittal and A. Sureka, Process mining software repositories from student projects in an undergraduate software engineering course. *Companion Proceedings of the 36th International Conference on Software Engineering (ICSE)*, Hyderabad, India, May 2014, pp. 344–353, 2014.
34. C. Schoor and M. Bannert, Exploring regulatory processes during a computer-supported collaborative learning task using process mining, *Computers in Human Behavior*, **28**(4), pp. 1321–1331, 2012.
35. P. Porouhan and W. Premchaiswadi, Process mining and learners' behavior analytics in a collaborative and web-based multi-tabletop environment, *International Journal of Online Pedagogy and Course Design (IJOPCD)*, **7**(3), pp. 29–53, 2017.
36. C. Argyris and D. A. Schön, Participatory action research and action science compared: A commentary, *American Behavioral Scientist*, **32**(5), pp. 612–623, 1989.
37. C. Robson and K. McCartan, *Real world research: A resource for users of social research methods in applied settings*, 4th edn, Wiley, Chichester, 2016.
38. W. Carr and S. Kemmis, Becoming critical: Education, knowledge and action research, Falmer Press, London, 1986.
39. StatMediaWiki, https://meta.wikimedia.org/wiki/StatMediaWiki, Accessed 6 May 2019.
40. AssessMediaWiki, https://assessmediawiki.forja.cica.es/index_en.html, Accessed 6 May 2019.
41. F. Lima, A. C. Prado, A. A. Massote and F. Leonardi, Automation course for industrial engineers: An approach based on Petri Nets, software tools and laboratory experiments, *International Journal of Engineering Education*, **30**(5), pp. 1271–1279, 2014.
42. W. M. Van Der Aalst, A. Adriansyah, A. K. A De Medeiros, F. Arcieri, T. Baier, T. Blickle et al., Process mining manifesto, *International Conference on Business Process Management*, Clermont-Ferrand, France, August 2011, pp. 169–194, 2011.
43. figShare, https://figshare.com/projects/IJEE_19_-_Teamwork_Assessment_in_Collaborative_Projects_through_Process_Mining_Techniques/63434, Accessed 28 Oct 2019.

Juan Antonio Caballero-Hernández graduated with a degree in Computer Science and is a PhD candidate at the University of Cadiz. His main research interest is skill assessment in learning experiences based on serious games. Outside the academic environment, he has been working in different positions of the IT sector, including managing teams and projects.

Antonio Balderas received his MSc in Computer Science and his PhD degree at the University of Cadiz, Spain. He also works at the University of Cadiz as a Lecturer of the Department of Computer Science. His research is focused on the field of Technology-Enhanced Learning.

Manuel Palomo-Duarte is an Associate Professor in the Computer Science Department at the University of Cadiz (Spain). His primary research interests are learning technologies and collaborative engineering using open-source software, fields in which he has published different contributions in peer-reviewed journals.

Pablo Delatorre is an Associate Professor in the Computer Science Department at the University of Cadiz (Spain). His main research is focused in modelling emotional responses applied to narrative and education. Outside the academic profession, he worked in the ICT sector for ten years, managing national and international corporate projects.

Antonio J. Reinoso works as an Assistant Professor in the Department of Information and Communications Technologies at Alfonso X University (Madrid, Spain). His current research interests include the study of free and open knowledge management systems, particularly Wikipedia, and the application of e/b/m/u-learning technologies in academic activities.

Juan Manuel Dodero is a Professor of Computer Systems & Languages at the University of Cadiz. He holds a Computer Science and Engineering PhD from the University Carlos III of Madrid. His main research interests are creative computing and technology-enhanced learning, with a focus on software technologies for computer-aided creation and assessment.