Predicting teamwork group assessment using log data-based learning analytics

- Accepted version.
- If you would like to obtain the published version of the article, please click on the following link: https://doi.org/10.1016/j.chb.2018.07.016.
- Citation: Hernández-García, Á., Acquila-Natale, E., Chaparro-Peláez, J., & Conde, M. Á. (2018). Predicting teamwork group assessment using log data-based learning analytics. Computers in Human Behavior, 89, 373-384. https://doi.org/10.1016/j.chb.2018.07.016

Predicting teamwork group assessment using log data-based learning analytics

Abstract

The application of learning analytics techniques to log data from Learning Management Systems (LMS) has raised increasing interest in the past years. Advances in this field include the selection of adequate indicators and development of research frameworks. However, most research has focused on individual students, which has hampered the development of learning analytics for team assessment in collaborative learning contexts. From a four-dimensional view of teamwork, this study proposes a set of log data-based indicators to facilitate group assessment in project-based learning courses, and identify relevant predictors of final project results.

Keywords

Learning Analytics, Learning Management Systems, Teamwork, Indicators, Log data

Introduction

The growing use of Learning Management Systems (LMS), advances in statistical analysis, the rapid development of software and analytics methods, and the increasing interest in the education field to apply the principles of business analytics to learning processes, have led to the emergence of educational data-mining and learning analytics as one of the most promising research fields in computer-supported education. The main principle of learning analytics lies on the extraction of useful and actionable information from the large amount of data generated in online learning systems–i.e. LMS log systems–to inform the different learning actors–institutions, instructors and students–in order to improve learning processes.

While the objective of the application of learning analytics techniques may vary from case to case–prediction of academic success, implementation of early-warning systems, reduction of attrition rates, etc.– the analysis mostly relies on one source: the data stored on LMS logs. The main challenge is to decide which data can provide useful information or how to aggregate and present data in a format that may offer any additional value to the different actors. Agudo-Peregrina et al. (2014) define interactions, and their representation as log database records, as the basic contextualized data units needed for learning analytics. However, specific and less-studied problems, such as data dimensionality, still need to be addressed in log data-based learning analytics of online courses.

Data dimensionality is especially relevant in online computer-supported collaborative learning. So far, log data-based learning analytics mostly focuses on one dimension or specific aspect of data: frequency of interactions, at individual or course levels. Such a perspective leaves out essential information about collaborative learning processes, where social construction of learning is expected to happen. For instance, teamwork assessment in a learning context requires individual and team monitoring and assessment. However, monitoring and continuous assessment are intensive, time-consuming tasks for instructors (Buckingham-Shum & Ferguson, 2011; Fidalgo-Blanco et al., 2015). Usually, teamwork assessment involves observational methods, but these methods demand a long time that instructors do not often have (Fidalgo-Blanco et al. 2015), or require the presence of experts who may guarantee a more objective assessment (Hobson et al. 2014). Other team assessment practices include peer-assessment using questionnaires (Salas, Burke, Fowlkes & Priest, 2004), but they also demand a lot of time and lack objectivity due to the individual nature of perceptions. Therefore, two of the main problems for teamwork behavior observation and teamwork assessment are the lack of time and of objectivity of the methods.

Technology-supported collaborative activities facilitate observation of teamwork behaviors by logging the activity of each team member in a database (Davies et al., 2011). These logs, when combined with adequate teaching methodologies, provide objective evidence of teamwork behaviors without requiring peer-assessment, and facilitate monitoring and tracking tasks for instructors. However, log data are just raw data, and therefore they may be very difficult to understand without the required technical training or knowledge; further, their big volume makes it difficult to manage those data (Dominguez et al., 2013).

Learning analytics techniques provide helpful support in handling these digital traces. Learning analytics may be applied to log data stored in the LMS database for collection, selection, analysis and interpretation of the information stored in those records (Ferguson & Clow, 2015; Long & Siemens, 2011). Learning analytics has so far confirmed that students with higher participation (MacFadyen & Dawson, 2010), higher levels of interaction with their peers (Crawford & Lepine, 2011) or higher levels of overall interaction (Robinson, 2010) perform better and achieve better learning outcomes. Therefore, indicators identified in prior research may predict student achievement (Hung & Zhang, 2008), and some indicators may even successfully detect at-risk learners in the first weeks of a course (Dekker et al., 2009). Time-related information present in log data is also a useful source of information for instructors, as students that reply promptly to questions from their peers (Liu et al. 2011), those who have higher levels of reciprocity (Haya et al. 2015), share knowledge timely (Navimipour & Charband, 2016) or deliver their assignments in due time (Vermeulen, 2014) also show better outcomes.

Learning analytics also enable revealing behaviors that may be associated with lower academic performance, such as logging in the LMS without participating in activities (Hernández-García et al., 2015), late submission of assignments or accessing the different resources only in the last days prior to an exam (Romero et al., 2016). Students with longer time intervals between accesses to the LMS also have lower grades (Cocea, 2011).

Nevertheless, the examples above offer overwhelming evidence that learning analytics has so far focused on the analysis of individual behaviors. However, and considering that teamwork is an integrated set of individual and group behaviors, the same principles could apply to the application of learning analytics techniques. A strand of research has already focused on the application of learning analytics to individual behaviors in team-based learning (e.g. Fidalgo-Blanco et al., 2015; Conde et al., 2016). However, because research on learning analytics in teamwork contexts at a group level is scant or almost non-existent, this study aims to investigate whether LMS log data-based learning analytics might be suitable for teamwork assessment, and proposes the following research question:

RQ: Is it possible to predict final team grade in teamwork project-based learning using log data from a LMS?

In order to do so, it will be necessary to provide a conceptual framework to determine the different dimensions of teamwork and identify potential indicators of teamwork behaviors, both at team and individual levels. After establishing the conceptual the study details the research methodology and empirically tests the conceptual framework using real data of a course run in a Moodle LMS following the

CTMTC (Comprehensive Training Model of the Teamwork Competence) method (Lerís, Fidalgo, & Sein-Echaluce, 2014). Further, the study presents an overview of the data extraction process required to obtain teamwork indicators from the traces of collaborative interactions stored in the LMS log database using data mining techniques through extraction, transformation and loading (ETL) processes. Finally, the data analysis is followed by a discussion of results and a presentation of the main conclusions of the study.

Conceptual framework

Teamwork

Teamwork refers to behavioral patterns emerging from the dynamic interaction between two or more individuals (Boyatzis, 1982; Stevens & Campion, 1994). Teamwork involves effort coordination and regular and constant negotiation to reach to an agreement in order to achieve shared goals. In a teamwork context, goals are accomplished through knowledge exchanges and problem solving (Hobson et al., 2014; Villa & Poblete, 2007; Loparev, 2016).

Teamwork is observable when a task is being performed (Hobson et al., 2014; Gillies, 2007), which implies that teamwork behavioral patterns can be recognized through observation and can also be differentiated from other group actions. Teamwork is stable, even though group member changes may alter the degree of success of the outcomes derived from the teamwork (Prichard et al., 2011). Most important, teamwork is a cause and predictor of outcomes, given an established behavioral pattern or interaction model (Loparev, 2016; Robbins & Judge, 2012; González-Morales et al., 2011; Earnest & Landis, 2014).

This study makes a distinction between group teamwork and the individual work of team members. Even though both share some inputs –goal definition and available resources–, process –task execution– and outputs –delivered result or outcome (Mathieu et al., 2008), the main difference lies in interrelation and interdependence: real teamwork requires that all team members share the goals and objectives. Interdependence involves interaction between team members, cooperation and social skills that are not necessary when performing an individual task.

Teamwork is therefore a multidimensional concept, as it includes regular communication between team members in order to coordinate efforts and monitor tasks during a period of time, and requires cooperation, idea-sharing and knowledge exchanges. The former implies that teamwork emerges as a result of different behaviors, such as communication, coordination, cooperation and monitoring, which are complementary and observable, and are developed by every team member.

Dimensions of teamwork

There are two different conceptual models of interest for this study: the KSA (Knowledge, Skills and Attitudes) model (Stevens & Campion, 1994) and the 3C (Communication, Coordination and Cooperation) model (Fuks et al., 2007), both of which support the multidimensional definition of teamwork used in this study.

The KSA model proposes that there are two different levels regarding teamwork: individual or team member individual level –i.e. the tasks that each team member needs to perform in order to work effectively as part of a team– and team level, which refers to the tasks that all members should perform to successfully achieve the team goals and objectives. At an individual level, Stevens and Campion (1994) distinguish four dimensions: cooperation, monitoring, coordination and responsibility; at a team level, they identify three dimensions: communication skills, information exchanges and conflict solving, all of which refer to communication among team members.

While the KSA approaches teamwork in a broad sense, Fuks et al. (2007) propose their 3C model for online contexts. In the 3C model, there are three necessary requisites for effective teamwork to happen: communication, in the form of message exchanges to help negotiation and decision-making; coordination, which refers to the different mechanisms to manage people, tasks or activities and resources; and cooperation, which involves the harmonious elaboration of the tasks in a shared space. It is worth noting that in the 3C model, cooperation becomes a substitutive of collaboration in prior literature (Ellis, Gibbs & Rein, 1991), which refers to the interrelation of communication, coordination and joint participation. The 3C model considers that teamwork should not focus just on the product, but rather on the dynamic process of harmonious knowledge construction. From this perspective, teamwork is not only defined by achieving a shared goal through cooperation, but is also the result of social interactions and knowledge exchanges between team members by means of communication and coordination. An additional element, present in the KSA and 3C models, is supervision, monitoring or tracking of the tasks performed and the work done (DeJong & Elfring, 2010). This dimension is highly dependent on the three dimensions of the 3C model, as monitoring activities are extended along the whole process (Stevens & Campion, 1994) in a regular way (Ellis, Gibbs & Rein, 1991; Fuks et al., 2007). Monitoring facilitates control over the state of the project, including pending tasks, and are a reflection of responsibility and commitment to the team (DeJong & Elfring, 2010).

The multidimensional and interdependent nature of teamwork requires that all dimensions or behaviors should occur for effective teamwork to happen, even though the frequency, intensity or duration of each one may vary depending on the stage of the process. Two of the dimensions are directly visible for all team members and instructors –communication and cooperation–, while the remaining two are not directly observable.

Communication

Communication, or interaction between team members, initiates every teamwork process and is present during the whole duration of the process, albeit with different levels of intensity. In an LMS, message exchanges in the team forum or chat are the primary evidence of communication (Fuks et al., 2007). Therefore, the main unit of measurement of communication in LMS should include creation and/or update of messages posted.

Further, effective communication involves frequent, immediate, lengthy and timely replies between all participant members (Gillies, 2003). Frequency, measured as the total number of message exchanged in a period of time –typically, the duration of the teamwork– (Pargman et al., 2013; Aghae & Hansson, 2013) has a positive relation with student achievement; in addition, frequent interactions are an indication of an atmosphere of trust between team members (Shen et al., 2008). Immediacy and timeliness are related to the idea of reciprocity (Romero & Ventura, 2013; Haya et al., 2015), which can be measured by observing the average response time to messages posted by other team members and average response time of replies by other team members to messages posted by a given team member (Loparev, 2016). Longer messages generally involve higher levels of elaboration of the discourse, whether the content of the message focuses on knowledge exchanges, social interaction, team coordination or conflict solving (Wise, Zhao & Hausknecht, 2013; Romero et al., 2016). Moreover, message length has a positive relation with student outcomes (Romero, Lopez, Luna & Ventura, 2013).

Additionally, communication in effective teams must also be persistent. Regular communication shows that each team member is involved and committed (Haya et al., 2015). Regularity is related to the distribution of messages during the elaboration of the teamwork (Fuks et al., 2007). Thus, communications that are limited to just the days prior to a deliverable due deadline indicate poor teamwork (Mlynarska, Greene & Cunningham, 2016), and participation during the initial stages –e.g. the first week of the work– and early communication in the message board have a high predictive relevance in final outcomes (Jiang et al., 2014; Dawson, Macfadyen & Lockyer, 2009).

Cooperation

Cooperation refers to the collection and combination of contributions of all the team members in order to complete, in a cohesive and harmonious way, the final product of the teamwork. The main element of cooperation is then the production of work. In online spaces, production is strongly linked to interactions in a common workspace, where members may combine their contributions through additions, changes and corrections, as part of the collaborative activity (Fuks et al., 2007). The 3C model limits cooperative interactions to actions that are specific of the collaborative action, without consideration for other potential interactions that might occur in the platform.

In an LMS, these cooperative interactions are saved and stored in the different workspaces –e.g. wiki, glossary, workshop– and are visible to all members (Dingsoyr, Faegri, Dyba, Haugset, & Lindsjorn, 2016), allowing them to revise the different contributions and reflect about the process at any time. As a product of this revision and reflection, team members may propose further changes, accept the contributions or demand further required information to complete the content (Fuks et al., 2007). Analogously to the case of communication, the most visible indicators of cooperation are related to frequency and length of contributions, as well as regularity or consistency, and timeliness.

Cooperation involves all sharing of knowledge in the workspace (Liu et al., 2011), and has a positive relation with learning outcomes (Navimipour & Charband, 2016). Cooperation means working together, and requires constant presence –regularity or consistency– for effective teamwork to happen (Hung & Zhang, 2008). Furthermore, an indicator of effective cooperation is timeliness of contributions, including how early team members start to contribute to the team (Villa & Poblete, 2007), whether the contributions are shared when needed (Hollenbeck, Beersma & Shouten, 2012) and whether the tasks are completed in due time and the deliverables are finished before the deadline. Contrarily, delays in sharing the contributions in the shared workspace evidence poor time management skills, both at individual and team levels (Klassen et al., 2010; Levy & Ramim, 2012; You, 2015).

Monitoring/Tracking

Monitoring or tracking of teamwork is defined as the process of observation of team activities. This observation allows to detect errors and differences of opinion, and therefore it promotes the generation of suggestions and corrections that act as an essential feedback element to the rest of the team (Marks, Mathieu & Zaccaro, 2001). From the definition, monitoring and tracking is also linked to the amount of interactions in the LMS (DeJong & Elfring, 2010), as the information about observation actions is stored in the log database. An important distinction between monitoring and tracking activities is that they do not require additional interactions with the rest of team members, or at least not in a way that is directly observable, and therefore they are sometimes referred to as "passive" interactions (Agudo-Peregrina et al., 2014).

A key element in monitoring and tracking behaviors is regularity. For a team to work effectively, tracking activities should happen during the whole execution of the work. Monitoring and tracking, especially in online environments, serve as an indication of revision of, and refection upon the different tasks, changes done and the different agreements achieved by the team (Gillies & Ashmann, 2007). This revision has a reflection in the data about passive interactions in terms of time devoted by team members to read the different messages in the communication –forum, chat– and cooperation–wiki, workshop, glossary– spaces. It is important that the duration of observations should ideally be proportional to the quantity of information that team members have to read (Cocea & Wibelzahl, 2006). That is, if the duration is too short–difference between the click that gives access to the information and the next action–or too long –due to timeout– these records would not provide useful information about monitoring and tracking (Vermeulen, 2014).

Coordination

Coordination of teamwork refers to the level of synchronization of interactions among team members (Steven & Campion, 1994). A coordinated effort translates to a constant work pace by all team members (Kozlowski et al., 2016). Coordination links communication and cooperation by harmonizing and integrating individual efforts to achieve the team goals (DeJong & Elfring, 2010; Fuks et al., 2007). For instance, message exchanges that follow the pace of execution of the different tasks are indicative of a successful completion of the team project (Jiang et al., 2014). Furthermore, the process of coordination becomes visible when both communication and cooperation –message exchanges and evidence of task completion in the shared workspace– happen, as it shows that the team members are reaching to an agreement (Alonso, 2012). As coordination is all about integrating individual efforts, it manifests when team work involves effort by all team members, in terms of time devoted to send messages and contributions (DeJong & Elfring, 2010). That is, coordination needs to take into account the time devoted by the team and its members to communication and cooperation-related activities.

Finally, an analogy may be established for the coordination of monitoring and tracking activities. While students –and instructors– are generally unable to directly observe monitoring and tracking activities, the individual effort –or time devoted to monitoring actions– is also stored in the log database, and it is also possible to quantify that effort.

Method

Sample

The data used for the empirical analysis comprise all the activity of an Operating Systems course in the year 2016/17. A total of 115 students were enrolled in this mandatory course of the Bachelor in Computer Science degree, of which 53 (46.1 percent) belonging to 23 different groups got a final grade. The course has an on-hands approach, and most of the classes entail performing different tasks associated to management and programming of an operating system in the computer lab. Practice sessions are complemented with lectures on theoretical foundations of operating systems. The estimated workload for the course is 6 ECTS (European Credit Transfer System), equivalent to 150 hours, of which around 22 hours correspond to theory, 30 hours are lab sessions, 8 hours are allocated for office hours and around 100 hours should be dedicated to individual and group programming work.

The assessment includes four questionnaires of theory and practice (35 percent of the final grade) and two practical assignments (65 percent of the final grade), of which the first assignment (or intermediate assignment) accounts for 35 percent of the practical of the final grade and the second final assignment accounts for the rest. This research focuses on the results of the final assignment and final group grade awarded by the two course instructors, and analyzes student activity in the LMS during the whole course.

The methodology of the course follows the CTMTC methodology (Lerís, Fidalgo, & Sein-Echaluce, 2014), and requires students to work on a project during the whole course forming teams of three or four members using Moodle LMS. Every team must adhere to the stages defined in CTMTC –mission and goals, team normative, responsibility map, planning, implementation, and final outcomes–, using the LMS message boards for team communication and a team wiki in the LMS to provide evidence of the work and deliver their solution and final project. After delivery of the assignment, all students must pass an exam consisting on a simple modification–no more than 10-15 lines of code– to the solution delivered by the team. Students who pass the exam get a final grade, or else they receive a final score of zero and do not pass the course.

Measures

Based on the work by Ruiz-de-Azcárate et al. (2017), this study proposes 20 team-level indicators of the different dimensions of teamwork, most of which are average group scores of individual indicators (see Appendix A).

Communication measures include average number of messages exchanged by the team, standard deviation –divided by the square root of the number of posts to account for posting activity– of the distribution of communication effort made by the team members –it gives an approximate idea of whether or not the messages posted by the different team members are evenly distributed–, average message length, average posting frequency –number of messages divided by the time available to complete the project–, average reciprocity –which computes the time spent between replies–, and average regularity –how evenly spaced in time are the different messages posted by the team members, while also considering the number of messages posted.

Measures of cooperation include the average number of contributions to the shared workspace –wiki– by all team members, evenly distribution of contributions among team members, average contribution length, average regularity –analogous to the case of communication, but considering wiki-related data–, average team earliness –time passed between the initial day and the first contribution to the workspace– and average team delay –time between the last contribution and the final deadline.

Monitoring and tracking measures include the average number of passive interactions and evenly distribution of observations, average regularity of observation activities, average tracking time of communication actions, average tracking time of cooperation actions, and average frequency of monitoring activities.

Coordination indicators consist of the average time devoted to communication, cooperation and monitoring activities, and the corresponding evenness of distribution among team members.

The conceptualization of the dependent variable of the study, final team grade, requires some previous considerations. As mentioned in the conceptual framework, teamwork behaviors happen at individual and team level. Therefore, it makes sense to take into account individual-specific and team-global contributions to calculate final grade. In the context of this study, and following an approach similar to that of other courses that use CTMTC (e.g. Conde, Colomo-Palacios, García-Peñalvo, & Larrucea, 2017), a learning analytics tool already assists instructors in calculating the individual contribution to the grade, and therefore the use of indicators of individual teamwork behaviors would be redundant. Team final grade, however, is based on the instructors' subjective assessment of the different stages of CTMTC using a rubric; consequently, two sets of final team grades are available for analysis: group-level scores for each component/stage of the CTMTC, and single-score final team grade awarded by instructors. The existence of these two sets of grades facilitates predictive analysis of final team grade using different statistical procedures, which will be detailed next.

Data collection

Because the data stored in the Moodle database does not directly provide the indicators used in this study, it was necessary to define an ETL (Extraction, Transformation and Loading) process, using Rapidminer to define and execute the different processes. The definition and implementation of the 189 ETL processes requires to identify the different Moodle modules affected –e.g. team forum and team wiki identifiers–, as well as the different tables with relevant data about the different dimensions of teamwork. The total number of log data records is 65.887.

Analysis techniques and statistical procedures

Because of the exploratory nature of this study, and considering the size of the dataset, we consider different exploratory and predictive statistical procedures in order to identify relevant predictors of teamwork results. First, it is important to note that not all indicators belonging to each dimension share common themes, but at the same time high intercorrelation and multicollinearity might be expected between some indicators. Therefore, an initial exploratory factor analysis using principal component analysis (PCA) with Varimax rotation seeks to identify common factors among the indicators belonging to each of the four dimensions (Figure 1), aiming for variable reduction and to provide higher understanding of the potential underlying structure of the each set of indicators included in each dimension (Field, 2013; Hair, Black, Babin & Anderson, 2014).

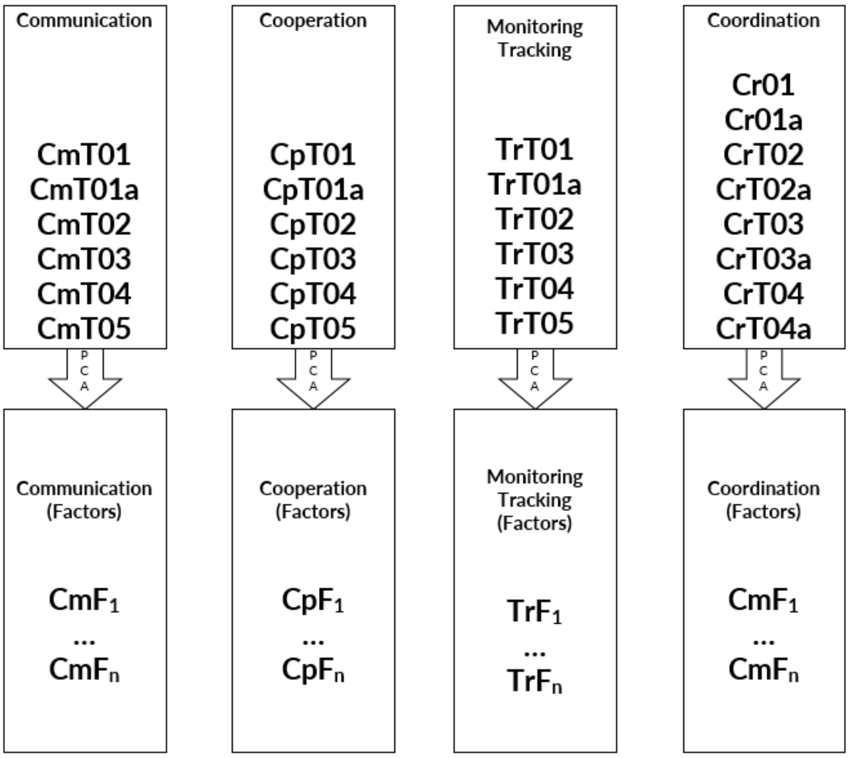

Figure 1. Diagram describing the exploratory factor analysis, used as reference for the subsequent statistical procedures.

Following the exploratory factor analysis, two different regression analyses aim to identify relevant predictors of final results, using the total final project grade awarded to the team by the instructor as dependent variable. The multiple regression analyses aim to predict the dependent variable –final team grade awarded by the instructor– based on the different indicators of teamwork –independent variables– across all dimensions. Because high collinearity among some of the indicators is expected, an initial multiple regression (figure 2a) introduces the factors from the exploratory factor analysis as independent variables. However, and for confirmatory purposes, the research also investigates the relations between all indicators and final grade by means of a second multiple regression analysis (figure 2b) that introduces all the indicators in sequential steps in the following order: communication, cooperation, monitoring and coordination. This sequence acknowledges the relevance of indicators from the learning analytics studies –presented in the conceptual framework– in the case of individual student log data.

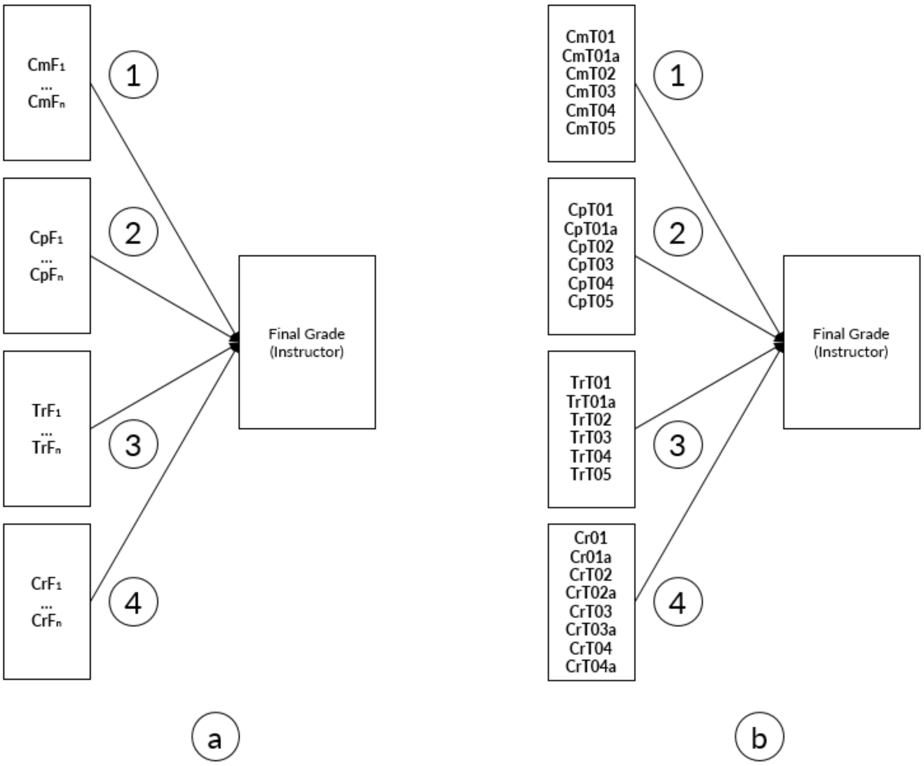

Figure 2. Diagrams of the multiple regressions, using the factors from the exploratory factor analysis (left) and team-level indicators of teamwork (right). The enclosed numbers indicate the order of introduction when using blockwise entry multiple regression.

Finally, the study proposes the analysis of a Partial Least Squares Structural Equation Model (PLS-SEM). PLS-SEM is an adequate prediction-oriented method for exploratory purposes (Hair, Hult, Ringle & Sarstedt, 2016). PLS-SEM facilitates exploration of the relationships between different latent variables or constructs, which are measured by manifest variables or indicators as composite variables. Thus, the analysis technique fits the purpose of this study. Further, PLS-SEM makes it possible to analyze hierarchical component models, in case higher order constructs are included in the model. The proposed research model consists of four endogenous latent variables, each corresponding to one dimension, and one exogenous variable, final team grade. In this case, in order to better characterize the endogenous variable and given that the final team grade awarded by the instructors comprises different components of teamwork following the CTMTC rubric (Conde et al., 2017), these components are included as indicators (Figure 3).

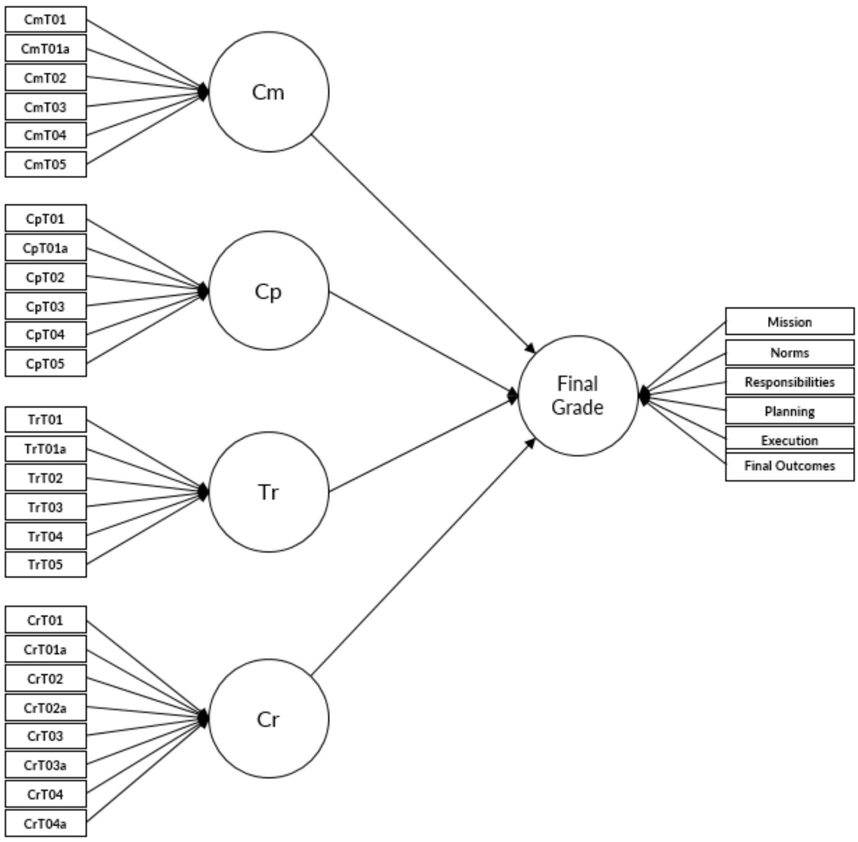

Figure 3. Partial Least Squares (PLS-SEM) Structural Model.

Data analysis

Exploratory Factor Analysis (Principal Component Analysis)

First, the PCA returns three different factors in the communication dimension, accounting for 86.9 percent of the variance (Table 2 in Appendix B). The first factor (CmF01) groups team message exchanges, length and frequency of posting, thus accounting for frequency and depth of message exchanges; the second factor (CmF02) measures even or uneven distribution of posting, and the third factor (CmF03) measures reciprocity. Regularity spreads with medium loadings along factors 1 and 2, and will be excluded from the regression. The second PCA, cooperation, also differentiates three different factors accounting for 81.7 percent of the variance (Table 3 in Appendix B). The first factor (CpF01) groups distribution of contributions and team earliness, the second factor (CpF02) groups number and length of contributions –i.e. frequency and depth–, and the third factor (CpF03) groups team delay and regularity. This result suggests that how soon students begin to cooperate in the shared workspace is related with how even is the workload shared among them (CpF01), in what could be labeled as promptness to cooperate, and that non-regular contributions are associated with how close to the deadline they deliver their work –with less regularly cooperating teams having more "last-minute" issues. The third PCA returns four factors accounting for 83.6 percent of the variance (Table 4 in Appendix B). However, the resulting factors, except for coordination of communication interactions (CrF03), mix different interactions and are not easy to explain. Because coordination is tightly connected to the other three dimensions, we shall proceed with caution and exclude these factors from the analysis. This result will be discussed later on. Finally, from the fourth PCA, monitoring/tracking, unveils two different factors accounting for 68.3 percent of the variance (Table 5 in Appendix B). The first one (TrF01) groups total number of observations, regularity, duration of monitoring/tracking forum interactions and frequency of monitoring, with observations and frequency having a perfect correlation, as expected by observing the definition; the second factor (TrF02) groups distribution of monitoring activities within the team and duration and monitoring of cooperative actions in the shared workspace. Based on the results of the PCA, Figure 4 depicts the procedure for the multiple regression analyses.

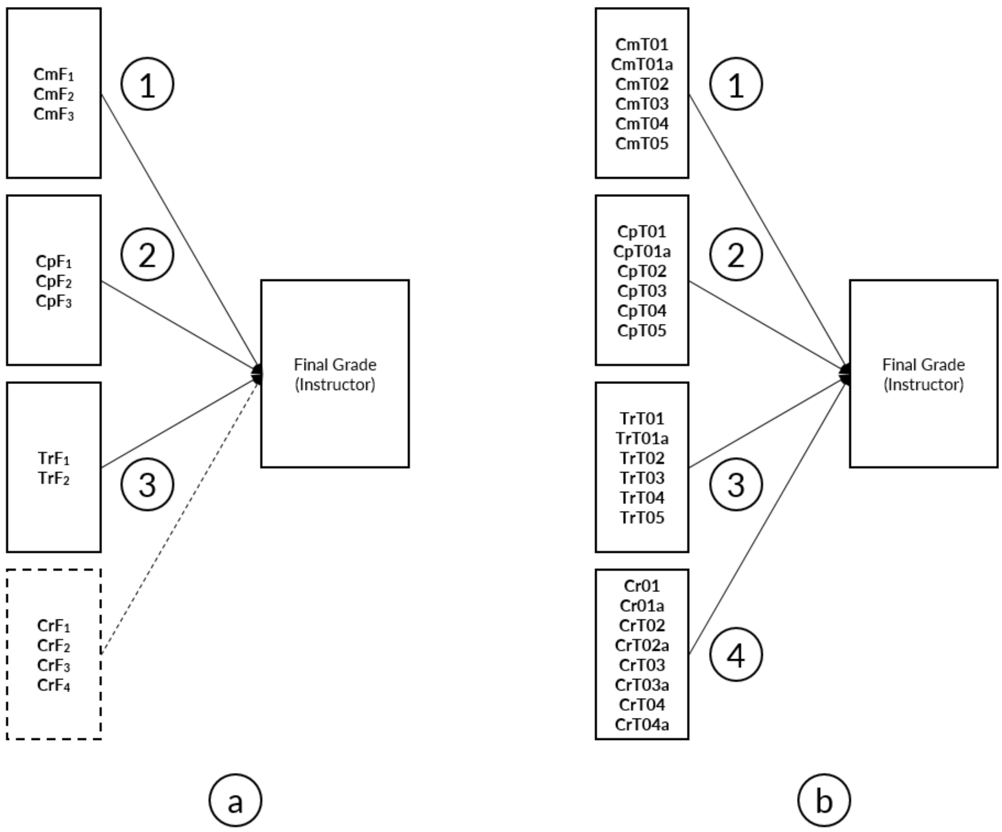

Figure 2. Diagrams of the multiple regressions, using the factors from the exploratory factor analysis after exclusion of coordination (left) and team-level indicators of teamwork (right). The enclosed numbers indicate the order of introduction when using blockwise entry multiple regression.

Multiple Regression Analysis

The regression analysis using both backward and forward stepwise regression with entry/deletion threshold of $p<0.05$ yield the same results (Table 6 in Appendix B), with only TrF01 significantly predicting final result (b_0=6.25; B=1.364, Beta=0.632, p=0.001; R^2=0.40, Adj. R^2=0.371). The total variance explained is similar to that of the model using forced introduction of variables. This result emphasizes the relevance of monitoring/tracking activities, especially those related to number, frequency and length of observation interactions –i.e. passive interactions.

The second regression, using blockwise entry, yields number of team messages exchanged as the only relevant predictor (b_0=4.88; B=0.112; Beta=0.561, p=0.005; R^2=0.315, Adj. R^2=0.283; Table 7 in Appendix B). This result is consistent with prior studies at individual levels, suggesting that the more messages a team exchanges, the better their results.

Partial Least Squares Structural Equation Model

The PLS-SEM analysis follows the different steps proposed by Hair et al. (2016). Upon the results from the exploratory factor analysis, and considering the nature of the factors that emerged, two alternative models are proposed: the first considers the endogenous constructs defined as formative in Mode A, allowing for correlations between their respective indicators (following Figure 3), whereas the second uses second order constructs for each dimension based on the factors identified in the PCA (Figure 5). For the latter, the repeated indicators approach (Becker, Klein, & Wetzels, 2012) has been followed.

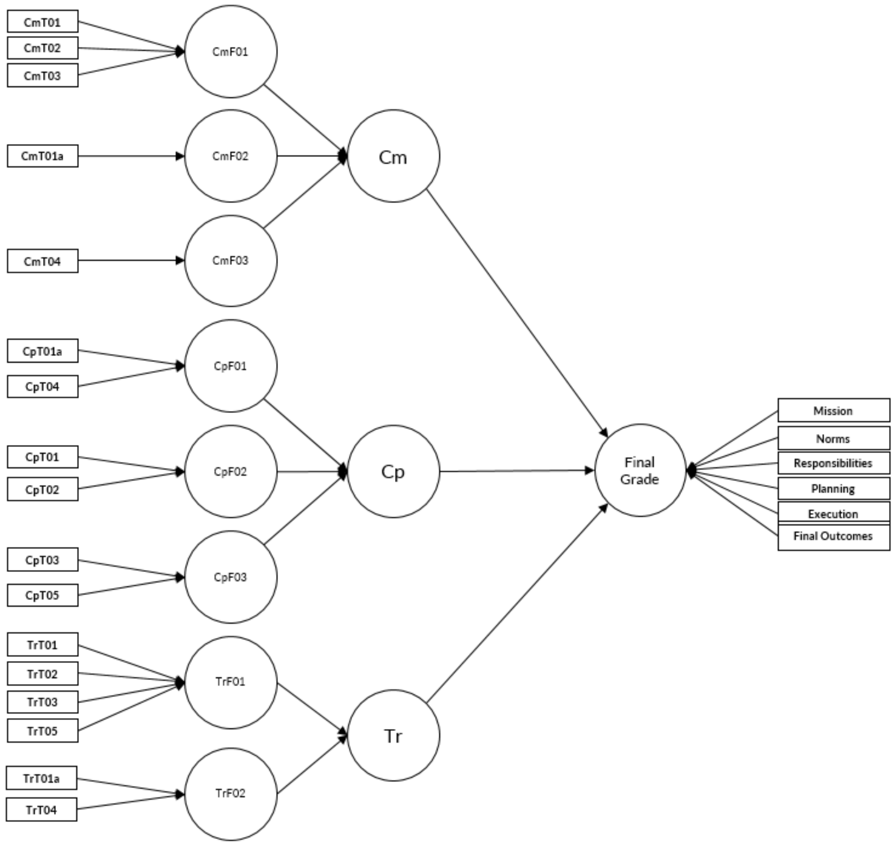

Figure 5. Alternative Partial Least Squares (PLS-SEM) Structural Model using the factors from the exploratory factor analysis.

PLS-SEM analysis follows a series of steps, including measurement model assessment –which evaluates the reliability and validity of the latent variable measures– and structural model assessment –which evaluates the relationships between the latent variables, as well as the model's predictive capability. In the following, and due to space limitations, only the main results of the PLS-SEM analysis are provided.

The first analysis (Figure 3), after item depuration, reveals high multicollinearity issues among constructs and no significant paths, suggesting the inadequacy of the model. The second analysis (Figure 5), after item depuration, reveals no effect of cooperation in the final result, non-significant paths between CmF02, CmF03 and Communication, and non-significant path between TrF02 and monitoring/tracking. In other words, CmF01 and TrF01 largely explain communication and monitoring/tracking behaviors, respectively. In addition, the results show high collinearity between CmF01 and TrF01, suggesting that both posting and post-monitoring behaviors reflect just one behavior in teamwork contexts. Based on the initial exploratory results, a final model is proposed (Figure 6). This simple model considers a single latent variable consisting of Cm01, Tr02 and Tr03 –that is, number of team message exchanges, tracking regularity and forum tracking duration. The final model has a path coefficient of β=0.683 (f^2=0.877, large effect), and R^2=0.467, Adj. R^2=0.442. A blindfolding procedure with a distance omission of 7 confirms predictive relevance of the model (Q^2>0).

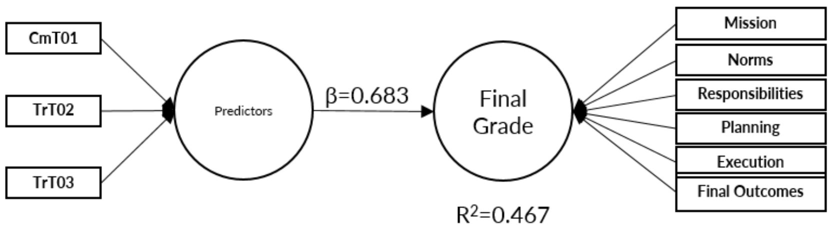

Figure 6. Final model.

Conclusion

The application of LMS log data-based learning analytics has focused primarily on the definition of predictors of individual student outcomes or academic performance. While this approach is necessary in order to make further advances in prediction techniques to model early-warning systems or give precise and informative feedback to both learners and instructors, it might not be so useful in collaborative learning scenarios, such as teamwork project-based learning. The possibility to explore, use and take advantage of learning analytics techniques in collaborative spaces has therefore been neglected by academic research so far.

From a theoretical perspective, this research pioneers the study of team-specific indicators based on LMS log data. Given the exploratory nature of this study, the results just offer a glimpse of which indicators might better reflect effective teamwork behaviors by observing LMS data. The study's theory-based proposal of indicators is promising and opens a new path for further research on this topic. Predictive power aside, the visualization of the different indicators presented in this study might greatly help learners and instructors to continuously and accurately supervise the current state of the work done by a team. For instance, an adequate visualization of evenness of communication exchanges, contributions and monitoring activities may be of great help to identify unbalance in the distribution of effort among team members in communication, cooperation or monitoring-related tasks.

An additional theoretical contribution of the study for future research is the difficulty to easily conceptualize and integrate the dimension of coordination in such a multi-dimensional approach. The main difficulty, as the analysis confirmed, lies mainly on two considerations: first, the tight relation with the other three dimensions, which might complicate statistical analysis due to confounding variables; second, the idea of synchronization, inherent to the concept of coordination, is quite difficult to translate to the digital context of LMS, where collaborative learning relies heavily on the asynchronous capabilities of online learning.

From a practical perspective, the results suggest that communication spaces are still the most critical environment where teamwork happens, confirming previous results in individual-centered applications of learning analytics techniques showing that students who participate the most obtain better academic results. However, the results also suggest that passive interactions –i.e. those that are not visible for students and teachers, related to monitoring and tracking behaviors– may play a more important role than that of individual behaviors. From the results, an effective team not only communicates regularly, but its members also regularly monitor communication exchanges, in order to improve coordination. This finding may establish a big depart from individual student-centered studies, where the importance of passive interactions cannot compare to active behaviors, such as posting (Agudo-Peregrina et al., 2014).

Surprisingly, cooperative actions –i.e. contributions to the shared workspace– do not seem to have predictive power over final results. After close inspection of the dataset, the reason behind this finding might be that the different teams delegated the responsibility of adding or editing the contributions to the wiki to one single member, who acted as representative of the team. CTMTC encourages that team members reflect their work in the wiki, but it does not specify whether one person or all team members should take charge of editing the wiki. Instead, the different content added to the wiki seems to be discussed and agreed in the message boards, and ultimately uploaded to the wiki once an agreement has been reached. Therefore, only a change in the method that could transform the collaborative space into a living and up-to-date state of the work, including content edition by all team members, might be observable through LMS data logs.

The study is not exempt from limitations. First, the most obvious is the sample size. Even though the number of data points used to elaborate the team indicators is quite large, comprising more than sixty-five thousand records, the number of students and groups is relatively small, and the results might prove context-specific. Therefore, confirmatory analyses with larger data sets are required to confirm the findings of this research. Finally, the analysis uses exploratory and predictive techniques for theory building –exploratory factor analysis and regression-based methods, such as multiple linear regression and PLS-SEM. It might be worth exploring how additional methods used in educational data mining –e.g. classification– compare to the methods used in this study.

As a final remark, the elaboration of teamwork indicators at a team level required computation of teamwork indicators at individual level. Because individual grading was already done in the course with the help of different analysis techniques with predefined rubrics, it was not considered appropriate to analyze the correspondence between final individual grade in the teamwork and teamwork indicators at individual level. However, the adequacy of indicators of teamwork at individual level may be worth exploring further when applied to other courses.

Acknowledgement

Blinded for peer review.

References

Aghaee, N., Hansson, H. (2013). Peer Portal: Quality enhancement in thesis writing using self-managed peer review on a mass scale. The International Review of Research in Open and Distributed Learning 14(1), 186-203.

Agudo-Peregrina, Á.F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in

VLE-supported F2F and online learning. Computers in human behavior, 31, 542-550.
Alonso, M.J. (2012). Competencias de una materia de Recursos Humanos: operativización y medidas. Revista de Formación e Innovación Educativa Universitaria 5(4)4, 191-208.
Becker, J.-M., Klein, K., & Wetzels, M. (2012). Hierarchical Latent Variable Models in PLS-SEM: Guidelines for Using Reflective-Formative Type Models. Long Range Planning, 45(5–6), 359–394.
Boyatzis, R.E. (1982). The competent manager: A model for effective performance. John Wiley & Sons.
Buckingham-Shum, S., & Ferguson, R. (2012). Social Learning Analytics. Journal of Educational Technology & Society 15(3), 3-26.
Cocea, M., & Weibelzahl, S. (2006). Can log files analysis estimate learners' level of motivation? In: Lernen-Wissensentdeckung-Adaptivität, 32-35.
Conde, M.Á., Colomo-Palacios, R., García-Peñalvo, F.J., Larrucea, X. (2018). Teamwork assessment in the educational web of data: A learning analytics approach towards ISO 10018. Telematics and Informatics 35(3), 551–563.
Conde, M.Á., Hernández-García Á., García-Peñalvo F.J., Fidalgo-Blanco Á., & Sein-Echaluce M. (2016). Evaluation of the CTMTC Methodology for Assessment of Teamwork Competence Development and Acquisition in Higher Education. In: Zaphiris P., Ioannou A. (Eds.) Learning and Collaboration Technologies. LCT 2016. Lecture Notes in Computer Science, vol 9753. Springer, Cham
Crawford, E.R., & Lepine, J.A. (2013). A configural theory of team processes: Accounting for the structure of taskwork and teamwork. Academy of Management Review 38(1), 32-48.
Davies, A., Fidler, D., & Gorbis, M. (2011). Future work skills 2020. Institute for the Future for University of Phoenix Research Institute, 540.
Dawson, S. P., Macfadyen, L. & Lockyer, L. (2009). Learning or performance: predicting drivers of student motivation. In R. Atkinson & C. McBeath (Eds.), Ascilite 2009: Same places, different spaces (pp. 184-193). Auckland, NZ: Ascilite.
De Jong, B. A., & Elfring, T. (2010). How does trust affect the performance of ongoing teams? The mediating role of reflexivity, monitoring, and effort. Academy of Management Journal, 53(3), 535-549.
Dekker, G.W., Pechenizkiy, M., & Vleeshouwers, J.M. (2009). Predicting Students Drop Out: A Case Study. Educational Data Mining 41.
Dingsøyr, T., Fægri, T. E., Dybå, T., Haugset, B., & Lindsjørn, Y. (2016). Team Performance in Software Development: Research Results versus Agile Principles. IEEE Software 33(4), 106-11.
Domínguez, A., Saenz-de-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J.J. (2013). Gamifying learning experiences: Practical implications and outcomes. Computers & Education 63, 380-392.
Earnest, D.R., & Landis, R.S. (2014). The Importance of Applicant Perceptions When Recruiting Employees to Teams. International Journal of Business and Social Science, 5(2).
Ellis, C.A., Gibbs, S.J., & Rein, G. (1991). Groupware: some issues and experiences. Communications of the ACM 34(1), 39-58.

Ferguson, R., & Clow, D. (2015). Consistent commitment: Patterns of engagement across time in Massive Open Online Courses (MOOCs). Journal of Learning Analytics 2(3), 55-80.

Fidalgo-Blanco, Á., Lerís, D., Sein-Echaluce, M.L., & García-Peñalvo, F.J. (2015). Monitoring Indicators for CTMTC: Comprehensive Training Model of the Teamwork Competence in Engineering Domain. International Journal of Engineering Education 31(3), 829-838.

Fidalgo-Blanco, Á., Sein-Echaluce, M.L., García-Peñalvo, F.J., & Conde, M.Á. (2015). Using Learning Analytics to improve teamwork assessment. Computers in Human Behavior 47, 149-156.

Field, A. (2013). Discovering Statistics using IBM SPSS Statistics (4th Ed.). Thousand Oaks, California: Sage.

Fuks, H., Raposo, A., Gerosa, M.A., Pimentel, M., & Lucena, C.J. (2007). The 3C collaboration model. In: Ned Kock (ed.), The Encyclopedia of E-Collaboration, 637-644. IGI Global, Hershey PA (2007).

Gillies, R.M. (2007). Cooperative Learning: Integrating Theory and Practice: Integrating Theory and Practice. Sage Publications.

Gillies, R.M. (2003). Structuring co-operative learning experiences in primary school. In: Adrian Ashman and Robyn Gillies (eds.), Co-operative learning. The social and intellectual outcomes of learning in groups, pp. 36-53. London: Routledge.

González-Morales, D., de Antonio, L. M. M., & García, J. L. R. (2011)Teaching “Soft” skills in software engineering. In Global Engineering Education Conference (EDUCON), 2011 IEEE (pp. 630-637).

Greene, J.A., Moos, D.C., Azevedo, R., & Winters, F.I. (2008). Exploring differences between gifted and grade-level students’ use of self-regulatory learning processes with hypermedia. Computers & Education 50(3), 1069-1083.

Hair Jr., J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2014). Multivariate Data Analysis (7th Ed.). Harlow: Pearson Education Limited.

Hair, J.F., Hult, G.T.M., Ringle, C.M., & Sarstedt, M. (2016). A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM), Second Edition. Thousand Oaks, California: Sage Publications Ltd.

Haya, P.A., Daems, O., Malzahn, N., Castellanos, J., & Hoppe, H.U. (2015). Analysing content and patterns of interaction for improving the learning design of networked learning environments. British Journal of Educational Technology 46(2), 300–316.

Hernández-García, Á., González-González, I., Jiménez-Zarco, A.I., & Chaparro-Peláez, J. (2015). Applying social learning analytics to message boards in online distance learning: A case study. Computers in Human Behavior 47, 68-80.

Hobson, C.J., Strupeck, D., Griffin, A., Szostek, J., & Rominger, A.S. (2014). Teaching MBA students’ teamwork and team leadership skills: An empirical evaluation of A classroom educational program. American Journal of Business Education 7(3), 191.

Hollenbeck, J.R., Beersma, B., Schouten, M.E. (2012). Beyond Team Types and Taxonomies: A Dimensional Scaling Conceptualization for Team Description. Academy of Management Review 37(1), 87-106 (2012).

Hung, J.-L., Zhang, K. (2008). Examining mobile learning trends 2003–2008: a categorical meta-trend analysis using text mining techniques. Journal of Computer Higher Education 24(1), 1-17.

Jiang, S., Williams, A., Schenke, K., Warschauer, M., & O'Dowd, D. (2014). Predicting MOOC performance with week 1 behavior. In Educational Data Mining 2014.

Klassen, R.M., Ang, R.P., Chong, W.H., Krawchuk, L.L., Huan, V.S., Wong, I.Y.F., & Yeo, L.S. (2010). Academic Procrastination in Two Settings: Motivation Correlates, Behavioral Patterns, and Negative Impact of Procrastination in Canada and Singapore. Applied Psychology 59(3), 361–379.

Kozlowski, S.W., Chao, G.T., Chang, C.H., & Fernandez, R. (2016). Team dynamics: Using "big data" to advance the science of team effectiveness. Big data at work: The data science revolution and organizational psychology. New York, NY: Routledge Academic.

Lerís, D., Fidalgo, Á., & Sein-Echaluce, M.L. (2014). A comprehensive training model of the teamwork competence. International Journal of Learning and Intellectual Capital 14, 11(1), 1-19.

Levy, Y., Ramim, M.M. (2012). A study of online exams procrastination using data analytics techniques. Interdisciplinary Journal of E-Learning and Learning Objects 8(1), 97-113.

Levy-Leboyer, C. Evaluación del personal: los métodos a elegir. Ediciones Díaz de Santos (1992).

Liu, Y., Keller, R.T., & Shih, H. A. (2011). The impact of team-member exchange, differentiation, team commitment, and knowledge sharing on R&D project team performance. R&D Management 41(3), 274-287.

Long, P.D., & Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education. EDUCAUSE Review 46(5), 31–40.

Loparev, A. (2016). The impact of collaborative scaffolding in educational video games on the collaborative support skills of middle school students (Doctoral dissertation). University of Rochester.

Macfadyen, L.P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. Computers & Education 54(2), 588-599.

Marks, M.A., Mathieu, J.E., & Zaccaro, S.J. (2001). A temporally based framework and taxonomy of team processes. Academy of management review 26(3), 356-376.

Mathieu, J., Maynard, M.T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. Journal of Management 34(3), 410-476.

Młynarska, E., Greene, D., Cunningham, P. (2016). Indicators of Good Student Performance in Moodle Activity Data. ArXiv

Navimipour, N.J., Charband, Y. (2016). Knowledge sharing mechanisms and techniques in project teams: literature review, classification, and current trends. Computers in Human Behavior 62, 730-742.

Pargman, D., Eriksson, E. (2013). "It's not fair!"-making students engage in sustainability. In: Proceedings of the EESD13 Conference, Cambridge, 2013

Prichard, J.S., Bizo, L.A., & Stratford, R.J. (2011). Evaluating the effects of team-skills training on subjective workload. Learning and Instruction 21(3), 429-440.

Robbins, S.P., Judge, T. (2012). Essentials of organizational behavior. Essex: Pearson.

Romero, C., Cerezo, R., Bogarín, A., & Sánchez-Santillán, M. (2016). Educational Process Mining: A tutorial and case study using Moodle data sets. Data Mining and Learning Analytics: Applications in Educational Research, 1.

Romero, C., & Ventura, S. (2013). Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 3(1), 12-27.

Ruiz-de-Azcárate, C., Hernández-García, Á., Iglesias-Pradas, S., & Acquila-Natale, E. (2017). Proposal of a system of indicators to assess teamwork using log-based learning analytics. In: Manuel Caeiro-Rodríguez, Ángel Hernández-García, Pedro J. Muñoz-Merino and Salvador Ros (eds.). Proceedings of the Learning Analytics Summer Institute Spain 2017: Advances in Learning Analytics. CEUR Worshop Proceedings 1925, 78-92.

Salas, E., Burke, C. S., Fowlkes, J. E., & Priest, H. A. (2004). On measuring teamwork skills. Comprehensive handbook of psychological assessment, 4, 427-442.

Shen, W., Hao, Q., & Li, W. (2008). Computer supported collaborative design: Retrospective and perspective. Computers in Industry 59(9), 855-862.

Stevens, M.J., & Campion, M.A. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. Journal of Management 20(2), 503-530.

Vermeulen, A.E. (2014). Improving Student Engagement Through Visualization of Course Activities (Doctoral dissertation).

Villa, A., & Poblete, M. (2007). Aprendizaje basado en competencias. Una propuesta para la evaluación de las competencias genéricas. Bilbao: Mensajero.

Wise, A.F., Zhao, Y., & Hausknecht, S.N. (2013). Learning analytics for online discussions: a pedagogical model for intervention with embedded and extracted analytics. In: Proceedings of the Third International Conference on Learning Analytics and Knowledge (pp. 48-56). ACM.

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