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

Ángel F. Agudo-Peregrina [a], Santiago Iglesias-Pradas [a], Miguel Ángel Conde-González [b], Ángel Hernández-García *,[a]

[a] Universidad Politécnica de Madrid, Departamento de Ingeniería de Organización, Administración de Empresas y Estadística, Escuela Técnica Superior de Ingenieros de Telecomunicación, Despacho A-126, Av. Complutense, 30, 28040 Madrid, Spain
[b] GRIAL Research Group, Information and Communications Services, Universidad de León, Campus de Vegazana s/n, 24071 León, Spain

ARTICLE INFO

ABSTRACT

Learning analytics is the analysis of electronic learning data which allows teachers, course designers and administrators of virtual learning environments to search for unobserved patterns and underlying information in learning processes. The main aim of learning analytics is to improve learning outcomes and the overall learning process in electronic learning virtual classrooms and computer-supported education. The most basic unit of learning data in virtual learning environments for learning analytics is the *interaction*, but there is no consensus yet on *which* interactions are relevant for effective learning. Drawing upon extant literature, this research defines three system-independent classifications of interactions and evaluates the relation of their components with academic performance across two different learning modalities: virtual learning environment (VLE) supported face-to-face (F2F) and online learning. In order to do so, we performed an empirical study with data from six online and two VLE-supported F2F courses. Data extraction and analysis required the development of an *ad hoc* tool based on the proposed interaction classification. The main finding from this research is that, for each classification, there is a relation between some type of interactions and academic performance in online courses, whereas this relation is non-significant in the case of VLE-supported F2F courses. Implications for theory and practice are discussed next.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Interactions between students and teachers, as well as interaction among students, may lead to effective learning by means of intellectual stimulation and exchange of ideas. However, it is very difficult to identify what is the net contribution of each type of interaction to the learning process; and even in the field of Information Technology (IT) supported learning, where interactions are easier to identify, this debate still remains open after two decades of research (Vrasidas & McIsaac, 1999; Hirumi, 2006).

Traditional in-class or face-to-face (F2F) learning has historically been a teacher-centric process, where students learnt by interacting with instructors, be it directly – through face-to-face interaction between teachers and students – or indirectly – with teachers adopting a mediation role between students and content,

mainly focused on the interpretation and explanation of content. The emergence of distance learning fostered a shift towards the decentralization of the learning process, with higher effort put on course structure and content creation – in order to reach a higher audience and reduce time consumption for instructors –, resulting in higher interaction levels between the students and the rest of course elements and agents. In recent years, the introduction of IT-supported or electronic learning (e-learning) has made it possible to focus the learning process in students by enabling multiple interactions among all the different agents involved – learners, instructors and course designers, tutors, contents, interfaces, administrative staff, code, environments, etc.

As Sims (1999) notes, interaction – in terms of interactivity – in electronic learning processes has many educational functions, related to learner control over system responses, adaptation to user's input, allowing for participation and communication and helping to provide meaningful learning. Hence, interactions have become an essential part of learning processes in electronic learning (Donnelly, 2010) and, according to McNeil, Robin, and Miller (2000), this variety of possible interactions is one of the biggest differences

* Corresponding author. Tel.: +34 913367366x2112.
E-mail addresses: af.agudo@upm.es (Á.F. Agudo-Peregrina), s.iglesias@upm.es (S. Iglesias-Pradas), miguel.conde@unileon.es (M.Á. Conde-González), angel.hernandez@upm.es (Á. Hernández-García).

between face-to-face and electronic learning. Anderson (2003) also notes the differences in the interactions between traditional face-to-face learning – with medium levels of student–teacher interaction, usually low levels of student–student interaction, and medium to low levels of student-content interaction –, traditional distance education – with high levels of student-content interaction and minimized student–teacher and student–student interaction – and web based courses – which support enhanced levels of student–student and student-content interactions while allowing for reduced student–teacher interaction.

Since in-class learning interactions are complex and informal by nature, it is difficult to systematically quantify them. On the other hand, information and communication technologies (ICT) – and more specifically the use of Learning Management Systems (LMS) or Virtual Learning Environments (VLEs) – make it easier to retrieve and process a high volume of data from every interaction among the different agents in a given course, especially those related to the traces that learners leave behind.

Surprisingly, despite the availability of this massive amount of data, usage logs from e-learning applications have been under- utilized in e-learning research (Phillips, Maor, Preston, & Cumming-Potvin, 2012) and it has only been recently that learning data analysis has raised the interest of scholars and educational institutions, be it under specific projects – e.g. "The Indicators Project"[1] – or joint initiatives like "SoLAR"[2], leading to the generation of new disciplines and research areas such as learning analytics (Ferguson, 2012).

There seems to be a consensus on what is the object of study of learning analytics: the analysis of VLE interaction data through the use of data extraction and data mining techniques, so that relations and useful information and knowledge about the learning processes may be inferred. However, the concept of learning analytics has different meanings for different people (Anderson, 2003). Therefore, although there is agreement that the ultimate goal of learning analytics is to improve teaching and learning – e.g. Macfadyen and Dawson (2012) –, the concept of learning analytics itself, and how it may contribute to improvements in teaching and learning, may differ from one person to another (Van Barneveld, Arnold, & Campbell, 2012).

1.1. Learning analytics and interactions in VLEs

Learning data analysis, most popularly known as learning analytics, is an emerging and promising – the estimated time-to-adoption horizon is between 2 and 3 years (Johnson, Adams, & Cummins, 2012) – discipline in educational research. In a first and broad approximation, learning analytics focuses on the analysis of automatically captured data to study student behavior (Phillips et al., 2012).

Learning analytics emerges from two converging trends: the increasing use of VLEs in educational institutions, on the one hand, and the application of data mining techniques to business intelligence processes in organizational information systems, on the other. From that perspective, learning analytics is commonly identified with Educational Data Mining (EDM), although there are some differences between the two concepts both in terms of purpose and scope; thus, while EDM is related to the development of methods for analysis of learning data (Baker & Yacef, 2009) from a mostly technical point of view – e.g. Romero and Ventura (2007, 2010) –, learning analytics has to do with the interpretation and contextualization of those data for the improvement of learning.

Therefore, learning analytics may be defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Ferguson, 2012). From this definition it follows that, in order to be able to understand and optimize learning processes in VLEs, it is necessary to know which data are stored by the system and to integrate them in a context which gives them a useful meaning for analysis.

The underlying idea in learning analytics comes from the large quantity of data – known as "big data" – about the activity of all the agents involved in the learning process which is registered by the VLE and stored in its database. Although these data are usually easy to extract, the volume of data is considered to be far too big to perform an analysis using typical database tools (Manyika et al., 2011). Moreover, the resources needed to transform the data are often scarce (Van Barneveld et al., 2012), which makes it necessary to develop *ad hoc* tools which allow filtering of these data so that useful information may be extracted from them (Johnson et al., 2012).

Learning analytics, as was mentioned in the introductory section, has many possible meanings. Van Barneveld et al. (2012) made a compilation of different definitions found in learning analytics contexts to make a clear difference between different types of educational data analysis, depending mainly on the subject and purpose of the analysis, leading to a conceptual framework of analytics in business and higher education. Fig. 1 depicts a simplified conceptual framework applied to educational institutions.

From Fig. 1, analytics in education is a form of data-driven decision making. While academic and action analytics consider similar data and decisions at an institutional level (Phillips et al., 2012), learning analytics is focused on the learner, gathering data from course management and student information systems in order to manage student success, including early warning processes where a need for interventions may be warranted (Van Barneveld et al., 2012).

It is not a surprise then that the most common approaches to learning analytics focus on the study of the underlying relations between interactions and students' academic performance – e.g. Ramos and Yudko (2008), Beer, Jones, and Clark (2009), Pascual-Miguel, Chaparro-Peláez, Hernández-García, and Iglesias-Pradas (2011) – or between interactions and participation levels and attrition rates in online courses – e.g. Cocea and Weibelzahl (2007), Macfadyen and Dawson (2010) – in order to explain learning behaviors (Van Barneveld et al., 2012), increase student success, improve enrolment and retention, and detect at-risk students (Macfadyen & Dawson, 2012).

As an analogy to in-class learning, where students usually interact with the teacher and other students, the term *interaction* was also applied to the first ICT-based learning systems, making reference to the different elements related to participation of students
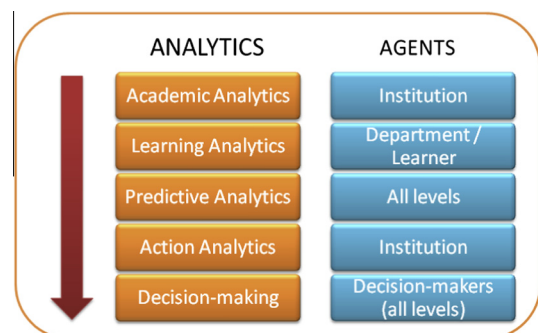

Fig. 1. Conceptual framework of analytics in education (adapted from Van Barneveld et al., 2012).

---

in a VLE. Steuer (1992) defines interaction in the VLE as "the extent to which users can participate in modifying the form and content of a mediated environment in real time" (p. 84). Later on, McNeil et al. (2000) introduce the idea that interactions group mutual actions among instructors, students and learning contents; this concept was expanded afterwards to include any exchange of information among agents in a course (Johnson, Hornik, & Salas, 2008), regardless that this exchange happens between human beings or between human and non-human agents (So & Brush, 2008).

Students leave a data trail while they interact with other people and information through the VLE (Siemens et al., 2011), and those interactions are intrinsically related to the different learning activities, bringing learners into contact with content, other learners and instructors (Richards & DeVries, 2011). Since that data trail, corresponding to records stored in the VLE, refers precisely to the actions among every agent involved in the learning process – both human and non-human –, and that those data are processed and stored in real time, we may infer that interactions – represented as data log records – are the basic contextualized data units needed for learning analytics.

But despite the growing number of studies on learning analytics, there is no agreement on which interaction data may be meaningful – or even if interactions have any pedagogical or educational value (Anderson, 2003), for that case –, or what are relevant criteria for selection (Duval & Verbert, 2012), and this lack of consensus is causing a greater fragmentation and lack of structure in this research area.

Unfortunately, and in spite of the rapid advances in data extraction techniques and data visualization tools, the number of studies dealing with how to profit from this information in the redesign of VLE is still scant (Beer et al., 2009), and the heterogeneity of educational contexts used in empirical studies makes it more difficult to offer any possible replication of the experiments for further generalization.

Moreover, the novelty of this research field implies that there is not yet solid theoretical framework available when it comes to decide which specific data must be analyzed, and more fragmentation is introduced by the great diversity of learning systems and tools available in the different educational institutions and the lack of standardization on which data are being captured and logged for later analysis. However, we find it necessary that these interactions are defined, identified and structured in a reference framework for analysis.

Therefore, since the relevance of interaction data on student performance and academic achievement remains an open question, this study aims to cover this research gap by addressing three questions:

R1. Is it possible to define a system-independent framework of interaction data characterization for learning analytics?

R2. If the answer to the previous research question is affirmative, is there any relation between interaction data and academic performance?

R3. If so, do results depend on course characteristics, such as instruction mode? (e.g. VLE-supported face-to-face learning versus on-line learning).

If we can give an answer to these questions, we might be able to move a step forward towards predictive analytics and use the results from this study as a basis for the design and development of systems which may give a preventive feedback to both students and teachers, based on real-time analysis of interactions within VLEs. In other words, depending on the interactions of a particular student in the VLE, the system might be able to respond in an automated – or semi-automated – way, generating corrective or reinforcing actions, in order to improve that student's academic performance, stimulate his or her participation in the course and avoid course withdrawal.

Therefore, building a categorization of educational interaction data based on extant literature becomes a critical requirement for this research. Nevertheless, the present study will still try and go a step further and, instead of being limited to only one classification, cover different typologies of interactions, irrespective of the VLE being used; for this research, we will demand one prerequisite for each classification proposal in order to assure system independence: that each one of the typologies must allow a univocal assignment of each interaction in the VLE to only one category. By so doing, we expect that the findings from the research and empirical work will help to assess the validity of each typology and their relative usefulness for learning analytics.

1.2. A proposal for system-independent classification of interactions in VLEs

1.2.1. Based on the agent

The first classification of interactions in learning processes to reach wide acceptance was proposed by Moore (1989), who identifies three different types of interactions associated to distance learning:

- Student–student interactions: they refer to the exchanges between the students enrolled in a course (Arbaugh & Benbunan-Fich, 2007). It includes the ability to establish a synchronous or asynchronous communication at the most convenient time or place, which may turn learning into a cooperative, socially constructed activity, rather than a solitary, isolated assignment (McNeil et al., 2000); this may be done, for example, through the use of chats and messages in forums or workgroups.
- Student–teacher interactions: these interactions are related to the participation level of teachers and the extent to which students perceive a teacher's proximity through online presence. Examples of these interactions are synchronous and asynchronous tutoring, exchanges of messages in the VLE between teachers and students – for example, teachers answering questions asked by the students about course-related topics via the VLE direct messaging system, etc.
- Student-content interactions: these interactions happen when students make use of many of the traditional content resources, such as textbooks, documents, research materials, videos, audios and other learning materials. In the context of a VLE, they are usually associated to browsing and accessing the different resources, tasks, etc.

In e-learning, every student must use the specific technologies, platforms, applications and templates available in order to interact with other students, teachers and content. Consequently, Hillman, Willis, and Gunawardena (1994) proposed an additional type of interaction which may reflect the information exchanges between students and system via the VLE interface, and they called it student-system interaction. The relevance of this kind of interaction relies on its role as facilitator or limiting factor in the quantity and quality of the other three types of interactions (Arbaugh & Benbunan-Fich, 2007).

Soo and Bonk (1998) added a new type of interaction to Moore's classification, named self-interaction, which refers to the self-regulation ability of each student as part of the self-directed learning process which is e-learning. This interaction is based on a reflexive thinking process by the student and does not generate any data in the VLE in a natural way; therefore, it has not been considered for this study.

Another addition proposal to Moore's typology was introduced by Hirumi (2002), who identified four kinds of interactions: self-interaction, student-human, student-"non-human" and student-instruction; however, this classification makes it possible to assign

each of the proposed categories to one of the original types, and it also does not differentiate the interactions the student has with the teacher from the interactions with his or her fellow students.

Muirhead and Juwah (2003) argued that teachers interact with contents too – mainly in creation/edition tasks, and also with the system; therefore, they added two more interactions to the previous four, namely teacher–content and teacher-system, and they included an additional one, since it is possible that some contents interact with one another (Muirhead & Juwah, 2003). Anderson and Garrison (1998) also extended Moore's classification to include other three types of interaction (teacher–teacher; teacher–content; content–content) but, as these interactions are not directly related to the students, they have not been included in this research.

1.2.2. Based on the frequency of use

Malikowski, Thompson, and Theis (2007) offered a perspective which integrates the technological perspective of the VLE and the conceptual aspects of online learning processes. Hence, they presented a categorization of interactions depending on the different activities which take place and features which are present in VLEs attending to their frequency of use in online courses. This does not mean that interactions are classified according to how much they are actually used but to how often they are present in a typical VLE – i.e., feature adoption rate. Thus, Malikowski et al. identified three different levels of use and a total of five categories:

- *Most used*: this level groups interactions related to the transmission of content. The category includes delivery and access to learning resources, general announcements and information about course grades.
- *Moderately used*: this level includes two different types of interactions: creating class discussions and evaluating students. Creation of class discussions, also known as creation of class interactions (Beer et al., 2009), refers to synchronous and asynchronous interactions between course members; on the other hand, interactions related to evaluating students or student assessment have to do with completing and sending individual and group assignments, quizzes, questionnaires, or other similar tasks.
- *Rarely used*: in this level we may find interactions related to the evaluation of courses and teachers – e.g. course/teaching quality or satisfaction surveys – or to computer-based instruction – self-assessment quizzes, checking of prerequisites for access to contents, adaptive learning elements, etc.

Malikowski et al.'s classification is consistent with later research. Thus, Dawson, McWilliam, and Tan (2008) proposed four categories representing the core activities within VLEs – working with content, administrative tasks, engagement with learning community and assessment – corresponding to the most used – the former two – and moderately used – the latter two – levels. Furthermore, empirical data from almost 53,000 students showed a great difference between the four most used functionalities of a VLE – content pages, discussions, course organization and assessment – and the rest of activities (Macfadyen & Dawson, 2012).

1.2.3. Based on the participation mode

A third possible classification is based on how the student interacts within the VLE. According to this criterion, Rovai and Barnum (2003) differentiate between two types of interaction: active and passive. Although at first research was limited to participation in message boards, Pascual-Miguel et al. (2011) made an extension of this classification to include synchronous media in their analysis, such as chats; and, ultimately, this classification may be extended to any type of interaction in the VLE, depending on

whether it requires the active participation of the student or not – e.g. reading an assignment might be considered a passive task, in contrast to completing it, which would be an active task.

2. Method

Once we have defined the three classifications which will be used in this research, we have designed an empirical exploratory experiment in order to determine the existing relationships between the different interactions and students' academic performance in VLE-supported F2F and online courses. The results from this experiment will help us to confirm the validity of the proposed classifications for learning analytics and to identify the interactions which have influence on academic performance by doing a comparative study of the relationship between interactions and academic performance for each of the three typologies.

In order to perform the empirical analysis, data were collected from VLE usage data logs in six online lifelong education courses and two VLE-supported F2F courses, delivered in two different Spanish universities (Universidad Politécnica de Madrid and Universidad de Salamanca). The diversity in the modality of content delivery responds to the objective of giving a meaningful answer to the third research question, which aims to evaluate if course characteristics have influence in the relation between interaction data and student performance, if there is any.

The total number of students enrolled in the online courses was 138, and there were 218 students in the VLE-supported F2F courses. The VLE used for all eight courses was Moodle,[3] an open-source Learning Management System (LMS).

In the following sections we describe the characteristics of these courses, the data extraction technique and the statistical analysis method used for this study.

2.1. Description of the courses and participants

The six online courses – which we will identify as courses 1–6 – selected for analysis are part of the lifelong learning offer of the Universidad Politécnica de Madrid, covering ICT-related subjects as well as business administration or organizational topics. The courses are comprised of virtual classes of 20–30 students and two or three teachers, and are structured in ten units taught during 10 weeks – for an estimated total dedication of 100 h per student. Course units are grouped in blocks of two units per block, and activities in each block are open to students for a period of 2 weeks – contents are made available as course progresses; once available, contents remain accessible during the whole course. The course also includes one face-to-face opening session; in this session, teachers explain the course objectives and methodology, and they present the VLE which will be used during the course. Technical support is provided mainly via in-course discussion board, although e-mail and telephone technical support is also available.

In order to assess student performance, each unit usually has one quiz and one written assignment – short answer question or essay. The course also includes one teamwork assignment – groups are randomly configured and have four or five components. Furthermore, and in order to foster a "classroom-like" feeling and to increase social presence in the VLE, teachers periodically post different topics for discussion in a discussion message board. Students are also encouraged to generate discussions about topics related with the course subjects.

The two VLE-supported F2F courses – courses 7 and 8 – were two Software Engineering third year mandatory graduate courses taught in two consecutive years at the Faculty of Computing

---

[3] http://www.moodle.org.

Sciences in Universidad de Salamanca. Face-to-face theory classes – 45 h – were comprised of three units; the course also included four workshops along the semester, in which students formed groups of three people and had to make a presentation of their work, and a final group assignment which also had to be presented at the end of the semester.

Final grade was distributed equally between theory and practice. Theory assessment included two tests during the course and a final exam at the end of semester, and the practice assessment included presentations made in the four workshops and an oral defense of the final assignment. The VLE was used mostly for content delivery – course syllabus, technical notes, documents and web references – as well as for communication purposes – discussion boards were available for intra-group, student–student and student teacher communication. The VLE, however, was not used for delivery of assignments or grading information.

2.2. Materials and procedure: data extraction tool

Due to the lack of standardization in systems, measures and procedures, until recently researchers have only been able to access, aggregate, analyze, visualize and interpret educational data via slow and cumbersome manual processes (Macfadyen & Dawson, 2012). After completing the interaction classifications proposal already covered in the previous section, we faced the task of actually extracting the desired data from the VLE logs; nevertheless, the structure and formatting information provided by Moodle's reporting tool is not suitable for learning analytics and requires a transformation process.

Moodle includes a tool – named *Log* – which generates activity reports for every user in the VLE. In Moodle, every user action – every click – is captured and stored as a database record. This allows users – provided that they have enough system privileges – to query the database using a functionality called *report logs* (Nagi & Suesalawuk, 2008). Although this tool has certain filtering capabilities, the information requires further processing to be analyzed in terms of the defined interaction categories, as was mentioned before.

Since we defined system-independent classifications of interactions, a system-specific implementation of the three typologies was required. The procedure followed for this implementation is as follows: in a first stage, two experts independently associated each possible interaction in the VLE to one category within each classification; after completion of this task, assignations between interaction and categories from the two experts were compared. From the 140 possible interactions, there was agreement in 94.3% of the cases for classification based on agent, in 96.9% of the cases for classification based on frequency of use – 12 interactions were not assigned to any category by either one or both experts – and in 100% of the cases for classification based on mode of interaction. The results of both assignations were reviewed by three additional experts, so as to resolve conflicts, and a majority criterion was adopted for final classification of interactions.

Finally, in order to perform the additional data processing, we developed an extracting and reporting tool in PHP programming language – called *Interactions* –, and integrated it in the VLE as a Moodle plug-in. This tool automatically makes the association of each of the possible interactions in the VLE to each category within each of the three types of classifications. The output from this module retrieves and stores each interaction in the VLE in a MS Excel spreadsheet - like *Log* –, and also information on how many interactions of each type occurred for each user – with all individual identifiers removed – during the course. Output data from the *Interactions* module was then imported in the statistical software package SPSS 18 (PASW Statistics), which was the tool used to perform the data analysis. Additionally, supplementary descriptive graphs were created from *Interactions* output data in MS Excel.

2.3. Data analysis

The first part of the analysis consists of an observation of the bivariate correlations between the different categories and the students' final grade for each learning delivery mode – i.e., VLE-supported F2F and online courses. Since simply relying on correlations for predictive power may lead to error, we proceeded to confirm the results with a multiple regression. This approach is similar to previous research in learning analytics with prediction purposes (Macfadyen & Dawson, 2012).

Multiple linear regression was used to find the different relations between student interactions in the VLE and their academic performance. In the context of this research, the independent variables were the number of interactions of each type registered for each user on the VLE – i.e., the output from the *Interactions* tool – and the dependent variable – academic performance – was represented by the final course grade achieved by each student.

Multiple regression methods are used to calculate the variance of the dependent variable as linear combinations of the independent variables. This makes it possible to create predictive models for the dependent variable based on data from the independent variables. This method also provides values of goodness-of-fit for the model and variance explained of the dependent variable. Furthermore, it assigns regression coefficients to each independent variable which will allow us to assess their relative importance in the predictive model (Brace, Kemp, & Snelgar, 2006).

More precisely, a backwards multiple regression was performed in this study. The advantage of this type of regression is that the initial equation includes all the independent variables, making it possible to find a set of variables with significant predictive capability even if none of its subsets have it; another advantage of this method is that there is no suppression effect, which occurs when independent variables interact with opposite effects (Garson, 2011) and may lead to Type II errors (Field, 2005, in Macfadyen & Dawson, 2012).

3. Results

Activities involving adaptive learning elements and assessment of course or teachers – i.e., surveys – were absent from all courses, so there were no interactions of the "computer-based instruction" and "evaluation of courses and teachers" types for the classification based on the frequency of use. Therefore, they both were excluded from the analysis.

Fig. 2 shows a graph with the average interactions per course. This figure shows similar tendencies in the behavior of students in all courses. Courses 1 and 2, with less technical contents, had a higher volume of interactions than the rest. Interestingly, the courses in both delivery modes seem to follow a similar VLE usage pattern – with lower values for VLE-supported F2F courses –, focused on student-system, student-content – i.e., non-human elements – and passive interactions. It is also worth noting that, regarding frequency of use, the two types of interactions in the "moderate use" category occur more frequently than content transmission-oriented interactions.

Unsurprisingly, the exception to this shared behavior patterns is the number of student assessment interactions – category "evaluating students" –, since tests and assignments were not submitted or handed over via the VLE in VLE-supported F2F courses.

Table 1 below shows the bivariate correlations of each type of interaction with final course grades for VLE-supported F2F and online courses.

From Table 1, there is a significant relation between the different types of interactions and the student's academic performance – all of them significant at $p < 0,01$ except for *transmission of content*,
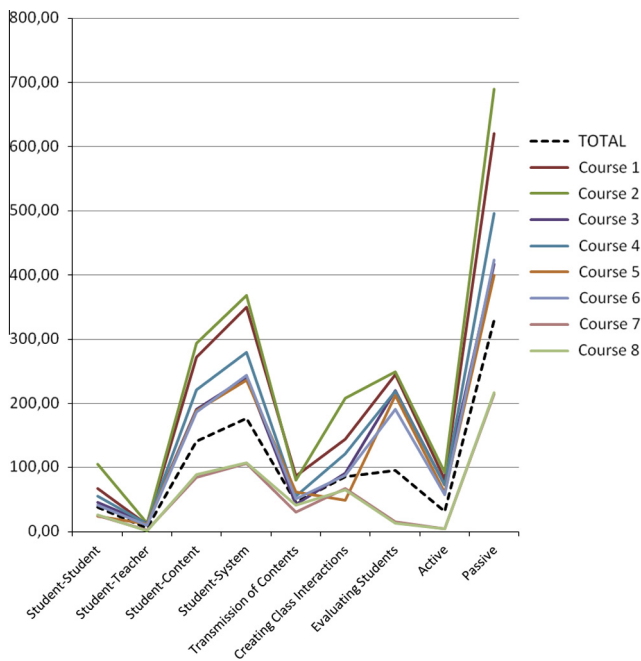
Fig. 2. Number of interactions of each type for each course.

Table 1
Bivariate correlations between interaction types and final course grade.

Classification	Interaction	Pearson's r^2	
		F2F	Online
Agent	Student–student	.091	.327[**]
	Student–content	.034	.438[**]
	Student–teacher	−.030	.540[**]
	Student-system	.031	.360[**]
Frequency	Transmission of content	.101	.191[*]
	Creating class interactions	.049	.321[**]
	Evaluating students	−.021	.526[**]
	Evaluation of teachers and courses	–	–
	Computer-based instruction	–	–
Mode	Active	−.014	.453[**]
	Passive	.039	.411[**]

[*] Significant at $p < 0.05$ (two-tailed).
[**] Significant at $p < 0.01$ (two-tailed).

significant at $p < 0,05$ – while, on the other hand, no significant relations were found for any interaction in VLE-supported F2F courses. However, simple correlation must not be mistakenly interpreted as causation and, as discussed before, cannot be heavily relied on for prediction purposes.

The final results of the backwards multiple regression are shown in Table 2 – model comparison – and Table 3 – final regression models for each classification.

In Tables 2 and 3, we have not been able to statistically establish any predictive model for final student learning outcomes in the context of VLE-supported F2F courses. Nevertheless, the results show that the three different classifications may be useful to predict student achievement in online courses.

From Table 2, the classification based on the type of agent offers a better explanation of students' academic performance than the other two classifications. The Durbin–Watson coefficient value indicated that there are no auto-correlation problems in any of the three models. As it was somehow expected, variance explained for the classification based on interaction mode, which includes the least elements, is lower than for the other two classifications.

Table 3 shows the final models for each of the classifications; criteria for exclusion in each step of the backwards stepwise regression was $p < 0.1$. According to the results, it follows that final students' academic performance is determined, depending on the classification used, by: (1) the interactions they have in the VLE with their peers and – mainly – with the teachers; (2) the interactions related to student assessment; and (3) interactions involving active participation. Moreover, the values of VIF (Variance Inflation Factor) suggest that multicollinearity effects may be ruled out in this analysis.

4. Discussion

Learning analytics is a rapidly evolving discipline, still undergoing its first stages of development. Although – as mentioned earlier in this paper – there is no consensus to this date on how actual learning analytics might be implemented – e.g. which data is useful, what different considerations have to be made regarding course and individual characteristics, teaching and learning styles, etc. – and empirical results from research may differ largely, we consider that only through a joint effort from the research community to combine theory and practice it will be possible to build a comprehensive learning analytics framework, which may be useful for all the agents involved in learning processes and educational decision making.

In this sense, our effort to keep a broad scope in this study is aimed to offer a theoretical and empirical basis for future research in learning analytics, open to critique and evaluation, by making this research as system and course-independent as possible. Hence, despite the high difficulty to extract general conclusions of application in every electronic learning context from our specific experiment, we strongly believe that this study leads a necessary path for effective use of learning analytics.

The most relevant findings from this research are directly related to the three research questions posed in the introductory section. Thus, we have proved that it is possible to define not one, but three system-independent characterizations of learning interactions – or educational data from VLE usage logs – which might be adequate and useful for learning analytics.

More importantly, we have found that the different types of interactions within each classification are related to student academic performance, but only in courses delivered online and not in VLE-supported F2F courses. This finding, if confirmed by future studies, may have an important implication for the practice of learning analytics, since it would mean that numerical-only

Table 2
Model comparison.

Classification	Model parameters							
	F2F				Online			
	R^2	Corrected R^2	Est. typ. err.	Durbin-Watson	R^2	Corrected R^2	Est. typ. err.	Durbin–Watson
Agent	–	–	–	–	0.356	0.346	1.396	1.894
Frequency	–	–	–	–	0.317	0.312	1.433	1.795
Mode	–	–	–	–	0.239	0.234	1.512	1.804

Table 3
Final models after multiple backwards stepwise regression.

	Regression parameters									
	F2F					Online				
	B	β	t	Sig.	VIF	B	β	t	Sig.	VIF
Based on agent										
Student–student						0.007	0.209	2.94	0.004	1.069
Student–teacher						0.154	0.508	7.14	0.000	1.069
Student-system	All variables excluded					Excluded from the regression				
Student-content										
Based on frequency of use										
Transmission of content	All variables excluded					Excluded from the regression				
Creating class interactions										
Evaluating students						0.012	0.563	7.97	0.000	1.000
Evaluation of teachers and courses	–	–	–	–	–	–	–	–	–	–
Computer-based instruction	–	–	–	–	–	–	–	–	–	–
Based on mode										
Active						0.028	0.489	6.56	0.000	1.000
Passive	All variables excluded					Excluded from the regression				

analysis of educational data might only be used for prediction in online courses; notwithstanding the foregoing, this does not mean that these classifications would be useless for the study of student outcomes in VLE-supported courses which are not exclusively delivered online, but that more advanced tools and different approaches would be required to perform effective learning analytics in these cases – for example, learning analytics based on data visualization on a student-by-student basis. This kind of tools might also help to discern how much does course structure affect predictability, a factor not taken into account in this research but which may have had influence on the results – online courses had a similar structure across the five different blocks, except for group assignments, while VLE-supported F2F was less rigid in terms of content delivery and assignment dates.

Furthermore, if we do not focus on prediction but rather on *ex post* analysis of courses, these advanced tools might help to examine student behaviors and detect uncommon or undesired behaviors which might have passed unnoticed; then, instructors might be able to use this knowledge for the redesign of courses by including new elements and activities or making changes to existent ones in order to correct those behaviors.

As for the final predictive model for online courses, the results from this study emphasize the importance of having teachers involved in the course (An, Shin, & Lim, 2009) and the promotion of active student participation as a lever to improve the learning process and its results. In other words, faultless operation of the VLE and high quality learning contents are considered a fundamental element in online learning processes – most of the interactions were made with system and contents –; but promoting interactions between VLE users plays an even more critical role in the planning and development of reinforcing and corrective actions in learning processes. This is consistent with Laurillard's (1997) view on how interaction between students and teachers plays a key role in educational processes, seen as a "conversation" which must be facilitated by learning technologies. However, teacher–student interaction is generally the least scalable type of interaction – as it is highly constrained by time availability. Thus, there is a tendency to substitute it by student-content interaction in mass education systems (Anderson, 2003), although there is evidence that the provision of additional content does not contribute to student achievement (Means, Toyama, Murphy, Bakia, & Jones, 2010).

Student–student interactions in online learning have been found to be the most important predictors of student success in prior studies (Macfadyen & Dawson, 2012). The results of our analysis show a lower influence than that exerted by student–teacher interactions, but student interaction with their peers may have been a less critical factor because the design of the courses was more based on cognitive learning principles than on constructivist theories (Anderson, 2003). Furthermore, the lack of significance of passive interactions and student-content interactions suggest that we have no effect of "vicarious interaction", where non-active students learn by observing the participation of active students (Sutton, 2000). However, this result should be further investigated before completely discarding it as a significant predictor of academic performance.

It is also worth noting that we found no relation between the "creating class interactions" type and final academic performance. This finding is especially interesting in light of the results for the other two models, and it may have been caused by the existence of slightly atypical results in two courses – courses 1 and 5 in Fig. 2 –, and therefore would require confirmation through the analysis of a larger volume of data. This result, together with the significant influence of "evaluating students" interactions, suggests the convenience of using multiple approaches and classifications simultaneously for learning analytics.

This research is not exempt from limitations, the most important of which is its exploratory nature. As we have stated in the introductory section, there is a great heterogeneity of experimental research on learning analytics; in this sense, although we have proposed a theoretical framework to characterize interaction data for learning analytics, our empirical study needs confirmation in other educational settings for further generalizability.

But the exploratory nature of this research also means that this study is a first step towards the formalization and definition of valid indicators for learning analytics processes, leaving an open door to the expansion of this field of research in the near future. It is the authors' belief that these research efforts should be focused in six main areas: (1) the study of the moderating factors of interactions in online courses, such as user experience in the use of VLE; (2) capturing data originated in informal learning processes which take place outside the VLE (Contreras-Castillo, Favela, Pérez-Fragoso, & Santamaría-del-Angel, 2004) or in other contexts – such as personal learning environments (PLE) –, and which are therefore not stored in the VLE database; (3) the analysis of interactions not only based on their nature but also by examining their semantic load – e.g. evaluating the content of human–human interactions, usage patterns of terms related to the learning objectives, etc.; (4) the inclusion of static or semi-static user data which are already present in the VLE – e.g. related to academic curricula – to allow for greater

customization when defining and applying corrective or reinforcing actions; (5) the complementary use of data visualization techniques, which add value and help to explain student behaviors and steer the learning process (Duval & Verbert, 2012); and (6) the development of recommender systems (e.g. Verbert et al., 2011), adapted to educational contexts, which might help to further expand the potential of learning analytics for decision-making for all learning agents.

5. Conclusion

The study of the relation between interactions and academic performance in VLE-supported courses is the object of study of learning analytics and becomes a key issue for learning process planning and deployment. This research field has lacked a structured view over time, leading to different results and implementations. This research provides a systematic approach to the study of these relations, applicable to all kinds of VLEs, irrespective of the system used. As a result of this study, we have presented three different classifications of student interactions based on: (a) the agents involved in the e-learning process; (b) the frequency of use of activities, features and functionalities in the VLE; and (c) participation mode.

We have also performed an exploratory analysis with data from eight courses – six online courses and two VLE-supported F2F courses. The results from this analysis show similar trends in VLE usage behavior by students across the different courses, and have helped to identify which interactions may have actual influence on students' academic performance in VLEs. These findings, which should be confirmed by further studies, provide a first theoretical basis for the selection of relevant data in learning analytics processes.

References

An, H., Shin, S., & Lim, K. (2009). The effects of different instructor facilitation approaches on students' interactions during asynchronous online discussions. *Computers & Education, 53*(3), 749–760.

Anderson, T. (2003). Getting the mix right again: An updated and theoretical rationale for interaction. *The International Review of Research in Open and Distance Learning, 4*(2), 1–14.

Anderson, T., & Garrison, D. R. (1998). Learning in a networked world: New roles and responsibilities. In C. Gibson (Ed.), *Distance Learners in Higher Education* (pp. 97–112). Madison, WI: Atwood Publishing.

Arbaugh, J., & Benbunan-Fich, R. (2007). The importance of participant interaction in online environments. *Decision Support Systems, 43*(3), 853–865.

Baker, R., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17.

Beer, C., Jones, D., & Clark, K. (2009). The indicators project identifying effective learning: Adoption, activity, grades and external factors. In: *Proceedings of ascilite* (pp. 60-70). Auckland, New Zealand.

Brace, N., Kemp, R., & Snelgar, R. (2006). *SPSS for psychologists: A guide to data analysis using SPSS for windows*. Routledge.

Cocea, M., & Weibelzahl, S. (2007). Cross-system validation of engagement prediction from log files. In E. Duval, R. Klamma, & M. Wolpers (Eds.), *Creating new learning experiences on a global scale, second European conference on technology enhanced learning, EC-TEL 2007, 14–25*. Berlin/Heidelberg: Springer.

Contreras-Castillo, J., Favela, J., Pérez-Fragoso, C., & Santamaría-del-Angel, E. (2004). Informal interactions and their implications for online courses. *Computers & Education, 42*(2), 149–168.

Dawson, S., McWilliam, E., & Tan, J.P.L. (2008). Teaching Smarter: How mining ICT data can inform and improve learning and teaching practice. *Proceedings of ascilite* (pp. 221–230). Melbourne, Australia.

Donnelly, R. (2010). Interaction analysis in a 'learning by doing' problem-based professional development context. *Computers & Education, 55*(3), 1357–1366.

Duval, E., & Verbert, K. (2012). Learning Analytics. *eleed, 8*.

Ferguson, R. (2012). *The state of learning analytics in 2012: A review and future challenges. Technical report KMI-12-01*. UK: Knowledge Media Institute, The Open University.

Field, A. (2005). *Discovering statistics using SPSS*. London: Sage.

Garson, G.D. (2011). Regression analysis, from Statnotes: Topics in multivariate analysis. <http://faculty.chass.ncsu.edu/garson/pa765/regress.htm> Retrieved 11.01.11.

Hillman, D. C., Willis, D. J., & Gunawardena, C. N. (1994). Learner–interface interaction in distance education: An extension of contemporary models and strategies for practitioners. *American Journal of Distance Education, 8*(2), 30–42.

Hirumi, A. (2002). A framework for analyzing, designing and sequencing planned e-learning interactions. *The Quarterly Review of Distance Education, 3*(2), 140–160.

Hirumi, A. (2006). Analysing and designing e-learning interactions. In Charles Juwah (Ed.), *Interactions in online education: Implications for theory and practice* (pp. 46–71). Abingdon, New York: Routledge.

Johnson, L., Adams, S., & Cummins, M. (2012). *The NMC horizon report: 2012 Higher education edition*. Austin, Texas: The New Media Consortium.

Johnson, R. D., Hornik, S., & Salas, E. (2008). An empirical examination of factors contributing to the creation of successful e-learning environments. *International Journal of Human-Computer Studies, 66*(5), 356–369.

Laurillard, D. (1997). *Rethinking university teaching: A framework for the effective use of educational technology*. London: Routledge.

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.

Macfadyen, L., & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology & Society, 15*(3), 149–163.

Malikowski, S. R., Thompson, M. E., & Theis, J. G. (2007). A model for research into course management systems: Bridging technology and learning theory. *Journal of Educational Computing Research, 36*(2), 149–173.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). *Big data: The next frontier for innovation, competition and productivity*. McKinsey Global Institute.

McNeil, S. G., Robin, B. R., & Miller, R. M. (2000). Facilitating interaction, communication and collaboration in online courses. *Computers & Geosciences, 26*(6), 699–708.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies. US Department of Education, Office of Planning, Evaluation, and Policy Development, Policy and Program Studies Service.

Moore, M. G. (1989). Three types of interaction. *American Journal of Distance Education, 3*(2), 1–6.

Muirhead, B., & Juwah, C. (2003). Interactivity in computer-mediated college and university education: A recent review of the literature. *Educational Technology & Society, 7*(1), 12–20.

Nagi, K., & Suesalawuk, P. (2008). Research analysis of moodle reports to gauge the level of interactivity in e-learning courses at Assumption University. In: Proceedings of the International Conference on Computer and Communication Engineering 2008 (ICCCE08), 772–776.

Pascual-Miguel, F., Chaparro-Peláez, J., Hernández-García, A., & Iglesias-Pradas, S. (2011). A characterisation of passive and active interactions and their influence on students' achievement using Moodle LMS logs. *International Journal of Technology Enhanced Learning, 3*(4), 403–414.

Phillips, R., Maor, D., Preston, G., & Cumming-Potvin, W. (2012). Exploring learning analytics as indicators of study behaviour. In: World Conference on Educational Multimedia, Hypermedia and Telecommunications (EDMEDIA), 2861–2867.

Ramos, C., & Yudko, E. (2008). "Hits" (not "discussion posts") predict student success in online courses: A double cross-validation study. *Computers & Education, 50*(4), 1174–1182.

Richards, G., & DeVries, I. (2011). Revisiting formative evaluation: Dynamic monitoring for the improvement of learning activity design and delivery. In: *Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK'11)*, 157–162.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications, 33*(1), 135–146.

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 40*(6), 601–618.

Rovai, A. P., & Barnum, K. T. (2003). On-line course effectiveness: An analysis of student interactions and perceptions of learning. *Journal of Distance Education, 18*(1), 57–73.

Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., et al. (2011). *Open learning analytics: An integrated & modularized platform. Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques*. Society for Learning Analytics Research.

Sims, R. (1999). Interactivity on stage: Strategies for learner–designer communication. *Australian Journal of Educational Technology, 15*(3), 257–272.

So, H., & Brush, T. (2008). Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & Education, 51*(1), 318–336.

Soo, K.S., & Bonk, C.J. (1998). Interaction: What does it mean in online distance education? In: *Proceedings of the world conference on educational multimedia and hypermedia & world conference on educational telecommunications ED-MEDIA/ED-TELECOM 98*, Freiburg, Germany.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication, 42*(4), 72–93.

Sutton, L.A. (2000). Vicarious interaction: A learning theory for computer-mediated communications. In: *Annual Meeting of the American Educational Research Association*, New Orleans.

Van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). Analytics in higher education: Establishing a common language. *Educause Learning Initiative, 1*, 1–11.

Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R., & Duval, E. (2011). Dataset-driven research for improving recommender systems for learning. In: *Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK'11)*, 44–53.

Vrasidas, C., & McIsaac, S. M. (1999). Factors influencing interaction in an online course. *The American Journal of Distance Education, 13*(3), 22–36.