Learning analytics and interactions in virtual learning environments

A comparative study of typologies and their relationship with academic performance

Ángel F. Agudo-Peregrina, Ángel Hernández-García, Santiago Iglesias-Pradas

Departamento de Ingeniería de Organización, Administración de Empresas y Estadística

Universidad Politécnica de Madrid

Madrid, Spain

{af.agudo, angel.hernandez, s.iglesias}@upm.es

Abstract—**Analysis of learning data (*learning analytics*) is a new research field with high growth potential. The main objective of *Learning analytics* is the analysis of data (*interactions* being the basic data unit) generated in virtual learning environments, in order to maximize the outcomes of the learning process; however, a consensus has not been reached yet on which interactions must be measured and what is their influence on learning outcomes. This research is grounded on the study of e-learning interaction typologies and their relationship with students' academic performance, by means of a comparative study between different interaction typologies (based on the agents involved, frequency of use and participation mode). The main conclusions are a) that classifications based on agents offer a better explanation of academic performance; and b) that each of the three typologies are able to explain academic performance in terms of some of their components (student-teacher and student-student interactions, evaluating students interactions and active interactions, respectively), with the other components being non-relevant.**

Keywords—*interactions; learning analytics; academic performance; typologies.*

I.   INTRODUCTION

Intellectual stimulation and exchange of ideas are key elements to achieve effective learning. Those may be reached through interactions between students and teachers, as well as among students, along the learning process. But identifying the specific contribution of each interaction has been, and still is, a subject of analysis and debate in the educational context in general, and in e-learning in particular [1].

In-class learning processes have traditionally been centred on the teacher; in these scenarios, students generally interact directly with the teacher, who also acts as a mediator in their interactions with learning contents through a process of interpretation of said contents. Nevertheless, e-learning allows interaction among multiple agents –students, teachers, tutors, contents, interfaces, characteristics, code, environments–, and this fact makes interactions become an essential part of learning processes [2]. McNeill et al. consider that it is precisely this diversity of interactions which constitutes one of the biggest differences between both learning scenarios [3].

Given their informal nature and their complexity for quantification purposes, the characterization of in-class learning interactions between students and teachers, and between students and contents, has proven a very difficult task. In e-learning, however, information and communication technologies (ICT), and more specifically the use of Learning Management Systems (LMS) or Virtual Learning Environments (VLE), have made it possible to retrieve a high volume of information about all the interactions among the different agents in a given course.

Despite the availability of this massive amount of data, it has only been recently that scholars have focused on this topic, be it under the form of specific projects –e.g. "The Indicators Project[1]"– or through joint initiatives like "SoLAR[2]", able to generate new disciplines and research areas such as Learning Analytics [4].

In general, the final objective of research on this field is the analysis of VLE interactions using data mining techniques, so that relations may be inferred from the chosen interactions. These relations usually try to establish a link between interactions and students' academic performance –e.g. [5-7]– or between interactions and participation levels and attrition rates in online courses –e.g. [8-9]–.

The main idea behind the study of these relations is the development of systems which may give a preventive feedback to both students and teachers, based on real-time analysis of interactions in VLEs. In other words, depending on the interactions of a particular student in the VLE, the system must 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.

Nonetheless, despite the growing number of studies focused on this topic, there is a surprising diversity in regards to which interactions are being considered for analysis, as well as to the results from these analyses, which in turn causes a greater fragmentation and lack of structure in this research field.

---

[1] http://indicatorsproject.wordpress.com/
[2] http://www.solaresearch.org/

Taking into account this dispersion, the present research has a twofold objective:

- First, and in order to establish a common reference framework for future studies, we aim to develop a formal set, based on literature research, of classifications of interactions in VLEs.

- Then, we will try to identify the nature of the relations between the different kinds of interactions and the students' academic performance by means of an exploratory empirical analysis. With this data analysis from actual courses we seek to identify the necessary relevant information to support the design of systems which may help to improve learning processes in the near future.

The remainder of this study has the following structure: in section 2 we present a theoretical background on learning analytics and interaction typologies in e-learning processes; in section 3, we will detail the characteristics of the empirical study, including descriptions of the courses, analysis tools and methodology. Section 4 will show the main results from the analysis and section 5 will summarize the main conclusions from this study, offering guides for future studies.

II. THOERETICAL BACKGROUND

A. Learning analytics and interactions in VLE

Learning data analysis (*LA, learning analytics)*, sometimes also known as *Educational Data Mining (EDM)* is a relatively new research field, with an estimated time-to-adoption horizon of between two and three years [10].

LA emerges from two converging trends: the increasing use of VLE 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.

The underlying idea in LA 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. This volume of data is considered to be far too big to perform an analysis using typical database tools [11], and this fact makes it necessary to develop *ad-hoc* tools which allow filtering of these data so that useful information may be extracted from them [10].

LA entails the processing of these data, and it 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" [4]. From this definition it may be assumed that, in order to be able to understand and optimize learning processes in VLE, 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.

As an analogy to in-class learning, where the 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 elements related to the participation of the students in the VLE. Thus, Steuer 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" [12]. Later on, McNeill et al. introduce the idea that interactions group mutual actions among instructors, students and learning contents [3]; this concept was expanded afterwards to include any exchange of information among agents in a course [13], regardless that this exchange happens between humans or between human and non-human agents [14].

Therefore, taking into account that the records stored in the VLE refer precisely to the actions among every agent involved in the learning process –both human and non-human–, and that these data are processed and stored in real time, we may infer that interactions are the basic contextualized data units needed for LA.

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 LMS/VLE is still scant [6], and due to the novelty of this research field there is not a solid theoretical base yet when it comes to decide which specific data must be analyzed. However, and since these data correspond to the different student interactions in the VLE, these interactions may be defined and identified.

Therefore, building a categorization of interactions based on extant literature becomes a critical requirement for this research. Nevertheless, the present study will still go a step further and, instead of being limited to only one classification, it will review various typologies of interactions. By so doing, results from the analysis will help to assess both the validity of each typology and their relative usefulness for LA in regards to students' academic performance. Naturally, a prerequisite will be that each one of the typologies must allow the univocal assignment of each interaction in the VLE to only one category.

B. Interaction types

1) Based on the agent

The first classification of interactions in learning processes to reach wide acceptance was proposed by Moore, who identifies three different types of interactions associated to distance learning [15]:

- **Student-student interactions**: they refer to the exchanges between the students enrolled in a course [16]. 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 [3]; 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 in 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 answering questions from the students about course topics, 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 et al. 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** [17]. 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 [16].

Soo and Bonk 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 [18]. 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 offered by Hirumi, who identified four kinds of interactions: self-interaction, student-human, student-"non-human" and student-instruction [19]; 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 argued that teachers interact with contents –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 [20]. As these interactions are not directly related to the students, they have not been included in this research.

2) Based on the frequency of use

Malikowski et al. 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 [21]. 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 comprises of two different types of interactions: creating class discussions and evaluating students. Creation of class discussions, also known as **creation of class interactions** [6], refers to synchronous and asynchronous interactions between the course members; on the other hand, interactions related to **evaluating students** 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, prerequisite checks for access to contents, adaptive learning elements, etc.–

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 differentiate between two types of interaction: **active** and **passive** [22]. Although at first this study was limited to participation in message boards, Pascual-Miguel et al. make an extension to include synchronous media in their analysis, such as chats [7]; 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–.

III. EMPIRICAL STUDY

Once the first research objective has been covered –i.e. describing the different typologies of interactions in VLEs from literature research–, we must proceed to the exploratory empirical analysis, in order to determine the existing relationships between the different interactions and students' academic performance in online courses.

This analysis allows us to achieve two different objectives. In first place, the identification of the interactions which have influence on academic performance will facilitate the development of LA tools; secondly, the comparative study of the relationship between interactions and academic performance for each of the three typologies will help to confirm their validity for LA purposes.

In order to perform the empirical analysis, data were gathered from six online lifelong education courses at the Universidad Politécnica de Madrid, with a total of 139 students. All six courses were delivered through the VLE Moodle. In the following section we will describe the characteristics of these courses, the data extraction technique and the statistical analysis method used for this study.

A. Description of the courses

The six courses which were selected for analysis are part of the lifelong learning offer at the Universidad Politécnica de

Madrid, and they cover ICT-related subjects as well as business administration or organizational topics. The courses are comprised of virtual classes of 20 to 30 students and two or three teachers, and are structured in ten units taught during ten weeks –for an estimated total dedication of 100 hours per student–. Course units are grouped in blocks of two units per block, and each block is open to students for a period of two weeks. 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.

In regards to evaluation and assessment, 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, the 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.

B. Data extraction tool

Moodle includes a tool –named *Log*– which generates activity reports for every user in the VLE. In Moodle, every user action 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* [23]. However, although this tool has certain filtering capabilities, the information available in this report requires further processing to be analyzed in terms of the defined interaction categories.

In order to perform this additional processing, we developed a plug-in for Moodle –called *Interactions*–, which automatically makes the association of each of the possible interactions in the VLE to each of the three types of classifications. The output from this module retrieves and shows in a MS Excel spreadsheet each interaction in the VLE -like *Log*–, but also information on how many interactions of each type occurred for each user during the course.

C. Methodology

Multiple linear regression was used to find the different relationships between student interactions in the VLE and their academic performance. In this study, the independent variables were the number of interactions of each type made by each user on the VLE –the output from the *Interactions* tool– while 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 prediction 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 [24].

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 [25].

IV. RESULTS

The statistical software package SPSS 18 (PASW Statistics) was used to perform the data analysis. Input data were provided by the *Interactions* module. In the absence of activities involving assessment of course or teachers –i.e. surveys– and adaptive learning elements, there were no interactions of the "evaluation of courses" and "teachers and computer-based instruction" types for the classification based on the frequency of use.

Fig. 1 shows a graph with the average interactions per course. This figure shows a similar behavior of students in every course. Interestingly, courses 1 and 2, with less technical contents, had a higher volume of interactions than the rest.

As shown in Fig. 1, a look at agents and interaction mode shows a tendency for students towards developing a higher number of interactions with non-human elements of the VLE -contents and system–, and predominantly of the passive kind. It is also worth noting that, according to the frequency of use, and contrary to the theoretical approach, the two types of interactions with moderate use occur more frequently than content transmission-oriented interactions.
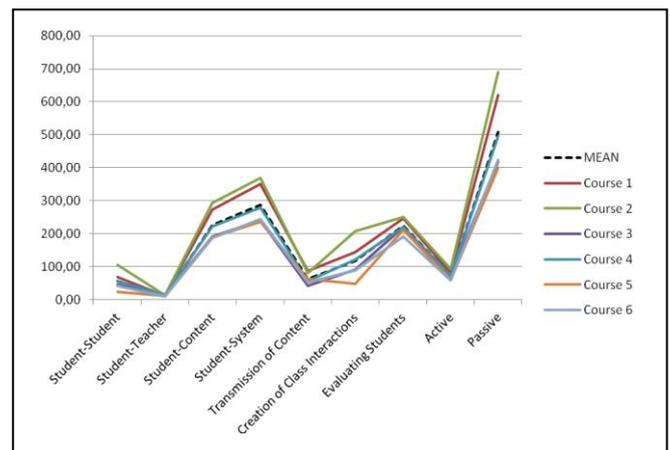

Figure 1. Number of interactions of each type for each course

Results of the backwards multiple regression are shown in Tables I –model comparison–-and II –final regression models for each classification–.

TABLE I. MODEL COMPARISON

Classification	Model parameters			
	R^2	Corrected R^2	Est. Typ. Err.	Durbin-Watson
Agent	0,356	0,346	1,396	1,894
Frecuency	0,317	0,312	1,433	1,795
Mode	0,239	0,234	1,512	1,804

From Table I, the classification based on the type of agent offers a better explanation of students' academic performance than the other two typologies. From the value of the Durbin-Watson coefficient, 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 II shows the final models for each of the classifications; this is, once the non-significant independent variables were excluded from the regression. According to the results, it follows that the 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 evaluating students; and 3) those interactions involving active participation. Moreover, the values of VIF (Variance Inflation Factor) suggest that multicollinearity effects may be ruled out in this analysis.

TABLE II. FINAL MODELS

	Regression parameters				
	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
Based on frequency of use					
Evaluating students	0,012	0,563	7,97	0,000	1,000
Based on mode					
Active	0,028	0,489	6,56	0,000	1,000

V. DISCUSSION OF RESULTS AND CONCLUSIONS

The study of the relationship between interactions and academic performance in online courses, a key issue for LA process planning and deployment, has lacked a structured view over time, which has led to very different results and implementations. This research provides a systematic approach to the study of these relationships, applicable to all kinds of VLE. As a result of this study, we have presented three different classifications of student interactions: based on the agents involved in the e-learning process, on the frequency of use of activities and features in the VLE, and on participation mode.

We have also performed an exploratory analysis with data from six courses; the results from this analysis have shown a similar behavior of students across different courses, and have helped to identify which interactions have actual influence on the students' academic performance in VLE. These findings, which should be confirmed by further studies, provide a first theoretical basis for the selection of relevant data in LA processes.

The two main results from this research are: a) the convenience of adopting a classification of interactions based on the agents involved for LA, if choice must be made; and b) the influence of student-teacher, student-student, evaluating students and active interactions in academic performance.

Although, as mentioned above, there is no consensus to this date neither on the results achieved in this field of research nor on which specific interactions should be measured, the results from this study emphasize the importance of highly involving the teachers in the course [26] and the promotion of active student participation as a lever to improve the learning process and its results. In other words, the operation of the VLE and the quality of the learning contents are a fundamental element in the support of online learning processes –most of the interactions were made with system and contents–; but promoting interactions between users of the VLE plays an even more critical role in the planning and development of reinforcing and corrective actions in learning processes.

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

Finally, we have to emphasize the exploratory nature of this research, which constitutes a first step towards the formalization and definition of valid indicators for LA processes, leaving an open door for the expansion of this field of research in the near future. It is the authors' belief that these research efforts should be focused in four 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 [27] or in other contexts –such as personal learning environments (PLE)–, and which are therefore not stored by the VLE; 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.–; and 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.

REFERENCES

[1] C. Vrasidas and S.M. McIsaac. "Factors influencing interaction in an online course". The American Journal of Distance Education, Vol. 13, No. 3, pp. 22-36, 1999.

[2] R. Donnelly. "Interaction analysis in a 'learning by doing' problem-based professional development context". Computers & Education, Vol. 55 No. 3, pp. 1357-1366, 2010.

[3] S.G. McNeil, B.R Robin, and R.M. Miller. "Facilitating interaction, communication and collaboration in online courses", Computers & Geosciences, Vol. 26, No. 6, pp. 699-708, July 2000.

[4] R. Ferguson. "The State Of Learning Analytics in 2012: A Review and Future Challenges". Technical Report KMI-12-01, Knowledge Media Institute, The Open University, UK, 2012.

[5] C. Ramos and E. Yudko. ""Hits" (not " discussion posts") predict student success in online courses: A double cross-validation study". Computers & Education, Vol. 50, No. 4, pp. 1174-1182, 2008.

[6] C. Beer, D. Jones, and K. Clark. "The indicators project identifying effective learning: Adoption, activity, grades and external factors". Proceedings of Ascilite. Auckland, 2009.

[7] F. Pascual-Miguel, J. Chaparro-Peláez, Á. Hernández-García, and S. Iglesias-Pradas. "A characterisation of passive and active interactions and their influence on students' achievement using Moodle LMS logs". International Journal of Technology Enhanced Learning, Vol. 3, No. 4, pp. 403-414, 2011

[8] M. Cocea and S. Weibelzahl. "Cross-System Validation of Engagement Prediction from Log Files". In Duval, E., Klamma, R. and Wolpers, M. (Eds.) Creating New Learning Experiences on a Global Scale, Second European Conference on Technology Enhanced Learning, EC-TEL 2007, pp. 14-25, Springer Berlin/Heidelberg, 2007.

[9] L.P. Macfadyen and S. Dawson. "Mining LMS data to develop an ''early warning system" for educators: A proof of concept". Computers & Education, Vol. 54, pp. 588-599, 2010

[10] L. Johnson, S. Adams, and M. Cummins. The NMC Horizon Report: 2012 Higher Education Edition. Austin, Texas: The New Media Consortium, 2012.

[11] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, et al. (2011). "Big Data: The Next Frontier for Innovation, Competition and Productivity". McKinsey Global Institute, May 2011.

[12] J. Steuer. "Defining Virtual Reality: Dimensions Determining Telepresence". Journal of Communication, Vol. 42, No. 4, pp. 72-93, 1992.

[13] R.D. Johnson, S. Hornik, and E. Salas. "An empirical examination of factors contributing to the creation of successful e-learning environments". International Journal of Human-Computer Studies, Vol. 66, No. 5, pp. 356-369, 2008.

[14] H. So and T. Brush. "Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors". Computers & Education, Vol. 51, No. 1,pp. 318-336, 2008.

[15] M.G. Moore. "Three types of interaction". American Journal of Distance Education, Vol. 3, No. 2, pp. 1-6, 1989.

[16] J. Arbaugh, and R. Benbunan-Fich. "The importance of participant interaction in online environments". Decision Support Systems, Vol. 43, No. 3, pp. 853-865, 2007.

[17] D.C. Hillman, D.J. Willis, and C.N. Gunawardena. "Learner–interface interaction in distance education: an extension of contemporary models and strategies for practitioners". American Journal of Distance Education, Vol. 8, No. 2, pp. 30–42, 1994.

[18] K.S. Soo and C.J. Bonk. "Interaction: What does it mean in online distance education?". Paper presented at the Ed-Media and EdTelec, June 1998.

[19] A. Hirumi. "A framework for analyzing, designing and sequencing planned elearning interactions". The Quarterly Review of Distance Education, Vol. 3, No. 2, pp. 140–160, 2002.

[20] B. Muirhead and C. Juwah. "Interactivity in computer-mediated college and university education: a recent review of the literature". Educational Technology & Society, Vol. 7, No. 1, 2003.

[21] S.R. Malikowski, M.E. Thompson, and J.G. Theis. "A model for research into course management systems: Bridging technology and learning theory". J. Educational Computing Research, Vol. 36, No. 2, pp. 149-173, 2007.

[22] A.P. Rovai and K.T. Barnum. "On-Line Course Effectiveness: An Analysis of Student Interactions and Perceptions of Learning". Journal of Distance Education, Vol. 18, No. 1, pp. 57-73, Spring 2003

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

[24] N. Brace, R. Kemp, and R. Snelgar. "SPSS for psychologists: A guide to data analysis, using SPSS for windows". Routledge, 2006.

[25] G.D. Garson. "Regression analysis", from Statnotes: Topics in Multivariate Analysis. Retrieved 11/01/2011 from http://faculty.chass.ncsu.edu/garson/pa765/regress.htm.

[26] H. An, S. Shin, and K. Lim. "The effects of different instructor facilitation approaches on students' interactions during asynchronous online discussions". Computers & Education, Vol. 53, pp. 749–760, 2009

[27] J. Contreras-Castillo, J. Favela, C. Pérez-Fragoso, and E. Santamaría-del-Angel. "Informal interactions and their implications for online courses". Computers & Education, Vol. 42, No. 2, 149-168, 2004.