

# Proposal of an Environment for Evaluation of EAD Course Communication System Based on Business Intelligence Techniques and Education Quality Referential

Ana T. Ichihara<sup>1</sup>, Nizam Omar<sup>2</sup>

<sup>1</sup>Faculty of Technology, Pça Cel Fernando, 30, São Paulo, SP

<sup>2</sup>Mackenzie Presbyterian University, Rua da Consolação, 930, São Paulo, SP

(<sup>1</sup>anatrava@uol.com.br, <sup>2</sup>nizamomar51@gmail.com)

**Abstract-** In Brazil, due to its extensive geography and social and economic differences, the expansion of distance education involves various aspects of social inclusion, democratization and customized training. This presents a number of challenges, including administration, an environment for evaluation, and understanding and measurement of the critical factors of a distance course – such as its communication system, essential in this teaching modality – in order to support the actions of decision makers and contribute to improving educational administration quality. This article presents a proposal for an environment to evaluate the communication systems of courses taught in Virtual Learning Environments (VLE), based on criteria of the “Quality Referential in Distance Higher Education” of the Brazilian Ministry of Education (MEC), which underlies educational evaluation processes, and Business Intelligence (BI) techniques. The work methodology consists of seeking Communication System Critical Success Factors (FCSSC) and measurable and applicable metrics based on this quality referential. A BI environment includes techniques that structure these metrics into dimensional models, which are created from processes of extracting and transforming data from VLE operational databases, which when quantitatively analyzed using graphic tools, makes results and visualizations flexible. Two case studies are presented in this article: the first shows quantitative analyses that characterize FCSSC “Interactivity” by means of the Pentaho Report Designer BI tool [21], and the second offers visual analyses of the relationship between elements of “Interactivity and Interaction” and student performance, by means of the Kohonen Self-Organizing Map (SOM) neural network.

**Keywords-** *Evaluation, Business Intelligence, Distance Higher Education*

## I. INTRODUCTION

In a country like Brazil, rich in diversity, geographic extension, social and economic inequalities and special needs, the Distance Education (EAD) modality is a topic that involves essential aspects of inclusion and customized training [1], and is present today, and expanding, in all Brazilian regions and states. According to Cunha [9], “EAD grows even in times of crisis”. Along with this growth come questions about how to

open spaces, innovate and improve distance education. According to the Brazilian Association of Distance Education (ABED), there is still much to be expanded and tried in this area, both in terms of program offers and technological and administrative innovation [1].

For Turrioni and Stano [33], evaluation is an administrative tool that ensures quality teaching and the factors critical to course success, such as communicative space. For Santos [30], to evaluate is to judge, appreciate someone or something, understand scenarios and contexts based on the interpretation of quantitative and qualitative data collected and interpreted using a scale of previously-defined values and criteria.

Therefore, information is essential, and by making it available in a Business Intelligence (BI) environment that permits exploratory analyses and has easy-to-use and visually attractive interactive features, it contributes to administrative excellence – an essential coordination skill [11] – and “helps to understand the social or scientific context and effectively discuss the meaning of distance education” [1].

This article presents an environment for evaluating the communication systems of distance education courses, based on criteria defined by the Quality Referential in Distance Higher Education published by the Distance Learning Secretariat of the Ministry of Education [18], and BI techniques, demonstrating this proposal with case studies that characterize some interactivity and interaction elements of distance. This is an extended version of a paper previously published in [12].

## II. METHODOLOGY

The work methodology consists of seeking Communication System Critical Success Factors (FCSSC) in the MEC/SEED Quality Referential in Distance Higher Education and mapping them in specific, measurable and applicable metrics.

These metrics are structured using the BI dimensional modeling technique that, according to Kimball [14], presents data to users in a standardized and intuitive manner, with performance advantages of access, greater integration and navigability using Business Intelligence graphics. The

dimensional data model is physically created using a dimensional database called Data Mart or Data Warehouse and loaded “populated,” through extraction, transformation and loading processes (ETL), with data from Virtual Learning Environment (VLE) operational databases. These operational databases store both data of participants’ registration and their actions and accesses to this environment. Thus, operationalization from obtaining the measurements related to each metric is accomplished by analyzing the VLE environment database, identifying the attributes and tables of the specified dimensional model.

The data provided by the dimensional model may be explored and analyzed using BI techniques that allow generation of consolidated and analytical reports, permitting quantitative exploration by decision makers. In this article, two case studies are presented: A hybrid distance learning course available in the Moodle VLE, with operational and dimensional databases, in MySQL, using the Pentaho Report Designer tool to characterize the “Interactivity” factor of the FCSSC; and a distance learning course in the Blackboard VLE, with Oracle operational and dimensional databases, in MySQL, that allows visual analyses of the relationships between the “Interactivity and Interaction” factors of the FCSSC and student performance using the Kohonen Self-Organizing Map (SOM) neural network.

### III. EVALUATION OF EAD COURSE COMMUNICATION SYSTEM

The success of a distance education course is anchored on a communication system that permits effective dialogue between all agents of the educational process, with principles of interaction and interactivity ensured by use of any technological means made available, creating conditions to reduce the sense of isolation, evasion and quality loss in distance education courses [18]. Therefore, it is necessary to evaluate this dimension to understand and monitor it, ensuring its efficacy. For Bertolin and De Marchi [4], there are various approaches to evaluating a communication system depending on the dynamics of the learning process and by the differences in space and time between course participants: by questionnaire, comparison with live courses, a hypothetical case or pre-determined indicators, using data mining techniques in education, statistical techniques, and Business Intelligence and Learning Analytics, among others.

Lesjak Florjančič [16] use the comparative evaluation method for three courses constructed for study, and questionnaires given to students at three different times. In their results, they mention the dependence between a course's success and its communication, observed through various channels, such as forums, videoconferences and others; and forms of interaction, for example, interaction with a tutor, his/her comments, interaction between students and between a student and professor. Martins and Zerbini [17] evaluate Reaction to Instructional Procedures and Reaction to Tutor Performance tools, seeking evidence of their validity through questionnaires and statistical techniques. Rochefeller and Costa [26] describe a teaching system using the Conversation and Pragmatism Analysis Theory designed to evaluate mediations in collaborative forums in VLE environments. Santana et al.

[29] evaluate use of interactivity tools in a Moodle VLE environment and their impact on student course performance through data mining techniques and quantitative access attributes whose results show underutilization, even with student success. Turrioni and Stano [33] propose a quantitative model to monitor and review EAD courses, which uses ten criteria, including interactivity and communication, focused on the discussion forum. Ramos, Rodrigues and Silva [25] emphasize the importance of communication in this teaching modality, referencing various authors, including Desai, Hart, Richards[10], and Rabello [24], and analyze the critical factors of an EAD course that affect student performance, conducting research on students with subsequent statistical analyses. Among these critical factors are mentioned those related to interactivity, with regard to the ability of the student to use the computer resources.

An evaluation alternative is to analyze the large amount of data stored in VLEs, using management systems to provide information to improve teaching quality [3]. This process can be automated by tools that use BI techniques [39]. This article proposes use of a BI environment, with metrics oriented by the quality referential established by the MEC/SEED [18], to evaluate the communication systems of distance education courses. This promotes understanding of the forms of interaction, participation patterns, their uses, strengths, weaknesses, difficulties and consequences, exposing needs, challenges and possibilities for improvements. With information, it is possible to understand, associate, judge and contribute with new forms that provide greater course quality [23].

### IV. MEC/SEED QUALITY REFERENTIAL IN DISTANCE HIGHER EDUCATION

For Junior [13], “operationalization of the evaluation process results from pre-determined criteria and indicators, ensuring uniformity of the elements to be judged.”

The “Quality Referential in Distance Higher Education” document published by the Distance Learning Secretariat of the Ministry of Education in Brasilia in August 2007 defines quality references through principles, guidelines and criteria [18]. The document states that it is not legally binding, but it recommends that its guidelines serve as a base or guide to help prepare government legal documents that relate to the specific processes of distance education regulation, supervision and evaluation. This quality referential is organized in eight interconnected categories or dimensions that fundamentally encompass teaching aspects, human resources and infrastructure, which should be addressed in the teaching policy of courses offered through the distance education modality. One of these dimensions is an EAD course’s communication system, the focus of this article, which according to the document should promote efficient communication channels, ensured by Information and Communication Technology (ICT) infrastructure, so that agents participating in the educational process (tutors, professors and students) can take advantage of and benefit from “two-way interactivity,” anywhere, at any time.

Silva, Melo and Muylder [31] emphasize the importance of identifying the critical success factors related to EAD, essential items in ensuring competitive performance, as well as tactical, executive and strategic objectives, in other words, the authors note that there are factors that need to be managed, regardless of the teaching institution or VLE platform being used. Conducting a bibliographic review of 71 articles cited in scientific productions on these factors, in the area of administration, they list among them communication providing unsatisfactory interaction.

Using the MEC/SEED quality referential model as a base and analyzing its criteria and reasoning, critical success factors related to an EAD course's communication system are mapped. According to this quality model: 1) Collaboration facilitates interaction between students by means of collective activities and through adequate instructional designs; 2) Interaction is related to the teaching and learning process, with ensured interactivity between participants of an EAD course, as well as course content; 3) Interactivity should take advantage of synchronous and asynchronous communication modalities, such as videoconferences and chats, among others, to promote real-time interaction between professors, tutors and students; 4) Cooperation and groups includes cooperation between students in study group formations and learning communities; 5) Deadlines and Punctuality means a communication system that allows students to quickly get answers to questions related to the teaching material and its content, as well as quantify the number of professors/hours available for the assistance required by students; 6) Participants and their Relations means understanding the profiles of those involved and quantifying the relations between tutors/students; 7) Etiquette is related to the adoption of attitudes of respect for and solidarity with others and different cultures; 9) in Development, students receive incentives and guidance on the progress of their studies; 10) the Calendar indicates locations, exam dates and deadlines for different activities; and finally, 11) Evaluation of the Institution and Professor or Tutor is related to supervision and analysis of tutors and improvement of the services provided by the institution.

The FCSSCs should be mapped in metrics that, continuously measured and accumulated over time, visualized and analyzed, transform into knowledge through human analysis. The metrics are derived from questions such as how, who, how much, which and when, for example: "What types of synchronous activities are designed in the course?" "Who participates in chats?" "How often is a page accessed?" "Who does the tutor give feedback to?" "Was the calendar accessed?" and "Did the professor access the student profile?"

These metrics, monitored and corrected, provide the basis for resolutions necessary to ensure the quality of the communication processes. The criteria to select the metrics are: easy comprehension, measurability, reliability, stability, administration, comparison and low maintenance cost (see Figure 1).

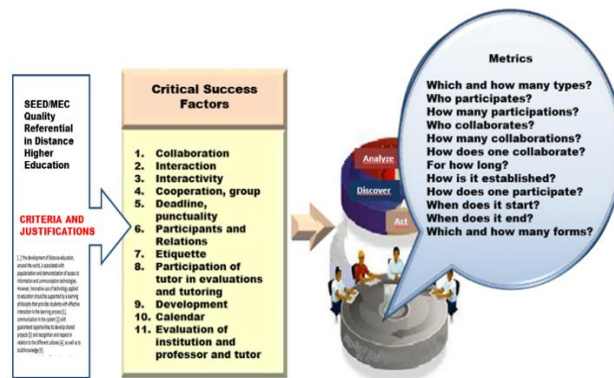


Figure 1. Critical success factors and metrics of the EAD course communication system

## V. BUSINESS INTELLIGENCE TECHNIQUES

Institutions' management systems can be automated by tools that extract, transform, analyze and mine data and make it available as evaluation instruments for managers. They are designed in accordance with profiles and needs, reflecting past, present and future situations of the institution [19]. These tools use Business Intelligence techniques to discover patterns and visualizations that make information available through consolidated and analytical reports, dashboards and warning reports, among others, allowing more assertive teaching, administrative, technical and humane management [39]. Sykurska [32] uses BI techniques that generate an environment where managers can monitor student, professor and tutor performance, compiling data on all classrooms, making management of the virtual environment possible in a way that VLE Moodle reports cannot.

BI architecture is composed of a variety of components, including: Data Warehouse and Data Mart, databases of this environment; extraction, transformation and loading tools; analytic tools that allow various types of analysis; data mining; report generation; and multidimensional visualizations; among others [22]. In this context, it is necessary to use the dimensional data modeling BI technique to structure the data of this environment and format it in dimensional databases, incorporating the needs and particularities of the business. Borges [5] addresses the need for a data model that represents the relationship between EAD course participants, their accesses, notes, forum and messages, but which is independent of the exploration technique used. Dimensional modeling is a technique that designs this data in structures made up of dimensions (dimension tables) and facts (fact tables), creating dimensional models (star, snowflake or constellation). The facts are formatted by metrics, generally numeric values. The dimensions have attributes that direct the analysis perspective and, according to Kimball [14], have a corporative character, since they are common to the various fact tables and can be shared.

The metrics prepared based on the FCCSCs should be mapped in these dimensional data models, therefore, in accordance with the guidelines of the MEC/SEED Quality Referential in Higher Distance Education, such that they are independent of the VLE environment being analyzed, but relate to the quality aspects selected for this referential and the success of the business. Other important components of this platform are the analysis, visualization and information acquisition tools. They may vary in proposal, type, technology and functionalities. They are basically Online Analytical Processing (OLAP), tools for data mining, statistics or report preparation and they benefit from the advantages offered by dimensional models, such as performance data navigation and the ease and comprehension they allow users, providing autonomous use and empowering information. In the first case presented here, the Pentaho Report Designer tool is used. This is a consolidated report builder – a tool for quantitative analysis of the “Interactivity” critical success factor. The second case, on the other hand, uses the Kohonen Self-Organizing Map (SOM) neural network to discover relationships between the variables: student performance (final grade point average), age, and “Interactivity and Interaction” variables, with visual results displayed in graphs, unique to this tool (see Figure 2).

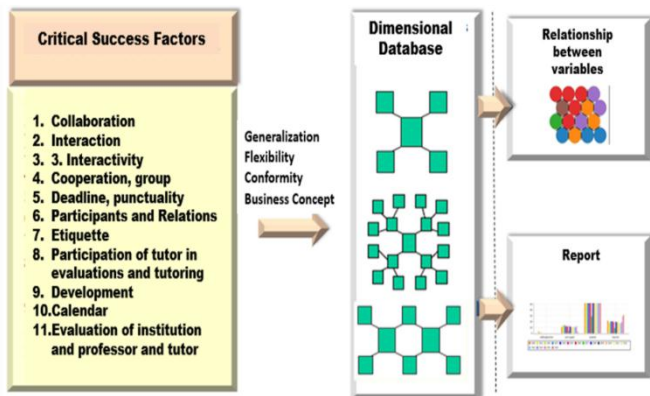


Figure 2. Critical success factors of the communication systems (FCSSC) and Business Intelligence techniques.

## VI. VIRTUAL LEARNING ENVIRONMENTS

Virtual Learning Environments offer a set of technological tools that make it possible for professors to develop courses on the most varied topics, allowing management, access control and recording of activities and evaluations. VLEs provide new educational possibilities, such as adjustment of work schedules to course activities and involvement of geographically disperse people, making the building of knowledge through the interactivity and interaction provided by the VLE feasible. There is also the option of customizing the educational process, with the student having greater autonomy to pursue his/her learning path, interacting with colleagues, tutors, professors and content in a collaborative and contextualized learning process, actively involving him/her in this process [2].

The databases of this type of environment store operational data (records of activity and resource access, communication and interaction data, registration data, and course data, such as modules, participants, group formations and performance, grades and participation) that can be processed [28]. This gives rise to certain questions: Which attributes related to the EAD course communication system are relevant and can be extracted from the VLE platform database? Are they sufficient and significant enough to support an evaluation in line with the MEC/SEED Quality Referential in Distance Higher Education? Are they verifiable, in other words, maintained in log registers for later analysis? Chen et al. [7] highlight the importance of choice and selection of the best data – that which supports instructor decisions – as well as the need for correct and visual understanding of the information offered. Another three procedures, important in building the VLE communication system evaluation process, and are illustrated in Figure 3:

- Analysis of conformity, identification and mapping of the FCSSC, constructed from the Higher Distance Learning Quality Referential [18], in the VLE platform of the course being evaluated with regard to application and use of its various functionalities and resources;
- Selection and identification of the VLE database attributes and tables, being evaluated, which will support the metrics and measures proposed by the FCSSC: are they verifiable? Are they current? Are they precise?;
- Construction of Extraction, Transformation and Loading (ETL) procedures, using BI techniques that will populate the dimensional models based on VLE environment operational data.

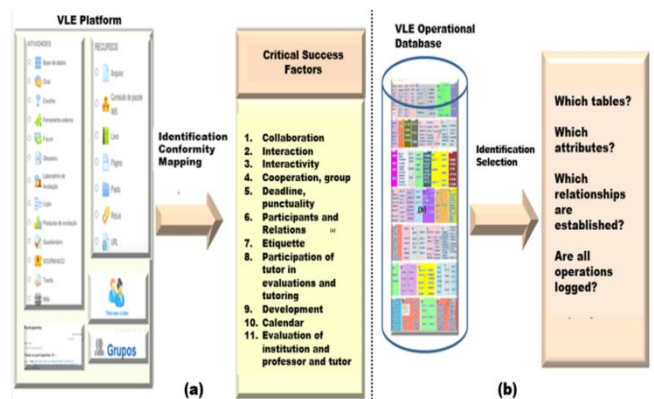


Figure 3. (a) Analysis of conformity, identification and mapping of the FCSSCs in the VLE platform being evaluated. (b) Selection and identification of database attributes and tables of the VLE being evaluated.

## VII. FIRST CASE STUDY

The purpose of this case study is to show the evaluation proposal presented here, quantitatively characterizing some

elements of the “Interactivity” critical success factor of the communication system of a course offered by a well-known teaching institution with extensive operations in the state of São Paulo. This study conducts field research by using data collection techniques to analyze data, done posteriorly, stored in a Moodle VLE database platform, coming from registration of activities of 2,961 participants in 17 subjects and 113 classes, and from quantitative exploratory research conducted in the following steps:

#### A. Software Installation

MySQL, a relational database server, (server version: 5.7.17-log); MySQL Workbench, a display tool to provide data modeling, development of queries in SQL, and administration of the database server (version: 6.2.3.12312 build 2280 (64 bits)); Pentaho Report Designer, a collection of tools to generate reports, (version: 7.0.0.0-25); Moodle, a software package provided free-of-charge as Open Source software (under a GNU Public License) (version: 3.0.12).

#### B. Identification, conformity and mapping of the “Interactivity” FCSSC in the Moodle VLE platform

The guidelines established by the MEC quality referential for the “Interactivity” FCSSC have support and complete compliance in the Moodle VLE, since this platform establishes communication mechanisms, both synchronous and asynchronous, allowing interactivity between course participants and content, through resources and activities that can be configured by the professors. Among the activities, forum, chat, glossary, data, wiki, choice, assign and quiz stand out. With regard to the resources, there are: Page, Folder, Book and Resource, all quantifiable [6] and [27].

#### C. Definition of metrics and the Dimensional Model

The physical design of the dimensional model, with its Fact and Dimension tables, was implemented in MySQL and is described in Table 1, with the respective attributes, metrics and tables involved in this study.

TABLE I. DIMENSIONAL MODEL, ATTRIBUTES, METRICS AND TABLES

Dimension and Fact	Attributes, Metrics and Tables of the Moodle VLE
<b>Dimension:</b> Dim_Course Information on the course, subject and classes	<b>Attributes:</b> Surrogatekey_course, nome_course, code_name, name_subject, name_class, date_start, date_end, date_update. <b>Tables:</b> course, modules, course_sections, course_modules
<b>Dimension:</b> Dim_Participant Information on course participants	<b>Attributes:</b> Surrogatekey_participant, code_participant, name_participant, code_role, name_role, date_start, date_end, data_update. <b>Tables:</b> user, role, context, user_enrollments, enroll, role_assignments
<b>Dimension:</b> Dim_Resource_Activity Information on the resources made available in the VLE environment: folder, page, book and resource Information on the activities made available in the VLE environment: assign, chat, choice, data, glossary, forum, wiki and qui.	<b>Attributes:</b> Surrogatekey_resource_activity, code_origin, name, code_type, name_type, date_start, date_end, date_update. <b>Tables:</b> logstore_standard_log, glossary, quiz, quiz_attempt, glossary_entries, comments, course_sections, course_modules, course, modules, chat, chat_messages, data_records, data, data_content, data_fields, page, url, folder, book, book_chapters, resource, wiki, wiki_pages, wiki_subwikis, forum_posts, forum_discussions, forum, choice_options, choice, assign, assign_submission, comments
<b>Dimension:</b> Dim_Time Information on time: indicates the time the facts occur.	<b>Attributes:</b> Surrogatekey_time, year, month, day, hour, minute, date_start, date_end, date_update
<b>Dimension:</b> Dim_Action Information on the action of the operation performed by the participant: viewed, created, uploaded, printed, submitted, among others. The “no access” action indicates an activity or resource available in the course that was not accessed. The “role without access” action indicates that the student belonged in the course, but did not access a resource or activity.	<b>Attributes:</b> Surrogatekey_action, code_action, name_action, date_start, date_end, date_update. <b>Table:</b> logstore_standard_log
<b>Fact: Interactivity</b> Measures which resources and activities are configured in the course to allow interactivity between participants. Measures the number of accesses of resources or activities by course participants. In the case of a “role without access” action, the quantity is equal to one (1), indicating that the student did not access the resource or activity. In the case of a “no_access” action, the quantity is equal to one (1), indicating that the resource or activity was not accessed by any student.	<b>Attributes:</b> Surrogatekey_course Surrogatekey_participant Surrogatekey_resource_activity Surrogatekey_time Surrogatekey_action <b>Metric:</b> Quantity. <b>Table:</b> logstore_standard_log

#### D. Extraction, transformation and loading

The processes of extracting, transforming and loading data coming from the Moodle VLE database and loading in the Dimensional Model use the programs constructed in the SQL language, through the MySQL Workbench tool.

#### E. Reports

The Pentaho Report Designer is a set of tools to generate reports, from the Pentaho suite, able to distribute them in various formats: html, pdf and xml, among others. Pentaho Report is used for complex solutions in both large and small

companies due to its easy use. It can connect to sources through a JDBC driver, offering link and graphic (pie, bar...) and formula possibilities, and can use the BI server as a repository of reports published for later access over the internet [21]. Figures 4, 5 and 6 summarize the display options of this tool.

- Course characterization:** Figure 4(a) shows a course consisting of 16 subjects, D00 to D16, and the number of classes for each subject. Figure 4(b) indicates the average resources configured per subject, where it can be seen that the course design prioritizes use of pages and folders, and other configurations in lower proportion, such as book and resource. The Figure 4(c) illustrates the design of the activities, by subject, in the form of assign, chats, choice, data, forums, wikis, glossary and quiz, with predominant use of the forum, assign and chat, present in all subjects. The classes of a subject have the same instructional design.
- Characterization of participants:** Figure 5(a) shows the number of students per subject, which varies from 100 to

125 students. The roles defined for the participants in this course are: student, teacher and editing teacher. The environment allows more granular displays: roles of participants in the D14 subject see Figure 5(b), or the relationship between the participants of a class (see Figure 5(c)).

- Characterization of accesses made by course participants:** there are countless possibilities for quantifying and visualizing interactivity in the VLE environment. Figure 6(a) indicates the percentage of access to forums by course participants, showing distinct use of this mechanism in the D16 subject.

However, it can be seen in Figure 6(b) that this resource is little used to analyze the number of forums configured for each subject and the number of respective participations. Figure 6(c) details the messages sent, in forums, by each student in the T1 class, in the D00 subject, during the month of February 2014. Figure 6(d) illustrated the average access of resources in the subjects.

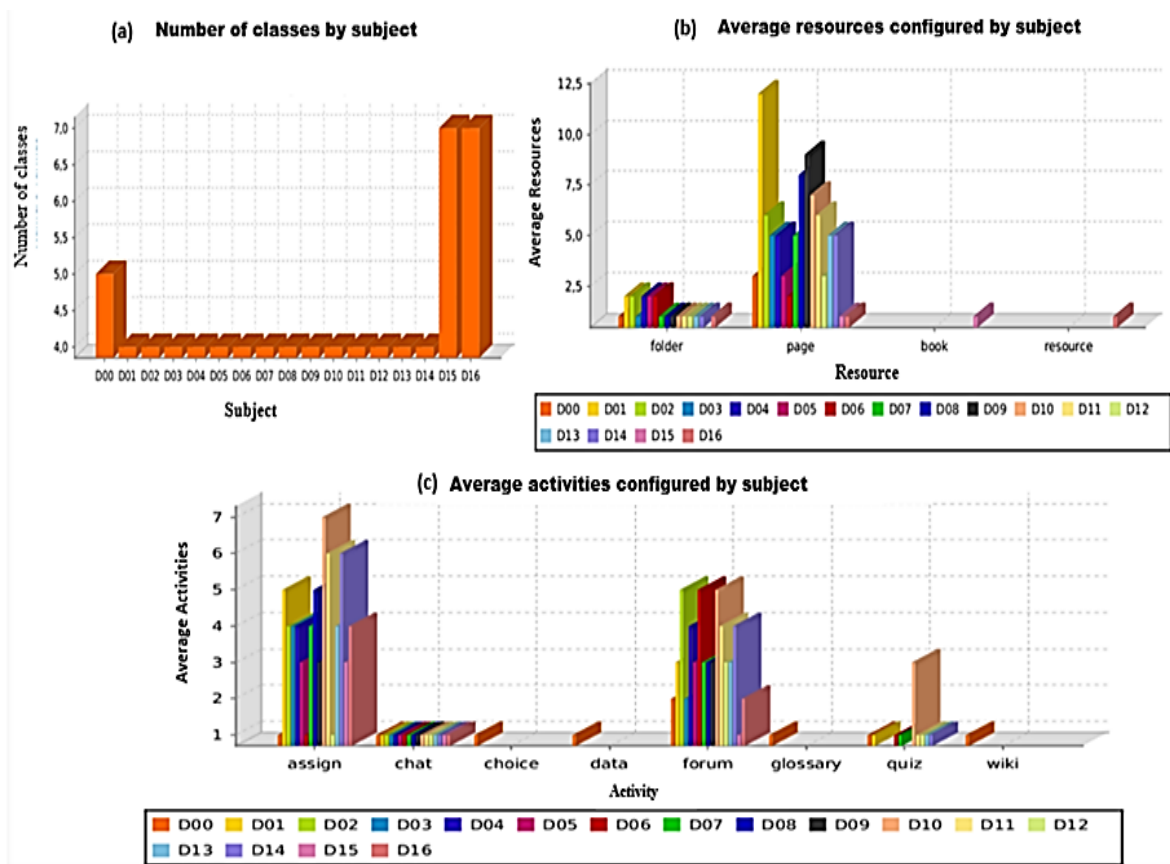
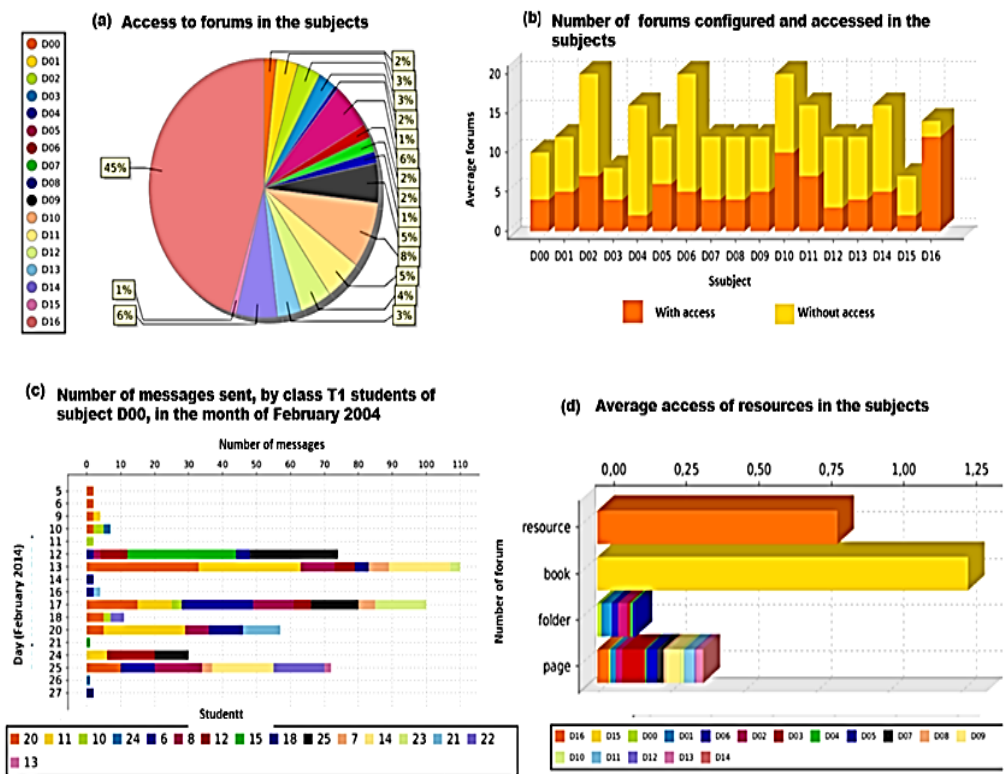
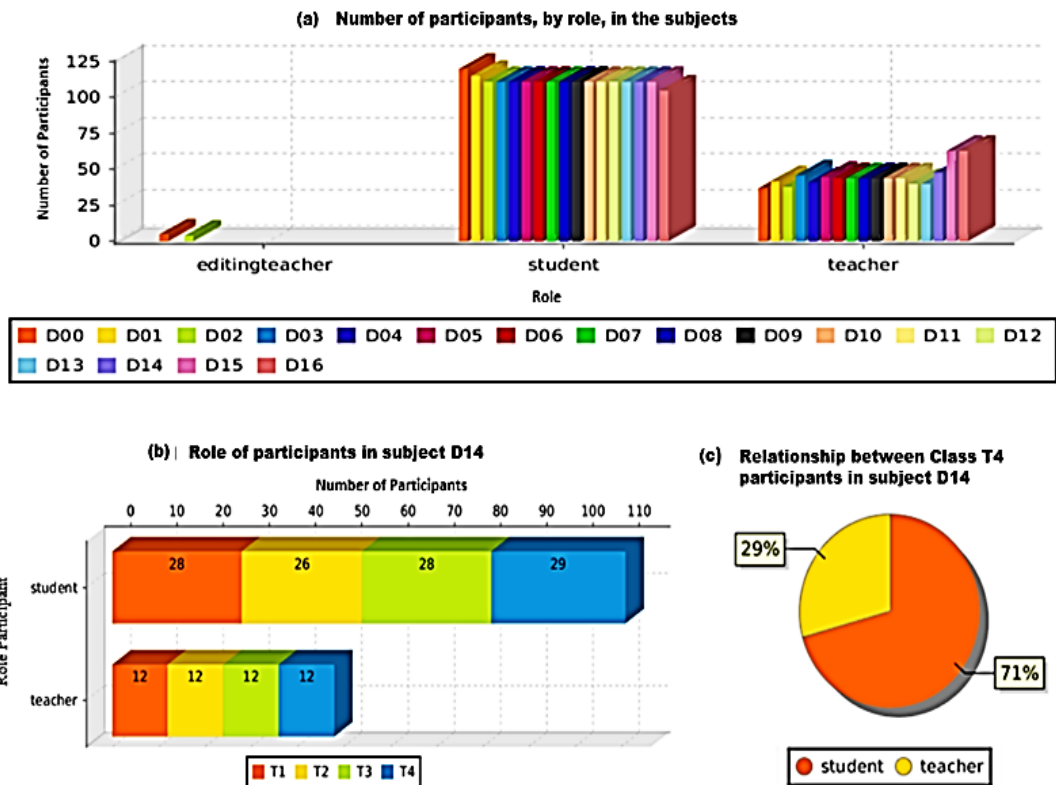


Figure 4. Course characterization. D00 to D16 are subjects.



## VIII. SECOND CASE STUDY

The second case study seeks to visually characterize the relationship between performance, age, and interactivity and interaction elements of 145 beginning students, using the Kohonen Self-Organizing Map neural network. This deals with a Bachelor's degree distance learning course offered by a private teaching institution that uses the Blackboard VLE platform. The study consists of 572 observations, in five subjects, with a duration of one month, offered during the first semester of 2017. Each subject is associated to one or more offers (classes).

### A. Software Installation

MySQL, a relational database server (server version: 5.7.17-log); MySQL Workbench, a display tools to provide data modeling, development of queries in SQL, and administration of the database server (version: 6.2.3.12312 build 2280 (64 bits); R version 3.4.1 (06-30-2017) with the Kohonen\_3.0.4 package [34] and [35]. R is free software that provides high-quality, flexible data display, calculation, manipulation and analysis functions that are easy to use and solid [8]. It uses various packages, including the Kohonen self-organizing map, which provide functions to create and process neuron network, and rich ways to display results [20].

### B. Identification, conformity and mapping of the "Interactivity" and "Interaction" FCSSC in the Blackboard VLE platform

This is a learning technology platform with three basic functionalities: teaching, communication and evaluation. When a student accesses the Blackboard VLE with user ID and password, the system exhibits course notices and subjects, providing a standard menu of tools, with options for announcements, calendar, tasks, view grades, personal information, goal performance and goals. The "Announcement" option has the function of exhibiting announcements that communicate important information on the course, subjects and institutional events. The "Calendar" option has the function of managing an activity calendar where the professor establishes the activities to be completed by the students. The "Tasks" option makes it possible for the professor to create tasks with established priorities and due dates, as well as monitor the status of each. The "View grades" option makes it possible for students to click on reports to see performance statistics and grades. The "Personal information" option allows students to enter data such as name and address, and change passwords. The "Goal performance" option allows students to check their performance, to see where they may need to improve. The "Goals" option indicates the objective of each subject chosen. After the initial screen appears, students may choose a subject to start the learning activities. By clicking on the online subject, the system displays the initial screen, with two blocks of options that can also be observed in the form of folders, and applications that are tools of the platform, for example: when students click on the "Click here to send a message to the teacher" option, the system presents a course message tool, allowing the sending and receiving of messages from/to the teacher; when students click on the forum option, the system presents a discussion area for related topics and for sharing ideas (discussion board), in an asynchronous form.

Here students can create multiple threads that are associated to an initial post and all its replies. In general, a subject has its content organized in six folders: general information, teaching material, calendar, subject content, talk to your teacher and technical support, which provide resources and activities that vary in number and type, depending on the design and intended purpose. For example, the Fundamentals of Mathematics subject provides external links to the virtual library and mental map, folders to organize learning units, links to platform tools, such as announcements, messages sent, tasks and evaluations, forums and documents such as manuals, theoretical material, tutorials and guidelines, among others. An example: when students click on the "Technical support" option, the system exhibits a document with information that includes the support line telephone number and guidelines on how to access the online academic center [38].

The interactivity and interaction elements, in conformity, mapped and identified for this study, include *assignments*, which are online works, tasks and evaluations; *external links and documents* related to course content; *announcements* of important course information; institutional subjects and events; a *discussion board*, which is a discussion area characterized by asynchronous communication through the forum; and *message to teacher*, which provides communication between the student and professor. These elements will be the variables of this study, hereinafter called: **AMTU**, the number of accesses to the "Click here to send a message to the teacher" option; **AFPA**, the number of responses sent by the professor to the student, in the forums; **AFAP**, the number of answers sent by the student to the professor, in the forums; **AFOR**, the number of accesses to discussion board; **ADOC**, the number of accesses to documents; **AAVI**, the number of accesses to announcements; **AASS**, the number of accesses to tasks and evaluations. Another two elements of this study are: **AGE** (of the student) and **MEFI**, final average grade of the student in the subject.

### C. Definition of metrics and the Dimensional Model

The dimensional model uses the same physical design as the first case study, and includes an additional fact table, that for performance. This table contains primary keys, surrogate keys, coming from the participant, course and time dimensions, as well as a grade metric (final average). The participant birthdate is included in the participant dimension table [14].

### D. Identification and selection of Blackboard VLE database table:

The dimensional model is loaded with data from the following Blackboard VLE operational tables: *course\_main*, with information about the course, subject and class; *course\_users*, users and institution\_roles, with information on the course participants; *course\_content*, *content\_handlers*, *application*, *forum\_main*, *conference\_main*, *msg\_main* and *navigation\_item*, with information on the resources and activities made available in the course; and *activity\_accumulator*, with information on the user actions in the Blackboard VLE platform.

### E. Data preparation

During the cleaning procedure, all observations with access values and final grades equal to zero, whose result was the



reduction of the number of entry records from 572 to 286, were eliminated. In the treatment of magnitudes procedure, the standard deviation normalization technique [37] is used to adjust the scales of the different magnitudes of the study variables.

#### F. Descriptive statistics of the study variables

Figure 7 shows the descriptive statistics that characterize each variable of this study, obtained using the R software [8]. Through the boxplots in Figure 7(a), one can see the differences between the various types of accesses made by students in the course activities and resources. The resources and activities most accessed are documents, tasks, announcements and forums, in this order. The maximum and minimum values, quartiles, mean and medians of each dataset are presented in Figure 7(c). Although the access amplitudes are significant, 75% have values much lower than the maximum value. There are high values that, although they indicate possible outliers, were considered normal data in the study. The age of the students varied between 18 and 61. The correlation between age and grade (final average) is -0.12, meaning that when one rises, the other falls. The final grades (final average) vary between 0 and 10, with 50% of the data between 0 and 5, in other words, the number of fail is high. There is a very large variation between access types: the maximum AFPA value is 2 and the maximum document access value is 393. However, this variation also depends on the number of resources available in each subject. The standard deviation shows dispersal around the data average of each variable, with document accesses standing out. The CV variation coefficient, or relative standard deviation, measures data homogeneity and is comparable among samples; in this study, all access variables have data that is not very homogeneous.

#### G. The Kohonen Self-Organizing Map

Or simply SOM is an artificial neural network with non-supervised learning that organizes patterns based on similarities found in the entry data, mapping the original space (entry) in a space defined by a neuron grid generally restricted to one or two dimensions, preserving notions of proximity (topological preservation) [36]. The SOM learning process is competitive and cooperative and throughout the learning the network creates specialized neurons in specific entries, which self-organize, producing important relationships between data, in other words, the network discovers, on its own, correlations, regularities and categories in the entry data and incorporates them into its internal structures [15].

#### H. Network Training

To train the network, 286 entry patterns were presented for each experiment, and the following network parameters were adjusted: learning rate, dimension, grade grid and number of iterations. These are important parameters, since there is no adequate progression in the network training if the number of iterations is insufficient; when the learning rate, which determines the network learning pace, is inadequate, it can cause instability or result in one neuron not learning over the others; the grade format determines the number of neighboring neurons (rectangular are four and hexagonal are six) and if

poorly dimensioned, it can lead to information loss or concentration of entries, or empty neurons. The resulting network has 42 neurons organized in a 7x6 grid with hexagonal topology. Figure 8(a) illustrates the number of observations mapped in each node, and Figure 8(b) exhibits the U-Matrix or Neighbor Distance, which shows the distance between each node and its neighbors. The competitive learning groups correlate data, thus dark (red) areas signify groups of similar nodes and light (yellow) areas signify dissimilar nodes, and are indicative of limits for possible clusters.

#### I. Visualization of patterns and relationships

Figure 9 shows the weight vector (codes) that represent the mapping of entries in a specific neuron and allow identification of patterns, magnitudes and relationships between variables. Figure 10 shows the heatmap for each variable, which makes it possible to display the distribution of a single variable, its relevance and regions of interest, in the different areas of the topological map. An analysis of the regions of these figures shows areas at the top, to the left, in which the final grades are very low, almost without accesses, and the ages are quite varied. Figure 8(a) also indicates a high number of entries mapped in the neurons of this area and Figure 8(b) shows that the neurons of this area are similar. The area located at the base, to the left, is characterized by high grades (final grades – MEFI). It concentrates low and median ages (AGE), and access to tasks (AASS) and forums (AFOR) stand out. In the area on the right, at the base, are several specific cases that combine various types of accesses, with regular grades (final average). For example, node 6 concentrates high values of document, task and announcement accesses, as well as access of student responses to professors, in the forums (ADOC, AASS, AAVI and AFAP). In the upper area, to the right, access to messages (answers) in the forums stands out, however the grades (final average) are low. In general, it can be seen that not having accesses by students characterizes, in fact, low final averages, however, the opposite is not seen.

Another type of analysis is to observe the gradation of heatmap colors, individually, by study variable. For example, observing the grade (final average) variable map, the colors range from red, which means lower values, to lighter yellow, which means higher values. The variation in the normalized value scale for this variable is shown at the left side of the map. A heatmap can be related to one or more variables. For example, the AASS and MEFI variables have more uniform distributions on the map; nodes 10 and 17 have the highest grade (final average) values, which can be observed in the heatmaps of all the other variables, and in both cases, orange-colored neurons for the task access, almost red for AAVI, ADOC and IDADE, and red colors for all the other variables stand out. Combining information from the heatmaps and code vector, with those of Figure 8(a), it can be seen that there is a concentration of lower grades (final averages) in the denser areas to the left, at the top. As the grade (final average) increases, the density decreases.

The environment allows exploration and evaluation of the data, with rich visual observations that promote discovery of relationships and patterns.

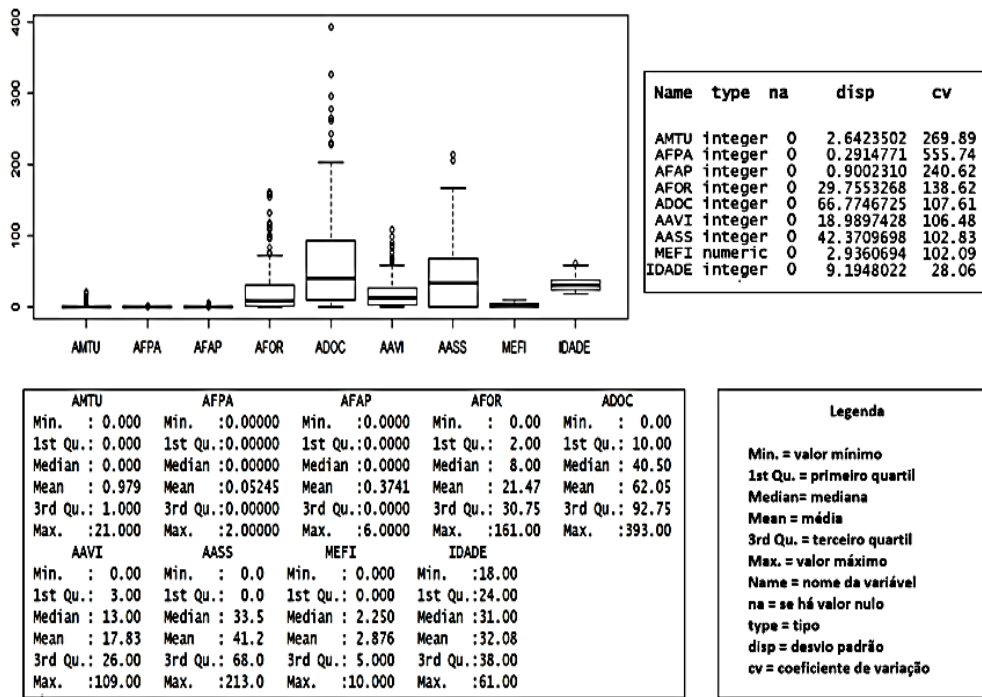


Figure 7. Descriptive statistics of the study variables. (AMTU, the number of accesses to the “Click here to send a message to the teacher” option; AFPA, the number of responses sent by the professor to the student, in the forums; AFAP, the number of answers sent by the student to the professor, in the forums; AFOR, the number of accesses to discussion board; ADOC, the number of accesses to documents; AAVI, the number of accesses to announcements; AASS, the number of accesses to tasks and evaluations. Another two elements of this study are: AGE (of the student) and MEFI, final average grade of the student in the subject).

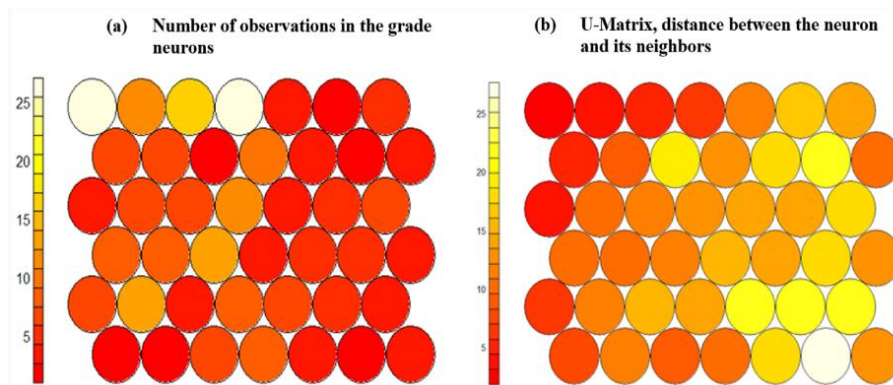


Figure 8. Descriptive statistics of the study variables (a) Number of observations in the grade neurons. (b) U-Matrix, distance between the neuron and its neighbors.

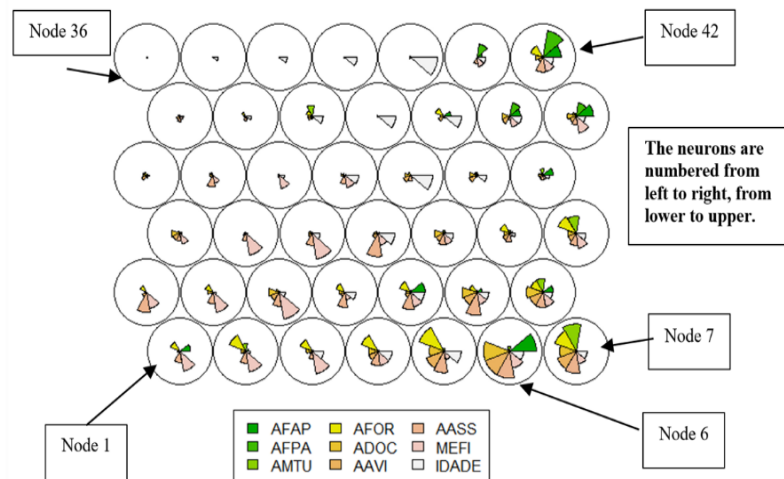


Figure 9. Weight vector. (AMTU, the number of accesses to the “Click here to send a message to the teacher” option; AFPA, the number of responses sent by the professor to the student, in the forums; AFAP, the number of answers sent by the student to the professor, in the forums; AFOR, the number of accesses to discussion board; ADOC, the number of accesses to documents; AAVI, the number of accesses to announcements; AASS, the number of accesses to tasks and evaluations. Another two elements of this study are: AGE (of the student) and MEFI, final average grade of the student in the subject).

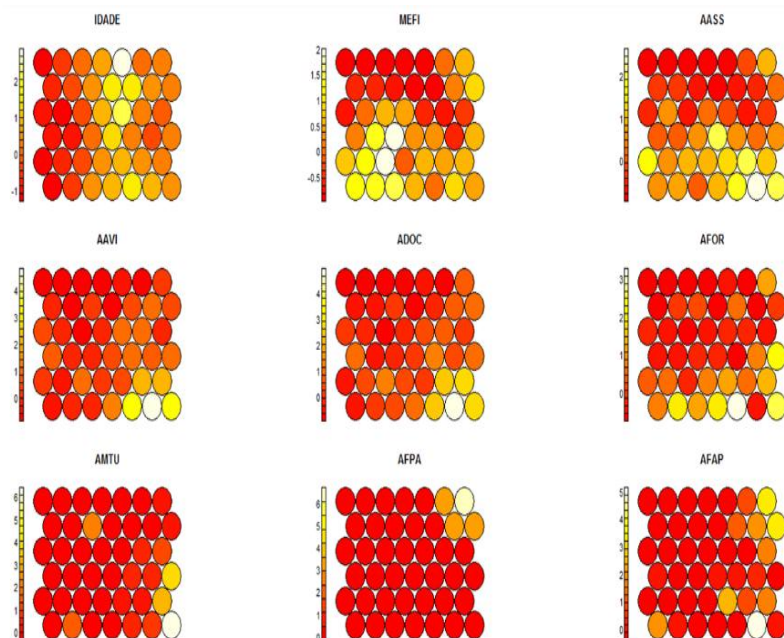


Figure 10. Heatmap of each variable. (AMTU, the number of accesses to the “Click here to send a message to the teacher” option; AFPA, the number of responses sent by the professor to the student, in the forums; AFAP, the number of answers sent by the student to the professor, in the forums; AFOR, the number of accesses to discussion board; ADOC, the number of accesses to documents; AAVI, the number of accesses to announcements; AASS, the number of accesses to tasks and evaluations. Another two elements of this study are: AGE (of the student) and MEFI, final average grade of the student in the subject).

### IX. CONCLUSION

The evaluation process is a complex matter, with several approaches by different authors and their works. This article presents a proposal for an evaluation environment for Higher Distance Education Course Communication Systems, guided by the MEC/SEED quality referential and BI techniques, as well as characterization of the “Interactivity and Interaction”

Communication System Critical Success Factors (FCSSC) of distance education courses, obtained through this environment.

This article reviews several works observed in the scientific literature and inserts the Quality Referential in Distance Higher Education of the MEC/SEED [18] in the context of the BI environments found, since as a document established by the MEC/SEED, whose criteria and recommendations guide

educational evaluation processes, it serves to establish the critical success factors of a communication system for EAD courses and, consequently, their mapping in metrics that, through monitoring and evaluation, support continuing efforts to improve the quality of these courses.

Virtual Learning Environments (VLE) provide resources and activities that promote structuring of individual and collective knowledge and are one of the pillars of communication system success, providing means and forms of interaction, interactivity and communication between course participants and content, promoting collaboration, cooperation and formation of individual and collective knowledge. The evaluation of communications systems in these environments, from the perspective of critical success factors, leads to a process of analyzing conformity, adherence and mapping between the communication mechanisms existing in the environment and the critical success factors.

A BI environment is composed of tools capable of extracting, transforming, cleaning and loading data into dimensional models that structure the data in dimensions and facts, allowing and facilitating differentiated and flexible views of the business that, when manipulated by intelligent tools, empower and give user autonomy to course managers and participants. As a first case study, this work prepared reports characterizing the "Interactivity" factor, based on data produced by a Moodle VLE database, using the market-leading Pentaho Report Designer, a tool to prepare reports whose results allow one to conclude that the evaluation environment proposed in this work helps manager make better decisions. The second case study is presented using Kohonen self-organizing maps, which allow exploratory visual analysis of the relationships between elements of interactivity and interaction and student performance and age.

One way to follow-up on this work would be to expand the study of critical success factors and metrics, addressing the other dimensions of the MEC/SEED Quality Referential, in a study of other VLEs, in the enrichment and expansion of the dimensional model with new attributes and hierarchies, facts and dimensions, focusing on completeness, flexibility and adherence to the other FCSSCs and use of other BI tools.

#### REFERENCES

- [1] Associação Brasileira de Educação a Distância Censo, "EAD.BR: Analytic Report of Distance Learning in Brazil 2015", [http://abed.org.br/arquivos/Censo\\_EAD\\_2015\\_POR.pdf](http://abed.org.br/arquivos/Censo_EAD_2015_POR.pdf).
- [2] Almeida M. E. B. de, "Educação a distância na internet: abordagens e contribuições dos ambientes digitais de aprendizagem," *Educação e pesquisa*, Vol. 29, No. 2, 2003, pp. 327-340.
- [3] Baker R. S. J. de, Isotani S., Carvalho A. M. J. B. de, "Mineração de dados educacionais: Oportunidades para o Brasil," *Revista Brasileira de Informática na Educação*. Vol. 19, No. 2, 2011, pp. 2-13.
- [4] Bertolin J. C. G., De Marchi A. C. B., "Instrumentos para avaliar disciplinas da modalidade semipresencial: uma proposta baseada em sistemas de indicadores," *Avaliação: revista da avaliação da Educação Superior*, Vol. 15, No. 3, 2010, pp. 131-146.
- [5] Borges V. A., Nogueira B. M., Barbosa E. F. 2016. "A multidimensional data model for the analysis of learning management systems under different perspectives" In *Proceedings of IEEE Frontiers in Education Conference, Pennsylvania, US, 2016*, pp. 1-2.
- [6] Büchner A., *Moodle 3 Administration*, 3rd ed. Birmingham: Packt Publishing Ltd, 2016.
- [7] Chen Y. et al. 2016. "Instructional dashboards to support deep learning in an online problem-based learning environment," In *Proceedings of the Pre-conference Workshop*, June 2016, 6-12, [12th International Conference of the Learning Sciences, Singapore, Asia, 2016], <https://pdfs.semanticscholar.org/5955/8831abff01342779cbbb49f8b9758ad0f83d.pdf#page=11>.
- [8] Crawley M.J., *The R book*, Chichester, West Sussex United Kingdom: John Wiley & Sons, Ltd., Publication, 2013, pp.1-951.
- [9] Cunha J., "Educação a distância cresce apesar da crise," *Folha de S. Paulo*, São Paulo, 14 jun. 2016.
- [10] Desai M S., Hart, J., Richards, T., "E-learning: Paradigm shift in education," *Education*, No 2, 2008, pp. 327-334.
- [11] Fiorillo A., Maccari E. A., Martins C. B., "A Ead no Brasil e a importância de competências subjacentes para o reconhecimento de cursos de acordo com a percepção de coordenadores de curso," *RAEP Administração: Ensino & Pesquisa*, Vol. 16, No. 1, 2015, pp. 141-178.
- [12] Ichihara, A. T. and Omar, N. 2017. "Proposta de um ambiente para avaliação do sistema de comunicação de cursos ead baseado em técnicas de business intelligence e em referências in de qualidade do mec/seed," In *Atas das conferências IADIS Íbero-Americanas www/internet 2017 e computação aplicada 2017*, Vilamoura, Algarve, Portugal, 2017 pp. 45-55.
- [13] Junior G. R. 2009. "Referenciais para avaliação da qualidade de cursos de educação superior a distância nas universidades públicas brasileiras", In *Proceedings of IX Colóquio Internacional sobre Gestão Universitária na América do Sul, Florianópolis, Brasil, 2009*, pp. 25-27.
- [14] Kimball R. and Ross M., *The Dattawarehouse Toolkit: the complete guide to dimensional methodology*. New York: Wiley Compute Publishers, 2002.
- [15] Kohonen T., "The Self-Organizing Map," *IEEE Transactions on Data and Knowledge Engineering*, Vol. 78, No. 9, 1990, pp 1464-1480.
- [16] Lesjack B and Florjanči V.. 2014. "Evaluation of mooc: hands-on project or creative use of ICT in teaching," In *Proceedings of the Management, Knowledge and Learning International Conference, Zadar, Croatia, 2014*, 1147-1155.
- [17] Martins, L. B., Zerbini, T., "Evidências de validade de instrumentos de reações no ensino superior à distância," *Estudos e Pesquisas em Psicologia*, Vol. 15, No. 1, 2015, pp. 116-134.
- [18] MEC/SEED - Ministério da Educação. Secretaria de Educação a Distância, "Referenciais de qualidade para educação superior a distância. Brasília". MEC/SEED, 2007.
- [19] Pascal G. et al., "Aplicación de Business Intelligence para la toma de decisiones en instituciones Universitarias. Implementación de Boletines Estadísticos en la Universidad Nacional de Lomas de Zamora (UNLZ)," *Revista Electrónica sobre Tecnología, Educación y Sociedad*, Vol. 4, No. 7, 2017.
- [20] Peng R.D., *R Programming for data science*, <https://www.cs.upc.edu/~robert/teaching/estadistica/rprogramming.pdf>.
- [21] Pentaho. Pentaho Reporting User Manual, <https://wiki.pentaho.com/display/Reporting/Pentaho+Reporting+User+Manual>.
- [22] Pifarré K. T., "Metodología para la creación de conocimiento en los entornos virtuales de enseñanza mediante herramientas de Business Intelligence. Caso de estudio en una institución educativa que emplea Moodle," PhD diss, Universitat Politècnica Catalunya, 2015.
- [23] Polak Y. N. de S.; Munhoz, A. S.; Duarte, E. V. G., "Referenciais de qualidade para cursos em EAD: dificuldades e desafios," *Rev. Diálogo Educ.*, Curitiba, Vol. 8, No. 24, 2008, pp. 473-483.
- [24] Rabello C. R. L., "Aprendizagem na Educação a Distância – dificuldades dos discentes na licenciatura em ciências biológicas na modalidade semipresencial," Masters diss, Universidade Federal do Rio de Janeiro, 2007.
- [25] Ramos J. L. C, Rodrigues R. L., Silva J. C. S., Gomes, A. S., "Analisando fatores que afetam o desempenho de estudantes iniciantes em um curso a distância," In *Proceedings of III Congresso Brasileiro de*

- [26] Rochefeller A. N. N. and Costa, I. L., “Sistema de tutoria a distância: análise da mediação em fóruns colaborativos em ambientes virtuais de aprendizagem – AVA,” EAD em FOCO, [S.l.], Vol. 6, No. 2, 2016, pp. 133-154.
- [27] Romain S., Moodle for teachers: reference manual. Genebra, Suíça: Graduate Institute of International and Development Studies, 2015, pp.1-31.  
[http://graduateinstitute.ch/files/live/sites/iheid/files/sites/students/users/admin\\_it/public/it\\_public/moodle/it\\_301\\_moodle\\_teachers\\_reference\\_manual\\_en.pdf](http://graduateinstitute.ch/files/live/sites/iheid/files/sites/students/users/admin_it/public/it_public/moodle/it_301_moodle_teachers_reference_manual_en.pdf).
- [28] Romero C.; Ventura, S.; García, E., “Data mining in course management systems: Moodle case study and tutorial,” Computers & Education, No. 51, 2008, pp. 368-384.
- [29] Santana M. A., Neto B. F. dos Santos, Costa, Evandro de Barros, 2014. “Avaliando o uso das ferramentas educacionais no ambiente virtual de aprendizagem Moodle”. In *Anais do XXV Simpósio Brasileiro de Informática na Educação. São Paulo, Brasil, 2014*, pp.1-10.
- [30] Santos A. A. da S, Fagundes A. I. J., Oliveira C. das G., Costa S. G.da., 2016. “Avaliação de polos de apoio presencial de Ead: um estudo comparado”. In *Anais of Simpósio Internacional de Educação a Distância-Encontro Internacional de Educação a Distância, São Carlos, Brasil. 2016*.
- [31] Silva M. P. D., Melo, M. C. O. L., Muiylder, C. F. de, “Educação a distância em foco: um estudo sobre a produção científica brasileira. RAM”, Rev. Adm. Mackenzie, São Paulo, Vol. 16, No. 4, 2015, pp. 202-230.
- [32] Sykurska J., 2016 “Data mining in learning analytics” Masters diss, Universitat Polytècnica de Catalunya, 2016.
- [33] Turrioni A. M. S., Stano, R. C. M. T., 2009. “Critérios de avaliação para a educação a distância”. In *Proceedings of XXII Congresso Ibero-Americano de Educação a Distância*, Manaus. 2009, Vol. 22, pp. 56-62.
- [34] Venables W.N. and Smith, D. M., An introduction to R.network: notes on R: A programming environment for data analysis and graphics. *Version 3.5.1*, 2018, <https://cran.r-project.org/doc/manuals/R-intro.pdf>.
- [35] Wehrens R 2018. Supervised and unsupervised Self-Organising Map. Version 3.0.6, 2018, <https://cran.r-project.org/web/packages/kohonen/kohonen.pdf>.
- [36] Wu S.; Chow, T.W.S., “Clustering of the Self-Organizing Map using a clustering validity index based on inter-cluster and intra-cluster density,” in *Pattern Recog*, No. 37, 2014, pp. 175–188.
- [37] Zaki M.J. and Junior W.M., Data mining and analysis: fundamental concepts and algorithm. New York, USA: Cambridge University Press, pp. 52-53.
- [38] Zanden P. V. D. and Veen W.. 2007. “Educational parameters revealed from vle logging data. In Proceedings of 7th International Educational Technology Conference (IETC), Nicosia, Turkish Republic of Northern Cyprus. 2007.
- [39] Zapparolli L. S., Stiubiener I., 2016. “Management support tool in virtual learning environments using Moodle as a case study”. In *IEEE Conference Publications of XI Latin American Conference on Learning Objects and Technology*, San Carlos, Alajuela, Costa Rica., 2016.