EduVis: Interactive Visualization of Educational Data

Vilma Jordão, Daniel Gonçalves

Instituto Superior Técnico, Universiade de Lisboa, Lisboa, Portugal

e-mail:vilma.jordao@tecnico.ulisboa.pt

Abstract

A successful analysis of the educational processes have a great relevance for the academic community, providing useful tools for identifying problems and atypical situations. Through data mining techniques, it is possible to identify patterns, and make performance predictions that students may have. Despite these advantages, such analyses consist in very large data sets, which easily reach magnitudes of the orders of thousands, hindering or preventing its overview. Thus, information visualization may overcome this kind of limitations due to its potential to display large quantities of data while alleviating cognitive load.

For this thesis, a multi-matrix visualization mechanism is proposed and implemented, which allows to observe the academic journey of students as well as to infer the various relationships and precedence of existing subjects. The proposed solution is within the Educare Project, and uses extracted patterns through data mining techniques, from collected data from the course of Informatic Engineering and Computers in Instituto Superior Técnico, University of Lisbon.

In this context, our solution was developed under an already existing visualization, in which the proposed mechanism aims to complement with new functionalities, with the ultimate goal of creating an integrated solution, able to give an effective response to the visualization of educational patterns.

I. Introduction

The number of students has grown considerably over the last decades in traditional and online teaching [2]. As a result, the CMS (Course Management Systems) and LMS (Learning Management Systems) became a very popular and had a large impact on distance learning [5].

With the difusion of this kind of tools, together with the increased rate in school enrollment, [2], emerges a big quantity of raw data from the students curriculum, so this information is increasingly accessible to be collected and analyzed. Doing so is essential for providing the academic community means to discover and analyze patterns, thus unlocking the possibility to the gain of knowledge and understanding about the academic performance of students.

The application of data mining techniques in this context, is an emerging research field, that has been providing means to analyze educational data, from the behavior of the students until teaching strategies, and courses coordination, assuming the denomination of EDM [18] (Educational Data Mining). However, the result usually consists of an extensive set of behaviors, which are described in the form of textual patterns , that are typically difficult to understand and relate due to its visual complexity.

Additionally, the understanding of this kind of information, often requires a reasonable knowledge of complex algorithms and mathematical statistics, whitch the user may not possess [6]. In that field, the information visualization has gained prominence in data representation. Its aim is to help understand large amounts of data, increasing the capacity of the human visual system to discover trends, patterns and outliers [4].

Hereupon, if there is a well designed visualization, cognitive mechanisms can be replaced only with perceptual deductions, enabling the discovery of emergent properties, which were then unknown because of their complexity, showing new problems [11], where just looking for a set of statistics, would be more difficult to identify. Taking into account the requirements described above, it was decided to use the advantages that information visualization offers, creating a visual metaphor, that effectively represents educational patterns, making its evaluation and analysis clear and effective, benefiting the professors and students.

In this sense, within the project Educare, were collected data from the first three years of the academic journey from the course Informatic Engineering and Comupters, by the tem of Professor Cláudia Antunes, from Instituto Superior Técnico.

Using the collected data, we present in this paper, a new mechanism, composed by several matrices, which highlight what are the subjects with more or less precedences, and that are present in most patterns. This new mechanism is intended to complement the previous visualization, also developed within project Educare, provinding the interpretation of the various existing relationships and dependencies between different subjects in several semesters. It also shows the subjects involved in more patterns and, which may be considered the most important ones, to get approval in future disciplines.

II. Related Work

With the growing number of students in traditional and online teaching several tools have been created within the framework of graphical representation of educational information.

Taking into account the online education, the CMS allows the creation of virtual classrooms where students and teachers can share information, provinding remote participation in discussions and management of the classes.

This type of interaction generates large amounts of data that need to be managed in a manner that provides relevant information to teachers about the students' performance.

To overcome this challenge, CourseVis [1] was created and is being used as an extension to the CMS that enables interactive data exploration and manipulation through different visualization mechanisms. One of the mechanisms is a three-dimensional representation of participants in a forum, in which topics are represented as spheres proportional to the number of students involved, allowing zoom and panoramic view. Another mechanism is the cognitive matrix, consisting of a matrix in which the names of the students are represented along one axis and the concepts of the discipline on the other axis, being associated with a color scale to performance, ranging from green (success) to red (failure).

A third tool, which depicts the behavior of students through graphics and text arranged in matrix form, aims to show various types of information such as contents, attendance and academic progress. A study with users showed that graphical representations of CourseVis allows to obtain information about the cognitive aspects of students quickly and accurately. However, users had some problems in understanding the information due to overlapping of visual elements, hampering the interpretation of the graphs that represent the behavior of students [8].

These limitations led to the creation of GISMO (Graphic Interactive System Monitoring Students) [7], which visually represents data from LMS, quite complex and difficult to read and to understand. The GISMO is integrated in LMS Moodle but can also be adapted to other learning platforms. Focused on a two dimensional viewing behaviors, GISMO offers a single view to specific parameters. It allows interactive exploration of access and details of resources, and provides the means for exploring the behaviors of students who had been considered relevant in CourseVis [1]. Having been used in an online course, the GISMO has proved to be effective for the understanding of individual behavior, as well as the methods of assessment, steering to a redesign of the course according to the needs of students.

Given those traditional teaching results from evaluation processes such as evaluation scores and other competencies, such as attendance and participation, the resulting information can be complex and thus difficult to interpret. Additionally, the discovery of patterns and information exchange can be almost impracticable in this context. One of the studies that have been done in the context of information visualization to support educational processes is AVOJ [3]. In order to reflect the different abilities of students, this tool allows the exploration and visualization of performance data, providing the ability to group students according to their grades and other features such as study habits. However, it makes use of a color scheme with maximum levels of brightness and saturation, competing for visual attention of the user, which makes it more difficult to find patterns [8].

The system provides an additional mechanism to compare general trends, consisting of a bar graph that shows real-time statistics display, making it possible to understand some specifics, such as how students manage their study sessions. The authors believe that once providing a broad set of data that was not previously available, it will be possible to obtain more information about educational processes and improve the effectiveness of teaching. Trimm et al [9] created a display in which students are grouped according to their scores.

These groups can be viewed using compositions that show their characteristics and variations in time. In order to use these composition, historical information of each student shall consist of a two-dimensional trajectory, represented in two axes. This representation, although being simple, is not very efficient to reveal trends and characteristics of a large amount of data from students. The visualization uses spatial capabilities to represent students' trajectories over time using the level set algorithm. A technique of color blending is used to show features such as the average and standard deviation in relation to the results of the students, representing the information in a natural form, and is used the color weaving technique to show the value of a given attribute for a randomly selected trajectory. However, the use of gradients can add noise to the image and the variation of brightness and saturation can also cause the illusion of proximity, leading to an inaccurate interpretation of the data. However, testing with users has shown that this view provides a means to find new relevant standards in the data. In order to understand the school drop-out of a significant number of students for a course of computer science, a tool has been developed, that allows the visualization of repeating patterns for success or failure [12].

Due to the large amount of information available, that tool used a visual structure of nodes and edges. The nodes represent events, with width proportional to the number of students from each event, and the edges represent the trajectory of the students in relation to the events (exams or assignments), being the width representative of the number of students to be described in this behavior. The colors represent the performance of students, allowing to differentiate groups of students, identifying those who have similar behavior. The display is interactive, allowing the selection of categories of students, such as *students who repeat at least one discipline* or *students who never failed a subject*. As a result, it was possible for teachers to obtain a set of conclusions regarding repetitive failures or implications of a particular subject in the success of other disciplines. The visual structure has, however, an important gap regarding the overlapping of thicker lines over the thinner ones, which can make it difficult to understand any information.

All the approaches mentioned focus on the visualization of information regarding educational processes, traditional or online. They present different techniques and methodologies to visualize the data in this context. Given the scope of the present study and the characteristics of the data we want to analyze, in which a visualization has increased relevance for studying the relation of disciplines and their interdependencies. The work of *Wortman* and *Rheingans* [12] stands out and, more recently, the Trimm et al. [9]. An interactive visualization that allows interactive mechanisms, such as featuring, comparison and filtering, is of extreme importance in the context of this study. However, the later work above mentioned, Trimm et al. [9], offers limited interaction on the particularities of our context. In addition, the study of Wortman and Rheingans [12], although providing interaction mechanisms, does not allow comparison between disciplines or highlight specific patterns. In addition, the color scheme used does not alleviate visual clutter that is present when many different edges are represented. In that way, a visualization technique was developed that represents the relationships between the disciplines of a study program in an attempt to bridge the gaps in existing solutions in this context.

III. Educare Project

Educare project consists in the visualization and modelling of the behaviour in education. Its aim is to provide to the academic community a decision support system, having as one of its main goals, a set of techniques for discovery and information visualization. On a first stage, Eng[a] Sandra Gama developed a visualization designated by Multi-Layer, illustrated in Figure 1

Figure 1: *Multi-Layer Visualization*

This mechanism allows to easily explore the temporal dimension, which corresponds to the choice of various disciplines, held by students at a given instant of time, giving rise to immediate visualization of the various routes followed by the them, throughout the course, as illustrated in Figure 2

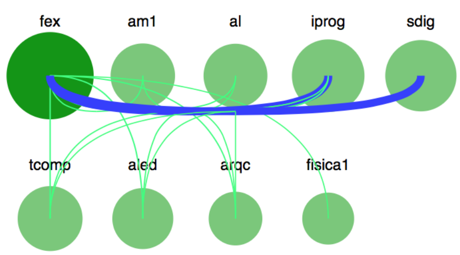

Figure 2: *Exploitation of temporal and dynamic aspects of the discipline fex*

Despite these features, the available data may show more information, which this mechanism fails to cover, due to the lack of structure from it's visual artifacts, to represent it.

In order to solve that problem, a new visualization has been developed in order to complement the previous one, filling its gaps. This new mechanism is designated by Multi-Matricial Visualization and is tackles the challenge proposed in section *1. Introduction*, built on the basis of the teachings of section *2. Related Work*. It was developed with the intention of being integrated into the previous mechanism,so that both worked in a complementary way, fluidly and coordinated.

I. Educational Patterns

In this work, the result of sequential data mining, which had previously been applied to data collected during 9 years, over the first three years of the academic path of the students of the degree in Informatic Engineering and Computers. The purpose of sequencial data mining is, given a set of sequences and a minimum support threshold, to discover the set of sequences that are contained in at least δ data set sequences, that is, the set of frequence sequences [15]. This method not only meets the expected standards based on background knowledge but, with the use of relaxations, it also allows the discovery of patterns that correspond to deviations from the expected behavior, making evident a few potentially relevant trends that were previously unknown.

Thus, to obtain the final data, three support threshold values were used, which represent the percentage of students, among all, that satisfies the pattern. In this case, three different set of patterns resulted. The lower the value of the threshold of support, the greater the number of patterns resulting from the application of sequential data mining. The patterns are generated in textual form and have the following structure:

Pattern i = (semester 1,...,semester, N total students),
Semester j = Subject1 OR (subject1,..., subjectN);

Example:

['am2', '1683'] : 1683 students who got approval in *am2*.

['fex', ['tcomp', 'aled'], '1168'] : 1168 students who got approval in *fex* in the first semester, and in *tcomp* and *aled* in the second semester.

['fex', ['∼arqc', '∼fisica1'], '591'] : 591 students who got approval in *fex* in the first semester but failed in *fisica1* in the second semester. In this last case, the disapprovals are represented by

the symbol "∼".

Although the textual information makes it difficult to understand specific patterns and provide little information overall, this pattern structure provides information about the relationships between the various semesters of the course that an effective visualization will be able to make evident.

IV. Proposed Solution

To arrive to the final visualization, there was the need to overcome various challenges, for it to be effective. These challenges eventually become requirements so that the usability of the visualization would be compromised. These are their descriptions:(i) Give the notion of sequenciality between semesters. The user should understand the relations among subjects in a specific pattern, over several semesters. (ii) See the patterns that correspond to each discipline. (iii) Distinguish the amount of patterns in which a subject is involved, through the brightness of colors. (iv) Understand which are the subjects that are necessary to have approval in order to succeed in future subjects. (v) Filter patterns. (vi) Have two patterns selected simultaneously, so that it is possible to compare the relations of two subjects with distinct patterns. (vii) Do the blend of two colors, when there are disciplines that overlap in different patterns.

I. Low Fidelity Prototypes

Taking into account the different types of information that is possible to extract from the educational patterns, and the need to create a visualization that shows them properly, we used an iterative and incremental design with strong emphasis on prototyping.

In this sense, severeal low fidelity prototypes have been created, in order to explore different ideas and views. These sketches allowed the testing of different visions and perspectives, and helped to decide what would be the best starting point to begin the development of the system.

Figure 3 displays the prototype that, after later iterations, led to the final visualization. In this prototype, the semesters are separated by vertical columns, in the center of the matrix there is a horizontal partition that separates the patterns referring to approvals and failures. The lines shows the relationships between the different subjects

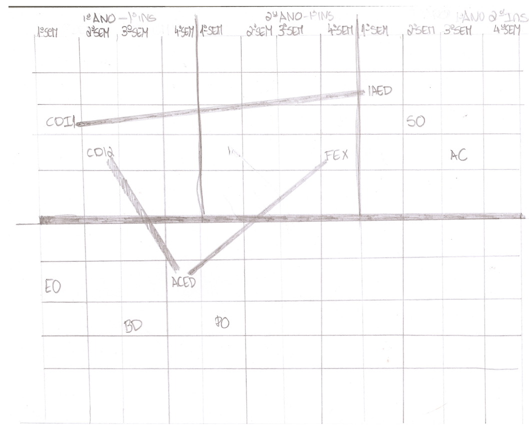

Figure 3: *LFP Matrix*

II. EduVis

After having decided how the structure of the visualization would be, through the analysis of the low fidelity prototypes, its development started. This new visualization, after being completed, was integrated into the mechanism that had already been developed, Multi-Layer Visualization, in the context of the Educare project. Thus appeared EduVis, which corresponds to a coordinated view, consisting of two integrated mechanisms, and that takes advantage of its conjugation. The two visualizations have different characteristics, and are intended to complement each other in the information that they extract and transmit from the analyzed patterns

III. Multi-Matricial Visualization

In this visualization, the matrices represent the subjects of the course, and each subject corresponds to a square, which is divided into two triangles: the upper triangle represents approvals and the lower one the failures. Before any kind of interaction, *i.e.* the initial state of the visualization, as seen in Figure 4, the subjects on which information exists are represented by colors between yellow and dark blue, while the disciplines without associated patterns are represented in grey. The brightness represents the number of patterns in which the subjects are involved: the lower the brightness(darker color), the higher the number of relations between the subject under analysis and the other subjects. If there are few or no dependencies, then the subject is associated with a higher brightness(lighter color).

Figure 4: *Heatmap, initial state*

IV. Interaction

The Multi-Matricial mechanism allows some kinds of interactions, in a way to motivate the user to explore the visualization by himself, and making the interpretation of the results more effective [10].

IV.1 Mouseover

This type of interaction highlights all the dependencies of a particular disicpline. In this way, to observe what patterns are associated with the approval of a subject, the cursor can be placed over the upper triangle of the subject in question, getting visible all its relations at the level of approvals and failures. To demonstrate these dependencies, two colors are used, blue and red, representing respectively, approvals and failures, as displayed in Figure 5. The luminance of these two colors reflects the quantity of relations/dependencies. So, the brightness decreases with the growth of relations and increases in the inverse situation. Fully saturated colors are not used, in order to avoid competition for the user's attention.

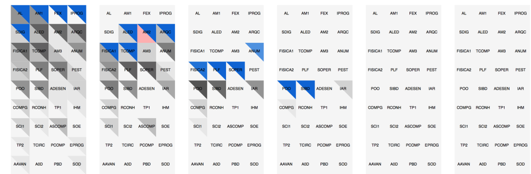

Figure 5: *Relations of the subject am1*

When the cursor leaves the visualization, if there is no selected discipline, the visualization returns to its initial state, as shown on Figure 4.

IV.2 Selection of Subjects

Regarding the selection of subjects, it is possible that the user wants to do a simultaneous comparison of distinct disciplines, or relationships for approvals and failures of the same discipline, and observe how passing or failing them can influence the academic career of the students. For this case, it is necessary to click on one of the triangles of two different subjects or on the same subject. The patterns of that subject become visible again, but in this case they are fixed and do not disappear with the scroll of the cursor. It is not possible to have more than two selections simultaneously, in order to prevent the generation of visual noise.

The colors used for these selections are the green and the blue, by being part of the most natural color space with better blending results [[13] and [14]]. Figure 6 illustrates the selection of the subjects *fex* and *soper*

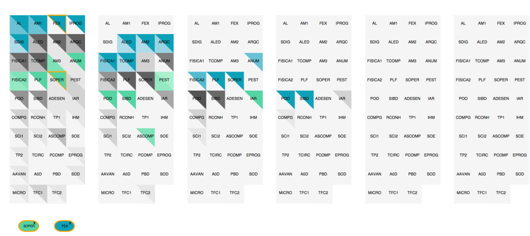

Figure 6: *Double selection*

Still concerning the selection of subjects, it is possible to apply certain restrictions to them. When the cursor passes over a subject, its patterns previously selected become visible in shades of lime, which are the common relationships between the initial selection and the the subject where the cursor currently is. This color was chosen because it is one of the intermediate colours between green and blue, and does not add visual noise, but at the same time it stands out from the other two tones, making more perceptible the result of the user interaction.

Finally, it may happen that two subjects from different patterns overlap. This happens when there are subjects in common from two patterns. In this case, the colors used initially, blue and green, are blended and the respective triangles are filled accordingly. In Figure 7 it is visible the result of the integration of both mechanisms, which corresponds to the final phase of EduVis.

Figure 7: *Vision of the overlapping cells in the patterns of the subjects sdig and aled*

V. Evaluation

A study was conducted with users to evaluate the solution created for the visualization of educational patterns. As a result, taking into account the context of the aforementioned study, it is intended to infer: (i) the effectiveness and efficiency of the developed solution, having registered for that time and number of errors associated with the performance of a set of tasks; (ii) the usability in general, and learning ability, in particular, that the contact with the created visualization provides; (iii) the degree of satisfaction of respondents regarding the performance of the tasks.

I. Representative Tasks of the System

Regarding the selected tasks, in order to test the developed visualization, a set of questions representative of the major components and features have been created: *): (1) How many semesters are represented? (2) Globally, what are the two subjects with more students? (3) What is the set of subjects involved in more patterns in the second semester? (4) What are the subjects related to iar in the fourth semester? (5) Concerning the students who did am1 and fisica1, what are the other subjects in which they also succeeded during the second semester? (6) What are the common subjects to who succeeded in sdig, in the first semester, and failed aled, in the second semester? (7) Whether the subject compg, in the fourth semester, has the lowest number of associated patterns, and is this the subject with less students on that semester?*

Questions 1, 2 and 3 correspond to general aspects immediately noticeable in the visualization: the number of semesters is visible in both views, represented by layers of circles in the upper and matrices in the lower.

In the first view it's possible to immediately check, through the size of the circles, the subjects with more students. Finally, on the lower representation it's represented the subjects involved in more patterns, through the prominence that is made with use of color.

The question 4 requires exploration of any of the visualization mechanisms by selecting the subject of interest in any of the tools. The point 5 requires the application of filters in the lower visualization, while the question 6 requires using filters in order to understand the blend of the colors.

To conclude, question 7 requires the use of both visualizations, the bottom to realize that in fact the subject has more patterns and the superior to realize the number of students of the mentioned subject.

II. Test Protocol

Before starting the tests, participants were approached individually and were taught about the context of the present study, being then, requested for their cooperation. After a verbal description and visualization of its fundamental components, the main features have been exemplified. Then, it was delivered to the participants a questionnaire with few questions and a list of other questions, corresponding to a set of tasks to perform. These questions were answered when interacting with the visualization via a laptop computer (common to all users). During this test phase, the time and number of errors for each task were measured. At the end, participants were encouraged to make oral comments about the system. Finally, they were asked to complete an online questionnaire of satisfaction, consisting of two parts: the first part matches the SUS (System Usability Scale) [17] and the second relates to a small set of questions that are intended to assess, also using a *Likert* scale of 5 points, the degree of difficulty felt by participants to perform each task in the understanding of the following aspects: (i) the number of semesters represented, (ii) the success and failure, (iii) the subjects involved in more patterns, (iv) the precedence relations between the subjects, and (v) the blend of patterns.

The study was conducted with 20 participants, of whom 15 (75%) are males and 5 (25%) are females. Of the total participants, 3 (15%) have between 18 and 24 years old, 10 (50%) were between 25 and 34 years old and 3 (15%) have between 35 and 44 years old, plus 2 (10%) have between 45 and 54, 1 (5%) were between 55 and 64 years old and 1 (5%) is situated in the upper age to 65 years old. As for education, 16 (80%) have a college degree, while 2 (10%) had completed secondary education and 2 (10%) had completed primary education studies.

Table 1: *Time and number of average errors*

Task	Time(seconds)	Number of errors
1	3.20	0.35
2	17.60	0.10
3	28.30	0.55
4	23.95	0.30
5	27.70	0.30
6	49.45	0.75
7	45.70	0.30

The average times and number of average errors relating to the performance of each of these tasks are summarized in Table 1. Although the tasks 1, 2 and 3 are immediate and therefore do not require additional exploration, the measured values do not allow to create a generalization about the time and number of errors for performance of these tasks with regard to the tasks of exploration (4 to 7), with possible exception of tasks 6 and 7. For deepening the results, a statistical analysis was conducted. It was applied a *Shapiro-Wilk* statistical test that showed evidence against a normal distribution in most data (p < 0.05) so it was applied a non-parametric test, the test *Wilcoxon signed-rank* for discovering the significant differences between samples. In terms of time, the task 1 is significantly faster than the others, $(z_{1-2}) = -3.72, z_{1-3} = -3.92, z_{1-4} = -3.92, z_{1-5} = -3.92, z_{1-6} = -3.92, z_{1-7} = -3.92, p < 0.05)$, While task 2 is significantly faster than 3 tasks, 5, 6 and 7, $(z_{2-3} = -2.39, z_{2-5} = -2.54, z_{2-6} = -3.32, z_{2-7} = -3.25, p < 0.05)$.

On the other hand, tasks 3, 4 and 5 are significantly faster than the tasks 6 and 7 $(z_{3-6} = -2.31, z_{3-7} = -2.17, z_{4-6} = -3.45, z_{4-7} = -2.76, z_{5-6} = -3.36, z_{5-7} = -2.63, p < 0.05)$. However, regarding the number of errors, there are significant differences between the task 6 and those with fewer errors, task 2 $(W = 6, cv = 17, p < 0.05)$ and task 4 $W = 0, cv = 3, p < 0.05)$, and between tasks 2 and 3, $(W = 5, cv = 8, p < 0.05)$.

In an attempt to find a correlation between the time and the number of errors, the *Pearson* coefficients have been estimated, finding a correlation, though relatively weak, in task 2 $(r = 0.45, p < 0.05)$ and in task 7 $(r = 0.49, p < 0.05)$, which does not allow to generalize a correlation between a greater time for task performance and a greater number of errors. It can be concluded that the temporal complexity of a task in the EduVis, does not lead to a greater number of errors in it's the performance.

Concerning the satisfaction questionnaire, the score on the SUS according to the corresponding parameters [16], was 79.47 points, showing high results with regard to usability and learning capacity of the system. Using the same method to calculate the answers of contextual satisfaction, was obtained a result of 92.11 points, showing that the proposed objectives were achieved.

III. Discussion

The results of the evaluation showed that users were able to obtain the necessary information to complete the tasks in a reasonable time and with low errors. For more complex tasks, the need for greater interaction with the system reflects a natural increase in the time of task performance which, however, this does not translate into an increase in the number of errors, which shows that the visualization allows to obtain the required information, necessary for the performance of tasks in an effective way. These results were corroborated by surveys of satisfaction, whicht proved the usability of the system and showed high satisfaction by users. The combination of these aspects shows the effectiveness of EduVis for visualizing educational patterns.

VI. Conclusion

A large amount of data emerges from the educational activities with the growing number of students in traditional and online education. Analyzed effectively, this information can help refine the processes of education. In this context, data mining techniques have shown to be relevant to find data patterns, but the result of the application of the same techniques often results in data sets difficult to read, interpret and analyze. Once overcome this limitation, it will be possible to represent this information as a whole and to have consistent notion of particular aspects in the data. Faced with this challenge, a visualization was created, that coordinates two different complementary interaction mechanisms to view relations between curricular units of a university course. The EduVis provides operating mechanisms, which allow simultaneous comparison between disciplines, and filtering, allowing the user to manage the information presented. It takes advantage of the color to highlight Visual elements in a way that emphasizes the relevant information and alleviates visual clutter that is typically present when a large number of articles with similar properties are represented.

Testing with users showed that the visualization turns immediately evident a relevant set of general information and allows easy data collection through interactive exploration and comparison mechanisms. The participants immediately realized aspects such as the number of semesters represented and the disciplines with more or less success. Using the mechanisms available to compare, filter and collect information the users showed satisfaction in the performance of tasks. It is possible to conclude that EduVis provides the means to represent the results of educational information in a way that provides the educational community to realize a set of standards that would not be obvious otherwise. By providing the means to diagnose certain problems, the system will be useful so that the coordinators of programs of study, such as teachers, have the means to find solutions to the existing limitations, promoting success in education.

References

[1] R. Mazza and V. Dimitrova, Generation of Graphical Representations of Student Tracking Data. In *Course Management System 9th International Conference on Information Visualization*, 2005

[2] Y. Gao, A study on Mass Higher Education in the world Based on Comparative Perspective. In *Education Institute of Xiamen University*, 2010

[3] Y. Guodong and W. Huan,Visual Exploration for Time Series Data Using Multivariate Analysis Method. In *Computer Science and Education (ICCSE)*,2013

[4] J. Heer and M. Bostock and V. Ogievetsky,A Tour Through the Visualization Zoo. In *Communications of the ACM*, 2010

[5] J. Kay and P. Reimann and E. Diebold and B. Kummerfeld,MOOCs: So Many Learners, So Much Potential. In *University of Sydney*,2013

[6] D. Keim,Information Visualization and Visual Data Mining, Visualization and Computer Graphics. In *IEEE Transactions on (Volume:8 , Issue: 1)*, 2002

[7] R. Mazza and L. Botturi, Monitoring an Online Course with the GISMO Tool: A Case

Study. In *Journal of Interactive Learning Research*,Volume 18, Issue 2,2007

[8] J. Steele and N. Iliinsky, *Designing data visualizations*,2011,O'Reilly Media,Gravenstein Highway North

[9] D. Trimm and P. Rheingans and M. DesJardins, Visualizing Student Histories Using Clustering and Composition. In *IEEE Transactions on Visualization and Computer Graphics*,2012

[10] E. Valiti, Avaliação de Usabilidade de Técnicas de Visualização de Informações Multidimensionais. In *Universidade Federal do Rio Grande do Sul,Instituto de informática, Programa de Pós-Graduação em Computação*, 2008

[11] C. Ware,*Information Visualization, Perception for Design, Second Edition*, 2004,Morgan Kaufmann Publishers

[12] D. Wortman and P. Rheingans, Visualizing Trends in Student Performance across Computer Science Courses Categories and Subject Descriptors. In *SIGCSE '07 Proceedings of the 38th SIGCSE technical symposium on Computer science education*, 2007

[13] S. Gama, D. Gonçalves, Studying Color Blending Perception for Data Visualization. In *Eurovis 2014 - Eurographics Conference on Visualization*, 2014.

[14] S. Gama, D. Gonçalves, Studying the perception of color components relative amounts in blended colors. In *NordiCHI 2014, 8th Nordic Conference on Human-Computer Interaction*, 2014

[15] R. Agrawal, R. Srikant, Mining sequential patterns. In *Proceedings of the IEEE International Conference on Data Engineering*, pag. 3-14, 1995.

[16] J. Sauro. *A Practical Guide to the System Usability Scale: Background, Benchmarks and Best Practices*. CreateSpace, 2011.

[17] J. Brooke. Sus: *A quick and dirty usability scale*, 1996.

[18] Romero, C., Ventura, S., Educational Data mining: A review of the state of the art. In *IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews, Vol 40, No.6, November 2010*