

Investigating the Suitability of Multi-dimensional Data Visualization as an Instrument for Assisting Distance Learning Instructors

Nathan Alves de L. Silva¹, Raphael A. Dourado², Rodrigo L. Rodrigues³, Aldo A. Cavalcanti¹

¹Computer Department – Federal Rural University of Pernambuco - Brazil

²Computer Center – Federal University of Pernambuco - Brazil

³Department of Education – Federal Rural University of Pernambuco - Brazil

nathan.alves.r2n@gmail.com, rasd2@cin.ufpe.br,
rodrigo.linsrodrigues@ufrpe.br, aldoacf@gmail.com

Abstract. *This paper investigates the suitability of using Data Visualization, specifically Parallel Coordinates Plots (PCPs), as an instrument to help distance learning instructors identify students with poor performance. To answer this question, we developed a web application that lets users generate and interact with PCPs and evaluated it through usability tests inspired by the NOVIS framework — a usability framework specifically designed for evaluating data visualizations. The results show that PCPs and Data Visualization are perceived by distance learning instructors as a useful tool for following students' progress on Virtual Learning Environments.*

1. Introduction

The advancement of technology changed the way we live and communicate in society, which reflects in the behavioral patterns of the most diverse social profiles. Sharing and searching for information became faster and simpler, thus favoring the growth of Distance Education. Distance Education proposes a more complete, dynamic and efficient teaching-learning process using technologies such as Virtual Learning Environments (VLEs). However, with the growth and adoption of Distance Education by many educational institutions, instructors are now facing the challenge of following the learning progress of a high number of students in a single class.

As a response to this challenge, an emerging field of research known as Learning Analytics is providing new techniques to collect and analyze data stored by VLEs in order to help instructors identify behavioral patterns in the learning process. Among the many techniques proposed by Learning Analytics researchers, Data Visualization offers the possibility to present data in such a way that a user can quickly find patterns and make better decisions. This can help instructors identify, for instance, students that are in a risky situation and then make pedagogical interventions [Vieira et al, 2018; Emmons et al, 2017].

In this context, our work proposes the development and evaluation of a multi-dimensional data visualization prototype — using Parallel Coordinates Plots (PCPs) — as an instrument to help teachers follow their students' online activity. Among the variety of multi-dimensional plots available in literature, we chose the PCP due to its simplicity and ability to show many dimensions simultaneously.

The remainder of this paper is organized as follows: Section 2 presents the theoretical foundations of Learning Analytics, Data Visualization, and the Parallel Coordinates Plot; Section 3 presents related work; Section 4 presents the methodology adopted in this study; Section 5 presents and discusses our experiment; finally, Section 6 draws some conclusions about the study.

2. Background

2.1 Learning Analytics and Data Visualization

According to [Conde and Hernández-García 2015], to optimize learning it is necessary not only to know how the learning process works and how to extract information about students' activities in VLEs, but also how to transform all this data into something useful and understandable, so the instructor can take appropriate action.

One emergent research area targeting this issue is Learning Analytics. This area proposes methods for collecting and analyzing data that can be used to discover relevant information about students and their context, to understand and optimize the teaching-learning process [Jivet et al. 2018]. Learning Analytics is a multidisciplinary area that uses several methods such as data visualization, social network analysis, and semantic and predictive models to extract knowledge from datasets. These techniques can help to identify students' behavioral patterns, improve their performance, and reduce retention [Nunn et al. 2016].

In distance learning courses, it is common for instructors to have to deal with a huge number of students, which make it difficult to keep track of the students' learning process in the virtual learning environment. In this context, Learning Analytics techniques can help instructors easily follow each student progress, identify learning problems, and take appropriate action [Augusto et al. 2015]. According to [Nunn et al. 2016] and [Jivet et al. 2018], the use of Learning Analytics in educational scenarios could bring benefits for teachers and students and help to reformulate the way learning is conducted. As benefits we can list:

- Identify students' learning gaps in relation to what is required in the course;
- Understand in a clear and objective way the students' development process to improve the relationship between student-instructor and student-student;
- Help students in keeping on track with the course demands, by monitoring their progress;
- Help instructors improve their practice through feedback on students' learning progress, learning materials, etc.;
- Stimulate the exchange of information between instructors, thus strengthening the educational community.

One of the most effective ways of delivering educational data to students and instructors is through interactive data visualization interfaces [Vieira et al, 2018]. Data Visualization can be defined as a way of representing abstract data in visual forms, which gives users a clearer comprehension of the phenomena being represented and the ability to dynamically interact with the data to discover relevant patterns and knowledge — something that is not easy while dealing with tabular or textual data, for instance [Gama et al. 2014] [Webber et al. 2013].

In the context of distance learning, Data Visualization can provide instruments to effectively analyze data such as forum participation, completion of required activities, and the number of times a student access the platform. Thus, with visualization interfaces instructors can visualize their students' engagement, motivation, and performance, which can be used as feedback for adjusting the course instructional design [Silva et al. 2016].

Among the wide array of visualization techniques available in literature, a particular interesting category is multidimensional data visualization [Gintautas et al. 2013]. These visualizations aim to map multi-dimensional data into a bi-dimensional plane, making it easier visualize patterns. The next subsection describes one approach to visualize multidimensional data: the Parallel Coordinates Plot.

2.2 Parallel Coordinates Plot

According to [Heinrich e Weiskopf 2013], Parallel Coordinates Plot (PCP) is a visualization technique commonly used to represent multidimensional data. The PCP is built by the composition of three domains: the *spatio-temporal domain*, which is responsible for the events projection inside of the chart; the *data domain*, which is responsible for the arrangement of data on the chart; and the *parallel-coordinate domain*, which is responsible for the construction of the chart on cartesian coordinates, using the x and y axes of the spatio-temporal domain. Figure 1 presents a generic representation of the parallel coordinates plot.

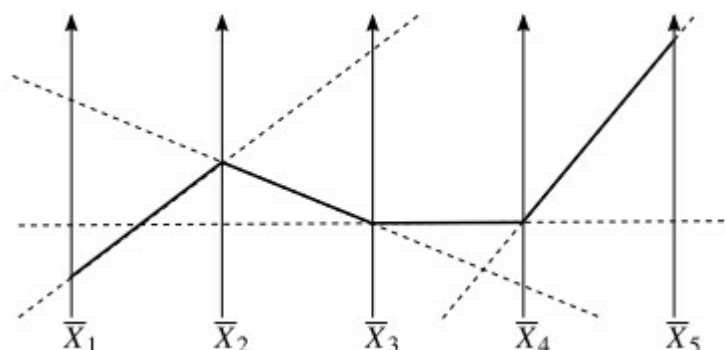


Figure 1. Representation of the construction of the Parallel Coordinates Plot

According to [Martins 2012], the construction of a PCP is achieved by mapping the n-dimensional space to a two-dimensional structure, which uses parallel axes that are vertically equidistant, as shown the Figure 1. The axes X1 to X5 represent the data dimensions and the sharp lines represents the connection between data and the other dimensions. A user can interact with the PCP by: inclusion and removal of axes (coordinates), selection of an interval on one of the axes to highlight part of the data (also known as *brushing*), and multiple brushing —the selection of intervals in two or more axes simultaneously.

Regarding the data mapping process, [Heinrich and Weiskopf 2013] present the concept of *data interpolation*, according to which it is possible to map points of lines and curves of a cartesian coordinate to points distributed on the axes of the PCP, preserving the data patterns.

In the next section, we discuss related works where the authors used Learning Analytics, Data Visualization, and the Parallel Coordinates Plot to represent multidimensional data about students' academic activities on VLEs.

3. Related Work

In the work of [Siirtola et al. 2009], the authors conducted a study to verify the visual perception of the PCP with users that never used this kind of visualization. To collect the experimental data, the authors used visual tracking, observing the areas in which the users most looked during the task execution and if this pattern was correlated with the task's expected results. After analyzed the data, the authors concluded that the middle part of the PCP, where the interaction (brushing) happens, requires more attention from the user than the top and bottom areas. The data related with fixation time and time spent on areas of interest of each task showed that the PCP has a quick learning curve and that is possible to make simple searches right after the first interaction with the plot.

In the same perspective, the study developed by [Tervakari et al. 2014] investigated the usefulness of Learning Analytics tools developed by the Tampere University of Technology on social medias inside the "TUT Circle" VLE. 27 teachers and academic managers participated in the study and were submitted to two phases of data collection: a focus group experiment and a usability test with a Learning Analytics dashboard, which included, among other charts, the parallel coordinates plot. The results showed that teachers considered the use of Learning Analytics more useful than traditional reports for keeping track of the students' progress and as a way of scheduling interventions when some student is in risk of failing.

Another work related to our proposal in this paper is the study developed by [Govaerts et al. 2010], where the authors tried to solve issues related to the collection of data for analyzing class' performance and providing self-reflection with Learning Analytics and Data Visualization techniques. The authors developed a tool that tracks the academic activities developed by each student in real time, including statistics about the time spent with a determined resource and the issues related to it. To present the data to users, the authors used Line charts, PCPs, and Bar charts. 12 computer science students participated in the experiment and expressed their experience with the proposed tool through the Think-aloud method. The results showed a 73% rate of user satisfaction.

Lastly, in [Lee et al. 2016] the authors performed an investigative study trying to understand people's reaction and interaction with certain types of data visualizations for the first time. The authors investigated the following characteristics about visualizations: *chart format, visual characteristics, knowledge about graphs, and knowledge about the context of the data represented on the chart*. In this study, the participants were submitted to an experimental session where they interacted with three types of charts: PCP, Chord Diagram and Treemap. At the end of the interview, they also participated in a semi-structured interview. The results showed that: i) giving extra information about the chart before the user interacts with it affects the way he/she builds their mental model; ii) the participants had issues to continue the task when the mental model was incorrectly built; and iii) the data context directly affects the interest of the participants in exploring the chart.

Different from the works discussed in this section, our work focuses on analyzing distance learning instructor's visual perception of PCPs, and to which extent a Visual Analytics tool based on PCPs can help them improve their practice.

4. Methodology

In this study, we used as a starting point the NOVIS evaluation method, proposed by [Lee et al. 2016]. This model focus on evaluating how the user interacts with textual objects (alphanumeric format), non-textual objects (the axes of a PCP), and how they build a mental model to interpret the data represented on the plot. We conducted an experiment with professors and tutors that work or have worked with distance education to analyze their ability to recognize patterns and extract information from PCP, using a set of proposed tasks and a semi-structured interview.

4.1 Prototype development

To perform the proposed experiment, we first built a low fidelity prototype (Figure 2 (a)), and then developed a high-fidelity prototype using HTML5, CSS3, JavaScript, and the D3.js library to provide a full interactive PCP.

To use the prototype, the user first chooses the course, semester, period, discipline, construct, and the variables related with the chosen construct that the participant wants to analyze with the PCP, as shown on Figure 2 (b).

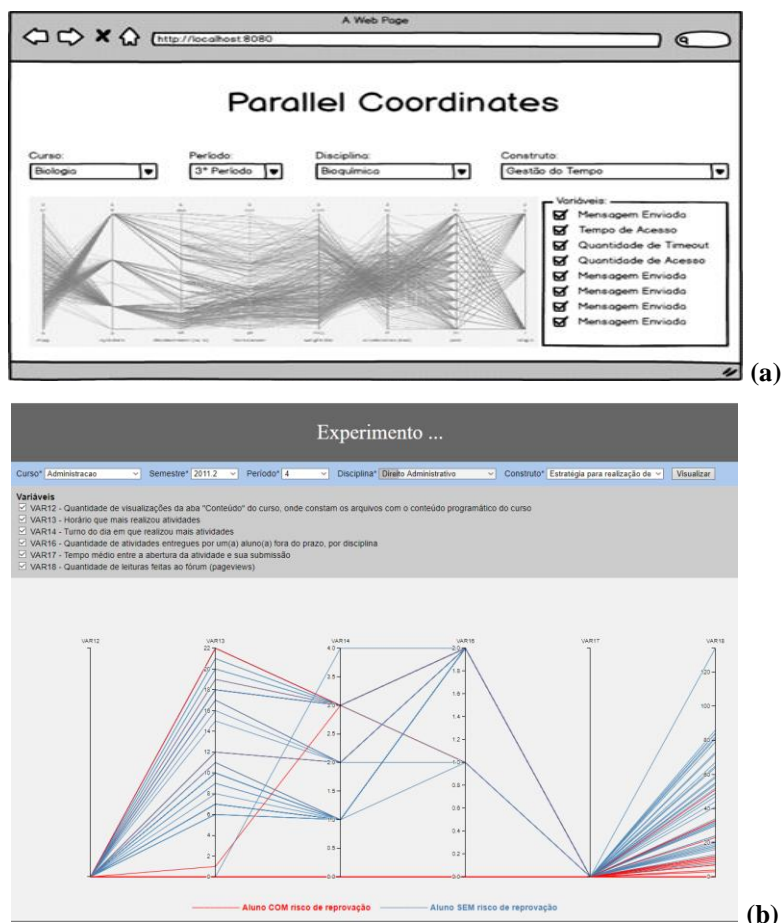


Figure 2. (a) Representation of the low fidelity prototype. (b) High fidelity prototype.

The dataset available on the prototype for plotting PCPs corresponds to data from distance learning undergraduate students' interactions with a VLE, comprising a period of 7 years, provided by the NEAD/UPE from University of Pernambuco.

4.2 Target Audience

Five distance learning instructors participated in this experiment (among professors and tutors, 4 men and 1 woman) between 28 to 47 years old, that are currently working on Distance Education or had earlier experience with it. The data used in the experiment are related to student activity within VLEs, a common domain to all participants.

4.3 Activity Execution

At the beginning of the experiment, participants were introduced to the context of the data, the concept of Learning Analytics, and the rationale of a PCP. This step was fundamental to provide all participants with the necessary background to use the prototype. After that, we introduced the Think-aloud method described by [Preece et al. 2005], following a demonstration on how to do it.

The experiment began with the following question "Please, describe your thoughts and behaviors while you try to understand the chart". To evaluate the participants' experience while using the prototype, we proposed the four tasks shown on the Table 1.

Nº	Description
1	Task I
	a) Select the course Pedagogy , Semester 2014.2 , in the 1st period , and the discipline Philosophical Foundations of Education . Select the construct Setting goals . b) Identify the students who are at risk of failing and interpret their behavior in relation to the variables. c) Change the discipline to Language and Production of Knowledge and reinterpret the student's behavior.
2	Task II
	a) Select the course Pedagogy , Semester 2014.2 , in the 1st period , and the discipline Anthropology . Select the Search for Help construct. b) Select the range from 0 to 7 for the variable VAR02 . Move the range bar slowly from bottom to top; c) Make a free interpretation of what you understood about the presented results.
3	Task III
	a) Select the course Letters , Semester 2014.2 , in the 5th period , and the discipline Brazilian Literature II . Select the Task Achievement strategy construct. b) Select the range from 18 to 23 for the variable VAR13 . Move the range bar slowly from top to bottom; c) Make a free interpretation of what you understood about the presented results.
4	Task IV
	a) Select the course Letters , Semester 2014.2 , in the 5th period , and the discipline Brazilian Literature II . Select the Search for Help construct. b) Select the interval of 7 to 10 for the variable VAR03 . Move the range bar slowly from top to bottom;

c) Make a free interpretation of what you understood about the presented results.

Table 1. Task list for the experiment

By performing the tasks described in Table 1, we could detect the participants' first impressions about the PCP and about their mental model construction while using its functionalities.

4.4 Semi-structured interview

After the experiment, we conducted a semi-structured interview with the participants, asking the following questions:

- *While lecturing or tutoring a given course, do you give feedback to the students about their performance? If so, which method do you use the most?*
- *Regarding the data collected by VLEs, do you think it is important to use it to improve course instructional design? What about using this data to improve student's educational experience?*
- *Should teachers, in their professional practice, have an obligation to intervene if a student shows unsatisfactory performance?*
- *In your opinion, which method is best for assisting you in following students' progress: the traditional VLE reports or data visualization?*
- *In your opinion, should teachers receive training in analyzing educational data as a resource to give feedback on student performance?*

The next section discusses the results obtained from the analysis of the experiment data.

5. Results and Discussion

5.1 PCP Experiment

To analyze the experiment results, we selected a set of metrics to understand the participants experience with the PCP and how they organized their mental models. Table 2 shows the results for each of the following metrics:

- *NC: Total number of clicks to complete the task;*
- *NH: Total number of times that the participant needed help or intervention;*
- *T1 to T5: Time in minutes that the participant spent on each task;*
- *RMM: Number of times that the participant reformulated the mental model;*
- *RDC: Number of times that the participant reported difficulty with the PCP;*

Participant	NC	NH	T1	T2	T3	T4	RMM	RDC
P1	80	7	07.37	02.94	05.81	04.20	6	6
P2	62	8	06.36	06.34	02.37	02.51	5	6
P3	50	3	04.59	05.26	03.13	01.30	3	1
P4	86	10	08.24	11.59	04.07	04.09	5	1
P5	80	16	15.43	12.42	12.34	11.56	11	2

Table 2. Score of each participant on the experiment with the PCP

Based on these variables, one can see that although the P4 and P5 needed more help or attention from the mediator, in general all the participants did not need so much help to understand the task, interact with the PCP, and discover the patterns on the chart

to complete the four tasks. One common issue related to the **NH** metric was that, although all participants had received an explanation about how to interact with the chart, what Learning Analytics is, and about the variables related to each construct, they showed some difficulty, in the first trial, in understanding the relationship between all the axes in the plot because the variables had different scales.

The **NC** metric reveals that the participants had a high number of interactions with the PCP, which lead us to think that when the participants know the context of the data, they are more motivated to explore and discover relevant information. The variables **T1** to **T5** show that despite the amount of information provided by a PCP, the participants did not spent much time on each task, which indicates that the use of Learning Analytics together with Data Visualization can offer the possibility to extract relevant data in a few minutes after exploring the PCP.

The metrics **RMM** and **RDC** show that the number of times a participant reformulated his/her mental model to complete and understand the tasks does not necessarily say that he/she has difficulty to understand the chart, but rather that he/she wants to find the right answer for the task and for what he/she is seeing on the chart. Moreover, the fact that only the participants **P1** and **P2** demonstrated higher difficulty with the PCP indicates that this kind of visualization could be a useful tool with which instructors can take their own conclusions about the class' performance and their pedagogical practices.

Another thing should be highlighted is that the higher scores showed by **P5** on the metrics **NH**, **T1**, **T2**, **T3**, **T4** and **RMM** can be a result of his personal interest in visualizations (as he declared during the experiment) — something that can be explored in a future work.

5.2 Semi-structured interview

In this section, we will analyze the answers for the semi-structured interview in a question by question basis.

Q1. While lecturing of tutoring a given course, do you give feedback to the students about their performance? If so, which method do you use the most?

All the answers show that the participants care about giving feedback to students, and the frequency depends on the institution rules. The feedbacks are usually given when the students hand out an activity, participate in the forum and when they spend a long time without accessing the VLE. The method is always text messages giving the students accounts on their performance as a form of motivation or alert of poor performance. It is more usual for tutors rather than professors to give personalized feedback to students.

Q2. Regarding the data collected by VLEs, do you think it is important to use it to improve course instructional design? What about using this data to improve the student's educational experience?

All the participants reported that the collected data can be useful and bring improvements to the class. Only **P1** said that never used the kind of the data used in the experiment to evaluate the students' academic performance. **P2** and **P5** said that the data needs to be well interpreted to fulfill the goal of providing teachers a real account of the class performance as soon as possible. **P3** and **P4** said that a Learning Analytics/Data

Visualization tool can be useful to interpret the data quicker and to know precisely each student's problem and fix it.

Q3. Should teachers, in their professional practice, have an obligation to intervene if a student shows unsatisfactory performance?

All the participants agreed that instructors have an obligation to intervene when the student is in a risky situation. According to **P1**, external factors are the main cause of failure and evasion. **P3** and **P4** also agreed with the instructor obligation to give feedback and added that when the student feels “alone” in the course, he/she is at greater risk of poor engagement. **P5** said that, speaking as a teacher it all depends on the “feeling” that the teacher has about the class, but as an “academic professional”, from a pedagogical point of view, the instructor has always the obligation to intervene.

Q4. In your opinion, which method is best for assisting you in following the students' progress: the traditional VLE reports or data visualization?

All the participants said that Learning Analytics together with Data Visualization makes it easier and effective to follow the students' progress compared to the traditional mode. Only **P1** said that had a little difficulty to understand all the information on the PCP and from a tutor point of view, only a handful of variables are sufficient to convey the most important information.

Q5. In your opinion, should teachers receive training in analyzing educational data as a resource to give feedback on student performance?

Only **P3** said that depends on the background that the user has with technology; in his opinion, a short introduction about the fundamentals of Learning Analytics, Data Visualization and PCP are sufficient to use de chart. **P4** and **P5** said that the training is important because there are a lot of people unfamiliar with plots like PCP.

6. Conclusion

In this work, we tested the suitability of using a multi-dimensional data visualization as instrument to help instructors in following their students' progress. The results suggest that Learning Analytics and Data Visualization empowers distance learning instructors, by giving them the ability to understand VLEs interaction data, gauge students' performance, give informed feedback, and act on time, before the student fails.

Despite the small number of participants in the study, the results showed that the use of PCP can help instructors to identify behavioral patterns quicker than traditional methods. As future work, we plan to compare the effectiveness of the PCP with other multi-dimensional plots to find which one works best with distance learning instructors.

References

- Augusto, M., Schmitt, R., Margaria, L., et al. (2015). Uma Ferramenta de Learning Analytics para o Moodle. *Anais do XX Congresso Internacional de Informática Educativa (TISE)*, v. 11, p. 821–824.
- Conde, M. Á. and Hernández-García, Á. (2015). Learning analytics for educational decision making. *Computers in Human Behavior*, v. 47, p. 1–3.
- Gama, S., Jordão, V. and Gonçalves, D. (2014). EduVis: Visualização Interativa de Dados Educacionais. p. 25–32.

- Gintautas, D., Olga, K. and Žilinskas, J. (2013). *Multidimensional Data Visualization. Methods and Applications.*
- Govaerts, S., Verbert, K., Klerkx, J. and Duval, E. (2010). Visualizing Activities for Self-reflection and Awareness. *Advances in Web-Based Learning – ICWL 2010*, v. 6483, n. June 2014, p. 91–100.
- Heinrich, J. and Weiskopf, D. (2013). State of the Art of Parallel Coordinates. *Eurochart Conference on Visualization (EuroVis)*, p. 95–116.
- Jivet, I., Scheffel, M., Specht, M. and Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. v. 18, n. 10.
- Lee, S., Kim, S. H., Hung, Y. H., et al. (2016). How do People Make Sense of Unfamiliar Visualizations?: A Grounded Model of Novice’s Information Visualization Sensemaking. *IEEE Transactions on Visualization and Computer Chart*, v. 22, n. 1, p. 499–508.
- Martins, R. (2012). Revista do programa de mestrado em cultura visual.
- Nunn, S., Avella, J. T., Kanai, T. and Kebritchi, M. (2016). Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review. *Online Learning*, v. 20, n. 2.
- Siirtola, H., Laivo, T., Heimonen, T. and Rähkä, K. J. (2009). Visual perception of parallel coordinate visualizations. *Proceedings of the International Conference on Information Visualisation*, n. 978, p. 3–9.
- Silva, J., Ramos, J. L. C., Rodrigues, R., et al. (2016). Uma abordagem para integração do Moodle com o framework Shiny para Learning Analytics. p. 930.
- Tervakari, A., Kuosa, K., Koro, J., Paukkeri, J. and Kailanto, M. (2014). Teachers’ learning analytics tools in a social media enhanced learning environment. *2014 International Conference on Interactive Collaborative Learning (ICL)*, n. December, p. 355–360.
- Webber, C. G., Cini, G. and Lima, M. de F. W. do P. (2013). Facilitando a Análise se Dados Educacionais através se Ferramentas se Visualização. *Revista Novas Tecnologias na Educação (RENOTE)*, v. 11, n. 3, p. 1–10.
- Preece, J., Rogers, Y. and Sharp, H. (2005). Interaction Design: Beyond Human Computer Interaction. *Leonardo*, v. 38, n. 5, p. 404.
- Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*.
- Emmons, S. R., Light, R. P., & Börner, K. (2017). MOOC visual analytics: Empowering students, teachers, researchers, and platform developers of massively open online courses. *Journal of the Association for Information Science and Technology*, 68(10), 2350-2363.