



Visualizing Learning Analytics and Educational Data Mining Outputs

Ranilson Paiva^(✉), Ig Ibert Bittencourt, Wansel Lemos, André Vinicius,
and Diego Dermeval

Federal University of Alagoas (UFAL) - Computing Institute (IC), Maceió, Brazil
ranilsonpaiva@ic.ufal.br, ig.ibert@gmail.com,
wansel.lemos@arapiraca.ufal.br, andreducap@gmail.com,
diegodermeval@gmail.com

Abstract. The increase in supply and demand of on-line courses evidences a new educational paradigm mediated by information and communication technologies. However, an issue in this new paradigm is the high number of students who drop out (85% on average). Some of them blame the lack of instructor support. This support needs the analysis of students' data to guide teachers' decision-making. Learning Analytics (LA), Educational Data Mining (EDM) and Data Visualization (DataViz) are some tools for this analysis, but teachers do not receive appropriate technological support to use them. So, we used DataViz to help teachers understand the output from the application of LA and EDM algorithms on the students' data. We evaluated if instructors understood the information in the visualizations, and asked their opinion about the visualizations' (1) utility; (2) ease of use; (3) attitude towards use; (4) intention to use; (5) aesthetics; (6) the color scheme used; and (7) the vocabulary used. The results indicate that instructors understood the information in the visualizations and the majority of them had favorable opinions, but we noticed the vocabulary used needs improvement.

Keywords: Data visualization · Educational data science · e-Learning

1 Introduction

The increase in supply and demand of on-line courses [2,6] evidences a new educational paradigm, which relies on digital information and communication technologies (DICT) [4]. However, this new paradigm poses some issues for teachers. One issue is the high number of dropouts (85%, on average) [8,11]. Learners blame the “Lack of Instructor Support” [11], but such support demands educational data analysis to guide educational decision-making [3,7,13]. Learning Analytics (LA), pedagogical Data Mining (EDM) and Data Visualization (DataViz) are a set of tools to do that, but teachers are not, normally, trained nor receive appropriate technological support to use them [10,13]. Thus, the need to assist

teachers using technology to guide pedagogical decision-making is latent. This aid should process learners' educational data in search for relevant information, showing the characteristics of the issues, guiding teachers on what they should do [3, 9, 12, 13]. For that, we created 3 visualizations to: (1) measure the amount of interactions, from a group of students, with each educational resource (called segmented bar chart and coded as Viz1); (2) show the most impactful interactions on students' performance (ordered weights, Viz2); and (3) show the most impactful combination of interactions on students' performance (combined interactions, Viz3).

2 Proposal

We used data visualization to help teachers understand the output from the application of data mining and learning analytics on educational data from 196 students (an on-line high-school math course), consisting of the amount of: (1) problems solved correctly, incorrectly and in total; (2) accesses to the learning environment; (3) videos watched; (4) points earned (gamification); (5) badges/trophies achieved (gamification); and (6) level (gamification). For that, we created 3 visualizations associated with the "RAG Colors" technique [1], to analyze students as groups, based on their performance¹. The visualizations are explained below:

Visualization 1 - Segmented Bar Graph. In this visualization, the interactions are counted and compared to the mean of all interactions of the same kind. Learners with scores below -1 standard deviation, were in the inadequate class; those with scores between -1 and +1 standard deviation, were in the insufficient class; and those with scores above +1 standard deviation, were in the adequate class. The aim was to isolate the interactions and facilitate comparison (Fig. 1 - Top).

Visualization 2 - Ordered Weights. In this visualization, we ran the *SimpleLogistic*² algorithm on the data to build a linear regression model [14]. The output is not "teacher-friendly". Thus, we transformed the textual output, considering the weights of each variable and the 3 classes of results: 0 = inadequate, 1 = insufficient and 2 = adequate. Variables with negative weights repel learners from the class. We ordered interactions that repelled students from the inadequate class (class 0) and attracted the adequate class (class 2), see (Fig. 1 - Middle).

Visualization 3 - Combined Interactions. In this visualization, we ran the *JRip* algorithm to infer association rules [5] based on frequent and relevant patterns in the data. The output shows some combinations of interactions leading to

¹ RAG stands for: red, amber, green. Red = Inadequate class: learners need urgent attention and well-planned pedagogical interventions [1]; Yellow/Amber = Insufficient class: learners need attention, monitoring and guidance [1] to progress; Green = Adequate Class: learners need incentives and/or challenges to keep them motivated and progressing well.

² Available at: <https://www.cs.waikato.ac.nz/ml/weka/>.

a particular class of results. Teachers can identify sequences of interactions that affect learning, which is potentially informative for the teachers. We calculated the “importance score”, adding a point for the occurrence of a resource in the rules returned and subtracted a point for each non-occurrence. The result was the combination of the four resources with highest (green) and four resources with the lowest (red) scores (Fig. 1 - Bottom).

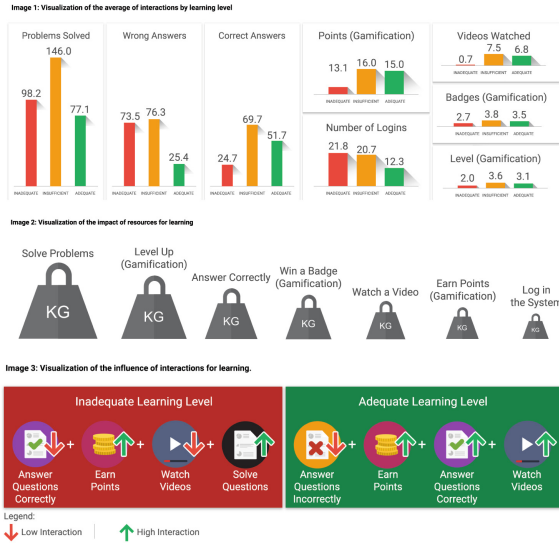


Fig. 1. The 3 Visualizations Created: Segmented Bar Graph (top), Ordered Weights (middle) and Combined Interactions (bottom). (Color figure online)

3 Design of the Experiment

The experiment was operationalized as an on-line questionnaire. We invited instructors (professors, teachers and tutors) to evaluate the visualizations, answering some questions to check if they understood the information displayed. We also asked them their perceptions on the visualizations, considering the: (1) perceived utility - PU³; (2) perceived ease of use - PEU⁴; (3) attitude towards use - ATU⁵; (4) intention to use - IU⁶; (5) perception about the aesthetics - AES⁷; (6) perception about the color scheme used (RAG Colours) - RC⁸; (7)

³ If the participants considered the visualizations would be useful for helping them with their professional activities.
⁴ If the participants considered the visualizations easy to use.
⁵ If participants showed a positive attitude regarding the use of the visualizations.
⁶ If participants would use the visualizations if they were available for them in their workplace.
⁷ If the visualizations were beautiful and attractive.
⁸ If the colors (red, yellow and green) helped them understand the results in the visualization.

perception about the terms used (inadequate, insufficient, adequate) to classify students' results - TU⁹, all following a Likert scale from 0 to 6¹⁰.

4 Results and Discussion

The questionnaire was available for one month and we had 116 valid records. First, we evaluated the answers about the visualizations. We called the metric Understandability and the results showed high values for all visualizations, indicating teachers understood the information they provided. After that, we compared the visualizations among themselves, testing for statistically significant differences regarding the understandability (Table 1).

Table 1. Comparison between visualizations.

VIZ	WILCOXON TEST	BONFERRONI	BEST
Viz1 vs. Viz2	0.0001185024	0.0007110142*	Viz1
Viz1 vs. Viz3	7.216746e-06	4.330048e-05*	Viz1
Viz2 vs. Viz3	0.01171951	0.07031707	No difference

As displayed in Table 1 Viz1 provided greater understandability to teachers. The order was: Viz1 > Viz2 = Viz3. One explanation is that Viz1 resembles a bar graph, which is a traditional kind of graph so it was more familiar to the participants.

The median result of the participants' perceptions, for all metrics, was around 4, meaning the participants "slightly agree" that the visualizations were easy to use (ease of use), interesting (attitude towards use), they would use them if they were available (intention to use), beautiful/attractive (aesthetics) and the color scheme was appropriate (color scheme used). Regarding the perceived utility, participants "neither agree nor disagree" the visualizations would increase their productivity. Regarding the vocabulary, the participants "neither agree nor disagree" the vocabulary was appropriate (vocabulary used), signaling a need for improve these last two metrics.

5 Conclusion

We created 3 visualizations to help teachers understand the output from data analysis techniques, using the RAG Colors technique to group learners according to their class of results. We asked highly competent and experienced instructors to evaluate them. The participants, overall, perceived the visualizations as easy to use, interesting, attractive and that they would use it, if they were available. For the perceived utility and the vocabulary used, the results show that these metrics need improvement.

⁹ If the terms used, helped them understand the results in the visualization.

¹⁰ 0 = I Strongly Disagree; 3 = I Neither Agree nor Disagree; 6 = I Strongly Agree.

The visualizations were effective (about 84% of all answers were correct) in making teachers understand the information extracted from the outputs of educational data mining and analytics (understandability), suggesting the visualizations are an objective and simple way for teachers to interpret what is going on with their groups. This is important to assist teachers' daily decision-making tasks, making it evidence-based.

Some topics that need further research: (1) how can we improve the visualizations' utility? (2) what kind of vocabulary is appropriate to be used? (3) are there algorithms that are easier to visualize than others? (4) what are the other algorithms we can visualize? (5) how can we visualize different information from a single educational data mining/analytics' output?

Acknowledgments. We would like to acknowledge CNPQ (Brazilian National Council for Scientific and Technological Development) and PROCAD (Brazilian National Programme for Academic Cooperation) for supporting this work.

References

1. Alexander, M., Yuk, M., Diamond, S.: *Data Visualization for Dummies*. Wiley, Chichester (2014)
2. Allen, I., Seaman, J., Poulin, R., Straut, T.: *Online Report Card: Tracking Online Education in the United States*. Babson Survey Research Group and Quahog Research Group, LLC, Babson Park, MA (2016)
3. Bienkowski, M., Feng, M., Means, B.: *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*, pp. 1–57. US Department of Education, Office of Educational Technology pp (2012)
4. Bittencourt, I.I., Costa, E., Silva, M., Soares, E.: A computational model for developing semantic web-based educational systems. *Knowl. Based Syst.* **22**(4), 302–315 (2009)
5. Cohen, W.W.: Fast effective rule induction. In: *Proceedings of the Twelfth International Conference on Machine Learning*, pp. 115–123 (1995)
6. de Educação à Distância, A.B.: *Censo ead br (2015). Relatório Analítico da Aprendizagem a Distância no Brasil* (2016)
7. Kowalski, T., Lasley, T.J.: *Handbook of Data-Based Decision Making in Education*. Routledge, New York (2010)
8. Liyanagunawardena, T.R., Parslow, P., Williams, S.: *Dropout: Mooc participants' perspective* (2014)
9. Mandinach, E.B., Honey, M., Light, D.: A theoretical framework for data-driven decision making. In: *Annual Meeting of the American Educational Research Association, San Francisco, CA* (2006)
10. Mandinach, E.B., Jackson, S.S.: *Transforming Teaching and Learning Through Data-Driven Decision Making*. Corwin Press, Thousand Oaks (2012)
11. Onah, D.F., Sinclair, J., Boyatt, R.: Dropout rates of massive open online courses: behavioural patterns. *EDULEARN14 Proceedings*, pp. 5825–5834 (2014)
12. Romero, C., Ventura, S.: *Educational data science in massive open online courses. Data Mining and Knowledge Discovery, Wiley Interdisciplinary Reviews* (2016)

13. Schildkamp, K., Lai, M.K., Earl, L.: Data-based decision making in education: Challenges and opportunities, vol. 17. Springer Science & Business Media (2012)
14. Sumner, M., Frank, E., Hall, M.: Speeding up logistic model tree induction. In: Jorge, A.M., Torgo, L., Brazdil, P., Camacho, R., Gama, J. (eds.) PKDD 2005. LNCS (LNAI), vol. 3721, pp. 675–683. Springer, Heidelberg (2005). https://doi.org/10.1007/11564126_72