PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

André Luiz de Brandão Damasceno

Supporting Instructors in Analyzing Student Logs from Virtual Learning Environments

Tese de Doutorado

Thesis presented to the Programa de Pós–graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Ciências – Informática.

Advisor: Prof.ª Simone Diniz Junqueira Barbosa

Rio de Janeiro
September 2020

André Luiz de Brandão Damasceno

Supporting Instructors in Analyzing Student Logs from Virtual Learning Environments

Thesis presented to the Programa de Pós–graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Ciências – Informática. Approved by the Examination Committee.

Prof.ª Simone Diniz Junqueira Barbosa
Advisor
Departamento de Informática – PUC-Rio

Prof. Alberto Barbosa Raposo
Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

Prof. Hélio Côrtes Vieira Lopes
Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

Prof. Carlos de Salles Soares Neto
Universidade Federal do Maranhão – UFMA

Prof.ª Milene Selbach Silveira
Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS

Rio de Janeiro, September 1st, 2020

André Luiz de Brandão Damasceno

Bachelor's in Computer Science (2013) and Masters' in Computer Science (2015) at the Federal University of Maranhão (UFMA), with emphasis on Applied Computing. Worked for the Laboratory of Advanced Web Systems (LAWS) at UFMA from 2011 to 2015 and Telemidia at PUC-Rio from 2016 to 2018. Since 2018 works at IDEIAS and DaSLAB, at PUC-Rio.

To everyone who encourages me in this journey: my family, professors, and dear friends.

Acknowledgments

First, I would like to thank my family: my parents Cláudia and Ronald (*in memoriam*) that went out of their way to give a good education, my grandfather Alcir (*in memoriam*) for share great values that shaped myself, my godmother Lurdinha and Almerice by their love and words of encouragement, my sisters Rita and Raquel that also contributed to what I am today.

To Júlia, my fiancée, I want to say a huge thank you for all the support, love, and care that she gave to me throughout these years. She also supported my decisions and helped me with my emotional issues. I will never forget our life in Rio de Janeiro and I hope to be with her for the rest of my life.

I am eternally grateful to my advisor Simone Barbosa for her advice and guidance. She always encouraged me in my research and helped me in many ways. I learned so much with her attitude and example, which contributed to my education as a researcher.

To Hélio Lopes for a great lecture about Data Science, which convinced me to study this area, and contributed with the topic approached in this thesis. I appreciate his good humor and enthusiasm with research, as well as his caring about the well-being of everyone.

I would like to thank everyone from TeleMídia Lab: Sergio Colcher, Álvaro da Veiga, Álan Lívio, Guilherme Lima, Roberto Gerson, Rodrigo Costa, Marcos Roriz (a honorary member), Antonio Busson, Pedro Almeida, Paulo Mendes, João, Jorge, Lucas, and others. I will never forget our discussions and good moments with coffee, chocolates, and cakes.

I would like to thank a lot of people that made my Ph.D. journey enjoyable and happy. My roommates Álan, Ruberth, Derlyane, and Thais. My friends from departament of informatics Alysson, Anderson Uchôa, André Davys, André Moreira, Bianca, Dalai, Daniel, Dieinison, Gabriel Homsi, João Vitor, Lauro, Lucas, Luiz Eduardo, Micaele, Paulo Henrique, Pedro Torres, Raul, Renatha, Rodrigo Almeida, Rômulo, Sérgio, Suellen, Vinícius and Wallas. My friends from other departaments Isabella Frajhof, Lisseth Saavedra and Patrick Saar. To all others not mentioned here.

I cannot forget to cite two professors that influenced me to pursue the doctorate: Alexandre César and Carlos Salles. Alexandre who taught me what research is during my period in PET Comp. Carlos Salles who believed in my work and invited me to conduct the Cacuriá Project (GT-VoA), which yielded my master's degree, my first academic paper, a company (Mediabox) and great learning. I will never be able to fully repay you, all that I can give you is my immense gratitude and admiration.

Abstract

Damasceno, André Luiz de Brandão; Diniz Junqueira Barbosa, Simone (Advisor). **Supporting Instructors in Analyzing Student Logs from Virtual Learning Environments**. Rio de Janeiro, 2020. 152p. Tese de doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Online education has broadened the avenues of research on student's behavior and performance. In this thesis, we shed light on how to support instructors in analyzing student logs from Virtual Learning Environments. Firstly, we conducted interviews with instructors and a systematic mapping of the state-of-art about Education Data Mining and Learning Analytics. Then, we analyzed logs from online courses offered in Brazil and compared our findings with results presented in the literature. Moreover, we gathered instructors' preferences in regard to visualization of both students' behavior and performance. However, we noted a lack of work showing models to support the development of learning analytics tools. In order to bridge this gap, this thesis presents a model connecting both Visual Analytics theories and models as well as instructors' requirements, their visualization preferences, literature guidelines and methods for analyzing student logs. We instantiated and evaluated this model in a tool to assemble dashboards. We captured evidence of their acceptance of our proposal and obtained instructors' feedback about the tool such as their both analysis and visualization preferences. Finally, we present some considerations and discuss gaps in existing works that can ground and guide future research, such as new instances of our model, as well as deploying them at Brazilian institutions and evaluating whether there are changes in students' performance when instructors are able to see information about their behavior and performance, and act accordingly. It is worth highlighting that the majority of studies presented in this thesis were conducted before the COVID-19 pandemic. Only the last study was performed in the beginning of the pandemic in Brazil.

Keywords

Learning Analytics;   Model;   Dashboard;   Virtual Learning Environments;   E-learning.

Resumo

Damasceno, André Luiz de Brandão; Diniz Junqueira Barbosa, Simone. **Apoiando Instrutores na Análise de Logs dos Estudantes de Ambientes Virtuais de Aprendizagem**. Rio de Janeiro, 2020. 152p. Tese de Doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Cursos online têm ampliado as possibilidades de pesquisa sobre comportamento e performance de estudantes. Esta tese investiga como apoiar instrutores na análise de *logs* de estudantes em Ambientes Virtuais de Aprendizagem. Primeiro, conduzimos entrevistas com instrutores e realizamos um mapeamento sistemático do estado da arte sobre *Education Data Mining* e *Learning Analytics*. Em seguida, analisamos logs de cursos online oferecidos no Brasil e comparamos nossas descobertas com resultados apresentados na literatura. Além disso, capturamos as preferências dos instrutores em relação a visualização de comportamento e performance de estudantes. Contudo, notamos uma lacuna de trabalhos mostrando modelos para o desenvolvimento de ferramentas de *Learning Analytics*. Com base nesses estudos, esta tese apresenta um modelo conectando teorias e modelos de visualização, assim como requisitos dos instrutores, suas preferências de visualização, diretrizes da literatura e métodos para análise de *logs* dos estudantes. Instanciamos e avaliamos esse modelo em uma ferramenta para montar *dashboards*, capturamos evidências de aceitação da nossa proposta e obtivemos feedbacks dos instrutores sobre a ferramenta tais como suas preferências de análise e visualizações. Por fim, apresentamos algumas considerações e discutimos lacunas existentes no trabalho que podem fundamentar e guiar futuras pesquisas, tais como desenvolvimento de novas instâncias e implantações do nosso modelo em instituições de ensino brasileiras e avaliação de eventuais mudanças na performance dos estudantes quando instrutores visualizam informações sobre o comportamento e performance deles, e agem de acordo. É importante ressaltar que a maioria dos estudos apresentados nessa tese foram conduzidos antes da pandemia de COVID-19. Somente o último estudo foi executado no início da pandemia no Brasil.

Palavras-chave

Learning Analytics; Modelo; Dashboard; Ambientes Virtuais de Aprendizagem; E-learning.

Table of contents

List of figures

List of tables

Science is more than a body of knowledge.
It is a way of thinking.

Carl Sagan, *The Demon-Haunted World.*

1
Introduction

Distance Learning (DL) is no longer a novelty. From the records of correspondence courses in Brazil in the 19th century (Saraiva, 1996) to the rise of computers and evolution of the Internet, instructors and students have been experiencing new ways of teaching and learning. In particular, improvements in multimedia technology has contributed to make online courses feasible. According to Christel (2009), the late 1980s were marked by efforts in content digitalization, such as video and image. Since 1990s, digital data has proliferated exponentially and accumulated on the World Wide Web (Christel, 2009). Then, in the 2000s we had continual improvements in technology for content authoring, storage, distribution, and presentation, which supported the emergence of Learning Management Systems (LMS), also known as Virtual Learning Environments (VLE), such as Moodle,[1] which was released in 2001. For instance, authoring tools provide teachers with different ways to create content (*e.g.*, slideshows, videos, and games) (Damasceno et al., 2017a); and platforms such as Videoaula@RNP[2] allow for storage, distribution, and presentation of learning content (Busson et al., 2017; Damasceno et al., 2017b), such as Learning Objects (LO), defined as any entity, digital or not, that can be used, reused or referenced during a learning supported by computer(s) (Wiley, 2000; IEEE, 2002). Figure 1.1 illustrates content authoring, storage, distribution, and presentation performed by VLE.

One of the reasons for the increase in supply and demand of online courses is that students can determine their own study pace and participate in courses regardless of geographic distance limitations (Seaton et al., 2014). Today, we are witnessing the success and scale of online courses and enrollment of hundreds of people from around the world, each one with different behaviors, outcomes, and levels of involvement, inspired the Massive Open Online Course (MOOC) definition (Fini, 2009). Many institutions are now providing courses based on MOOCs, such as Coursera,[3] Khan Academy,[4] and Udemy.[5] In

---

[1]http://moodle.org
[2]http://videoaula.rnp.br
[3]http://coursera.org
[4]http://khanacademy.org
[5]http://udemy.com
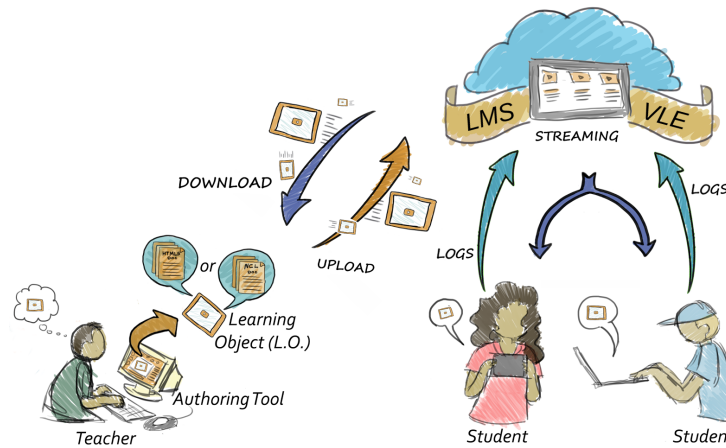
Figure 1.1: An overview of resources used by LMS and VLE (Damasceno et al., 2017b).

addition, it is worth highlighting the increase of institutions, from schools to universities, that shifted to VLEs during the COVID-19 pandemic (Chick et al., 2020; Cucinotta and Vanelli, 2020; Zhou et al., 2020). We believe that DL resources may be increasingly used, which could make students self-regulated in learning.

Virtual Learning Environments are not exclusive to distance education. Some VLEs are used together with face-to-face learning (*aka* blended learning). The usage of VLE in all classroom modes (*i.e.*, face-to-face, blended, and online education) has leveraged much research on Informatics in Education. As students' interactions with VLEs can be captured in logs, by analyzing these logs we can evaluate their learning achievements in a course, identify behavior patterns, and even predict their performance (Romero and Ventura, 2010; Dutt et al., 2017), whereas in the traditional classroom instructors face more difficulty in assessing how much effort students put into their learning.

Two research groups have related goals regarding the exploration of education data: the International Educational Data Mining Society (IEDMS)[6] and the Society for Learning Analytics Research (SoLAR),[7] which define their research area as Educational Data Mining (EDM) and Learning Analytics (LA), respectively. In both areas, we can find research on automated discovery and on human analysis using visualizations and other techniques. While EDM employs human judgment as a tool to improve the automated discovery, LA focuses on automated discovery to empower human judgment (Siemens and Baker, 2012). In other words, LA has a more pedagogical focus aiming to improve data visualization and human interpretation, whereas EDM has a more technological focus to enhance the automated discovery methods.

---

[6]http://educationaldatamining.org
[7]http://solaresearch.org

Nevertheless, online courses continue to face a serious problem of identifying students at risk (*e.g.*, low performance and drop out), which can be attributed to several reasons, such as lack of feedback to instructors. For instance, low completion rates of MOOC participants has been a central criticism (Kizilcec et al., 2013; Seaton et al., 2014). In turn, visual analytics research has developed visualization tools to aid instructors in monitoring students' achievement and to make meaningful decisions based upon the interpretation of the underlying data (Hu et al., 2014; Chen et al., 2016b; Robal et al., 2018; Sun et al., 2019). According to Gómez-Aguilar et al. (2015), tools based on visual analytics techniques exploit the human visual capabilities of achieving rapid understanding of the correlations found in the data or the comparison of data elements and the free scan by means of visual tools. Therefore, they advocate that it is essential to inform instructors about students' activities to identify potential problems and support modifications needed to optimize eLearning activities (Gómez-Aguilar et al., 2015). In line with them, Chatti et al. (2012) highlight that the Learning Analytics challenge is to design and develop useful tools to help learners, instructors, and institutions to achieve their analytics objectives without the need for having an extensive knowledge of the techniques underlying these tools.

1.1
Research Goal

Given the context above, we argue that, when instructors understand the student's behavior and performance, they can design better courses and improve the learning effectiveness. Therefore, we address in this proposal the following research question:

RQ: *How can we enable Virtual Learning Environments to assist instructors in gaining insights about both students' behavior and performance?*

Next, we briefly present existing works on this subject, each one addressing one research sub-question, contributing to the answer to our main research question.

Taking into account the last three years, we have found in the literature some review papers that presented the state-of-the-art in EDM and LA (Dutt et al., 2017; Na and Tasir, 2017; Sergis and Sampson, 2017; Vieira et al., 2018). However, each one focuses on either EDM or LA, but not both, and highlights a single research topic, such as an intervention during learning, support to instructor, data clustering, or data visualization. In addition, some papers

pose the need for research to identify which data about the learning process and students can be useful for learning analytic tools (Verbert et al., 2014; Schwendimann et al., 2017). This leads to the first sub-question:

SQ1: *What do the literature and instructors say about the analysis of student interaction logs on Virtual Learning Environments?*

In Chapter 2, we present interviews with instructors who work in Brazil and a systematic mapping on EDM and LA. This study aims to identify which kinds of information about students the instructors regard as meaningful (*e.g.*, performance, behavior, engagement); how these kinds of information are gathered; and how they drive requirements for improving their analyses (Damasceno et al., 2019b).

However, all papers found in the systematic mapping are related with course data offered outside Brazil, which drove the second sub-question:

SQ2: *How do student interaction logs of online courses offered in Brazil relate to findings in the literature?*

To identify which results in the literature are reflected in online courses offered in Brazil, in Chapter 3 we explore and analyze, using statistical methods and machine learning techniques, a dataset provided by a Brazilian institution that offers large-scale online courses (Damasceno et al., 2019).

Furthermore, we have not found in the literature works about instructors' visualization preferences of student logs. This gap is also emphasized in the literature review about learning analytics tools conducted by Schwendimann et al. (2017), where they highlight the lack of specific visualizations and visual metaphors that address the activities of learning and teaching. Therefore, the third sub-question is thus related to visualization:

SQ3: *What are the instructors' preferences regarding the visualization of student access and performance in courses using VLEs?*

In order to answer this question, Chapter 4 presents a study to identify how much the instructors take into account topics related to both students' behavior and performance, as well as their visualization preferences (Damasceno et al., 2019a).

Aligning learning design with learning analytics tools has been identified as a key issue within the learning analytics community, which requires collective effort (Bakharia et al., 2016). However, we noted that there is a lack

of works presenting conceptual models of learning analytics tools (*e.g.*, dashboards) to support the gathering and analysis data from VLEs, as well as the presentation of visualizations to provide insights and support instructors in their pedagogical decisions. This gap inspires the fourth sub-question:

SQ4: *What do dashboards require to support, from the gathering data from VLEs to the presentation of visualizations?*

To that end, Chapter 5 presents a model for dashboards that connects Visual Analytics theories and models, requirements, and guidelines uncovered by the interviews with instructors and the systematic mapping, methods used to analyze students' logs from VLEs and instructors' visualization preferences. In addition, we instantiate and evaluate this model in a tool to assemble dashboards, and obtain instructors' feedback on the tool.

It is worth highlighting that the majority of studies shown in this thesis were conducted before the COVID-19 pandemic. Only the study presented in Chapter 5 was performed in the beginning of the pandemic in Brazil.

1.2
Expected Contribution

The combination of the answers to the research questions yields a framework to enable Virtual Learning Environments to assist instructors in gaining insights about both students' behavior and performance. In addition, we hope that our proposed model might be a guide to the development of new dashboards and ground future research.

1.3
Thesis Structure

The remainder of this thesis is structured as follows. Chapter 2 reports on the interviews with instructors and a systematic mapping on EDM and LA. Chapter 3 shows a data analysis of some courses offered in Brazil. Chapter 4 details a study on instructors' topics and visualization preferences related to students' logs. Chapter 5 describes our proposed model for Learning Analytics dashboards and reports an evaluation of our proposal. Finally, Chapter 6 presents our final considerations.

2
Interviews with Instructors and Literature Review

As previously mentioned, the purpose of this chapter is to identify which kinds of information about students the instructors regard as meaningful (*e.g.*, performance, behavior, engagement); how these kinds of information are gathered; and how they drive requirements for improving the analyses.

To achieve our goal, we led interviews with instructors who work in Brazil and conducted a systematic mapping on EDM and LA. The aim of this mapping was to uncover papers that discuss the use of logs to analyze and predict both student behavior and performance. We then triangulated the answers obtained in the interviews with the instructors and the paper results found in the systematic mapping. The main outcome of this triangulation is a broad assessment of the area, which can ground and guide future research.

The remainder of this chapter is structured as follows. Section 2.1 describes the interviews with instructors and Section 2.2 describes the systematic mapping. Section 2.3 triangulates and discusses the answers obtained from the instructors and the results found in the literature. Lastly, Section 2.4 presents some final considerations, guidelines and recommendations for designing of learning analytics tools.

2.1
Interviews with Instructors

Between November 2017 and April 2018, we invited 37 instructors from 11 education institutions located in different Brazilian regions to participate in an individual interview. This sample was selected aiming to answer the following questions:

– *Which resources have instructors been using in VLEs to analyze student behavior and performance?*
– *What do the instructors need to improve those analyses?*

The interviews were conducted with 18 university instructors (13 men and 5 women) from institutions located in six Brazilian states (Goiás, Maranhão, Minas Gerais, Pernambuco, Piauí and Rio de Janeiro). Before to start the interviews, all instructors signed an informed consent form allowing us to

share their information (available at Appendix A). The interviews were one-hour long, semi-structured (available at Appendix B), conducted remotely or in person (depending on the availability and location of the participant). All instructors work in institutions that make use of VLEs.

We categorized the instructors' statements using a process of open coding Flick (2018). Table 2.1 compiles the interview results relating each instructor[1] to the topics they discussed. Most instructors (except I06, I07, I08 and I09) teach in STEM[2] courses. In particular, I03 teaches engineering courses, and I02, I04, I05, I10, I11, I12, I13, I15 and I18 teach computing courses. They all had experience with Distance Learning (DL): nine had taught exclusively distance learning courses, one taught a blended learning course, and eight taught both types of courses. In total, the instructors mentioned having worked with 10 different VLEs, and Moodle was the most often cited one (by 15 of the 18 instructors).

Table 2.1: General results of the interviews.

Category	Total values
Used Resources in VLEs (Communication tools; Postings; Assigned activities; Content repository; References from Web; Videoconference; Wiki; Blog; Badges and rankings; Dashboard for monitoring the student access; Other)	18 10 11 4 4 4 3 2 1 1 7
Analysis of students' interaction with the content (Access logs; Logs of interactions with the materials; Do not analyze student interactions; Forum postings and participation; Evaluating assessments and exams; Other; Student feedback; Observing which students are online)	11 9 6 5 5 3 2 1
Meaningful informations about students (Background; Interaction; Performance; Learning; Course feedback; Expectations, Intentions, and Motivations; Knowledge usefulness; Profile; Perception; Age; Difficulties; Participation; Fulfilled goals and expectations; Drop out; Other)	16 14 12 12 10 8 7 6 4 3 3 2 2 2 3
Behavior patterns to be identified (Access; Interest; Forum usage; Interaction with content, instructor and each other; Participation; Performance; Responsibility; Other)	8 6 5 4 4 2 2 4
Evaluation of students' motivation (Communication; Assignment completion; Access; Reaction; Participation; Does not know; Performance; Feedback; Badges; Interest; Other)	13 7 5 3 3 3 2 2 2 2 2
Factors that affect the student to watch a video lecture (Video length; Way in which instructors pose and express themselves in videos; Format in which the content is presented; Closeness to an exam date; Video production and editing; Interactivity and the use of images to present the content; Video content; Being able to watch a video on a mobile device)	15 8 8 7 7 5 3 1
Students' interaction and performance (Students' access to materials; Students' participation in forum; Assignment completion; Video access; This information is not available to the instructor)	14 7 5 3 2

I01 reported that he had been using VLEs in face-to-face teaching and he noticed that students' performance improved when compared with the period that he did not use VLE, and I06 remarked that students participate more in online courses.

All instructors stated they make use of some communication tool with their students. Most of them use the VLE to assign activities (*e.g.*, quiz, exercises) and to post educational content and news. I04 uses gamification

---

[1]In this chapter, instructors are identified in the format I99.

[2]Acronym of Science, Technology, Engineering, and Math.

techniques in the VLE and noticed an improvement in the class. According to them, students like to earn badges, participate in rankings, and compete with one another. In line with this statement, I10 said that competitive students tend to perform better.

We asked how they analyze students' interaction with the content. Although only I17 informed the use of a tool (*i.e.*, dashboard) for monitoring student access in the VLE, 11 interviewees reported that they examine the access logs, 9 analyze the logs of students interactions with the materials available in the VLE, and 6 do not analyze student interactions. In particular, I05 said that he had used a dashboard to oversee students and another tool designed for students to highlight text, from which he gathered information to support his pedagogical decisions.

Not all students behave the same way. I04 said older and younger students have distinct interaction patterns. Furthermore, some instructors (I03, I04, I05, I09, and I11) reported that there are students who only turn in their assignments, without any other interaction. According to I09, this "virtual silence" is challenging, whereas I03 and I05 reported that in these cases the only feasible analysis is through the assignments and exam results. Regarding students who achieve a good grade despite having very little interaction with the VLE, some instructors provided the following interpretations:

– not every student likes to interact with instructors (I03, I04, and I11);

– there are students who do not interact because they already know the content (I04).

We asked instructors which kinds of information they find meaningful. Besides interaction data analysis, 16 instructors in this study believe that identifying student background is important. They mentioned, for instance, the importance of knowing how much students know about the course material and their individual background knowledge. They stated that such information helps the overall planning of the course and to match the difficulty level of the content. They also mentioned their interest in student performance, learning, expectations, and course feedback.

When asked about how they evaluate student learning and motivation, 15 instructors answered that they use exams and assignments to assess student learning; and 13 instructors estimate students motivation by observing their communication in the VLE, such as the use of the forum and chats, and students' contributions and questions. Seven said they also evaluated motivation by observing which students completed assignments and which ones were accessing the VLE. I07 believes motivation to be one of the main

factors influencing student learning. I06 stated that the instructor has to make dynamic lectures and incentivize students in order to engage and motivate them. For instance, I04 said that they motivate students using gamification techniques. Nevertheless, I03 and I09 said that it is hard to evaluate student motivation in DL. With face-to-face teaching, this evaluation is made easy by analyzing facial expressions and voice intonation. In turn, I02, I03 and I05 claim that, in DL, the signals emitted by students are their interaction with the materials and forum. In line with them, I06 reported that student motivation cannot be quantified. In contrast, I16 said face-to-face teaching is more difficult because a student is a "black-box", whereas a VLE provides more student information (*e.g.*, participation and contributions, performance, engagement).

We asked the instructors what student behavior patterns they would like to identify. They reported access patterns, interest (inferred, for instance, from students' questions (I18) and from watching video segments more than once (I09)); forum usage; interaction with the content, with the instructor, and with one another; and participation, among others. Only one instructor claimed to use a tool to analyze student access logs; most instructors said that they do not have any analytic tool or information besides student access logs, which makes it impractical to analyze students' interaction with the VLE in classes with a large number of students.

Table 2.2 and Table 2.3 show requirements to VLE and video lecture mentioned by the interviewees. A total of 12 instructors pointed out the need for tools to analyze students access and interaction logs with the material and other VLE resources, such as the forum, video player, quiz, and e-books. The instructors believe that systems who satisfy those needs will allow identifying student interaction patterns and predicting student performance and drop out. The analysis suggestions made by the instructors were:

- to analyze emoticons to identify the students' mood;
- to identify the students' background through their behavior and reaction;
- to adapt content for students using decision trees;
- to correlate student access to materials and drop out;
- to correlate student navigation on video and their performance.

In addition, I12 and I16 mentioned that these pieces of information should be presented in a dashboard. According to I16, in face-to-face teaching, instructors are very often overwhelmed with preparing their classes, assessing coursework and evaluating students. The usage of VLEs has mitigated this

Table 2.2: Requirements gathered to VLE.

Requirement description	Number of instructors
Identify student access patterns (*e.g.*, login, materials)	9
Identify student interest patterns on the course	6
Identify student usage patterns on the forum	5
Identify student performance pattern	5
Identify student interaction patterns (*e.g.*, materials)	4
Identify student participation patterns on the course	4
Capture students' reactions with materials	3
Identify student navigate patterns on the VLE	2
Identify self-regulated students	2
Predict student performance	2
Identify pace learning student	2
Identify student drop out patterns and reasons	2
Know when students has connection problems	1
Identify students' expectations	1
Identify student mood patterns (analysing emoticons)	1
Identify student learning patterns	1
Automatic content recommendation to student	1
Self-regulated content to student (using tree decision)	1
Resources to students evaluate the materials	1
Provide a course progress bar to student	1
Provide achievements to engage the students	1
Relate content presentation with student behavior	1
VLE show a schedule customized to student	1

Table 2.3: Requirements gathered focused on video lecture.

Requirement description	Number of instructors
Statistics of interactions on video (*e.g.*, access, re-watch, seek)	17
Know if the student has understood the video	3
Know if the student is watching the video	2
Relate both students' navigation and performance	2
Relate video length and student performance	2
Markers on the video showing where begins each topic	2
Know what videos (or segments) the students have liked	2
Relate video script and student performance	1
Know the backjump reason	1
Video's content be interactive	1
Know video segments that the students have difficulty	1
Know how students navigate among the video lectures	1
Know how students navigate in the video	1
Know if another material was accessed with the video	1

problem and, by using a dashboard that provides student behavior information, they could spend more of their time working on teaching methods and materials. The instructors also said that a dashboard could help them in several ways:

– to produce educational material that takes into consideration students' background and performance;

– to know what materials to upload in the VLE;

– to understand the reason for students to drop out;

– to compare the performance of students in his/her class;

– to make pedagogical decisions to enhance students' performance and reduce drop out.

In line with this, I18 stated that, without such a tool, it is unfeasible to evaluate students in an online course, whose number of enrollments can be huge.

Out of the 18 instructors, only I11 had never created online educational content, because they only tutored students and had a content manager to create the course materials. Conversely, 14 instructors reported having experience in the authoring of a video lecture. Some of them emphasized the value of using video in education. They reported that video is the preferred content format by students because it allows students to watch (part of) a class more than once, and the video lecture is often the gateway to knowledge. I11 and I12 also remarked that video lectures improved both student understanding and performance. Moreover, I01 and I11 noticed that students were less engaged in videos using just slides or captured from a live classroom lecture. By contrast, I01, I16, and I18 stated that the video format in Khan Academy and Talking Head engaged students the most.

When asked about student interaction with videos, most instructors (17) said they would like to know the number of views of particular parts of the videos, and which segments were skipped or re-watched by students. According to most of them, this information could provide insights to the instructor about what the students found relevant and about which segments the students had more doubts. In addition, some instructors stated their needs explicitly:

– to know whether students are watching and understanding the video;

– to add markers to the video player showing students where each content topic begins and ends, thereby enabling them to find the content more efficiently;

– to identify (parts of) videos the students liked.

The instructors cited the following factors that influence a student to watch a video:

– the closeness of an exam date;

– the quality of video production and editing;

– the format in which the content is presented;

– the way in which the instructors pose and express themselves in videos.

For instance, I04 and I12 noticed that students are less engaged on videos where instructors speak slowly. The video length was the factor most often mentioned (by 15 instructors). According to them, students do not usually access long videos and, when they do, they do not watch the whole video. However, there was no consensus on the ideal length of a video lecture. Responses ranged from 5 to 30 minutes.

Another point noticed by the instructors is that students are less engaged with videos that have more theoretical content. I01 and I06 said students are more engaged on math video lecture presenting exercise solutions. I06 also remarked that students provided feedback through questionnaires, reporting that they do not like video lectures presenting only the instructor speaking, without demonstrations and images related to the content. Furthermore, the instructors highlighted that:

– the most important point is the content presented;

– a super production of a video lecture requires a reason;

– watching a video lecture does not ensure that a student has learned the content;

– there is no single video format that performs well with all the students.

Regarding the relationship between students' usage of the VLE and their performance in the course, 14 instructors reported that students' access to materials is related to their grades, 7 said that students who participate more (*e.g.*, making questions, using the forum, chatting) perform better, and 5 use assignment completion as a cue that a student will perform well. I01 said that showing the solution after students have answered questions negatively affects their performance. In addition, I03, I04, and I05 stated that each student has their own study style and, according to I07, the instructors should provide the content using more than one format (*e.g.*, text, video). I05 said students who already know the content prefer text material instead of other media format, and I02 reported that lecture notes are one of the contents that students like the most.

We also asked instructors whether they had identified any relationship between student interaction and drop out rate in VLEs. Half of them reported not having identified any relationship, and 7 perceived a correlation between student access and course completion. Ten other answers were provided for this

question, but none was mentioned by more than 2 instructors. These answers relate student drop out to:

– affinity with the content;

– accumulation of homework assignments;

– problems with network bandwidth limitations;

– student time available;

– students with problems of interaction with another student or group of students (*aka* peer interaction);

– problems in the usage of VLE

– students interested in part of the content;

– access in only the first weeks (*i.e.*, drop outs do not usually occur later in the course).

I07, I10, I11, and I13 reported problems with network bandwidth limitations, which may complicate such analysis. For instance, I10 and I11 said that, in general, the students tend to download the course materials instead of accessing them online.

2.2
Systematic Mapping of Educational Data Mining and Learning Analytics

The systematic mapping conducted in this paper was based on the method described by Kitchenham and Charters (2007), following a well-defined protocol to highlight the main problems, objectives, methods, case studies, and results presented in the gathered papers. The whole procedure was performed by only one person. The first step was to define the research questions to guide the mapping:

– *Which results related VLE logs analysis and student behavior and performance?*

– *Which tools (e.g., dashboards) are used by instructors to analyze logs of student interactions with VLEs?*

Next, we defined a search string stemming from the combination of keywords related to the research questions: *(education OR course OR MOOC OR "e-learning" OR teaching OR "virtual learning environments" OR "virtual training environments" OR "learning management systems" OR LMS)* **AND** *(engagement OR behavior OR behaviour)* **AND** *(analysis OR analyses OR analytics OR analytic OR visualisation OR visualization OR "data mining" OR "learning analytics").*

We used the advanced search systems of three digital libraries: ACM,[3] Elsevier[4] and IEEE[5]. We set filters in all libraries to return only papers published after 2009, in PDF format, and written in English or Portuguese. As Figure 2.1 shows, this procedure followed four steps:

Step 1	Step 2	Step 3	Step 4
Search on digital libraries	Remove duplicate papers	Analysis of titles and abstracts	Full text reading
Total: 2,174	Total: 1,835	Total: 320	Total: **136**
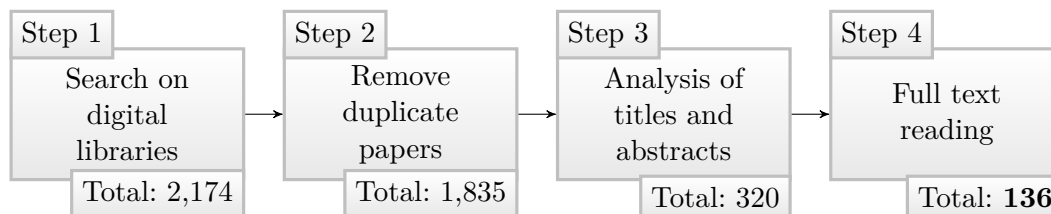
Figure 2.1: Study selection process.

1. the search on the digital libraries returned 2,174 papers;

2. we removed duplicate papers, leaving 1,835 papers;

3. we analyzed titles and abstracts of each paper using the inclusion and exclusion criteria, resulting in 320 papers;

4. we read the 320 papers in full, also applying the inclusion and exclusion criteria. Finally, 136 papers were selected for this mapping.

Inclusion and exclusion criteria were used to filter irrelevant papers that, despite including the defined keywords, do not present results to answer the research questions. The inclusion criteria were:

– papers that present results, methodologies or case studies related to data analysis (*e.g.*, logs) to measure the student performance, motivation, participation, or drop out in VLEs;

– papers that present ways to detect students' behavior pattern in VLEs;

– papers that show results, methodologies, or case studies to view logs in VLEs;

– papers that evaluate students' interaction problems in VLEs.

The exclusion criteria were:

– call for papers or keynotes;

– papers focused on face-to-face teaching (without VLE) or course recommendation;

---

[3] http://dl.acm.org/advsearch.cfm
[4] http://sciencedirect.com/science/search
[5] http://ieeexplore.ieee.org/search/advsearch.jsp

– papers aiming to improve the accessibility for people with special needs;

– papers that present results related to emoticons analysis;

– papers that analyze data only from questionnaires;

– papers without results.

Table 2.4 shows the paper distribution by year and digital library. It is worth noting that, although only one paper appears in 2019, this is expected because the search was performed on 25 February 2019.

Table 2.4: Papers distribution by year and digital library.

Library	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019*	Total
ACM	-	2	3	6	14	13	16	21	23	1	99
Elsevier	1	-	-	3	2	6	2	3	2	0	19
IEEE	1	-	-	-	3	2	2	8	2	0	18
Total	2	2	3	9	19	21	20	32	27	1	**136**

Table 2.5 lists the main paper sources (*i.e.*, at least 5 papers), including conference proceedings and journal articles.

Table 2.5: Main paper sources.

Source (conference proceedings and journals articles)	Total
ACM Conference on Learning Analytics & Knowledge (LAK)	36
ACM Conference on Learning @ Scale (L@S)	21
Elsevier Computers in Human Behavior	13
ACM Technical Symposium on Computer Science Education (SIGCSE)	6
IEEE Frontiers in Education Conference (FIE)	5
ACM Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM)	5

We classified the papers according to problems, objectives, methods, case studies, and results.[6] We also classified the type of course analyzed by each paper: 38 papers analyzed theoretical courses, 29 practical, and 32 both. However, 37 papers do not describe the type of course analyzed.

As we can see in Tables 2.6 and 2.7, most papers focus on students' behavior or performance, aiming at identifying student behavior or performance patterns. Table 2.8 shows that most of these papers (101 out of 136) focused on STEM courses as case studies (*e.g.*, Computing and Engineering courses), and 16 papers had miscellaneous areas as case studies, without specifying them.

We also classified papers according to the methods they adopted. We found 63 different methods. Table 2.9 presents the main found methods

---

[6]Available at: `https://github.com/andrelbd1/research-edm-la/blob/master/README.md`

Table 2.6: Found problems.

Problems	Total
Analyze student behavior (*e.g.*, interaction, engagement)	64
Analyze student performance or learning	38
Students at risk (drop out or bad performance)	24
Prediction of student performance	22
Difficult to analyze, manage or view data from e-learning systems	14
Online learner participation	9
Evaluation teachers process (*e.g.* time consuming, difficult)	7
Lecture video in order to assist the learning	4
Prediction of student behavior	4
Language proficiency	3
Generation of feedback to teacher	2
Learning to program	1
Literature review	1

Table 2.7: Found objectives.

Objectives	Total
Identify learners' patterns of behavior or trajectory	68
Identify learners' patterns of performance or learning	41
Make or improve a tool to analyze data from the e-learning systems	24
Identify learners' patterns of engagement and disengagement	17
Identify predictors of performance	15
Analyze online discussion activity (*e.g.*, forum, chat)	13
Predict students at risk (*e.g.*, drop out, bad grades)	12
Identify learners' profiles in online discussion (*e.g.*, forum)	7
Design visualizations	5
Identify learners' motivations	5
Make students aware of their behavior or performance	5
Allow teachers know students' learning performance during the course	4
Develop a theoretical model	4
Analyze the use of badges	2
Identify learners' personality	2
Identify reasons for drop out	2
Relate students' goals at the course with their performance	2
Teachers' behavior (*e.g.*, patterns, interventions to increase academic performance)	2
Collect student interactions	1
Literature review	1
Offer action recommendation to teacher	1
Predict students certification	1

(mentioned by at least 2 papers). Most papers used methods based on clustering (*e.g.*, K-means) and prediction (*e.g.*, Logistic Regression, Decision Tree). We found 18 papers that applied methods using metrics defined by their own authors, relating the students' interaction with communication tools (*e.g.*,

Table 2.8: Found case studies.

Case Studies	Total
Computer	52
Business	24
Engineering	17
Misc. areas	16
Health and Biology	13
Mathematics	7
Psychology	6
Sociology	6
Art	5
Chemistry	5
Education	5
Law	3
Physics	3
Design	2
English	2
History	2
Sport	2
Geoscience	1
Linguistics	1
Material Science	1
Pharmacy	1

forum, chat) and their grades, and 20 papers relating students' interaction or performance with data from a qualitative questionnaire.

We categorized 137 different results in 14 topics, as shown in Appendix C. We noticed that 53 papers show that it is possible to cluster students based on their access and interaction patterns.

Regarding the students' performance and interaction with forums, we found:

– models to predict student performance;

– correlations between student access, completion assessment, participation, and performance;

– evidences that:

  – materials available in distance courses are ignored even by good performance students;
  – student groups that use forums more tend to have a good performance.

Papers addressing tools yielded the following main result categories (with at least 5 papers each):

Table 2.9: Main found methods.

Methods	Total
Correlating questionnaire data with interaction or performace	20
Logistic regression	19
Only indicators defined by own authors;	18
K-means clustering	17
Social network analysis (*e.g.*, forum, chat), Correlation between online discussion activity (*e.g.*, forum, chat) and grades achieved.	16
Decision tree	13
Linear regression	12
SVM	10
Spearman correlation	9
ANN - Artificial neural network (*e.g.*, RNN - Recurrent Neural Networks (*e.g.*, Long Short Term Memory (LSTM)))	8
ANOVA	8
Pearson correlation	8
Text mining techniques (*e.g.*, SPSS Clementine's, Latent semantic analysis (LSA), Probabilistic latent semantic analysis (PLSA))	7
Analysis using tools (without specify method) as excel, weka , matlab	6
Chi-squared test	6
Random forest	6
Naïve Bayes	4
Bayesian Network	3
Bayesian Information Criterion (BIC)	2
Calculates the vector Euclidean distance	2
Fuzzy Inductive Reasoning (FIR)	2
Fuzzy Miner algorithm (*e.g.*, Disco Miner)	2
Gaussian Processes (GP)	2
Hidden Markov model (HMM)	2
JRIP	2
K-Nearest Neighbor (K-NN)	2
MANOVA	2
N-gram models	2
RandomTree	2
Semantic Spiral Timeline (SST)	2
Stochastic gradient descent (SGD)	2

- analytics in learning systems used to provide both auditing and interventions in student learning;

- tools to aid instructors to analyze student behavior;

- proposals of tools that use logs from eLearning systems for instructors to monitor students behavior, motivation, or performance;

- tools to aid students in analyzing their own performance.

The term *engagement* is used in 23 papers, most of which measure engagement as:

Table 2.10: Grouped found results.

Results	Total
Cluster	53
Students' performance	47
Forum	31
Tools	30
Students' engagement	25
Course completion	20
Videos	15
Others	14
Students' drop out	12
Students with self-regulated learning	5
Students' intention	3
Students' personality	3
Attending	2
Click activity	2

– assignments posted on the VLE;

– materials accessed on the VLE;

– total hits, readings, and postings on communication tools (*e.g.*, forum).

Some papers also related engagement to how long students take:

– watching each video;

– taking notes.

Most papers related to the *Course completion* topic found that:

– MOOCs typically have lower completion rates;

– assignment completion can be used as a predictor of student course completion;

– students that initiate threads in forums tend to complete the course.

We found 23 result categories related to the *Video* topic. However, the only category including more than 5 papers is the one stating that students often do not watch the entire videos. The others include just 1 or 2 papers. In regard to the *Drop out* topic, all papers present proposals to predict student drop out, but only 4 papers show models to predict drop out through data analysis of student interaction logs. Six other result categories are assigned to this topic, but none of them has more than 2 papers.

Papers assigned to the *Students with self-regulated learning* topic state that students with negative self-regulation have poor academic performance. In addition, 3 papers claim that having self-regulation does not necessarily

imply good performance. The *Intention* topic presents 13 result categories, in which only one category has 2 papers. In particular, the others has only the same paper as result. The *Personality* topic includes 2 papers showing that the students' personality can be identified through the interaction logs. We also identified the *Attending* topic to include papers presenting results of students' attendance detection in DL and correlations between attendance and demographic data (*e.g.*, country of origin and education level). The 2 papers assigned to the *Click activity* topic claim that the click count increase and decrease are related with a probability of passing and failing a course, respectively. The *Students' intention* topic includes 2 papers relating the students' intention in the course with their behavior and performance. In those papers, the students' intention was assessed through questionnaires. Lastly, the *Other* topic presents 5 categories, but only one of them has more than 5 papers: it states that access to the online environment resources increases in periods close to exams or assignment deadlines.

2.3
Triangulation of Results

In this section, we present the main relations between the results of the interviews and of the systematic mapping on EDM and LA. We relate the instructors' statements with the results of the selected papers. The results mentioned in both studies (*i.e.*, by the interviewed instructors and in the literature) are listed in Table 2.11. We highlight that STEM courses are the most prevalent in each topic presented in that table.

I04 noted that older and younger students have distinct interactions with VLEs, in line with the results presented by Guo and Reinecke (2014), which shows that older students tend to have a good performance and realize backjump actions more frequently on video players than younger students. Papers have shown that older students participate more in forums (Kizilcec et al., 2013; Ransdell, 2013; Guo and Reinecke, 2014). I03, I04, I05, I09, and I11 stated that there are students who do not interact either with the instructor or in forums; they just access the materials and achieve a good grade. In line with this statement, Hogo (2010); Cobo et al. (2012); Wilkowski et al. (2014); Xu et al. (2014); Ferguson and Clow (2015); Hernández-García et al. (2015); Kovanović et al. (2016); Wang et al. (2016b); Corrin et al. (2017); Sunar et al. (2017) show clustering techniques of VLE data and present a similar student profile (*aka* lurkers), among others. To motivate students, I04 makes use of gamification techniques (*e.g.*, badges and rankings) and has noticed improvements in the class. In line with this result, Anderson et al. (2014);

Table 2.11: Triangulation results.

Topics	Instructors' statements and paper results	N instructors	N papers	Total
Performance & Course Completion	Assignment completion is a cue that the student will achieve a good performance	6	29	35
Performance & Forum	Students who use the forum more tend to achieve good grades	7	20	27
	There are students who do not interact either with the instructor or in forums, access the materials and achieve a good grade	5	10	15
Video	Students engage less with long video lectures	15	5	20
	Format how video lecture is presented influences the student to watch it	8	2	10
	Video editing and production affect the student who watches it	7	1	8
Engagement & Forum	Students' interest can be identified by their interactions on the forum	4	18	22
Others	Access close to exam dates or assignment deadlines increase	7	10	17
	Each student has his/her own study style	3	9	12
	Older and younger students have distinct interaction patterns	1	3	4
Performance	There is a positive correlation between student access to the materials and performance	14	2	16
	There is a positive relation between video access and student performance	3	3	6
	Gamification improves student performance	1	2	3
Course Completion	There is a positive correlation between student access and course completion	7	3	10
Engagement & Video	Students are less engaged in videos with more theoretical content	4	2	6
	Students are less engaged in videos where instructors speak slowly	2	1	3
	Students are more engaged in videos with Khan Academy format	2	1	3
	Students are more engaged in videos with Talking Head format	2	1	3
	Students are less engaged in videos captured from a live classroom lecture	1	1	2
Engagement	The instructor has to encourage students to keep them motivated	1	2	3
	Motivation is one of the main supporters of student learning	1	1	2

Dicheva et al. (2019) identified improvements in student engagement when compared to distance courses that use and do not use badges.

I04, I05, I09, I10, I11, I16, and I17 noticed that students using the forum more tend to achieve good grades. In addition, I02, I03, I05, and I09 identify the students' interest by their interactions on the forum. In line with these answers, we found the following results:

– student groups that use more forums tend to have a good performance (Nandi et al., 2011; Cobo et al., 2012; Paredes and Chung, 2012; Haig et al., 2013; Ransdell, 2013; Anderson et al., 2014; Feng et al., 2015; Kizilcec and Schneider, 2015; Zhu et al., 2016; Carter et al., 2017; Corrin et al., 2017; Wan et al., 2017; Sunar et al., 2017; Bosch et al., 2018);

– student groups that post more replies in forums tend to complete the course (Feng et al., 2015; Hernández-García et al., 2015; Carter et al., 2017; Andres et al., 2018) and have a good performance (Feng et al., 2015; Hernández-García et al., 2015; Carter et al., 2017);

– student groups that initiate threads in forums tend to complete the course (Feng et al., 2015; Hernández-García et al., 2015; Carter et al., 2017; Houston et al., 2017; Andres et al., 2018) and have a good performance (Feng et al., 2015; Hernández-García et al., 2015; Carter et al., 2017; Houston et al., 2017);

– there is a positive correlation between the number of questions students asked the instructor and their final grade (He, 2013);

– student groups that have more posts are more likely to complete the course (Yang et al., 2015; Qiu et al., 2016; Sunar et al., 2017; Bosch et al., 2018; Andres et al., 2018);

– engagement with the online environment can be measured by total hits, readings, and postings (Haig et al., 2013; Ransdell, 2013; Wells et al., 2016; Zhu et al., 2016; Hamidon, 2018; Gong et al., 2018);

– student groups that complete more assignments tend to use more forums (Kizilcec et al., 2013);

– forum usage can be used as a predictor of students completing the course (Qiu et al., 2016; Hermans and Aivaloglou, 2017; Sunar et al., 2017; Andres et al., 2018);

– comments can be used as a predictor of student performance (Sorour et al., 2015; Bosch et al., 2018).

The increase of access to VLE resources and materials (*e.g.*, videos) in periods close to exam dates or assignment deadlines was noted by I01, I02, I03, I04, I06, I12, and I13, as well as by Nandi et al. (2011); Wolff et al. (2013); Haig et al. (2013); Wen and Rose (2014); Nespereira et al. (2014); Park et al. (2017); Shi et al. (2017); Cicchinelli et al. (2018); Raad and McKay (2018). Complementing that, Nandi et al. (2011); Wells et al. (2016) show an increase in forum posts close to deadlines.

According to I01, I02, I10, I11, I14, I15 assignment completion is a cue that the student will achieve good performance. In line with this, Haig et al. (2013); Mödritscher et al. (2013); Bogarín et al. (2014); Guo and Reinecke (2014); Yu and Jo (2014); Wilkowski et al. (2014); Champaign et al. (2014); Nespereira et al. (2014); Anderson et al. (2014); Kennedy et al. (2015); Gómez-Aguilar et al. (2015); Samson (2015); Pardo et al. (2016); Chen et al. (2016a); Khosravi and Cooper (2017); Shi et al. (2017); Wan et al. (2017); Tempelaar et al. (2017); Feild et al. (2018); Cicchinelli et al. (2018); Nguyen et al. (2018); Al-Shabandar et al. (2018); Boulton et al. (2018) show that successful students are more frequently engaged with online assignments and participate regularly. These instructors and Wilkowski et al. (2014); Kizilcec and Schneider (2015); Samson (2015); Qiu et al. (2016); Hermans and Aivaloglou (2017); Feild et al. (2018); Cicchinelli et al. (2018); Andres et al. (2018); Al-Shabandar et al. (2018); Maldonado-Mahauad et al. (2018) also state that assignment completion can be used as a predictor of course completion. For instance, Rienties et al. (2015); Feild et al. (2018); Cicchinelli et al. (2018); Nguyen et al. (2018); Al-Shabandar et al. (2018) presents positive correlations between productive, assignment completion, and pass rates. I06, I10, I11, I12, I13, I15, I17, Zhang and Zhu (2017); Laveti et al. (2017); Al-Shabandar et al. (2018) also noticed that there is a positive correlation between student access and course completion. Additionally, almost all instructors (except I06, I08, I10, and I18) stated that there is a positive correlation between student access to the materials and performance. In line with them, Elbadrawy et al. (2015); Al-Shabandar et al. (2018) found that viewing the course materials and the students' previous performance contribute to predicting grades.

I07 claims that motivation is one of the main supporters to student learning. In line with them, Wang et al. (2016a) stated that online learning requires even more learner motivation and self-direction than traditional classroom-based instruction. According to I06, the instructor has to encourage students in order to keep them motivated, and Brinton et al. (2014) showed that an active participation of the teaching staff in the forum is associated with

a higher post volume. This issue was also explored by Dias et al. (2015), finding that instructor participation (*e.g.*, posts, activity) leads to student engagement (*e.g.*, module, wiki, blog, form, forum). Moreover, I03, I04, and I05, as well as Hogo (2010); Bos and Brand-Gruwel (2016); Wang et al. (2016a); Davis et al. (2018); Bosch et al. (2018); Sheeba and Krishnan (2018); Cicchinelli et al. (2018); Boroujeni and Dillenbourg (2018); Maldonado-Mahauad et al. (2018), report that each student has their own study style. For instance, Bos and Brand-Gruwel (2016) identified different clusters of students based on differences in their regulation strategies. They also report that the use of the same learning resources to the same extent may have a different impact on different groups of students.

In regard to video, most interviewees (except I02, I08, and I13) highlighted video length as one of the main reasons that influence a student to watch a video lecture (or not). In line with this, Guo et al. (2014); McGowan et al. (2016) stated that short videos promote more engagement. Furthermore, McGowan et al. (2016) found a correlation between audience retention and video length. Another meaningful result was identified by Qiu et al. (2016); Wang et al. (2016a); Al-Shabandar et al. (2018), which found a positive correlation between the amount of time students watch a video and learning results. Besides, I04 and I12, as well as Guo et al. (2014), note students generally engage less with videos where instructors speak slowly.

I04, I06, and I16, as well as Qiu et al. (2016); Wang et al. (2016a); Al-Shabandar et al. (2018), stated that there is a positive relation between video access and student performance. However, there are clues that student retention on videos is related to the video authoring. According to I01, I05, I06, I10, I12, I17, I18 and Guo et al. (2014), video editing and production affect the student watching it. In particular, I01 believes that videos using slides do not achieve a good result. I01, I06, I09, and I14 have noticed that there is little student engagement with videos with more theoretical content, without solving exercises. In addition, we found in the literature that students re-watch tutorials more frequently than lectures (Guo et al., 2014). Moreover, McGowan et al. (2016) also stated that theoretical videos perform worst in terms of holding the students' attention on video than videos with code demonstration and videos with coding walkthroughs tend to have a higher engagement than the active coding sections.

Lastly, I01, I06, I07, I08, I09, I11, I17, and I18 reported that the video lecture format affects the student watching it. I01 and Guo et al. (2014) support that students generally engaged less with video captured from a live classroom lecture. They noted that, when a pre-production was made, the

students engaged more with video lectures. Guo et al. (2014) also revealed that students engaged more with videos filmed informally with the instructor sitting at their office desk when compared with a video filmed in a multi-million dollar TV production studio. In line with this result, I03 believes that a super production of a video lecture requires some justification. Additionally, I01, I16, and I18, supported by Guo et al. (2014), said that Khan Academy and Talking Head video formats perform better in terms of student engagement. By contrast, McGowan et al. (2016) claimed that video lectures using slides with audio in the background perform worst.

2.4
Concluding Remarks

This chapter reported an analysis of the answers to our research questions obtained through interviews with course instructors who use VLEs. Some instructors said they would like to compare performance and drop out with the student interaction in the VLE. Normally, instructors get information from observing what students say and do on VLEs. A few challenges interfere with this analytical process because most instructors are not statistics experts or do not receive training to extract key information from VLEs. Therefore, tool support is called for. The instructors suggested tools to analyze student interaction logs in the forums, to capture how students react to course content, and to detect patterns of student navigation in the VLE materials and resources. They also want to identify the relationship between student access and the drop out rate. Some instructors emphasized the importance of visualizing these data through dashboards that presented, for example, the weekly summary of a class and of each student (*e.g.*, who accessed, who participated in the forum or chat, who submitted the assignment). For instance, I10 and I12 highlight the difficulty of doing such an analysis in the platform they use because, even using the available filters, the logs are presented in a confusing way.

We also presented the results of a systematic mapping of EDM and LA. A total of 136 papers were selected and categorized according to problems, objectives, methods, case studies, and results. 137 results were categorized related to student logs analysis and tools. The mapping shares limitations with similar mappings in the literature: some important work may not have been included, such as thesis and dissertations, books, and even some papers, which may not have been found in the digital libraries using the search and selection protocol, such as papers from Brazilian digital libraries. In addition, our mapping uncovered papers published from 2010 to 25 February 2019.

Although the number of instructors is small and their answers are anecdotal, the interviews showed close correspondence between their statements and the paper results. It was not our goal to achieve statistical significance, but even with the small number of interviewees we uncovered existing gaps in the literature. We found in the mapping 30 tools to support instructors in analyzing logs. Because Moodle is the most cited VLE by the instructors, we looked into the documentation[7] of version 3.6, finding 37 additional tools to provide information based on student logs. In Table 2.12 we summarize the instructors' needs and the characteristics of existing tools. We note that most of those needs are satisfied by one tool or another. However, none of the existing tools fulfills all requirements raised by the instructors. From what we have learned, we show in Table 2.13 some design guidelines for student log dashboards. These guidelines are sorted by the number of papers that refer to them. Obviously, the requirements and guidelines lists are not exhaustive, and further research is called for.

Some instructors' statements (mentioned by at least 2 instructors) were not addressed in any of the papers found:

– the way teachers pose and express themselves in the video lectures affects the student watching it;

– videos need to show images related to the content to engage the student watching it;

– video content affects the student watching it,

– videos using slides are not effective;

– video lectures improve both student understanding and performance;

– student drop out occurs only in the first few weeks;

– students without affinity with the content tend to drop out of the course;

– the course needs to match the students' learning styles.

Although the instructors had defined requirements to support them in their decision making, we have not found in the literature any evidence relating the use of the analytic tools supporting such requirements and the improvement of student performance. We also noticed a gap in regard to analyzing instructor behavior in VLEs. Most papers we found analyze only student behavior.

---

[7]https://docs.moodle.org

Table 2.12: Requirements gathered with instructors and implemented by tools described on found papers and Moodle documentation.

Requirement description	N instructors	N tools	Total
Statistics of interactions on video (*e.g.*, access, re-watch, seek)	17	1	18
Identify student interest patterns on the course	6	9	15
Identify student access patterns (*e.g.*, login, materials)	9	5	14
Identify student performance patterns	5	8	13
Predict student performance	2	10	12
Identify self-regulated students	2	8	10
Provide a course progress bar to students	1	9	10
Identify student usage patterns on the forum	5	3	9
Identify student drop out patterns	2	7	9
Identify student interaction patterns (*e.g.*, materials)	4	3	7
Identify student participation patterns on the course	4	3	7
Know whether the student has understood the video	3	4	7
Identify student navigation patterns on the VLE	2	2	4
Capture students' reactions to materials	3	1	4
Know which videos (or video segments) the students have liked	2	2	4
Know if another material was accessed with the video	1	3	4
Know if the student is watching the video	2	1	3
Know in which video segments the students have difficulty	1	2	3
Resources for students to evaluate the materials	1	2	3
Provide achievements to engage the students	1	2	3
Identify pace learning student	2	-	2
Identify student drop out reasons	2	-	2
Relate both students' navigation and performance	2	-	2
Relate video length and student performance	2	-	2
Markers on the video showing where begins each topic	2	-	2
Know how students navigate in the video	1	1	2
Identify students' expectations	1	1	2
Relate video script and student performance	1	-	1
Know the backjump reason	1	-	1
Video's content be interactive	1	-	1
Know how students navigate among the video lectures	1	-	1
Know when students has connection problems	1	-	1
Identify student mood patterns (analysing emoticons)	1	-	1
Identify student learning patterns	1	-	1
Automatic content recommendation to student	1	-	1
Self-regulated content to student (using tree decision)	1	-	1
Relate content presentation with student behavior	1	-	1
VLE show a schedule customized to student	1	-	1

Table 2.13: Guidelines for student log dashboards.

Guidelines description	Total
Identify behavior patterns by students access and interaction	53
Identify successful students by access and assessments	23
Identify performance by student groups that use more forums	14
Predict the students' performance from their interaction	12
Identify student engagement by materials accessed	10
Identify student engagement by assessments done	10
Predict course completion by students that complete assessements	10
Identify increase of resources access with deadlines closeness	9
Predict drop out from students' interaction	6
Identify student engagement by forum interactions	6
Identify course completion by students that have more posts	5
Identify performance by student groups that initiate more threads in forums	4
Predict course completion from students that use forum	4
Identify performance by student groups that post more replies in forums	3
Identify personality by students interaction logs	3
Identify student engagement by how long students watch videos	3
Identify self-regulated students	3
Predict performance by student comments in forums	2
Identify pass and failing students	2
Identify performance by students that make more questions on forum	1
Identify student engagement by taking notes	1
Identify learner motivation	1

3
Student Interaction Logs of Online Courses Offered in Brazil

Students' interactions with Virtual Learning Environments (VLEs) are often stored in logs. As we can see in the literature review presented in the previous chapter, these logs can predict the students' performance, evaluate their learning achievement in a course, and even identify behavior patterns. However, all results presented in the review were from courses offered outside Brazil. The purpose of this chapter is to identify which results in the literature can be found in online courses offered in Brazil. To achieve our goal, we have explored and analyzed, using statistical methods and machine learning techniques, a dataset provided by a Brazilian institution that offers large-scale online courses. In general, results found in the literature related to student access and course completion can also be found in courses offered in Brazil. However, most analyses about student performance did not show results in line with existing work.

This chapter is structured as follows. Section 3.1 describes both the courses and dataset used, as well as the methods adopted in the data analysis. Section 3.2 presents the procedure adopted to analyze the student interaction logs and discusses the results of the analysis. Finally, Section 3.3 discusses our findings.

3.1
Dataset and Methods

Our dataset was provided by UNASUS-UFMA[1], an institution that offers large-scale online courses, mostly in the Healthcare area. They include undergraduate, specialization, and training-on-demand courses. All the courses use the Moodle[2] software platform. The dataset comprises three file types:

– *log file*, which provides a timestamped log of every student's interaction with the system, *e.g.*, viewing course materials, interacting with the forum, or any other activity;

– *grades file*, which provides both the students' grades and gender;

[1] http://www.unasus.ufma.br/
[2] https://moodle.org/

– *schedule file*, describing the course schedule, including the task deadlines.

All user data were anonymized and the students are represented by unique ids in both *logs* and *grades files*.

Table 3.1 shows an overview of the dataset, consisting of 5 specialization courses[3] (in a free translation: Elderly Care, Family Care, Basic Care, Nephrology and Home Care) offered between 2013 and 2017, with a total of 755,869 records. In particular, three courses were offered more than once (Elderly Care, Family Care, Basic Care). Each course is organized in two or three cycles and all cycles are composed of a set of modules. In general, these modules are led by instructors (*i.e.*, teachers and tutors), who provide the content in ebooks (available to download in PDF format) and some modules (1 or 2 per course) make use of video and audio resources. Instructors evaluate the students through their postings related to topic discussions in the forum, assignments submitted and quizzes answered. Such evaluations result in numerical values (from 0 to 10) used as part of the module grade. At the end of each cycle, students take a test in a physical classroom, and this grade is used to calculate the final grade of the modules included in the cycle. As a requirement to conclude the course, the students have to achieve module grades greater than or equal to 7 and write a final paper, whose presentation is also in a physical classroom. Apart from the test and final presentation in the classroom, all other course activities are online.

In order to verify whether the literature results presented in Chapter 2 could be reproduced with our dataset, we first conducted an exploratory data analysis to understand, extract, and organize the meaningful data. For instance, the dataset provides 95 user interaction types (*e.g.*, course view, assignment submission, posting on forum). However, only 64 have records related to student interactions. The others were records of instructor interactions, whose analysis is outside the scope of this work. Therefore, only 64 interaction types were used as features in our analysis. Moreover, we noticed that there are no access records to resources included by teachers (*e.g.*, ebooks, videos and audios), because all of them were accessed through links to other pages outside of the VLE (*e.g.*, YouTube, ebook repository system).

Then, we counted the records of each feature by student, arranging them per course, edition, and cycle. To improve our analyses, we derived features to identify:

[3]In Portuguese: *Saúde da Pessoa Idosa, Saúde da Família, Atenção Básica, Nefrologia* and *Atenção Domiciliar*

Table 3.1: Overview of the dataset.

Course	Edition	Cycle	Modules	Period	Students	Log entries
Elderly Care	1	2	23	11/2013 - 01/2015	206	61,632
	2	2	23	07/2014 - 09/2015	253	74,842
Family Care	1	3	15	12/2013 - 01/2015	179	27,442
	2	3	15	04/2014 - 02/2015	200	30,618
Basic Care	1	3	15	03/2014 - 06/2015	279	42,696
	2	3	15	07/2014 - 08/2015	224	34,275
	3	2	15	09/2014 - 08/2015	332	50,800
	4	2	15	12/2014 - 12/2015	839	128,372
	5	2	15	04/2015 - 02/2016	402	61,512
	6	2	15	04/2015 - 02/2016	112	17,139
	7	2	15	10/2015 - 09/2016	146	28,541
	8	2	15	05/2016 - 07/2017	335	57,289
Nephrology	1	3	12	10/2014 - 03/2016	454	77,642
Home Care	1	2	13	04/2015 - 07/2016	289	63,069

– *students who completed the module*, a categorical attribute (with true or false values) based on the grades file, which does not present grades of students who dropped out;

– *students who completed the course*, a categorical attribute (with true or false values) verifying who completed all modules and has final paper grades;

– *students' results in each module*, a categorical attribute (with true or false values) identifying whether the student has a final module grade greater than or equal to 7;

– *students' results in the course*, a categorical attribute (with true or false values) checking whether the student has all final module grades and final paper grade greater than or equal to 7;

– *number of days each student accessed the modules*, a numerical attribute based on students' access to the course during the cycle period;

– *number of days accessed until the in-class test day*, a numerical attribute based on students' access to the course during the cycle period leading up to a test;

– *student inactivity for three or more continuous weeks in the course*, a categorical attribute (with true or false values).

We also built a dataset with the number of student accesses by day, arranging them per course and edition.

As all features are represented by either categorical or numerical data, statistical methods (*e.g.,* Pearson correlation, Wilcoxon rank test) were used in their analysis and interpretation. The statistical analyses allows us to make inferences about each result presented in the literature on student interactions.

Finally, we also aimed to identify student clusters based on the interaction patterns and develop models to predict both students' performance and drop out behavior. Traditional statistical analyses develop accurate prediction models based on human input in making assumptions about the relationships between variables. Therefore, we used machine learning techniques due to their capabilities to analyze high dimensional log data, of arbitrary form, characterized by both noise and complex non-linear pattern components. The choice of techniques such as K-means, Logistic Regression, Random Forest and Decision Tree used in the development of clustering and prediction models was based on techniques widely used in the literature. In addition, we evaluate our prediction models using as quality measure Precision (Prec.), Recall (Rec.), and F1-Measure (F1).

3.2
Analysis and Results

In this section, we present the analysis procedure and discuss the results. We took into consideration the following students' data:

- – access and interaction patterns;

- – forum usage;

- – performance;

- – course completion.

In general, we correlated access, interaction patterns, and forum usage with Moodle features that recorded student interactions. We define as student performance their result in the in-class test and course completion, if they presented the final paper. Moreover, the analyses used the dataset grouped by cycle (total of 33) and course edition (total of 14).

First, we used a K-means clustering algorithm to analyze whether the students could be clustered based on their interaction logs in the VLE. The elbow method was used to detect the number of student interaction clusters per cycle through computing and plotting the sum of squared errors in order to identify where the marginal gain drops significantly, producing an angle

(elbow) in the graph. In line with existing work (Guo and Reinecke, 2014; Park et al., 2017; Khosravi and Cooper, 2017), we found 3 or 4 clusters (depending on the course, edition, and cycle). We then noticed that, in all clusters, the feature that records course access (*i.e.,* course views) was the most meaningful one to discriminate the clusters. However, there were no significant differences of student performance across the clusters. In other words, the means of test grades were similar in all of them. For instance, Figure 3.1 shows the dispersion of course access and test grades by cluster in the Family Care course, edition 1, cycle 1.
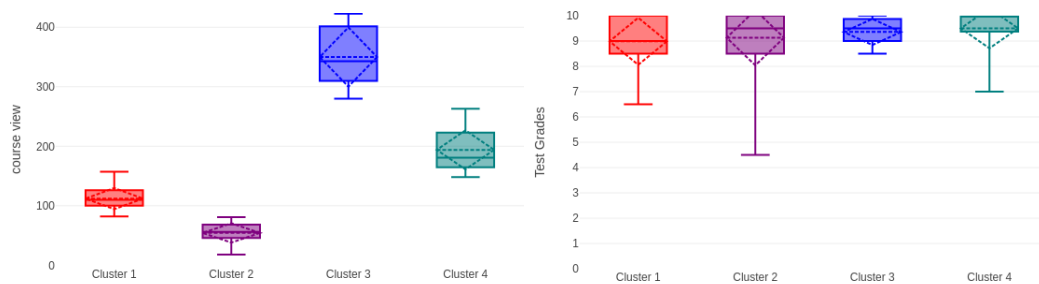

Figure 3.1: Distribution of course views and test grades by cluster in *Saúde da Família* course, edition 1 and cycle 1.

Complementing the previous analysis, we analysed whether the number of page views (*i.e.,* course views) had a significant positive impact on students' final score. First, we removed from the dataset the no-show students in the in-class test. Using Pearson correlation, we did not find a significant correlation (at the significance level $\alpha = .3$). Next, we analyzed the dataset grouped both by course edition and by cycle. As we can see in Table 3.2, only 2 course editions and 13 cycles presented relevant correlation. The majority of the results did not show meaningful correlation, contradicting the result found by Zhang and Zhu (2017).

We noticed that the number of accesses increased in periods close to in-class tests and assignment deadlines, in line with existing work (Nandi et al., 2011; Park et al., 2017; Shi et al., 2017). In particular, for in-class tests, we analyzed a period of 15 days before each test date, whereas for the assignments, we analyzed a period from the date the assignment was available until its deadline. To do this, we used the generalized linear mixed-effects model (GLMM) (Bates et al., 2015) and we fit it to the dataset with the number of student accesses per day (*i.e., number of days accessed until the in-class test day*). Our analysis used the slope to identify the access trend and we defined as increasing trend the positive slopes with significance level of $\alpha = 0.05$. As described in Table 3.3, 67% of the periods preceding assignment deadlines and

Table 3.2: Analysis of whether number of access on VLE had a significant positive impact on the students' final grade.

Course	Edition	p (Edition)	p (Cycle I)	p (Cycle II)	p (Cycle III)
Elderly Care	1	< .001	< .001	< .001	-
	2	.001	**.941**	< .001	-
Family Care	1	.014	.126	**.528**	**.337**
	2	.149	**.621**	.033	.003
Basic Care	1	.099	.196	.137	.015
	2	**.714**	.214	**.620**	.015
	3	.019	.041	.217	-
	4	< .001	.003	< .001	-
	5	**.887**	.213	**.494**	-
	6	.086	**.415**	**.349**	-
	7	< .001	**.381**	**.343**	-
	8	< .001	**.489**	.129	-
Nephrology	1	< .001	< .001	**.390**	**.801**
Home Care	1	.003	.0243	.286	-

82% of periods close to the in-class tests had an increasing trend. In general, the number of accesses increased on average 1.13 and 1.07 per day before the assignments and in-class tests, respectively.

Table 3.3: Trends analysis of access in the VLE in periods preceding in-class tests or assignment deadlines.

Periods close to	Number of periods in which accesses...		
	increased	decreased	showed no trend
assignment deadlines	**289 (67%)**	50 (12%)	91 (21%)
in-class tests	**27 (82%)**	1 (03%)	5 (15%)

Besides, as presented in Table 3.4, we found in only three courses students who were inactive for three or more consecutive weeks (*i.e*, Elderly Care edition 1 and 2, Basic Care edition 2). Most inactive students did not complete the course, so *student inactivity for three or more continuous weeks in the course* can be used as a cue of course drop out, in line with existing work (Chen and Zhang, 2017).

In regard to forum interactions, we analyzed whether the use of forums was related to students' performance. It is worth noting that forum access is required for all forum activities (*e.g.*, add a post, view comments). Therefore, to this analysis, we correlated the students' forum access (*i.e.*, forum discussion views) and test grades. In addition, we removed no-show students in the in-class

Table 3.4: Analysis of inactive students for continuous three weeks or more.

Course	Edition	Dropped out	Concluded
Elderly Care	1	**72**	48
	2	**22**	17
Family Care	1	0	0
	2	0	0
Basic Care	1	0	0
	2	**1**	0
	3	0	0
	4	0	0
	5	0	0
	6	0	0
	7	0	0
	8	0	0
Nephrology	1	0	0
Home Care	1	0	0

test from the dataset, grouped by cycle. Table 3.5 shows that we did not find a significant correlation in most courses, contradicting existing work (Nandi et al., 2011; Carter et al., 2017). We used Pearson correlation and we found a significant correlation (at the significance level $\alpha = .3$) only in the Elderly Care course, albeit a low or moderate correlation, ranging from .32 to .46, depending on the cycle.

Using the dataset grouped by course edition, we analyzed whether students who had more posts were more likely to complete the course (Andres et al., 2018). As illustrated in Table 3.6, we split the students by course completion (*i.e.*, *students who completed the course*) and, through the Wilcoxon rank test and Pearson correlation, we noticed a strong correlation in all courses ($p < .001$). However, in all courses, the forum was used in assignments and as part of the module grade and the students cannot add forum topics. We therefore did not analyze whether students who completed more assignments tended to use forums more (Kizilcec et al., 2013), whether there was correlation between initiating threads and course completion (Andres et al., 2018), and whether the number of forum posts rose in periods close to deadlines (Nandi et al., 2011).

Table 3.7 presents results of the analysis of whether women had more postings than men using the dataset grouped by course edition, keeping in mind that in most courses there were more women than men. Using the Wilcoxon rank test and Pearson correlation, we have not found a significant difference of postings according to gender ($p < .05$): in some courses, on average men posted more than women, and in others the opposite occurred. This result

Table 3.5: Analysis whether student groups that use more forums tend to have a good performance.

Course	Edition	p (Cycle I)	p (Cycle II)	p (Cycle III)
Elderly Care	1	**.391**	**.464**	-
	2	.042	**.328**	-
Family Care	1	.100	.066	< .001
	2	.066	.027	< .001
Basic Care	1	.092	.101	< .001
	2	.096	< .001	< .001
	3	.081	.166	-
	4	.090	.075	-
	5	.047	.058	-
	6	.095	< .001	-
	7	< .001	< .001	-
	8	.009	< .001	-
Nephrology	1	.234	.072	.037
Home Care	1	.020	.130	-

contradicts existing work (Crues et al., 2018).

Using the same dataset and statistical methods of the previous analysis, we analyzed whether there was a correlation between successful students (*i.e.*, who passed in the course using *students' results in course*) and number of accesses ($p < .05$). As shown in Table 3.8, our results showed that successful students had more page views than failing students, in line with existing work (Guo and Reinecke, 2014; Shi et al., 2017; Al-Shabandar et al., 2018; Boulton et al., 2018).

Next, we verified whether there is a correlation between students' success and completion rate of online assignments. To do this, we used the dataset grouped by cycle without no-show students in the in-class test and split the students by whether they passed or failed. We applied the same statistical methods taking into account the students' grades (*i.e.*, *students' results in each module*) and assignment features (*i.e.*, quiz attempts, assignment submissions, assignment submissions for grading, forum posts). As presented in Table 3.9, we noticed that only 6 of the 33 cycles presented a significant correlation ($p < .05$). In most courses, there was no positive correlation between online assignments and pass rates, contradicting existing work (Al-Shabandar et al., 2018).

We also built classification models to predict both students' course

Table 3.6: Analysis whether students who have more posts are more likely to complete the course.

		Number of students who		Mean of posts of who		
Course	Edition	Concluded	Dropped out	Concluded	Dropped out	p
Elderly Care	1	**128**	78	**17.507**	9.730	$<$ **.001**
	2	**132**	121	**11.628**	7.140	$<$ **.001**
Family Care	1	**154**	25	**5.363**	3.440	$<$ **.001**
	2	**178**	22	**6.297**	4.363	$<$ **.001**
Basic Care	1	**233**	46	**8.712**	7.913	$<$ **.001**
	2	**162**	62	**10.932**	7.193	$<$ **.001**
	3	**181**	151	**8.646**	6.158	$<$ **.001**
	4	**710**	129	**8.666**	6.085	$<$ **.001**
	5	**312**	90	**4.512**	2.955	$<$ **.001**
	6	**99**	13	**4.767**	3.230	$<$ **.001**
	7	**104**	42	**12.259**	6.761	$<$ **.001**
	8	**276**	59	**14.423**	10.915	$<$ **.001**
Nephrology	1	**328**	126	**5.472**	3.761	$<$ **.001**
Home Care	1	**146**	143	**10.363**	5.426	$<$ **.001**

Table 3.7: Analysis whether women had more postings than men.

		Number of		Mean of posts		
Course	Edition	Women	Men	Women	Men	p
Elderly Care	1	**174**	32	**22.218**	17.781	.162
	2	**209**	44	15.377	**17.318**	.085
Family Care	1	**111**	68	17.468	**17.602**	.866
	2	**119**	81	**16.042**	14.617	**.011**
Basic Care	1	**168**	111	**20.041**	16.891	**.013**
	2	**113**	111	**22.283**	19.630	.160
	3	**209**	123	**13.598**	12.934	.123
	4	**553**	286	**18.488**	16.045	**.001**
	5	197	**205**	**16.446**	12.829	$<$ **.001**
	6	**61**	51	**17.032**	15.196	.150
	7	**81**	65	15.296	**16.046**	.705
	8	**202**	133	**17.519**	15.909	**.016**
Nephrology	1	**374**	80	**16.540**	16.037	.831
Home Care	1	**234**	55	11.931	**12.000**	.916

completion and performance. In general, the dataset used was grouped by course, and by course and cycle. However, we grouped the Basic Care course

Table 3.8: Analysis of correlation between number of accesses and student success (passed/failed).

Course	Edition	Passed	Failed	p
Elderly Care	1	**124**	82	< .001
	2	**130**	123	< .001
Family Care	1	**153**	26	< .001
	2	**177**	23	.024
Basic Care	1	**222**	57	< .001
	2	**154**	70	< .001
	3	158	**174**	< .001
	4	**638**	201	< .001
	5	199	**203**	< .001
	6	**71**	41	< .001
	7	**91**	55	< .001
	8	**252**	83	< .001
Nephrology	1	**321**	133	< .001
Home Care	1	123	**166**	.009

Table 3.9: Analysis whether there is correlation between students' success and completion rate of online assignments.

Course	Edition	p (Cycle I)	p (Cycle II)	p (Cycle III)
Elderly Care	1	**< .001**	**< .001**	-
	2	.161	**< .001**	-
Family Care	1	**.038**	.429	.650
	2	.916	.311	.891
Basic Care	1	.153	.888	.270
	2	.907	.324	.239
	3	.279	.385	-
	4	.486	.381	-
	5	**.047**	.458	-
	6	*	.810	-
	7	.246	.226	-
	8	.498	.531	-
Nephrology	1	**.001**	.148	.218
Home Care	1	.956	.670	-

*All the students have passed in Basic Care, edition 6 and cycle 1.

into 5 distinct groups, because it had been offered in 5 distinct ways:

1. editions 1 and 2 had 3 cycles, whereas the others had 2;

2. editions 3 and 4 were designed with 6.5 modules in the first cycle and 8.5 in the second cycle;

3. editions 5 and 6 had 4.5 modules in the cycle 1, 7.5 in the cycle 2, and 3 independent modules (outside of cycles);

4. edition 7 had the same number of modules (total of 7.5) in both cycles;

5. in edition 8, the cycles comprised 8.5 and 6.5 modules respectively.

We split our data per course into 80% to train and 20% to test.

As mentioned in Section 3.1, we created three models for the prediction tasks: Logistic Regression, Random Forest, and Decision Tree. In order to calibrate our models, we used a 5-fold cross-validation over the training set. Table 3.10 presents the quality of the best model for the proposed prediction tasks. Each column presents the best model (on average) for each task.

Table 3.10: Mean of the best model results to predict course completion using: (i) all features, (ii) features related to assignment submissions and quiz attempts, (iii) data accesses, and (iv) features related to forum. (v) Results of the prediction model of students' performance using all features.

Method	(i) Random Forest	(ii) Decision Tree	(iii) Decision Tree	(iv) Decision Tree	(v) Decision Tree
Prec.(%)	84.58	81.93	83.18	74.65	37.23
Rec.(%)	79.76	78.33	78.63	71.72	35.13
F1(%)	79.34	78.08	78.28	70.43	32.72

First, we built models to **predict students completing the course**, using the dataset grouped by course with all features. Our Random Forest classifier had $F_{\beta=1} = 79.34\%$ (Table 3.10(i)), in line with some papers that presented accurate classification models to predict student drop out rates through their interaction in the VLE (Shi et al., 2017; Laveti et al., 2017). Next, using only features related to assignment submissions and quiz attempts, we analyzed **whether assignment submissions could be used as a predictor of course completion** (Andres et al., 2018; Al-Shabandar et al., 2018). As presented in Table 3.10(ii), the Decision Tree had $F_{\beta=1} = 78.08\%$, only 1.26% less than the model that used all features (Table 3.10(i)). We also built models using only features related to **data access (*i.e.*, course views) to predict course completion** (Laveti et al., 2017; Al-Shabandar et al., 2018). As we see in Table 3.10(iii), the best performing model was the Decision Tree, with $F_{\beta=1} = 78.28\%$, 0.2% greater than the model using only features related to

assignment submissions and quiz attempts Table 3.10(ii), and 1.06% less than the model that uses all features (Table 3.10(i)). Then, we used all **forum features to predict course completion** (Andres et al., 2018). Table 3.10(iv) shows the best model, Decision Tree, with $F_{\beta=1} = 70.43\%$, a decrease of 8.91% compared to models that use all features (Table 3.10(i)), 7.65% compared to models that use features related to assignment submissions and quiz attempts Table 3.10(ii), and 7.85% compared to models that use features related to data access (Table 3.10(iii)). That shows that using only forum features had a negative impact on model quality. Lastly, we built models to **predict student performance through their interaction in the VLE** (Carter et al., 2017; Shi et al., 2017; Al-Shabandar et al., 2018). To do this, we used the dataset grouped by course and cycle. We segmented the grades into four groups: (i) lower than 7, (ii) between 7 and 7.9, (iii) between 8 and 8.9, and (iv) between 9 and 10. As we see in Table 3.10(v), $F_{\beta=1}$ was strongly impacted, decreasing to 32.72%. However, we could not directly compare this model to the other models, since the prediction objective was different: this model aimed to predict student final score, whilst the other ones aimed to predict student course completion.

3.3
Concluding Remarks

This chapter compared the results found in the literature about the use of VLE logs to identify student behavior patterns and performance with those found in some online courses offered in Brazil in the Healthcare area. We used data exploration, statistical methods, and machine learning techniques.

Our analyses showed evidences that the students could be clustered by their accesses and successful students had more page views than failed students. We also found results showing that the number of accesses increased in periods close to exams and assignment deadlines. In addition, students who were inactive for three or more consecutive weeks could be used as a cue of course drop out. Besides, we noticed that students who had more posts were more likely to complete the course. In regard to the prediction models, we found good results related to course completion. Conversely, we did not find significant differences of student performance across the clusters. In regard to forum interactions, we also did not find correlations with student performance, nor significant differences in postings across gender. Another result was that, in most courses, there was no positive correlation between online assignment submissions and pass rates. Lastly, the models built to predict the student performance did not achieve reasonable results.

4
Instructors' Visualization Preferences

According to Schwendimann et al. (2017), there is gap in the literature about specific visualizations and visual metaphors that address the activities of learning and teaching. In this chapter, we use as basis the uncovered requirements and proposed guidelines presented in Chapter 2 to identify instructors' preferences regarding the visualization of student access and performance in courses using VLEs. To achieve our goal, we elaborated a survey about which visualizations answered efficiently a set of questions that instructors found relevant. The visualizations in the survey were developed to support the analysis of the logs from students' interactions with VLEs and to provide insights to instructors. Then, we asked instructors from Brazilian education institutions to respond to the survey and analyzed the collected data. The main outcomes of the study were the types of visualization to show certain VLE data, the instructors' visualization preferences, and their evaluations of each type of visualization.

This chapter is organized as follows. Section 4.1 shows the design process of the VLE data visualizations investigated. Section 4.2 describes the exploratory study performed with instructors about their visualization preferences. In Section 4.3 we present the results of the study. Lastly, Section 4.4 discusses the results.

4.1
Proposed Visualizations

In previous work, we described 38 requirements and 22 guidelines for tool support for analyzing VLE logs (Damasceno et al., 2019b). The guidelines were compiled from the literature and the requirements resulted from interviews with course instructors. From that work, 3 guidelines and 15 requirements were not related to student access or performance, and were therefore removed from the study reported here.

After selecting the relevant requirements and guidelines, we arranged them in 11 visualization groups (VGs) with common VLE data logs and elaborated questions to be answered for each VG.[1], as shown in Appendix D. For

---

[1]In this chapter, visualization groups are identified in the format VG-99.

instance, VG-05 is related to correlations between student grades and activity logs (*e.g.*, access, assignments completed, forum posts). Table 4.1 presents the visualization groups, their corresponding questions, and visualization task. The task is based on the taxonomy of visualization tasks proposed by Amar et al. (2005).

Table 4.1: Visualization groups and the corresponding questions and task.

VG	Questions
VG-01	Task: Filter and Retrieve Value. Q-01: Which students completed the assignments? Q-02: Which assignments were completed by the students?
VG-02	Task: Filter and Retrieve Value. Q-01: Which students accessed the materials? Q-02: Which materials were most accessed by the students?
VG-03	Task: Filter and Retrieve Value. Q-01: How many student accesses, posts, and likes were there?
VG-04	Task: Determine Range. Q-01: For how long did the students watch each video?
VG-05	Task: Cluster, Compute Derived Value and Correlate. Q-01: What is the relation between students' grades and VLE access? Q-02: What is the relation between students' grades and materials access? Q-03: What is the relation between students' grades and assignments completed? Q-04: What is the relation between students' grades and forum access? Q-05: What is the relation between students' grades and forum posts? Q-06: What is the relation between students' grades and forum replies? Q-07: What is the relation between students' grades and forum threads?
VG-06	Task: Cluster, Compute Derived Value and Correlate. Q-01: What is the relation between students' age and VLE access? Q-02: What is the relation between students' age and forum access? Q-03: What is the relation between students' age and forum posts? Q-04: What is the relation between students' age and forum replies? Q-05: What is the relation between students' age and forum threads?
VG-07	Task: Cluster, Compute Derived Value and Correlate. Q-01: What is the prediction of students' grades and drop out?
VG-08	Task: Cluster, Compute Derived Value, Correlate, Determine Range. Q-01: How many students' accesses were there each day? Q-02: How many students' accesses were there per week?
VG-09	Task: Determine Range. Q-01: What are the statistics of interactions with video (*e.g.* play, seek)?
VG-10	Task: Filter and Retrieve Value. Q-01: Which videos were understood by students?
VG-11	Task: Cluster, Compute Derived Value and Correlate. Q-01: What were the students' navigation patterns on the VLE?

Next, based on the information visualization literature (*e.g.*, Abela (2008); Kirk (2012); Munzner (2014)) and online data visualization cata-logues,[2] we designed visualizations in order to handle the different kinds of data (*e.g.*, numerical, categorical, time series), their cardinality (*e.g.*, 1, 2, ... N), and the task supported by the chart (*e.g.*, compare, correlate) used by each VG.

We developed the visualizations using Plotly,[3] a framework for building data visualization applications in Python, R, and Javascript. In particular, this framework provides interactive mechanisms, such as zoom in, zoom out, pan, select, toggle spike lines, and hover selector. In the Appendix E, Figures E.1 to E.11 show a sample of all types of visualizations designed for each VG (total of 90 charts). These visualizations present different ways to view the same data (*e.g.*, Bar Chart, Box & Whisker Plot, Bubble Chart, Flow Chart, Heatmap, Lollipop, Grouped Bar Chart, Scatterplot, Stacked Area Graph, Stacked Bar Chart, Violin Plot, and Table). Although the data depicted in the figures are fictional, they are in line with actual study results presented in the Chapter 2 (*e.g.*, number of clusters presented in Figure E.5, student access patterns in Figure E.8). We assumed a scenario where all the results uncovered in the Chapter 2 are true. Each figure is assigned to one VG question (*i.e.*, questions shown in blue shown in Table 4.1). Even though different VG questions may use the same type of visualization, they have distinct parameters or axis values. For instance, Figure E.8 presents visualizations related to VG-08's second question (*i.e.*, related to student access per week), whereas VG-08's first question uses the same visualization type, but presents data access per day instead of week.

We did not consider visualizations that encode values either as angles (as in a pie chart) or areas as separate shapes (as in a cartogram) and subrectangles (as in a treemap). According to Cleveland and McGill (1984); Heer and Bostock (2010), we tend to misjudge quantities encoded in such charts. For instance, the size of acute angles tends to be underestimated, and the size of obtuse angles overestimated (Heer and Bostock, 2010). In addition, area-based comparisons are even worse and could be easily misinterpreted or exaggerated (Heer and Bostock, 2010).

4.2
Study Procedure

Although VLE instructors are domain experts, they do not necessarily know how to use information visualizations to answer their questions (Cox

---

[2]For instance: `https://datavizcatalogue.com` and `https://www.data-to-viz.com`

[3] `https://plotly.com/`

et al., 2001). It is necessary to identify which questions they take into account and how those questions can be answered through graphical means. To achieve this goal, we developed an online survey using all the types of chart designed for each VG. This survey was designed to assess what is relevant for our study. It included fields for instructors identify themselves (*e.g.*, name, age, institution) and to answer questions about their experience with VLE, meaningful student information, VLE data logs, data visualization, and their visualization preferences. We also presented all the questions of Table 4.1 asked them which questions they found relevant to them. Next, the survey displayed all the VG charts related to the questions they selected. As illustrated in Figure 4.1, for each chart, a Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree) was presented for the instructors to evaluate how efficiently they believed the chart answers the corresponding question. To break possible ties, the instructors were then asked to select the chart that, in their opinion, better answers the question.
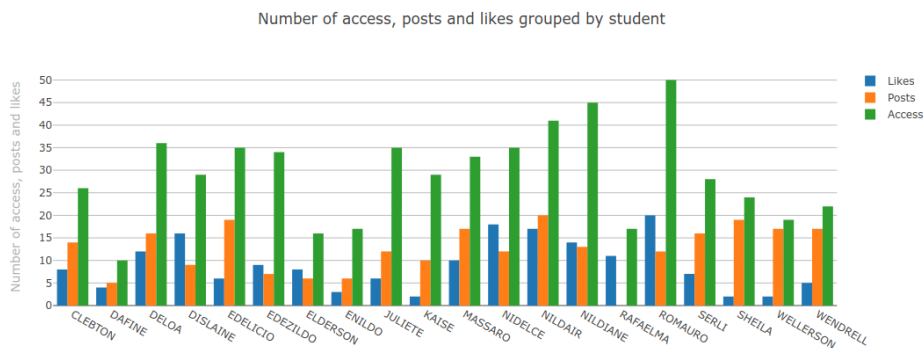
Figure 4.1: A fragment of the online survey.

We highlighted that only questions shown in blue in Table 4.1 were used in this survey. As mentioned in the previous section, the same visualization type is used for all questions related to the same VG. To avoid repetitive visualizations, we assigned the preference of visualization style to the corresponding VG.

4.3
Analysis and Results

The survey was online between December 2018 and January 2019. We sent emails inviting the participation of 88 instructors and 15 universities from different Brazilian regions. The survey was responded by 21 instructors (15 men and 6 women) from institutions located in eight Brazilian states (Amazonas, Ceará, Maranhão, Paraíba, Paraná, Piauí, Rondônia, and São Paulo).

Table 4.2 shows an overview of the instructors' profiles.[4] They all have experience with Virtual Learning Environments: 10 had taught courses that were entirely at a distance, 1 a blended learning course, and 10 both categories. In total, the instructors mentioned having worked with three different VLEs (*i.e.*, Blackboard, Moodle, SIGAA), and Moodle was the most often cited one (by all the instructors). In regard to educational roles performed, most had experience as teacher or tutor (except I-06) and four of them reported experience as coordinator (I-04, I-06, I-09, I-17). Most of the instructors have a computing degree, and two other instructors have programming experience (I-03, I-18).

Table 4.3 shows an overview of the students' age and educational background provided by each instructor. Most of them had taught students who were at least 18 years old. In particular, some instructors reported that their students had various educational backgrounds, without specifying them. However, like the instructors, most of the students had a computing background.

All the instructors claimed to use some VLE communication resource (*i.e.*, chat, forum); most of them reported the use of videos (except I-06, I-09, I-10, I-11, I-18) and 7 mentioned they used e-books (I-02, I-04, I-04, I-11, I-12, I-13, I-16, I-20). They also mentioned other resources, such as text (I-05, I-15, I-19), audio recording (I-15), questionnaires (I-14, I-17), poll (I-19), badges (I-18), and wikis (I-19). Only I-07 mentioned he oversaw student access. Nevertheless, when we asked how they monitor the students and what student data they analyze, most of the instructors reported student access (except I-01, I-06, I-11, I-14, I-16, I-17, I-18, I-20). In addition, they also oversaw the student completion assignments (except I-02, I-06, I-07, I-14, I-16, I-18) and forum usage (except I-04, I-06, I-13, I-14, I-15, I-16, I-17, I-18, I-21).

We asked how the VLE log data were presented to the instructors. Most of them said through tables (I-02, I-04, I-07, I-08, I-11, I-13, I-19, I-21) and charts (I-02, I-03, I-04, I-07, I-08, I-10, I-11, I-13, I-15, I-20), such as Bar Chart (I-11, I-15, I-20), Histogram (I-10), Line Graph (I-07), and Pie Chart (I-02). Some instructors mentioned that the VLE shows reports (I-09, I-14,

---

[4]In this chapter, instructors are identified in the format I-99.

Table 4.2: Overview of the instructors' profiles.

I	Age	Gender	Education area	VLEs	Using VLEs	Mode
I-01	45	M	Philosophy	Moodle	5 years	Both
I-02	52	F	Pedagogy	Moodle	6 years	Both
I-03	37	M	Geography	Moodle	6 months	Distance
I-04	35	F	Odontology	Moodle	10 years	Both
I-05	29	M	Computing	Moodle	18 months	Both
I-06	40	M	Computing	Moodle, SIGAA	2 years	Distance
I-07	35	M	Computing	Moodle	2 years	Distance
I-08	30	M	Computing	Moodle, Blackboard	4 years	Both
I-09	37	M	Computing	Moodle	9 years	Both
I-10	37	F	Computing	Moodle	1 year	Distance
I-11	28	M	Computing	Moodle	14 months	Distance
I-12	30	M	Computing	Moodle	9 months	Distance
I-13	42	F	Industrial Chemistry	Moodle	2 years	Distance
I-14	28	M	Law	Moodle	1 year	Distance
I-15	33	M	Computing	Moodle	9 years	Both
I-16	39	M	Computing	Moodle, SIGAA	5 years	Both
I-17	34	F	Physiotherapy	Moodle	5 years	Distance
I-18	29	M	Design	Moodle	1 year	Blended
I-19	46	F	Computing	Moodle	8 years	Both
I-20	31	M	Computing	Moodle	3 years	Distance
I-21	31	M	Computing	Moodle	2 years	Both

I-21) without specifying their presentation. In particular, I-06 said that the VLE does not show these data and I-17 claimed to use a resource called course progress (also without specifying the data presentation).

In regard to the periodicity of reading and interpretation of charts, 6 instructors said to realize this activity more than once per week (I-01, I-05, I-06, I-11, I-15, I-16), 7 once per week (I-03, I-04, I-08, I-09, I-10, I-14, I-20), 3 once per month (I-13, I-17, I-21) and 5 seldom (I-02, I-07, I-12, I-18, I-19), whereas the periodicity of making charts was reported by 4 as once per week (I-03, I-08, I-10, I-11), 8 as once per month (I-01, I-04, I-05, I-06, I-09, I-13, I-16, I-20), 7 as seldom (I-02, I-14, I-15, I-17, I-18, I-19, I-21) and 2 as never (I-07, I-12).

Table 4.4 presents, grouped by VG, the questions that instructors take or would like to take into account. Questions related to VG-01 and VG-03 were the most often chosen by them. Except for VG-06, at least one question of each VG was selected by most instructors. In other words, most instructors do not (or would not) take into account the correlation between students' age and their VLE interactions, even though they had said they taught students at a

Table 4.3: Overview of the students' ages and educational background, per instructor.

I	Range of age	Education area
I-01	23 to 65	Miscellaneous
I-02	18 to 60	Education and Health
I-03	30 to 55	Public administration
I-04	23 to 60	Health
I-05	18 to 50	Computing
I-06	18 to 60	Computing
I-07	18 to 28	Computing
I-08	25 to 60	Agronomy, Computing, Economy, Forest Engineer and Mathematics
I-09	13 to 45	Computing
I-10	18 to 50	Computing
I-11	20 to 30	Does not know
I-12	25 to 40	Computing
I-13	18 to 50	Agronomy, Business administration, Chemistry, Computing and Health
I-14	25 to 45	Miscellaneous
I-15	18 to 60	Computing
I-16	18 to 30	Computing and Health
I-17	25 to 60	Health
I-18	18 to 60	Miscellaneous
I-19	15 to 50	Computing
I-20	18 to 50	Business administration and Computing
I-21	22 to 60	Computing

wide range of ages. We also asked what other questions they take into account and how they prefer that such questions be answered. In the first question, only three were mentioned: (i) why the students did not access the VLE, (ii) why the students did not meet the assignment deadline, and (iii) what the relation of both students grades and interaction in in-presence classes is. In regard to the second question, almost all the instructors (except I-12, I-16, I-18, I-19) said through tables (I-01, I-02, I-08, I-09, I-11, I-13, I-17, I-21) and charts (I-01, I-03, I-04, I-05, I-06, I-07, I-08, I-09, I-10, I-11, I-13, I-14, I-15, I-17, I-20, I-21), such as Bar Chart (I-01, I-07, I-11, I-13, I-14), Line Graph (I-07), and Pie Chart (I-15, I-21). I-16 and I-19 did not answer, I-12 said what he wanted to see without specifying how and I-18 mentioned that he wanted to receive by e-mail. Moreover, Table 4.4 shows the type of visualization most often selected by them. Most instructors chose Table (VG-02, VG-05, VG-06, VG-08, VG-11) or Grouped Bar Chart (VG-01, VG-03, VG-04, VG-10) for presenting data. Only in the case of VG-07 (Scatterplot) and VG-09 (Stacked Area Graph) another way of data presentation was selected by most of them.

Table 4.4: Overview of both the questions chosen and visualization style more selected by the instructors.

I	VG-01 Q-01	VG-01 Q-02	VG-02 Q-01	VG-02 Q-02	VG-03 Q-01	VG-04 Q-01	VG-05 Q-01	Q-02	Q-03	Q-04	Q-05	Q-06	Q-07	VG-06 Q-01	Q-02	Q-03	Q-04	Q-05	VG-07 Q-01	VG-08 Q-01	Q-02	VG-09 Q-01	VG-10 Q-01	VG-11 Q-01
I-01																								
I-02																								
I-03																								
I-04																								
I-05																								
I-06																								
I-07																								
I-08																								
I-09																								
I-10																								
I-11																								
I-12																								
I-13																								
I-14																								
I-15																								
I-16																								
I-17																								
I-18																								
I-19																								
I-20																								
I-21																								
Total	20	17	14	15	17	14	11	15	13	13	12	13	6	5	5	8	8	7	13	8	15	12	11	16
Selected	j		a		b	h	a							a					b	a		c	k	a

Figure 4.2 shows the instructors' evaluation of each VG chart about whether they answer the VG question efficiently, such as Figure 4.1. We considered good evaluations as those whose total count of agreeing responses is greater than 50% of the responses, whereas bad evaluations are those whose total count of disagreeing responses is greater than 50%. Only in 5 VGs (VG-02, VG-05, VG-08, VG-09, VG-10) the charts with the best evaluations were the same as the selected ones. However, the other charts were among the 3 best evaluated. In line with the instructors' visualization preferences, data presented in Tables, Bar Chart, Grouped Bar Chart, Stacked Bar Chart, Lollipop, and Stacked Area Chart received good evaluations, whereas the use of Bubble Chart, Flow Chart, and Heatmap received poor evaluations. In general, the chart used to show both data distribution and variation, Box & Whisker Plot, received more good evaluations than bad ones. Another interesting point is that all Heatmap charts with good evaluations had the corresponding numeric data values presented in each cell.
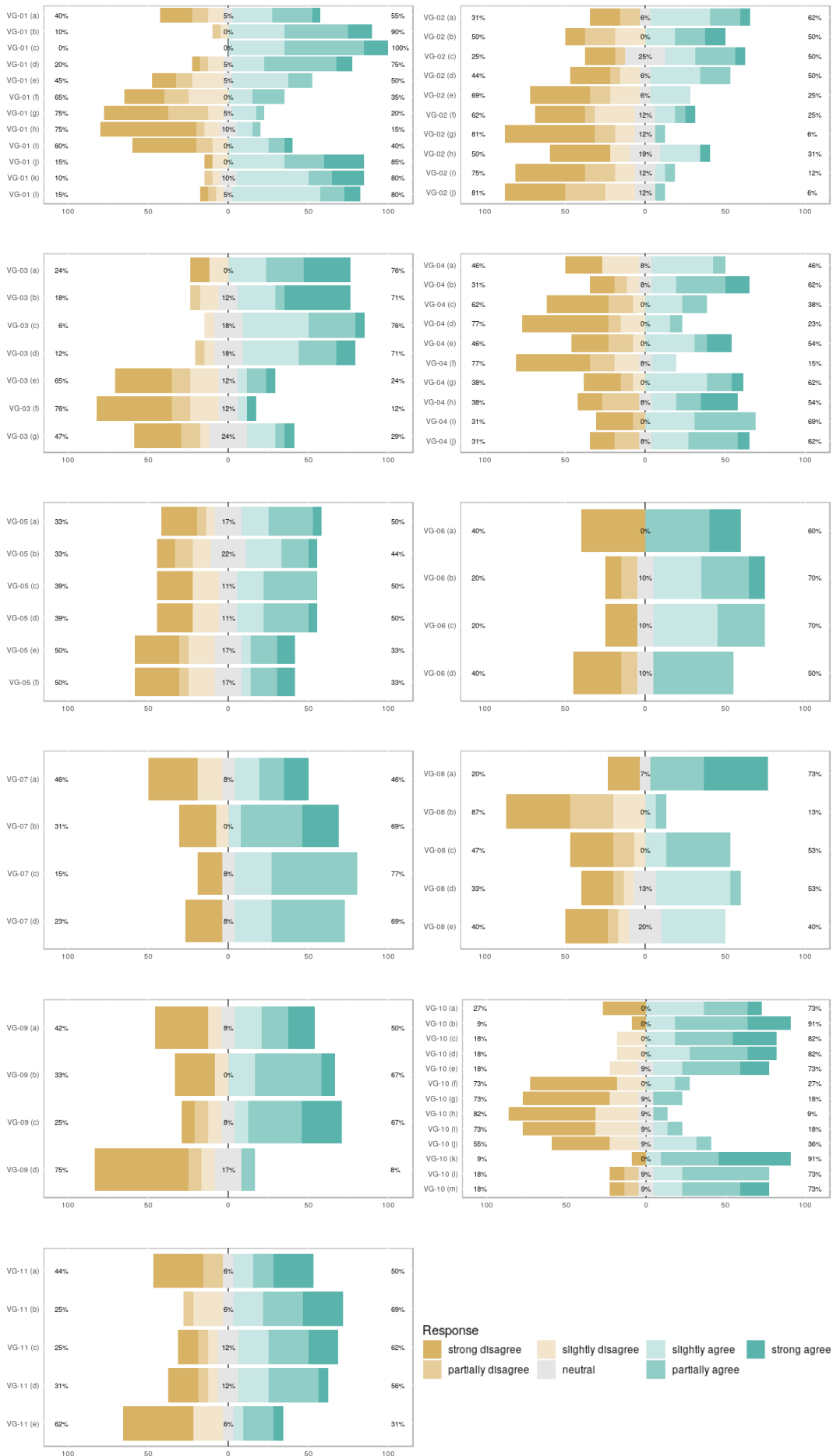
Figure 4.2: Results of the evaluation of each group of visualizations.

4.4
Concluding Remarks

Our aim in this chapter was to understand which visualizations better support instructors. We argue in this thesis that understanding the learning process through visualizations can help instructors design better courses and improve learning effectiveness.

The instructors reported that they take into account more data related to forum usage, assignment completion, and student access, whereas data of correlation between students' age and their VLE interactions had less importance for them, even many of them have said that taught students with ages from 18 to 50. Moreover, data presented in Tables, Bar Chart, Grouped Bar Chart, Stacked Bar Chart, and Lollipop received good evaluations, whereas the use of Bubble Chart, Flow Chart, Heatmap, and Violin Plot received bad evaluations. We note that even though instructors were presented different types of the visualization, the ones they selected the most and evaluated better are in line with both their preferences mentioned before responding the survey and the charts they already used.

5
A Model for Learning Analytics Dashboards

One way to present visualizations in this domain is through learning dashboards, which Schwendimann et al. (2017) define as a single display that aggregates different indicators about learner(s), learning process(es), and/or learning context(s) into one or multiple visualizations. However, there are reports that such dashboards commonly feature poor interface design and lack of usability testing (Reimers and Neovesky, 2015; Schwendimann et al., 2017). In addition, the data presented in the visualizations do not commonly correspond with what instructors and students really want or need (Bodily and Verbert, 2017), in part, because they are not involved in the design process. This problem is pointed by Bakharia et al. (2016), who present considerations about the alignment of learning design with learning analytics tools as a key issue within the learning analytics community, which requires collective effort.

To bridge this gap, this chapter presents a model to develop dashboards supporting a range of activities, from gathering the data from VLEs to presenting visualizations to provide insights and support instructors in their pedagogical decisions. To achieve our goal, we investigated works of models and frameworks for developing dashboards and similar tools. Based on these works, we propose a model connecting Visual Analytics theories and models with found results presented in the previous chapters, such as requirements and guidelines obtained by interviews with instructors and systematic mapping (Chapter 2), methods used to analyze students' logs from VLEs (Chapter 3), and instructors' visualization preferences (Chapter 4). Then, we instantiated the model in EDUVIS, a tool for instructors to assemble dashboards considering their analyzing goals and visualization preferences. In addition, we asked instructors from Brazilian education institutions to respond to a survey and use EDUVIS. We evaluated our proposal through a web-based evaluation form used by instructors and their feedback about the tool.

This chapter is organized as follows. Section 5.1 presents related works of models and frameworks for developing dashboards and similar tools. Section 5.2 shows our proposed model to build dashboards. Section 5.3 presents EDUVIS. Section 5.4 details our evaluation procedure and results. Finally, Section 5.5 discusses our findings.

5.1
Related Work

Visual Analytics research made great strides over the last years, with several works presenting successes in helping domain experts explore large and complex datasets. Among these works, Sacha et al. (2014) present the "human is the loop" concept and propose a model integrating system and human aspects. As we can see in Figure 5.1, for them, human and machine are a loop in the knowledge generation process using visual analytics, and the exploration loop is the basis in that process. The model is demonstrated undertaking a comparative assessment of four applications (Sacha et al., 2014).
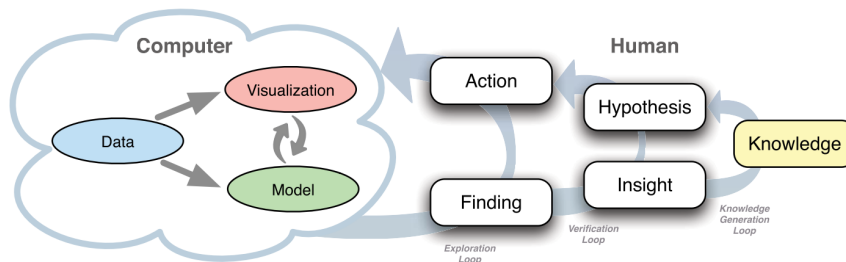
Figure 5.1: Knowledge generation model for visual analytics (Sacha et al., 2014). It can adopted for Visual Learning Analytics, integrating both (i) data processing and presenting flow and (ii) instructors' sensemaking process.

Figure 5.2 presents a model proposed by Zahalka and Worring (2014). According to them, this model was designed to combine:

– literature methods of data visualization, interaction, and analysis;

– methods which express individual intentions of the analyst.
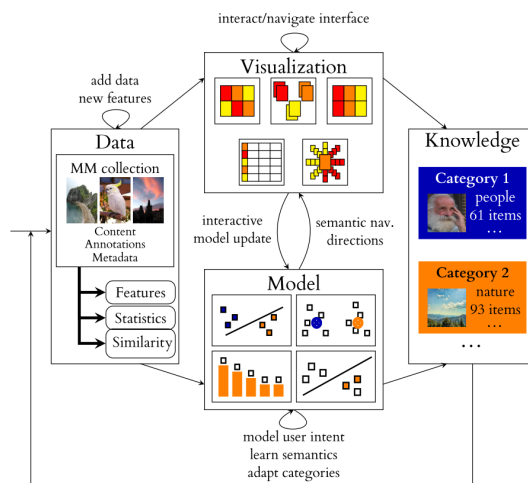
Figure 5.2: The multimedia analytics model expanding upon a diagram process (Zahalka et al., 2015).

In addition, it is instantiated in a recommender system for suggesting places of interest to visit to a user based on their preferences (Zahalka et al., 2015). However, because that model was built to be generic, it presents limitations to support specific domain models, and they do not consider requirements and guidelines such as those uncovered in Chapter 2.

We also highlight some models, frameworks, and guidelines for Visual Analytics, focusing on Learning Analytics (*aka* Visual Learning Analytics (Gómez-Aguilar et al., 2015; Chen et al., 2018; Vieira et al., 2018)), aimed to support instructors about their actual needs and issues. For instance, Chatti et al. (2012) presented a model (Figure 5.3) for learning analytics based on four questions:

1. **What** kind of data does the system gather, manage and use for the analysis?

2. **Why** does the system analyze the collected data?

3. **How** does the system perform the analysis of the collected data?

4. **Who** is targeted by the analysis?
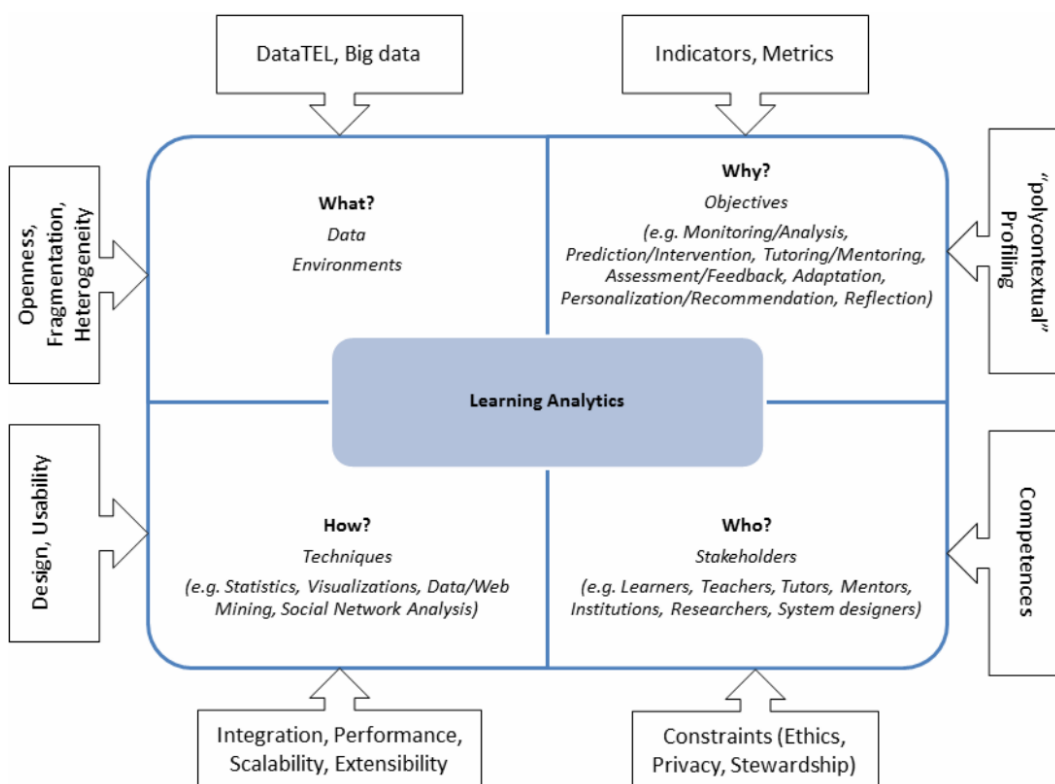
Figure 5.3: Learning analytics model based on four questions (Chatti et al., 2012).

In line with these questions, Klerkx et al. (2017) suggested general considerations and guidelines regarding the design of visual and interaction properties of learning dashboards. Orduña et al. (2014) described what information can be extracted from remote laboratories and how to show them on a dashboard. Fulantelli et al. (2015) proposed a framework to support instructors in the analysis of student interactions in mobile devices, showing indicators for a learning scenario. Ruipérez-Valiente et al. (2015) presented some visualization guidelines to help teachers in making decisions about the learning process and implemented them in a MOOC system.

In particular, some papers show how to use storytelling in Visual Learning Analytics. For instance, Echeverria et al. (2018) proposed a model (illustrated in Figure 5.4) that considers the teacher's pedagogical intentions translated into rules which can be read by a data processing system.
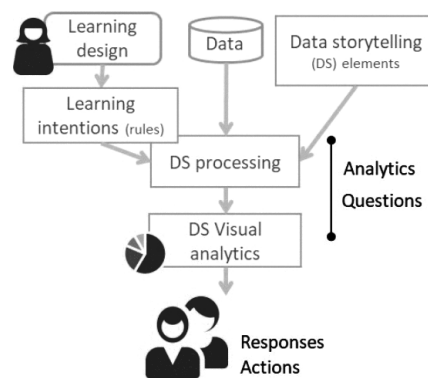
Figure 5.4: A model for Visual Learning Analytics using storytelling concepts (Echeverria et al., 2018).

Moreover, Echeverria et al. (2018) described a dashboard that instantiates this model, uses data from a system that captures students' interactions, and shows their performance through visualizations using storytelling concepts to instructors, taking into account their intentions. Evaluations of their dashboard showed that the instructors could easily interpret data. However, they mentioned that they did not involve the instructors in the design process of dashboards and that further evaluations are needed to understand how teachers might adjust their learning activities based on their insights (Echeverria et al., 2018).

We have found in the literature a clear trend invested into building learning dashboards to improve student learning and performance (Verbert et al., 2014; Bodily and Verbert, 2017; Schwendimann et al., 2017). Complementing that, Mangaroska and Giannakos (2019) present a literature review about Learning Design for Learning Analytics where they claim that future research should consider developing a framework on how to systematically capture and

structure learning design data, and track what learning design choices made by educators influence subsequent learning activities and performances over time. According to them, aligning learning analytics with learning design is an essential condition for creating more meaningful tools, methods, and representations of data for educators and learners (Mangaroska and Giannakos, 2019).

Chen et al. (2016b) introduced PeakVizor, a dashboard to show students' interaction patterns in MOOC videos and views, displaying correlation between those interactions and students' performance. They also proposed glyphs to represent those interactions, but all the users found it difficult to understand its views and, therefore, a usage tutorial was required (Chen et al., 2016b).
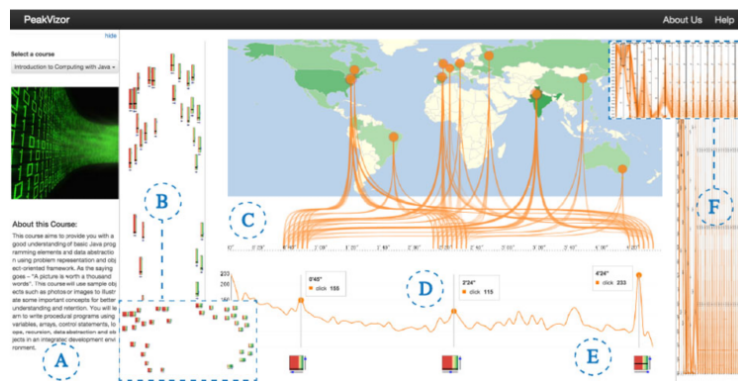
Figure 5.5: PeakVizor's user interface (Chen et al., 2016b).

Fu et al. (2017) presented LAPLE (Figure 5.6), a dashboard for instructors to identify the weaknesses of novice programmers in order to improve teaching materials supporting C education in the classroom through visualizations (*e.g.*, Heatmap, Bar Chart, Line Chart) showing high-frequency errors and those which may need more time to rectify. However, they did not present any evaluation of the dashboard (Fu et al., 2017).
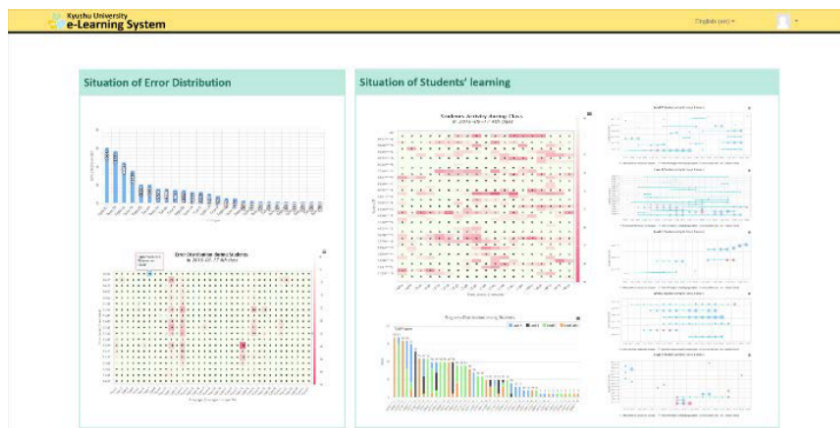
Figure 5.6: LAPLE's user interface (Fu et al., 2017).

Molenaar and Knoop-van Campen (2019) conducted a study to understand how teachers use dashboards in the classroom. They used a dashboard called Snappet (Figure 5.7), where math teachers from primary schools received feedback about students' performance on pre-selected exercises; they could see whether students' answers were correct or incorrect. After each lesson, teachers were asked about both their dashboard consultations and their pedagogical decision made. They concluded that teachers use more dashboards to see individual student performance. In addition, their pedagogical actions were most likely to occur after dashboard consultations (Molenaar and Knoop-van Campen, 2019).
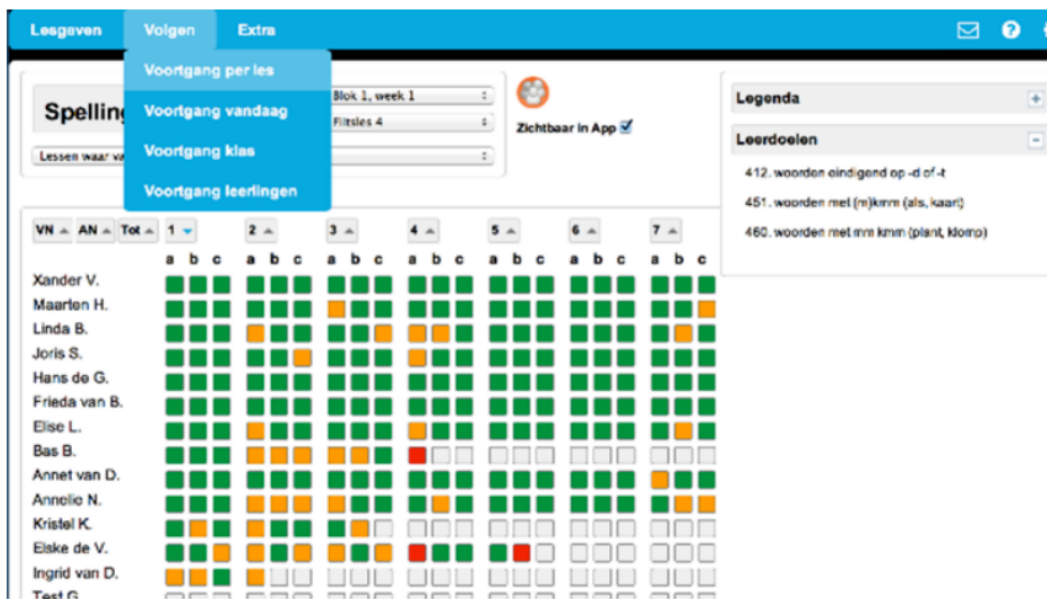
Figure 5.7: Snappet's user interface (Molenaar and Knoop-van Campen, 2019).

Weiand et al. (2019) carried out a study focused on reducing drop out rates through the analysis of VLE logs using some machine learning techniques, such as Decision Tree, K-means, and Naive Bayes. First, they conducted interviews with instructors to gather feedback about a proposed approach for the visualizations of mined data. They identified some requirements that support our findings in Chapter 2. They also presented a dashboard (Figure 5.8) to assist instructors in identifying students at risk (*e.g.*, low performance and drop out). This dashboard was used in a course with 486 students enrolled. As a result, they noted that students who had shown a tendency to drop out or failure, due to low access to resources, were contacted and ended up being approved (Weiand et al., 2019).
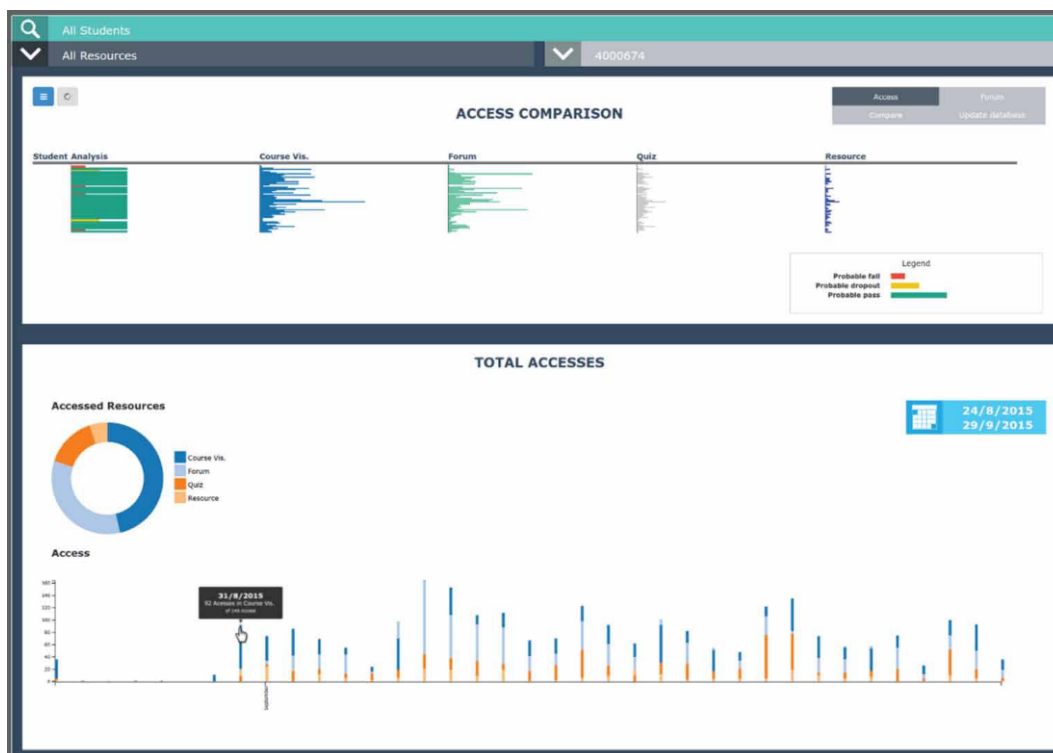
Figure 5.8: Dashboard's user interface proposed by Weiand et al. (2019).

As mentioned previously, there are many works aiming to improve instructors' analyses of student logs. However, the majority of them share problems, such as no evaluation and not involving instructors and stakeholders in the design of the visualizations (Chatti et al., 2012; Reimers and Neovesky, 2015; Schwendimann et al., 2017; Bodily and Verbert, 2017). In addition, the models we have found do not include solutions to fill those gaps in their approaches. Our proposal aims at developing a model that reduces such gaps.

5.2
Proposed Model

To overcome the gaps mentioned previously, we discuss in this section our proposal to extend the analyses of student interactions in Virtual Learning Environments. More precisely, we define a model to be instantiated into dashboards to support instructors in analyzing students' logs. To better detail which works we used as a basis for building our proposal, this section presents each one in the corresponding step we took to build our model.

The first step of our model is based on the work of Echeverria et al. (2018), where the raw data from VLE logs is processed and displayed through visualizations for the instructors. Nevertheless, as shown in Figure 5.9, we have not yet considered using the storytelling concepts described in Figure 5.4. In addition, we highlight that the selected data is processed taking into account

all requirements related to both student access or performance uncovered in Chapter 2, as well as methods described in Chapter 3 to predict the students' performance, evaluate their learning achievement in a course, and even identify behavior patterns.
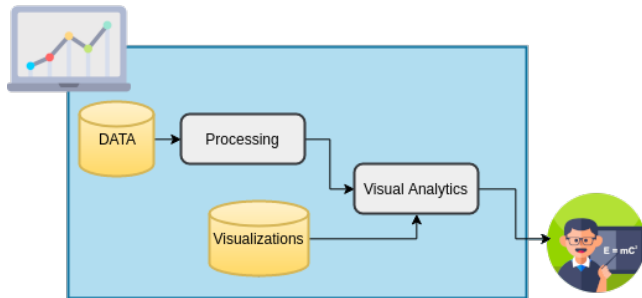
Figure 5.9: First step: raw data processing.

Figure 5.10 depicts the second step of our model, which is based on the work of Sacha et al. (2014) and illustrated in Figure 5.1. Visualizations are presented for instructors to understand students' logs and begin the sensemaking process for knowledge generation. Our goal in this step is empower instructors in this process and keep them in feedback loop to support their pedagogical decisions. To do this, it is important to allow instructors to take actions, such as filtering results, correlating data, setting range, detecting patterns, and so on, to achieve their analytics objectives without the need for an extensive knowledge of the techniques underlying visual analytics tools.
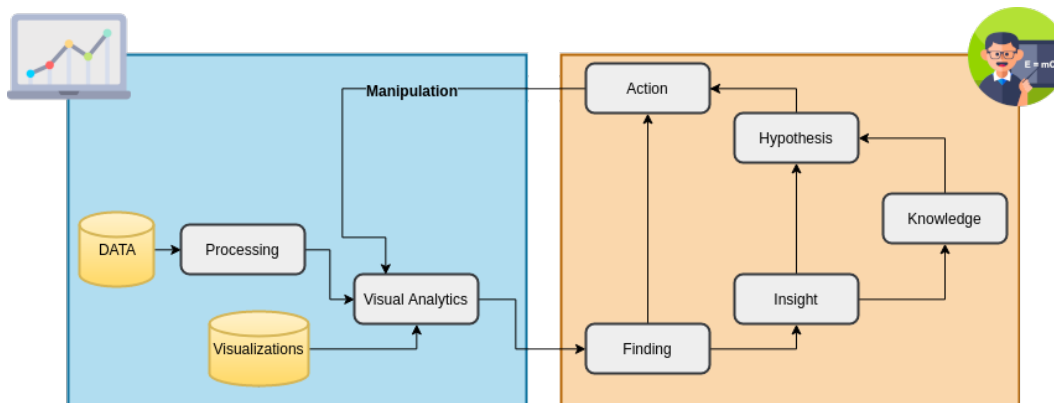
Figure 5.10: Second step: sensemaking process.

The next step is concerned with design of visualizations. We based it on the guidelines reported by Klerkx et al. (2017), which present some questions to drive the visualization design process, such as:

– What is the goal of the visualization? What questions about data should it answer?

- For whom is the visualization intended? Are the people involved specialists in the domain, or in visualization?

- What data will the visualization display? Do these data exhibit a specific internal structure, like time, a hierarchy or a network?

- How will the visualization support the goal? How will people be able to interact with the visualization? What is the intended output device?

In line with them, we argue the visualization designer has a central role to develop representations that answers the questions taking into account issues related to visualization design mentioned in Chapter 4, such as encoding values performed in some charts, as well as instructors' visualization preferences to handle the different kinds of data (*e.g.*, tables, bar chart, heatmap). This step is illustrated in Figure 5.11, which shows a designer building a set of charts and make available in the visualization storage.
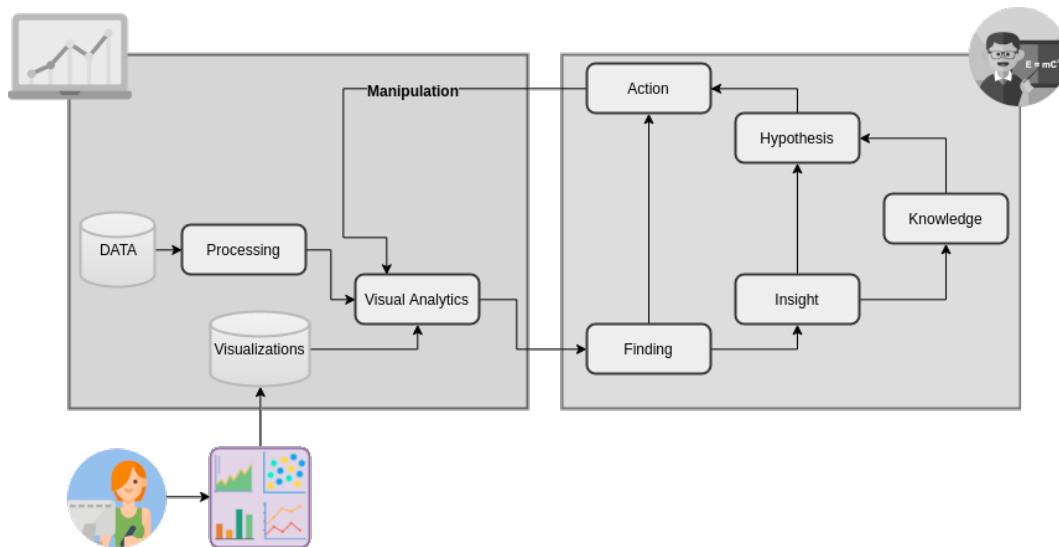
Figure 5.11: Third step: visualization design.

The last step of our model is presented in Figure 5.12. In contrast to Echeverria et al. (2018), the layer of learning intentions is set by instructors and their visualization preferences. In order to improve the instructors' experience, this step allows them to customize dashboards showing only what they want to analyze and how they prefer visualize students' data. However, we highlight that instructors' learning intentions and their visualization preferences are constrained by both requirements uncovered in the first step and visualizations previously made available by the designer.
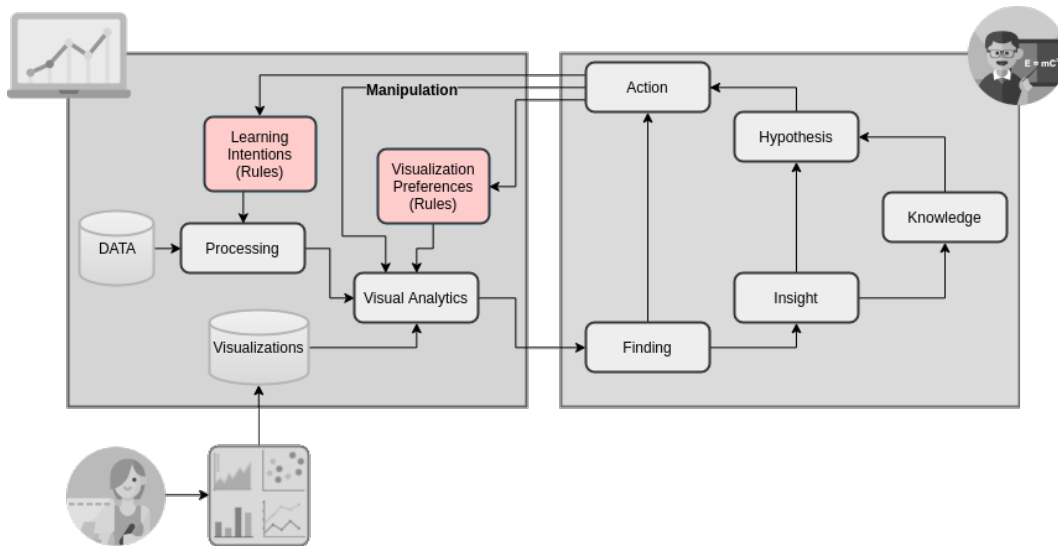
Figure 5.12: Fourth step: customizing instructors' experience.

Figure 5.13 depicts our final model to Visual Learning Analytics, integrating:

– data processing and presenting flow (blue box);

– instructors' sensemaking process (orange box).
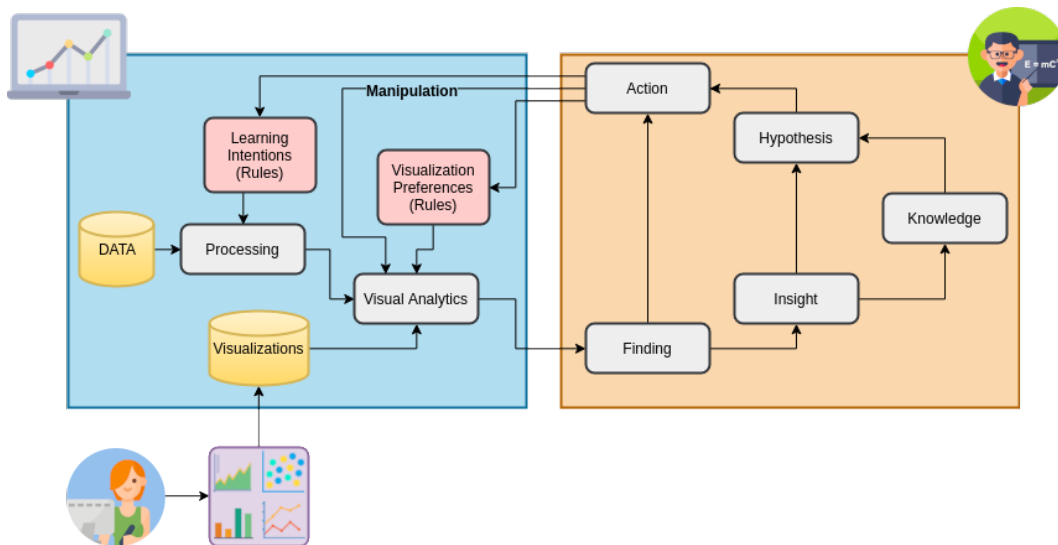
Figure 5.13: A model for Visual Learning Analytics integrating all steps.

As described above, our proposal is based on models and guidelines introduced by Sacha et al. (2014); Klerkx et al. (2017); Echeverria et al. (2018); and findings presented in Chapters 2, 3, and 4. In line with these works, our model aims to support instructors in analyzing student logs from VLE, taking into account the instructor as a central role in generating knowledge to deal efficiently and effectively with a large amount of data and gain new insights.

5.3
Study Procedure

This section presents the procedure we followed to instantiate and evaluate our model described in the previous section. This instance was called EDUVIS, a tool for instructors to assemble dashboards. Then, we developed an online survey to obtain feedback about EDUVIS and evaluate our model.

5.3.1
EDUVIS

We developed EDUVIS[1] as an open source online tool to assemble dashboards based on our proposed model. This tool was built using Python and Javascript. In particular, all the visualizations (total of 141) were designed using Plotly, which provides interactivity, such as zoom in, zoom out, pan, select, toggle spike lines, and mouse hover.

In general, the interface of EDUVIS is composed of a menu and a visualization area. The menu presents two kinds of options: dashboard and visualization groups. In particular, EDUVIS has 11 visualization groups options based on the same VG topics presented in Chapter 4 and detailed on Table 5.1

In addition, we organized the 90 visualizations in the same structure used in the previous chapter to arrange common VLE data logs. Each group presents the same information in different visualizations, such as Arc Diagram, Bar Chart, Box & Whisker Plot, Bubble Chart, Flow Chart, Heatmap, Line Chart, Lollipop, Grouped Bar Chart, Scatterplot, Stacked Area Graph, Stacked Bar Chart, Violin Plot, and Table. We also added more charts in EDUVIS based on the same requirements and guidelines uncovered in Chapter 2, and we designed them using the same considerations followed in Chapter 4. For instance, Figure 5.14 shows the page of a visualization group (Student navigation), which is displayed in different ways to view and analyze students' navigation. Each chart is in a card showing both the visualization group title and a selector option, which allows a instructor to add (or remove) the chart to (from) the dashboard. This mechanism enables instructors to assemble dashboards adapted to their needs. In particular, each number presented in the menu indicates the number of charts included in the dashboard from that visualization group.

Figure 5.15 presents the EDUVIS's dashboard user interface, where the charts selected by instructors are displayed. It is worth highlighting that the layout is almost the same of a visualization group page. The difference is that, instead of a selector option, there are move buttons (top, up, down, bottom),

---

[1]An trial version is available at: http://eduvis.pythonanywhere.com/eduvis/dashboard/

Table 5.1: Visualization groups and the questions related to them.

VG	Topics
VG-01	Menu label: Assignments completion. T-01: Students who completed the assignments. T-02: Assignments completed by students.
VG-02	Menu label: Materials accessed. T-01: Students' accesses to materials (*e.g.*, videos, ebooks). T-02: Materials most accessed by the students (*e.g.*, videos, ebooks).
VG-03	Menu label: Forum usage. T-01: Number of student forum accesses, posts, and likes.
VG-04	Menu label: Video accessed. T-01: How long the students are watching each video.
VG-05	Menu label: Student clusters. T-01: Relation between students' grades and VLE access. T-02: Relation between students' grades and materials access. T-03: Relation between students' grades and assignments completed. T-04: Relation between students' grades and forum access. T-05: Relation between students' grades and forum posts. T-06: Relation between students' grades and forum replies. T-07: Relation between students' grades and forum threads.
VG-06	Menu label: Student profiles. T-01: Relation between students' age and VLE access. T-02: Relation between students' age and forum access. T-03: Relation between students' age and forum posts. T-04: Relation between students' age and forum replies. T-05: Relation between students' age and forum threads.
VG-07	Menu label: Performance prediction. T-01: Prediction of students' grades and drop out.
VG-08	Menu label: Student access. T-01: Number of students' accesses per day. T-02: Number of students' accesses per week.
VG-09	Menu label: Video interaction. T-01: Statistics of interactions with video (*e.g.* play, pause, drop out).
VG-10	Menu label: Video understood. T-01: Videos were understood by students.
VG-11	Menu label: Student navigation. T-01: Students' navigation patterns on the VLE.

a settings button and a remove button. The move buttons allow instructors to change the charts' order on the dashboard according to their preferences, the remove button deletes a chart from the dashboard, and the settings button opens the chart's visualization group page.

EDUVIS enables instructors to assemble dashboards based on their learning intentions and visualization preferences, constrained only by the data available and charts designed. However, this problem might be overcome by adding more visualization groups or additional charts.

Figure 5.14: Selecting a chart for the dashboard.
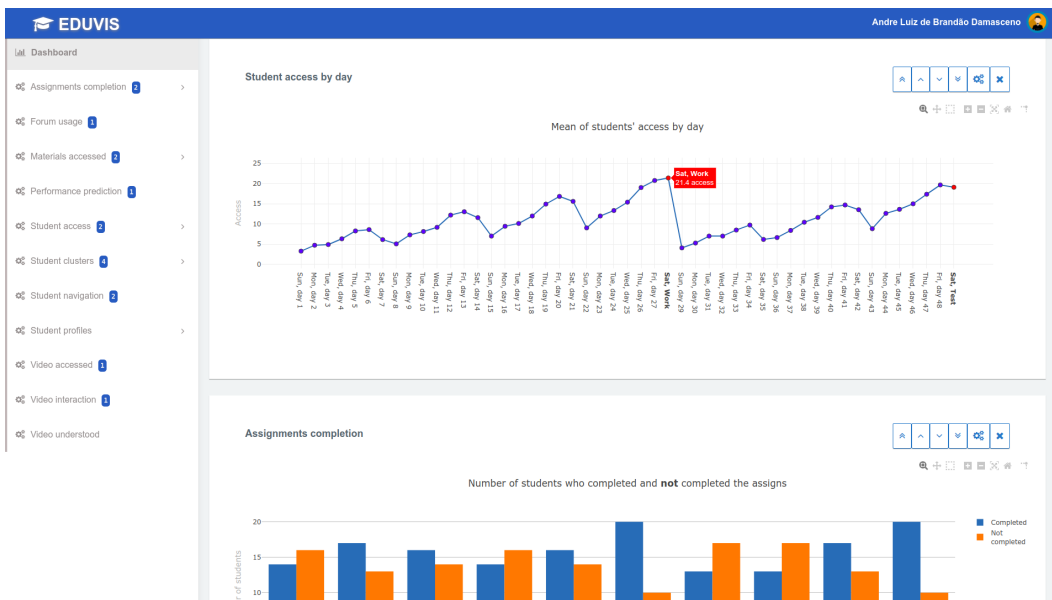
Figure 5.15: Eduvis' dashboard user interface.

5.3.2
Online Survey

To obtain feedback about EDUVIS and evaluate our model, we developed an online survey. This survey was designed to assess just what is relevant for our study. In Appendix F, Figures F.1 to F.12 show a sample of all screens used in this survey. It begins like the online survey presented in the Chapter 4, asking instructors to identify themselves (*e.g.*, name, age, institution), answer questions about their experience with VLE, meaningful student information, and data visualization.

Our online survey displayed a static dashboard with 13 charts, assembled based on results presented in Table 4.4. In other words, we selected charts that, in the opinion of the instructors we had interviewed, better answer the most relevant questions chosen by them:

1. **VG-01(Q-01)**: Which students completed the assignments? - using Grouped Bar Chart;

2. **VG-01(Q-02)**: Which assignments were completed by the students? - using Grouped Bar Chart;

3. **VG-02(Q-01)**: Which students accessed the materials? - using Table;

4. **VG-02(Q-02)**: Which materials were most accessed by the students? - using Grouped Bar Chart;

5. **VG-03(Q-01)**: How many student accesses, posts, and likes were there? - using Grouped Bar Chart;

6. **VG-04(Q-01)**: For how long did the students watch each video? - using Grouped Bar Chart;

7. **VG-05(Q-02)**: What is the relation between students' grades and materials access? - using Table;

8. **VG-05(Q-03)**: What is the relation between students' grades and assignments completed? - using Table;

9. **VG-05(Q-04)**: What is the relation between students' grades and forum access? - using Table;

10. **VG-05(Q-06)**: What is the relation between students' grades and forum replies? - using Table;

11. **VG-07(Q-01)**: What is the prediction of students' grades and drop out? - using Scatterplot;

12. **VG-08(Q-02)**: How many students' accesses were there per week? - using Table;

13. **VG-11(Q-11)**: What were the students' navigation patterns on the VLE? - using Table;

Next, in order to familiarize the instructors with the charts and data, we asked them to browse the static dashboard. Next, as illustrated in Figure F.7,

for each chart, we asked instructors to evaluate the importance of the chart using a Likert scale ranging from 1 (irrelevant) to 7 (relevant), and to provide two facts that were possible to obtain from the chart. To avoid the influence of the chart sequence on the instructors' evaluations, we ordered them randomly for each instructor.

We asked the instructors to watch a video presenting EDUVIS.[2] Then, they could interact with the tool. EDUVIS begins by presenting, as a default dashboard, the same charts from the static dashboard. The instructors may remove these charts, add others, and make the changes they deem necessary to assemble dashboards that better suit their visualization preferences and learning intentions. Then, like in the evaluation of the static dashboards, for each chosen chart in EDUVIS that had not been evaluated previously (in the static dashboard part of the survey), a Likert scale ranging from 1 (irrelevant) to 7 (relevant) was presented for the instructors to answer how important is the chart, and the instructors were asked to provide two facts that are possible to obtain from the chart. In addition, in case they had removed one or more charts from the default dashboard, we also asked them to provide feedback about why they removed charts from the default dashboard. Figures F.8, F.9, and F.10 presents these screens, in sequence.

All data depicted in both the static dashboard and EDUVIS in the study were fictional. They were in line with actual study results presented in the Chapter 2 (*e.g.*, number of clusters presented in Figure E.5, student access patterns in Figure E.8). In other words, in contrast to our findings in Chapter 3, we assumed a scenario where all the results uncovered in Chapter 2 are true.

Finally, as depicted in Figure F.11, we finish our online survey with a section based on the Technology Acceptance Model (TAM) (Gefen and Keil, 1998), a questionnaire widely used in the literature to evaluate interactive systems (Hornbæk and Hertzum, 2017). As we asked instructors to perform tasks using EDUVIS, we used TAM to complement our evaluation and to obtain feedback on the instructors' acceptance of our tool, based mainly on its perceived usefulness and its perceived ease of use.

5.4
Analysis and Results

Between March and April 2020, we deployed our online survey with 30 instructors (24 men and 6 women) from institutions located in 9 states (Amazonas, Ceará, Goiás, Maranhão, Minas Gerais, Pará, Piauí, Rio de Janeiro, and São Paulo). Before they started the survey, we invited them to

[2]Available at: `https://youtu.be/7F_tZ0gZ6b8`

video meetings where we presented the study, our goal, and made available a link to access the survey. After they agreed to fill out our survey, we asked them to share their screen to allow us to record and monitor remotely their progress. We also told them we could answer any questions they might have. The majority of them spent about one and a half hour to complete the study.

Table 5.2 shows an overview of the instructors' profiles.[3] Only I-03, I-11 and I-21 had not had experience with Virtual Learning Environments, 3 had taught courses that were entirely at a distance, 7 a blended learning course, and 17 both categories. In addition, 24 instructors said they had experience in VLE for at least one year. In total, the instructors mentioned having worked with 7 different VLEs (*i.e.*, Blackboard, Google Classroom, Maxwell, Moodle, Q-Acadêmico, Socrative, SIGAA), and Moodle was the most often cited one. In particular, 3 instructors said they had worked with an institutional VLE from their school/university. In regard to educational roles performed, most had experience as teacher or tutor (except I-03, I-11, I-21, I-22, I-23, I-26), 4 cited experience as a coordinator or manager (I-01, I-08, I-23, I-24), two other instructors said to have experience as a moderator (I-07) and monitor (I-22), and one mentioned to be a designer of educational content (I-24). Most of the participants have a degree either in Computing, Management or Pedagogy.

Table 5.3 presents an overview of the students' age and educational background provided by each instructor. Most of them had taught students 18 years old or above and most of those students had a computing background. In addition, only two participants were not university professors, I-06 and I-10. The first one is an instructor of a theology course, whereas I-10 teaches in a high school.

In regard to instructors that had experience with VLE, most of them claimed to have used some VLE communication resource such as chat and forum (except I-10, I-16, I-17, I-23, I-25); only I-02 did not report the use of videos, and 11 mentioned they used e-books (I-04, I-05, I-06, I-08, I-09, I-12, I-16, I-17, I-18, I-20, I-26). They also mentioned other resources, such as assignments (I-07, I-14, I-18, I-24, I-26, I-29, I-30), videoconference (I-01, I-02, I-07, I-08, I-13, I-27, I-30), slides (I-06, I-13, I-15, I-17, I-28, I-29, I-30), text (I-09, I-20, I-27, I-29, I-30), questionnaires (I-14, I-16, I-23, I-27, I-29), links (I-07, I-12, I-14, I-30), games (I-10, I-19, I-29), file sharing (I-02, I-09), and badges (I-08, I-24). Some instructors claimed the COVID-19 pandemic made them shift from a blended learning model to entirely at a distance suddenly, as well as increase the use of resources from VLE. Nevertheless, I-14 said that, even when trying to engage his students with different resources, the decrease

---

[3]In this chapter, instructors are identified in the format I-99.

Table 5.2: Overview of the instructors' profiles.

I	Age	Gender	Education area	VLEs	Using VLEs	Mode
I-01	31-40	M	Computing	Moodle, SIGAA	10 years	Both
I-02	31-40	M	Computing	Moodle	1 year	Distance
I-03	41-60	M	Statistics	-	-	-
I-04	31-40	M	Computing	Moodle	10 years	Both
I-05	31-40	M	Computing	Moodle, SIGAA, Q-Acadêmico	7 years	Both
I-06	31-40	M	Physics	Moodle	3 years	Distance
I-07	31-40	F	Computing	Moodle, Institutional tool	13 years	Both
I-08	31-40	M	Computing	Moodle, Blackboard Google Classroom	5 years	Both
I-09	31-40	M	Geoprocessing	Moodle	1 year	Distance
I-10	21-30	F	Pedagogy	Socrative, Institutional tool	4 or 5 months	Blended
I-11	21-30	M	Pedagogy	-	-	-
I-12	41-60	M	Management	Moodle Google Classroom	4 years	Blended
I-13	41-60	M	Computing	Institutional tool	1 year	Both
I-14	31-40	F	Marketing	Moodle, SIGAA	5 years	Blended
I-15	31-40	M	Computing	Moodle	2 years	Both
I-16	41-60	M	Computing	Moodle	7 years	Blended
I-17	21-30	M	Law	Moodle	2 years	Both
I-18	41-60	M	Math	Moodle, Blackboard Maxwell	8 years	Both
I-19	31-40	M	Computing	Moodle, Google Classroom	4 years	Both
I-20	21-30	F	Psychology and Pedagogy	Google Classroom	2 years	Both
I-21	31-40	M	Computing	-	-	-
I-22	31-40	M	Computing	Moodle	1 year	Both
I-23	31-40	M	Safety Engineering	Moodle, SIGAA	4 years	Blended
I-24	41-60	M	Management	Moodle	more than 10 years	Both
I-25	41-60	F	Languages	Moodle	1 month	Blended
I-26	41-60	F	Computing	Moodle	2 years	Both
I-27	21-30	M	Control and Automation Engineering	Moodle, Google Classroom	6 months	Blended
I-28	41-60	M	Design	Moodle	2 years	Both
I-29	41-60	M	Management	Moodle	10 years	Both
I-30	31-40	M	Computing	Moodle	14 years	Both

Table 5.3: Overview of the students' ages and educational background, per instructor.

I	Range of age	Education area
I-01	13 to 56	Computing
I-02	23 on average	Computing and Scientific methodology
I-03	-	-
I-04	17 to 60	Computing
I-05	18 to 25	Computing
I-06	25 to 55	Theology
I-07	17 to 60	Computing
I-08	18 to 60	Computing
I-09	Above 20	Public Management
I-10	6 to 8	Portuguese and Math
I-11	-	-
I-12	17 to 35	Computing, Management
I-13	18 to 35	Computing
I-14	18 to 20	Computing
I-15	20 to 50	Computing
I-16	17 to 23	Computing
I-17	20 to 40	English and Law
I-18	18 to 30	Math
I-19	18 to 48	Computing
I-20	18 to 60	Pedagogy
I-21	-	-
I-22	18 to 22	Computing
I-23	15 to 25	Safety Engineering
I-24	Above 18	Computing, Management
I-25	30 to 45	Portuguese
I-26	20 to 60	Geosciences
I-27	18 to 50	Control and Automation Engineering
I-28	18 to 50	Computing
I-29	18 to 25	Management
I-30	21 to 50 , 18 to 30	Computing

in students' attendance was inevitable. According to I-14, one reason for this is that several students either do not have computers or they have low access to the Internet.

The instructors mentioned 26 attributes about students' interactions that they consider meaningful, and in line with Chapter 2, the most cited were related to background (I-06, I-07, I-12, I-13, I-15, I-24, I-27), number of accesses to the VLE (I-02, I-18, I-20, I-22, I-24, I-29, I-30) and content (I-04, I-17, I-20, I-24, I-26, I-29), assignments completed (I-04, I-17, I-22, I-30), and grades (I-02, I-07, I-24, I-26). In particular, I-29 said to be in favor of gathering all students' data, since that does not compromise their privacy.

Afterward, we presented topics related to students' logs arranged in 11

visualization groups (Appendix D), and we asked for instructors to answer how relevant to them these topics are. As illustrated in Appendix F, Figure F.4, a Likert scale ranging from 1 (irrelevant) to 7 (relevant) was used for each topic. Figure 5.16 shows that almost all the topics were considered relevant by the instructors. Only those related to students' age had more neutral evaluations. In line with Chapter 4, most instructors do not take into account the correlation between students' age and their VLE interactions, even those who said to teach students of a wide range of ages. We also asked them whether there is some topic that they consider relevant beyond those presented. However, the answer most often mentioned by them was none. In particular, I-24 would like to see students' historical navigation to know what extra material they access to answer their doubts. He also said that this information would be relevant to complement the content.
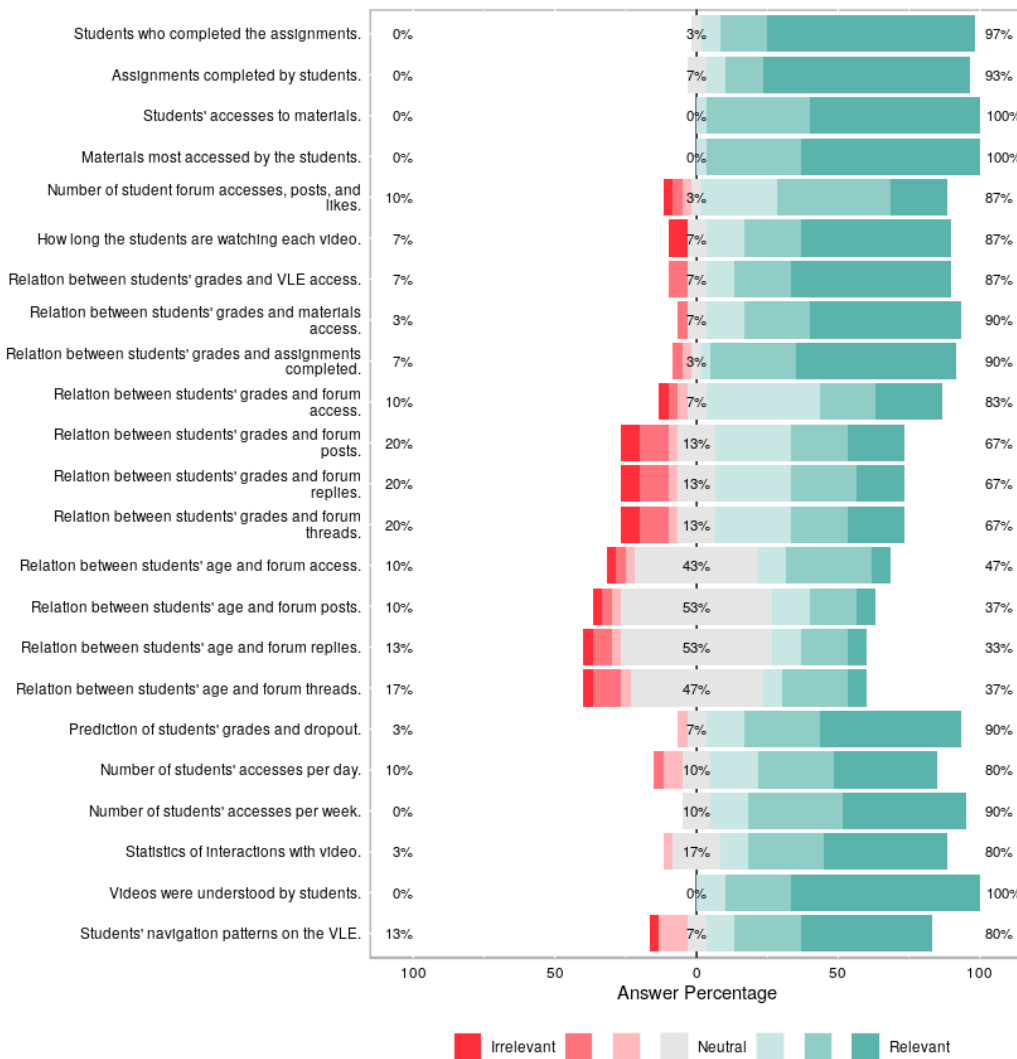
Figure 5.16: Results of the evaluation for each topic from visualization groups.

We also asked instructors how they would like to visualize these VLE log data. Most of them said through unspecified graphical means (I-02, I-03, I-04, I-07, I-09, I-12, I-14, I-15, I-16, I-18, I-21, I-22, I-24, I-26, I-27, I-28, I-30), followed by tables (I-01, I-03, I-07, I-08, I-12, I-13, I-14, I-16, I-22, I-24, I-26, I-28), Bar Chart (I-01, I-05, I-06, I-10, I-11, I-13, I-17, I-19, I-20, I-25), Pie Chart (I-01, I-13, I-20), Scatter plot (I-13), and Dashboard (I-29). Complementing that, some features were mentioned as requirements, such as presenting interactive charts (I-21) and allowing users to export data shown in tables (I-24). In addition, I-02 said that users should be able to choose their visualizations among a range of options. According to him, visualizations have to change according to the information that a user would like to analyze. In line with him, I-14 and I-21 stated that graphics make data analysis easier, but tables are essentials to make deep analyses.

In regard to the periodicity of reading and interpreting charts, 12 instructors said to realize this activity more than once per week (I-02, I-03, I-05, I-08, I-10, I-12, I-13, I-14, I-17, I-21, I-23, I-28), 12 once per week (I-01, I-06, I-09, I-15, I-16, I-18, I-19, I-22, I-24, I-25, I-27, I-30), 3 once per month (I-04, I-11, I-26) and 3 seldom (I-07, I-20, I-29), whereas the periodicity of making charts was reported by 10 as seldom (I-02, I-05, I-06, I-07, I-09, I-10, I-15, I-17, I-20, I-25), by 8 as once per month (I-01, I-04, I-08, I-12, I-14, I-19, I-21, I-28), by 5 more than once per week (I-18, I-22, I-24, I-27, I-30), by 4 as once per week (I-03, I-13, I-16, I-23), and by 3 as never (I-11, I-26, I-29).

Figure 5.17 presents the results related to the instructors' evaluation of the static dashboard. All the charts received more positive feedback. They also gave positive feedback in regard to the charts present interactivity (I-05, I-07, I-25, I-26), which was pointed out as a feature that supports analyses in courses with a large number of students. In addition, sorting a table by column (I-09, I-14, I-15, I-16, I-22, I-29) was the feature most cited by them to improve the dashboard. They also gave feedback about how data was presented. For instance, I-24 said that he is a table's fan and, therefore, he liked the dashboard to show a lot of tables, whereas I-25 would like the dashboard had used more graphical means, as well as I-21, I-29 and I-30, who prefer to make correlation analyses using charts instead of tables. Complementing that, I-21 prefers to see access data through graphics, and I-14 said that a Line Chart is more suitable in this topic. I-22 prefers to view "students who completed the assignments" and "assignments completed by students" through Stacked Bar Chart instead of Bar Chart. It is worth highlighting that these kinds of visualizations are available in EDUVIS. Furthermore, some instructors said that they prefer to analyze certain categories of data. I-07 stated that she

prefers to see an overview of students, whereas I-26 considers more important data related to performance. On average, the instructors extracted about 13 distinct facts obtained for each chart analysis, ranging from 9 mentioned in "Relation between students' grades and forum replies" to 27 in "Students' navigation patterns on the VLE".
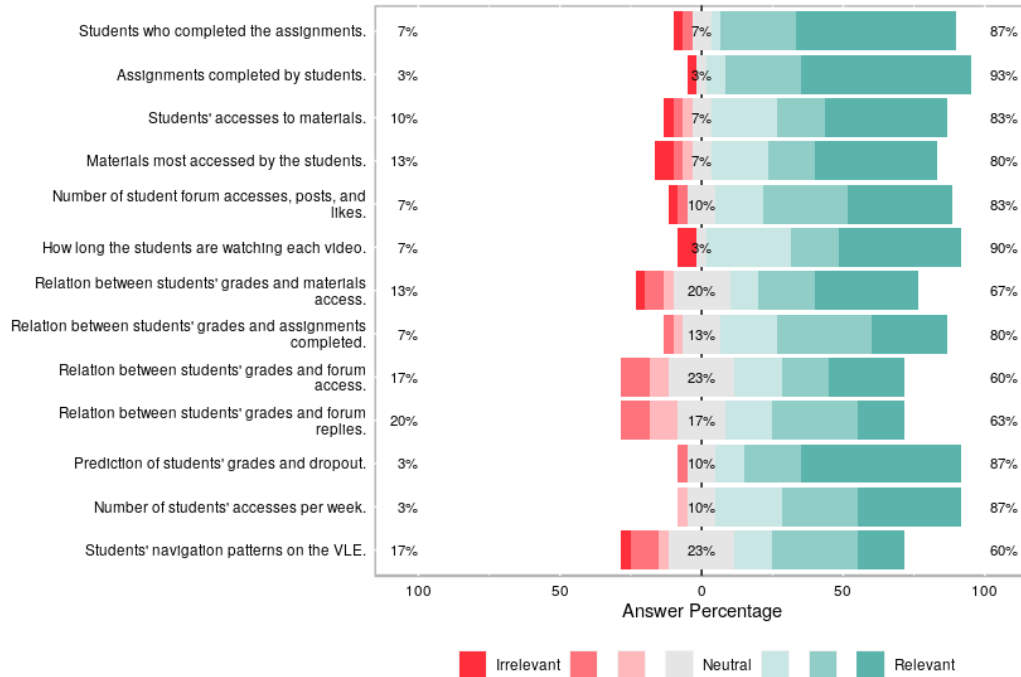
Figure 5.17: Results of the evaluation of the static dashboard.

Table 5.4 shows the number of charts and topics for each VG selected into the instructors' dashboards using EDUVIS. Due to a bug in the implementation, we realized at the end of our study that we did not consider the T-01 topic from VG-06. However, this problem does not significantly affect our goal. The instructors assembled dashboards taking into account topics they would like to analyze and charts based on their visualization preferences. In general, the less chosen topics were the same ones evaluated with relevance lower than 80% in Figure 5.16, such as all those related to students' age (VG-06) and a few attributed to the relation between students' grades (VG-05) and forum posts (T-05), forum replies (T-06) or forum threads (T-07). Some reasons for this were mentioned by instructors: I-23 stated that age is not relevant because his students do not have a big difference of age; I-24 said that he does not like to label students by age because there are several younger students more engaged and self-regulated than older students. It is worth noting that except for VG-06, at least one topic for each VG was selected by most instructors. In line with the results presented in Chapter 4, topics related to assignments completion were the most often chosen ones. According to I-17, this is the topic whose

instructors would use more often. We also realized, in VG-08, that instructors prefer to see more accesses per week than per day. For instance, even though I-22 selected charts from both topics, he said that they would view daily charts related to access per week, whereas access per day he would see seldom.

Table 5.4: Overview of both the number of charts and topics from each VG selected to Eduvis dashboards by instructors.

I	VG-01		VG-02		VG-03	VG-04	VG-05							VG-06					VG-07	VG-08		VG-09	VG-10	VG-11
	T-01	T-02	T-01	T-02	T-01	T-01	T-01	T-02	T-03	T-04	T-05	T-06	T-07	T-01*	T-02	T-03	T-04	T-05	T-01	T-01	T-02	T-01	T-01	T-01
I-01	5	1	1	1	1	1		1	1	1		1							1		1		1	1
I-02		1		1		1		1	1	1	1										1	1		
I-03	1	1		1			1	1	1												2			2
I-04	1	1		1	1	1	1	1			1	1			1				1	1				
I-05	1	1	1					1											1		1	1	1	1
I-06	1	1	2	1	1	1	1	2	1	1					1				1	2	1	1	1	1
I-07	1	1	1	1				2	1	1		1		1					1		1	1	1	
I-08	1	1	1	1		1	1	1						1	1				1	1		1	1	1
I-09	1	1	1		1	1	1	1	1	1	1	1	1	1			1		1				1	1
I-10	1	1	1			1			1	1													1	
I-11	1	1	1	2	2	3	1	1		1	1									1		1	1	1
I-12	1	1	1	1		1		1	1	1		1							1		2	2		
I-13	2	2	1	1	1	1	1	1	1										1	1	2	2	1	2
I-14	2		2	1	1	1	2		2										1		3	1	1	1
I-15	1	1	1		1	1		1	1	1		1							1		1		1	2
I-16	3	1			1	1	3												1		1			
I-17	1	1	1	1		1		1	1										1		1	2		
I-18	1	1		2	1	1	1	1		1									1	1	1	2		1
I-19	1	1	1	1	1	1	1	1	1				1		1	1	1	1	1	1	1	1	2	1
I-20	1	2	1	3	2	3	1	1	1	1	1	1	1		1	1	1	1	2	1	1	2	3	1
I-21		1		2		1		1				1									1	1		
I-22	1	1	1		1	2		1	1										1	1	2	1	1	1
I-23	2	1	3	2	2	2	1	1	1	1		1							2		1	3	2	3
I-24	4	1	2	1	1	1		1	1	1	1	1	1							1	1		3	3
I-25	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1		1		1	1
I-26		1	2		2	1	1								1	1			1		1		2	1
I-27	4	1	1				1	1	1						1				1	2	4	2	3	
I-28	1	1	1	2		1	2	2	2										1		2	1		1
I-29	1	1		1	1	1													1	1	1	1	1	1
I-30	1			1	1	1			1										1		2		1	2
Topics	27	28	23	22	19	26	17	22	21	14	8	12	5	-	8	6	3	4	24	12	25	19	21	21
Charts	42	30	29	29	23	32	21	25	23	14	8	12	5	-	8	6	3	4	26	14	36	27	30	29

*No considered in this study.

We realized that the dashboards were assembled either by exploring all the charts from each topic (*e.g.*, I-02, I-29) or by looking only those from topics deemed relevant (*e.g.*, I-15, I-21). They assembled dashboards with different purposes. I-04 said that his dashboard aims to show mainly forum usage, grade prediction, video accesses, and materials accessed. I-23 reported

that his selected charts are related to both students' access and frequency. I-29 claimed his dashboard is focused on information about the class and only three charts are related to see data by students. In addition, some instructors selected charts in contrast to their visualization preferences. For instance, although I-20 stated to prefer Bar Charts before using EDUVIS, she noted that tables were the most selected visualizations. I-21 also mentioned in the static dashboard that he would like to see some chart presenting patterns of students' navigation flow. However, when he saw this chart in EDUVIS (Appendix E, Figure E.11e), he said that, although this chart is in line with his expectations, it is not a chart that he would put on his dashboard. In contrast, I-24 introduced himself as a table's fan when asked about his visualization preferences, and tables were the most selected visualizations by him. I-15 said that even though he had not selected a chart to analyze video interaction, he would use the Stacked Area Graph to see students' interactions on videos in some situations.

Table 5.4 shows that some instructors selected more than one chart from the same topic into their dashboards. I-21 said that he added both a Bar Chart and a table because the latter provides more information about materials most accessed by the students. I-22 mentioned that he selected both a Box & Whisker Plot and a Heatmap to see the number of students' accesses per week, because the former presents an overview of the class, whereas the latter shows the same information in more detail for each student.

Some instructors reported that they selected more charts than they needed. For instance, I-24 believes he added many charts, but if he had used EDUVIS during a course, his dashboard would present only the suitable charts, and I-25 also mentioned that she would not use all selected charts, but she likes knowing that those charts are available. Complementing that, I-14 said that her visualization preferences might change throughout the semesters. In line with this statement, I-25 said that essential information to one course could be irrelevant to another one because they have different contexts. We also have some instructors stating that they would like to have more than one dashboard. For instance, I-05 said that he would want two dashboards because there are charts that he would use during a course and others they would see at the end of a course as feedback to improve the next courses.

Table 5.5 shows the types of charts most often selected by instructors. In line with the results presented in Chapter 4, Table, Stacked Bar Chart, Grouped Bar Chart, and Bar Chart were the most often presented in instructors' dashboards. In contrast to the previous study, Scatterplot was also one of the most chosen ones. From 141 charts in EDUVIS, 106 were selected at least once, and most instructors selected the following charts:

– Scatterplot from VG-07, T-01 (total of 20), shown in Appendix E, Figure E.7b;

– Grouped Bar Chart from VG-01, T02 (total of 19);

– Table presenting all topics from VG-05 (total of 18), shown in Appendix E, Figure E.5a;

– Area Chart from VG-09, T01 (total of 17), shown in Appendix E, Figure E.9c;

It is worth highlighting that these charts are the same evaluated in Chapter 4 as those that better answer the questions related to their VGs.

Table 5.5: Overview of the most type of charts selected by instructors.

I	Table	Grouped Bar	Scatterplot	Stacked Bar	Bar Chart	Box & Whisker	Heatmap	Flow Chart	Bubble Chart	Area Chart	Line Chart	Violin	Lollipop
I-01	5	3	1		1				2				2
I-02		1	4	1						1	1		1
I-03			3	2	1	1	1	1				1	
I-04	2	4	2			1					1		
I-05	1	1	1	2			1	1	1	1			
I-06	2	3	5	3	1		2			1	1		
I-07	1	3	5	1	1	1				1	1		
I-08	1	3	3	1		3		1	1	1			
I-09		6	11										
I-10	2	3			1								
I-11	8	3		1	1					1			
I-12	3	3	1			1		1		1			
I-13	1	2	1	6	1			2		1	2	2	
I-14	3	3	3	1	1	1	1	1	2	1	1		
I-15	3	4	1	1			1	1					
I-16	1	4	2			1			2			1	
I-17	4	3	1							1			
I-18	2		1	3	1			2	1	1	2		
I-19	5	1	3	1	5	1	1			1			
I-20	14	3	1	1	4								
I-21	2	1	1	1	1					1	1		
I-22		2	3	3		2	1	1	2	1			
I-23	9	5	2		2		2	2		1			
I-24	6	5			3	1	2	1					
I-25			2	2	1	11	2		2		1		
I-26	2	2	2		1	1	2	1	3				
I-27	3				2	4	9			1	1		
I-28	4	1	4	2	1					1	1		
I-29			1	5	1	1		2			1		
I-30		1	2		1		3	1	2			2	
Total	84	70	66	37	31	30	28	18	18	17	14	6	3

After assembling the dashboards, the instructors evaluated the charts they selected that were not in the static dashboard. As mentioned previously, these evaluations were performed using a Likert scale ranging from 1 (irrelevant) to 7 (relevant). From 97 charts and 313 evaluations, only 2 evaluations received negative feedback (lower than 4), and 10 presented a neutral evaluation (equal to 4). Therefore, most evaluations (about 96%) were positive.

Some instructors gave positive feedback to EDUVIS. I-02 and I-24 said the interactivity available in charts allows them to view just what they wanted to analyze. In particular, I-02 and I-21 highlighted the feature of arranging charts' order, because it allows putting the most important charts on the top of the dashboard. I-03 and I-04 liked assembling a customizable dashboard. I-08 would like to use EDUVIS in his work. I-16 said that EDUVIS surprised him due to both the number and the quality of the presented charts, thinking of how it could improve his evaluation of students' performance. I-22 also liked EDUVIS and uncovered some types of charts that he had never seen before. I-23, I-28 and I-29 mentioned that EDUVIS is very useful and they would like it to be a Moodle plugin. In particular, I-23 also said that EDUVIS would help him to manage better his courses and it would reduce his labor.

Some instructors mentioned improvements to EDUVIS. I-02 would like tables to have filters and students' grade average. Although grades are presented in some tables, I-29 and I-30 suggested to have a chart to show only grades. I-21 also suggested a chart presenting whether students understood other materials beyond videos (VG-10). I-16 encouraged us to think of more possibilities of interacting with the charts, such as allowing to sort data and extracting data from a chart to be displayed on another one. In line with him, I-30 also highlighted sorting options as essential in data analysis. I-20 said the colors used by the charts could be customizable as well. I-21 would like to play videos and watch them like some students to gather insights about their experience and identify which parts need to be improved. I-15 and I-21 reported difficulties in arranging charts on the dashboard. In particular, I-15 suggested to improve the usability of EDUVIS, whereas I-21 proposed to enable drag and drop. In addition, I-28 would like to automatically group all charts on the dashboard by topic. I-21 pointed out that most metrics used by charts are focused on quantitative data and suggested us to provide more charts with qualitative metrics, such as students' feedback about why they quit watching a video.

Table 5.6 shows that the majority of instructors removed or replaced almost all the charts from EDUVIS' default dashboard (except T-02 from VG-01, and T-01 from VG-07). When asked why they removed charts from

the default dashboard, most of them said either there was another chart they deemed better (VG-01, VG-02, VG-07, VG-08, VG-11), difficulty of analyzing the information (VG-04, VG-05), or irrelevant information (VG-03).

Table 5.6: Overview of the removed charts from Eduvis' default dashboard by instructors.

I	VG-01 T-01	VG-01 T-02	VG-02 T-01	VG-02 T-02	VG-03 T-01	VG-04 T-01	VG-05 T-02	VG-05 T-03	VG-05 T-04	VG-05 T-06	VG-07 T-01	VG-08 T-02	VG-11 T-01
I-01													
I-02													
I-03													
I-04													
I-05													
I-06													
I-07													
I-08													
I-09													
I-10													
I-11													
I-12													
I-13													
I-14													
I-15													
I-16													
I-17													
I-18													
I-19													
I-20													
I-21													
I-22													
I-23													
I-24													
I-25													
I-26													
I-27													
I-28													
I-29													
I-30													
Total	16	11	20	18	19	16	16	19	21	22	10	16	23

Figure 5.18 depicts the instructors' answers in the questionnaire based on TAM. In general, most answers were positive. We had some negative feedback only in the questions related to configuring a dashboard quickly (I-08, I-13, I-14, I-25), and to whether EDUVIS could make their job easier (I-16). In particular, I-16 said that EDUVIS could not make his job easier because it would encourage him to understand more, he would try to obtain more information, and consequently it would demand more labor and consume more time.
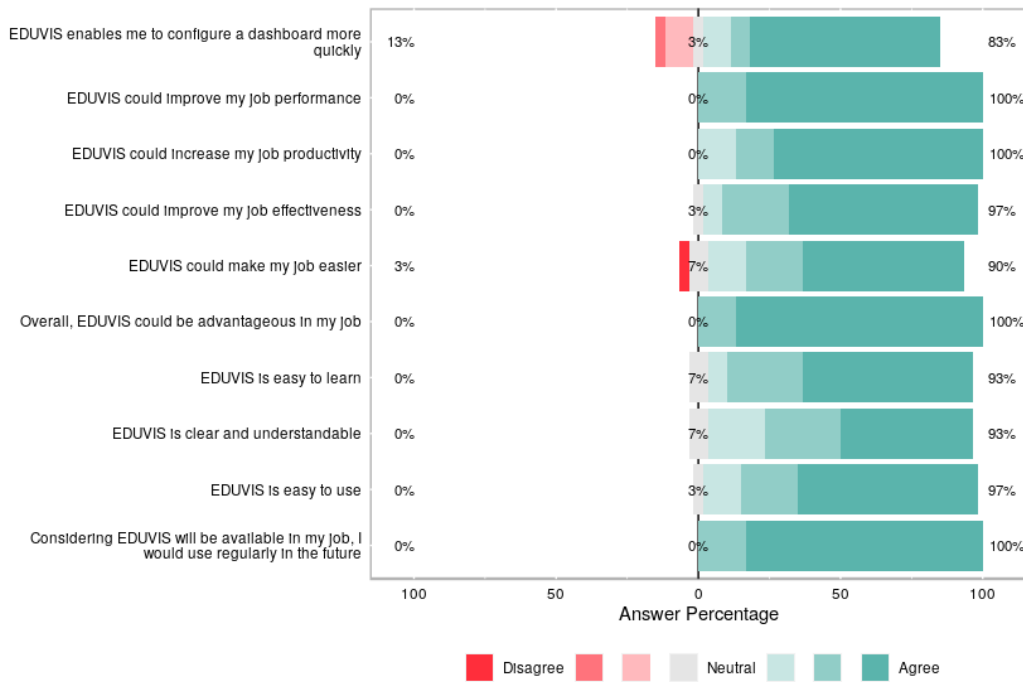
Figure 5.18: Instructors' answers in the questionnaire based on TAM.

5.5
Concluding Remarks

In this chapter, we introduced our model for a learning analytics visualization design process, supporting raw data analysis and effectively deploying learning dashboards. It is worth noting that our model is based on models and guidelines presented by Sacha et al. (2014); Klerkx et al. (2017); Echeverria et al. (2018), and findings presented in Chapters 2, 3 and 4. Then, we instantiated this model in EDUVIS, a tool for instructors to assemble dashboards considering aspects related to their learning intentions and visualization preferences.

We evaluated our proposal through tasks performed by instructors using EDUVIS and an online survey to obtain their feedback. In general, our results show a good acceptance by the instructors. They mentioned positive aspects of our tool and some of them said EDUVIS is very useful and called for releasing it as a Moodle plugin. Regarding topics they take into account to analyze students' logs, most of them do not consider aspects related to students' age; they would like to analyze the relevant topics through graphical means, in line with our findings in Chapter 4. However, dashboards assembled by them presented more visualizations using Tables. A reason for this could be the same one hypothesized in the previous chapter, which is most of them are using to see these data through tables. Their dashboards were assembled with different

intentions and four charts were frequent in most of them.

The instructors gave mostly positive answers in the TAM questions and our results present evidence of instructors' understanding and acceptance. However, we highlight that the evaluation was performed using one instance of our model, and further studies are required to assess more instances, as well as their underlying development process. Some instructors presented concerns related to students' privacy regarding their personal data. For instance, I-21 pointed out the importance of empowering instructors to intervene on the learning process immediately and effectively. Nevertheless, he raised the following question: if students had known that instructors analyze their logs, how would students deal with it? This topic has been indicated as a relevant issue and can guide future research.

6
Conclusions

This chapter presents the final considerations about our work. In the Introduction of this thesis (Section 1.1) we posed the main research question we aimed to answer with this work. Based on this question, we defined sub-questions to lead each one of the subsequent chapters.

Firstly, we shed light on how student interaction logs on VLEs have been analyzed. We led interviews with instructors who work in Brazil and conducted a systematic mapping on EDM and LA. We found evidence that several instructors do not have any learning analytics tool or information besides the student access logs. Normally, they get information from observing what students say and do on VLEs. A few challenges interfere with this analytical process, because most instructors are not statistics experts nor do they receive training to extract key information from VLEs. Therefore, tool support is called for. In addition, we found in the mapping 30 tools to support instructors in analyzing logs. Because Moodle was the most cited VLE by the instructors, we found 37 additional tools to provide information based on student logs. We note that most of the instructors' needs are satisfied by one tool or another. However, none of the existing tools fulfills all requirements raised by the instructors and guidelines presented in the literature.

Next, we aimed to identify which results in the literature were reflected in online courses offered in Brazil. Our analyses showed evidence that the students can be clustered by their accesses and successful students have more page views than failed students. We also found results showing that the number of accesses increased in periods close to exams and assignment deadlines. In addition, students who were inactive for three or more consecutive weeks had a high probability of dropping out. Besides, we noticed that students who had more posts were more likely to complete the course. In regard to the prediction models, we found good results related to course completion. Conversely, we have not found significant differences of student performance across the clusters. Regarding forum interactions, we also have not found correlations with student performance, nor significant differences in postings across gender. Another result was that, in most courses, there was no positive correlation between online assignment submissions and pass rates. Lastly, the

models built to predict the student performance did not achieve good enough results.

Afterward, we elaborated a survey about which visualizations answered efficiently a set of questions that are relevant to instructors. The visualizations in the survey were developed to support the analysis of the logs from students' interactions with VLEs and to provide insights to instructors. We asked for instructors from Brazilian educational institutions to respond to the survey and analyzed the collected data. The main outcome of the study were the types of visualization to show certain VLE data, the instructors' visualization preferences, and their evaluations of each type of visualization. In particular, we note that, even though instructors were presented different types of visualizations, the ones they selected the most and evaluated better are in line with both their preferences mentioned before and the charts they already used and with types of visualizations already integrated into the learning analytic tools found in the literature.

Finally, we proposed a model to instantiate dashboards, based on models and guidelines presented in the literature, as well as instructors' learning intentions and visualization preferences. Then, we developed EDUVIS, an instance of our model to assemble dashboards. We evaluated our proposal through tasks performed with instructors using EDUVIS and an online survey adopting questions based on TAM. In general, we obtained positive feedback of both instructors' understanding and acceptance. Some outcomes are in line with what we have found in previous studies. Besides, we have found evidence of instructors have different learning intentions and visualization preferences, resulting in several dashboards with distinct purposes. However, we highlight that the evaluation was performed using one instance of our model, and further studies are required to assess more instances, as well as their development process.

6.1
Publications

Until the present moment, our work yielded the following publications:

– Chapter 1

  1. Damasceno, A. L. B., Soares Neto, C. S., and Barbosa, S. D. J. (2017). Integrating Participatory and Interaction Design of an Authoring Tool for Learning Objects Involving a Multidisciplinary Team. In Marcus, A. and Wang, W., editors, Design, User Experi-

ence, and Usability: Theory, Methodology, and Management, pages 554–569, Cham. Springer International Publishing.

2. Damasceno, A. L. B., Soares Neto, C. S., and Barbosa, S. D. J. (2017). Lessons Learned from Evaluating an Authoring Tool for Learning Objects. In Zaphiris, P. and Ioannou, A., editors, Learning and Collaboration Technologies. Novel Learning Ecosystems, pages 77–89, Cham. Springer International Publishing.

– Chapter 2

1. Damasceno, A. L. B., Guedes, A. L. V., Colcher, S., Barbosa, S. D. J., Rabêlo, D. J. L.; Monier, E. B. (2018). VideoViz - Uma proposta de ferramenta de visualização de logs de interação em players de vídeo. In Workshop "O Futuro da Videocolaboração" - XXIV

2. Damasceno, A. L. B., Ribeiro, D. S., and Barbosa, S. D. J. (2019). What the Literature and Instructors Say about the Analysis of Student Interaction Logs on Virtual Learning Environments. In 2019 IEEE Frontiers in Education Conference (FIE), pages 1–9. IEEE.

– Chapter 3

1. Damasceno, A. L. B., Almeida, C., Fernandes, W., Lopes, H., and Barbosa, S. D. J. (2019). What Can Be Found from Student Interaction Logs of Online Courses Offered in Brazil. In Anais do XXX Simpósio Brasileiro de Informática na Educação (SBIE 2019), pages 1641 – 1650, Brasilia. Brazilian Computer Society (Sociedade Brasileira de Computação - SBC).

– Chapter 4

1. Damasceno, A. L. B., Ribeiro, D. S., and Barbosa, S. D. J. (2019). Visualizing student interactions to support instructors in Virtual Learning Environments. In Proceedings of the International Conference on Human-Computer Interaction. Springer International Publishing.

6.2
Future Work

An interesting research direction would be to use requirements and guidelines gathered in our research to support other works such as designing more models of LA and building of other Visual Analytics tools. We believe that this could empower instructors with concurrent information about students' abilities, behavior, progress, and performance. This type of research using such tools is fundamental to improve human interaction and enhance students' learning.

We also hypothesize that instructors aware of the data are able to take suitable pedagogical actions. It is important to evaluate whether dashboards indeed impact the way instructors teach and actually support advanced educational effectiveness. Therefore, we intend to assess whether there are changes in students' performance when instructors are able to see information about their behavior and performance, and act accordingly. First, we will compare both the students' drop out rates and grades before and after instructors use the dashboard. In addition, we will periodically ask for instructors, through a questionnaire embedded in the dashboards, what they plan to do and have done to improve the student performance after seeing the reports on the dashboard.

As mentioned previously, another future work would be to develop more instances of our model to obtain more feedback and to assess the development process. Human interaction with visualizations is crucial for better analyzing students' logs. Therefore, we suggest some improvements, such as maintaining preloaded charts and adding recommendations for visualizations. In addition, we see that many improvements could be made to further support human interactions. A possible way to do this is that dashboards can actively learn from the instructors' behavior and adapt its interface, menus, and set of visualizations. Alternatively, it could add a search bar where the user can pose questions for the dashboard to answer them through visualizations according to the instructor's preference. However, for this latter feature, a domain-specific ontology is required.

Visual Analytics tools processes could be collaborative. Multiple instructors and stakeholders may analyze data together, asynchronously or synchronously, to gain more insights. Our model currently assumes an individual's analytic process, so it is lacking collaborative components. We also highlight the possibility of extending our model to support storytelling concepts, for example.

We left for future work another analysis to identify students' study patterns in Brazilian courses through methods of Process Mining, to verify

whether there is a correlation between that behavior and student performance, and to compare behavior patterns and performance between student logs from distance courses with instructors overseeing students' assignments, forum posts and tests, and courses without such tracking (*e.g.* Coursera, EdX, Udemy).

So far our work has focused on empowering and assisting instructors in gaining insights about students' logs. However, the learning process depends on both instructors and students. We also believe in empowering students to make them aware of their behavior and performance. Therefore, another key research is to assist students in gaining insights about their own behavior and performance in a course. A possible approach could be the same as we have used in this work, but considering students as the central focus of research instead of instructors.

Lastly, we recommend more research on ethics, data privacy, and stewardship issues relating to the educational context. These issues are not a novelty; they have been emphasized by many instructors and highlighted by prestigious scientific conferences and magazines as an important research topic (Chatti et al., 2012; Sun et al., 2019).

Bibliography

Abela, A. (2008). *Advanced Presentations by Design: Creating communication that drives action.* John Wiley & Sons.

Al-Shabandar, R., Hussain, A. J., Liatsis, P., and Keight, R. (2018). Analyzing Learners Behavior in MOOCs: An Examination of Performance and Motivation Using a Data-Driven Approach. *IEEE Access*, 6:73669–73685.

Amar, R., Eagan, J., and Stasko, J. (2005). Low-Level Components of Analytic Activity in Information Visualization. In *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*, INFOVIS '05, page 15, USA. IEEE Computer Society.

Anderson, A., Huttenlocher, D., Kleinberg, J., and Leskovec, J. (2014). Engaging with Massive Online Courses. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 687–698, New York, NY, USA. ACM.

Andres, J. M. L., Baker, R. S., Gašević, D., Siemens, G., Crossley, S. A., and Joksimović, S. (2018). Studying MOOC completion at scale using the MOOC replication framework. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*, LAK '18, pages 71–78, New York, New York, USA. ACM Press.

Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gašević, D., Mulder, R., Williams, D., Dawson, S., and Lockyer, L. (2016). A Conceptual Framework Linking Learning Design with Learning Analytics. In *Proceedings of the 6th International Conference on Learning Analytics & Knowledge*, LAK '16, pages 329–338, New York, NY, USA. ACM.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1).

Bodily, R. and Verbert, K. (2017). Trends and Issues in Student-facing Learning Analytics Reporting Systems Research. In *Proceedings of the 7th International Conference on Learning Analytics & Knowledge*, LAK '17, pages 309–318, New York, NY, USA. ACM.

Bogarín, A., Romero, C., Cerezo, R., and Sánchez-Santillán, M. (2014). Clustering for Improving Educational Process Mining. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, LAK '14, pages 11–15, New York, NY, USA. ACM.

Boroujeni, M. S. and Dillenbourg, P. (2018). Discovery and temporal analysis of latent study patterns in MOOC interaction sequences. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*, LAK '18, pages 206–215, New York, New York, USA. ACM Press.

Bos, N. and Brand-Gruwel, S. (2016). Student Differences in Regulation Strategies and Their Use of Learning Resources: Implications for Educational Design. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, LAK '16, pages 344–353, New York, NY, USA. ACM.

Bosch, N., Crues, R. W., Henricks, G. M., Perry, M., Angrave, L., Shaik, N., Bhat, S., and Anderson, C. J. (2018). Modeling Key Differences in Underrepresented Students' Interactions with an Online STEM Course. In *Proceedings of the Technology, Mind, and Society - TechMindSociety '18*, TechMindSociety'18, pages 1–6, New York, New York, USA. ACM Press.

Boulton, C. A., Kent, C., and Williams, H. T. (2018). Virtual learning environment engagement and learning outcomes at a 'bricks-and-mortar' university. *Computers & Education*, 126:129–142.

Brinton, C. G., Chiang, M., Jain, S., Lam, H., Liu, Z., and Wong, F. M. F. (2014). Learning about Social Learning in MOOCs: From Statistical Analysis to Generative Model. *IEEE Transactions on Learning Technologies*, 7(4):346–359.

Busson, A. J. G., Damasceno, A. L. d. B., Azevedo, R. G. d. A., Neto, C. d. S. S., Lima, T. d. S., and Colcher, S. (2017). A Hypervideo Model for Learning Objects. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, HT '17, page 245–253, New York, NY, USA. Association for Computing Machinery.

Carter, A. S., Hundhausen, C. D., and Adesope, O. (2017). Blending Measures of Programming and Social Behavior into Predictive Models of Student Achievement in Early Computing Courses. *ACM Trans. Comput. Educ.*, 17(3):12:1–12:20.

Champaign, J., Colvin, K. F., Liu, A., Fredericks, C., Seaton, D., and Pritchard, D. E. (2014). Correlating Skill and Improvement in 2 MOOCs with a Student's Time on Tasks. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 11–20, New York, NY, USA. ACM.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., and Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5/6):318–331.

Chen, G., Davis, D., Hauff, C., and Houben, G.-J. (2016a). On the Impact of Personality in Massive Open Online Learning. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, UMAP '16, pages 121–130, New York, NY, USA. ACM.

Chen, Q., Chen, Y., Liu, D., Shi, C., Wu, Y., and Qu, H. (2016b). PeakVizor: Visual Analytics of Peaks in Video Clickstreams from Massive Open Online Courses. *IEEE Transactions on Visualization and Computer Graphics*, 22(10):2315–2330.

Chen, Q., Yue, X., Plantaz, X., Chen, Y., Shi, C., Pong, T.-C., and Qu, H. (2018). ViSeq: Visual Analytics of Learning Sequence in Massive Open Online Courses. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–14.

Chen, Y. and Zhang, M. (2017). MOOC Student Dropout: Pattern and Prevention. In *Proceedings of the ACM Turing 50th Celebration Conference - China*, ACM TUR-C '17, pages 4:1–4:6, New York, NY, USA. ACM.

Chick, R. C., Clifton, G. T., Peace, K. M., Propper, B. W., Hale, D. F., Alseidi, A. A., and Vreeland, T. J. (2020). Using Technology to Maintain the Education of Residents During the COVID-19 Pandemic. *Journal of Surgical Education*, 77(4):729 – 732.

Christel, M. G. (2009). *Automated Metadata in Multimedia Information Systems: Creation, Refinement, Use in Surrogates, and Evaluation*, volume 1.

Cicchinelli, A., Veas, E., Pardo, A., Pammer-Schindler, V., Fessl, A., Barreiros, C., and Lindstädt, S. (2018). Finding traces of self-regulated learning in activity streams. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*, LAK '18, pages 191–200, New York, New York, USA. ACM Press.

Cleveland, W. S. and McGill, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, 79(387):531–554.

Cobo, G., García-Solórzano, D., Morán, J. A., Santamaría, E., Monzo, C., and Melenchón, J. (2012). Using Agglomerative Hierarchical Clustering to Model Learner Participation Profiles in Online Discussion Forums. In *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge*, LAK '12, pages 248–251, New York, NY, USA. ACM.

Corrin, L., de Barba, P. G., and Bakharia, A. (2017). Using Learning Analytics to Explore Help-seeking Learner Profiles in MOOCs. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, LAK '17, pages 424–428, New York, NY, USA. ACM.

Cox, K., Grinter, R. E., Hibino, S. L., Jagadeesan, L. J., and Mantilla, D. (2001). A Multi-Modal Natural Language Interface to an Information Visualization Environment. *International Journal of Speech Technology*, 4(3):297–314.

Crues, R. W., Henricks, G. M., Perry, M., Bhat, S., Anderson, C. J., Shaik, N., and Angrave, L. (2018). How do Gender, Learning Goals, and Forum Participation Predict Persistence in a Computer Science MOOC? *ACM Transactions on Computing Education*, 18(4):1–14.

Cucinotta, D. and Vanelli, M. (2020). WHO Declares COVID-19 a Pandemic. *Acta bio-medica : Atenei Parmensis*, 91(1):157—160.

Damasceno, A. L. B., Almeida, C., Fernandes, W., Lopes, H., and Barbosa, S. D. J. (2019). What Can Be Found from Student Interaction Logs of Online Courses Offered in Brazil. In *Anais do XXX Simpósio Brasileiro de Informática na Educação (SBIE 2019)*, pages 1641 – 1650, Brasilia. Brazilian Computer Society (Sociedade Brasileira de Computação - SBC).

Damasceno, A. L. B., Ribeiro, D. S., and Barbosa, S. D. J. (2019a). Visualizing student interactions to support instructors in Virtual Learning Environments. In *Proceedings of the International Conference on Human-Computer Interaction*. Springer International Publishing.

Damasceno, A. L. B., Ribeiro, D. S., and Barbosa, S. D. J. (2019b). What the Literature and Instructors Say about the Analysis of Student Interaction Logs on Virtual Learning Environments. In *2019 IEEE Frontiers in Education Conference (FIE)*, pages 1–9. IEEE.

Damasceno, A. L. B., Soares Neto, C. d. S., and Barbosa, S. D. J. (2017a). Integrating Participatory and Interaction Design of an Authoring Tool for Learning Objects Involving a Multidisciplinary Team. In Marcus, A. and Wang, W., editors, *Design, User Experience, and Usability: Theory, Methodology, and Management*, pages 554–569, Cham. Springer International Publishing.

Damasceno, A. L. B., Soares Neto, C. d. S., and Barbosa, S. D. J. (2017b). Lessons Learned from Evaluating an Authoring Tool for Learning Objects. In Zaphiris, P. and Ioannou, A., editors, *Learning and Collaboration Technologies. Novel Learning Ecosystems*, pages 77–89, Cham. Springer International Publishing.

Davis, D., Kizilcec, R. F., Hauff, C., and Houben, G.-J. (2018). The half-life of MOOC knowledge: a randomized trial evaluating knowledge retention and retrieval practice in MOOCs. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*, LAK '18, pages 1–10, New York, New York, USA. ACM Press.

Dias, S. B., Hadjileontiadou, S. J., Hadjileontiadis, L. J., and Diniz, J. A. (2015). Fuzzy cognitive mapping of LMS users' Quality of Interaction within higher education blended-learning environment. *Expert Systems with Applications*, 42(21):7399–7423.

Dicheva, D., Irwin, K., and Dichev, C. (2019). OneUp: Engaging Students in a Gamified Data Structures Course. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education - SIGCSE '19*, SIGCSE '19, pages 386–392, New York, New York, USA. ACM Press.

Dutt, A., Ismail, M. A., and Herawan, T. (2017). A Systematic Review on Educational Data Mining. *IEEE Access*, 5:15991–16005.

Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C., and Shum, S. B. (2018). Driving Data Storytelling from Learning Design. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, LAK '18, pages 131–140, New York, NY, USA. ACM.

Elbadrawy, A., Studham, R. S., and Karypis, G. (2015). Collaborative Multi-regression Models for Predicting Students' Performance in Course Activities. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, LAK '15, pages 103–107, New York, NY, USA. ACM.

Feild, J., Lewkow, N., Burns, S., and Gebhardt, K. (2018). A generalized classifier to identify online learning tool disengagement at scale. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*, LAK '18, pages 61–70, New York, New York, USA. ACM Press.

Feng, Y., Chen, D., Zhao, Z., Chen, H., and Xi, P. (2015). The Impact of Students and TAs' Participation on Students' Academic Performance in MOOC. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, ASONAM '15, pages 1149–1154, New York, NY, USA. ACM.

Ferguson, R. and Clow, D. (2015). Examining Engagement: Analysing Learner Subpopulations in Massive Open Online Courses (MOOCs). In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, LAK '15, pages 51–58, New York, NY, USA. ACM.

Fini, A. (2009). The Technological Dimension of a Massive Open Online Course: The Case of the CCK08 Course Tools. *The International Review of Research in Open and Distributed Learning*, 10(5).

Flick, U. (2018). *An Introduction to Qualitative Research.* SAGE.

Fu, X., Shimada, A., Ogata, H., Taniguchi, Y., and Suehiro, D. (2017). Real-time Learning Analytics for C Programming Language Courses. In *Proceedings of the 7th International Conference on Learning Analytics & Knowledge*, LAK '17, pages 280–288, New York, NY, USA. ACM.

Fulantelli, G., Taibi, D., and Arrigo, M. (2015). A framework to support educational decision making in mobile learning. *Computers in Human Behavior*, 47:50–59.

Gefen, D. and Keil, M. (1998). The Impact of Developer Responsiveness on Perceptions of Usefulness and Ease of Use: An Extension of the Technology Acceptance Model. *SIGMIS Database*, 29(2):35–49.

Gómez-Aguilar, D. A., Hernández-García, Á., García-Peñalvo, F. J., and Therón, R. (2015). Tap into visual analysis of customization of grouping of activities in eLearning. *Computers in Human Behavior*, 47:60–67.

Gong, L., Liu, Y., and Zhao, W. (2018). Using Learning Analytics to Promote Student Engagement and Achievement in Blended Learning: An Empirical Study. In *Proceedings of the 2nd International Conference on*

E-Education, E-Business and E-Technology - ICEBT 2018, ICEBT 2018, pages 19–24, New York, New York, USA. ACM Press.

Guo, P. J., Kim, J., and Rubin, R. (2014). How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 41–50, New York, NY, USA. ACM.

Guo, P. J. and Reinecke, K. (2014). Demographic Differences in How Students Navigate Through MOOCs. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 21–30, New York, NY, USA. ACM.

Haig, T., Falkner, K., and Falkner, N. (2013). Visualisation of Learning Management System Usage for Detecting Student Behaviour Patterns. In *Proceedings of the Fifteenth Australasian Computing Education Conference - Volume 136*, ACE '13, pages 107–115, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

Hamidon, Z. (2018). The Learner's Engagement in the Learning Process Designed Based on the Experiential Learning Theory in Post Graduate Program at Open University Malaysia. In *Proceedings of the 2018 2nd International Conference on Education and E-Learning - ICEEL 2018*, ICEEL 2018, pages 26–31, New York, New York, USA. ACM Press.

He, W. (2013). Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*, 29(1):90–102.

Heer, J. and Bostock, M. (2010). Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, page 203–212, New York, NY, USA. Association for Computing Machinery.

Hermans, F. and Aivaloglou, E. (2017). Teaching Software Engineering Principles to K-12 Students: A MOOC on Scratch. In *2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering Education and Training Track (ICSE-SEET)*, pages 13–22. IEEE.

Hernández-García, Á., González-González, I., Jiménez-Zarco, A. I., and Chaparro-Peláez, J. (2015). Applying social learning analytics to message

boards in online distance learning: A case study. *Computers in Human Behavior*, 47:68–80.

Hogo, M. A. (2010). Evaluation of e-learning systems based on fuzzy clustering models and statistical tools. *Expert Systems with Applications*, 37(10):6891–6903.

Hornbæk, K. and Hertzum, M. (2017). Technology Acceptance and User Experience: A Review of the Experiential Component in HCI. *ACM Transactions on Computer-Human Interaction*, 24(5):33:1–33:30.

Houston, II, S. L., Brady, K., Narasimham, G., and Fisher, D. (2017). Pass the Idea Please: The Relationship Between Network Position, Direct Engagement, and Course Performance in MOOCs. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, L@S '17, pages 295–298, New York, NY, USA. ACM.

Hu, Y.-H., Lo, C.-L., and Shih, S.-P. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior*, 36:469–478.

IEEE (2002). IEEE Standard for Learning Object Metadata. *IEEE Standard 1484.12.1-2002*, pages 1–40.

Kennedy, G., Coffrin, C., de Barba, P., and Corrin, L. (2015). Predicting Success: How Learners' Prior Knowledge, Skills and Activities Predict MOOC Performance. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, LAK '15, pages 136–140, New York, NY, USA. ACM.

Khosravi, H. and Cooper, K. M. (2017). Using Learning Analytics to Investigate Patterns of Performance and Engagement in Large Classes. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, SIGCSE '17, pages 309–314, New York, NY, USA. ACM.

Kirk, A. (2012). *Data Visualization: a successful design process.* Packt Publishing Ltd.

Kitchenham, B. and Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering.

Kizilcec, R. F., Piech, C., and Schneider, E. (2013). Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online

Courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, LAK '13, pages 170–179, New York, NY, USA. ACM.

Kizilcec, R. F. and Schneider, E. (2015). Motivation As a Lens to Understand Online Learners: Toward Data-Driven Design with the OLEI Scale. *ACM Trans. Comput.-Hum. Interact.*, 22(2):6:1–6:24.

Klerkx, J., Verbert, K., and Duval, E. (2017). Learning Analytics Dashboards. In *Handbook of Learning Analytics*, pages 143–150. Society for Learning Analytics Research (SoLAR).

Kovanović, V., Joksimović, S., Gašević, D., Owers, J., Scott, A.-M., and Woodgate, A. (2016). Profiling MOOC Course Returners: How Does Student Behavior Change Between Two Course Enrollments? In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, L@S '16, pages 269–272, New York, NY, USA. ACM.

Laveti, R. N., Kuppili, S., Ch, J., Pal, S. N., and Babu, N. S. C. (2017). Implementation of learning analytics framework for MOOCs using state-of-the-art in-memory computing. In *2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH)*, pages 1–6. IEEE.

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., and Munoz-Gama, J. (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Computers in Human Behavior*, 80:179–196.

Mangaroska, K. and Giannakos, M. (2019). Learning Analytics for Learning Design: A Systematic Literature Review of Analytics-Driven Design to Enhance Learning. *IEEE Transactions on Learning Technologies*, 12(4):516–534.

McGowan, A., Hanna, P., and Anderson, N. (2016). Teaching Programming: Understanding Lecture Capture YouTube Analytics. In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education*, ITiCSE '16, pages 35–40, New York, NY, USA. ACM.

Mödritscher, F., Andergassen, M., and Neumann, G. (2013). Dependencies Between E-Learning Usage Patterns and Learning Results. In *Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies*, i-Know '13, pages 24:1–24:8, New York, NY, USA. ACM.

Molenaar, I. and Knoop-van Campen, C. A. N. (2019). How Teachers Make Dashboard Information Actionable. *IEEE Transactions on Learning Technologies*, 12(3):347–355.

Munzner, T. (2014). *Visualization Analysis and Design.* CRC press.

Na, K. S. and Tasir, Z. (2017). A Systematic Review of Learning Analytics Intervention Contributing to Student Success in Online Learning. In *2017 International Conference on Learning and Teaching in Computing and Engineering (LaTICE)*, pages 62–68. IEEE.

Nandi, D., Hamilton, M., and Haland, J. (2011). How active are students in online discussion forums? In *Proceedings of the 30th Australasian Computing Education Conference - Volume 114*, ACE '11, pages 125–134, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

Nespereira, C. G., Dai, K., Redondo, R. P. D., and Vilas, A. F. (2014). Is the LMS Access Frequency a Sign of Students' Success in Face-to-face Higher Education? In *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality*, TEEM '14, pages 283–290, New York, NY, USA. ACM.

Nguyen, Q., Huptych, M., and Rienties, B. (2018). Linking students' timing of engagement to learning design and academic performance. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*, LAK '18, pages 141–150, New York, New York, USA. ACM Press.

Orduña, P., Almeida, A., López-de-Ipiña, D., and Garcia-Zubia, J. (2014). Learning Analytics on federated remote laboratories: Tips and techniques. In *IEEE Global Engineering Education Conference (EDUCON)*, pages 299–305.

Pardo, A., Han, F., and Ellis, R. A. (2016). Exploring the Relation Between Self-regulation, Online Activities, and Academic Performance: A Case Study. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, LAK '16, pages 422–429, New York, NY, USA. ACM.

Paredes, W. C. and Chung, K. S. K. (2012). Modelling learning & performance: a social networks perspective. In *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge*, LAK '12, pages 34–42, New York, NY, USA. ACM.

Park, J., Denaro, K., Rodriguez, F., Smyth, P., and Warschauer, M. (2017). Detecting Changes in Student Behavior from Clickstream Data. In *Proceedings of the 7th International on Learning Analytics and Knowledge*, LAK '17, pages 21–30, New York, NY, USA. ACM.

Qiu, J., Tang, J., Liu, T. X., Gong, J., Zhang, C., Zhang, Q., and Xue, Y. (2016). Modeling and Predicting Learning Behavior in MOOCs. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, WSDM '16, pages 93–102, New York, NY, USA. ACM.

Raad, M. A. and McKay, D. (2018). ESL student engagement in an introductory blended learning course in chemistry. In *Proceedings of the 2018 The 3rd International Conference on Information and Education Innovations - ICIEI 2018*, ICIEI 2018, pages 15–20, New York, New York, USA. ACM Press.

Ransdell, S. (2013). Meaningful posts and online learning in Blackboard across four cohorts of adult learners. *Computers in Human Behavior*, 29(6):2730–2732.

Reimers, G. and Neovesky, A. (2015). Student Focused Dashboards - An Analysis of Current Student Dashboards and What Students Really Want. In *Proceedings of the 7th International Conference on Computer Supported Education - Volume 1: CSEDU*, pages 399–404. INSTICC, SciTePress.

Rienties, B., Toetenel, L., and Bryan, A. (2015). "Scaling Up" Learning Design: Impact of Learning Design Activities on LMS Behavior and Performance. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, LAK '15, pages 315–319, New York, NY, USA. ACM.

Robal, T., Zhao, Y., Lofi, C., and Hauff, C. (2018). IntelliEye: Enhancing MOOC Learners' Video Watching Experience through Real-Time Attention Tracking. In *Proceedings of the 29th on Hypertext and Social Media - HT '18*, pages 106–114, New York, New York, USA. ACM Press.

Romero, C. and Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6):601–618.

Ruipérez-Valiente, J. A., Muñoz Merino, P. J., Leony, D., and Kloos, C. D. (2015). ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform. *Computers in Human Behavior*, 47:139–148.

Sacha, D., Stoffel, A., Stoffel, F., Kwon, B., Ellis, G., and Keim, D. (2014). Knowledge Generation Model for Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1604–1613.

Samson, P. J. (2015). Can student engagement be measured? And, if so, does it matter? In *2015 IEEE Frontiers in Education Conference (FIE)*, pages 1–4. IEEE.

Saraiva, T. (1996). Educação a distância no Brasil: lições da história. *Em Aberto*, 16(70):17–27.

Schwendimann, B. A., Rodriguez-Triana, M. J., and Vozniuk (2017). Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. *IEEE Transactions on Learning Technologies*, 10(1):30–41.

Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., and Pritchard, D. E. (2014). Who Does What in a Massive Open Online Course? *Commun. ACM*, 57(4):58–65.

Sergis, S. and Sampson, D. G. (2017). Teaching and Learning Analytics to Support Teacher Inquiry: A Systematic Literature Review. In Peña-Ayala, Alejandro, editor, *Learning Analytics: Fundaments, Applications, and Trends: A View of the Current State of the Art to Enhance e-Learning*, pages 25–63. Springer International Publishing.

Sheeba, T. and Krishnan, R. (2018). Prediction of student learning style using modified decision tree algorithm in e-learning system. In *Proceedings of the 2018 International Conference on Data Science and Information Technology - DSIT '18*, DSIT'18, pages 85–90, New York, New York, USA. ACM Press.

Shi, Y., Peng, Z., and Wang, H. (2017). Modeling Student Learning Styles in MOOCs. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, pages 979–988, New York, NY, USA. ACM.

Siemens, G. and Baker, R. S. J. d. (2012). Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge*, LAK '12, pages 252–254, New York, NY, USA. ACM.

Sorour, S., Goda, K., and Mine, T. (2015). Correlation of Topic Model and Student Grades Using Comment Data Mining. In *Proceedings of the 46th*

ACM Technical Symposium on Computer Science Education, SIGCSE '15, pages 441–446, New York, NY, USA. ACM.

Sun, K., Mhaidli, A. H., Watel, S., Brooks, C. A., and Schaub, F. (2019). It's My Data! Tensions Among Stakeholders of a Learning Analytics Dashboard. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pages 1–14, New York, New York, USA. ACM Press.

Sunar, A. S., White, S., Abdullah, N. A., and Davis, H. C. (2017). How Learners' Interactions Sustain Engagement: A MOOC Case Study. *IEEE Transactions on Learning Technologies*, 10(4):475–487.

Tempelaar, D. T., Rienties, B., and Nguyen, Q. (2017). Towards Actionable Learning Analytics Using Dispositions. *IEEE Transactions on Learning Technologies*, 10(1):6–16.

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., and Klerkx, J. (2014). Learning dashboards: An overview and future research opportunities. *Personal and Ubiquitous Computing*, 18(6):1499–1514.

Vieira, C., Parsons, P., and Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, 122:119–135.

Wan, H., Yu, Q., Ding, J., and Liu, K. (2017). Students' behavior analysis under the Sakai LMS. In *2017 IEEE 6th International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, pages 250–255. IEEE.

Wang, H., Hao, X., Jiao, W., and Jia, X. (2016a). Causal Association Analysis Algorithm for MOOC Learning Behavior and Learning Effect. In *2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, pages 202–206. IEEE.

Wang, X., Wen, M., and Rosé, C. P. (2016b). Towards Triggering Higher-order Thinking Behaviors in MOOCs. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, LAK '16, pages 398–407, New York, NY, USA. ACM.

Weiand, A., Manssour, I. H., and Silveira, M. S. (2019). Visual analysis for monitoring students in distance courses. *International Journal of Distance Education Technologies*, 17(2):18–44.

Wells, M., Wollenschlaeger, A., Lefevre, D., Magoulas, G. D., and Poulovassilis, A. (2016). Analysing Engagement in an Online Management Programme and Implications for Course Design. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, LAK '16, pages 236–240, New York, NY, USA. ACM.

Wen, M. and Rose, C. P. (2014). Identifying Latent Study Habits by Mining Learner Behavior Patterns in Massive Open Online Courses. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pages 1983–1986, New York, NY, USA. ACM.

Wiley, D. (2000). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. Technical report.

Wilkowski, J., Deutsch, A., and Russell, D. M. (2014). Student Skill and Goal Achievement in the Mapping with Google MOOC. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 3–10, New York, NY, USA. ACM.

Wolff, A., Zdrahal, Z., Nikolov, A., and Pantucek, M. (2013). Improving Retention: Predicting At-risk Students by Analysing Clicking Behaviour in a Virtual Learning Environment. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, LAK '13, pages 145–149, New York, NY, USA. ACM.

Xu, Z., Goldwasser, D., Bederson, B. B., and Lin, J. (2014). Visual Analytics of MOOCs at Maryland. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 195–196, New York, NY, USA. ACM.

Yang, D., Wen, M., Howley, I., Kraut, R., and Rose, C. (2015). Exploring the Effect of Confusion in Discussion Forums of Massive Open Online Courses. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, L@S '15, pages 121–130, New York, NY, USA. ACM.

Yu, T. and Jo, I.-H. (2014). Educational Technology Approach Toward Learning Analytics: Relationship Between Student Online Behavior and Learning Performance in Higher Education. In *Proceedings of the Fourth*

International Conference on Learning Analytics And Knowledge, LAK '14, pages 269–270, New York, NY, USA. ACM.

Zahalka, J., Rudinac, S., and Worring, M. (2015). Interactive Multimodal Learning for Venue Recommendation. *IEEE Transactions on Multimedia*, 17(12):2235–2244.

Zahalka, J. and Worring, M. (2014). Towards interactive, intelligent, and integrated multimedia analytics. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 3–12. IEEE.

Zhang, M. and Zhu, J. (2017). A Data-driven Analysis of Student Efforts and Improvements on a SPOC Experiment. In *Proceedings of the ACM Turing 50th Celebration Conference - China*, ACM TUR-C '17, pages 1:1–1:6, New York, NY, USA. ACM.

Zhou, L., Wu, S., Zhou, M., and Li, F. (2020). 'School's Out, But Class' On', The Largest Online Education in the World Today: Taking China's Practical Exploration During The COVID-19 Epidemic Prevention and Control As an Example. *SSRN Electronic Journal*.

Zhu, M., Bergner, Y., Zhang, Y., Baker, R., Wang, Y., and Paquette, L. (2016). Longitudinal Engagement, Performance, and Social Connectivity: A MOOC Case Study Using Exponential Random Graph Models. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, LAK '16, pages 223–230, New York, NY, USA. ACM.

A
Informed Consent Form

This form is mentioned in Chapter 2.

1ª via participante

Pᴏɴᴛɪғíᴄɪᴀ Uɴɪᴠᴇʀsɪᴅᴀᴅᴇ Cᴀᴛóʟɪᴄᴀ
ᴅᴏ Rɪᴏ ᴅᴇ Jᴀɴᴇɪʀᴏ

Termo de Consentimento Livre e Esclarecido

Eu, _____,

autorizo o uso, com fins estritamente acadêmicos, das informações fornecidas durante a entrevista

sobre o uso de tecnologias na educação, realizada no dia ____/____/_____, a partir das ____h.

Estou ciente que:

(1) minha participação é voluntária;

(2) a sessão será registrada com anotações e captura de áudio;

(3) esta sessão visa prover informações a uma pesquisa acadêmica relacionada ao uso de tecnologias educacionais, desenvolvido por André Luiz de Brandão Damasceno, sob orientação da professora Simone Diniz Junqueira Barbosa, na Pontifícia Universidade Católica do Rio de Janeiro;

(4) será garantido o anonimato no uso das informações capturadas nesta sessão;

(5) todos os dados brutos serão acessados somente pelos pesquisadores envolvidos nesta pesquisa.

(6) a qualquer momento, até dois ano após o término da pesquisa, poderei solicitar mais informações sobre o estudo ou cópias dos materiais divulgados, entrando em contato com o pesquisador através do e-mail **andre@telemidia.puc-rio.br**.

(   ) Autorizo o uso dos dados coletados conforme as condições supracitadas.

(   ) Não autorizo o uso dos dados coletados.

Pesquisador	Participante
André Luiz de Brandão Damasceno	
(nome)	*(nome)*
(assinatura)	*(assinatura)*

2ª via pesquisador

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Termo de Consentimento Livre e Esclarecido

Eu, _____,

autorizo o uso, com fins estritamente acadêmicos, das informações fornecidas durante a entrevista

sobre o uso de tecnologias na educação, realizada no dia ____/____/_____, a partir das ____h.

Estou ciente que:

(1) minha participação é voluntária;

(2) a sessão será registrada com anotações e captura de áudio;

(3) esta sessão visa prover informações a uma pesquisa acadêmica relacionada ao uso de tecnologias educacionais, desenvolvido por André Luiz de Brandão Damasceno, sob orientação da professora Simone Diniz Junqueira Barbosa, na Pontífícia Universidade Católica do Rio de Janeiro;

(4) será garantido o anonimato no uso das informações capturadas nesta sessão;

(5) todos os dados brutos serão acessados somente pelos pesquisadores envolvidos nesta pesquisa.

(6) a qualquer momento, até dois ano após o término da pesquisa, poderei solicitar mais informações sobre o estudo ou cópias dos materiais divulgados, entrando em contato com o pesquisador através do e-mail **andre@telemidia.puc-rio.br**.

(   ) Autorizo o uso dos dados coletados conforme as condições supracitadas.

(   ) Não autorizo o uso dos dados coletados.

Pesquisador   Participante

André Luiz de Brandão Damasceno   _____
(nome)   *(nome)*

_____   _____
(assinatura)   *(assinatura)*

B
Instructors Interview Script

This script is mentioned in Chapter 2.

First, I would like to thank you for participating in this study. This is a questionnaire regarding the use of educational technologies for distance education. Questions do not have right or wrong answers. The purpose here is to collect your opinion on the topic.

1. Name

2. Occupation

3. Age

4. What is your experience with teaching? What roles did you undertake (*e.g.*, teacher, tutor)?

5. Have you ever used Virtual Learning Environments (VLE) to study or teach? Which ones? (if not) Do you know any? (if not) Have you heard of Moodle, for example?

6. How do you use VLE? What features do you use or have used to (*e.g.*, upload content, post assignments, forum)?

7. What experience do you have in building online educational content? How do you produce it (*e.g.*, alone, in collaboration)? How is the course content planning? How do you use video, audio, animation, etc.? Does the instructor video is included on the material? How and when? How is the material produced (*e.g.*, homemade, with support from other departments, improvised studio, professional studio)?

8. What do you think that good educational content requires?

9. Do you have experience on building video lectures? How long you worked with educational video?

10. How do you analyze students' interactions of the course content (*e.g.*, by exams and exercises, by interaction logs)? Do you analyze students' interaction logs? How do you analyze logs?

11. What students' information do you consider relevant? What about their use on VLEs? Regarding the student's interaction with the VLE, what data do you analyze and what data would you like to have?

12. How do you evaluate students' learning? What about the students' interactions with the course material?

13. What would you like to know about the student before, during, and after a class? And before, during and after a course?

14. If you knew what the learner knows about the subject of a class, how would you use this information? Do you consider it relevant to identify the background of the student subject covered in course/video content?

15. What else would you like to know about yours students? Why?

16. Unlike face-to-face education, instructors have no idea about how students react to course content. What kind of behaviour student pattern would you like to identify?

17. How do you measure the students' motivation? Do you see any other means of capturing this through the VLE?

18. How would you use information about students' interactions with a video lecture? What information interest you? Why are they relevant? What would you do with them?

19. And about specific sections of a video lecture? If you knew that a student does not watch everything linearly, from beginning to end, what would you do? What information interests you? Why are they relevant? What would you do with them?

20. In your experience/opinion, what factors influence how the student sees a video lecture (*e.g.*, video duration, part of the content covered by the video, proximity to the exams)? Which of these factors could you measure/analyze today? What would you like to measure? How would you analyze them? How do you influence or would you like to influence the student about this?

21. Have you ever identified any relationship between students' use on VLE and grades achieved? And specifically on the use of video lectures, could you identify any relation with the grades obtained by students? If you could not identify, what would you need for this?

22. Have you identified any relationship between the students' use on VLE or video lecture and students' drop out?

C
Mapping Results (MR)

Found Mapping Results (MR) categorized in 14 topics. This results are described in Chapter 2.

Students with Self-regulated learning

1. MR-001 Online learning requires even more learner motivation and self-direction than traditional classroom-based instruction.

2. MR-002 Students with self-regulation not necessarily imply in good performance.

3. MR-114 Students with Negative Self-regulated have negative relation with academic performance.

4. MR-128 Students with self-regulation spend has more access in the VLE.

Click Activity

1. MR-009 Students who increased or has a high click activity have a higher probability of passing the course.

2. MR-010 Students who decreased or has a low click activity have a higher probability of failing the course.

Students' Engagement

1. MR-004 It is possible to obtain accurate classification models to predict students' engagement through the analysis of the students' interaction.

2. MR-011 Engagement in the online environment, as measured by assessment did on the environment.

3. MR-012 Engagement in the online environment, as measured by materials accessed on the environment.

4. MR-013 Engagement in the online environment, as measured by total hits, readings, and postings.

5. MR-014 Engagement in the online environment, as measured by how long students are watching each video.

6. MR-015 Engagement in the online environment, as measured by taking notes.

7. MR-016 Engagement on the discussion forum is associated with behaviors that are on the pathway to earning a certificate in MOOCs.

8. MR-094 The use of badges in course had higher engagement compared to course in which badges were not used.

9. MR-101 Learning design and learning design activities influence how students are engaging in our LMS.

10. MR-136 Teacher intervention can improve student engagement and achievement compared to teacher without intervention.

Cluster

1. MR-017 Students can be clustered into different groups based on their access or interaction patterns.

Forum

1. MR-026 Older students participate more in forums

2. MR-027 Students fluent in English MOOCs start more forum threads make more comments, and votes more on other learners' posts than to students not fluent.

3. MR-028 Student groups that use more forums tend to have a good performance.

4. MR-029 Students groups that do more assessment tend to use more forums

5. MR-030 Students groups that do more replies in forums tend to have a good performance.

6. MR-031 Students groups that init threads in forums tend to have a good performance.

7. MR-032 Classrooms with the most active consultant teachers – higher number of posts and replies – correspond to the worst performing group.

8. MR-052 Bachelors post significantly more questions in non-science courses.

9. MR-084 The number of posts in forums rises in periods close when assignments and examinations are due.

10. MR-087 Students in Health Sciences, Education and Arts & Letters make more questions than students in other colleges (*e.g.*, Business & Public Administration, Engineering & Technology, and Sciences).

11. MR-088 There is a positive correlation between the number of questions students asked to the instructor and their final grade for both courses.

12. MR-092 Students with Ph.D. degrees participated less in forums than non-Ph.D. holders.

13. MR-096 Students of humanities and social sciences courses make more threads classified as small-talk (*e.g.*, self-introductions)

14. MR-097 Teaching staff active participation in the forum is associated with a higher volume of post

15. MR-098 Peer-reviewed homework increase posts

16. MR-106 Student groups that have more posts are more likely to complete the course

17. MR-112 Teacher's participation (*e.g.*, posts, activity) lead to the engagement of the student (*e.g.*, module, wiki, blog, form, forum)

18. MR-051 Forum can be used as a predictor of students completing the course

19. MR-102 Comments can be used as a predictor of students performance

20. MR-116 The amount of comments produced by the two genders is related to the topic of the course, the course itself and the week of the course.

21. MR-121 Women has more postings than men.

22. MR-122 The proportion of women who post is higher than to those who only view the forum.

23. MR-123 The proportion of men who only view the forum is higher than to those who post.

24. MR-124 Women who participate more in forum are those one who indicate they took the course due to being a student and wanting to learn.

25. MR-125 Men who participate more in forum are those one who indicate they took the course due to career aspirations.

26. MR-130 Students who use a narrower variety of words are more likely to complete the course.

27. MR-131 Students that write longer posts are more likely to complete the course.

28. MR-132 Students that do more replies in forums tend to complete the course.

29. MR-133 Students that init threads in forums tend to complete the course.

30. MR-134 Students characterized by grades lower than average, has their questions with more votes than average.

31. MR-135 Students characterized by grades higher than average, has their questions with less votes than average.

Students' performance

1. MR-006 Older students taking online courses tended to perform better than younger students.

2. MR-019 Page viewing times has a significant positive impact on students' final score

3. MR-020 Students with a satisfatory performance ignore part of the materials in distance courses

4. MR-021 There are demographic differences in student behavior, interaction patterns and performance

5. MR-022 Students not fluent in English MOOCs has certification rates lower than students fluent.

6. MR-023 Higher achievement is related, counter-intuitively, to being in a non-English speaking country.

7. MR-024 Patterns of Students performance are related with Human Development Index (HDI)

8. MR-035 Successful students are more frequently and regularly participating and engaged in online activities.

9. MR-041 Researches in predicting performance of students

10. MR-042 Viewing of the course material and students previous performance contribute the most to the predicted grades.

11. MR-079 It is possible to obtain accurate classification models to predict students' performance through the analysis of the students' interaction

12. MR-085 Good performance prediction based on basic demographic information and academic history

13. MR-086 Good performance prediction based on the ratio of credits earned to credits attempted

14. MR-100 There are positive correlations between productive and assessment activities and pass rates

15. MR-104 Conditional Random Fields (CRF) can be used as a predictor of students performance

16. MR-109 Dataset consisting of time-based data can be used to predict performance

17. MR-126 Student performance has significant decrease in the curse goes through.

18. MR-127 Students who earned high grades could be distinguished from other students based on interaction patterns in the first two weeks of class.

Course completion

1. MR-005 MOOCs typically have lower completion rates.

2. MR-034 Planning prompts increased course completion

3. MR-037 Completing activities or assessment can be used as a predictor of students completing the course

4. MR-038 Event logs from students can be used as a predictor of students completing the course

5. MR-039 Data Access can be used as a predictor of students completing the course

6. MR-091 Students who earned a certificate on online course most commonly held a master's degree

7. MR-093 Students who will go on to complete the course are on a rapid increase in their advanced search skills use

8. MR-103 Support Vector Machines (SVM) classifier with Different Error Costs (DEC) can be used to predict certification

Students' drop out

1. MR-033 Classrooms where the teacher intervened the least had high drop out rates.

2. MR-040 It is possible to obtain accurate classification models to predict students' drop out through the analysis of the students' interaction

3. MR-043 Researches in predicting drop out of students in distance learning

4. MR-047 Student inactive for more than 3 weeks, likely drop out from the course

5. MR-095 Researches in identify reasons for drop out

6. MR-105 Helping resolving or providing responses to student confusion reduces their drop out in the courses

7. MR-107 The extent to which different types of confusion affect drop out is determined by specific courses.

8. MR-108 Students that show confusion and don't have their doubts resolved are more likely to drop out.

Tools

1. MR-025 Analytics in learning systems can be used to provide both auditing and interventions in student learning.

2. MR-048 Suggest a tool that uses logs from e-learning system to teachers monitore students behavior, motivation or performance.

3. MR-049 Tools to aid teachers analyze student behavior.

4. MR-050 Tools to aid students analyze their self performance.

5. MR-117 Tools to aid students analyze their self behavior.

Students' Intention

1. MR-053 Students who intended to earn a certificate in MOOCs are more likely to watch most video lectures and attempt most assessments in the course.

2. MR-054 Students with the intention to earn a certificate in MOOCs are more likely to actively engage on the discussion forum.

3. MR-055 Students with the intention to meet new people are engaged on the discussion board and even of receiving social recognition from peers (*e.g.*, votes) in MOOCs.

4. MR-056 Students with the intention to meet new people are less engaged with lectures and assessments in MOOCs.

5. MR-057 Students with the intention to take the course with colleagues are more likely to watch at least 10% of lecture videos and assessments in MOOCs.

6. MR-058 Students with the intention to take the course with colleagues are more likely to more likely to earn a certificate in MOOCs.

7. MR-059 Students with the intention to take the course with colleagues are less likely to engage in the discussion forum in MOOCs.

8. MR-060 Students with the intention to take the course due to job relevance were more likely to watch at least 10% of lecture videos in MOOCs.

9. MR-061 Students with the intention to take the course due to job relevance were less likely to actively engage on the forum.

10. MR-062 Students with the intention to change its career were more likely to watch more than 80% of video lectures

11. MR-063 Students with the intention to change its career were more likely to complete more than half of the assessments in the course.

12. MR-064 Students with the intention to dedicate more than 6 hours on average to the course and aimed for achieving a higher grade in the course were less likely to be sporadic users.

13. MR-065 Students with the intention to participate in all aspects of the course were also less likely to be sporadic in their use of course material.

Attending

1. MR-018 Students from countries with lower student-teacher ratios seem to participate in MOOCs later in life than students from countries with higher student-teacher ratios

2. MR-036 Researches in attending detection of students in distance learning

3. MR-090 Students who participating on online course most commonly held a bachelor's degree

Students' personality

1. MR-110 Researches in identify students' personality

2. MR-111 Students' personality can be identified through interaction logs.

Videos

1. MR-007 Streaming is the dominant form of access to video lectures.

2. MR-066 There is a positive correlation, within a certain range, between the lengths of online time and watching video with learning results

3. MR-067 The Audience Retention (AR) decreases with video length.

4. MR-068 Students often do not watch the entire videos

5. MR-069 The shortest videos has the highest engagement.

6. MR-070 Students also engaged less frequently with assessment problems that followed longer videos.

7. MR-071 Students usually engaged more with talking-head videos.

8. MR-072 Students are more engaged with Khan-style tutorials.

9. MR-073 Students engaged more with pre-production videos.

10. MR-074 Students generally engaged more with videos where instructors spoke faster.

11. MR-075 Students re-watch tutorials more frequently than lectures.

12. MR-076 Students with satisfactory performance interact more in timeline than others.

13. MR-077 Older students backjump more frequently in timeline.

14. MR-078 A higher student-teacher ratio corresponds to fewer backjumps;

15. MR-080 Videos often are where learners spend a great deal of time in a MOOC

16. MR-081 Videos with slide (theory) format performs worst in terms of holding the students' attention on video than videos with slide (code)

17. MR-082 Students not fluent in English MOOCs create significantly more transcript events than students fluent.

18. MR-083 Students not fluent in English MOOCs take significantly more video play, video pause, and seek actions compared to students fluent.

19. MR-089 Students engaged more on videos filmed informally with the instructor sitting at his office desk, when compared with a video filmed in a multi-million dollar TV production studio.

20. MR-099 Student dwelling increases with a high or complexity information rate

21. MR-113 Videos with walkthroughs coding tend to have a higher engagement than the active coding sections

22. MR-115 Certificate-earners also engage heavily with in-video quizzes.

23. MR-118 Students watching videos has high levels of inattention.

   Others

 1. MR-003 Courses with less than 1,000 participants are likely to fit with the Zipf's law.

 2. MR-008 Access in the online environment resources increase in periods close exams or assessment deadline.

 3. MR-044 There is herding phenomenon in online learning

 4. MR-045 The difficult level of a course is positively correlated with the rational herding of learners will be.

 5. MR-046 The experience of learners is positively correlated with rational herding behavior.

 6. MR-119 Most learners use hardware and software setups which are capable to support Webcam-based attention tracking techniques.

7. MR-120 Most learners are reluctant to allow the use of Webcam-based attention tracking techniques.

8. MR-129 The time spent on VLE by students is on average less than the number of hours recommended by instructors.

D
Grouping Requirements (RQ) and Mapping Results (MR)

Grouping RQ and MR uncovered in 11 visualization groups (VGs) with common VLE data logs and elaborated questions to be answered. This Appendix is described in Chapter 4.

VG-01

1. MR-011: Engagement in the online environment, as measured by assessment did on the environment.

2. RQ-04: Identify student interest patterns on the course.

3. RQ-11: Identify self-regulated students.

    – **Topics:**

    (a) Students who completed the assignments.
    (b) Assignments completed by students.

    – **Questions:**

    (a) Which students completed the assignments?
    (b) Which assignments were completed by the students?

VG-02

1. MR-012 Engagement in the online environment, as measured by materials accessed on the environment.

2. RQ-04 Identify student interest patterns on the course.

    – **Topics:**

    (a) Students' accesses to materials (*e.g.*, videos, ebooks).
    (b) Materials most accessed by the students (*e.g.*, videos, ebooks).

    – **Questions:**

    (a) Which students accessed the materials?
    (b) Which materials were most accessed by the students?

VG-03

1. MR-013 Engagement in the online environment, as measured by total hits, readings, and postings.

2. RQ-04 Identify student interest patterns on the course.

   – **Topics:**

     (a) Number of student forum accesses, posts, and likes.

   – **Questions:**

     (a) How many student accesses, posts, and likes were there?

VG-04

1. MR-014 Engagement in the online environment, as measured by how long students are watching each video.

2. MR-068 Students often do not watch the entire videos.

3. RQ-04 Identify student interest patterns on the course.

   – **Topics:**

     (a) How long the students are watching each video.

   – **Questions:**

     (a) For how long did the students watch each video?

VG-05

1. MR-017 Students can be clustered into different groups based on their access or interaction patterns.

2. MR-020 Students with a satisfactory performance ignore part of the materials in distance courses.

3. MR-028 Student groups that use more forums tend to have a good performance.

4. MR-030 Students groups that do more replies in forums tend to have a good performance.

5. MR-031 Students groups that init threads in forums tend to have a good performance.

6. MR-035 Successful students are more frequently and regularly participating and engaged in online activities.

7. MR-106 Student groups that have more posts are more likely to complete the course.

8. MR-114 Students with Negative Self-regulated have negative relation with academic performance.

9. RQ-02 Identify student access patterns (*e.g.*, login, materials).

10. RQ-03 Identify student performance patterns.

11. RQ-04 Identify student interest patterns on the course.

12. RQ-05 Identify student usage patterns on the forum.

13. RQ-07 Identify student interaction patterns (*e.g.*, materials).

14. RQ-08 Identify student participation patterns on the course.

15. RQ-09 Identify student drop out patterns.

16. RQ-14 Identify pace learning student.

17. RQ-17 Relate both students' navigation and performance.

18. RQ-18 Relate video length and student performance.

19. RQ-21 Relate video script and student performance.

   – **Topics:**

   (a) Relation between students' grades and VLE access.
   (b) Relation between students' grades and materials access.
   (c) Relation between students' grades and assignments completed.
   (d) Relation between students' grades and forum access.
   (e) Relation between students' grades and forum posts.
   (f) Relation between students' grades and forum replies.
   (g) Relation between students' grades and forum threads.

   – **Questions:**

   (a) What is the relation between students' grades and VLE access?
   (b) What is the relation between students' grades and materials access?

(c) What is the relation between students' grades and assignments completed?

(d) What is the relation between students' grades and forum access?

(e) What is the relation between students' grades and forum posts?

(f) What is the relation between students' grades and forum replies?

(g) What is the relation between students' grades and forum threads?

VG-06

1. MR-026 Older students participate more in forums.

    – **Topics:**

    (a) Relation between students' age and VLE access.

    (b) Relation between students' age and forum access.

    (c) Relation between students' age and forum posts.

    (d) Relation between students' age and forum replies.

    (e) Relation between students' age and forum threads.

    – **Questions:**

    (a) What is the relation between students' age and VLE access?

    (b) What is the relation between students' age and forum access?

    (c) What is the relation between students' age and forum posts?

    (d) What is the relation between students' age and forum replies?

    (e) What is the relation between students' age and forum threads?

VG-07

1. MR-009 Students who increased or has a high click activity have a higher probability of passing the course.

2. MR-037 Completing activities or assessment can be used as a predictor of students completing the course.

3. MR-040 It is possible to obtain accurate classification models to predict students' drop out through the analysis of the students' interaction.

4. MR-051 Forum can be used as a predictor of students completing the course.

5. MR-079 It is possible to obtain accurate classification models to predict students' performance through the analysis of the students' interaction.

6. MR-088 There is a positive correlation between the number of questions students asked to the instructor and their final grade for both courses.

7. MR-102 Comments can be used as a predictor of students performance.

8. MR-114 Students with Negative Self-regulated have negative relation with academic performance.

9. RQ-06 Predict student performance.

   – **Topics:**

     (a) Prediction of students' grades and drop out.

   – **Questions:**

     (a) What is the prediction of students' grades and drop out?

VG-08

1. MR-008 Access in the online environment resources increase in periods close exams or assessment deadline.

   – **Topics:**

     (a) Number of students' accesses per day.
     (b) Number of students' accesses per week.

   – **Questions:**

     (a) How many students' accesses were there each day?
     (b) How many students' accesses were there per week?

VG-09

1. RQ-01 Statistics of interactions on video (*e.g.*, access, re-watch, seek).

2. RQ-16 Know if the student is watching the video.

3. RQ-20 Know which videos (or video segments) the students have liked.

4. RQ-24 Know in which video segments the students have difficulty.

5. RQ-26 Know how students navigate in the video.

6. RQ-28 Know when students has connection problems.

– **Topics:**

(a) Statistics of interactions with video (*e.g.*, play, pause, seek backward, seek forward).

– **Questions:**

(a) What are the statistics of interactions with video (*e.g.*, play, pause, seek backward, seek forward)?

VG-10

1. RQ-12 Know whether the student has understood the video.

2. RQ-20 Know which videos (or video segments) the students have liked.

   – **Topics:**

   (a) Videos were understood by students.

   – **Questions:**

   (a) Which videos were understood by students?

VG-11

1. RQ-10 Identify student navigation patterns on the VLE.

2. RQ-25 Know how students navigate among the video lectures.

   – **Topics:**

   (a) Students' navigation patterns on the VLE.

   – **Questions:**

   (a) What were the students' navigation patterns on the VLE?

E
Visualizations

Visualizations to obtain instructors' preferences, as described in Chapter 4.

(a) Table

(b) Bar Chart
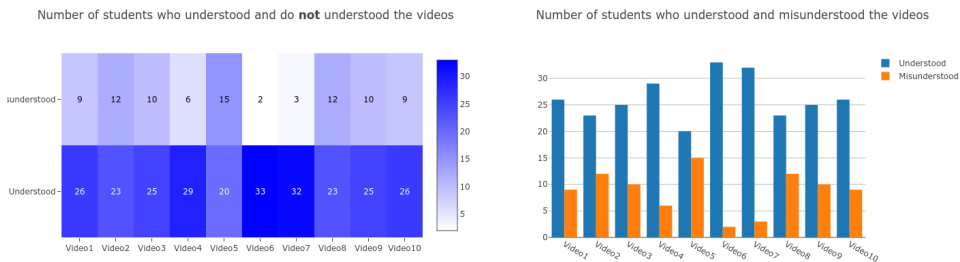
(c) Bar Chart
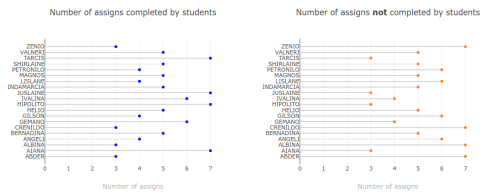
(d) Lollipop
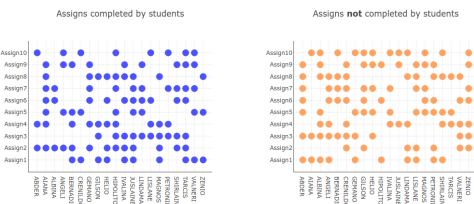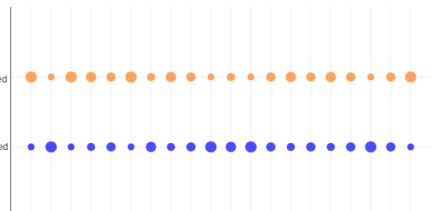
(e) Bubble Chart

(f) Bubble Chart
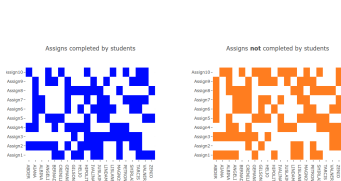
(g) Heatmap
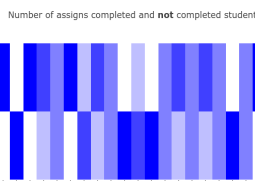
(h) Heatmap
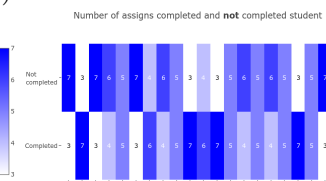
(i) Heatmap
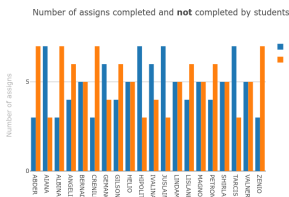
(j) **Grouped Bar Chart**
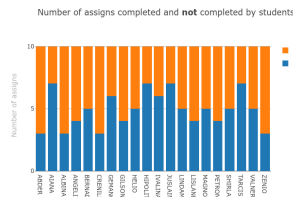
(k) Stacked Bar Chart
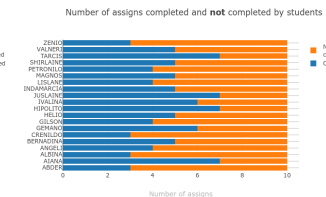
(l) Stacked Bar Chart

Figure E.1: Sample of visualizations used in **VG-01** question: "Which students completed the assignments?"

(a) **Table**

(b) Grouped Bar Chart

(c) Stacked Bar Chart

(d) Stacked Bar Chart

(e) Heatmap

(f) Heatmap

(g) Heatmap

(h) Heatmap

(i) Bubble Chart

(j) Bubble Chart

Figure E.2: Sample of visualizations used in **VG-02**: "Which students accessed the materials?"

PUC-Rio - Certificação Digital Nº 1621797/CA

Students	Likes	Posts	Access
ALISSA	17	10	31
ALZINEI	20	7	45
ANIZA	3	17	32
ARLUCE	15	1	19
ARTENILDO	5	2	15
CAIE	2	7	10
CRISELITE	4	11	28
DELUBIA	15	9	43
DULCIELE	18	16	38
EVANIZIA	2	16	32
GLEISE	20	5	31
JEDESON	19	10	31

(a) Table
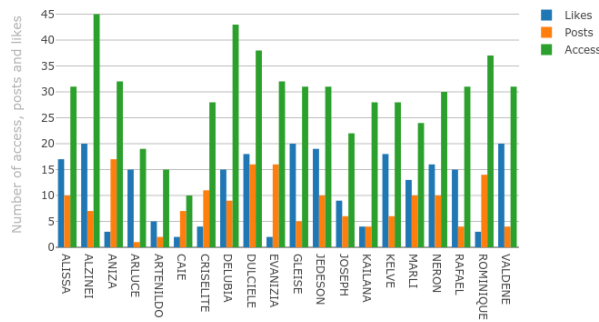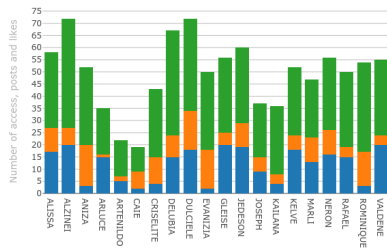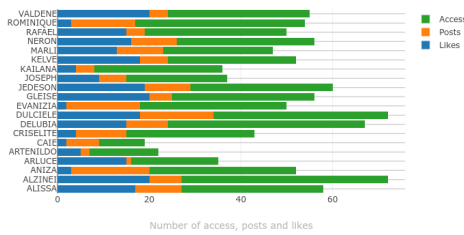
(b) **Grouped Bar Chart**

(c) Stacked Bar Chart

(d) Stacked Bar Chart
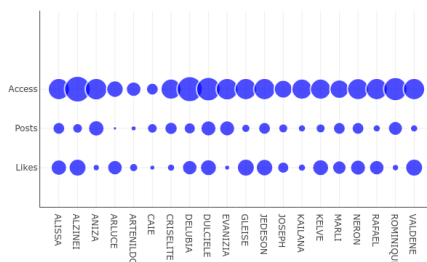
(e) Bubble Chart
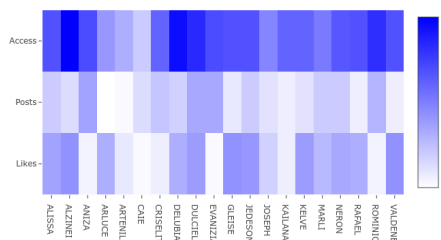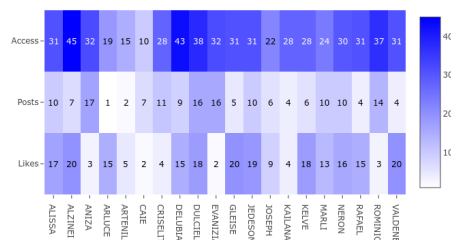
(f) Heatmap

(g) Heatmap

Figure E.3: Sample of visualizations used in **VG-03**: "How many student accesses, posts, and likes were there?"

Students	Video1 (485s)	Video2 (550s)	Video3 (444s)	Video4 (391s)	Video5 (409s)	Total
ACIEL	491,246	342,280,97	13,133,55	324,316,58	430,482,406	3673
DULCIMARI	09,255,37,9	71	167,319	6,401,202,1		2083
ENAILE		51,93,136,12	200,318	84,188,163		1572
GENESSON	19,323,460	2,192,415,4	23			2038
GEOVANEID	86,55,326	76,167,90,2	116,240,32,2	112	425,115,127	2613
GILSON	9,337,372,4	131,41			6,296,100,3	2643
IATA	301,315,387	241,268	80,160,297		361,122,322	2854
IGUATEMI		274	366,47	6,277	97,65,177,40	1912
JOCIVANDO	240,379,18,8			303,267,296	270	1781
LAUREA	431,520	60,54,106,5	308	169,158	89,272,221	2638
LERINALDO	8,381,355,50	0,337,298,34	4,273,221,12	2,133,116,2	136,158,454	4740

(a) Table

Students	Video1	Video2	Video3	Video4	Video5	Total
ACIEL	737	719	201	698	1318	3673
DULCIMARI	493	71	486	1033	0	2083
ENAILE	0	619	518	435	0	1572
GENESSON	802	1213	23	0	0	2038
GEOVANEID	467	850	517	112	667	2613
GILSON	1554	172	0	0	917	2643
IATA	1003	509	537	0	805	2854
IGUATEMI	0	274	413	283	942	1912
JOCIVANDO	645	0	0	866	270	1781
LAUREA	951	470	308	327	582	2638
LERINALDO	1440	1344	930	278	748	4740
MAELLY	672	1175	986	932	0	3765

(b) Table
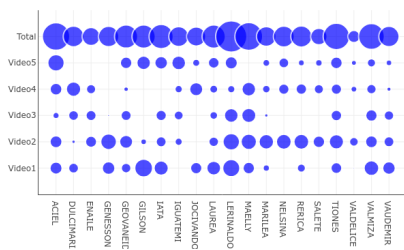
(c) Bubble Chart
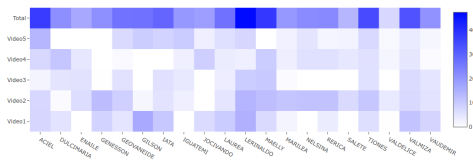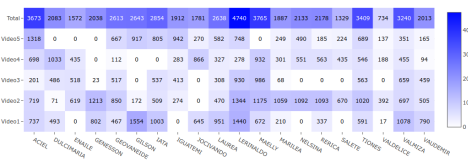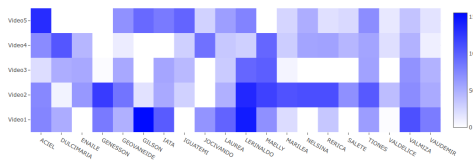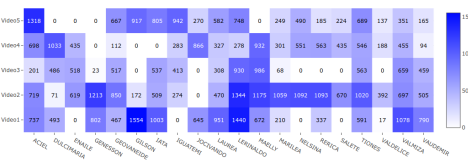
(d) Heatmap

(e) Heatmap

(f) Heatmap

(g) Heatmap
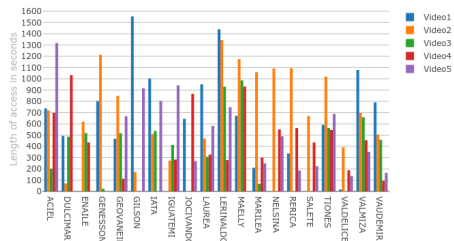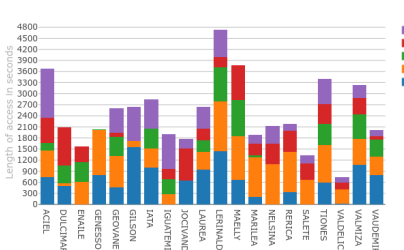
(h) **Grouped Bar Chart**
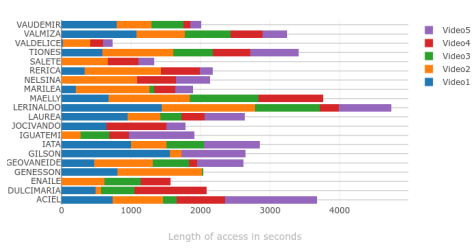
(i) Stacked Bar Chart

(j) Stacked Bar Chart

Figure E.4: Sample of visualizations used in **VG-04**: "For how long did the students watch each video?"

Students	Grade	AssignTotal	MaterialTotal	AVA Access	Forum Access	Forum Post	Forum Replies	Forum Add Thread
AGLAISE	53	4	8	23	28	6	0	3
ALANA	83	3	58	82	68	27	20	5
ARISTAO	48	0	6	22	6	3	2	0
ATALITA	76	3	51	60	37	3	10	2
CAROLANE	77	2	36	54	67	15	18	4
CECI	53	2	10	38	9	4	3	0
CEJANA	61	3	23	50	27	5	10	3
CENO	50	0	3	16	7	3	2	2
CORINTO	69	2	26	45	37	6	10	7
DENILSON	50	0	3	22	6	3	0	2

(a) **Table**
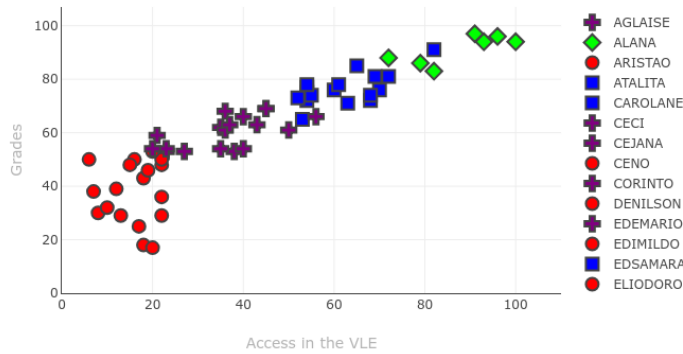
(b) Scatterplot

(c) Box & Whisker Plot

(d) Box & Whisker Plot
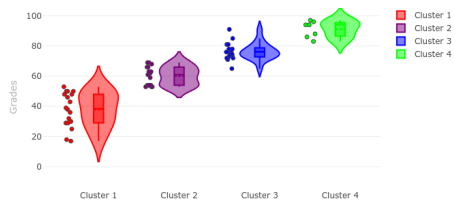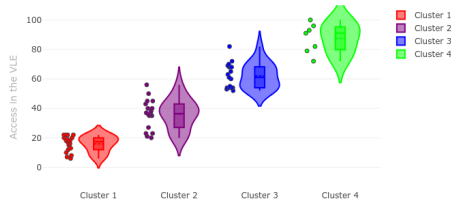
(e) Violin Plot

(f) Violin Plot

Figure E.5: Sample of visualizations used in **VG-05**: "What is the relation between students' grades and VLE access?"

Students	Age	Forum Access	Forum Post	Forum Replies	Forum Add Thread
AFONSO	57	87	25	28	5
ALBARI	38	39	7	10	2
ALOISIO	28	30	6	2	2
ANNELIESE	29	23	5	7	3
ARILZA	53	68	11	31	8
AUDENISE	51	71	10	29	6
BABIANA	31	35	2	10	2
CLEISSIANE	25	8	2	3	0
ELI	38	34	10	11	2
ELZIMEIRE	32	43	7	9	6
ERIO	49	29	8	9	2

(a) **Table**
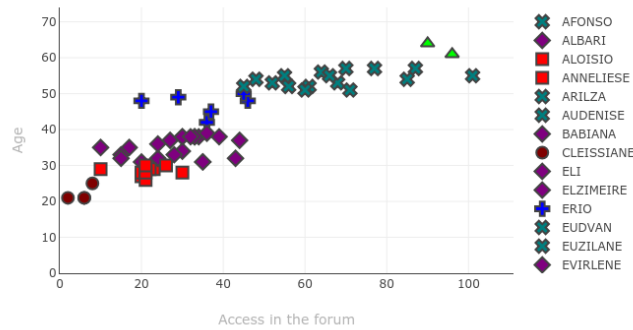
(b) Scatterplot
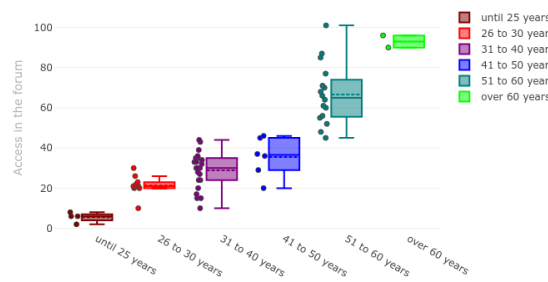
(c) Box & Whisker Plot
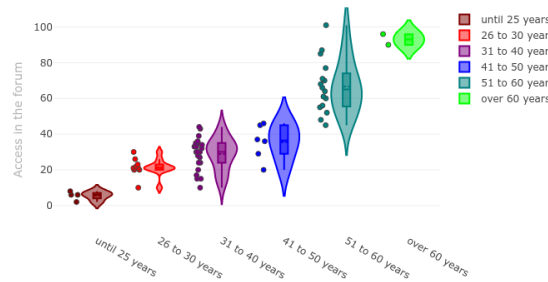
(d) Violin Plot

Figure E.6: Sample of visualizations used in **VG-06**: "What is the relation between students' age and forum access?"

Students	Predicted Cluster	Predicted Dropout	Predicted Grade	AVA Access	Forum Post	Forum Replies	Forum Add Thread	Forum Access
ADGERS	Dropout	true	1	22	1	1	0	7
ALANA	61 - 70	false	66	49	2	10	0	19
ALFONSI	61 - 70	false	66	51	3	5	1	16
ALISTON	0 - 60	false	15	25	2	5	0	20
ANAIDA	71 - 80	false	75	55	4	2	6	39
BARBAR	81 -	false	81	76	20	20	1	56

(a) Table
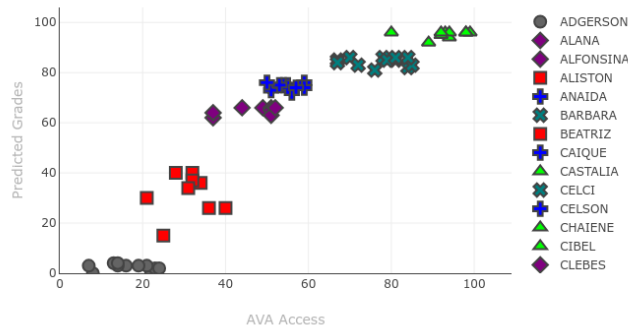
(b) **Scatterplot**
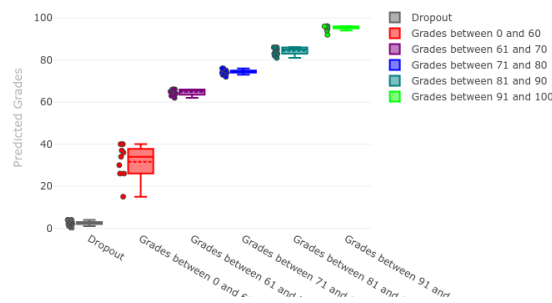
(c) Box & Whisker Plot
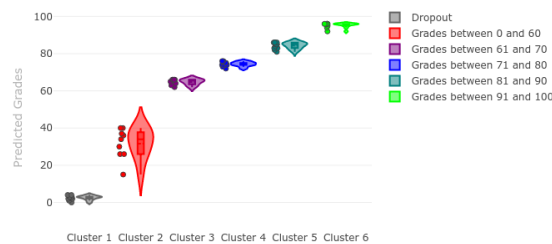
(d) Violin Plot

Figure E.7: Sample of visualizations used in **VG-07**: "What is the prediction of students' grades and drop out?"

Students	Week1	Week2	Week3	Week4	Week5	Week6	Week7
ALCEDO	42	71	94	109	49	84	109
ALDALUCI	38	56	65	112	46	85	96
ALISON	57	74	98	117	27	52	99
BERNARD	25	48	84	112	55	88	124
CRISLEIN	51	71	87	118	52	74	99
DEUSILET	29	50	65	97	41	64	93
DIANARI	43	72	86	115	55	70	95
DILMA	36	73	85	101	54	72	94
DIOGE	42	74	90	119	47	92	115
DULCILA	39	61	76	108	53	76	115
GERDAN	36	47	77	98	39	55	81
GEVALDIN	40	59	72	103	52	91	116

(a) **Table**
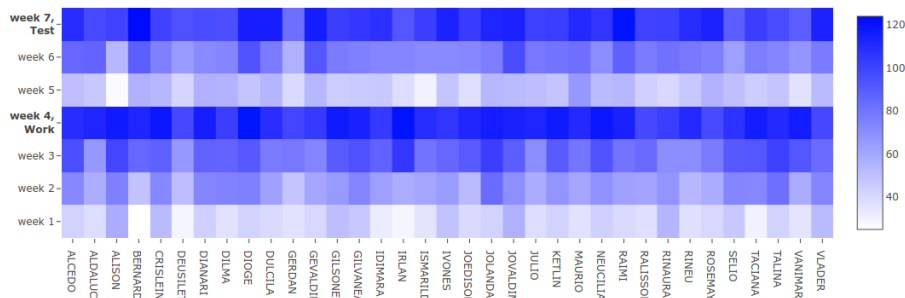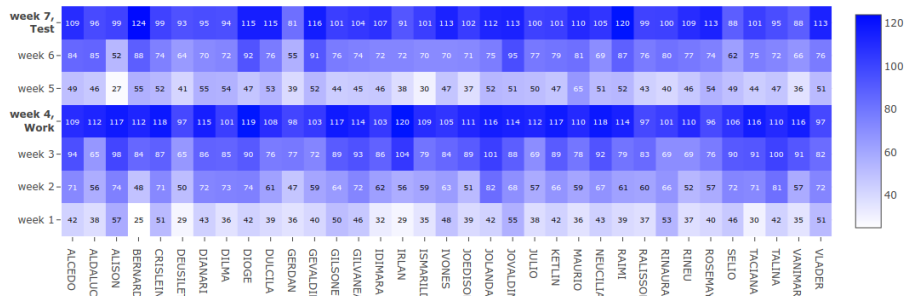
(b) Heatmap

(c) Heatmap
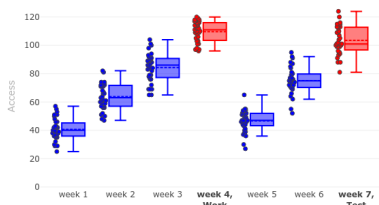
(d) Box & Whisker Plot
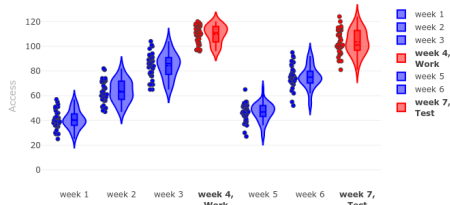
(e) Violin Plot

Figure E.8: Sample of visualizations used in **VG-08**: "How many students' accesses were there per week?"

Students	Action
ABIDIEL	Play at 0s
ABIDIEL	Pause at 17s
ABIDIEL	Seek from 17s to 13s
ABIDIEL	Pause at 13s
ABIDIEL	Seek from 13s to 12s
ABIDIEL	Pause at 20s
ABIDIEL	Play at 20s
ABIDIEL	Dropout at 20s
ACELIA	Play at 0s
ACELIA	Pause at 7s
ACELIA	Play at 7s
ACELIA	Seek from 17s to 11s

Time	Play	Pause	Seek from	Seek to	Dropout
0	24	0	0	2	3
1	0	0	1	0	0
2	1	1	1	2	0
3	2	3	0	1	0
4	0	0	1	2	0
5	0	0	0	0	0
6	0	1	2	1	1
7	1	1	0	0	0
8	0	0	1	3	0
9	3	3	1	1	0
10	1	2	0	2	0
11	0	1	3	3	0

(a) Table
      (b) Table
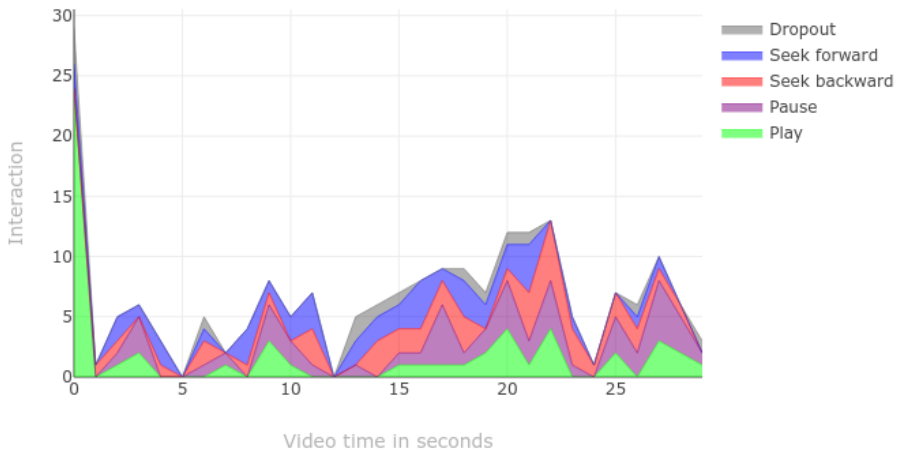
(c) **Stacked Area Graph**
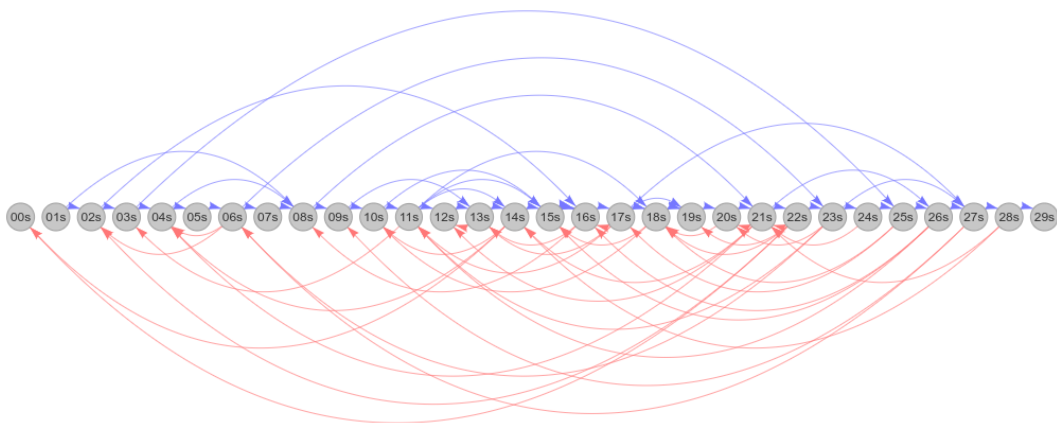
(d) Flow Chart

Figure E.9: Sample of visualizations used in **VG-09**: "What are the statistics of interactions with video (e.g. play, pause, seek)?"

(a) Table

(b) Table

(c) Bar Chart

(d) Bar Chart

(e) Lollipop
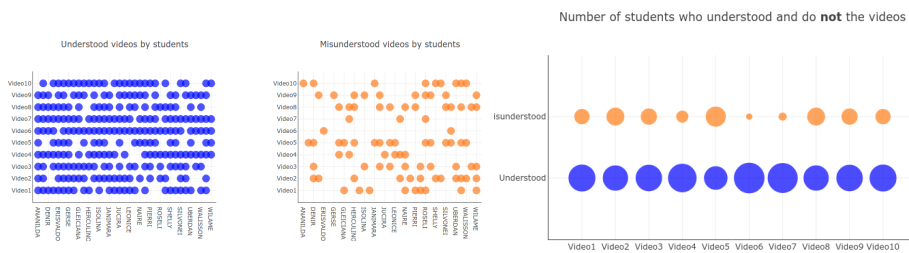
(f) Bubble Chart

(g) Bubble Chart
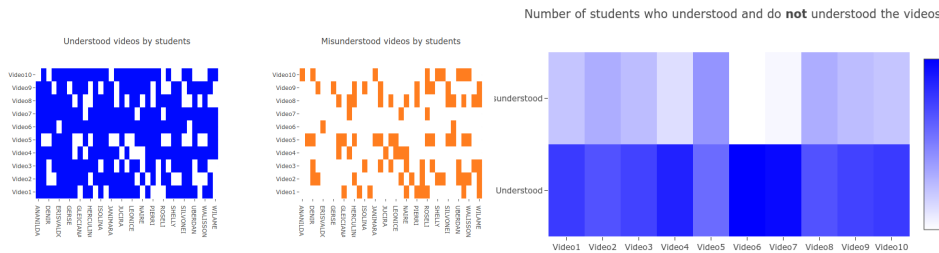
(h) Heatmap

(i) Heatmap

(j) Heatmap

(k) **Grouped Bar Chart**
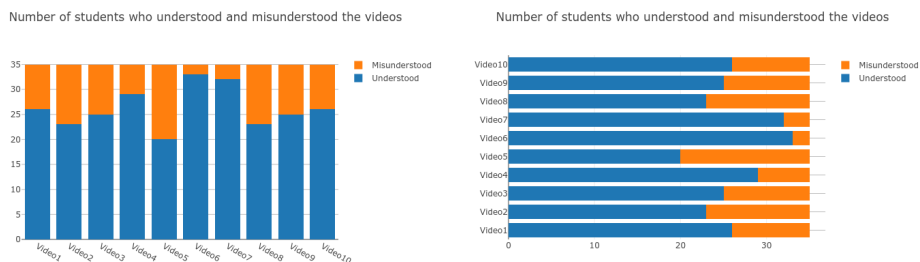
(l) Stacked Bar Chart

(m) Stacked Bar Chart

Figure E.10: Sample of visualizations used in **VG-10**: "Which videos were understood by students?"

Students	Grade	AVA Access	Forum Post	Forum Replies	Forum Add Thread	Forum Access
ALANA	59	51	2	5	7	24
ALVANO	81	70	16	16	6	47
ANDREICE	84	68	12	23	3	64
ANDY	78	52	14	20	2	56
ARUZA	97	89	39	28	8	87
BAYARD	63	41	8	7	4	27
BIBIANO	91	92	26	20	13	86
CELSIO	92	94	25	32	8	76
DALVANISE	79	53	17	15	5	46
DEBORA	20	34	0	0	2	10

(a) **Table**
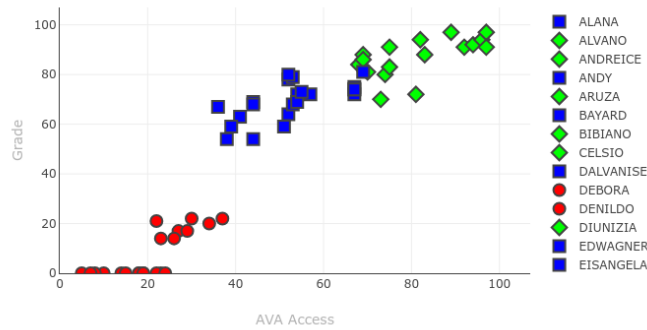
(b) Scatterplot
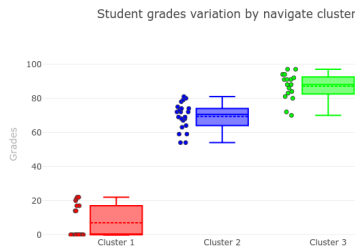
(c) Box & Whisker Plot   (d) Violin Plot
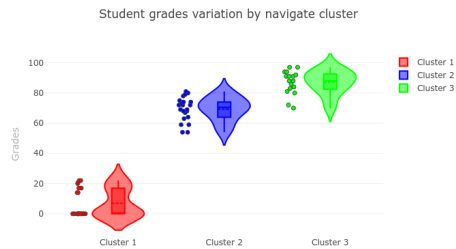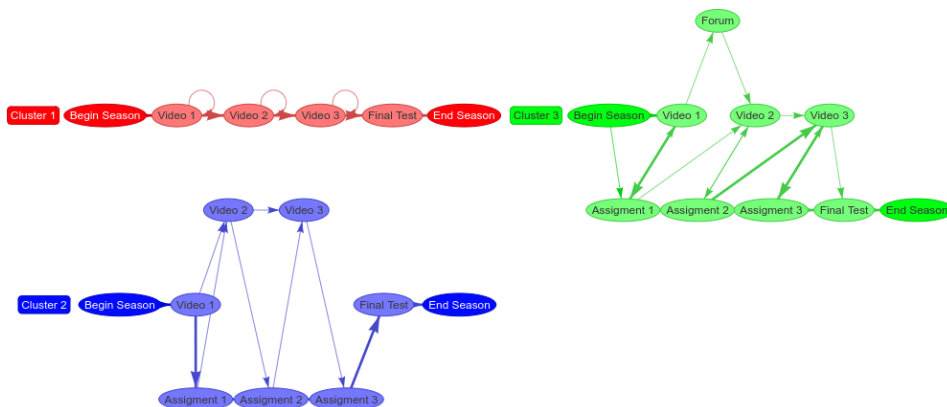
(e) Flow Chart

Figure E.11: Sample of visualizations used in **VG-11**: "What were the students' navigation patterns on the VLE?"

F
Online Survey for Evaluate our Proposed Model

Survey mentioned in Chapter 5 to evaluate our proposal.
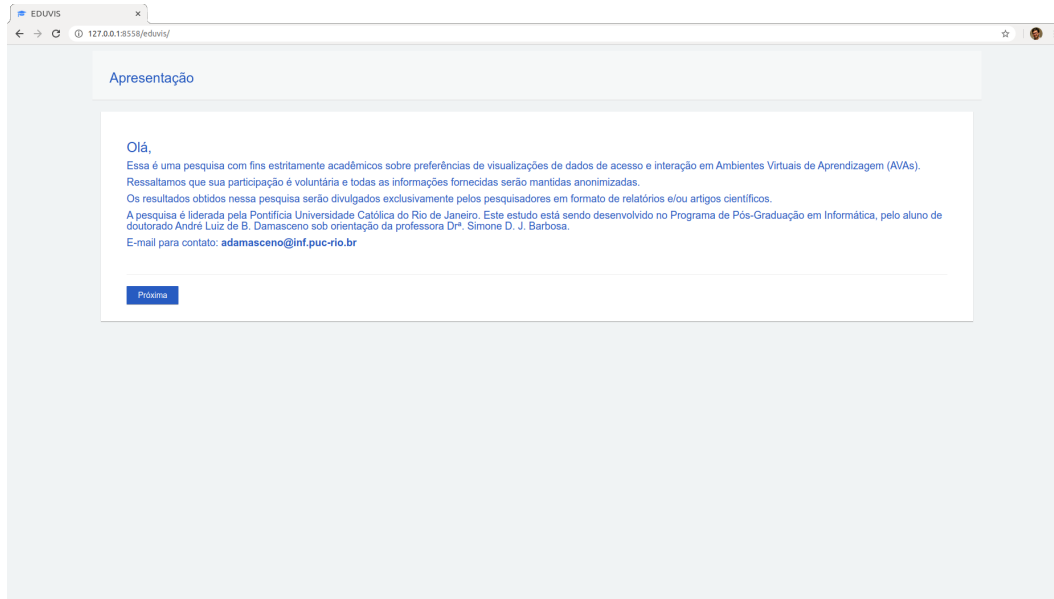
Figure F.1: Welcome screen.

Figure F.2: Screen to instructors identify themselves.

Figure F.3: Screen to instructors answer about their experience with VLE.

Figure F.4: Screen to instructors answer about meaningful student information.
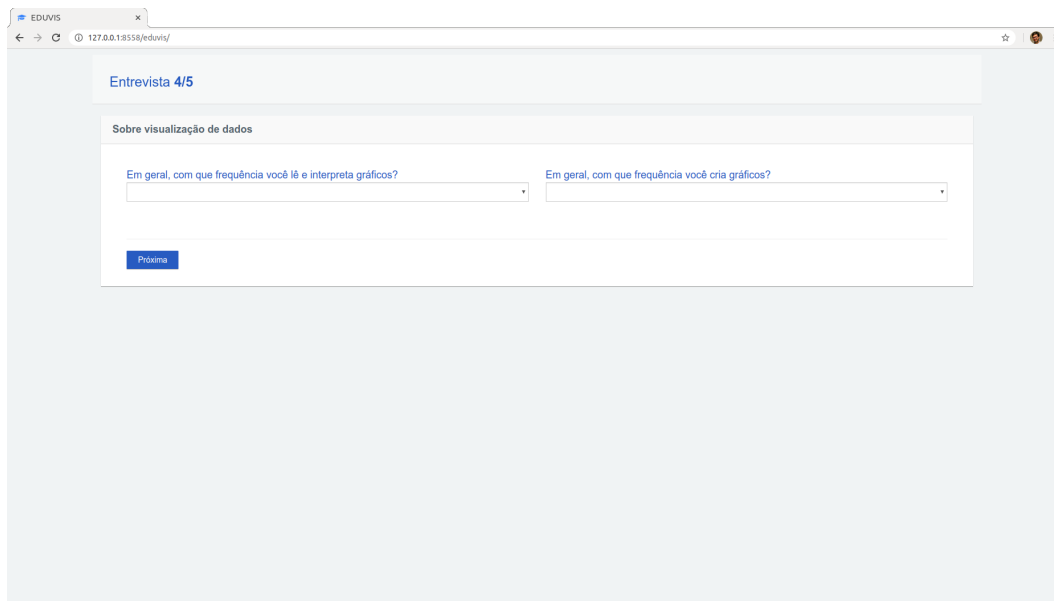
Figure F.5: Screen to instructors answer about data visualization.
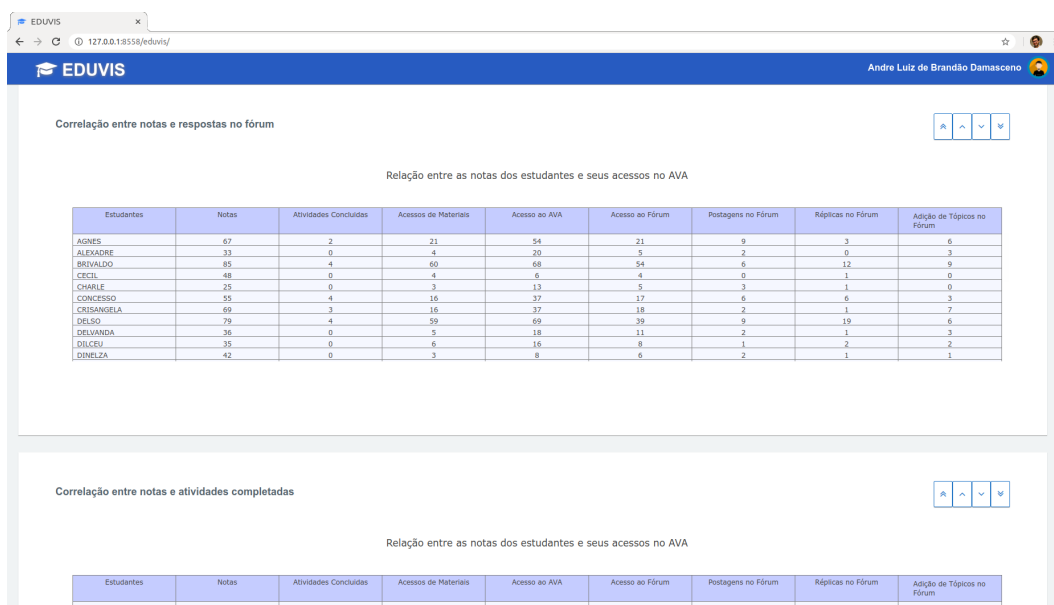
Figure F.6: Screen to instructors analyze a static dashboard.
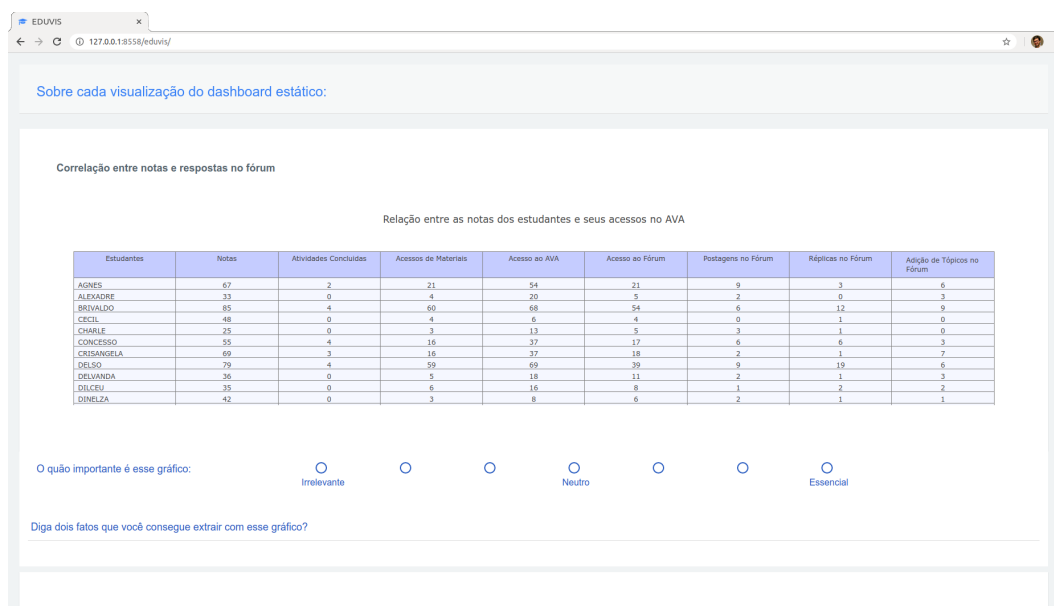
Figure F.7: Screen to instructors provide a feedback about the static dashboard.
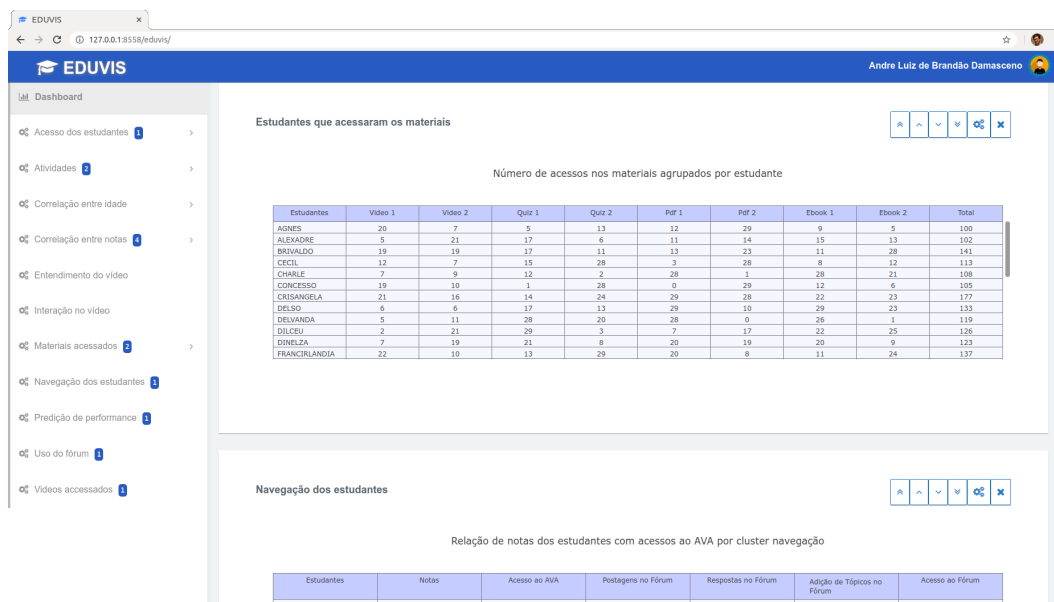
Figure F.8: Screen to instructors make use of the Eduvis and assemble dashboards.

Figure F.9: Screen to instructors provide a feedback about their dashboards.
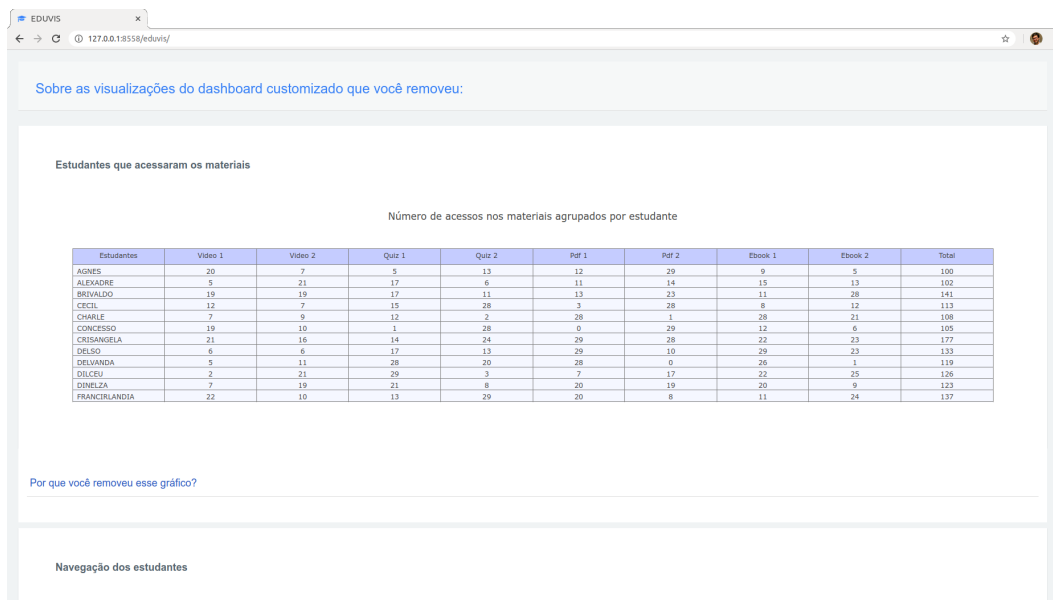
Figure F.10: Screen to instructors provide a feedback about why they removed charts from the default dashboard.

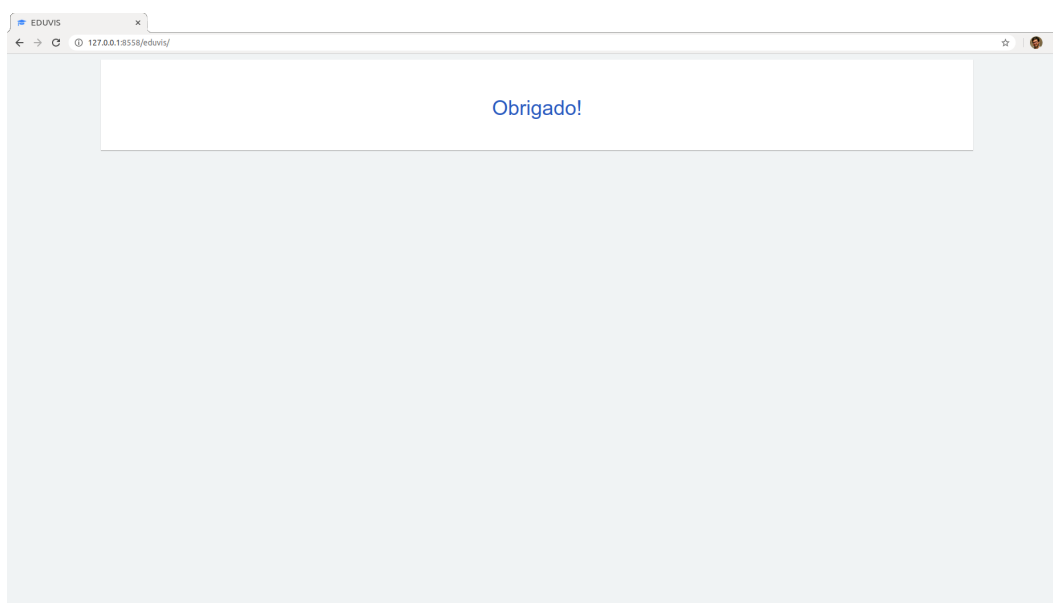Figure F.11: Screen to instructors answer a questionnaire based on TAM (Gefen and Keil, 1998).

Figure F.12: Final screen.