Factors Related to the Quality of Computer Science Teacher Education Courses: A Public Data Analysis

André Menolli
Universidade Estadual do
Norte do Paraná
Universidade Estadual de
Londrina
menolli@uenp.edu.br

Andrey Bragagnolo
Universidade Estadual do
Norte do Paraná
andreyb.bragagnolo@gmail.com

Rafael F. Gonçalves
Universidade Estadual
de Londrina
rafael.goncalves@uel.br

João Coelho Neto
Universidade Estadual
do Norte do Paraná
joaocoelho@uenp.edu.br

Abstract

Computational thinking has been gaining importance in Basic Education, as an instrument for the development of learning, and the Computer courses, focusing on training teachers are responsible for training professionals to foster it. Thus, using the Design Science Research method, it was defined and deployed a public data analysis method. Once the method was deployed, data from Higher Education Institutions in Brazil and data from Brazilian National Student Performance Exam (ENADE) were analyzed. It was explored the profile of courses and students in Computer Science Teacher Education courses, focus on understand, how the different findings about these courses are related to the course quality. The results show that many factors are related to the course performance in ENADE, both factors associated with students enrolled in the courses, as well as factors associated with the courses and institutions. Finally, this article presents an overview of the computer science teacher education courses in Brazil, considering the course qualities.
Keywords: Computer Science Teacher Education Courses; Computational Thinking; Basic Education; Design Science Research Method

Resumo

O pensamento computacional vem ganhando importância na Educação Básica, como instrumento para o desenvolvimento da aprendizagem, e os cursos de Informática, com foco na formação de professores, são responsáveis pela formação de profissionais para fomentá-la. Assim, utilizando o método Design Science Research, foi definido e implantado um método de análise de dados públicos. Uma vez que o método foi implantado, foram analisados dados de Instituições de Ensino Superior do Brasil e dados do Exame Nacional de Desempenho de Estudantes (ENADE). Explorou-se o perfil dos cursos e alunos de Licenciatura em Computação, com foco em compreender, como as diferentes constatações sobre esses cursos se relacionam com a qualidade do curso. Os resultados mostram que muitos fatores estão relacionados ao desempenho do curso no ENADE, tanto fatores associados aos alunos matriculados nos cursos, quanto fatores associados aos cursos e instituições. Por fim, este artigo apresenta uma visão geral dos cursos de formação de professores de informática no Brasil, considerando as qualidades do curso.
Palavras-chave: Cursos de Licenciatura em Computação; Pensamento Computacional; Educação Básica; Método de Pesquisa Design Science.

1   Introduction

In Brazil, an important agent to improve the computational thinking are the Computer Science courses, focus on training teachers, or Computer Science Teacher Education Courses. Since the teaching of computing, especially in Basic Education, has been gaining importance, the introduction of computational and algorithmic thinking at this level of education provides the necessary cognitive resources for problem solving, across all areas of knowledge (Brasil, 2016).

Many schools have adopted initiatives to strengthen computational thinking, with educational robotics classes, programming clubs, creation of digital games, simulators and interdisciplinary activities with "unplugged computing" (Bell et al., 2011).

Considering this scenario, it is important an overview of the Computer Science Teacher Education Courses in Brazil, understanding how different phenomenon occurs in these courses, for instance the dropout rate. Dropout in the Higher Education has been addressed by several works over the past few years, e.g., (Rodrigues; et al., 2015; Damasceno and Carneiro, 2018).

Specifically, in the Computer Science Teacher Education Courses, some works explore the phenomenon of dropout, such as (Prietche and Pazeto, 2010) which aims to understand the reasons that dropout occurs, and Calixto (2015), that conduct a study on the main factors that contribute to dropout in e-learning courses.

However, dropout is only one phenomenon that can be explored using public data, as provide by INEP (National Institute of Education Anísio Teixeira Studies and Research). Furthermore, in Brazilian education context, there are other public data available, as data about ENADE[1] (National Student Performance Exam). Some work explores the data available by ENADE, as Hinterholz et al. (2014) and Freitas et al. (2019), that aims to understand the women profile in the computing area through an analysis of microdata referring to ENADE. However, despite data from ENADE and INEP's census be open and available, few studies try to explore them and understand which factors are related to courses qualities, and with factors may influence the ENADE results.

Thus, this work intends to bring an overview of Computer Science Teacher Education Courses in Brazil and understand how the different findings about these courses are related to the quality, using as quality criteria the classification of the course in the ENADE.

Nevertheless, the challenges now are many. To start with, the available systems for data sharing are very confusing with not a clear data model and visual interface that would support a proper analysis, and thus decision-making. Secondly, it is needed to analyze together data collected in different time and purposes (INEP and ENADE data) to give a clear overview how ENADE results are associated to the courses characteristics. Finally, there is a continuous need to take the analysis of such data to a broader level (e.g., National Level - Brazil), and from that understand the overall impact of public policies and existing efforts.

In this manner, this paper aims to answer the follow research question "What factors may bee related to the classification of Computer Science Teacher Education Courses at ENADE?". In order to answer it, we used we use public data, from the years 2017 and 2018, from INEP and data from ENADE 2017. From the data, to perform the analyzes, we use several computational techniques, such as transformation, cleaning and data integration, dimensional modeling and DW (Data Warehouse), as well as BI (Business Intelligence) tools, building a solution that allows countless analyzes on different metrics of courses and students.

---

[1] the ENADE abbreviation will be written in Portuguese throughout the text.

2   Background

2.1   National Census of Higher Educations and ENADE data

The main source of data on higher education in Brazil is available from INEP, called the Higher Education Census. This census gathers information about higher education institutions, e-learning and on-site courses, as well as information about students and vacancies offered at graduation. Data is collected from questionnaires filled out by and carried out by the MEC (Ministry of Education) system. During the questionnaire filling period, institutional researchers can make the necessary changes or additions to their institutions' data at any time. After this period, INEP checks the consistency of the information collected and the system is recovered for verification and data validation by the HEI (Higher Education Institutions).

ENADE is the national exam, which is applied every year to a portion of the country's graduation courses, and each course is evaluated every three years. To calculate the ENADE concept, the grade of graduates in undergraduate courses is considered, using both the grade of FG (General Training) and the CE (Specific Components) grade (INEP, 2017).

In this work, only the ENADE concept was used to classify the quality of courses. However, there are other indexes that can be used, such as the IDD or Difference Indicator between Observed and Expected Performance. IDD is a grade used in quality measuring that includes the students' results to calculate the net effect by the HEI. Another concept existing is the CPC or Preliminary Course Concept. CPC uses the ENADE, IDD, information on professor and other factors to calculate a grade for the previous year.

2.2   Related Works

Regarding the course quality, some studies it was found in the literature. The quality of courses may be related to different aspects, and several studies trying to understand the impact of these aspects on course quality. For instance, Yin et al. (2014) present a survey of the quality of teaching in higher education in Chinese universities and Karpinski et al. (2017) seeks to identify what are the critical factors for the success of offering a specialization course in distance education, from the perspective of academics. However, as the quality can be measured from different methods, the related works present studies using ENADE as quality criteria of the course.

Among the studies exploring ENADE, many studies focus to identify the main factors related to the student's performance at the test. Silva et al. (2015) present a study of factors that impact the performance of business students in the ENADE score. The study presented by Rocha et al. (2018) aims to verify the association between the performance of graduating students of Nutrition at ENADE and socioeconomic factors, academic trajectory and profile of the institution. Towards in the same direction, the study of Moriconi and Nascimento (2011), seeks to identify, using ENADE data, factors associated with the performance of new engineers trained in Brazil. The work of Brito (2016) aims is to determine how the characteristics of the teaching staff of an HEI influence the performance of the graduates of their undergraduate courses in Administration in the ENADE test. Finlay, the study of Simm (2015) presents an analysis of the quality of undergraduate courses in Business Administration in Brazil and Paraná in the period from 2019 to 2012, through the analysis of the results of ENADE in these two periods.

Although the previous works have addressed important and relevant questions, they conducted studies using only ENADE data. In the researched literature, there is no study using the INEP and ENADE databases to understand the Computer Science Teacher Education courses. The first study towards this direction is presented by Menolli et al. (2020), which presents a study about the System Information course in Brazil. However, this study uses only the INEP database, to understand Brazilian Information Systems (IS) courses. In particular, it analyzes the phenomenon of dropout from different perspectives. This work instead integrates data from

different databases using BI-method, to provide general analysis about aspects related to the quality of Computer Science Teacher Education courses, as well as describes a BI-based approach to conduct such analysis, including data sources, Extract-Transformation-Load (ETL) process, and data model.

3   Method

This work adopted the Design Science Research (DSR) method as a means to develop the BI-Based Methodology. In particular, DSR method was choose since it focuses on "designing and evaluating IT artifacts that can be used to address practice and research problems" (Gregor and Hevner, 2013). We understand an artifact as "a thing that can be transformed into a material (e.g., model) or process (e.g., method)" (Hevner et al., 2004) and, therefore, our methodology comprises the artifact utilized to investigate the quality of Computer Science Teacher Education Courses (i.e., problem of practice).

For proper elaborating and evaluating the artifact, we followed the six-steps DSR method proposed by Peffers et al. (2007). First, we identified the research problem that was already explored in Section 1, i.e., a comprehensive and suitable methodology to support in the analyzes of open education data. Next, we developed the IT artifact (i.e., BI-Based Methodology) to address the identified problem. Through the results allowed by our proposed methodology, we definitely would be able to understand the critical factors about the problem. Ultimately, this paper itself consists the communication of obtained results, i.e., the last step of Peffers et al. (2007)'s DSR method.

3.1   BI-Based Methodology

In order to build an overview of Computer Science Teacher Education Courses in Brazil and try to understand how the ENADE concept is related to particular characteristics of courses, we constructed a DW (Data Warehouse), and from this we deployed a BI (Business Intelligence) solution, using Pentaho Business Analytics[2].

It was established a software architecture, presented in Figure 1, that utilizes existing analytics (e.g., Mondrian), ETL (e.g., Kettle), storage (e.g., Postgres), and web-based visualization (e.g., OLAP) tools. Having developed the IT artifact, we carried out a case study method in which the context was the Brazilian Computer Science Teacher Education Courses, and the unit of analysis was the ENADE concept of these courses (Yin, 2002).

---

[2]   https://www.hitachivantara.com/en-us/products/data-managementanalytics/pentaho-platform/pentaho-business-analytics.html
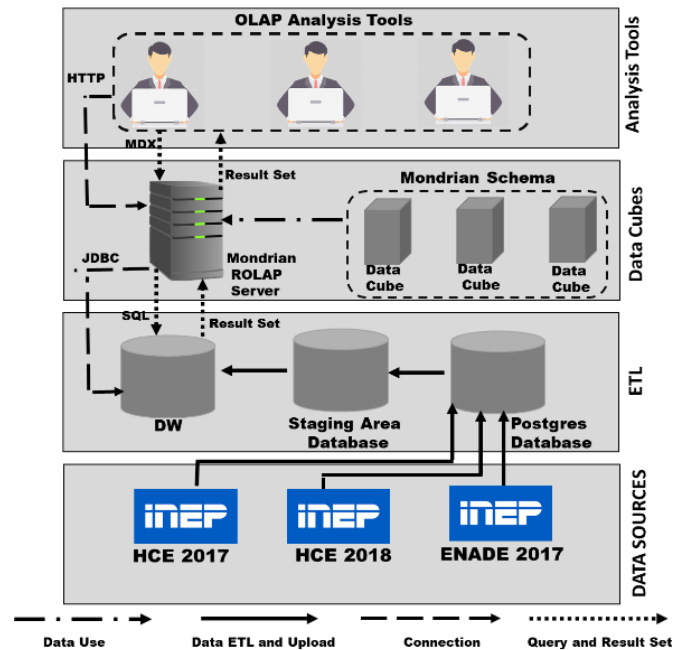
Figure 1. Architecture to Deploy BI Solution. Adapted from Menolli et al. (2020)

3.1.1 Data Sources

We used the public data from ENADE (National Student Performance Exam) of the year 2017 and public data from Higher Education Census (HCE) of the year 2017 and 2018. HCE presenting data compiled provided by all HEI and each database is comprised by more than 11 million records from all students and courses of superior education on Brazil. In turn, the ENADE database contains data on the performance of each course in the national exam.

3.1.2 Extract-Transformation-Load (ETL) process

The HCE and ENADE data are available on CSV (Comma-Separated Values) format. Hence, the first task to start building the DW was to understand the data and import them to a database in a DBMS (Database Management System). After we have had all data in Postgres tables, we migrate them to tables in the staging area. This area is an intermediate step between the extraction of the data source and the load in DW. It is responsible for filtering the data, as well as for conducting transformations and integrations. A fundamental concept that greatly simplifies DW projects and facilitates maintenance is the use of data staging areas. Starting from the logical project of the staging area, it is possible to have a good idea of the attributes and source tables necessary to populate the DW. The data staging should just contain the necessary information to populate the warehouse.

The staging area, defined in this work, uses a normalized data model to allow the best data consistency and to facilitate the integration process among different databases. Furthermore, in this step, it was created one table for each table in Postgres Database.

In this step, two main tasks were executed. First, to integrate data from different years of HCE and ENADE. Second, apply several data processes and transformation rules. The ETL tool for Data Integration (Kettle) platform was used [3]. Once the staging area was created and the most data processing and transformation were performed, we created a DW using a multidimensional model.

---

[3] https://community.hitachivantara.com/

3.1.3 Dimensional Model

The dimensional model was used because, according to Song et al. (2001), there are two main advantages of using a dimensional model in data warehouse environments. First, a dimensional model provides a multidimensional analysis space in relational database environments. Second, a typical and not normalized dimensional model has a simple schema structure, which simplifies the processing of query and improves performance.

In relation to the definition of the DW data model, the first step is the formulation of the data marts that are in a single data source initially. In the second step, the dimensions are identified for these data marts; and the intersection between them and their dimensions are marked. Next, the granularity of the fact tables of these data marts must have to be declared. Also, it is verified which the dimensions are related to these tables. Finally, the facts that compose the fact tables are defined.

The final physical dimensional model is composed of four fact tables (gray color) and five dimensions (white color), as shown in Figure 2. In the design, we choose to use dimensions unnormalized, with many attributes, but that can be used more than once with different names in the cubes in the Data Cubes layer. This kind of dimension is called the role-playing dimension. For instance, we used the table dim_data to create a dimension called Census Year (Ano do Censo) and other called Year of Admission (Ano de Ingresso).
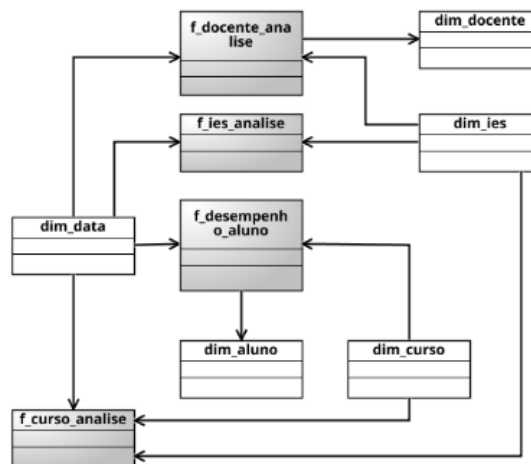

Figure 2. Dimensional Data Model.

3.1.4 Data Cubes and Data Analysis

Initially, our solution has used the Mondrian ROLAP server, which is an open source OLAP (online analytical processing) server, written in Java. The Mondrian receives and parses the MDX language into Structured Query Language (SQL) to retrieve answers to dimensional queries.

The Mondrian Server requires the Mondrian Schema that describes a logical model. It consists of cubes, hierarchies, and members, and a mapping of this model onto a physical model. The first task before creating the Mondrian schema was to define the cubes' granularity and to understand which dimensions were common to the cubes. We used the Bus matrix to define it. After that, the cubes, measures, calculated measures, dimensions, hierarchies, members, and levels were defined in order to create the Mondrian Schema, as shown in Figure 4 (it was used the Schema Workbench tool). The Mondrian Schema is composed of four cubes with the following granularity, number of dimensions, and number of measures:

- AlunoCurso - student per course by year, 32 dimensions and 28 measures.
- CursoAnalise - course by year, 21 dimensions and 40 measures.

- IESAnalise - university by year, 5 dimensions and 57 measures.
- DocenteInstituicao - professor by year, 20 dimensions and 49 measures.

In the Analysis Tools layer, we set the Pentaho Community tools to enable final users to make OLAP queries. The OLAP tools allow analysis and management, providing a great profit of performance, as fast access to a great variety of data views organized through the multidimensional database.

In a second moment, we migrate the solution to the Power BI tool, due to better performance to make the analysis. The data cubes were created in Power BI[4], providing the same analysis, and showing the proposed architecture is flexible, be possible to use different BI tools.

4   Results

In order to achieve the objective proposed in this work, this section presents the most relevant analyzes to answer the proposed research questions. It is worth mentioning that, for a better understanding of the results, many of the analyzes were reorganized, after the results generated by the tool and others went through simple statistical processes.

In addition, for the purpose of comparison and integration between databases, it is important to highlight some aspects. The ENADE database consists only of courses that participated in the ENADE test, even if it has not obtained a concept, described by (NC - No Concept). Thus, in 2017, 96 Computer Science Teacher Education Courses responded to the INEP census, while only 59 participated in the ENADE test. Therefore, for some analyzes, only the courses participating in the ENADE test were considered, while in others, all courses that responded to the Census were considered. Another important fact for the understanding of the analyzes, is that only one course in Computing obtained concept 1. So many analyzes have been disregarded, since we consider that a course does not reflect the panorama of this category.

The first results aim to contextualize the Computer Science Teacher Education Courses in comparison with other courses in the Computing area (Computer Science Teacher Courses, Computer Science, and Information Systems). The Table 1 shows the numbers of courses, students, enrollments, and dropout group by course type in the Computing area in the years 2017 and 2018. As can be seen in Table 1, the Computer Science Teacher Education degree has around 17% of the number of courses existing in Information Systems, and around 25% of the number of Computer Science courses. In the following tables, the items "Comp. (Lec)", refers to the term "Computer Science Teacher Education Course".

---

[4] https://powerbi.microsoft.com/

Table 1: The numbers of courses, students, enrollments and dropout group by course type in the Computing area in the years 2017 and 2018.

Course	2017				2018			
	N. Courses	N. of Students	Enroll. Total	Dropout (%)	N. Courses	N. of Students	Enroll. Total	Dropout (%)
All	1047	216.751	147.361	18,51%	1085	214.165	145.220	19,18%
Comp. Sci.	376	91690	63668	17,84%	396	93.887	64.486	19,11%
Comp. (Lec)	96	16.081	12.084	16,02%	100	16.650	11.707	20,81%
Inf. Systems	575	108.980	71.609	19,45%	589	103.628	69.027	19,23%

Table 2: The numbers of on-site and e-learning courses group by course type in the Computing area.

Course	2017			2018		
	N. of Courses	on-site	e-learning	N. of Courses	on-site	e-learning
All	1053	1012	41	1195	1138	56
Comp. Sci.	376	375	1	396	393	3
Comp. (Lec)	96	71	25	100	73	27
Inf. Systems	575	560	15	589	568	21

Also, through Table 1, it is observed that the dropout rate is similar between courses. However, the dropout rate in Computer Science Teacher Education Courses increased more than in the two other courses, between the 2017 and 2018 census, from the lowest dropout rate to the highest. On the other hand, through Table 2, it is possible to observe that the teacher education courses, despite being fewer in number than the Information Systems and Computer Science courses, have a greater number of distance courses.

Table 3 presents the consolidated data of the computer courses related to the data of the evaluation of the courses of 2017, showing, within the total number of graduates, the percentage that took the test. Through Table 3, the courses with the higher concept level, had a higher number of students who took the test. Finally, Table 4 shows the distribution of courses in each ENADE concept range, for the three computing courses. Through Table 4, it is observed that the Computer Science Teacher Education Degree has the highest rate of courses in the higher concepts (considering only the courses that participated in the evaluation of ENADE), reaching more than 45% of the courses in concept levels 4 and 5.

Table 3. Percentage of graduates who took the ENADE exam, per course.

Courses	ENADE Concept				
	1	2	3	4	5
Comp. (Lec)	-	81,19%	76,67%	89,98%	86,73%
Inf. System	87,13%	81,97%	83,36%	87,53%	93,41%
Comp. Sci.	82,42%	82,34%	84,70%	90,58%	89,40%
All courses	84,77%	81,84%	81,58%	89,37%	89,85%

Table 4. Distribution of the courses according to the ENADE concept for the year 2017.

Enade Concept	Course		
	Computer Science	Computer Lecturing	Information Systens
1	5,47%	1,69%	3,82%
2	30,55%	16,95%	29,72%
3	30,87%	28,81%	41,19%
4	22,51%	28,81%	19,11%
5	9,32%	16,95%	2,55%
NC	1,29%	6,78%	3,61%

* NC - Courses that participated but did not obtain a concept

In the Tables, the item "Computer Lecturing", refers to the term "Computer Science Teacher Education Courses".

4.1 Data from the Computer Science Teacher Education Courses

This section is dedicated to presenting specific results referring to the Degree in Computer Science Teacher Education Courses, in order to aid identifying factors that may be directly related to the classification of the course in the ENADE concept level.

Table 5 shows the distribution of courses in each ENADE concept range, considering different characteristics of the HEIs. Table 5 highlights that public HEIs present a large part of their courses in the higher concepts, Federal HEIs, which present almost 60% of courses in concepts levels 4 and 5.

Regarding the teaching method, both classroom and distance courses have more than 45% of the courses in concepts levels 4 and 5. However, distance courses have more than one third of their courses in concept level 2. Considering the region in which the course is located, the South region stands out, where around 90% of courses in this region are classified in concepts 4 or 5. As a negative point, the Midwest region presents only 9% of courses in concepts 4 or 5 and the North region presents only 33% of the courses in these concepts.

Table 5. Distribution of Computer Science Teacher Education Courses by ENADE concept according to Institution Type, Teaching Method and Location.

Analysis	Category	ENADE Concept					
		NC	1	2	3	4	5
Type of Institution	Private for-profit	0,00%	33,33%	33,33%	33,33%	0,00%	0,00%
	Private non-profit	10,00%	0,00%	20,00%	20,00%	20,00%	30,00%
	State Public	0,00%	0,00%	14,29%	50,00%	28,57%	7,14%
	Federal Public	9,38%	0,00%	9,38%	21,88%	40,63%	18,75%
Teaching method	E-learning	0,00%	0,00%	33,33%	16,67%	33,33%	16,67%
	On-Site	8,51%	2,13%	12,77%	31,91%	27,66%	17,02%
Region	North	0,00%	0,00%	22,22%	44,44%	33,33%	0,00%
	Northeast	5,56%	0,00%	11,11%	33,33%	44,44%	5,56%
	Southeast	16,67%	0,00%	16,67%	16,67%	33,33%	16,67%
	South	0,00%	0,00%	0,00%	11,11%	22,22%	66,67%
	MidWest	9,09%	9,09%	36,36%	36,36%	0,00%	9,09%

Table 6 shows the consolidated results of different metrics of the CPC, according to the ENADE concept level of the courses. Among the metrics presented, those in dark grey are highlighted. It is noticed that, although the average enrolment rate to participate in the ENADE test is above 75% in all concept levels, we see a big difference in the average rate of students who took the test and obtained a grade. In concept levels 4 and 5, on average, more than 75% of enrolled students participated and obtained a grade, while in concept levels 2 and 3 this rate was less than 60%.

Most of the metrics results as expected, and the courses with the higher concept levels present better grades (IDD, ENADE, FG, CE and CPC). However, the didactic and infrastructure grade is not related to the ENADE concept (the average for these metrics was higher for courses with concept level 2 than courses with concept level 4).

On the other hand, we see that all metrics that involve professors are strongly related to the concept level. Courses with higher concept levels, on average, have a higher number of professors, and them are more qualified. The grade for doctors stands out, which for courses with concept levels 4 and 5, is approximately 60% higher than courses with concept levels 3 and 2.

Table 6. Different CPC metrics according to the ENADE concept.

Analysis	Concept 5			Concept 4			Concept 3			Concept 2			NC		
	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.	Mean	Median	Std. Dev
Graduates	28,30	20,50	36,32	21,35	19,00	13,37	27,82	21,00	18,71	29,70	34,00	16,20	12,50	14,50	7,59
Enrolled	20,30	15,00	22,08	19,06	17,00	12,27	20,41	18,00	13,98	23,10	26,00	11,85	0,25	0,00	0,50
Enrolled (%)	86,73%	100,00%	0,18	89,98%	90,74%	0,12	76,67%	82,86%	0,18	81,19%	84,13%	0,14	-	-	-
Enrolled by grade	12,67	14,00	8,03	15,82	13,00	11,18	12,75	10,50	9,30	14,33	12,00	11,20	-	-	-
Enrolled by grade (%)	76,59%	92,86%	0,30	83,75%	94,74%	0,23	58,64%	60,56%	0,29	49,35%	42,86%	0,26	-	-	-
FG grade	59,78	59,91	6,41	55,33	55,36	5,53	46,67	48,10	5,48	40,88	39,09	5,15	-	-	-
CE grade	50,15	47,36	9,25	42,43	42,94	2,79	34,68	35,01	2,39	27,20	26,98	1,23	-	-	-
ENADE grade	4,23	4,11	0,31	3,56	3,65	0,28	2,52	2,62	0,28	1,60	1,64	0,19	-	-	-
IDD grade	3,59	1,41	5,44	0,80	0,59	1,30	-0,99	-0,51	1,22	-1,44	-1,29	3,08	5,61	5,73	0,39
Didactic grade	5,22	5,29	0,50	5,03	5,07	0,25	4,95	5,11	0,59	5,05	4,97	0,30	5,52	5,64	0,43
Infrastructure grade	5,08	5,01	0,57	4,78	4,77	0,36	4,63	4,72	0,76	4,82	4,81	0,32	25,75	24,50	18,25
Professor count	34,80	19,50	51,31	26,12	23,00	15,15	14,47	14,00	10,91	11,56	10,00	8,92	0,89	0,88	0,08
Master grade	0,96	0,98	0,05	0,88	0,91	0,12	0,76	0,74	0,19	0,75	0,80	0,23	0,42	0,39	0,20
Doctor grade	0,47	0,48	0,26	0,44	0,42	0,25	0,24	0,16	0,25	0,20	0,17	0,16	-	-	-
CPC	3,70	3,68	0,39	3,15	3,13	0,33	2,30	2,26	0,52	2,04	2,15	0,41	-	-	-
CPC Zone	4,10	4,00	0,57	3,65	4,00	0,49	2,82	3,00	0,64	2,70	3,00	0,48	-	-	-

Table 7 shows data from the courses, according to the ENADE concept. The dropout rate is not related to the ENADE concept of the course. It is observed public HEIs present most students enrolled, and the dropout rates are much lower than private HEIs. Regarding the teaching method, only in concept level 2, there are more students enrolled in distance courses than in classroom courses, an indication that the teaching method is related to the ENADE concept level. In addition, distance courses have a higher dropout rate than classroom courses. Finally, the average age of students enrolled in classroom courses is lower than the average age of students on distance courses.

Table 8 presents the data referring to students, according to the course ENADE concept in which they are enrolled. Among the data presented in Table 8, it is noteworthy that, in courses with concept level 5, there is an equivalent number of enrolled male and female students. In courses with concept level 5, the ethnicity with the largest number of students enrolled is white, unlike all other concept levels, which the mixed ethnicity concentrates most students enrolled.

Table 7. HEI data according to the ENADE concept level.

			2			3			4			5			NC			NP		
			Dropout (%)	Age (Avg)	Students Enrolled	Dropout (%)	Age (Avg)	Students Enrolled	Dropout (%)	Age (Avg)	Students Enrolled	Dropout (%)	Age (Avg)	Students Enrolled	Dropout (%)	Age (Avg)	Students Enrolled	Dropout (%)	Age (Avg)	Students Enrolled
Type of Institution	2017	Private	23,5%	34	7,6%	16,2%	31	20,0%	27,4%	28	1,9%	29,8%	34	24,2%			–	22,8%	31	20,0%
		Public	9,8%	30	92,4%	16,2%	30	95,0%	16,7%	30	98,1%	17,5%	25	75,8%	23,4%	29	100,0%	12,6%	30	86,1%
	2018	Private	56,8%	33	6,8%	28,8%	31	19,1%	63,2%	30	2,2%	30,5%	35	26,2%			–	23,9%	31	20,5%
		Public	22,6%	31	93,2%	19,4%	30	97,7%	21,6%	30	97,8%	19,3%	25	73,8%	30,2%	29	100,0%	15,6%	30	88,1%
Teaching Method	2017	e-learning	11,8%	33	68,6%	24,8%	34	18,5%	15,5%	33	30,3%	31,1%	35	20,3%	–	–	–	12,5%	31	76,6%
		on-site	13,7%	27	31,4%	15,5%	29	81,5%	17,6%	29	69,7%	17,8%	25	79,7%	23,4%	29	100,0%	18,9%	28	23,4%
	2018	e-learning	35,5%	33	62,2%	29,6%	35	15,7%	25,8%	33	25,7%	29,6%	36	24,0%	–	–	–	16,5%	32	71,3%
		on-site	12,9%	27	37,8%	19,0%	29	84,3%	22,9%	28	74,3%	20,0%	25	76,0%	30,2%	29	100,0%	18,4%	27	28,7%

Table 8. Data on students according to the ENADE concept.

| | | | ENADE Concept | | | | | | | | | | | | | | | | | | |
| | | | 2 | | | 3 | | | 4 | | | 5 | | | NC | | | NP | | |
			Dropout (%)	Age (Avg)	Students Enrolled	Dropout (%)	Age (Avg)	Students Enrolled	Dropout (%)	Age (Avg)	Students Enrolled	Dropout (%)	Age (Avg)	Students Enrolled	Dropout (%)	Age (Avg)	Students Enrolled	Dropout (%)	Age (Avg)	Students Enrolled
Gender	2017	Fem.	11,0%	31	41,9%	14,8%	31	35,2%	16,9%	30	29,7%	20,8%	28	19,6%	30,1%	30	37,0%	13,5%	30	33,8%
		Male	13,3%	31	58,1%	18,6%	30	64,8%	17,0%	30	70,3%	20,9%	27	80,4%	19,0%	29	63,0%	14,4%	30	66,2%
	2018	Fem.	27,0%	31	40,6%	23,1%	30	32,9%	25,2%	30	28,6%	22,2%	28	18,3%	27,7%	29	41,4%	17,3%	31	33,0%
		Male	29,0%	31	59,4%	19,6%	30	67,1%	23,3%	30	71,4%	22,4%	28	81,7%	31,8%	29	58,6%	17,0%	31	67,0%
Race	2017	Yellow	5,0%	33	1,2%	7,1%	30	1,4%	25,4%	29	1,6%	25,0%	30	1,8%	0,0%	34	1,0%	7,8%	31	1,4%
		White	6,9%	32	33,2%	18,5%	31	22,3%	17,8%	30	24,4%	20,8%	28	50,2%	12,5%	28	17,2%	15,7%	31	33,7%
		Indigenous	38,1%	30	1,1%	20,0%	23	0,3%	14,3%	27	0,6%	20,0%	27	0,2%	–	22	0,5%	21,6%	32	0,4%
		Not Declared	24,0%	32	11,3%	20,1%	29	25,3%	21,4%	28	20,3%	26,3%	24	12,2%	18,8%	30	6,3%	9,5%	32	18,6%
		Brown	12,3%	30	37,6%	16,5%	30	39,9%	13,9%	30	44,3%	18,3%	28	29,3%	31,8%	29	36,5%	14,5%	29	37,8%
		Black	8,2%	29	15,5%	11,1%	30	10,8%	17,1%	30	8,8%	18,8%	29	6,3%	20,0%	29	38,5%	15,8%	29	8,0%
	2018	Yellow	39,1%	32	1,4%	12,5%	33	2,7%	14,3%	28	1,4%	23,1%	33	1,4%	66,7%	29	0,6%	6,9%	31	2,0%
		White	29,0%	32	30,3%	22,4%	31	20,5%	24,3%	30	27,1%	21,2%	28	47,6%	33,3%	30	11,7%	24,0%	32	27,4%
		Indigenous	13,3%	29	1,4%	18,2%	25	0,6%	30,0%	28	0,6%	50,0%	27	0,3%	0,0%	23	0,6%	14,9%	31	0,6%
		Not Declared	40,4%	33	11,0%	22,5%	29	18,2%	31,1%	30	15,1%	21,0%	24	16,2%	20,0%	33	4,3%	15,4%	30	21,6%
		Brown	25,9%	30	38,3%	19,9%	30	46,9%	20,9%	30	47,4%	22,2%	29	28,9%	36,4%	29	40,7%	13,4%	30	40,2%
		Black	22,1%	29	17,6%	20,2%	30	11,1%	24,9%	29	8,4%	32,8%	30	5,7%	21,6%	29	42,0%	15,9%	30	8,3%
High Scholl	2017	Private	12,6%	34	10,5%	21,4%	29	18,0%	14,9%	32	38,3%	20,9%	27	36,5%	22,2%	27	10,4%	11,0%	32	36,6%
		Public	12,3%	30	89,4%	16,1%	30	80,4%	18,1%	29	60,6%	21,0%	28	62,0%	23,5%	29	89,6%	16,0%	29	58,7%
		Not Declared	33,3%	40	0,1%	24,4%	34	1,6%	25,0%	34	1,1%	13,6%	32	1,4%			–	13,3%	35	4,7%
	2018	Private	40,0%	35	8,7%	19,3%	29	20,4%	28,1%	32	34,2%	21,4%	28	42,0%	32,1%	26	10,5%	14,1%	31	35,5%
		Public	26,5%	30	91,3%	20,8%	30	78,7%	21,3%	28	65,2%	23,4%	28	57,1%	29,9%	29	89,5%	18,2%	30	61,8%
		Not Declared	–	–	0,0%	41,7%	37	0,9%	27,3%	36	0,6%	0,0%	–	0,9%	–	–	–	30,0%	36	2,7%

5 Discussion

From the data presented in Section 4, some discussions are raised in order to answer the research question proposed in this work. A first analysis concerns how the Degree Computer Science Teacher Education has its courses distributed among the ENADE concept levels. In Figure 3, among the courses in the computing area, Teacher Education has the highest rate of courses with the concept levels 4 and 5. However, through Figure 3, we observe that Computer Science Teacher Education has many courses that did not participate in the assessment, reaching 37% of the courses registered with INEP, that did not take the exam. Considering the courses that did not obtain a concept level, Computer Science Teacher Education presents more than 42% of the courses without an ENADE concept, which may influence the result of conclusions presented in this work.

Analyzing the data obtained, we found no evidence that the dropout rate of Computer Science Teacher Education Courses has any association with the ENADE concept level. What has been noticed is that public HEIs have a dropout rate lower than private HEIs and e-learning courses have evasion rates similar to on-site, both in public and private HEIs. However, comparing only public HEIs, regardless of the teaching method, dropout is not associated with the ENADE concept. This same analysis applies to private HEIs. It was not possible to find a pattern between dropout and the ENADE concept.
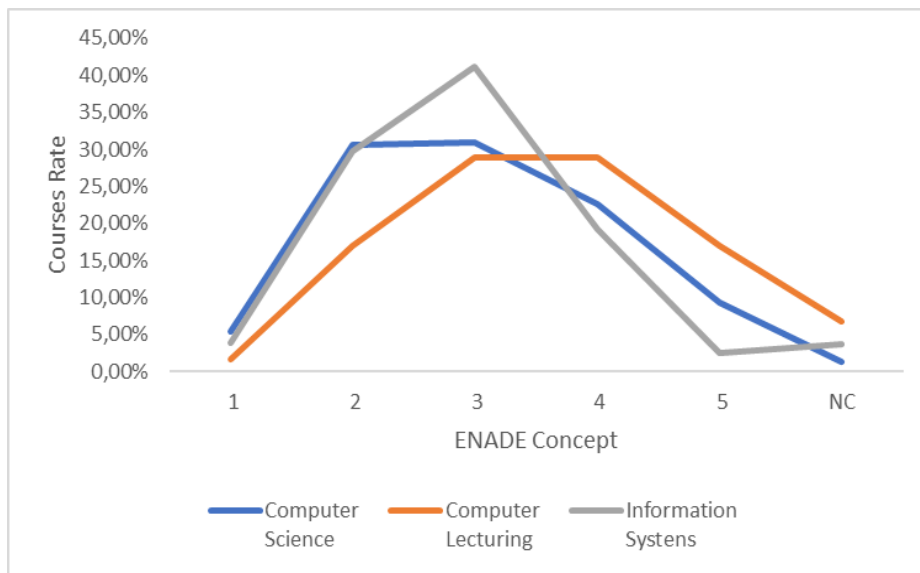
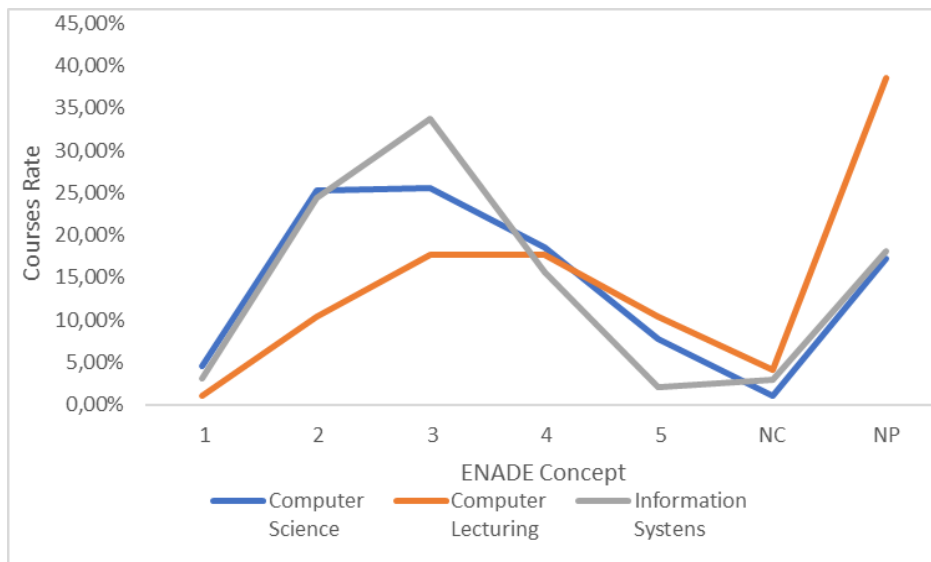Figure 3. Percentage of Courses by ENADE Concept.

Figure 4. Percentage of Courses by ENADE Concept, considering non-participating courses.
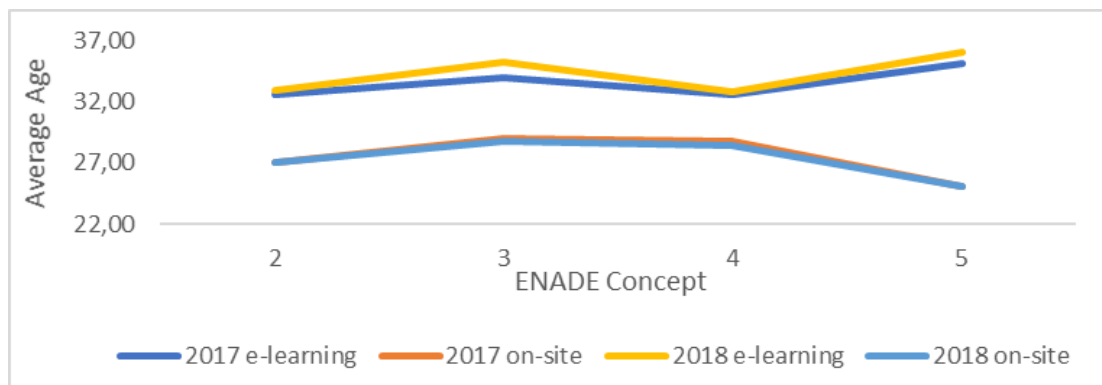

Figure 5. Average age of students according to the ENADE concept.

Considering the profile of the students, what can be observed is that the gender and type of high school that the student attended are associated with the ENADE concept level. Through

221

Figure 6, it is possible to see that as the concept level increases, the number of male students also increases. About secondary education, a similar phenomenon is also observed (Figure 7), the number of students who attended private high school increases as the concept level increases.

Analyzing the data presented in Table 8, courses with higher concept levels have most of their students enrolled from white ethnicity. However, looking at Table 5, the southern region concentrates almost 70% of the courses with concept 5. Thus, the concept level to ethnic relationship does not appear to be a cause-effect relationship, but a consequence of the courses with higher concepts being mostly located in the southern region, since according to IBGE (Brazilian Institute of Geography and Statistics), more than 78% of the population in the southern region is white.
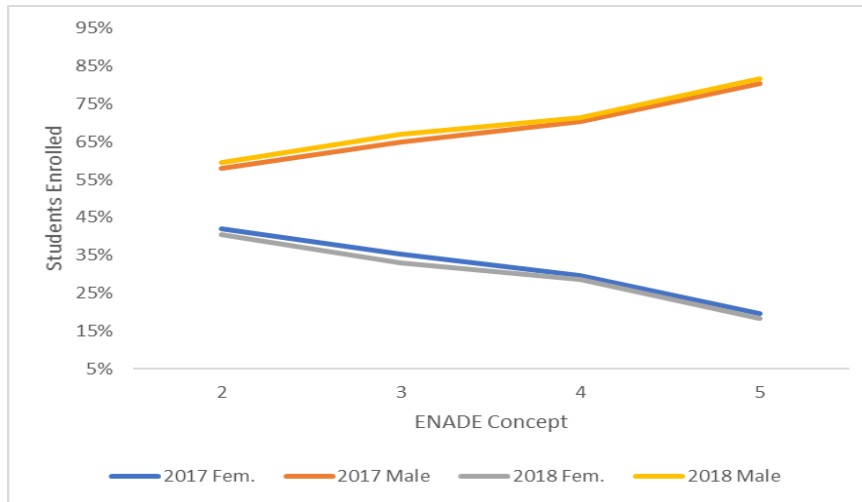
Figure 6. Percentage of students enrolled by gender and ENADE concept.
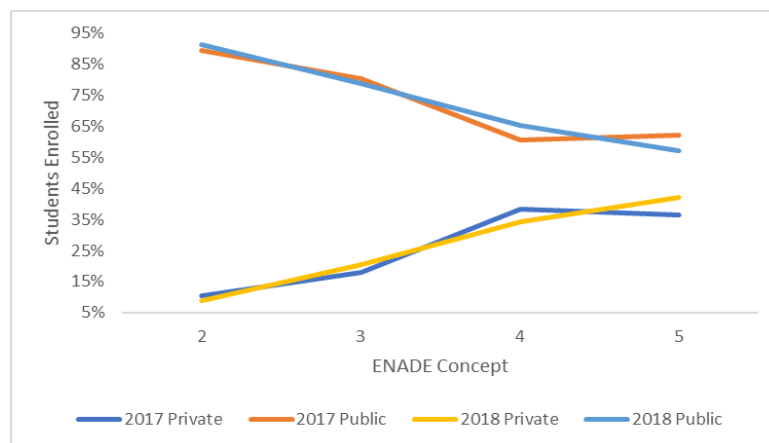
Figure 7. Percentage of students enrolled according to the high school and ENADE concept.

Analyzing Table 6, it was observed a possible relationship between some variables and the ENADE concept. In order to analyze more deeply some of these relations, it was used Multiple Correspondence Analysis (MCA). The MCA is the generalization of Correspondence Analysis (CA) to several categorical variables. The CA is a method of data visualization that is applicable to cross-tabular data such as counts, compositions, or any ratio-scale data where relative values are of interest (Greenacre, 2010).

To understand the relationship between professor's grades and the ENADE concept the Figure 8 presents the MCA analysis. Analyzing the Figure 8, it is possible to observe that the number of professors in the course, and their training there is a strong association with the concept level of the course. The courses with better ENADE grades present the best Ph.D. and master grades, and a bigger number of professors, this relation was also observed in the studies of Silva et al. (2015), Simm (2015) and Brito (2016).

Figure 8. Multiple correspondence analysis (MCA) of ENADE concept and number of professors and professor's grades of a course.

Another analysis using MCA is presented in Figure 9. The analysis presents the relation between and percentage of students who took the test and ENADE concept. It is possible observe that the percentage of students who took the ENADE there is an association with the concept level of the course. The courses with a better ENADE concept presented a bigger percentage of students who took the test.
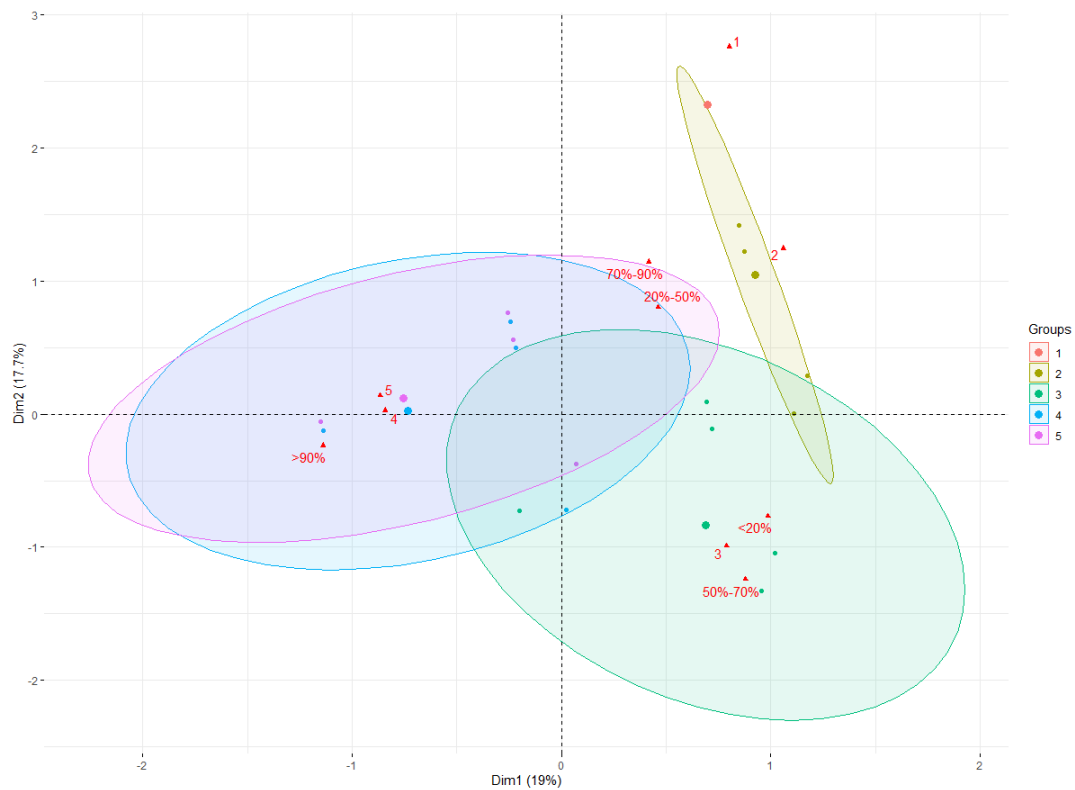
Figure 9. Multiple correspondence analysis (MCA) of ENADE concept and percentage of students who took the test.

Though Table 6, it is possible to observe, the didactic and infrastructure grade is not related to the ENADE concept. It was also observed in the study presented by Silva et al. (2017).

6   Conclusions

This work presented a data analysis using a BI-based method, in order to present an overview of the data of the Computer Science Teacher Education Courses in Brazil, with special emphasis on the ENADE concept level of the analyzed courses. Regarding the proposed research question "What factors may be related to the classification of Computer Science Teacher Education Courses at ENADE", we can classify the results in four categories of factors, using the obtained data:

- HEI: The location of the HEI and the type of HEI in the data analyzed were related to the concept level of the courses. HEIs in the southern region represent almost 70% of courses with level 5. Public HEIs have higher concept level than private ones, especially federal HEIs.

- Courses: the factor that most closely related to the concept level was the number of teachers and their training. Courses with a higher number of professors and higher degrees, showed better results.

- Students: Three factors were highlighted as more related to the concept level. Gender, the type of high school and age, the latter presented a weaker relationship.

- ENADE: the number of enrolled students who took the exam. Courses that obtained higher concepts, had a higher rate of students who took the exam.

Regarding the study, we point out three limitations. The first is related to the scope of the analysis, which considered only the last two years (2017 and 2018) of the census and only one year of ENADE (2017). The second refers to the high number of courses that are registered with INEP and that did not take the exam. The third is about the existence of different grading systems beyond ENADE, which were not used in this analysis. As a future work, we intend to expand the data analyzed and to apply machine learning techniques and statistics, to find relations not so evident and confirm the correlation found in this work.

Acknowledgements

References

Bell, T.; Witten, I. & Fellows, M. (2011). *Computer Science Unplugged: Ensinando Ciência da Computação sem o uso do computador*. 2011. Tradução coordenada por Luciano Porto Barreto. Disponível em: <https://classic.csunplugged.org/wp-content/uploads/2014/12/CSUnpluggedTeachers-portuguese-brazil-feb-2011.pdf>. [GS Search]

Brasil (2016). Ministério da Educação Conselho Nacional de Educação – Câmara de Educação Superior (Org.). *Resolução no. 5, de 16 de novembro de 2016*. Disponível em: http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=52101-rces005-16-pdf&category_slug=novembro-2016-pdf&Itemid=30192.

Brito T. (2016). "Corpo docente: fatores determinantes do desempenho discente no ENADE". Universidade de São Paulo - Faculdade de Economia, Administração e Contabilidade. *Master's Dissertation*. doi: 10.11606/d.12.2016.tde-21032016-115045 [GS Search]

Calixto, C. (2015). "Análise das causas de evasão discente no curso de Licenciatura em Computação: um estudo da UFPB virtual no formato UAB". *Revista Tecnologias na Educação*, 7 (12), 1-13. [GS Search]

Damasceno, I. and Carneiro, M. (2018) Panorama da Evasão no Curso de Sistemas de Informação da Universidade Federal de Uberlândia: Um Estudo Preliminar. *Anais do Xxix Simpósio Brasileiro de Informática na Educação (sbie 2018),* 1766-1779. doi: 10.5753/cbie.sbie.2018.1766 [GS Search]

Freitas, B.; Cosme, L and Nascimento, M. (2019). "Exame Nacional de Desempenho de Estudantes (ENADE): Análise do Perfil das mulheres dos cursos da área de computação". Anais do XIII Women In Information Technology. doi: 10.5753/wit.2019.6733 [GS Search]

Greenacre, Michael J. (2010). "Correspondence analysis". *WIREs Computational Statistics*, 2(5), 613-619. doi: 10.1002/wics.114 [GS Search]

Gregor, S. and Hevner, A. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *Mis Quarterly*, 37 (2), 337-355. doi: 10.25300/misq/2013/37.2.01 [GS Search]

Hevner, A.; March, S.; Park, J. and Ram, S. (2004). "Design science in information systems research". *MIS quarterly,* 28(1), 75–105. [GS Search]

Hinterholz. O; Valenzuela, G.; Silva, W. and Freire, A. (2014). "TRANSVERSAL: uma Abordagem Sistêmica para Predição de Desempenho de Estudantes no ENADE por meio do uso do Ciclo PDCA". *Anais do XXII Workshop Sobre Educação em Computação.* [GS Search]

INEP (2017). Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Nota Técnica N 12/2017/CGCQES/DAES.* Disponível em: http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2017/Nota_Tecnica_C GCQES_n12_2017_Calculo_da_nota_final_do_Enade.pdf.

Karpinski, J.; Mouro, N; Castro, M.; and Lara L. (2017) "Critical factors to the success of a course in distance learning: the perception of academics". *Revista da Avaliação da Educação Superior*, 22 (2). doi: 10.1590/s1414-40772017000200010. [GS Search]

Menolli, A.; Horita, F.; Dias, L. J. J.; and Coelho, R. (2020). "BI–based Methodology for Analyzing Higher Education: A Case Study of Dropout Phenomenon in Information Systems Courses". In: Proceedings of SBSI'20: XVI Brazilian Symposium on Information Systems. doi: 10.1145/3411564.3411636 [GS Search]

Moriconi, G. and Nascimento, P. (2014). "Fatores associados ao desempenho dos concluintes de Engenharia no Enade 2011. *Estudos em Avaliação Educacional*, 25(57). doi: 10.18222/eae255720142831 [GS Search]

Peffers, K., Tuunanen, T., Rothenberger, M., and Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal Of Management Information Systems*, *24*(3), 45-77. doi: 10.2753/mis0742-1222240302 [GS Search]

Prietche, S. and Pazeto, T. (2010). Estudo sobre a evasão em um curso de licenciatura em informática e considerações para melhorias. In: *Workshop De Educação Em Informática,* Anais do VIII do Workshop em Educação em Informática (WEIBASE). Maceió/AL. [GS Search]

Rocha, A.; Leles, C. and Queiroz, M. (2018). "Factors associated with the Nutrition students' academic performance in the Enade exam". *Revista Brasileira de Estudos Pedagógicos,* 99 (251). doi: 10.24109/2176-6681.rbep.99i251.3162 [GS Search]

Rodrigues, F.; Brackmann, C. and Barone, D. (2015). Estudo Da Evasão No Curso De Ciência Da Computação Da Ufrgs. *Revista Brasileira de Informática na Educação*, 23(01), 97-104. doi: 10.5753/rbie.2015.23.01.97 [GS Search]

Silva, C. M; Sinay, M. C. F; Rezende, J. F. and Araújo, G A. (2015). "Fatores Determinantes Para O Desempenho Dos Alunos De Administração no ENADE". *In: XV Colóquio Internacional de Gestão Universitária. Available at* https://repositorio.ufsc.br/xmlui/handle/123456789/136145. [GS Search]

Simm, E. (2015). "A qualidade dos cursos de administração no brasil e no paraná mensurada através dos indicadores do enade". C*adernos da Escola de Negócios*, 1(13). [GS Search]

Song Y., Rowen, W., Medsker, C., & Ewen, E. (2001). An Analysis of Many-to-Many Relatioships Between Fact and Dimension Tabels in Dimensional Modeling. *In: Proceedings of International Workshop on Design and Management of Data Warehouses (DMDW'01)*, Zurich, Switzerland. [GS Search]

Yin, R. (2013). *Case Study Research: Design and Methods*. (5. Ed). Sage Publications. [GS Search]