Review article

Correlation analysis using teaching and learning analytics

P.A.N. Prestes [*], T.E.V. Silva, G.C. Barroso

Department of Teleinformatics (DETI), Federal University of Ceará (UFC), Fortaleza, CE, Brazil

ARTICLE INFO

ABSTRACT

Data analytics techniques have been gaining more space in the scientific environment with applications in various areas of knowledge, including education. This paper aims to analyse data taken from a questionnaire of the Organization for Economic Development Cooperation (OECD) given to teachers and school managers. In this questionnaire, school environment issues are assessed, specifically: school environment, professional development, school leadership, and efficient management. As a methodology, Teaching and Learning Analytics (TLA) was used, particularly correlation analysis, which enables the extraction of useful information from raw data, relating issues that interfere with the teaching and learning relationship, besides specific analysis of student learning. The results obtained about the school environment are not linear. They do not present moderate or a solid linear correlation, making it impossible to validate and integrate answers related to the statements of the themes and sub-themes chosen for this analysis. In this sense, the research found dichotomous observations that mirrored many controversies and insecurities, enabling considerations about possible school scenarios and their effective practices.

1. Introduction

Educational research focusing on the perception of teachers and principals is an important initiative for the school environment, which allows them to express their views on relevant issues in their daily lives, both in the classroom and in school management.

Regarding the mechanisms, it is known that the teaching and learning process is not a one-way street that emphasises only the unilateral problem of teaching or learning, but is a dialogic process between teaching and learning and concerns the student, the teacher, the pedagogical and methodological practices, and the institutional conditions, fulfilling all the requirements of an educational policy.

Thus, it is essential to collect data covering more qualitative aspects of school routine, besides students' proficiency measurements. In this sense, when analysing large amounts of data containing qualitative aspects that demonstrate the occurrence of a given phenomenon, we see ourselves overloaded with information that, if not accompanied by actions, hinders us from reaching conclusions that can indicate paths to the learner, teachers, pedagogues, and managers. This burden is likely to curb the improvement of practices that may allow new dynamics to bring those actors closer to a greater degree of satisfaction, possible retention, less overload and, consequently, improvement in teaching and learning.

Therefore, we can say that even if the focus is only to assess learners -for example, when reflections of what led them to have low performance are not observed-, the educational ecosystem stops being fed with variables that were not reflected and, consequently, there were no referrals to possible decision-making to equate the problems faced by it, leaving incomplete the cycle of a possible assessment.

At this point, several authors deal with the various approaches to the teaching and learning process and didactic-methodological issues, which apply to the teaching of any discipline. Among these authors, we can mention [1, 2, 3].

But to do so, it is necessary to constitute educational models that consider reflections arising from a learning analysis that promotes emerging articulations and requires a new look into the teaching as mentioned earlier and learning process.

In this article, we will analyse these themes and sub-themes raised in a contextual research questionnaire applied by the Teaching and Learning International Survey (TALIS) in Brazilian schools [4], to validate the respondents' answers in assertions dependent on other assertions of previous themes and sub-themes and that somehow, when answered, may reflect logical non-dichotomic reasoning.

The results were obtained by calculating Spearmam's rank correlation coefficient, then computed from Pearson's linear correlation coefficient, which identified moderate and dichotomous correlations between the

* Corresponding author.
  E-mail address: pnprestes@hotmail.com (P.A.N. Prestes).

themes involved. The above corroborates previous results found in the literature [5] and establishes a quantitative parameter of analysis on the quality of learning.

Table 2 shows the statements and their respective nomenclatures. The themes and their sub-themes selected for this analysis are presented below:

✓ Basic Information and Qualification;
✓ Professional Development
- You have participated in any induction (integration) activity.
- When you started working at this school, were the following activities part of your induction (integration)?
- Are you currently engaged in any mentoring activity as part of a formal arrangement at this school?
- During the past 12 months, have you participated in any of the following professional development activities?
✓ General teaching;
- Considering the teachers at this school, to what extent do you agree or disagree with the following statements?
- Regarding your teaching, to what extent are you able to do the following?

1.1. Objective

The present study aimed to expose phenomena, if any, for decision-making in a predictive way. Throughout its implementation and assessment, this practice allowed adjustments so that in the final assessment, this proposal could contemplate the refinement needed for potential and satisfactory delivery to those educational networks

Data analysis techniques to compare answers in search of greater fidelity and lower dichotomy, such as Pearson's and Spearman's correlation compared to the other techniques mentioned in the following section, were considered the most appropriate because they demonstrate how strong or weak the degree of relationship between the variables analysed is.

It should also be noted that the TLA concept was applied in this work to validate the analysis made by the OECD, which aims, among others, to address the issues inherent to the teacher to answer to gaps that may collaborate with teaching and learning.

Below we present the studies focused on literature review supporting this work.

2. Literature review

The intentions of practices to be considered for this integration are presented below. Note that part of them is found in the questionnaires applied to teachers and principals by INEP/OECD. The premises of Teaching Learning Analytics – TLA will also be presented and, finally, how the correlation coefficients proposed in the objectives of this work will be calculated.

2.1. Talis

The Teaching and Learning International Survey (Talis), coordinated by the (OECD)[1] aims to assess the teaching and learning environment and the working conditions of teachers and principals in schools. Within these conditions are the school environment, professional development, school leadership, management, among others. In Brazil, the OECD has a partnership with the National Institute of Educational Studies and Research Anísio Teixeira (Inep).

From the research results, after analysis for possible decision-making, innovative propositions for good public policies aimed at the profession of educators, managers, and the very teaching institution are expected.

This approach also aims to explain the differences in learning outcomes revealed by the Programme for International Student Assessment (Pisa)[2].

This procedure began in 2008, focused on the final years of elementary school, being executed in 24 countries and expanding its action plan to 48 countries to date.

In Brazil, this plan was also expanded, allowing through Inep, the research in high school, to broaden the reflection scenario more accurately to respond with more suitable instruments for possible decision-making.

The volume of information generated led the OECD to divide the analysis and dissemination of results into two stages: the first issue was released in June 2019, and the second issue was scheduled for March 2020, but due to the pandemic scenario, it was postponed again, so in this study, we used only the data from the first issue.

The issue provided by Inep is organised by themes and sub-themes, with their respective assertions, for better understanding. In this mode, we followed the same organisational reasoning but selected only those that could best represent the universe for analysis.

2.2. Teaching and learning analytics

Before defining Teaching and Learning Analytics - (TLA)[3], we should explain Learning Analytics (LA). LA promotes the conversion of raw educational data into useful information related to the teaching and learning process, aiming to provide means for possible decision-making by allowing information projections in a more user-friendly and intuitive way through graphs and multidimensional interfaces. Hence, it allows improvement in the learning process [4]. For [6], LA is an emerging field of study in which sophisticated analytical tools can be used to improve the learning process. According to [7], LA acts for better practices to transform educational data into useful information for decision making.

When the LA technique is applied, we observe a considerable volume of raw data that enables pattern identification of relationships and learning styles to be treated. In addition, this treatment allows us to extract new information that until then was implicit [8, 9, 10].

Nevertheless, regarding the teacher, growing gaps remained unanswered due to the non-applicability of techniques that could actually observe their interaction with their students, their self-assessment about their engagement and their methodology, and their proposal to deliver the project, among other questions. In response to this gap, to observe this integration, a new analytical strand was proposed, the *Teaching and Learning Analytics* - TLA, which is presented as a synergy between teaching and learning analysis. More specifically, the TLA advocates the need to obtain methods and tools to assess teachers and also use students' educational data for evidence-based assessment, reflection, and improvement of this process [10]. This synergy has been considered one of the major research challenges in the field of technology-enhanced education.

TLA is an emerging field in which analytical tools are used to improve learning and education. For the methodology applied in this work, the conceptual map (see Figure 1) in question allows us to visualise the mediation of the TLA between the relationships of the actors involved and their attributions. Table 1 shows the legend by relevance.

According to the conceptual map (Figure 1), we can observe the new analytical strand, the TLA, interacting with the supporting actors, teachers and students, through the virtual means of teaching and learning, and their data logs visualised through graphs, printed matter, and plots, trying to foster a new interpretation of teaching and learning while mining educational data and using treatment

---

Figure 1. Conceptual map.

Table 1. Legend of the conceptual map presented in Figure 1 by relevance.

	"*Teaching and Learning Analytics*" and "Learning Assessment Tool", the concepts focused on the work.
	Work essential concepts.
	Important concepts, but not essential to the work.
	Related concepts, but not important to the work.

techniques, such as Business Intelligence[4], Action Analytics, Academic Analytics, Recommendation System, Web Mining, Social Media Analysis, Web Analytics, customised Adaptative Learning for the proper analysis of the processes generated throughout the educational interactions with their respective practices, and, most importantly, assessing this integration.

Thus, the TLA will serve as a parameter for our proposal by presenting the essential integration concepts related to Students, Teachers, Virtual Learning Environments, Logs and Data, Information, Educational Data Mining, and Information Visualisation.

Finally, it should be noted that through the conceptual map (Figure 1) presented, the teaching and learning trajectory mentioned in the previous paragraph meets the OECD proposal of assessing the teaching process directed to teachers through Talis.

2.3. Correlation

Correlation is the relationship, or the dependence, between people, things, ideas, etc. As an analogy, in the statistics, we can say: interdependence between two or more variables [11, 12].

---

[4] Business Intelligence - is a generic term that includes the applications, infrastructure, tools and best practices that allow the access and analysis of information to improve decision making and performance.

It is often necessary to evaluate the degree of relationship between two or more variables. We can discover accurately how much one variable interferes with the outcome of another.

However, in some situations, the relationship between two variables is not linear, or one of them is not continuous, or the observations are not randomly selected. In those cases, coefficient alternatives should be applied. Among the various options, there are Spearman's Coefficient and Contingency Coefficient.

Correlation is a bivariate analysis that measures the strength of the association between two variables and the direction of the relationship between the measurements obtained. In terms of the strength of the relationship, the value of the correlation coefficient (r) ranges between +1 and -1. A value of $|\pm 1|$ indicates a perfect degree of association between the two variables, where the sign indicates the direction of this association; a + sign indicates a positive relationship and a - sign indicates a negative relationship. For this study, we used Pearson's and Spearman's correlation coefficients.

Pearson's correlation coefficient is the most used correlation statistic to measure the degree of relationship between linearly related variables. The point-biserial correlation is presented in Pearson's correlation Eq. (1), except that one of the variables is dichotomous.

$$\rho_{pb} = \frac{\overline{S}_p - \overline{S}}{\sigma} \sqrt{\frac{\overline{p}}{\overline{q}}} \tag{1}$$

Notation:

$\overline{S}_p$ prepresents the average result of the test for those who hit the item; $\overline{S}$ is the general average result, σ is the general standard deviation $\overline{p}$ and $\overline{q}$ correspond, to the means of hit and error of the item, respectively.

It is also important to note that for Pearson's correlation, both variables should be distributed according to the normal curve. In addition, other assumptions include linearity and homoscedasticity. Linearity assumes a linear relationship between each of the two variables, and

homoscedasticity assumes that data are distributed equally over the regression line.

Spearman's correlation is a non-parametric test used to measure the degree of association between two variables based on the rank of the vectors obtained by the averages. Spearman's rank correlation test does not make any assumptions about data distribution and is the appropriate correlation analysis when variables are measured on a scale that is at least ordinal. We then assume that the scores of one variable should be monotonically related to the other variable.

3. Methodology

In this section, we present the procedures for data collection. The materials available consist of questionnaires with appropriate and harmonic themes and assertions, i.e., sequencing in parts the respective themes, defined in the introductory part of this work and made available in Table 2. Throughout this work, portraits/excerpts of those files are presented when necessary to facilitate understanding.

3.1. Sample

To apply the analytical techniques in this study, we used the data from the Inep open database, where they were processed and made available in CSV format[5] to facilitate reading through the following tools and libraries:

3.2. Assessment

Throughout its implementation and assessment, this practice allowed adjustments, when necessary, so that this proposal could cover the refinement needed for the potential and satisfactory delivery to those educational networks in the final evaluation.

In this phase, according to the structure of the questionnare, the basic information of the schools and the basic information of the teachers were observed through the variables that include:

- Their complementary initial education;
- Their updating to attend specific projects of the school;
- Their degree of satisfaction with the school and as a teacher;
- The interferences from inside and outside school they work;
- In relation to their peers, to develop a colaborative work to reach the objectives of an internal Project and even a broader proposed by their teaching network, as well as federal and international programmes.

The analysis, in general, when completed, will allow us to observe the applicability of this proposal in school units to compare its performance index before and after the delivery of the project.

3.3. Techniques used

The Colab platform was used as a storage and treatment environment, which is made available and integrated by Google Drive.

For this representativeness, cleaning was initially considered baseline, considering the absence of data above threshold >0.6, the randomness of filled data making columns of low data volume insignificant for the analysis, and if applicable, filled with statistical data, avoiding outliers.

Then, we move to dimensionality reduction to avoid computational complexity, impractical visualisation, and a very complex model to analyse. This technique allowed us to visualise the proportion of absence of values, their low variance, and work on data normalisation.

The following steps were performed for the data collection:

✓ Identifying the data fields to be and correlations.
✓ Backup of the database in CSV format and uploading to a specific folder created on google drive.
✓ Identifying the main tables linked to the data to be handled while maintaining their integritytreated for the empirical analysis initially.
✓ Importing appropriate libraries for the analysis of statistical data.

Through empirical inferences, based on systematic observations, we handled variables that at first could exhibit a possible affinity with each other, or even those that most present dichotomies in their answers between their themes and subthemes, and their extrapolation when compared with other themes and subthemes that address other questions. However, they may or may not confirm a given affinity that could naturally reinforce, when there is no contrast, statements of respondents avoiding controversies in the understanding.

Initially, we used the dimensionality reduction technique to reduce unnecessary features because there are columns with missing and partially filled data where their thresholds [13, 14, 15] are well above the indicated limit ranging between 60-75%. Besides that, the data presented in this study include many independent variables that could cause computational complexity, impractical visualisation, unnecessary variables. As we continue, we also examined its low variance so that we can then analyse its correlation.

Feature is an element of a (variable) whose value is higher among the other variables, making it the most important [16].

The Pandas, Sklearn, DecisionTreeClassifier, Numpy, MatplotLib, Seaborn and Boken libraries were used for the analysis of the database.

4. Data analysis

For this ongoing work, we used the correlation analysis technique to validate the hypothesis of the subjective dimensions of the answers that had not yet been tested in the field on the regulation of the teaching and learning process, considering the non-linearity of the procedural assessment.

For this analysis, variables were separated by context, and when possible, by the affinity between themes and sub-themes, to be crossed later and have their correlation degrees observed.

To seek a better understanding of the questionnaire applied by the OECD, Table 2 presents the meaning of the terminologies that can subsequently assist in the presentation of graphs and figures during the analysis:

Initially, the types of data manipulated were verified, noting that most of them were typed as *"objects,"* jeopardising any analysis by correlation techniques, for example.

In this mode, the necessary changes were made in the types of data for "categorical" and "whole," which allowed us to work the normalisation process and also the processing by the correlation technique chosen.

5. Applying the correlation technique (Pearson and Spearman)

The seaborn method was used to better understand the analysis. Pearson's correlations are presented in Graph 1. Spearman's correlations are presented in Graph 2.

5.1. Applying the correlation technique (Spearman)

The variables analysed in this study are seen in orange in Graph 1 and Graph 2 populated around the strongest point of correlation due to their crossing with the variables related.

In Graph 2, which presents the appropriate Spearman correlation technique to treat ordinal qualitative variables, we observe that there were no significant changes that could allow us to infer something new not detected by Pearson's correlation method, as in Graph 1.

In the following information range ('TT3G03'... 'TT3G06D2′), teacher's *"Basic Information and Qualification,"* we find the assertions:

Table 2. Questionaire provided by OECD.

As dictionary available, we will analyse the variables###
###########Basic Information and Qualification ###############

IDTEACH	Teacher Code) (Indexer)
TT3G01'	Sex
TT3G03	Higher level of formal education completed
TT3G04	First teacher education course
TT3G06D1	Classroom practices (teaching or pedagogical practices) in some or all of the disciplines I teach
TT3G06G1	Teaching cross-curricular skills (e.g., creativity, critical thinking, problem solving
TT3G06H1	Use of ICT (Information and Communication Technology) for teaching
TT3G06I1	Student behaviour and classroom management
TT3G06J1	Monitoring of students' development and learning
TT3G06B2	Pedagogy (didactics) of some or all of the disciplines I teach
TT3G06C2	Pedagogy (didactics) in general
TT3G06D2	Classroom practices (teaching or pedagogical practices) in some or all of the disciplines I teach

################# Professional Development ###############

You have participated in any induction (integration) activity###

TT3G19A1	Yes, during my first job
TT3G19A2	Yes, at this school
TT3G19A3	No

When you started working at this school, were the following activities part of your induction (integration)?####

TT3G20A	Classroom courses/seminars
TT3G20B	Online courses/seminars
TT3G20D	Scheduled meetings with the principal and/or more experienced teachers
TT3G20E	Supervision by the director and/or more experienced teachers

########Are you currently engaged in any mentoring activity as part of a formal arrangement at this school?######

TT3G21A	I currently have a designated advisor to support me
TT3G21B	I act as a designated advisor to one or more teachers

During the past 12 months, have you participated in any of the following professional development activities?######

TT3G22C	Conferences on education where teachers and/or researchers present results of their research or discuss educational issues
TT3G22D	Formal qualification program (e.g. specialisation, master's, doctorate)
TT3G22G	Peer observation and/or self-observation and training as part of a formal school arrangement

#################### General teaching ###############

Considering the teachers at this school, to what extent do you agree or disagree with the following statements?######

TT3G32A	Most teachers in this school seek to develop new ideas for teaching and learning
TT3G32B	Most teachers in this school are open to change
TT3G32C	Most teachers in this school seek new ways to solve problems
TT3G32D	Most teachers in this school support each other in applying new ideas

Regarding your teaching, to what extent are you able to do the following?

TT3G34A	Make students believe that they can do well in schoolwork
TT3G34B	Help students value learning
TT3G34D	Control disruptive behaviour in the classroom
TT3G34E	Motivate students who show low interest in schoolwork
TT3G34H	Make students follow the rules of the classroom
TT3G34I	Calm a student who interrupts too much or makes too much noise
TT3G34J	Use a variety of assessment strategies

✓ teacher's education;
✓ practices in the classroom;
✓ intercurricular teaching and skills;

✓ use of ICTs;
✓ qualification on classroom management regarding student behaviour and learning;
✓ their general didactic-pedagogical practice.

They should be strongly correlated, but we observe a moderate aspect for weak correlations tending towards disharmony between them.

For greater emphasis on analysis, we take a shorter range, contained in this same range ('TT3G06B2'... 'TT3G06D2', which, by the affinity of the assertions that deal with the same interest, verifies how weakly to moderately correlated they are with each other. Nevertheless, we also observe that the assertions ('TT3G06D1'... 'TT3G06G1') - which should be grouped and, as they are contained in the same theme, should have a reasonable correlation-have a weak correlation that, by the standard, should be discarded, ending any possibility of following with the analysis, since the significance of ρ is >0.05.

Continuing the analysis, on the theme '*Professional Development,*' whose range corresponds to ('TT3G19A1'...'TT3G22G'), they are subdivided into four subthemes as follows:

✓ Did you participate in any induction (integration) activity?
✓ When you started working at this school, were the following activities part of your induction (integration)?
✓ Are you currently engaged in any mentoring activity as part of a formal arrangement at this school?
✓ During the past 12 months, have you participated in any of the following professional development activities?

We can observe that the assertions ('TT3G19A1')...('TT3G20E') are contained as a subrange we will analyse below:

This subinterval clearly demonstrates that the teacher practically did not participate in courses, seminars, more formal guidance for the induction (integration) process by the school where he/she works at the time of the research and that his/her follow-ups are quite unassisted by supervision, directors, and more experienced peers, revealing a weak to moderate negative correlation between them.

Advancing, in this same range of assertions, but in different subranges ('TT3G20A'... ('TT3G20E'), we see they are strongly correlated, but we realise a strong dichotomy when we observe that while the teacher states that his/her integration was effective at the time he/she started to work in that school, he/she contradicts this statement, as we see in the other statements previously analysed, reflecting the lack of monitoring and formal events for his/her better integration.

The other assertions within this range with their respective subthemes were not analysed due to their weak correlation, as their significance is not (<0.05). The literature [22] points to a non-linear relationship between these variables; thus, they cannot be captured by Pearson's and Spearman's linear correlation models, illustrated in graphs 1 and 2.

On the theme *"General Education,"* with its respective subthemes:

✓ Considering the teachers at this school, to what extent do you agree or disagree with the following statements?
✓ Regarding your teaching, to what extent are you able to do the following?

We have the intervals of the following assertions ('TT3G32A')... ('TT3G34J'). Their strong correlation states that most teachers in their current schools seek to develop new ideas for teaching and learning, and that they are open to change, seeking new ways of solving problems. Also, the correlation shows that there is mutual support in applying new ideas, according to the intervals of assertions related to the first subtheme. About the second subtheme, the correlation reveals that teachers make students believe that they can do well in schoolwork, help students value learning, control disruptive behaviour in the classroom, motivate students who show low interest in
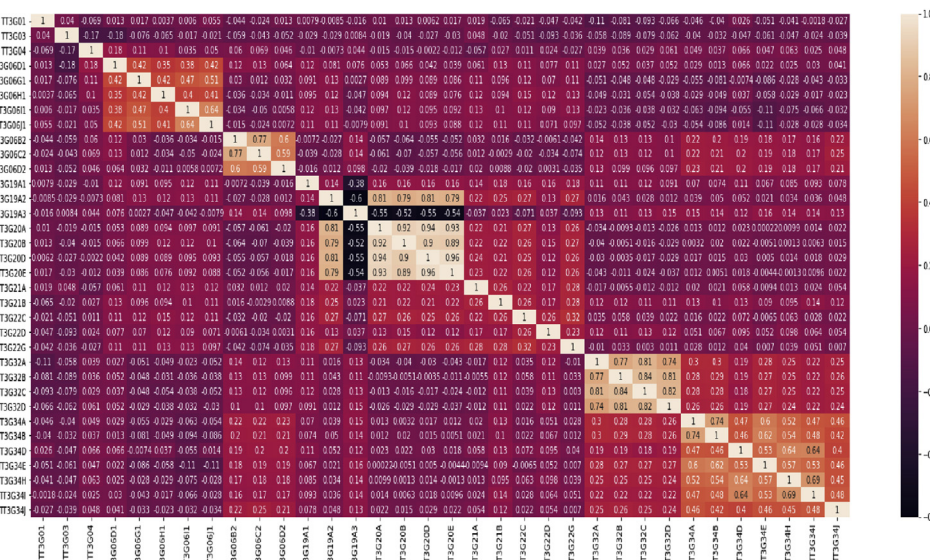
Graph 1. Pearson's Correlation
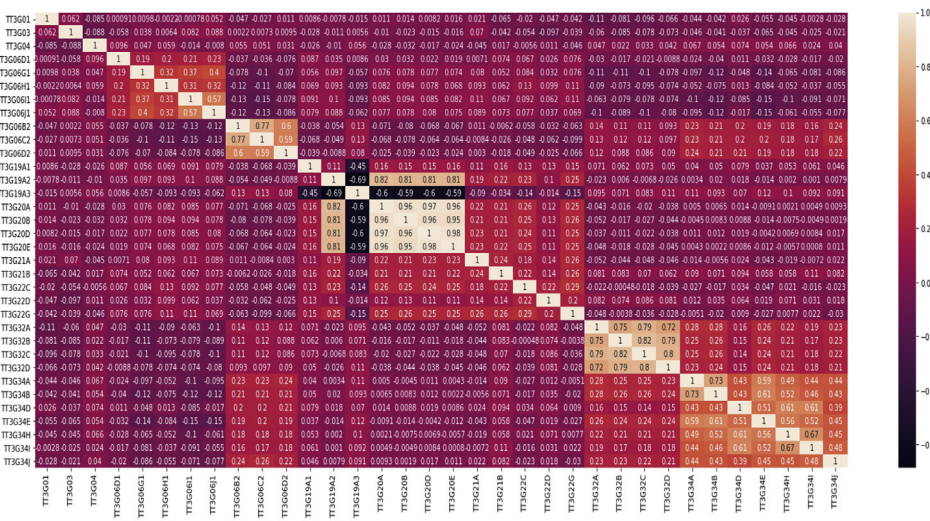
Graph 2. Spearman's Correlation

schoolwork, make students follow the rules of the classroom, calm a student who interrupts too much or makes too much noise, and use a variety of assessment strategies. We observed that both graphs have strong correlations, between 74% and 84% for Pearson's correlation and 72% and 82% for Spearman's correlation of the respondents' statements. However, when seeking a possible integration of the subthemes, the weak correlation between them is clearly mirrored, not allowing to infer any positive possibility aiming at better performance in the teaching and learning process, but only perceive the dichotomy that, by relying on the teacher, although well-intentioned, with specific purposes, is still a small action when phenomena that arise within the classroom disorganises any planning and action plan he/she idealised.

In fact, we enter here into the dichotomous process of analysis that allows us to emphasise how weak a correlation is with previous and subsequent themes, because we conclude that it is unreasonable to ensure that the assertions from the themes analysed make it possible to linearly and quickly identify the need to do things in different ways, to respond quickly to changes when necessary, to accept new ideas

promptly, and to readily provide support for the development of new ideas, and that they can in fact implement this set of actions that are more proactive, feasible, and quick in new referrals.

Thus, according to the analysis developed here, we observed that in this theme, just like in the previous themes, the behaviour of the process is identical, since it is worth mentioning again that its correlation with the other themes is also weak.

Notwithstanding, when we see Only the set of its assetives and its harmonic status, its correlations range from moderate to strong, without any novelty to add, as well as the other assertives.

This result once again corroborates the thesis that there is a strong dichotomous correlation between practices aimed at favouring teaching and learning, and if we want to analyse more data, there is a strong indication that it is very unlikely that the assertions, in harmony with their themes and subthemes, will provide, in practice, favourable actions for the correct performance of teaching and learning.

To ensure such statements and better understand qualitatively the analysis mentioned before, we used the technique of important features.

We present Graph 3 to confirm the proper analysis, where the most important features of all the assertions of the process whose themes contain the assertions harmonically are observed.

Note that by observing Graph 3 again, randomly, without delimiting themes or specifying given assertions, its degree of importance is greatly weakened, allowing discarding without much effort, except when one wants to demonstrate the dichotomous degree in its correlations.

To contribute to a better understanding of the relationships between the variables, a weak linear correlation can be observed between them when we cross their respective themes and sub-themes, which allows us a holistic view of the whole process by considering teaching and learning in general, which requires considerations of public policies, managers, and actors involved in the teaching and learning process.

Indeed, through the applied correlation methods and important features, such as the Chi-2 method, the results analysed show that the linearity of the process was compromised due to the dichotomy and the weakening of the relationships between its themes.

Other techniques such as RandonForestClassifier, f_classif, Sklearn. features, and Recursive Feature Elimination – RFE were not included in this work, as it could become quite extensive. However, they served as reinforcement for the analysis, where we can clearly perceive their convergence in determining that the important features in all their rounds are weak.

However, it is interesting to emphasise that the Chi-2 is the most suitable method for categorical and whole data.

Completing the analysis, the applicability of the Chi-2 method alone is presented. The other techniques, together with the technique chosen to specify this work better, can be visited on the dataset (see Table 3): Nery Prestes, Pedro Alexandre (2021), "Correlation Analysis using Teaching and Learning Analytics", Mendeley Data, V1, doi: 10.17632/ 5yb5r8bz98.1.

Having said that, by the practices exercised in this work, we can perceive how much attention is needed to see the gaps that could not be seen without the use of statistical techniques and methods [17, 18, 19, 20]. We also realised that now they can be part of new casts, and thus give greater robustness to possible decision-making.

To conclude this chapter, it is important to emphasise that at no time was a comparison made between the correlation methods and the classification methods; they only served to assess the degree of importance of the features, enabling us to affirm that the variables treated are, in fact, weakened.

6. Final considerations

Regarding the analysis techniques, this work used the correlation methods (Pearson and Spearman) and selection through feature techniques as defined above, and, from our point of view, Chi-2 as the most appropriate due to the types of data to be analysed.
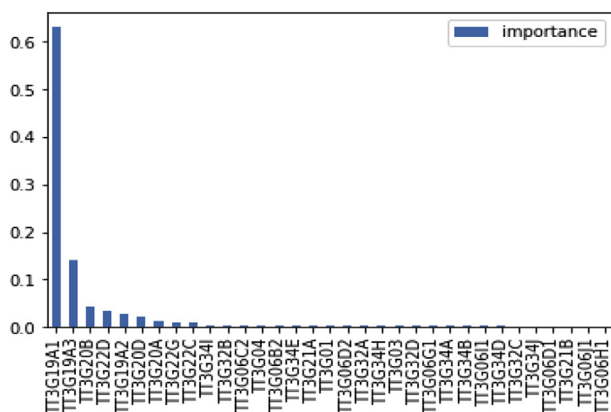
Graph 3. Most important features of the whole set of assertions represented by their respective themes.

Table 3. Sample of the extracted dataset for this work.

IDTEACH	TT3G06D1	TT3G19A1	TT3G19A3	TT3G21B
300102	0	0	1	0
300103	0	1	1	0
300104	0	1	0	1
300105	0	1	0	1
300106	0	1	0	1
...	...	...	...	...
319223	0	0	1	1
319224	0	1	0	1
319225	0	1	0	1
319226	0	1	0	1
319228	0	1	0	1

2047 rows × 4 columns.

As a data source, we used the data made available by Inep in partnership with the OECD, structured and made available in CSV standard, accompanied by a library and appropriate attachments for reading and interpretations related to themes and subthemes not limited to learning, but to teaching and its public policies to respond to possible concerns related to low performance and its probable causes on the teacher's side.

For this work, we considered the Teacher and Learning Analytics - TLA methodology, which allows analysing the process as a whole, taking into account the actors involved within the synergy of the teaching and learning process.

Therefore, for further analyses, it will be necessary to continue separating the actors by their respective specificities, through a new concept illustrated by the *Teaching and Learning Analytics* conceptual map, as shown in Figure 1, to enable new interpretations of the process of teaching and learning, since actors with different profiles and practices stand out and should be treated and analysed separately and integrated, where possible, for a better understanding of the educational process.

Regarding future work, this study serves as a baseline for analysis of other rounds of data made available by institutions that exercise the same premises and contexts of holistic interpretations of teaching and learning and, when possible, proposing a comparative analysis.

When Inep releases the final part of Talis, we will apply the statistical techniques again to observe the degree of satisfaction and the possible student retention methods related to the teaching and learning process. For this, elements that will collaborate with the possible methods and technological resources announced in this work will be analysed.

Another issue to be highlighted and that can be examined is the teacher's overload generated at the time of the regulation of the teaching and learning process. This point of analysis is of paramount importance to explain and detail the process of teaching and learning in a segregated and holistic manner, with elements that would hardly be explored without suitable methodologies, techniques, and analytical tools.

References

[1] MOHD Faiz Hilm– Disruptive Innovation in Education: Open Learning, Online Learning, MOOCs and what Next?, 2016. Avaliable at: http://www.ijhssi.org/papers/v5(10)/version-3/H5103049053.pdf.

[2] OECD, TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners, TALIS, OECD Publishing, Paris, 2019.

[3] OECD, TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners, TALIS, OECD Publishing, Paris, 2019.

[4] M. Taylor, et al., Teacher professional leadership in support of teacher professional development, Teach. Teach. Educ. 27 (2011) 85–94.

[5] EUROPEAN COMMISSION, Teachers' Professional Development: Europe in International Comparison: an Analysis of Teachers' Professional Development Based on the OECD's Teaching and Learning International Survey (TALIS), Office for Official Publications of the European Union, Luxembourg, 2010.

[6] Sergis Stylianos, Teaching and Learning Analytics to Support Teacher Inquiry: A Systematic Literature Review, 2017. Available at, https://goo.gl/iBqu5A. (Accessed 11 March 2018).

[7] L. Lockyer, E. Heathcote, S. Dawson, Informing pedagogical action: aligning learning analytics with learning design, Am. Behav. Sci. 57 (10) (2013) 1439–1459.

[8] M.A. Chatti, A.L. Dyckhoff, U. Schroeder, H. Thus, A reference model for learning analytics, Int. J. Technol. Enhanc. Learn. (IJTEL) 4 (5) (2012) 318–331.

[9] Anna Lea Dyckhoff, et al., Design and implementation of a learning analytics toolkit for teachers, Educ. Technol. Soc. 15 (3) (2012) 58–76.

[10] GIST Inference Scores Predict Cloze Comprehension "In Your Own Words" for Native, Not ESL Readers Health Communication, 2021.

[11] A.B. Costello, Jason Osborne, Best practices in exploratory factor Analysis: four recommendations for getting the most from your analysis, Practical Assess. Res. Eval. 10 (2005) 1–9.

[12] C. Spearman, Demonstration of formulae for true measurement of correlation, in: American Journal of Psychology, 1907. Available at: http://human-nature.com/nibbs/03/spearman.html. (Accessed 30 April 2020).

[13] T.E.V. Silva, F.H.L. Vasconcelos, A.L.F. Almeida, J.C.M. Mota, Multivariate analysis for the students' evaluation of teaching effectiveness in teleinformatics engineering, in: IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE), Hong Kong, 2012. H1A–1 – H1A–6.

[14] J.H.F. Meyer, R. Land, Threshold concepts and troublesome knowledge: linkages to ways of thinking and practising within the disciplines, in: ISL10 Improving Student Learning: Theory and Practice Ten Years on, Oxford Brookes University, 2003, pp. 412–424.

[15] J. Meyer, R. Land, Threshold concepts and troublesome knowledge (2): epistemological considerations and a conceptual framework for teaching and learning, High Educ. 49 (3) (2005) 373–388.

[16] M.L.A. Launius, Christie. Threshold, Concepts in Women's and Gender Studies : Ways of Seeing, Thinking, and Knowing, Routledge, New York, NY, 2015.

[17] Rodrigues, Célio Fernando de Sousa, Lima, Fernando José Camello de e Barbosa, Fabiano TimbóImportance of using basic statistics adequately in clinical research☆ ☆ Study performed at Universidade Federal de Alagoas, Revista Brasileira de Anestesiologia [online] 67 (2017) 619–625, n. 6 [Acessado 18 Junho 2021].

[18] E.S. Reis, M.T. Angeloni, F.R. Serra, Business Intelligence como Tecnologia de Suporte a Definição de Estratégias para a Melhoria da Qualidade de Ensino. Informação Sociedade: Estudos 20, 2010 (3). Recuperado de, https://periodicos.ufpb.br/ojs/index.php/ies/article/view/3783.

[19] R.S.J. Baker, S. Isotani, A.M.J.B. Carvalho, Mineração de Dados Educacionais: Oportunidades para o Brasil, Revista Brasileira de Informática na Educação, 2011. V.19, N.02.

[20] C.M. Romero, S. Ventura, E. Garcia, Data mining in course management systems: moodle case study and tutorial, Comput. Educ. 51 (2008) 368–384. Elsevier.