

Impact of teaching action on student interaction in virtual learning environments: canonical correlation analysis

David Benko Iseppon¹, Alex Sandro Gomes¹, Rodrigo Lins Rodrigues², Jorge Luis Cavalcanti Ramos³, João Carlos Sedraz Silva⁴

¹Centro de Informática – Universidade Federal de Pernambuco – UFPE

²Departamento de Educação – Universidade Federal Rural de Pernambuco – UFRPE

³Colegiado de Engenharia da Computação – Universidade Federal do Vale do São Francisco – UNIVASF

⁴ Colegiado de Engenharia Civil – Universidade Federal do Vale do São Francisco – UNIVASF

{dbi,asg}@cin.ufpe.br, rlr@ded.ufrpe.br, {jorge.cavalcanti, joao.sedraz}@univasf.edu.br

Abstract. *Some studies show that in a Virtual Learning Environment (VLE), it is important that the teacher adopts a proactive stance and mediating, managing the virtual classroom in order to integrate their students as a virtual learning community. These studies were conducted only from a qualitative point of view. The objective of this study is to verify the effectiveness of the stance of those involved under quantitative point of view, using distance undergraduate data provided by a public University. It was applied to Canonical Correlation Analysis technique for statistical data analysis. In the results of this study were identified database variables that representing the actions of the teacher and the interactions of students in the virtual environment. In addition, the results showed that the teacher's actions have significant influence on the interaction of the students in the VLE.*

Resumo. *Estudos apontam que em um Ambiente Virtual de Aprendizagem (AVA), é importante que o professor adote uma postura proativa e mediadora, gerindo a sala de aula virtual de forma a integrar seus alunos como uma comunidade virtual de aprendizagem. Tais estudos foram conduzidos apenas do ponto de vista qualitativo. O objetivo deste trabalho é verificar a eficácia da postura dos envolvidos do ponto de vista quantitativo, utilizando dados de cursos de graduação à distância fornecidos por uma IES pública. Em seguida, foi aplicada a técnica de Análise de Correlação Canônica para análises estatísticas dos dados. Como resultados deste trabalho foram identificadas variáveis no banco de dados que representavam as ações do professor e as interações dos alunos no ambiente virtual. Além disso, os resultados mostraram quais as ações do professor possuem bastante influência na interação de seus alunos no AVA.*

1. Introduction

Virtual learning environments (VLEs) are defined as computing environments that allow interaction and information exchanges with other participants and provide access to a wide range of resources [Wilson 1996]. A VLE allows students to access ubiquitously different learning tools, course information, content, discussion forums, document sharing and messaging between students and teachers [Martins *et al.* 2004]. The VLEs make possible the achievement of higher levels of control to the student during the whole learning process, providing an opportunity to restructure the learning experience in ways that are not feasible by the traditional processes [Piccoli *et al.* 2001].

In general, teachers are the main actors in any learning environment [Webster *et al.* 1997]. Studies have indicated that teachers who readily respond to students' questions have a significant influence on the satisfaction of the students [Arbaugh *et al.* 2002]. The logic behind this information is based on the fact that when students face problems in a course in a VLE, readiness in the teacher's assistance encourages students to continue learning [Thurmond *et al.* 2002]. The teachers' positive attitudes with regard to students adopting an interactive teaching style through technology produces an increasing effectiveness of learning by students [Piccoli *et al.* 2001].

In this paper, we verify the effectiveness of the posture of those involved from a quantitative perspective, through data mining techniques, using data from distance undergraduate courses in a VLE. In Section 2 we present the literature about the impact of teaching action on student interaction in virtual learning environments. In Section 3 we present the canonical correlation analysis technique that was performed, as statistical analysis technique in the present study. In Section 4 we present the methodology adopted. In Section 5 we present the results and in Section 6 a discussion about the results.

2. Impact of teaching action on student interaction in virtual learning environments

The interaction between participants in a VLE plays an important role in promoting effective learning, allowing students to evaluate their progress and their weaknesses. The interaction between students and teachers is also critical to the creation of so-called learning communities, which are defended by several studies [Bransford, *et al.* 1999] [Wenger *et al.* 2002] [Anderson 2008]. Participants of these communities, through interaction, help and challenge each other, encouraging the construction and sharing of knowledge that did not exist previously [Wenger *et al.* 2002].

In addition, the community participants together create a sense of belonging, trust, expectation of learning and commitment to participate and contribute in the community [Anderson 2008]. Such communities are perceived as a perspective of the learning environment, which is called community-centered learning environment [Bransford, *et al.* 1999].

In these communities, when given the opportunity to ask and answer questions, post comments and generally engage in an intellectual exchange with peers and with the teacher, students verbalize their current understanding of the course content. By engaging in this type of discussion in a VLE, students comment about the issues in which they are currently working and articulate their current understanding of the content [Piccoli *et al.* 2001]. This articulation process encourages students to assess their understanding and makes their decisions and more explicit problem-solving strategies. The verbalization activities are considered important to improve the learning process, because students discover knowledge gaps that were not previously known (*Ibid.*).

Participants in a VLE can also keep up with contributions by other students. This monitoring process also serves as a personal progress evaluation mechanism. While other students may or may not respond to initial contribution, it stimulates a cognitive response by readers, who will evaluate their understanding of the subject

(*Ibid.*). Teachers have a key role in the success of a course through a VLE and engagement of students, and therefore are target of the present study.

3. Canonical Correlation Analysis

The canonical correlation analysis is a useful statistical analysis when the object of study is measured in two sets of variables and we want to know if, and how, the two sets are related [Tabachnick 2012]. Often, the use of multiple independent and dependent variables provides a better perspective of the representation of a concept, and are employed in other techniques such as factor analysis and structural equation modeling [Hair 1998].

Proposed by Hotelling in 1936, the goal of canonical correlation is to develop a linear combination between sets of variables (dependent and independent) so as to maximize the correlation between the two sets. Stated differently, the process involves obtaining a set of weights for the variables, dependent and independent, which provides maximum correlation between the single set of dependent variables and the set of independent variables [Hair 1998].

The canonical correlation analysis is useful when the underlying dimensions representing the combinations of variables are unknown. Therefore, the canonical correlation analysis can be seen as an exploratory technique [Tabachnick 2012]. It is interesting to define some key terms for the complete understanding of canonical correlation analysis [Hair 1998]:

- **Canonical variables** - linear combinations, which are the weighted sum of two or more variables, for both dependent and independent variables.
- **Canonical correlation** - strength measure of the relationship between the linear combinations (canonical variables) of the independent and dependent variables. It is the bivariate correlation between the two canonical variables.
- **Canonical root** - Squared canonical correlation. It provides an estimate of the amount of shared variance between the dependent and independent canonical variables. Also known as eigenvalues.
- **Canonical function** - relationship (correlation) between two canonical variables. Each canonical function has two canonical variables, one for the set of dependent variables and one for the set of independent variables. The strength of the relationship is provided by canonical correlation.
- **Canonical loadings** - simple linear correlation between independent variables with its own canonical variable.
- **Canonical cross-loadings** - Correlation of each variable, independent or dependent, with the opposite canonical variable.

The result of the application of canonical correlation is a strength measure of the relationship between sets of canonical variables, expressed as a coefficient of canonical correlation. There are two outcomes of interest: the canonical variables that represent the optimal linear combinations of dependent and independent variables and the canonical correlation representing the strength of the relationship between them [Hair 1998].

The canonical correlation is not discontinued with the derivation of a unique relationship between the sets of variables. Instead, a number of canonical functions (canonical variable pairs) can be derived [Hair 1998].

4. Method

4.1 Objectives

The objectives of this study are:

1. Determining whether the two sets of variables from teacher and student behavior are independent of each other and determine the magnitude of the relationships that may exist between them.
2. Obtain a set of weights for the teachers' variables and the students' variables to produce a linear combination of each so that they are maximally correlated.
3. Explain the nature of any relationship between the teachers and students variables by evaluating the weight of each variable and other data provided by the canonical correlation.

4.2 Data Gathering

The data was provided by Universidade de Pernambuco (UPE), from the Núcleo em Educação a Distância (NEAD) department. The data covers one academic semester, from April 2014 to March 2015. The database was exported and provided by NEAD in the SQL format.

The definition of variables (Table 1) was based on discussions with experts and users of VLE, important factors found in previous studies, and the knowledge of which data were available in the database.

Table 1 - Description of variables used in this study.

<i>Variable</i>	<i>Actor</i>	<i>Description</i>
<i>X1</i>	Teacher	Average time in seconds to reply a post from a student in the forum
<i>X2</i>	Teacher	Number of topics on the doubts forum created without attachment
<i>X3</i>	Teacher	Number of topics on the doubts forum created with attachment
<i>X4</i>	Teacher	Number of posts in topics on the doubts forum
<i>X5</i>	Teacher	Number of posts in topics on the other forums
<i>X6</i>	Teacher	Number of distinct days that the teacher logged in the platform
<i>X7</i>	Teacher	Number of messages created on the bulletin board
<i>Y1</i>	Student	Average topics created on the doubts forum
<i>Y2</i>	Student	Average posts created on topics of the doubts forum

Teacher's actions on the platform that could somehow influence the dependent variables of students of the classes taught by the teacher were selected for the independent variables (X1 to X7). The students' actions on the platform that best represented the concept of interaction in VLE were selected for the dependent variables, or target variables (Y1 and Y2).

The courses offered on the platform by default have two forums. The first one, called "Thematic Forum", has mandatory participation, as it serves as the basis for one of the grades in the class. The second forum, called "Doubts Forum", is an open forum where students can create topics and posts on topics to discuss and answer their questions about the class. The spontaneous interaction of the student with the teacher and other students was best captured through their interactions in the doubts forum, represented by two variables containing the number of topics and posts created in that forum. After obtaining the two sets of variables, the process proceeds with the canonical correlation analysis.

4.3 Data analysis

The first step of the canonical correlation analysis is to derive one or more canonical functions. The maximum number of functions that can be drawn from the sets of variables is equal to the number of variables in the smaller set, independent or dependent. Each of the functions is orthogonal and independent of all other derived from the same data set.

The strength of the relationship between the pairs of canonical variables is reflected by the canonical correlation. When squared, the canonical correlation is the amount of variance in a canonical variable explained by the other canonical variable. This can be also called the amount of shared variance between the two canonical variables.

Before giving an interpretation of the results it is necessary to determine whether there is any relationship between the two sets of variables, because is possible that both sets of variables are completely independent of one another.

To test the independence between the sets of variables there is a multiple multivariate regression model; each multiple regression predicting one of the variables of the first group from the variables in the second group. We wish to test the null hypothesis that the regression coefficients are all equal to zero. This would be equivalent to the null hypothesis that the first set of variables is independent of the second set of variables. In this paper we use the F-test from the F-distribution to investigate the null hypothesis.

5. Results

After the initial data extraction, even without any transformation, it was found that there are classes that have no topic or teacher posting to any of the forums. This influenced the extraction of variable 6, which includes the average time for the teacher's answer.

This is not a nominal variable and so a numerical value should be placed where the data is lacking. This is a case of missing value. For this study it was decided to remove the instances that have missing values, because there were only eight classes in a 122 universe (6%).

Furthermore, it was found 8 outliers among the remaining data using the criterion of z-score higher than 3.29 ($p < .001$). No decision was made regarding the outliers at this stage. An experiment will be conducted without the outliers to verify that the change in results is significant.

The simple correlation between variables was calculated by the multicollinearity or singularity on data (Table 2). The criterion to consider two variables multicollinear was a correlation equal to or higher than 0.9. The variables X5 and X6 presented a high correlation of 0.87, but a preliminary analysis did not reveal any information redundancy between these variables. The experiments will be conducted with the removal of each variable to check if the change in results is significant.

Table 2 - Simple Correlation between variables.

<i>Variables</i>	<i>X1</i>	<i>X2</i>	<i>X3</i>	<i>X4</i>	<i>X5</i>	<i>X6</i>	<i>X7</i>	<i>Y1</i>	<i>Y2</i>
<i>X1</i>	1	-0.19	0.041	-0.09	-0.05	0.042	-0.03	-0.04	-0.01
<i>X2</i>	-0.19	1	0.212	0.406	0.17	0.101	-0.03	0.075	0.103
<i>X3</i>	0.041	0.212	1	0.369	0.493	0.49	0.421	0.434	0.424
<i>X4</i>	-0.09	0.406	0.369	1	0.576	0.492	0.007	0.481	0.48
<i>X5</i>	-0.05	0.17	0.493	0.576	1	0.876	-0.08	0.931	0.914
<i>X6</i>	0.042	0.101	0.49	0.492	0.876	1	0.042	0.905	0.929
<i>X7</i>	-0.03	-0.03	0.421	0.007	-0.08	0.042	1	-0.07	-0.05
<i>Y1</i>	-0.04	0.075	0.434	0.481	0.931	0.905	-0.07	1	0.964
<i>Y2</i>	-0.01	0.103	0.424	0.48	0.914	0.929	-0.05	0.964	1

The analysis was conducted with 114 classes, which involved 104 teachers and 1883 students. The recommendation for multivariate analysis techniques is to use at least 10 samples per variable to avoid over fitting [Hair 1998]. The conducted analysis had a ratio of approximately 16 samples per variable. A summary of the data is presented in Table 3.

Table 3 - Summary of analyzed variables.

<i>Variable</i>	<i>Min.</i>	<i>Max.</i>	<i>Median</i>	<i>Average</i>	<i>Variance</i>	<i>Stand. Dev.</i>
<i>X1</i>	39335.73653	2702489.556	413444.0081	575764.4289	2.74E+11	523603.9509
<i>X2</i>	0	8	0	0.371681416	1.646333755	1.283095
<i>X3</i>	2	103	24	28.02654867	350.0260746	18.708983
<i>X4</i>	0	7	0	0.292035398	1.119310999	1.0579749
<i>X5</i>	0	36	0	1.053097345	13.22929836	3.6372102
<i>X6</i>	0	176	2	6.707964602	315.5657396	17.764170
<i>X7</i>	0	991	55	114.7964602	25589.07427	159.96585
<i>Y1</i>	0	3.2222222	0.023622047	0.082573133	0.095252737	0.3086304
<i>Y2</i>	0	4.2592592	0.059139785	0.139189459	0.174392902	0.4176037

Two canonical functions were derived from the data. The first canonical function is derived in order to have the highest possible correlation, while the next functions are orthogonal (linearly independent) to previous, and try to capture as much variation that was not captured by the above functions.

Subsequently, significance tests were conducted and the results are presented in Table 4. The first test is applied on all pairs comprising an independence test between the two sets of variables. The second test is applied with the first and most important pair removed. If the first test is significant, but the second is not, only the first pair is interpreted.

Table 4 - Results of hypothesis tests from the canonical functions.

<i>Canonical Function</i>	<i>Likelihood Ratio</i>	<i>Approximate F value</i>	<i>Num. D.F.</i>	<i>Den. D.F.</i>	<i>Pr > F</i>
1	0.06438532	43.69	14	208	<.0001
2	0.86504525	2.73	6	105	0.0167

As the canonical correlations are derived in decreasing order of importance, usually only the first ones can be interpreted. Both pairs of this study were significant with $p < 0.05$ and will pass through further tests before being interpreted.

A third criterion for interpretation of the functions is the magnitude of the canonical correlation. The square of the canonical correlation (canonical root) is an estimation of the amount of explained variance among the canonical variables. A common rule is to do not interpret pairs whose correlation is less than 0.3 and therefore explain only 9% of the variance. The results from the canonical functions can be found in Table 6.

Table 6 - Canonical functions and correlations.

<i>Canonical Function</i>	<i>Canonical Correlation</i>	<i>Adjusted Canonical Correlation</i>	<i>Approximate Standard Error</i>	<i>Squared Canonical Correlation</i>
1	0.962065	0.959959	0.007033	0.92557
2	0.367362	0.316486	0.081739	0.134955

The first function had a fairly high canonical correlation, while the second had a low correlation, but sufficient to be interpreted.

To interpret the data produced by the canonical correlation technique, cross-canonical loads are the preferred way to analyze the results when available, because the standardized coefficients and the canonical loads suffer from changes in the variance of the variables. The cross-canonical loads are presented in Tables 7 and 8.

Evaluating the cross-canonical loads of Y1 and Y2 of the first canonical function, we found that both values are quite high. This indicates that both Y1 and Y2 variables are strongly influenced by independent variables (X1 to X7) representing teacher's actions.

Table 7 - Cross-canonical Loads of teacher variables.

	<i>Function 1</i>	<i>Function 2</i>
<i>X1</i>	-0.023	0.0973
<i>X2</i>	0.0907	0.1035
<i>X3</i>	0.433	-0.041
<i>X4</i>	0.4845	-0.0066
<i>X5</i>	0.9302	-0.071
<i>X6</i>	0.9261	0.0859
<i>X7</i>	-0.0599	0.0735

The variables X1 to X7, representing teacher's actions, vary in magnitude and sign. It is noteworthy that the variables X1 and X7 have a negative value, indicating a negative correlation. However, as their values are low, these variables have little influence to be interpreted. Similarly, the variable X2 has a positive correlation, but a low value, which is why it will not be interpreted.

The variables X3, X4, X5 and X6 presented positive and high values. This indicates that these variables, and the actions of the teacher they represent, have enough influence on the target variables, which represent the actions of students. In particular variables X5 and X6 had a fairly high correlation and explain most of the variance of the target variable.

Table 8 - Cross-canonical loads of student variables.

	<i>Function 1</i>	<i>Function 2</i>
<i>Y1</i>	0.9525	-0.0516
<i>Y2</i>	0.9542	0.0469

Evaluating the cross-canonical loads Y1 and Y2 of the second canonical function, it appears that both are quite low. The assessment of cross-canonical loads of X1 to X7 results a similar finding. This indicates that the second canonical function explains very little variance between the variables and their usefulness for interpretation is quite limited.

Tests were conducted removing the outliers identified in the pre-processing of the data phase. The first canonical function without outliers had a variation of approximately 0.14 in the canonical correlation and little variation in the canonical charges. This indicates that outliers have a significant influence on the model, but did not affect the practical results and interpretations.

The results indicated that there is only one way that the variables are related, due to the low practical significance of the second canonical function. By analyzing the first canonical function, we found that the two dependent variables, representing the number of topics and posts created in the doubts forum by students, are both strongly influenced by the teacher's actions.

By interpreting the independent variables that represent teacher's actions, we found that two variables have the greatest contribution: X5 and X6. They represent, respectively, topics created without attachments to the doubts forum and posts made to the doubts forum. This indicates that the teacher should prioritize these two actions in order to increase the participation of students in the VLE.

The X4 variable, which represents the number of threads created with attachments by the teacher to the doubts forum, also has a high influence on the participation of students. Finally, the X3 variable, which represents the amount of different days in which the teacher used the platform, also had a high contribution to the participation of students. This reinforces qualitative studies indicating that the perception of students that the teacher is always present results in greater student satisfaction with the course.

The other variables of the teacher did not have significant positive or negative influence and were not interpreted.

6. Conclusions and future works

This study aimed to verify the relationship between the actions of the teacher in student participation in virtual learning environments. Furthermore, it intended to identify variables that represent actions of teachers and students interactions. If the relation is

verified, to describe their behavior and identify which factors have the greatest influence on the relationship.

The canonical correlation analysis was used in this study since it identifies the dimensions between the dependent and independent variables and maximizes the relationship between the dimensions. This procedure provides information about the structure of the different sets of variables and how they relate.

This study aimed to verify the relationship between the actions of the teacher in student participation in virtual learning environments. The variables that have been defined represent both the actions of the teacher and student participation. It was found that four of the teacher's actions resulted in an increase in the participation of students in the class taught.

As future work, we want this relationship to be verified in other educational databases, preferably using the Moodle platform, what would allow the same set of variables to be used.

As future studies, we want to identify what are the minimum required levels of actions on the platform that the teacher should have to influence their students to have a significant participation. Regression models can be developed to accurately assess how the actions of the teacher influence the participation of students, enabling a way to provide the teacher, at any given time, feedback about the actions that yield the most results in the platform.

Additional qualitative studies can be conducted to try to explain, from students' point of view, how these four teacher's actions have influenced them to participate more actively in the class.

7. Acknowledges

Alex Sandro Gomes is Researcher DTI Level 2 CNPq in processes 310466 / 2012-1, 475634 / 2013-6 and 307202 / 2015-1. This research is also supported by FACEPE APQ 0525-1.03 / 14.

8. References

- Anderson, T. (2008) "The Theory and Practice of Online Learning" Athabasca University Press, 2nd edition
- Arbaugh, J. and Duray, R. (2002) "Technological and Structural Characteristics, Student Learning and Satisfaction with Web-Based Courses", Academy of Management Learning & Education
- Bransford, J. D., Brown, A. L. and Cocking, R. R. (2004), "How People Learn Brain, Mind, Experience, and School", The National Academies Press, Expanded Edition
- Hair, J. (1998) "Multivariate Data Analysis", Prentice Hall, 7th edition
- Martins L. and Kellermanns, F. W. (2004) "A Model of Business School Students' Acceptance of a Web-Based Course Management System", Academy of Management Learning & Education

Piccoli, G., Ahmad R. and Ives B. (2001) "Web-based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic IT skills training", MIS quarterly

Tabachnick, B. "Using Multivariate Statistics" (2012), Pearson, 6th edition

Thurmond, V. A., Wambach, K., Connors, H. R. and Frey, B. B. (2002) "Evaluation of Student Satisfaction: Determining the Impact of a Web-Based Environment by Controlling for Student Characteristics", American Journal of Distance Education, Vol. 16

Webster, J. and Hackley P. (1997) "Teaching effectiveness in technology-mediated distance learning", Academy of Management Journal

Wenger, E., McDermott, R. A. and Snyder, William (2002) "Cultivating Communities of Practice: A Guide to Managing Knowledge", Harvard Business Review Press, 1 edition

Wilson, B. G. (1996) "Constructivist Learning Environments: Case Studies in Instructional Design", Educational Technology Publications, 1st edition