A PCA Study of Student Performance Indicators in a Web 2.0-based Learning Environment

Carlo Giovannella, Federico Scaccia
ISIM Garage – ScuolaIaD
University of Rome Tor Vergata, Italy
mifav@roma2.infn.it, fedescaccia@gmail.com

Elvira Popescu
Computers and Information Technology Department
University of Craiova, Romania
popescu_elvira@software.ucv.ro

***Abstract*—Analyzing student performance indicators has been a long-standing research direction, nowadays demanding for new multidimensional approaches tailored to the recent advent of Web 2.0 in education. This paper investigates two such potential learning performance indicators: i) students' active involvement with the Web 2.0 tools and ii) students' learning styles. By using Principal Component Analysis (PCA) we were able to identify a subset of student activities that are relevant to build up an orthogonal space of representation and those activities that may support prediction of learner success (measured as final grades). On the other hand, no significant correlations were found between the Felder-Silverman learning styles and students' performance in our case study.**

Keywords-Principal Component Analysis (PCA), learning performance indicators, Web 2.0, learning styles

I. INTRODUCTION

Predicting students' performance is an essential part of the instructional process, with the aim to find adequate strategies to offer appropriate feedback, reduce student dropout and increase learner success. In this paper we focus on two types of potential multidimensional performance indicators: active online participation and learning styles.

The effect of active participation has long been investigated in computer assisted learning and, recently, in online learning settings, although with contradictory results. Some authors reported that the use of discussion forums was the strongest indicator of student performance [1] or concluded that online participation enhanced student engagement and thus improved learning effectiveness [14]; conversely, others found that students were able to learn equally well on online courses regardless of their level of online participation [6]. As far as learning styles (LS) are concerned, several researchers assumed them to be a predictor of student success [5], while others found that they had no significant effect on learning performance [6, 11]. Paper [10] provides a summary of previous research on the relationship between LS and learning performance and shows that no unanimous conclusion can be drawn.

The present study explores both indicators in the context of a Web 2.0-enhanced learning environment. We decided to use Principal Component Analysis (PCA) method to address this issue, focusing on two different approaches: i) selection of a set of student activities of potential relevance for the educational process, identification of an orthogonal space of representation and evaluation of possible relationships between the space dimensions and the student performance; ii) measurement of the students' LS (according to the four-dimensional Felder-Silverman model - FSLSM [4]) and identification of the relationship between the LS and the student performance.

The rest of the paper is structured as follows: the next section describes the context of the study, the instructional scenario based on Web 2.0 tools, and the methods used for data collection. The third section describes the data analysis and results. The final section outlines conclusions and future research directions.

II. CASE STUDY DESCRIPTION AND DATA COLLECTION

A. Experiment Settings

The context of the study is a course on "Web Applications' Design", delivered to 4th year undergraduate students in Computer Science from the University of Craiova, Romania. A project-based learning (PBL) scenario was used, in which students had to work in teams, throughout the semester, in order to design and implement an authentic Web application. 45 students were enrolled in the course, grouped into 11 teams. The project activity was done in a blended mode: there were weekly face-to-face meetings between each team and the instructor and for the rest of the time students had to use four Web 2.0 tools (Blogger, MediaWiki, Delicious, Twitter) as support for their communication and collaboration activities. More details regarding the PBL scenario can be found in [7].

B. Data Collection

To investigate the relation between students' performance, their involvement with the Web 2.0 tools and their LS, three kinds of data were collected: i) students' final grades; ii) students' activity on the blog, wiki, Delicious and Twitter; iii) students' LS. While the former were taken from the gradebook (as provided by the instructor), the latter were gathered explicitly from the students, by means of the Index of Learning Style questionnaire [12] (filled in by 42 of the students). Finally, learners' activity data on the Web 2.0 tools were automatically collected by means of a dedicated platform, called eMUSE (**e**mpowering **M**ash**U**ps for **S**ocial **E**-learning) [8]. The recorded student actions refer to various types of learning activities: creating content (*blog_post-entry, wiki_revise-page, wiki_upload-file*), social interactions (*delicious_add-friend-to-network*), organizing content (*delicious_post-bookmark*), communication and feedback (*blog_post-comment, twitter_post-tweet*). The number of actions of each type was computed for each student at the end of the semester; obviously, this is a quantitative indicator only, since it does not take into account the quality of the students' actions (e.g., correctness of the wiki page, content of the blog post, relevance of the tweet, pertinence of the blog comment, usefulness of the bookmark).

C. *Method*

The collected data was analyzed by using PCA, a method of factorial analysis, widely used whenever dealing with a multidimensional space of representation (multivariate distribution) characterized by an incertitude on the orthogonality of the axes forming its basis [2] (here the 7 dimensions given by the student actions).

III. Analytics and Discussion

Since the set of actions considered in this paper were not all carried out with equal frequencies, we had to normalize them to operate on a space of representation in which all the axes have the same scale and the same initial weight.

The calculation of the correlation matrix shows a high degree of correlation among the selected parameters with the exception of *blog_post-comment* (see Fig. 1).

e_1	e_2	e_3	e_4	e_5	e_6	e_7
1	0.0513	0.3369	0.3684	0.6195	0.5536	0.7265
0.0513	1	-0.0752	0.0347	-0.0394	-0.0116	-0.0894
0.3369	-0.0752	1	0.2213	0.3523	0.2275	0.2397
0.3684	0.0347	0.2213	1	0.4779	0.22	0.2431
0.6195	-0.0394	0.3523	0.4779	1	0.3511	0.545
0.5536	-0.0116	0.2275	0.22	0.3511	1	0.8478
0.7265	-0.0894	0.2397	0.2431	0.545	0.8478	1

Figure 1. Correlation matrix of student activities (e1 = *blog_post-entry*; e2 = *blog_post-comment*; e3 = *delicious_add-friend-to-network*; e4 = *delicious_post-bookmark*; e5 = *twitter_post-tweet*; e6 = *wiki_upload-file*; e7 = *wiki_revise-page)*

The very strong correlation between *delicious_add-friend-to-network* and *delicious_post-bookmark* suggested us to consider only one of these two parameters; the same goes for the parameters *wiki_upload-file* and *wiki_revise-page* on one hand and *twitter_post-tweet* and *blog_post-entry* on the other hand. According to these observations we decided to apply PCA on two reduced four-dimensional subspaces: a) subspace I composed by *blog_post-entry*, *blog_post-comment*, *delicious_add-friend-to-network* and *wiki_upload-file;* b) subspace II composed by *blog_post-entry*, *blog_post-comment*, *delicious_post-bookmark* and *wiki_revise-page*. The difference was not substantial, although in the case of subspace II the first two eigenvalues contribute to 74.3% of the total, while in the case of subspace I they contribute to 70.1%. We also tried to use a five-dimensional subspace adding the parameter *twitter_post-tweet* but no significant improvement was observed.

Figure 2 shows the activity of the students in the bi-dimensional space of the first two principal components: Y1_N and Y2_N. The position on the Y2_N axis is mainly determined by the parameter *blog_post-comment*, while the position on the Y1_N axis is determined equally by the parameters *blog_post-entry* and *wiki_revise-page* and to a lesser extent by *delicious_post-bookmark*.

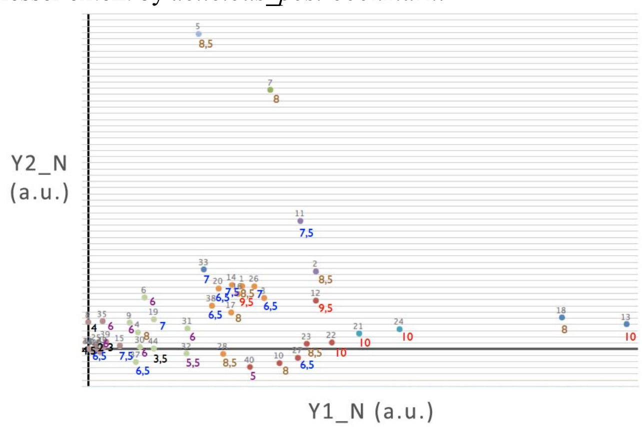

Figure 2. Students' positions in the space of the two principal components derived from a selected subset of student activities (subspace II)

From Fig. 2 we can observe that students with low average grade tend to cluster around the origin of the axis, while another small cluster tends to collect students with an average grade ranging between 6.5 and 7.5, with few exceptions. In general, the final score tends to increase along the Y1_N axis, with the very high ones located between the middle and the extreme right of the plot. No clear dependence between the average grade and the position on the Y2_N axis emerges. We may only observe that both very high and low grades tend to be located on the 0 value of the Y2_N axis. The dependence of the average grade on the value of the Y1_N component is clearly confirmed in Fig. 3 (left), where the log fit is characterized by R=0.79. Similarly, Fig. 3 (right) confirms the unclear dependence of Y2_N on the final average grade, with the tendency to peak around 8.

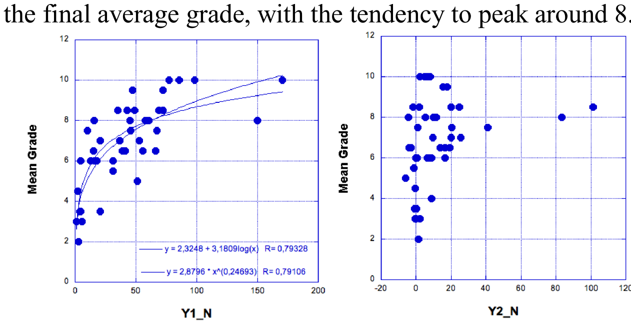

Figure 3. Value of the first and second principal component vs. student mean grade (left side and right side respectively)

Following our previous experience [9], we applied the PCA also to the four-dimensional space of FSLSM. The correlation matrix is not dissimilar from those observed in other studies [3, 13]. All are characterized by a quite relevant correlation between the *sensing-intuitive* and *sequential-global* dimensions (0.39 for the set of data considered in this work) and a weak correlation between *active-reflective* and *visual-verbal* dimensions (0.13). Peculiar to this set of data is a weak anti-correlation between *active-reflective* and *sensitive-intuitive* dimensions.

Figure 4 shows roughly the contribution that each LS gives to Y1 and Y2 and where the students that got very high or low marks are located: no meaningful correlation between the final score and the students' LS was identified. To cross-check, we tried to correlate the mean grade with each single dimension of FSLSM. No strong correlation between mean grade and *sensing-intuitive* and *sequential-global* dimensions has been observed (with a slight trend toward better results for *intuitive* and *sequential* students). A weak correlation (R around 0.2) emerges with the other two dimensions: more *verbal* and *active* students achieved a better grade (Fig. 5). Finally, we checked for possible correlations between the FSLSM dimensions and Y1 with no results.

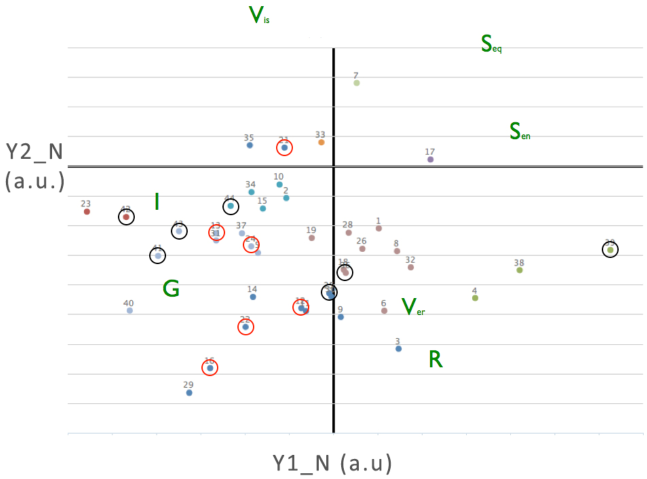

Figure 4. Students' positions in the space of the two principal components derived from students' learning styles. Black circles identify students with final average grade ≤ 4 (out of 10), while red circles identify those with final average grade ≥ 9.5 (out of 10)

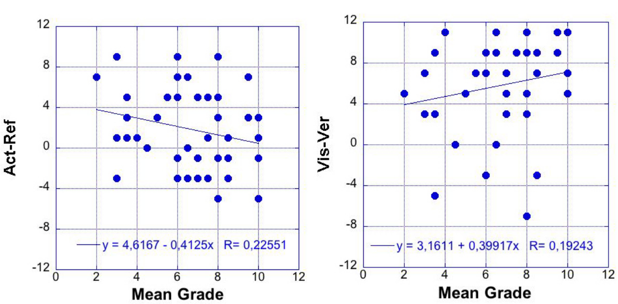

Figure 5. Student mean grade vs. student learning style dimensions (*active-reflective* on the left, *visual-verbal* on the right)

Overall, these results indicate that the measurement of the LS could provide information on individual characteristics at a given point but they are not correlated with the overall level of participation in the learning process (at least considering the parameters that have been measured here), although more active people seem to get higher grades.

IV. Conclusion and Future Work

The paper investigated the relations between students' performance, their involvement with Web 2.0 tools and their LS. The findings of this exploratory study include:

- PCA is very effective in identifying the principal component (Y1_N) that should be monitored to predict the student success. The actions that, in our case, contribute most to Y1_N are those that involve content production (i.e., actions that imply a previous appropriation process) and its organization (i.e., actions that also imply awareness about the content). Not so relevant seems to be the social interaction. Providing feedback to content produced by the peers appears to be a fully independent variable determining the values taken by the second principal component (Y2_N) but its correlation with the final grade is unclear.

- FSLSM dimensions do not seem to be reliable predictors of the student performance.

As future work, we plan to: i) conduct larger studies, including a higher number of students, and perform more in-depth analyses of the recorded student actions; ii) add a qualitative dimension (i.e., evaluate the relevance of each action and the quality of the learner-generated content).

References

[1] J.W. Alstete and N.J. Beutell, "Performance indicators in online distance learning courses: a case study of management education", Quality Assurance in Education, 12(1), 6–14, 2004.

[2] S. Bolasco, "Analisi Multidimensionale dei dati", Carocci, Roma, 1999.

[3] C. Giovannella, "What can we learn from long-time lasting measurements of Felder-Silverman's learning styles ?", Proc. ICALT 2012, pp. 647-649, 2012.

[4] R.M. Felder and L.K. Silverman, "Learning styles and teaching styles in engineering education", Engineering Education 78, 674–681, 1988.

[5] E. Y. Huang, S. W. Lin, and T. K. Huang, "What type of learning style leads to online participation in the mixed-mode e-learning environment? A study of software usage instruction", Computers & Education, 58(1), 338-349, 2012.

[6] J. Lu, C.S. Yu, and C. Liu, "Learning style, learning patterns, and learning performance in a WebCT-based MIS course", Information and Management, 40, 497–507, 2003.

[7] E. Popescu, "Project-Based Learning with eMUSE: An Experience Report", Proc. ICWL 2012, LNCS 7558, Springer, pp. 41-50, 2012.

[8] E. Popescu, "Providing Collaborative Learning Support with Social Media in an Integrated Environment", World Wide Web, Springer, DOI: 10.1007/s11280-012-0172-6, 2012.

[9] F. Scaccia and C. Giovannella, "How about using the PCA to analyze changes in learning styles ?", Proc. ICWL 2012, pp. 279-284, 2012.

[10] R.S. Shaw, "A study of the relationships among learning styles, participation types, and performance in programming language learning supported by online forums", Computers & Education, 58(1), 111-120, 2012.

[11] C.C. Shih, T. Ingebritsen, J. Pleasants, K. Flickinger, and G. Brown, "Learning strategies and other factors influencing achievement via web courses". ERIC. Document Reproduction, Service no. ED 422876, 1998.

[12] B. Soloman and R. M. Felder, "Index of learning styles questionnaire", http://www.engr.ncsu.edu/learningstyles/ilsweb.html, 1998.

[13] S.R. Viola, S. Graf, Kinshuk, and T. Leo, "Analysis of Felder-Silverman Index of Learning Styles by a Data-driven Statistical Approach", International Journal of Interactive Technology and Smart Education, vol. 4, 7-18, 2007.

[14] D. Zhang, L. Zhou, R.O. Briggs, and J.F. Nunamaker, "Instructional video in e-learning: assessing the impact of interactive video on learning effectiveness", Information and Management, 43, 15–27, 2006.