Analyzing the influence of online behaviors and learning approaches on academic performance in first year engineering[1]

Sergio Celis[1], Dany López[2], and Javier Silva[3]

[1] School of Engineering and Science, FCFM, Universidad de Chile, Santiago 8370456, Chile
[2] School of Education, Pontificia Universidad Católica, San Joaquín 7820436, Chile
[3] Physics Department, FCFM, Universidad de Chile, Santiago 8370456, Chile
scelis@ing.uchile.cl

Abstract. Over the last four decades, the study of academic performance in higher education has increased its number of information sources to understand phenomena such as student achievement or dropout. The first econometric models in the field commonly used student characteristics, pre-college achievement, and college performance. Then, a large range of psychosocial theories, with its respective instruments (typically questionnaires), added a new layer of analyses that complemented previous models. Recently, colleges and universities have dramatically expanded their capacity to capture student data through different systems. Such is the case of the learning management systems (LMS), which provide dynamic and a large amount of data about student online behavior. We are just beginning to explore how these layers of data come together to explain academic performance. In this study, we seek to understand and model these layers of data from a first year cohort at a large engineering school in Chile (784 students). First, we use support vector regressions to model second semester GPA on student characteristics, pre-college data, first semester grades, and online behavior. We then added to the model information extracted from the LEARN+ questionnaire, a psychosocial instrument that profiles different learning approaches (i.e., surface, strategic, and deep) and environmental perceptions. The results indicate that both online behavior and LEARN+ data increase prediction power. In addition to first semester performance, the features that seem to explaining academic achievement in the second semester to a significant extent are the LMS interaction distribution over the semester, perception of applied knowledge, and the score in the science score in the national admission test. These results are important for first year engineering, since in this field first year performance has long lasting effects on future persistence and achievement.

Keywords: Learning Analytics, First-year Engineering, Support Vector Machine, Learning Approaches, Online Behavior.

[1] This work was funded by Complex Engineering Systems Institute (CONICYT – PIA – FB0816) and by the CONICYT Fondecyt project #1161413.

1 Introduction

How students' individual characteristics, behaviors, and previous academic performance influence future achievement and other college outcomes is an old quest in higher education (Pascarella & Terenzini, 2005). Recently, this quest has become more comprehensive as management systems grow in complexity and the capacities for data analysis increase in colleges and universities. In this article we combine three types of datasets in order to explore how they complement to each other and how much information they add to our knowledge about student learning and achievement.

Traditionally, research on academic achievement in higher education has focused on outcomes such as dropouts and retention rates (Pascarella & Terenzini, 2005). This literature often discusses the impact of individual characteristics—e.g., gender, race, socioeconomic background—and institutional factors—school size, discipline, selectivity—(DesJardins, Ahlburg, & McCall, 1999) on persistence. Scholars have also proposed multiple non-observable factors for analyzing dropout and retention. Among the most influential perspectives are the dropout model of Bean and the integration process of Tinto. Bean (1982), relying on studies of organizational rotation, found that intentions to leave, academic performance, and instrumental values are the most important factors in predicting dropout. Based on tribal studies, Tinto (1988) proposed that institutional integration—during the first year, in particular—is key to understanding persistence. Despite some recent criticism (e.g., Núñez, 2004) these theoretical frameworks continue to be widely used in the United States and in many other countries. Quantitative studies model these frameworks through econometric techniques that use national and institutional data registered at an annual or semester base (e.g., DesJardins et al., 1999; Cabrera, Burkum, & La Nasa, 2005).

As the "granularity" of student data gets smaller, practitioners and scholars have moved beyond dropout and retention to create early warning (e.g, Celis et al., 2015) and recommendation systems to all students. Moreover, not only academic records are rapidly captured, but also a wide range of interactions with technology systems, from the now ubiquitous learning management systems (LMS) to college cards swiped around the campus to sensors and application logs located in labs or electronic textbooks. The research fields where many of these new analyses converge is called learning analytics. Learning analytics is a fairly new area of research that uses mathematical and computing tools to analyze educational data generated by the interaction between students and multiple platforms that universities use for supporting learning processes (Larusson & White, 2014; Romero et al., 2008). Among the learning analytics' common tasks are classification, clustering, text mining, and visualization (Romero & Ventura, 2010). What data platforms, and under what mathematical technique are key questions for explaining student academic achievement.

In addition to the systems that capture student achievement and behavior, higher education scholars have also used psychosocial models to explain different aspects of college learning. In this area, there is a wide spectrum of theoretical constructs, from those that consider psychological treats as grit or early attachment (Lavy, 2017) to those who are developed through academic experiences, such as self-regulated learning (Zimmerman, 1990) and self-efficacy (Bandura, 1977) to those that are shaped through continuous interaction with the institutional environment such as learning approaches (Marton & Saljo, 1976). Multiples studies have shown that these frameworks explain to some extent students' academic achievement (e.g., Campbell & Cabrera, 2014). Although there are some recent efforts in this line (e.g., Ellis, Han, & Pardo, 2017), there is still much room for exploring how these theoretical lenses perform or improve student achievement models that includes big data analysis, such as those in LMS. In this study, we analyze how the learning approach framework (Marton & Saljo, 1976) informs a model based on individual characteristics and online behavior in the context of a large engineering school.

1.1 First-Year Engineering

Authors who focus on engineering programs have depicted how multiple individual characteristics influence academic achievement. In particular, previous academic achievement is frequently reported as a significant predictor of future performance (French, Immekus, & Oakes, 2005). However, in some engineering schools, its predictive capacity is rather modest (Besterfield-Sacre, Atman, & Shuman, 1997). Moreover, as discussed before, the increasing need for early warning systems and accurate models for predicting performance have pushed the exploration for additional sources of information. Therefore, other factors have been closely investigated. Blumner and Richards (1997) collected study habits data from first-year students at a selective engineering school. Taking into account previous academic abilities, they found that students with greater GPA declared more inquisitiveness (i.e., deep learning strategies) and less distractibility (i.e., low concentration when working on a task) than their less successful peers.

Another venue of intense research is non-cognitive traits, with mixed results. French et al. (2005) found no difference between students' motivation and institutional integration on cumulative GPA. On the other hand, Vogt (2008) found a positive correlation between self-efficacy and engineering students' GPA. Scholars have also asked for the influence of race and gender. Interestingly, despite the fact women are underrepresented in engineering, they often report greater GPA than men (French et al., 2005). Vogt, Hocevar, and Hagedorn (2007) found that women in engineering exerted more effort and were more likely to ask for academic help than men.

In engineering and other STEM fields, the first year is particular important. For instance, academic achievement has a greater effect on persistence in first year than in the subsequent ones (Pascarella & Terenzini, 2005). Similarly, factors such as self-efficacy and approaches to learning are strongly shaped in the first-year of studies (Felder, Felder, Mauney, Hamrin, & Dietz, 1995; Marra, Rodgers, Shen, & Bogue, 2009; Meyer & Marx, 2014). Thus, institutional and research efforts for understanding how different factors influence first-year academic achievement in engineering are key for increasing persistence and better outcomes for future engineers and other STEM workers. Our study focuses precisely on first-year engineering at a large and selective school in Chile.

2 Conceptual Framework

In the previous section, we discussed literature on the influence of individual characteristics, online behavior, and psychosocial treats on academic achievement. In this study, we explore these three areas on the particular setting of first-year engineering. Next, we describe the key variables in each of these areas, presenting how they have been conceptualized by previous research.

2.1 Personal Characteristics and Academic Achievement

In engineering and other STEM fields, student *previous academic achievement* is the most significant factor for explaining future ones (Pascarella & Terenzini, 2005). Previous achievement can be divided in two kinds: precollege achievement and accumulated GPA during college. In Chile, pre-college achievement is usually measured by the national admission test scores (PSU in Spanish for *Prueba de Selección Universitaria*) and high school GPA and Ranking. During college, the grades in any semester are the best predictor for the following one.

Another important factor for predicting academic success and in particular persistent is student *socioeconomic background* (DesJardins et al., 1999). Students from less advantages families or environments often have a harder time socializing and creating future perspective on professional engineering. For instance, in Chile, Celis et al. (2015) found that among the students with risk of falling on probation, those coming from private or subsidized schools had greater chance of overcoming this risk than their counterparts from public high schools. In the Chilean highly segregated school system, school character is a good proxy for socioeconomic background, with public school serving the less advantaged population. Finally, another important aspect in engineering schools is *gender*, since this corresponds to a traditionally male dominated field (Marra et al., 2009). Even though women are usually underrepresented in engineering fields, those who persists show a certain academic advantage over their male counterparts (Celis et al., 2015).

2.2 Online Behavior

Currently, most colleges and universities support courses by means of digital online resources, such as: delivery of electronic educational resources, support of interactive environment, and LMS (Cavus & Zabadi, 2014; Graham, 2006). As a result, a large amount of student data has become available (De Freitas et al., 2015; Ferguson, 2012). Regarding LMS, researches have used different techniques for estimating online activity, such as frequency measures, like *total count* on each type of LMS interaction; measure of *student's distribution* related to the amount of interactions on each online activity, like the *mean*, *median, standard deviation, center mass* or *skewness*; time spent on a particular action; time of first log-in; time on-task, among others (Jo, Kim, & Yoon, 2015; Joksimović, Gašević, Loughin, Kovanović, & Hatala, 2015; Kovanović et al., 2015b, 2015a; Michinov, Brunot, Le Bohec, Juhel, & Delaval, 2011; You, 2016)

Another way to examine distributions of online activities is entropy measures. Among this type of variables, there exists the Shannon entropy, the Gini coefficient or the Atkinson's index (Harris et al., 2009; Kelleher, Mac Namee, & D'Arcy, 2015; Pena-Ayala, 2017). Thus, in order to capture different types of online activity patterns by a particular student, it is possible to include more understandable entropy estimators. These are: the *participation ratio* (RP), which gives a rough estimate of the number of the effective weeks in which the student's online interactions differ markedly from zero (Kramer & Mackinnon, 1993); and the *compact index* (Skokos, Krimer, Komineas, & Flach, 2009) which measures the sparseness of a particular online activity distribution over the semester.

Finally, researches has found that certain students' behaviors in a LMS can be associated with academic achievement (Asarta & Schmidt, 2013; Fritz, 2011; Michinov et al., 2011; Strang, 2017) and student engagement (Hart & Ganley, 2017; Howard, Meehan, & Parnell, 2018). In general, these studies found that those students who have higher levels of online activity are more likely to present higher course grades, whereas students who have lower levels of online activities are more likely to present lower course grades with respect to their peers.

2.3 Learning Approaches

The approaches to learning theory (Marton & Säljö, 1976) is among the most influential framework in the field of student learning research. This framework (developed latter by multiples researchers such as Ramsden, 1979) suggests that students approach learning in three distinctive ways: *surface*, *strategic*, and *deep.* These learning approaches are shaped in each student according to his or her personal characteristics and interactions with the institutional context, such as peers and teaching. *Surface learning* is associated with memorizing and repeating content and with studying for the test. *Deep learning* refers to understand the meaning of the subject matters, relating them with previous knowledge, and associating theory with practice. *Strategic learning* occurs when students organize their effort (between surface and deep approaches) in order to obtain the best possible grade.

Usually, these approaches are obtained through self-reported questionnaires. Previous research is mixed regarding the learning approached effects on academic achievement. Some scholars have found that surface and deep approaches are related to poor performance and to high achievement, respectively (Matthew, Ellis, & Taylor, 2011; Rytkonen et al., 2012). In the same way, Campbell and Cabrera (2014) found that surface learning is not necessarily related to better academic achievement.

3 Methods

3.1 The Setting

The School of Engineering and Science at the Universidad de Chile (FCFM) is a selective academic unit, which receives the best students in the nation. FCFM has nine engineering and three science undergraduate programs, as well as a geology program, and about 5,000 undergraduate students. Each year, FCFM receives an entry cohort of approximately 800 students. All FCFM students enroll in a common core program in their first two years of school and take the same courses in the first semester: introductions to calculus, algebra, Newtonian physics, engineering, chemistry, and computing tools for engineering and science. Approximately 30% of first-year students fail at least one course. In the last two decades, FCFM has implemented several strategies and actions for improving students' retention and academic performance. For instance, promoting active learning, improving infrastructure for student life, and launching specialized units to support students' academic achievement and wellbeing. Currently, first year's retention rates are close to 95%.

3.2 Data Collection and Sampling

Among the vast types of LMS—such as Moodle, Sakai, Desire2Learn or Canvas—this study addresses a particular one, namely, U-Cursos. This LMS was developed and launched at the Universidad de Chile in the 1990s. At first, it aimed to support engineer students, and now, it has expanded across other schools and Chilean institutions. Regarding U-Cursos usage, there are many online activities recorded in its log files. These U-Cursos activities reflect different types of students' operation: download files, upload files, assessments tasks, forum topic response, message read, to mention a few. These activities create a pattern of usage that might change as a student advances through the semester. So, in order to account for differences in instructional conditions across courses, general LMS categories are needed to be determined before data are merged (Gašević, Dawson, Rogers, & Gasevic, 2016). We encapsulate the large amount of U-Cursos online activity into five categories. Student activity related to accessing course contents are encoded as *Academic Content*. Similarly, *Administrative Content* refers to students' access to information related to syllabus or other course's rules. On the other hand, *Read* and *Write Comment* categories focus on both reading communications and writing or filling out information through the LSM, respectively. Finally, activities that involve operations to access to online test/questionnaires within the LMS are related to *Test* category. In this study we used the log files of the entire 2017 first-year entry cohort.

The data corresponds to students who in 2017 took first year courses (e.g., calculus, linear algebra, and physics), bringing 1,090 in the first semester and 871 in the second. In order to incorporate academic performance in the model, we are going to analyze students with information in both semesters. For this study, we define a normalized GPA (*z-score*) as the academic achievement variable. We predict second semester z-score, considering the first semester z-core as previous achievement and the variables defined above (see conceptual framework section). Then, the students who answered the LEARN+ questionnaire—built to measure learning approaches—are 478 from where 296 did it their second semester. The final dataset corresponds to the subset of these 296 students. Below, Table 1 and Table 2 show the descriptive statistics between the original dataset and the subset for categorical and continuous variables, respectively.

Table 1. Descriptive statistics for categorical variables. The first dataset corresponds to all the students and the subset to those who completed the LEARN+ questionnaire.[1] Regular admission refers to those admitted via PSU test. The admission types for students who were admitted through a different via (e.g., sport, equity access) were classified as Boundary or Special if they either had similar or lower scores in the PSU test than the regular ones, respectively.

Category: Gender	Dataset (%) (N=827)	Subset (%) (N=296)
Male	71.38	66.78
Female	28.62	33.22
Category: Admission Type[1]		
Regular	88.80	85.71
Boundary	7.59	9.30
Special	3.62	4.98
Category: School Type		
Private	41.56	40.20
Subsidized	34.44	32.56
Public	23.99	27.24

Table 2. Statistics for continuous variables. The 1 dataset corresponds to the full dataset (all students) and the 2 to the Subset of those who completed the LEARN+ questionnaire. The means and SD reported for the LMS variables represent the means and SD of each estimator based on students' weekly interaction distributions.

Variables	Mean 1*	Mean 2*	SD 1*	SD 2*
GPA	6.519	6.528	0.217	0.228
GPA Ranking	775.764	779.811	64.688	60.720
PSU: Maths	737.851	742.814	117.317	109.091
PSU: Language	666.956	673.066	115.379	108.816
PSU: Sciences	697.030	704.907	113.472	107.516
Enrolled credits	21.613	23.023	4.767	2.372
Failed credits	4.447	4.286	5.793	5.719
LMS: academic content	318.042	319.789	204.980	180.542
LMS: administrative content	7.896	7.595	9.401	8.497
LMS: write comment	32.589	34.505	28.460	29.203
LMS: read comment	598.486	654.140	396.869	374.057
LMS: Mass Center	-0.085	-0.093	0.129	0.095
LMS: Skewness	0.469	0.335	0.642	0.435
LMS: Mean	48.055	50.918	25.678	23.051
LMS: Standard deviation	34.969	36.219	15.809	14.430
LMS: Compact Index	0.811	0.865	0.402	0.362
LMS: RP	0.417	0.442	0.121	0.095
z-score (2nd semester)	0.000	0.216	1.214	0.990
z-score (1st semester)	0.191	0.411	1.443	1.241

3.3 Analysis

Support vector machines are a family of algorithms for classification, regression and outliers detection developed in the nineties at AT&T Bell laboratories by Vapnik and co-workers (Smola & Schölkopf, 2003). These algorithms have some advantages over linear regression, for example they are memory efficient, effective in high dimensional spaces and when the number of dimensions is greater than the number of samples, and also when there are issues of multicollinearity, among others. In the case of regression (SVR) the mathematical formulation corresponds to an optimization problem of find α and α^*.

$$\textit{Maximize} \quad -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) - \varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*)$$

$$\textit{Subject to} \quad \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \quad \textit{and} \quad \alpha_i, \alpha_i^* \in [0, C] \tag{1}$$

Where each x, y contain the students attributes and output variable respectively. These algorithms have also the versatility to use different kernel functions k, which is ideal for nonlinear problems. The parameters ε and C are from the model and define respectively the soft margin of the loss function and the trade-off between the flatness of the model and the amount up to which deviations larger than ε are tolerated. Finally, the model can be obtained with equation (2). Where b is the interception of the model.

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)k(x_i, x) + b \tag{2}$$

In our analysis the SVR determines the *second semester z-score* to understand the student academic performance at the end of the first year. We used SVR instead of linear regression because of two reasons. First, since some of the data can be contaminated with human error, a model with a soft margin in the loss function can be an improvement to avoid noise. Second, the data was collected without an experimental design and there could be strong correlations between variables that enter in the regression bringing issues of multicollinearity, which does not arise in SVR. Since we want a first approach to understand the academic performance, which mostly depends linearly on the variables, we are going to use a linear kernel. We used Scikit Learn, a Python library, for the implementation of SVR.

4 Results

We computed three SVR models to understand how each block of data adds explanatory power in predicting first year academic performance (See Table 3). In general, we found that including LMS online behavior data improves explanatory power, considerably. However, when we add the LEARN+ variables the contribution is minimal. In all models past performance, *1st semester z-score,* is the best predictor for first year GPA. Student's high school GPA ranking has also relative importance. Interestingly, this influence decreases as we include LMS variables. On the contrary, the influence of PSU Science increases. Regarding LMS activity, *LMS standard deviation,* which is base on weekly activity, and *center mass* have the most significant influence in the models. The former indicates that the more variation in weekly activity (e.g., weeks of intense work and other with low activity), the more likely to achieve a better grade at the end of the year. The latter has a negative sign, which indicate that the larger the *center mass* is, the less likely to obtain a better grade. *Center mass* is based on the semester activity, and a large value indicates a students that concentrate activity towards the end of the semester. Also, *LMS mean* has some significance. This variables indicates that the more students connects in a weekly based, the best for their grades. Finally, *LMS read comments* and *write comments* show no significant impact. Of note, it is the fact that those students with a higher value on a strategic approach to learning are more likely to achieve a successful academic performance than those with higher scores on the deep or surface learning.

Table 3. Results for the importance of each variable in three SVR models, using the subset (296 students). Model 1 incorporates pre-college and first semester academic performance variables. Model 2 adds online behavior variables. Model 3 adds LEARN+ variables.

Type	Variables	Model 1	Model 2	Model 3
Pre-College & 1st Semester Grades	z-score (1st semester)	2.196	1.792	1.765
	PSU: Sciences	0.257	0.492	0.538
	PSU: Language	0.179	0.041	0.353
	Enrolled credits	0.348	0.337	0.315
	Ranking GPA (High School)	0.708	0.288	0.289
	School Type: Partially Subsidized	0.060	0.064	0.100
	Admission Type: Regular	0.077	0.095	0.062
	Gender: Male	0.029	0.048	0.046
	Admission Type: Boundary	-0.077	0.004	0.012
	School Type: Public	0.004	0.016	-0.007
	Gender: Female	-0.029	-0.048	-0.046
	Admission Type: Special	0.000	-0.098	-0.074
	School Type: Private	-0.065	-0.080	-0.093
	High School GPA	-0.263	-0.010	-0.132
	PSU: Maths	-0.024	-0.051	-0.192
	Ratio of failed credits (1st semester)	-1.126	-0.961	-0.945
LMS	Standard Deviation		0.590	0.556
	Mean		0.370	0.279
	Administrative Content		0.076	0.158
	RP		0.114	0.141
	Compact Index		0.211	0.017
	Read Comment		-0.123	0.013
	Write Comment		0.043	-0.006
	Skewness		-0.135	-0.031
	Academic Content		-0.037	-0.242
	Center Mass		-0.546	-0.507
	Learning Approach: Strategic			0.312
	Professor's teaching			0.293
	Learning Approach: Deep			0.239
	Academic freedom			0.225
	Distance to university			0.151
	Peer support			0.059

	Constructive feedback			0.059
LEARN+ Questionnaire	Learning Approach: Surface			0.007
	First member of the family in the university			-0.017
	Family support			-0.076
	Cultural activities			-0.090
	Learning resources			-0.106
	E-learning			-0.115
	Workload			-0.297
	Age			-0.356
	Skills			-0.384
	Mean Squared Error	0.672	0.615	0.607
	Coefficient of determination	0.320	0.378	0.386

5 Discussion and Conclusions

Our findings suggest that student LMS online activity is a relevant factor to explain variances in engineering first year student performance. To some extent, student online activity works as a sign of perseverance, study methods, and engagement. Indeed, according to our results, when and how much activity a student had is more important than what kind of activity he or she did in the LMS. The LMS variables are also useful to understand the enduring relevance of pre-college characteristics. For instance, it is interesting to observe the increase of the *PSU science* (biology, chemistry, and physics) relevance in the model. This indicates that when controlling for LMS activity and other characteristics, the *PSU science* score stands out. In the Chilean systems, there is large variation in the science content among high schools. Perhaps, this part of the PSU test indicates a knowledge base that gives advantages to students when controlling for everything else.

The LEARN+ variables did not add much information when using LMS variables. An overlap between online activity and the psychosocial traits captured by LEARN+ might explain this small effect. Nevertheless, the results are consistent among them. For instance, the *strategic approach* is consistent with the LMS *standard deviation*. At FCFM, to engage in periods of intense academic activity seems to be more beneficial than pure perseverance and constant study work. We can also observe that the *strategic* and *deep approaches* show relatively close prediction power as compared to *surface approach,* which is also consistent with the literature [12]. The domain specific subject of the sample might also explain the LEARN+ small effect. Engineering students do engage more in online activity. We have some evidence that LEARN+ is significant in other disciplines with less frequency of online behavior.

References

1. Asarta, C. J., & Schmidt, J. R. (2013). Access Patterns of Online Materials in a Blended Course. *Decision Sciences Journal of Innovative Education*, *11*(1), 107–123.
2. Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review, 84*(2): 191–215.

3. Bean, J. P. (1982). Student attrition, intentions, and confidence: Interaction effects in a path model. Research in Higher Education, 14(4), 291-320.
4. Besterfield-Sacre, M., Atman, C. J., & Shuman, L. J. (1997). Characteristics offreshman engineering students: Models for determining student attrition and success in engineering. Journal of Engineering Education, 86, 139–149.
5. Blumner, H. N., & Richards, H. C. (1997). Study Habits and Academic Achievement of Higher. *Journal of Engineering Education, 86*(2), 125–132.
6. Cabrera, A. F. Burkum, K. R. & La Nasa, S. M. (2005). Pathways to a four-year degree: Determinants of transfer and degree completion (pp. 155-214). In Alan Seidman (Ed.). *College Student Retention: A formula for success*. ACE/Prager Series on Higher Education
7. Campbell, C., & Cabrera, A. (2014). Making the Mark: Are Grades and Deep Learning Related? *Research in Higher Education, 55*(5)494-507.
8. Cavus, N., & Zabadi, T. (2014). A Comparison of Open Source Learning Management Systems. *Procedia - Social and Behavioral Sciences*, *143*, 521–526.
9. Celis, S., Moreno, L., Poblete, P., Villanueva, J., & Weber, R. (2015). Un modelo analítico para la predicción del rendimiento académico de estudiantes de ingeniería. *Revista de Ingeniería de Sistemas, 29*(1), 5-24.
10. De Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., … Arnab, S. (2015). Foundations of dynamic learning analytics: Using university student data to increase retention. *British Journal of Educational Technology*, *46*(6), 1175–1188.
11. DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (1999). An event history model of student departure. Economics of Education Review 18(1), 375-390.
12. Ellis, R. A., Han, F., & Pardo A. (2017). Improving Learning Analytics – Combining Observational and Self-Report Data on Student Learning. *Educational Technology & Society, 20*(3), 158–169
13. Felder, R., Felder, G., Mauney, M., Hamrin, C., & Dietz, E. (1995). A Longitudinal Study of Engineering Student Performance and Retention. III. Gender Differences in Student Performance and Attitudes. *Journal of Engineering Education, 87*(4), págs. 469-480
14. Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304.
15. French, B. F., Immekus, J. C., & Oakes, W. C. (2005). An examination of indicators of engineering students ' success and persistence. *Journal of Engineering Education., 94*, 419–425.
16. Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *Internet and Higher Education*, *28*, 68–84.
17. Graham, C. R. (2006). *Blended learning systems*. The Handbook of Blended Learning.
18. Hart, S. A., & Ganley, C. M. (2017). Individual differences related to college students' course performance in Calculus II. *Journal of Learning Analytics*, *4*(2), 129–153.
19. Howard, E., Meehan, M., & Parnell, A. (2018). Contrasting prediction methods for early warning systems at undergraduate level. *Internet and Higher Education*, *37*(January), 66–75.
20. Lavy, S. (2017). Who benefits from group work in higher education? An attachment theory perspective. *Higher Education, 73*(2), 175–187.
21. Marra, R. M., Rodgers, K. A., Shen, D., & Bogue, B. (2009). Women Engineering Students and Self-Efficacy: A Multi-Year, Multi-Institution Study of Women Engineering Student Self-Efficacy. *Journal of Engineering Education, 98*(1), pp. 27-38.

22. Marton, F., & Saljo, R. (1976). On qualitative differences in learning: I. Outcome and process. *British Journal of Educational Psychology, 4*6(1), 4-11.
23. Meyer, M., & Marx, S. (2014). Engineering dropouts: A qualitative examination of why undergraduates leave engineering. *Journal of Engineering Education, 103*(4), 525–548.
24. Michinov, N., Brunot, S., Le Bohec, O., Juhel, J., & Delaval, M. (2011). Procrastination, participation, and performance in online learning environments. *Computers and Education*, *56*(1), 243–252.
25. Núñez, A. (2004). Using Segmented Assimilation Theory to Enhance Conceptualization of College Participation. *InterActions: UCLA Journal of Education and Information Studies, 1*(1), Art 4.
26. Pascarella, E. T. & Terenzini, P. T. (2005). How college affects students, volume 2. San Francisco, CA: Jossey-Bass.
27. Ramsden, P. (1979). Student learning and perceptions of the academic environment. *Higher Education, 8*(4)411-427.
28. Smola, A. J., Schölkopf, B. (2004). A Tutorial on Support Vector Regression. *Statistics and Computing archive,* 14(3), pp. 199-222.
29. Strang, K. D. (2017). Beyond engagement analytics: which online mixed-data factors predict student learning outcomes? *Education and Information Technologies*, *22*(3), 917–937.
30. Tinto, V. (1988). Stages of student departure: Reflections on the longitudinal character of student leaving. Journal of Higher Education, 59(4), 438-455.
31. Vogt, C. (2008). Faculty as a critical juncture in student retention and performance in engineering programs. *Journal of Engineering Education, 97*(1), 27–36
32. Vogt, C. M., Hocevar, D., & Hagedorn, L. S. (2007). A Social Cognitive Construct Validation: Determining Women's and Men's Success in Engineering Programs. *The Journal of Higher Education, 78*(3), 337–364.
33. You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *Internet and Higher Education*, *29*, 23–30.
34. Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3-17.