Combining University Student Self-Regulated Learning Indicators and Engagement with Online Learning Events to Predict Academic Performance

Abelardo Pardo, Feifei Han, and Robert A. Ellis

Abstract—Self-regulated learning theories are used to understand the reasons for different levels of university student academic performance. Similarly, learning analytics research proposes the combination of detailed data traces derived from technology-mediated tasks with a variety of algorithms to predict student academic performance. The former approach is designed to provide meaningful pedagogical guidance, while the latter is designed to identify event patterns and relations that can be translated into actionable remediation. The benefits of both approaches have motivated this study to investigate if a combination of the self-report data and data arising from an observation of the engagement of students with online learning events offers a deeper understanding and explanation of why some students achieve relatively higher levels of academic performance. In this paper we explore how to combine data about self-regulated learning skills with observable measures of online activity in a blended learning course to increase predictive capabilities of student academic performance for the purposes of informing teaching and task design. A case study in a course with 145 students showed that the variation of the students' final score for their course is better explained when factors from both approaches are considered. The results point to the potential of adopting a combined use of self-report and observed data to gain a more comprehensive understanding of successful university student learning.

Index Terms—Education, computer-assisted instruction, learning management systems, personalized e-learning

1 Introduction

For over a decade, higher education has experienced a rapid development through integration of Internet and Web-based technologies as part of the student learning experience. This change has resulted in the widespread adoption of blended learning contexts. Advances in research have redefined the boundaries of online learning, recognizing it as an ecological phenomenon which is made up of a number of interrelated aspects involving students, teachers, technologies and physical and virtual space [1].

Learning in a blended environment requires students to engage in not only cognitive processes, (e.g., adopting appropriate learning strategies and activating prior knowledge), but also metacognitive processes (e.g., self-regulated strategies), motivational processes (e.g., self-efficacy and intrinsic motivation), and affective processes (e.g., anxiety and joy) [2], [3], [4], [5], [6], [7], [8]. To account for the diversity of factors which affect quality of blended learning, self-regulated learning (henceforth SRL) theories in educational research offer a way into understanding why some groups of students are more successful than others.

The disciplines of Learning Analytics and Educational Data Mining present two distinct research paradigms. They seek to improve quality of online and blended learning by collecting massive data sets (i.e., big data) about students' learning processes through a wide range of conventional learning management platforms, such as Moodle and Blackboard, or through customer designed virtual appliances. Using complex algorithms, the numbers derived from the digital footprints of students reflect how students engage with a variety of online learning activities to provide insights to help understand and improve students' learning experience [9]. However, big data analysis from a purely technological perspective has its own limitations, which reduces the potential of its significance [10]. Without a proper educational theoretical framework, results produced by algorithms can be difficult to translate into meaningful pedagogical guidance. Analogously, generic guidelines about effective pedagogical strategies still require to be contextualized with as much information as possible from the learning environment. The complex interactions between different factors in a learning environments requires a holistic approach in which the data captured from student interactions is combined with a theoretical underpinning to enhance its reliability. In this paper we report a case study

- *A. Pardo is with the School of Electrical and Information Engineering, The University of Sydney, NSW 2006, Australia. E-mail: Abelardo.pardo@sydney.edu.au.*
- *F. Han and R.E. Ellis are with the Institute for Teaching and Learning, The University of Sydney, NSW 2006, Australia. E-mail: {Feifei.han, Robert.ellis}@sydney.edu.au.*

Manuscript received 28 Mar. 2016; revised 9 Nov. 2016; accepted 8 Dec. 2016. Date of publication 14 Dec. 2016; date of current version 16 Mar. 2017.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TLT.2016.2639508

that investigates the complex relationship between several variables reflecting self-regulated learning features, a set of indicators of student engagement with online learning events, and their academic performance. The results clearly show the benefits of this type of combined analysis. The factors based on self-regulated learning features and the events in a blended learning environment offer a statistically better prediction of academic performance when they are combined. This result provides robust evidence of the advantages of combining self-reported and observed data sources to gain more precise insight of the learning experience leading to more effective overall improvements.

The rest of the paper is organized as follows. Section 2 describes the related work in the areas of self-regulated learning and learning analytics. Section 3 describes the experimental method used, the context, and the instruments for data collection. The results of the study are described in Section 4 and discussed in Section 5. A set of conclusions and avenues for future research are presented in Section 6.

2 Related Work

2.1 Self-Regulated Learning

SRL is defined as the extent to which students are motivationally, metacognitively, and cognitively engaged in their learning processes [11]. Despite the fact that there are multiple camps in SRL placing emphases on different elements in the theoretical frameworks, e.g., [12], [13], all the theoretical models in SRL share the following four common underlying assumptions [14]. First, students are seen to play an active role to construct meaning from their prior knowledge and the context of their learning. Second, students have the potential to monitor and regulate their own cognition and motivation. Furthermore, students have abilities to modify their cognition and motivation to achieve the goals they set in learning. Lastly, SRL is shaped by an interaction between the learning context and students own characteristics.

Despite the basic assumptions shared by theoretical models of SRL, each model has their distinct features. Some models conceive SRL as an event-based phenomenon, e.g., [2], [3], whereas others view SRL as a progression through metacognitive monitoring and control [15], [16]. Winne [15] and Winne and Hadwin [17] model adopts an information processing perspective, whereas the model proposed by McCaslin and Hickey [18] starts from a sociocultural standpoint.

The study described in this paper adopts Pintrich's social-cognitive view of SRL where students' motivational, affective, and cognitive processes are jointly shaped by their perception of external environment (i.e., the face-to-face and online learning environment, and task characteristics) and internal environment (i.e., students' state of mind), which may either facilitate or hamper their SRL, and in turn affect their learning outcomes [14], [19], [20]. The questionnaire answered by the students contained questions about these two aspects. To obtain a more comprehensive picture of how SRL influence students' learning outcomes in a blended learning context, we investigated the complex relationship between a set of SRL variables including self-efficacy and intrinsic motivation, test anxiety, use of self-regulated learning strategies, and engagement with online learning events, with levels of academic performance.

In recent years, SRL has been used more widely than before to investigate students' learning in computer-based learning (CBL) and hypermedia learning in higher education [21]. These studies have focused on how learner characteristics relate to students' SRL in the CBL e.g., [13], [22], [23]; characteristics of hypermedia learning tasks [24], [25]; and how various learning support can enhance students' SRL, e.g., [26], [27], [28], [29], [30].

As the presented study is concerned with learner characteristics in SRL, we review empirical studies in this area. In general, past research has shown that learner characteristics, such as prior knowledge, ability, motivation, and self-regulated strategy use, affect students SRL processes and learning outcomes in CBL. To be more specific, students with more prior knowledge tended to plan and monitor their learning more often than students with less prior knowledge, e.g., [27]. Students who obtained a higher achievement were more likely to use appropriate and active learning strategies, such as summarizing and making inferences compared to their counterpart with a lower achievement, e.g., [22]. Students' motivational beliefs in SRL, such as self-efficacy and goal orientation, were found to be related to academic success in CBL and hypermedia learning. For instance, Joo, Bong and Choi [31] reported that students' self-efficacy for SRL was not only positively related to efficacy in academic learning, efficacy in using internet, and self-reported strategy use in web-based learning but also predicted academic performance. Students with higher-level of learner control were more likely to gain more in learning than students with lower level of learner control [32]. One of the limitations of past studies is that they often focused only on the metacognitive aspect of SRL "*but has paid little attention to motivation and self-reactions, the more social cognitive aspects of SRL*", failing to "*capture SRL in its entirety*" [21, p. 441]. In this study we attempted to investigate the combination of a number of SRL variables, namely self-efficacy, intrinsic motivation, test anxiety, and the use of self-regulated learning strategies.

2.2 Use of Learning Analytics in Higher Education

The collection of data from technology-mediated activities to create predictive models of user behavior has been used with increasing frequency in areas such as marketing, financial markets, sports, health, etc. In recent years, postsecondary educational institutions started to use the data captured in their technology-mediated activities and apply data mining algorithms to better understand students' learning processes. The early initiatives explored how to translate conventional business intelligence techniques to postsecondary institutions [33]. The first problem they addressed at the level of institutions was student retention [34] and the area was initially called *academic analytics*. Over time the data collected about students was used within institutions to gain a deeper insight on learning processes and improve the overall quality of the student learning experience at the level of degree programs and even individual courses. In this context the areas of *educational data mining* and *learning analytics* emerged [9], [35]. The data collected from learning environments is now being used for a wide variety of support mechanisms. For example, Arnold, Hall, Street, Lafayette and Pistilli [36] used data to derive a traffic-light indicating the risk of students failing a course. Similar initiatives are included in the systems known as *Early Warning Systems* [37], [38]. Dashboards containing data visualizations are provided to students and instructors to help them reflect on their learning [39], [40], [41], and some authors propose the use of data to improve learning design [42]. Other applications of learning analytics techniques rely on data to provide academic advising [43], but the vast majority focus on predicting academic performance at various levels, e.g., [44], [45], [46], [47].

But the use of data collected from technology mediated interactions and its use in algorithms is often insufficient to gain a comprehensive vision of a learning experience. Suthers and Vebert [48] proposed the need for a *multivocal* approach to combine the contributions of research areas such as education, learning sciences and computer science, to inform and guide educational design and learning. Pardo, Ellis and Calvo [49] presented some preliminary work on how the combination of self-reported and observed data could be combined, whose outcomes warranted further investigation and in part motivated this study. The study compared the insight extracted using only self-reported information, or only observational data and showed their differences.

The study described here examines the interrelationships amongst self-regulated learning, students' interaction with online learning events, and students' academic performance in a first year engineering course. A detail statistical analysis is provided to ascertain the effect of combining these data sources. The study was guided by the following research questions:

1. What is the relationship between self-efficacy, intrinsic motivation, test anxiety, self-regulated learning strategy (henceforth SRL variables), engagement with online learning events (henceforth observed variables), and academic performance?
2. To what extent do SRL and observed variables contribute to our understanding of variation in students' academic performance?

3 Method

3.1 Participants

The participants of the study were 145 first year undergraduate students, who majored in a four year Bachelor of Engineering degree in a large research-intensive university in an Australian metropolitan area. Among the 145 students, 74.5 percent were female, and 24.8 percent were male.

3.2 The Context of Research

The research was conducted in the 13-week course "Introduction to Computer Systems" that is compulsory for first year engineering students. The four major learning outcomes of the course are: 1) students should be able to design, build, configure, program, and test an electronic system for a specific engineering problem through observing common professional practice; 2) students should understand theoretical knowledge of how computers work from the digital logic level to basic programming constructs; 3) students should be equipped with abilities to draft reports about the design process; and 4) students will have hands-on experience in team-based design work and creative tasks in order to solve an engineering problem. The course also aimed at nurturing students' personal and intellectual autonomy, and an in-depth appreciation of ethical and professional issues.

The design of the course represents a typical blended learning experience in tertiary education. It consisted of compulsory face-to-face learning experiences and of a significant amount of online learning tasks. The face-to-face component involved a weekly two-hour lecture, a weekly two-hour tutorial, and a weekly three-hour laboratory session, whereas the online component was hosted in a custom-designed learning management system capable of monitor and recording a comprehensive set of interactions with the available resources. All the course resources were available through the online platform. Students were required to complete the online tasks in their own learning time. The online tasks were scheduled to prepare the face-to-face plenary sessions and tutorials. The lecture preparation activities required students to view several videos per week, read course notes, answer multiple-choice questions, and solve a sequence of exercises. A 10 percent final mark value together with a deadline to submit the answer to the exercises was set before the start of the lecture to encourage student engagement. The activities to prepare the tutorial session required students to answer an additional set of exercises before the beginning of the session and counted an additional 10 percent towards the final course mark. The weekly events captured by the learning management system in the preparation activities were summarized in almost real time and made available to the students through a dashboard showing their engagement and the average of the entire cohort. The data was reset at the start of each week. The rest of assessments in the course included a multiple-choice question midterm exam (20 percent), lab report and project design (20 percent), and a final exam (40 percent) with multiple-choice and open questions.

3.3 Instruments

Three types of data sources were used for the study:

1. Students' self-regulated learning variables, including motivational, affective, and cognitive aspects, were collected using a self-reported questionnaire.
2. Information about the interaction with the online learning events through the learning management system.
3. Academic performance as the final marks in the course.

The following sections provide a more detailed description of each data source.

3.3.1 Motivated Strategies for Learning Questionnaire

The motivational beliefs, affect in learning, and the use of self-regulated learning strategies by the students were collected using a subset of the Motivated Strategies for Learning Questionnaire, henceforth, MSLQ [19]. The original instrument has 44 items to be answered using a seven-point Likert scale answer. In our study we selected 31 items to focus in four of the five scales in the questionnaire, namely: Self-efficacy (9 items), Intrinsic value (9 items), Test anxiety (4 items), and Self-regulation (9 items). The section of the instrument with questions about the use of cognitive strategy was removed to focus on those features more directly related to self-regulation. The scale about test anxiety was retained due to the presence of two test-based components in the course assessment (midterm and part of the final exam).

3.3.2 Student Engagement with Online Learning Events

The engagement of students with the course resources was captured through the digital footprints left in the learning management system. The platform hosted all course notes, task descriptions, questions, videos and additional documents in electronic format. Every interaction with these resources was recorded with a time stamp and the user id. For this study the following seven event categories were considered:

1. Resource view (Resource). These events are recorded whenever a student accesses any page of any document in the system.
2. Collapse and expand a section of course notes (Col-Exp). Electronic documents with sections appear with their content collapsed or not visible, and are expanded by clicking in the section title. These events are recorded whenever any student collapses or expands the content of any of these sections.
3. Video event (Video). These events are recorded when a video is loaded within a page, the play button is pressed, the pause button is pressed or the end of the video is reached. The four events are not distinguished.
4. Video multiple-choice questions (VMCQ). These events are recorded whenever students interact with the multiple-choice questions next to the videos. Three different events are recorded in this category: question is answered correctly, question is answered incorrectly, and the student requests to see the answers (only possible after providing one answer).
5. Multiple-choice questions (MCQ). These events are identical to those in the previous category but for the questions in the course notes (not attached to the video activities). Both type of questions were formative (no impact in the course score).
6. Exercise sequences (Exercise). These events are recorded when the students answer any of the exercises in the sequences to prepare the lecture and tutorial sessions. The score of this sequence is part of the summative assessment. These exercises require the student to understand some stimulus content, engage in reasoning in a small case-study type example, and then select the correct answer.
7. Dashboard views (Dboard). This event is recorded every time a student reviews the levels of their weekly engagement on the dashboard.

The system recorded the events occurring during the 13 weeks of the semester. The accumulated event counts over this period were used for the study. The descriptive statistics of these seven variables for the students that participated in the study are presented in Table 1

The results show that the frequency of engagement with the online learning events exhibits large differences among them, but also between students. The mean (M) values ranged from 31.10 for dashboard views to 2723.49 for the number of sequence exercises answered. The high value of this type of events is related with the fact that they were part of the summative assessment. All event types also show large standard deviations (SDs) implying a large dispersion of values among individuals. To account for this disparity, all the variables were standardized (Z-score) with a mean of 0 and a standard deviation of 1 for the data analysis.

TABLE 1
Descriptive Statistics of Engagement with Online Learning Events

Variables	Min	Max	Mean	SD
Dboard	0	233	31.10	41.84
Col-Exp	59	1182	421.97	234.36
Resource	138	2492	818.07	443.00
Video	0	2890	338.59	395.48
MCQ	0	3054	233.01	300.50
VMCQ	0	5598	191.05	471.17
Exercise	353	9957	2723.49	1419.81

Note: Dboard = Dashboard views, Col-Exp = collapse and expand a section, Resource: access to any document, Video: any event in a video, MCQ = multiple-choice questions, VMCQ = multiple-choice questions embedded in videos, Exercise: answer to exercise in sequence.

3.4 Academic Performance

The final marks in the course were used as the measure of academic performance. The values in this variable were in the range 20 to 98.50 (out of 100), with a mean value of 65.50, and a standard deviation of 16.12. The pass mark for the course is set to 50, which means the average score is above the pass mark, but with a wide spread in the values. This scenario is fairly common in first-year university courses.

The final mark was computed as the accumulated sum of six assessments. The sequence of exercises to prepare each lecture had a weight of 1 percent over a period of 10 weeks for a total of 10 percent. The sequence of exercises to prepare the tutorial sessions had the same policy and accounted for another 10 percent. A written laboratory report documenting one of the laboratory sessions accounted for 5 percent. Students had access to the grading rubric for the report and were allowed to make multiple submissions (each of them graded) over a period of four weeks. A design project requiring the design of a computer system, programming, a written report describing its structure, a presentation and a demonstration accounted for 15 percent of the course mark. The scores of the midterm and the final exams contributed to 20 percent and 40 percent respectively. To facilitate interpretation of the results, the total mark was also transformed into a Z-score with a median of 0 and a standard deviation of 1.

3.5 Data Collection Procedure

Data collection followed the ethical procedures stipulated in the Human Ethics Committee of the University where the data was collected. Students were informed that participating in the study was completely voluntary and their written consent was obtained to use their answers to the questionnaire, their final marks, and their digital footprint in the Learning Management System.

3.6 Data Analysis

We used the following statistical methods to analyze the data. First, we performed a series of Exploratory Factor Analysis (EFA) using Principal Component procedures followed by Varimax rotation to examine the scale structure of the answers the MSLQ. This method was used to identify which items of the questionnaire were more relevant and how to group them according to the similarity of the answers. To determine the number of scales we reviewed a screen plot and deleted those items with coefficients less than .40 within a scale and those with high multiple coefficients loaded across scales [50]. The scales with the highest interpretability were retained [51]. To evaluate the internal consistency we calculated the Cronbach's alpha reliability analyses for each resulting scale. The result of this step was a set of four scales that grouped the items of the questionnaire.

To investigate the relationship between the scales in the questionnaire, the interactions with online learning events, and students' academic performance we adopted three methods. The objective was to show the relationships both at the level of variables and at the level of sub groups of students within the population sample. The use of multiple methods also contributes to the integrity of the whole analysis [52]. The methods selected were influenced by the presence of the academic performance. Hierarchical clustering analysis was chosen to identify which factors better distinguish the variation in academic performance. This technique calculates the ideal number of clusters to maximize the differentiation and the statistical descriptors characterizing each factor in the cluster. Additionally, regression analysis was chosen to explore the potential linear relationship between the numerical factors and the academic performance.

At the level of variable, we first used correlation analysis to display the interrelationship between pairs of variables. Principal Component Analysis was then used to explore the relationship between the scales from the questionnaire, the aggregated scores of frequency of all the engagement with events, and the academic performance results. At the level of student, we adopted a hierarchical cluster analysis using the motivational and self-regulated variables, and students' academic performance to identify subgroups of students to maximize similar learning experiences within groups and different learning experiences between groups. On the basis of the cluster membership computed with this method, one-way ANOVA analysis was performed to see whether students in different clusters differ from each other on the motivational, self-regulated, and academic performance. Then one-way ANOVA analysis was also conducted to examine whether students' engagement with online learning events differed between groups.

To investigate how motivation, self-regulation, and engagement with online learning events, contributed to the academic performance, we conducted a hierarchical multiple regression analysis. The construction of the model is based on the results from the bivariate correlation analyses so that only variables that significantly correlated with academic outcomes are considered. This method is especially adequate when handling a large number of factors and the most relevant ones need to be identified. Additionally, the model provides an indication of its quality which can be used to compare different solutions and quantify this difference.

Two models were constructed using multiple regression techniques. The first model included the academic performance (dependent variable) and the self-regulated learning variables from the questionnaires (independent variables) [12], [53], [54]. The second model was constructed adding the variables derived from the student engagement. In both models the effect size was calculated using Cohen's f^2. According to Cohen [55], a value of .02 is considered a small effect, and a value of .15 and .35 are indicative of medium and large effect respectively.

4 Results

4.1 EFAs and Reliability of Scales

The results of EFA for the MSLQ are presented in Table 2. Out of the initial 31 items in the questionnaire, 24 were retained generating five scales, namely: self-efficacy (EF), intrinsic motivation (IM), test anxiety (TA), positive self-regulation (PS), and negative self-regulation (NS). The remaining items were removed because the rotated factor loading were below the value of 0.4. The Eigen-values of the five scales were 3.88, 3.16, 2.80, 2.16, and 2.03 respectively, and they accounted for 17.61, 14.28, 12.73, 9.84, and 9.21 percent of the total variance. The reliability analysis showed that the five scales had Cronbach's alpha of .89, .85, .84, .68, and .72, indicating acceptable reliability for all of them.

4.2 Correlation Analysis

The pairwise correlations among variables are displayed in Table 3 (bold values denote statistically significant correlation). These numbers are indicators of how similar are the variations when considering pair-wise comparisons among factors. Positive correlation means similarity in both magnitude and sign, whereas negative variation means similarity in magnitude but with opposite sign (one factor varies at a similar rate than the negative of the other factor). As it can be seen, self-efficacy (SE) is significantly and positively associated with intrinsic motivation ($r = .45, p < .01$) and positive self-regulated strategy ($r = .48, p < .01$). Additionally, intrinsic motivation is also significantly and positively related to positive self-regulated strategy ($r = .35, p < .01$), even though the value was lower than those between self-efficacy and positive self-regulated strategy.

On the other hand, self-efficacy was found to have negative correlation with anxiety ($r = -.18, p < .05$) and negative self-regulated strategy ($r = -.22, p < .05$), so was the correlation between intrinsic motivation and negative self-regulated strategy ($r = -.23, p < .05$). The values of the three negative correlations were small. In contrast, test anxiety was positively and moderately associated with negative self-regulated strategy ($r = .48, p < .01$), but the correlations of test anxiety with intrinsic motivation ($r = -.09, p = .30$) and positive self-regulated learning were not significant ($r = .02, p = .77$). We can also see that the correlation between positive and negative self-regulated strategy did not reach significance ($r = -.13, p = .13$). These results point to relationships between the scales that have been described in the research literature.

The relationship between qualitative variables and those capturing the engagement with online learning events

TABLE 2
Results of EFA for MSLQ

Scales	Description of items	Rotated factor loadings				
Self-efficacy (.89)	Compared with other students in this class I expect to do well.	.59				
	I'm certain I can understand the ideas taught in this course.	.69				
	I expect to do very well in this class.	.76				
	Compared with others in this class, I think I'm a good student.	.65				
	I am sure I can do an excellent job on the problems and tasks assigned for this class.	.79				
	I think I will receive a good grade in this class.	.82				
	Compared with other students in this class I think I know a great deal about the subject.	.66				
	I know that I will be able to learn the material for this class.	.73				
Intrinsic Motivation (.85)	It is important for me to learn what is being taught in this class.		.73			
	I think I will be able to use what I learn in this class in other classes.		.69			
	I think that what I am learning in this class is useful for me to know.		.81			
	I think that what we will learn in this class is interesting.		.69			
	Understanding this subject is important to me.		.85			
Test Anxiety (.84)	I am so nervous during a test that I cannot remember facts I have learned.			.72		
	I have an uneasy, upset feeling when I take a test.			.85		
	I worry a great deal about tests.			.81		
	When I take a test I think about how poorly I am doing.			.74		
Positive Self-regulated strategy use (.68)	I ask myself questions to make sure I know the material I have been studying.				.53	
	I work on practice exercises and answer end of chapter questions even when I don't have to.				.62	
	Even when study materials are dull and uninteresting, I keep working until I finish.				.73	
	I work hard to get a good grade even when I don't like a class.				.71	
Negative self-regulated strategy use (.72)	When work is hard I either give up or study only the easy parts.					.69
	I often find that I have been reading for class but don't know what it is all about.					.75
	I find that when the teacher is talking I think of other things and don't really listen to what is being said.					.67

Values less than .40 removed; KMO (Kaiser-Meyer-Olkin measure of sampling adequacy): .83.

shows some interesting trends. Intrinsic motivation seems to have almost no relation with online activities (except with the use of video). However, positive self-regulation shows significant correlation with most of the quantitative variables. These results seem to suggest that students who adopted a positive self-regulated strategy tended to interact more frequently with online learning events. A different pattern emerges when considering negative self-regulation as no quantitative variable shows statistically significant correlation. This seems to suggest a lack of connection between negative self-regulation and engagement with learning events. The last row in the table also highlights the significant correlation of most of the quantitative variables with academic performance. These relationships point to a potential increase in the capacity of all the variables to explain a larger variation of academic performance.

4.3 Cluster Analysis and One-Way ANOVA

A hierarchical cluster analysis using Ward's method was conducted using the students' self-regulated learning variables (from MSLQ) and the academic performance. The aim of this analysis is to explore the presence of subgroups of learners that can be differentiated based on their self-regulation and academic performance. Based on the increasing value of the squared euclidean distance between clusters, a two-cluster solution was obtained. Obtaining the clusters, although relevant, it is not a satisfactory solution as it needs to be shown if the factors under study show variation among the clusters. On the basis of membership, a series of ANOVA were performed to examine whether there were significant differences of the mean for the self-regulation variables. The ANOVA analysis was also extended to

TABLE 3
Correlation Analysis

Variables	TA	IM	PS	NS	AP	DB	CE	RE	VI	MC	VM	EX
SE	−.18*	**.45****	**.48****	**−.22****	.10	.11	.02	.09	.09	.13	.12	.16
TA	—	−.09	.02	**.48****	**−.28****	.07	.06	.01	.07	.02	−.05	.04
IM	—	—	**.35****	**−.23****	.01	.06	−.01	.02	.21*	.10	.11	.14
PS	—	—	—	−.13	−.02	.18*	.21*	**.23****	.14	.17*	.06	**.28****
NS	—	—	—	—	−.20*	−.03	.15	.10	−.03	−.01	−.11	.02
AP	—	—	—	—	—	**.24****	**.35****	**.44****	.14	**.28****	.14	**.38****
DB	—	—	—	—	—	—	**.34****	**.35****	.19*	**.41****	**.35****	**.50****
CE	—	—	—	—	—	—	—	**.93****	**.24****	**.31****	.06	**.74****
RE	—	—	—	—	—	—	—	—	**.25****	**.28****	.06	**.76****
VI	—	—	—	—	—	—	—	—	—	.11	.08	**.52****
MC	—	—	—	—	—	—	—	—	—	—	**.85****	**.73****
VM	—	—	—	—	—	—	—	—	—	—	—	**.57****

Note: $^{**}p < 0.01$, $^{*}p < 0.05$, SE: Self-efficacy, TA: test anxiety, IM: intrinsic motivation, PS: Positive self-regulated strategy use, NS: negative self-regulated strategy use, AP: academic performance, DB: dashboard view event, CE: collapse-expand event, RE: access to a resource, VI: video event, MC: multiple-choice question, VM: multiple-choice question in videos, EX: exercise in sequence.

include the variables encoding the use of the online learning events. The results are shown in Table 4.

The 145 students in the study were classified in two groups: a group of 62 students denoted "High self-regulated and high achieving", and a group of 83 students denoted as "Low self-regulated and low achieving" cluster. Based on the group membership, we found that the two groups of students differed significantly on all self-regulation variables as well as academic performance with statistically significant differences and effect sizes as reported by eta square values.

Students in the "High self-regulated and high achieving" cluster had significantly higher ratings on self-efficacy ($M = 0.55$), intrinsic motivation ($M = 0.48$), positive self-regulated strategy use ($M = 0.64$), and achieved significantly better in the course ($M = 0.20$) than "Low self-regulated and low achieving" students. However, they tended to feel less anxious about tests ($M = -0.28$) and adopted less of negative self-regulated strategy ($M = -0.60$) than those in the "Low self-regulated and low achieving" group.

In contrast, the "Low self-regulated and low achieving" tended to have significantly lower ratings on self-efficacy ($M = -0.41$), intrinsic motivation ($M = -0.36$), positive self-regulated strategy use ($M = -0.48$), higher ratings on test anxiety ($M = 0.21$) and negative self-regulated strategy use ($M = 0.45$), and tended to perform significantly poorly ($M = -0.15$) than their counterparts in the "High self-regulated and high achieving" cluster.

In relation to engagement with online learning events between the two clusters, we found statistically significant differences on three out of the seven event types: Dashboard views $F(1,144) = 7.46, p < .05, \eta^2 = .05$, multiple-choice questions $F\,(1,\,144) = 4.15,\ p < .05,\ \eta^2 = .03$, and solving exercise sequences $F\,(1,\,144) = 12.58,\ p < .01,\ \eta^2 = .04$. The two groups of students did not show any significant difference with respect to the events Col-Exp $F(1,144) = 0.62, p = .43, \eta^2 = .00$, Resource $F(1,144) = 2.11, p = .15, \eta^2 = .02$, Video $F\,(1,\,144) = 1.65,\ p = .20,\ \eta^2 = .01$, and MCQ embedded in video activities $F\,(1,\,144) = 3.74,\ p = .06,\ \eta^2 = .03$.

TABLE 4
Result of One-Way ANOVA of the Student Clusters

Variables	High self-regulated and high achieving (N = 62) Mean (*SD*)	Low self-regulated and low achieving (N = 83) Mean (*SD*)	*F*	*p*	η^2
		Self-regulated Learning			
Self-efficacy	0.55 (0.80)	−0.41 (0.94)	41.15	.00	.22
Test Anxiety	−0.28 (1.02)	0.21 (0.94)	8.74	.00	.06
Intrinsic motivation	0.48 (0.79)	−0.36 (0.99)	29.72	.00	.17
Positive self-regulated strategy use	0.64 (0.75)	−0.48 (0.90)	63.46	.00	.31
Negative self-regulated strategy use	−0.60 (0.71)	0.45 (0.94)	55.00	.00	.28
		Engagement with online learning events			
Dboard	0.26 (0.93)	−0.19 (1.01)	7.46	.01	.05
Col-Exp	0.08 (0.98)	−0.06 (1.00)	0.62	.43	.00
Resource	0.14 (1.06)	−0.10 (0.95)	2.11	.15	.02
Video	0.12 (0.98)	−0.09 (1.01)	1.65	.20	.01
MCQ	0.19 (1.35)	−0.14 (0.60)	4.15	.04	.03
VMCQ	0.18 (1.49)	−0.14 (0.26)	3.74	.06	.03
Exercise	0.22 (1.11)	−0.16 (0.88)	5.41	.02	.04
		Outcomes			
Academic performance	0.20 (0.95)	−0.15 (1.01)	4.54	.04	.03

TABLE 5
Results of Multiple Regression Analysis

Variables	B	SE B	β	t	Adj. R^2	p	f^2
Model 1					.07**	.00	.08
Test Anxiety	−2.89	1.09	−.24**	−2.66		.01	
Negative Self-Regulation	−1.11	1.24	−.08	.90		.37	
Model 2					**.32****	.00	.37
Test Anxiety	−2.60	0.95	−.22*	-2.75		.01	
Negative Self-Regulation	−1.54	1.09	−.11	−1.40		.16	
Dboard	0.03	.03	.07	0.84		.40	
Col-Exp	−0.02	.01	−.32	−1.69		.09	
Resource	0.03	.01	.85**	4.12		.00	
MCQ	0.02	.01	.31*	2.56		.01	
Exercise	−0.01	.01	−.28	−1.49		.14	

*Note: **p < .01, *p < .05, SE B: Standard error in B, Adj R2: Adjusted R-square*

As it can be seen, the students in the "High self-regulated and high achieving" cluster interacted significantly more frequently with the dashboard and the exercise sequences than their classmates in the "Low self-regulated and low achieving".

4.4 Multiple Regression Analysis

The academic performance may have a linear relationship with a subset of the factors under consideration. Multiple regression methods allow you to calculate which factors are more likely to be linearly related to the dependent variable, and identify the linear coefficients of such dependency. Before performing the multiple regression analysis, a series of assumption tests were run to examine whether the data met the required assumptions. A standard residuals test showed that no outliers were found (Std. Residual $\mathrm{Min} = -2.21$, Std. Residual $\mathrm{Max} = 2.60$). Multicolinearity tests showed that the values of Tolerance and VIF were within the acceptable limits, thus that the data met the assumption of multicolinearity. The value of the Durbin-Watson suggested that the assumption of independence of errors were satisfied (Durbin-Watson = 2.20).

The results of the hierarchical multiple regressions are presented in Table 5. The first model (Model 1) only considered self-regulated learning variables as independent variables. The results show that only test anxiety and negative self-regulated learning significantly predicted the academic performance with a small effect $[F(1, 143) = 6.18, p <$ $, .01, f^2 = .08]$. The two variables accounted only for 7 percent of the variance.

A second model was obtained (Model 2) considering the additional independent variables encoding student engagement with online learning events. This model explains an additional 26 percent of variation in academic performance in the course, $F(7, 138) = 10.44, p < .01, f^2 = .37$, for a total of 32 percent, and a large effect size.

The second model also shows that not every independent variable was a significant predictor of students' academic performance. To be more specific, three variables, test anxiety ($\beta = -.22, p < .05$), Resource ($\beta = .85, p < .01$), and MCQ ($\beta = .31, p < .05$) significantly contributed to academic performance. Negative self-regulated strategy use ($\beta = -.11, p = .16$), interaction with Dashboard ($\beta = .07$, $p = .40$), and interaction with exercises ($\beta = -.28, p = .$ 14) were not significant predictors of academic performance.

5 Discussion

The aim of the study was to explore to what extent the self-report and observed variables of the students experience of learning could explain variation in academic performance, which would help to predict why some students are more successful than others. The first set of variables (SRL) contained self-reported qualitative self-regulation variables (including indicators for self-efficacy, intrinsic motivation, test anxiety, and positive and negative use of a self-regulated strategy). The second set (quantitative) contained observed indicators of the engagement of students with online learning events. The initial exploration of correlation among these variables shows an interesting point of departure for the discussion (see Table 3). Some of the SRL variables, as expected, show significant correlation among themselves, but also with academic performance. This is coherent with previously reported results, e.g., [21], [56]. More specifically, the variables capturing potentially negative aspects (test anxiety and negative use of self-regulation strategy) clearly show their negative relation with academic performance. When the correlation is compared with quantitative variables two interesting observations emerge. A significant portion of the quantitative variables correlates with the use of a positive self-regulation strategy as previously reported in other studies, e.g., [32], [57], [58]. This result confirms the connection between self-regulation and behavior in the online space. But even more revealing is the fact that five of the seven quantitative variables show a significant correlation with academic performance. The results of this first step of the analysis suggest the value of exploring how a combination of these variables may more fully account for why some students are more successful than others, knowledge which can help to inform teaching. SRL variables provide insight from the point of view of the student and suggest those factors that are shaping their learning experience, whereas the observed variables provide evidence of how students interacted with the online events in their course.

The clustering results shown in Table 4 also emphasized the value of combining the variables in the analysis of the student experience. While the correlation identified pairwise associations at the level of the variables, the cluster analysis identified sub-groups within the population sample that reported similar experiences of learning. One cluster of students (n = 62) reported a relatively higher achieving experience marked by self-efficacy, intrinsic motivation, positive self-regulation, relatively higher academic performance and greater use of the online learning environment. The second cluster of students (n = 83) reported relatively lower achieving experience marked by test anxiety, negative self-regulation, relatively lower academic performance and a lesser use of the online learning environment. These outcomes motivated the researchers to look for which variables were most likely to explain the differences in the learning experience of the two clusters of students using multiple regression analysis.

In this study, multiple regression analysis was used to identify those variables that accounted for the variation in the students learning experience. When the analysis included only the SRL variables (Table 5), only test anxiety provides a statistically significant coefficient in the regression and the amount of variation in academic performance was only 7 percent. When both the SRL and quantitative variables are used in the analysis three variables have statistically significant coefficients (test anxiety, engagement with resources, and multiple-choice questions in the course notes) and together explain 32 percent of the variation of academic performance. This last result offers solid evidence to answer the two research questions posed at the beginning of the paper. The students approach to self-regulation and their engagement with online learning events offer a more complete understanding that explains significant differences in the students' academic performance. Up to almost one third of the variation in academic performance is determined by the model when the two sets of variables are combined.

5.1 Implications

The results of this study offer a number of ideas for teaching students in blended courses. The associations amongst the variables in the correlation and cluster analyses suggest that a positive student experience of self-efficacy, tests, motivation, self regulation *and* positive interaction with many of the online events, particularly those that offer feedback (such as the dashboard and the multiple choice questions) and reflection and reasoning (such as the problem-solving exercise sequences) will correlate with relatively higher academic achievement compared with negative experiences of these aspects. These results offer strategies for teachers to explore task designs and approaches that involve revealing to students what positive self-regulation for learning involves, how to interact effectively with the online learning events, or redesigning the instructional materials accompanying these activities. The regression analysis reinforces these ideas and suggests that attention to negative aspects of the experience such as test anxiety and addressing poor student engagement with the online events through improved task design and modeling effective engagement online may improve student outcomes.

5.2 Limitations

Even though the current research provided a more nuanced picture of the relation between self-regulation, online engagement and academic performance, some limitations need to be pointed out. First, the self-reporting questionnaire was designed for the whole course experience, encompassing both face-to-face and online learning. No distinction was made between face-to-face and online learning for each variable. In future research, different items could be used so that a more detailed relationship can be revealed. Second, the self-reporting data were collected towards the end of the semester so that students had almost a complete learning experience for the course. However, the digital footprint was recorded throughout the whole course from the beginning to the end. Self-regulation is dynamic [14], [59] and students may have different perceptions of their level of self-efficacy, motivation, and test anxiety throughout the course. A potential improvement would consist on collecting self-reporting data at multiple times and take this aspect into account to identify potential dynamic relationships. Thirdly, the events recorded in the study were from a considered from a purely observational standpoint (summarized by event counts). It would be more interesting if the qualitative feature of how students engaged with different online learning activities were also traced. For example, when students engage with a video resource, pair that visualization with other resources that they use at the same time, or in the same session.

6 Conclusions and Future Work

The study described in this paper highlighted the importance of analyzing learning experiences combining the insight gained with self-reported data based in well established theoretical frameworks such as self-regulation, with those obtained by methods such as recording the interactions between students and course events in an online platform. The analysis of self-reported factors undoubtedly explains fhow students approach aspects of their learning experiences and identifies the factors that contribute to better academic outcomes. But when complemented with observed behavior in an online environment, there is a significant portion of the specific context that is brought into the analysis and the models offer a more accurate description. The study used a blended learning course in which students are required to prepare two of the three weekly sessions using online material. The relevance of this scenario is very important, as it places the combination of face-to-face and online tasks as one of the defining features. Although students reported their preferences with respect to self-regulation aspects of their learning approach, the combination with indicators of engagement online offered a linear model explaining 32 percent of the variance of the academic performance.

This result requires more extensive evaluation to see its replicability in other contexts and instructional designs. A second iteration of the case study would strengthen the insight and validity of the claims. Nevertheless, the result also suggests exciting avenues for further exploration. There is a vast range of possibilities to study models of student learning that combine self-report and observational sources of data that offer explanations and predict academic performance more adequately. Ultimately, we envision an approach in which these sources of data are used in increasingly useful combinations to provide a more nuanced description of learning to both teachers and students for the purposes of improving teaching and learning.

References

[1] R. A. Ellis and P. Goodyear, *Students' Experiences of E-learning in Higher Education*. New York, NY, USA: Routledge, 2013.

[2] R. Azevedo, D. C. Moos, A. M. Johnson, and A. D. Chauncey, "Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges," *Educ. Psychologist*, vol. 45, no. 4, pp. 210–223, 2010.

[3] J. A. Greene and R. Azevedo, "The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments," *Educ. Psychologist*, vol. 45, no. 4, pp. 203–209, 2010.

[4] D. C. Moos and E. Marroquin, "Multimedia, hypermedia, and hypertext: Motivation considered and reconsidered," *Comput. Human Behavior*, vol. 26, no. 3, pp. 265–276, 2010.

[5] D. C. Moos, "Note-taking while learning hypermedia: Cognitive and motivational considerations," *Comput. Human Behavior*, vol. 25, no. 5, pp. 1120–1128, 2009.

[6] D. C. Moos, "Self-regulated learning with hypermedia: Too much of a good thing?" *J. Educ. Multimedia Hypermedia*, vol. 19, no. 1, pp. 59–77, 2010.

[7] D. C. Moos, "Self-regulated learning and externally generated feedback with hypermedia," *J. Educ. Comput. Res.*, vol. 44, no. 3, pp. 265–297, 2011.
[8] R. Azevedo, "Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: A discussion," *Metacognition Learn.*, vol. 4, no. 1, pp. 87–95, 2009.
[9] R. S. Baker and G. Siemens, "Educational data mining and learning analytics," in *The Cambridge Handbook of the Learning Sciences*, R. K. Sawyer, Ed. Cambridge, U.K.: Cambridge Univ. Press, 2014.
[10] S. Knight, S. Buckingham Shum, and K. Littleton, "Epistemology, assessment, pedagogy: Where learning meets analytics in the middle space," *J. Learn. Analytics*, vol. 1, no. 1, pp. 23–47, 2014.
[11] B. J. Zimmerman, "A social cognitive view of self-regulated academic learning," *J. Educ. Psychol.*, vol. 81, no. 3, pp. 329–339, 1989.
[12] B. J. Zimmerman and D. H. Schunk Eds., *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*, 2nd ed. Milton Park, U.K.: Taylor & Francis, 2001.
[13] D. C. Moos and C. A. Stewart, "Self-regulated learning with hypermedia: Bringing motivation into the conversation," in *International Handbook of Metacognition and Learning Technologies*, R. Azevedo and V. Aleven, Eds. New York, NY, USA: Springer, 2013.
[14] P. R. Pintrich, "The role of goal orientation in self-regulated learning," in *Handbook of Self-Regulation*, M. Boekaerts, P. R. Pintrich, and M. Zeidner, Eds. San Diego, CA, USA: Academic Press, 2000.
[15] P. H. Winne, "Self-regulated learning viewed from models of information processing," in *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*, B. J. Zimmerman and D. H. Schunk, Eds. Milton Park, U.K.: Taylor & Francis, 2001.
[16] P. H. Winne and N. E. Perry, "Measuring self-regulated learning," in *Handbook of Self-Regulation*, M. Boekaerts, P. R. Pintrich and M. Zeidner, Eds. San Diego, CA, USA: Academic Press, 2000, pp. 531–566.
[17] P. H. Winne and A. F. Hadwin, "Studying as self-regulated learning," in *Metacognition in Educational Theory and Practice*, D. J. Hacker, J. Dunlosky, and A. C. Graesser, Eds. Abingdon-on-Thames, U.K.: Routledge, 1998, pp. 227–304.
[18] M. McCaslin and D. T. Hickey, "Educational psychology, social constructivism, and educational practice: A case of emergent identity," *Educ. Psychologist*, vol. 36, no. 2, pp. 133–140, 2001.
[19] P. R. Pintrich and E. V. de Groot, "Motivational and self-regulated learning components of classroom academic performance," *J. Educ. Psychol.*, vol. 82, pp. 33–40, 1990.
[20] P. R. Pintrich, C. A. Wolters, and G. P. Baxter, "2. assessing metacognition and self-regulated learning," in *Issues in the Measurement of Metacognition*, G. Schraw and J. C. Impara, Eds. Lincoln, NE, USA: Buros Institute of Mental Measurements, 2000.
[21] F. I. Winters, J. A. Greene, and C. M. Costich, "Self-regulation of learning within computer-based learning environments: A critical analysis," *Educ. Psychol. Rev.*, vol. 20, no. 4, pp. 429–444, 2008.
[22] R. Azevedo, J. T. Guthrie, and D. Seibert, "The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia," *J. Educ. Comput. Res.*, vol. 30, no. 1/2, pp. 87–111, 2004.
[23] P. E. Williams and C. M. Hellman, "Differences in self-regulation for online learning between first- and second-generation college students," *Res. Higher Educ.*, vol. 45, no. 1, pp. 71–82, Feb. 2004.
[24] D. H. Schunk and P. A. Ertmer, "Self-regulatory processes during computer skill acquisition: goal and self-evaluative influences," *J. Educ. Psychol.*, vol. 91, no. 2, pp. 251–260, 1999.
[25] G. Stahl, T. Koschmann, and D. Suthers, "Computer-supported collaborative learning: An historical perspective," in *Cambridge Handbook of the Learning Sciences*, R. K. Sawyer Ed. Cambridge, U.K.: Cambridge University Press, 2006, pp. 409–426.
[26] S. P. Lajoie, et al., "Technology-rich tools to support self-regulated learning and performance in medicine," in *International Handbook of Metacognition and Learning Technologies*, R. Azevedo and V. Aleven, Eds. New York, NY, USA: Springer, 2013, pp. 229–242.
[27] D. C. Moos and R. Azevedo, "Self-regulated learning with hypermedia: The role of prior domain knowledge," *Contemporary Educ. Psychol.*, vol. 33, no. 2, pp. 270–298, 2008.
[28] S. Narciss, A. Proske, and H. Koerndle, "Promoting self-regulated learning in web-based learning environments," *Comput. Human Behavior*, vol. 23, no. 3, pp. 1126–1144, 2007.
[29] A. Proske, S. Narciss, and H. Körndle, "Interactivity and learners' achievement in web-based learning," *J. Interactive Learn. Res.*, vol. 18, no. 4, pp. 511–531, 2007.
[30] V. Venkatesh, K. Shaik, A. Zuberi, K. Urbaniak, T. Gallant, and A. Lakhana, "Development of task understanding and monitoring in information retrieval environments: demystifying metacognitive and self-regulatory mechanisms in graduate learners using topic maps indexing technologies to improve essay-writing skills," in *International Handbook of Metacognition and Learning Technologies*, R. Azevedo and V. Aleven, Eds. New York, NY, USA: Springer, 2013, pp. 229–242.
[31] Y.-J. Joo, M. Bong, and H.-J. Choi, "Self-efficacy for self-regulated learning, academic self-efficacy, and internet self-efficacy in Web-based instruction," *Educ. Technol. Res. Development*, vol. 48, no. 2, pp. 5–17, 2000.
[32] W. Eom and R. A. Reiser, "The effects of self-regulation and instructional control on performance and motivation in computer-based instruction," *Int. J. Instructional Media*, vol. 27, no. 3, pp. 247–261, 2000.
[33] P. Goldstein and R. Katz, "Academic analytics: The uses of management information and technology in higher education," Educause Center for Applied Research, Boulder, CO, USA, vol. 8, 2005.
[34] J. P. Campbell, P. B. DeBlois, and D. Oblinger, "Academic analytics: A new tool for a new era," *EDUCAUSE Rev.*, vol. 42, pp. 40–57, 2007.
[35] G. Siemens, "Learning analytics: The emergence of a discipline," *Amer. Behavioral Scientist*, vol. 57, no. 10, pp. 1380–1400, 2013.
[36] K. E. Arnold, Y. Hall, S. G. Street, W. Lafayette, and M. D. Pistilli, "Course signals at purdue: Using learning analytics to increase student success," in *Proc. 2nd Int. Conf. Learning Analytics Knowl.*, 2012, pp. 267–270.
[37] A. E. Krumm, R. J. Waddington, S. D. Teasley, and S. Lonn, "A learning management system-based early warning system for academic advising in undergraduate engineering," in *Learning Analytics: From Research to Practice*, J. A. Larusson and B. White, Eds. New York, NY, USA: Springer, 2014, pp. 103–119.
[38] S. Lonn, A. E. Krumm, R. J. Waddington, and S. D. Teasley, "Bridging the gap from knowledge to action: Putting analytics in the hands of academic advisors," in *Proc. Int. Conf. Learn. Analytics Knowl.*, 2012, pp. 184–187.
[39] L. Corrin and P. de Barba, "Exploring students' interpretation of feedback delivered through learning analytics dashboards," in *Proc. ASCILITE Conf.*, 2014, pp. 629–633.
[40] K. Verbert, et al., "Learning dashboards: An overview and future research opportunities," *Personal Ubiquitous Comput.*, vol. 18, pp. 1499–1514, 2013.
[41] D. Verpoorten, W. Westera, and M. Specht, "A first approach to "Learning Dashboards" in formal learning contexts," in *Proc. 1st Int. Workshop Enhancing Learn. Ambient Displays Vis. Techn.*, 2011.
[42] L. Lockyer, E. Heathcote, and S. Dawson, "Informing pedagogical action: Aligning learning analytics with learning design," *Amer. Behavioral Scientist*, vol. 57, no. 10, pp. 1439–1459, 2013.
[43] R. Bramucci and J. Gaston, "Sherpa: Increasing student success with a recommendation engine," in *Proc. 2nd Int. Conf. Learn. Analytics Knowl.*, 2012, pp. 82–83.
[44] C. Antunes, "Anticipating student's failure as soon as possible," in *Handbook of Educational Data Mining*, C. Romero, S. Ventura, M. Pechenizkiy and R. S. J. d. Baker, Eds. Boca Raton, FL, USA: CRC Press, 2010, Art. no. 353.
[45] C. Romero, M.-I. López, J.-M. Luna, and S. Ventura, "Predicting students' final performance from participation in on-line discussion forums," *Comput. Educ.*, vol. 68, pp. 458–472, 2013.
[46] A. Essa and H. Ayad, "Student success system: Risk analytics and data visualization using ensembles of predictive models," in *Proc. 2nd Int. Conf. Learn. Analytics Knowl.*, 2012, pp. 158–161.
[47] A. Essa and H. Ayad, "Improving student success using predictive models and data visualisations," *Res. Learn. Technol.*, vol. 20, pp. 58–70, 2012.
[48] D. Suthers and K. Vebert, "Learning analytics as a "Middle Space"," in *Proc. 3rd Int. Conf. Learn. Analytics Knowl.*, 2013, pp. 2–5.
[49] A. Pardo, R. A. Ellis, and R. A. Calvo, "Combining observational and experiential data to inform the redesign of learning activities," in *Proc. Int. Conf. Learn. Analytics Knowl.*, 2015, pp. 305–309.
[50] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, London, U.K.: Sage Publications Inc, 2013.
[51] K. Preacher and R. MacCallum, "Repairing Tom Swift's electric factor analysis machine," *Understanding Statistics*, vol. 2, no. 1, pp. 13–43, 2003.
[52] M. Prosser, P. Ramsden, K. Trigwell, and E. Martin, "Dissonance in experience of teaching and its relation to the quality of student learning," *Studies Higher Educ.*, vol. 28, no. 1, pp. 37–48, 2003.

[53] L. N. Lynn, M. Cuskelly, M. J. o'Callaghan, and P. H. Gray, "Self-regulation: A new perspective on learning problems experienced by children born extremely preterm," *Australian J. Educ. Developmental Psychol.*, vol. 11, pp. 1–10, 2011.

[54] B. J. Zimmerman, "Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects," *Amer. Educ. Res. J.*, vol. 45, pp. 166–183, 2008.

[55] J. Cohen, "A power primer," *Psychological Bulletin*, vol. 112, no. 1, pp. 155–159, 1992.

[56] P. Shea and T. Bidjerano, "Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments," *Comput. Educ.*, vol. 55, no. 4, pp. 1721–1731, 2010.

[57] V. A. Romero Zaldívar, A. Pardo, D. Burgos, and C. Delgado Kloos, "Monitoring student progress using virtual appliances: A case study," *Comput. Educ.*, vol. 58, no. 4, pp. 1058–1067, 2012.

[58] S. Joksimović, D. Gasevic, V. Kovanović, B. E. Riecke, and M. Hatala, "Social presence in online discussion as a process predictor of academic performance," *J. Comput. Assisted Learn.*, vol. 31, pp. 638–654, 2015.

[59] T. García Duncan and W. J. McKeachie, "The making of the motivated strategies for learning questionnaire," *Educ. Psychologist*, vol. 40, no. 2, pp. 117–128, 2005.

Abelardo Pardo received the PhD degree in computer science from the University of Colorado, Boulder, in 1997. He is an associate professor in the School of Electrical and Information Engineering, The University of Sydney, and co-director in the Software Engineering Group. He is an author of more than 150 research publications in the area of educational technology. He is a member of the executive board of the Society for Learning Analytics Research, and member of the editorial boards of the *IEEE Transactions on Learning Technology* and the *Journal of Social Media and Interactive Learning Environments*.

Feifei Han received the BA degree from Xi'an International Studies University, the MA, the Master of Education, and the PhD degrees, all from the University of Sydney, in 2003, 2006, 2008, and 2014, respectively. She has worked as a lecturer, research officer, and currently is a research associate for education research at the University of Sydney. Her PhD research has received more than 10 scholarship and awards nationally and internationally. She is also a solo principle investigator on four grants in language and literacy education. As an early career research, she has published one book, 15 book chapters, and 13 journal articles. One of her current research interests focuses on educational technology.

Robert A. Ellis received the BA, MA, and PhD degrees from the University of Sydney, 1994, 1997, and 2003, respectively. He is currently an associate professor in the Deputy Vice-Chancellor Education Portfolio, The University of Sydney. He is the author of three books and more than 85 research publications in the area of Education. He is also coordinating editor of the *Higher Education Journal*, and series editor of *The Practice of Education*. His research interests are in the area of spaces for teaching and learning and the student experience in higher education.