

Finding Traces of Self-Regulated Learning in Activity Streams

Analia Cicchinelli

Know-Center GmbH

Austria

acicchinelli@know-center.at

Eduardo Veas

Know Center GmbH

Austria

eveas@know-center.at

Abelardo Pardo

The University of Sydney

Australia

abelardo.pardo@sydney.edu.au

Viktoria Pammer-Schindler

Know-Center GmbH

Angela Fessl

Know-Center GmbH

Carla Barreiros

Know-Center GmbH

Stefanie Lindstädt

Know-Center GmbH

ABSTRACT

This paper aims to identify self-regulation strategies from students' interactions with the learning management system (LMS). We used learning analytics techniques to identify metacognitive and cognitive strategies in the data. We define three research questions that guide our studies analyzing i) self-assessments of motivation and self regulation strategies using standard methods to draw a baseline, ii) interactions with the LMS to find traces of self regulation in observable indicators, and iii) self regulation behaviours over the course duration. The results show that the observable indicators can better explain self-regulatory behaviour and its influence in performance than preliminary subjective assessments.

CCS CONCEPTS

• **Applied computing** → **Interactive learning environments; Learning management systems; E-learning;** • **Social and professional topics** → *Informal education;*

KEYWORDS

Self regulation, Learning strategies, Blended-learning, Clickstream activity, Learning analytics,

ACM Reference Format:

Analia Cicchinelli, Eduardo Veas, Abelardo Pardo, Viktoria Pammer-Schindler, Angela Fessl, Carla Barreiros, and Stefanie Lindstädt. 2018. Finding Traces of Self-Regulated Learning in Activity Streams. In *LAK'18: International Conference on Learning Analytics and Knowledge, March 7–9, 2018, Sydney, NSW, Australia*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3170358.3170381>

1 INTRODUCTION

In the context of higher education, students find themselves among hundreds of peers, all free to choose and regulate their effort, while at the same time receiving less personalized attention. Participation in computer science first year courses, for example, ranges from 200

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
LAK'18, March 7–9, 2018, Sydney, NSW, Australia

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-6400-3/18/03...\$15.00

<https://doi.org/10.1145/3170358.3170381>

to 900 students. This is a difficult number for teachers and tutors to oversee and follow, much less provide personalized feedback. These reasons call for a blended learning approach, where digital resources complement face-2-face teaching such that students can consume content at their own pace, e.g., [2, 13].

To understand the effectiveness of the blended learning support system and the behavior of learners in the online system, learning analytics techniques have been proposed by many researchers, for example [8, 21]. The Society for Learning Analytics and Research (SoLAR) defined learning analytics as “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” [22]. In other words “learning analytics is about collecting traces that learners leave behind and using those traces to improve learning” [6]. Numerous studies successfully deploy learning analytics to relate academic performance with online activity [5, 14].

Nevertheless, blended learning or any other form of online support do not necessarily solve the problem: students still have less personalized feedback, particularly with regards to the strategies they employ in learning. It has been argued that effective online learning requires students to be self-disciplined and self-regulated [15]. Self regulated learning refers to the strategies used to control and manage learning as well as those used to regulate cognition [18]. These skills are crucial, as learners gain autonomy to choose their career paths. Yet, students with poorly developed self-regulatory skills choose suboptimal learning strategies [23], which makes it increasingly important to assess and scaffold self-regulation [10] particularly at early stages where students start to exercise their autonomy.

Beyond predicting academic performance, we contend that learning analytics can be used to identify traces of metacognitive and cognitive strategies in activity streams. Traces are defined as observable indicators about cognition that students create as they engage in a task [25]. In doing so, learning analytics can inform the design of learning environments and the methods to nudge learners towards using better strategies, improving learning.

This paper proposes an analytic methodology based on simple metrics to identify self-regulation in clickstream behavior, and thence establishes its influence in course performance. To find traces of self regulation in activity streams, we undertook the analysis of a first-year university course on knowledge technologies, consisting of face-2-face lectures, and web-based system with the

course content and quizzes on course content as control phases. We analysed i) students self-reported motivation and self-regulation strategies as a baseline, ii) students interactions with the learning system to extract traces of self-regulation, and iii) students' behavior throughout the course using descriptors of self-regulation strategies. Results lead to identify four types of behavior: inactive, procrastinators, probers and continuously active students that we compared in terms of performance.

2 BACKGROUND

In the absence of support and guidance from an instructor, the ability to regulate one's learning process is a critical skill to achieve personal learning objectives. Unlike in school settings, where time is typically structured around classes and everyone follows a fixed schedule, online learners need to determine when and how to engage with course content of their own accord [11].

This ability is known as Self-regulated learning (SRL) and is defined as the strategies that students use to regulate their cognition as well as the resource management strategies that they use to control their learning [18]. Learners are responsible for constructing their own meanings, goals, and strategies from the information available in the "external" environment as well as information in their own minds (the "internal" environment) [19]. A common conceptualization of these students has emerged as metacognitively, motivationally, and behaviorally active participants in their own learning [24].

Most models of metacognitive control or self-regulating strategies include three general types of strategies: planning, monitoring, and regulating.

Planning: planning involves activities that help the learner prepare the use of cognitive strategies, organize and comprehend better the material. Examples are: setting goals for studying, skimming a text before reading, generating questions before reading a text.

Monitoring: monitoring is in relation with the goals that guide the monitor process. Various monitoring strategies alert the learner to breakdowns in attention or comprehension that can then be repaired using regulation strategies. Monitoring activities include: tracking of attention while reading a text or listening to a lecture, self-testing through the use of questions about the text material, and using test-taking strategies (i.e., monitoring speed and adjusting to time available) in an examination situation.

Regulation strategies: As students monitor their learning and performance against some goal or criterion, this monitoring process suggests the need for regulation processes to bring behavior back in line with the goal. For example, as learners ask themselves questions as they read in order to monitor their comprehension, and then go back and reread a portion of the text, this rereading is a regulatory strategy. Reviewing any aspect of course material reflects a general self-regulatory strategy [18]

Regulatory activities during learning include orientation in order to obtain an overview over the task and resources, planning the course of action, evaluating the learning product and monitoring and controlling all activities [1]. In this way, self-regulated learning involves a complex interplay of cognitive, metacognitive, and motivational regulatory components [3]. Pintrich identified three motivational components: (a) an expectancy component, which

includes students' beliefs about their ability to perform a task, (b) a value component, which includes students' goals and beliefs about the importance and interest of the task, and (c) an affective component includes students' emotional reactions to the task [18].

Summing up, these three components refer to the students' perceptions of the task and beliefs about learning, both aspects relevant to cognitive engagement and classroom performance [18]. Hence, self-regulated learning theories are distinct by their emphasis on how students select, organize or create advantageous learning environments for themselves and how they plan and control the form and amount of their own instruction [24]. Motivation and self-regulation are commonly assessed subjectively, using the questionnaire introduced by Pintrich and De Groot [20].

The quest for the present work is to find observable indicators of self-regulation behavior in logs of online activity using learning analytics techniques. Learning analytics concerns the collection of data that students produce while learning to understand and improve learning and the environments where it occurs [6, 22]. Much of the work in the area seeks to model predictors of student performance from online activities [5], for example to identify students at risk [14].

Recent works seeking to understand the behavior of the learner in online learning environment take place in Massive Open Online Courses (MOOCs). For example Kizilcec defines learner trajectories as longitudinal patterns of engagement with two primary features of the course: video lectures and assessments [12]. To characterize student engagement they define a coding: T: *on track* (did the assessment on time), B: *behind* (turned in the assessment late), A: *auditing* (looking at content, watching a video or doing a quiz but not doing exercises), O: *out* (did not participate in the course at all). They first compute a descriptor of the way in which each student was engaged with course content throughout the course and then apply clustering to find subpopulations in these engagement descriptors. Using these patterns as a lens to more closely analyze learner behavior and backgrounds across the different trajectories.

Coffrin et. al. present an exploratory investigation of students' learning processes in two MOOCs which have different curriculum and assessment designs [4]. They classify student types and visualize patterns of student engagement in a transition diagram. They also divided the students into three mutually exclusive subgroups: auditors (students who watched videos in a particular week, but did not participate in any assessments), Active (students who participated in an assessment in a particular week), Qualified (students who watched a video or participated in an assessment). Along this line, Pérez San Agustín et.al. organize students into *active* and *non-active* depending on their usage of the platform in two periods: (1) before the diagnostic exam, and (2) during remedial courses [17]. They classified students into these two groups by analyzing the number of movements that each student registered on the different MOOCs in each phase.

In the context of blended learning, the work of Jovanovic et al. focused on flipped learning (FL) – a form of blended learning that requires students' active participation in learning activities both before and during face-to-face sessions [10]. The authors examined students' learning strategies with exploratory sequence analysis and agglomerative hierarchical clustering to detect patterns in student behavior that are indicative of the adopted learning strategies.

On the topic of SRL and blended learning, Pardo et al. studied how SRL affects students' online interactions with various learning activities. Their instrument identified positive and negative regulation strategies and clustered students therewith (high, and low). They found that students reporting positive self-regulating strategies interact more frequently with online activities and obtain higher scores [16]. This work was confined to subjective assessment of SRL, whereas our approach identifies SRL components in the data.

We build on the idea of coding activities and grouping students based on their trajectories [4, 10, 12, 17]. Some of these studies are based on MOOCs and MOOCs users [4, 12, 17]. Instead, our study is based on LMS blended learning and the users are college students, closer to the flipped classroom model in [10]. Nevertheless, we went beyond these works; we based our analysis on the SRL theory and we define the coding so as to identify, in the online activities, the phases corresponding to the SRL strategies explained above.

3 RESEARCH QUESTIONS

The theory of self-regulated learning encompasses the concepts that are relevant for our work. Hereby, our research questions are:

RQ1: Have motivated students better self-regulated strategies?

RQ2: Do students who report high self-regulation, evidence more self-regulated behavior on the platform?

RQ3: Do students who exhibit self-regulated behavior achieve better scores in the course?

Accordingly, Study I analyzes relations between motivational beliefs and self-regulation strategies using a standard self-reporting instrument [20]. Study II finds evidence of SR strategies in learning activity, and Study III demonstrates how SRL behavior influences performance in the class.

4 METHODS AND DATA

This section describes the course and system used to collect the data. This information together with the participants is common to the three studies that follow.

4.1 Course description

The studies reported in this paper were deployed in the Introduction to Knowledge Technologies (IKT) course from March to June 2015, a second semester course in the Bachelor programme on “Computer Science” and “Software Development and Business Management”.

Overall, the course was organized in a blended learning format, combining face-to-face class with the IKT LMS. In detail, organizational information includes the course organisation, objectives, important dates as well as the lecturer and tutor team. The content for the eight major topics was: Knowledge Technologies (KT), Semantic Network Ontologies (SNO), Semantic Web (SW), Information Retrieval (IR), Web 2.0 - Web 3.0 (WW), Recommender Systems (RS), Rule Based Systems (RBS) and Artificial Neural Networks (ANN). For each topic, the LMS contains an index with objectives and table of contents, a content page with a detailed script, and the slides used in the face-to-face lecture. The course script has *embedded questions* as quick control for students. Summary quizzes about each topic were available for one week after the face-to-face class. Each quiz contained ten randomly presented questions and

was considered successfully completed only when all questions are correctly answered.

4.2 System description

The LMS is a web based application with HTML pages created using a bespoke version of Sphinx-doc¹. It provides an automatic post-production process translating documents written in the mark-up format ReStructuredText into a self-contained HTML web site. This extensible process was enhanced with macros to embed multiple-choice questions, and with a JavaScript layer to sense various engagement actions with the resulting pages. The actions registered by the browser are transmitted using AJAX to a PostgreSQL² database.

4.3 Data

The studies in this paper use the following three datasets:

Motivational Beliefs and Self-Regulation Strategies (MBSRS) questionnaire: A printed version of the questionnaire on motivational beliefs and self-regulation [20] was used together with consent and demographics questions.

Activity logs: The LMS logs user interaction with course organization pages, content pages, the practical exercises pages, and the interaction with the quizzes.

Performance and control phases: We consider points obtained in quizzes as control measures and points in the final exam as performance measure. There were eight quizzes awarding a maximum of 800 points. The exam awarded a maximum of 58 points.

4.4 Participants

The study was conducted in a course with 392 first year (second semester) students, prior consent and anonymization of data. Out of 392 enrolled students, 170 (140M,40F) completed the MBSRS questionnaire. The number of students in studies that refer to this questionnaire is smaller than 170 after removal of incomplete responses.

5 STUDY I: MOTIVATION AND SELF-REGULATION

This study seeks to answer RQ1: Have motivated students better self-regulated strategies? It lays the foundation for subsequent studies: traces of SRL in activity streams in Section 6, and types of SRL strategies in Section 7.

5.1 Measurement instrument

To answer RQ1 we used responses to the MBSRS questionnaire. Altogether, it consisted of 44 questions covering: self-efficacy (9q), intrinsic value (9q), test anxiety (4q), cognitive strategy use (13q) and self-regulation (9q). We collected 170 answers of which a total of 160 were usable (due to missing answers in different sections). Responses were compiled and compared with the course control and performance phases.

¹ <http://www.sphinx-doc.org/>

² <http://www.postgresql.org/>

Table 1: Correlation table. Self-reported motivation and self-regulation

Variable	A.	B.	C.	D.	E.
Motivational Beliefs					
A. Self-efficacy	–				
B. Intrinsic value	.48 [‡]	–			
C. Test anxiety	–.20 [†]	–.06	–		
Self-regulation strategy					
D. Cog. strat. use	.41 [‡]	.61 [‡]	.04	–	
E. Self-regulation	.20 [*]	.32 [‡]	.17 [*]	.58 [†]	–
Mean	4.75	5.29	3.53	4.94	4.13
SD	.77	.81	1.43	.73	.63
Quiz	–.11	–.02	–.02	–.02	–.10
Test points	.28 [†]	.06	–.20 [*]	.16	.23 [†]
N = 160, p < .05 [*] , p < .01 [†] , p < .001 [‡]					

5.2 Analysis

We grouped responses to the pre-questionnaire as proposed in the original paper: motivational beliefs (intrinsic values, self-efficacy, test anxiety) and self-regulated learning strategies (cognitive strategy use, self regulation) [20]. Reliability of these constructs was validated with Cronbach's alpha (values above .7 indicate good reliability). Scores were averaged for each construct after inverting scores in reversed questions. Thereafter, we analyzed: correlations between motivation and self-regulation dimensions, and correlations with results in the control phase of the course (responses to quizzes; points obtained in the final exam). Finally, regression analysis was performed to assess if the results in control and performance can be explained from the scores in MBSRS.

5.3 Results

5.3.1 Construct reliability.

Motivation. After removing incomplete rows 160 of 170 responses were grouped in the dimensions: self efficacy ($\alpha = .85$), intrinsic value ($\alpha = .84$), test anxiety ($\alpha = .85$).

Self-regulation. After removing incomplete rows 162 out of 170 cases were left. We grouped responses in the dimensions: cognitive strategy use ($\alpha = .75$) and self-regulation ($\alpha = .73$).

These results indicate that the constructs in the MBSRS were found reliable in the case of our data ($\alpha > .7$).

5.3.2 Correlations between motivation and self-regulation. Our data shows the same pattern of correlations as exhibited in the original study, with the exception of test anxiety having different signs (c.f., [20]). Table 1 shows the correlations between motivational beliefs (rows 1–3) and self regulation strategies (rows 4 and 5). A.Self efficacy strongly correlates with D.Cognitive Strategy Use (.41[‡]); it also correlates with E. Self Regulation (.20^{*}). Students that believe in their efficacy, also use cognitive strategies appropriately and are good in self-regulation. B.Intrinsic Value correlates strongly with D.Cognitive Strategy Use (.61[‡]), with A.Self efficacy (.48[‡]) and with E.Self Regulation (.32[‡]).

C.Test anxiety correlates positively with E.Self Regulation (.17^{*}) and negatively with A. Self Efficacy (–.20[†]). Hereby, higher levels

of test anxiety accompany higher self regulation. But, at increasing test anxiety, efficacy is lower.

There is a strong correlation between D.Cognitive Strategy use and E.Self Regulation (.58[†]). This result confirms that higher levels of cognitive strategy use are associated to when and how these strategies are applied in order to achieve the goal (self-regulation).

Summarizing, these correlations indicate that students that believe they have self efficacy also use self regulations strategies. Our findings confirm what is suggested by expectancy-value theory [7], students who value their academic work are more willing to put forth effort and spend more time engaged in their school work. In addition, students who are more interested in and value their school work are less likely to be distracted by irrelevant thoughts.

5.3.3 Correlations with control phases. Table 1 also shows correlations between performance and motivational beliefs or self-regulation. We found a positive correlation between A.Self-efficacy and test points (.28[†]). We also found a positive correlation between E.Self-regulation and test points (.23[†]). These results hint that students who had higher scores in self-efficacy and self-regulation also scored higher in the test. Next, we found a negative correlation between C.test anxiety and test points (–.20^{*}), hinting that people who feel anxious about tests scored worse in the exam. This result confirms the negative correlation between test anxiety and self-efficacy found in the previous section. We found no correlations with the quiz points.

Quizzes are introduced as control for students to monitor their progress. They are not mandatory, but can only be taken at a specified period. Regression analysis did not reveal relationships between dimensions of motivation or self-regulation with quiz points.

Looking at the points obtained in the exam, univariate regression to predict test points from motivational beliefs was not significant ($F(1, 118) = 0.009, p = 0.92$). However, univariate regression to predict test points from self regulation was significant ($F(1, 118) = 6.6, p < .05, R^2 = 0.0449$), but the goodness of fit is not necessarily good ($R^2 > .2$) thus, we don't further explore this result.

While we found correlations between some dimensions of motivation and self-regulation with the points obtained in the final exam, these self-reported metrics cannot *explain* the results in quiz points or the results in the exam. In the next section, we attempt to find a connection through the analysis of behavior in the platform.

6 STUDY II: TRACES OF SELF-REGULATION

This study seeks to answer RQ2: Do students who self-report high self-regulation, evidence more self-regulated behavior on the platform?. Looking at the clickstream data logged during the course, we extract metrics that indicate self-regulated behavior and perform correlation analysis with self-reported scores.

6.1 Measurement instrument

The course required students to interact with an online environment providing organizational resources, content pages for each class, questions embedded in content pages, as well as quizzes and compulsory exercise. Quizzes and exercises were administered as described in section 4.1. The LMS hosting these resources recorded

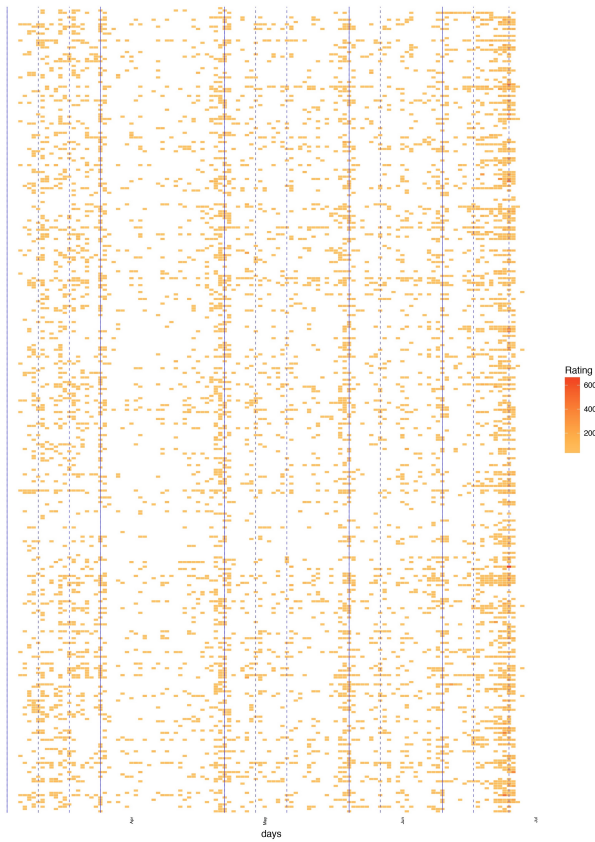


Figure 1: Clickstream graph. A heatmap with a row for each student and a column per day of the course. Color intensity indicates more activity by a student on that day. Dashed vertical lines are class days and solid vertical lines are exercise deadlines.

the interaction of the students while learning. The clickstream activity obtained at the end of the semester is a timestamped series of actions including:

- Access to any resource related to course organization: course dates, groups, exercise deadlines, etc.(summarized under the label V_O).
- Access to indices of content. Each content topic has an index page with objectives and table of contents (summarized under V_I).
- Access to content pages summarized under V_C.
- Access to pages describing the exercises, summarized under V_X.
- Solving questions embedded in content pages. Questions embedded in content pages serve as control, summarized under E_C.
- Solving quizzes. Interactions with each quiz serve as control, summarized under Q_Q.

A total of 136,384 interactions were collected over 116 days. After anonymization and removal of search and download actions, 25,154 V_O, 22,687 V_I, 33,965 V_C, 13,099 V_X, 12,579 E_C and

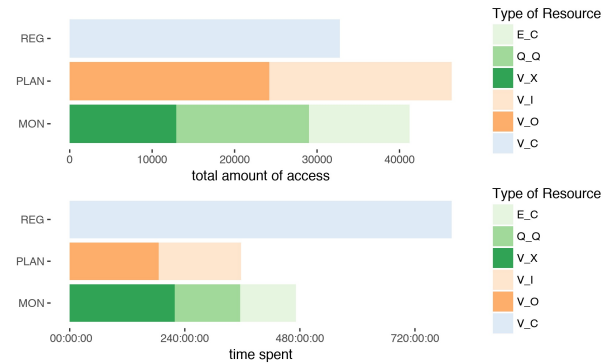


Figure 2: Self-regulation behavior: a high amount of activity concentrated on planning (top), while more time was spent on regulation activities (bottom).

16,262 Q_Q actions were kept for a total of 123,746 actions on the platform.

6.2 Analysis

Figure 1 shows a clickstream chart of the course activity. It pictures a regular participation in the first two classes and more access before exercise deadlines. The period with less participation matches the Easter holiday. Thereafter, participation grows towards the second exercise period and then decreases. A similar pattern can be seen until the fourth exercise. The last dashed vertical line indicates the exam date, where the highest number of accesses occur, the last minute rush before the exam (three days before exam). In general, while some students present a regular participation (e.g., 17 students were active 42 days in average.), the majority of the access take place before the last exam.

The goal is to track aspects that indicate self-regulating behavior in the log of *continuous* activity. Hence, we intend to extract indicators of three types of activity: planning, monitoring and regulating. Firstly, the clickstream activities were labeled according to the type of activity in self-regulatory behavior as:

Planning activities that help organize effort, V_I: viewing content indices, V_O: viewing course organization.

Monitoring activities that test the learner acquired knowledge, V_X: viewing exercises, Q_Q: solving quizzes, E_C: solving questions embedded in content.

Regulating activities taken to acquire or reinforce knowledge, V_C; viewing content.

Figure 2 shows the partition of activities in self-regulation behaviors. The top chart shows the total amount of access and the bottom one shows the time spent on each activity in the platform. Here it is interesting for a tutor to see mainly the time spent and which strategy concentrates the effort, in order to improve the management of these strategies. Just by counting clicks planning and monitoring, both surpass regulation activities. But more time is spent overall in activities related to regulation, as the bottom chart shows.

Secondly, the clickstream activity was partitioned in sessions, assuming that any two subsequent actions from a user that are

separated more than 30m belong to different sessions. With this approach, we computed the indicators shown in table 2 as:

- general indicators: *TC*: total clicks, *TS*: total sessions, *Tt*: total time, *aSD*: avg. session duration, *aCS*: avg. clicks per session,
- self-regulation strategy: *TP*: nr. of planning actions, *TM*: nr. of monitoring actions, *TR*: nr. of regulation actions, *aSP*: avg. planning per session, *aSM*: avg. monitoring per session, *aSR*: avg. regulation per session, *atP*: avg. time planning, *atM*: avg. time monitoring, *atR*: avg. time regulation,
- procrastination *atSm*: is the average time until a student returns to the platform after each class.

We performed correlation analysis with the self-assessment to determine if students with higher scores in self-regulated learning strategies evidence higher self-regulation behavior in the platform. We also performed correlation analysis between the indicators and self-reported motivational beliefs.

6.3 Results

The indicators in table 2 were computed for each student across the duration of the course. To find out if students reporting high self-regulation actually evidence it in their class behavior, we performed correlations between our computed indicators and the questionnaire scores.

Table 2 shows correlations between E-Self-regulation and TC: total clicks, TS: total sessions and Tt: total time. These correlations indicate that students with high self regulation not only access more, but also spend more time in the platform and divide the time in sessions, and so evidence a time management strategy.

Table 2 also shows positive correlations between E.Self-regulation and TP: nr. of planning actions and TM: nr. of monitoring actions. These correlations indicate that students reporting higher self-regulation evidenced more planning and monitoring. There is also a positive correlation between D. Cog. strat. use and TP: nr. of planning actions, which indicates that more planning actions helps select appropriate cognitive strategies in order to regulate the learning process. Self-reported motivational beliefs did not correlate with extracted indicators, except for procrastination. Regarding procrastination, there is a negative correlation with B. Intrinsic value. This correlation indicates that students with high intrinsic value, access the platform earlier, that is they delay less the access or, summarizing, at higher intrinsic value, shorter delays.

7 STUDY III: SELF-REGULATION AND PERFORMANCE

This study seeks to answer RQ3: Do students who exhibit self-regulated behavior achieve better scores in the course? We first analyze correlations between control phases and the indicators of self-regulation behavior from section 6. The indicators describe self-regulation behaviors at a global scale, across the whole course. But, we also intend to find how self-regulation behaviors manifest along the course. We investigate how users engage with the course by coding activities to identify study strategies.

Table 2: Correlation of self-regulation indicators with self-reports.

Indicators	A.	B.	C.	D.	E.
General indicators					
TC	.10	.01	.10	.11	.19*
TS	.05	.10.	.10	.18*	
Tt	.08	-.01	.04	.08	.18*
aSD	.01	-.01	-.04	.07	.16
aCS	.06	.03	.05	.10	.13
Self Regulating Behavior					
TP	.14	.07	.15	.16*	.21 [†]
TM	.07	-.01	.05	.10	.19*
TR	.05	-.03	.07	.04	.13
aSP	.14	.06	.06	.14	.13
aSM	-.03	.01	.01	.06	.08
aSR	.06	.01	.05	.05	.10
atP	-.00	-.08	-.03	.03	.04
atM	-.10	.01	.10	.01	.06
atR	.07	.01	-.08	.08	.06
Procrastination					
atSm	.08	-.18 [‡]	-.10	.04	.06
N=160 for columns A:E, $p < .05^*$, $p < .01^†$, $p < .001^‡$					
A. Self-efficacy, B. Intrinsic value, C. Test anxiety, D. Cog. strat. use, E. Self-regulation					
TC: total clicks, TS: total sessions, Tt: total time, aSD: avg. session duration, aCS: avg. clicks per session, TP: nr. of planning actions, TM: nr. of monitoring actions, TR: nr. of regulation actions, aSP: avg. planning per session, aSM: avg. monitoring per session, aSR: avg. regulation per session, atP: avg. time planning, atM: avg. time monitoring, atR: avg. time regulation, atSm: avg. time before session after milestone					

7.1 Measurement instrument

The course counted a number of different control mechanisms: questions embedded in content pages, optional quizzes available for a week after each class, exercises solved in groups and a final exam. The final note is computed with contributions from all these control phases, thus we will not consider it further. Exercises were assigned to groups and require an analysis of group behavior that goes beyond this paper. The present study concentrates on points obtained in quizzes and points obtained in the final exam. Each course topic, has a quiz consisting of 10 questions. Each correctly answered question awarded students 10, for a total possible of 100 points per quiz, and 800 possible points by the end of the course. The final exam consisted of 11 open-ended questions, awarding a max. of 58 points.

7.2 Analysis

Having extracted traces (indicators) of self-regulation and established their relation with the MBRSR questionnaire, we performed correlation analyses to establish how these descriptors link with the results obtained in class performance.

To find how self-regulation behaviors manifest along the course, beyond global indicators, we resort to activity coding. Considering each class as a milestone in the course and a phase as the time between milestones, a descriptor is created summarizing the behavior for a phase with one label $l_p = x \in \{O, P, M, R, PM, PR, MR, PRM\}$,

where: O =no activity, P=planning activities, M=monitoring activities, R=regulation activities, and their combinations: PM, PR, MR, PMR. Hereby, an *trajectory* is calculated characterizing a student self-regulation behavior as a sequence of discrete states at each phase. One possible trajectory could be: {PM, PM, O, PMR, O, O, PMR, O, PMR, O, PMR, PR}.

Activities are not ordered: performing a planning activity (check KT index) and then a regulatory activity (read the KT page) gains a student a PR label, same as performing first the regulatory activity and then the plan. Repetitions are not counted. These rules follow a similar approach to Kizilcec et al. and San Agustín et al.(c.f., [12, 17]). For each student there is one trajectory describing her/his behavior in the course, for a total of 321 trajectories.

A transition chart illustrates the trajectories of students over time, see Figure 3. The chart shows a large proportion of students inactive, that seem to resort to planning, monitoring and regulation activities at exercise deadlines or before the exam. To analyze behavior patterns and how they impact the course performance we performed exploratory sequence analysis. We used the TraMiner package in R to analyze behavior sequences. Agglomerative hierarchical clustering based on Ward's method was used, as it is deemed suitable to detect student groups from learning activities [10]. Optimal matching was used as distance metric for the clustering algorithm. It is the minimal cost in terms of insertions, deletions and/or substitutions needed to transform a sequence into another [9, 10]. We compared results obtained in quizzes and the final exam by students in each cluster using one-way, between groups ANOVA and Tukey Honest Significant Differences(Tukey HSD) for pair-wise comparisons.

7.3 Results

We report on the correlation analysis of indicators with MBSRS and the cluster analysis of self-regulation behavior.

7.3.1 Correlations with performance. Table 3 shows a positive and consistent correlation between quiz points, test points, general indicators, self regulating strategies and procrastination. For general indicators, quiz points and test points, the table indicate strong correlation with TC: total clicks, TS: total sessions and Tt: total time and moderate correlation with aSD: avg. session duration, aCS: avg. clicks per session. This indicates that higher amount of sessions and time spent interacting with the platform translate in higher points in quizzes and in the final test. For quizzes, the strong correlation with TC indicates a behavior where some students may repeat quizzes until they find the right answer instead of going back to review.

In relation to self regulating strategies, there is a strong correlation with TP: nr. of planning actions, TM: nr. of monitoring actions; moderate correlation between quiz points and aSM: avg. monitoring per session, atM: avg. time monitoring, but small correlation with atR: avg.time regulation and aSR: avg regulation per session . Hereby, the interaction with quizzes could alert the learner to improve the regulation strategies.

With regards to test points there is a meaningful and strong correlation with TP: nr. of planning actions, TM: nr. of monitoring actions and TR: nr. of regulation actions. This means, as expected, that higher planning, monitoring and regulation, translate in more

Table 3: Correlation of self-regulation indicators with control phases.

Indicator	Quiz	Test
General indicators		
TC	.70 [‡]	.66 [‡]
TS	.61 [‡]	.61 [‡]
Tt	.58 [‡]	.65 [‡]
aSD	.27 [‡]	.45 [‡]
aCS	.31 [‡]	.34 [‡]
Self Regulating Behavior		
TP	.69 [‡]	.60 [‡]
TM	.79 [‡]	.58 [‡]
TR	.47 [‡]	.63 [‡]
aSP	.24 [‡]	.18 [†]
aSM	.41 [‡]	.20 [†]
aSR	.11	.39 [‡]
atP	.04	.13 [*]
atM	.31 [‡]	.23 [†]
atR	.20 [‡]	.44 [‡]
Procrastination		
atasm	.01	.39 [‡]
N = 321, p < .05 [*] , p < .01 [†] , p < .001 [‡]		
TC: total clicks, TS: total sessions, Tt: total time, aSD: avg. session duration, aCS: avg. clicks per session, TP: nr. of planning actions, TM: nr. of monitoring actions, TR: nr. of regulation actions, aSP: avg. planning per session, aSM: avg. monitoring per session, aSR: avg regulation per session, atP: avg. time planning, atM: avg. time monitoring, atR: avg.time regulation, atasm: avg. time before session after milestone		

points in the final exam. The next session further investigates behaviors extracted using indicators of self-regulation along the course duration.

7.3.2 Self-regulation behavior and performance. This analysis uses the activity descriptors summarizing behavior for each phase of the course. After agglomerative hierarchical clustering, a solution with four clusters was chosen after inspecting the dendrogram and the different ways of cutting the tree structure. Adding more clusters resulted only in slight variations of one of the already existing clusters. Figure 4 is a density plot of activities in each cluster. Each step in the x axis is one of the phases and positions on the y axis represent the probability of an activity occurring in that phase in that cluster. For example, in Cluster 1 at the 3rd phase a PMR activity occurs with a 10% probability, whereas the same action has about 50% probability in Cluster 2. The clusters illustrates patterns of behavior as follows:

- Cluster 1 inactive: students who did not engage with the course material (N=94).
- Cluster 2 continuously active: students who engaged with course material before each class (N=90).
- Cluster 3 procrastinators: students who become active for deadlines (N=62).
- Cluster 4 probers: students that concentrate on exercises and quizzes repeatedly (N=75).

For each cluster, Figure 5 illustrates the differences in performance at control stages (quiz, exam). Table 4 shows mean and standard deviation for each cluster.

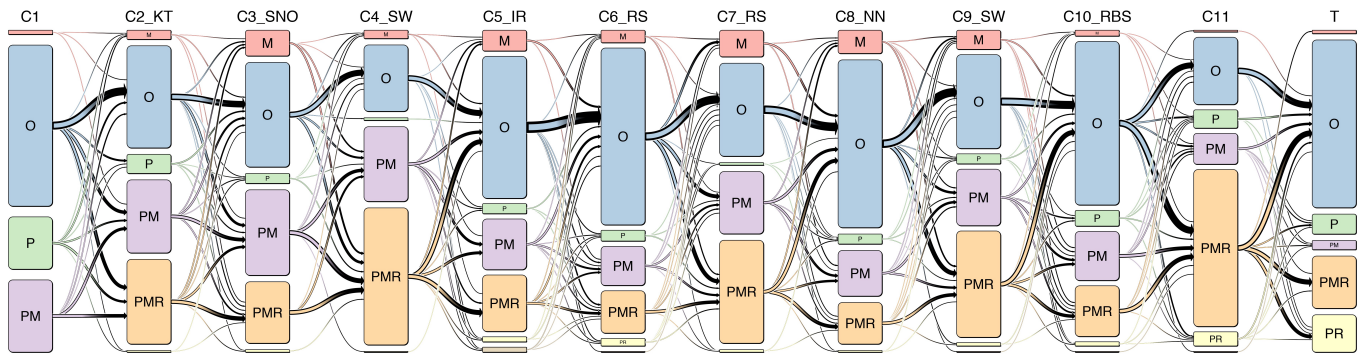


Figure 3: Transition plot of Self Regulating Behavior. Each vertical block represents the activity after a class. A large block of students is out. At deadlines, students turn to more PMR. Some students do more PM, where O=out, P=planning, M=monitoring, and R=regulation activity .

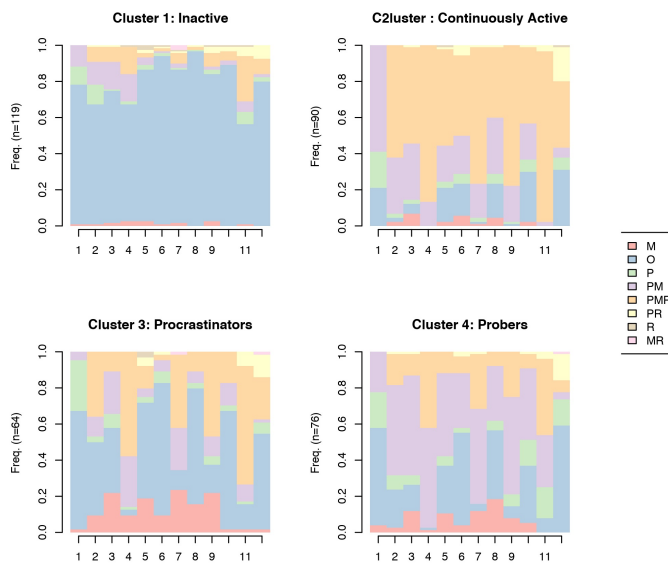


Figure 4: Density plot of SRL activities. Students in Cluster 1 are mostly out. Students in Cluster 2 perform regular PMR activities. Students in Cluster 3 perform those activities only close to exercise deadlines. Students in Cluster 4 do trial and error.

A one-way, between groups ANOVA revealed significant differences in quiz points obtained ($F(3, 310) = 68.35, p < .0001$). Tukey HSD post-hoc comparisons revealed significant differences: students that were *continuously active* obtained ($p < .0001$) more points ($M = 522.7$) than *inactive* ($M = 53.06$), also significantly ($p < .0001$) more points than *procrastinators* ($M = 160.8$) and *probers* ($M = 263.3, p < .0001$). *Procrastinators* obtained significantly better results than *inactive students* ($p < .0001$). *Probers* also obtained significantly more points than inactive students ($p < .0001$) and obtained significantly more points than procrastinators

Table 4: Control phases by group. Mean and standard deviation for control phases by group.

Cluster	C1	C2	C3	C4
Quiz M	53.1	522.7	160.8	263.3
Quiz SD	110.0	271.3	216.3	274.7
Exam M	9.1	32.3	19.9	18.2
Exam SD	13.3	11.3	14.8	14.0

C1: inactive, C2: cont. active, C3 procrastinators, C4 probers.

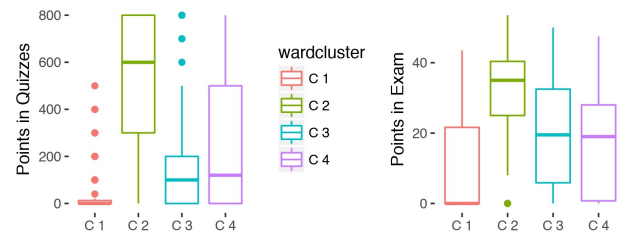


Figure 5: Performance plots. Points by cluster obtained in quizzes (left) and the final exam (right).

($p < .05$). All comparisons were significant. A one-way, between groups, ANOVA test revealed significant differences in test points obtained ($F(3, 317) = 47.12, p < .0001$). Tukey HSD post-hoc comparisons revealed significant differences between groups: students that were continuously active had significantly ($p < .001$) more points ($M = 32.36$) than inactive ($M = 9.14$), also significantly ($p < .0001$) more points than procrastinators ($M=19.91$) and probers ($M = 18.28, p < .0001$). Procrastinators obtained significantly better results than inactive students ($p < .0001$). Probers also obtained significantly more points than inactive students ($p < .0001$). However, there was no statistical difference in the scores between procrastinators and probers.

These results support the claim that students who exhibit self-regulated behavior achieve better scores in the course.

8 DISCUSSION AND OUTLOOK

We proposed three studies to analyze observable indicators and align them with components of self-regulation. Results show that the observable indicators and descriptors proposed here, better explain self-regulatory behavior and its influence on performance than subjective assessments.

In detail, the first study replicated well-established research to outline relations between motivational beliefs and self-regulation strategies [20]. While our results closely align with the original work, we found that this subjective assessment could not explain the performance of students in the course (i.e., regression analysis did not report significant results). This occurs frequently with subjective instruments administered once at the start of a course [16]. Other researchers have also found that motivation, as well as the strategies for self-regulation vary in time, depending of external factors [25]. This result prompts the next two studies seeking evidence of self-regulation in interactions of students with the LMS.

The second study aimed to find traces of self-regulation in activity streams. Firstly, general indicators were utilized following the common practice in learning analytics for clickstreams: counting clicks, sessions, total time in the platform [5]. Secondly, specific indicators were introduced to detect self-regulation activities: planning, monitoring and regulating. We found that these indicators correlate with responses to MBSRS. In general students reporting high self regulation evidenced a time management strategy with more accesses divided in time and sessions. Also, students that scored high in self-regulation evidenced more planning and monitoring activities.

The third study intended to describe self-regulating behavior and its correspondence with control phases and final performance. To do so, we introduced a descriptor of the learner self-regulation activity across the course and used it to analyze clusters of behaviors. We found four kinds of self-regulating behavior: continuously active, probers, procrastinators and inactive. Continuously active students evidenced mature self-regulation strategies and achieved significantly better scores than all others. This study shares with previous research [4, 12, 17] the approach of activity and trajectory coding in online learning environments. But, instead of active / inactive flags, our coding is based on metacognitive and cognitive strategies. Therefore, it allows us to identify SRL strategies.

Learning analytics applied in this way allows us to observe and follow the learning processes of students, which are not visible otherwise. Based on these methods an instructor can observe patterns of behavior and modify practice to foster proper self-regulated learning among students. This type of feedback about strategies employed is essential for the development of students, who are active participants of their own learning.

One of the limitations of our work is that we could not analyze external or contextual factors that also influence decisions about cognitive strategies. For example, offline activities of students who downloaded the study material, and their prior knowledge, both influence learning behavior. Other aspects like the planning and metacognitive phases are difficult to analyze. To facilitate their analysis, the system would need to integrate spaces or tools, whereby students could plan the learning process and reflect on what has been learned.

Future works should consider extending the findings and methodology used in this work to other course settings and learning experiences in general. It would be interesting to investigate also other factors around the learning process, such as goal setting, engagement and enjoyment and how they cause particular SRL strategies.

9 CONCLUSIONS

Online learning platforms offer a great opportunity to investigate the processes that learners employ to construct their knowledge. Taking this into account, this work focuses on analyzing traces of SRL strategies through the use of simple metrics, in order to understand the learning process of students in a blended learning environment and how they influences the course performance.

The main contribution of our studies is in finding observable indicators of the three components of SRL: planning, monitoring and regulating to evidence which combinations of these strategies that lead to higher scores. In addition, clustering behaviors according the descriptors of SRL along the course duration helps illustrate patterns of learning strategies used, and the interplay of the various cognitive processes. These findings can be used pedagogically to show which strategies work best to achieve the goal and also to promote student's awareness of their own learning process in order to transfer it and adapt it to different learning environments.

ACKNOWLEDGMENTS

This work was partially funded by AFEL (H2020-ICT-2015 687916). Know-Center GmbH is funded within the Austrian COMET Program under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency (FFG).

REFERENCES

- [1] Maria Bannert, Peter Reimann, and Christoph Sonnenberg. 2014. Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning* 9, 2 (01 Aug 2014), 161–185. <https://doi.org/10.1007/s11409-013-9107-6>
- [2] Yongmei Bentley, Habte Selassie, and Elizabeth Parkin. 2012. Evaluation of a global blended learning MBA programme. *The International Journal of Management Education* 10, 2 (2012), 75 – 87. <https://doi.org/10.1016/j.ijme.2012.03.001>
- [3] Monique Boekaerts. 1997. Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction* 7, 2 (1997), 161 – 186. [https://doi.org/10.1016/S0959-4752\(96\)00015-1](https://doi.org/10.1016/S0959-4752(96)00015-1)
- [4] Carleton Coffrin, Linda Corrin, Paula de Barba, and Gregor Kennedy. 2014. Visualizing Patterns of Student Engagement and Performance in MOOCs. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge (LAK '14)*. ACM, New York, NY, USA, 83–92. <https://doi.org/10.1145/2567574.2567586>
- [5] R. Conijn, C. Snijders, A. Kleingeld, and U. Matzat. 2017. Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS. *IEEE Transactions on Learning Technologies* 10, 1 (Jan 2017), 17–29. <https://doi.org/10.1109/TLT.2016.2616312>
- [6] Erick Duval. 2012. Learning Analytics and Educational Data Mining. (2012). <https://erikduval.wordpress.com/2012/01/30/learning-analytics-and-educational-data-mining/>
- [7] Jacquelynn S. Eccles. 2009. Expectancies, values, and academic behaviors. In *Achievement and achievement motives*.
- [8] Rebecca Ferguson. 2012. Learning Analytics: Drivers, Developments and Challenges. *Int. J. Technol. Enhanc. Learn.* 4, 5/6 (Jan. 2012), 304–317. <https://doi.org/10.1504/IJTEL.2012.051816>
- [9] Alexis Gabadinho, Gilbert Ritschard, Nicolas MÄijller, and Matthias Studer. 2011. Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software, Articles* 40, 4 (2011), 1–37. <https://doi.org/10.18637/jss.v040.i04>

- [10] Jelena Jovanović, Dragan Gašević, Shane Dawson, Abelardo Pardo, and Negia Mirriahi. 2017. Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education* 33, Supplement C (2017), 74 – 85. <https://doi.org/10.1016/j.iheduc.2017.02.001>
- [11] René F. Kizilcec, Mar Pérez-Sanagustin, and Jorge J. Maldonado. 2017. Self-regulated Learning Strategies Predict Learner Behavior and Goal Attainment in Massive Open Online Courses. *Comput. Educ.* 104, C (Jan. 2017), 18–33.
- [12] René F. Kizilcec, Chris Piech, and Emily Schneider. 2013. Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13)*. ACM, New York, NY, USA, 170–179. <https://doi.org/10.1145/2460296.2460330>
- [13] M. Victoria López-Pérez, M. Carmen Pérez-López, and Lázaro Rodríguez-Ariza. 2011. Blended learning in higher education: Students' perceptions and their relation to outcomes. *Computers & Education* 56, 3 (2011), 818 – 826. <https://doi.org/10.1016/j.compedu.2010.10.023>
- [14] Leah P. Macfadyen and Shane Dawson. 2010. Mining LMS data to develop an early warning system for educators: A proof of concept. *Computers & Education* 54, 2 (2010), 588 – 599. <https://doi.org/10.1016/j.compedu.2009.09.008>
- [15] Abelardo Pardo, Feifei Han, and Robert A. Ellis. 2016. Exploring the Relation Between Self-regulation, Online Activities, and Academic Performance: A Case Study. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK '16)*. ACM, New York, NY, USA, 422–429. <https://doi.org/10.1145/2883851.2883883>
- [16] Abelardo Pardo, Feifei Han, and Robert A. Ellis. 2016. Exploring the Relation Between Self-regulation, Online Activities, and Academic Performance: A Case Study. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK '16)*. ACM, New York, NY, USA, 422–429. <https://doi.org/10.1145/2883851.2883883>
- [17] Mar Pérez-Sanagustin, Josefina Hernández-Correa, Claudio Gelmi, Isabel Hilliger, and María Fernanda Rodríguez. 2016. *Does Taking a MOOC as a Complement for Remedial Courses Have an Effect on My Learning Outcomes? A Pilot Study on Calculus*. Springer International Publishing, Cham, 221–233. https://doi.org/10.1007/978-3-319-45153-4_17
- [18] P. Pintrich. 1999. The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research* 31, 6 (1999), 459–470. [https://doi.org/10.1016/S0883-0355\(99\)00015-4](https://doi.org/10.1016/S0883-0355(99)00015-4)
- [19] Paul R. Pintrich. 2004. A Conceptual Framework for Assessing Motivation and Self-Regulated Learning in College Students. *Educational Psychology Review* 16, 4 (01 Dec 2004), 385–407. <https://doi.org/10.1007/s10648-004-0006-x>
- [20] Paul R. Pintrich and Elisabeth V. De Groot. 1990. Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology* (1990), 33–40.
- [21] George Siemens. 2013. Learning Analytics. *American Behavioral Scientist* 57, 10 (2013), 1380–1400. <https://doi.org/10.1177/0002764213498851>
- [22] George Siemens, Dragan Gasevic, Caroline Haythornthwaite, Shane Dawson, Simon Buckingham Shum, Rebecca Ferguson, Erik Duval, Katrien Verbert, and Ryan S. J. d. Baker. 2011. Open Learning Analytics: an integrated & modularized platform. (2011). <https://solaresearch.org/core/open-learning-analytics-an-integrated-modularized-platform/>
- [23] Philip H Winne and Dianne Jamieson-Noel. 2002. Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology* 27, 4 (2002), 551 – 572. [https://doi.org/10.1016/S0361-476X\(02\)00006-1](https://doi.org/10.1016/S0361-476X(02)00006-1)
- [24] Barry J. Zimmerman. 1990. Self-Regulated Learning and Academic Achievement: An Overview. *Educational Psychologist* 25, 1 (1990), 3–17. https://doi.org/10.1207/s15326985ep2501_2
- [25] Barry J. Zimmerman. 2008. Investigating Self-Regulation and Motivation: Historical Background, Methodological Developments, and Future Prospects. *American Educational Research Journal* 45, 1 (2008), 166–183. <https://doi.org/10.3102/0002831207312909>