How to Quantify Student's Regularity?

Mina Shirvani Boroujeni(✉), Kshitij Sharma, Łukasz Kidziński, Lorenzo Lucignano, and Pierre Dillenbourg

EPFL-CHILI, Lausanne, Switzerland
{mina.shirvaniboroujeni,kshitij.sharma,lukasz.kidzinski, lorenzo.lucignano,pierre.dillenbourg}@epfl.ch

Abstract. Studies carried out in classroom-based learning context, have consistently shown a positive relation between students' conscientiousness and their academic success. We hypothesize that time management and regularity are main constructing blocks of students' conscientiousness in the context of online education. In online education, despite intuitive arguments supporting on-demand courses as more flexible delivery of knowledge, completion rate is higher in the courses with rigid temporal constraints and structure. In this study, we further investigate how students' regularity affects their learning outcome in MOOCs. We propose several measures to quantify students regularity. We validate accuracy of these measures as predictors of students' performance in the course.

Keywords: Regulation · Self-regulation · Time management · Massive open online courses · Procrastination · Engagement

1 Introduction

Massive Online Open Courses allow millions of students from all over the world to participate in top quality courses on-line. Due to a great number of distractions in the environment where MOOCs are usually watched, it is more difficult to grasp learners' attention in a MOOC than in a classroom.

In this paper we present a quantitative framework which simplifies analysis of time-related behaviours. From the full spectrum of variables reflecting consciousness, we focus on regularity of a student. We investigate three key dimensions of regularity: intra-course, intra-week and intra-day as well. The intra-course regularity refers to the repetitive participation in the lectures and responsiveness to course-related events, intra-week corresponds to participation on the same day(s) of the week whereas intra-day corresponds to daily behavioural pattern.

We hypothesize that there are two strategies for participating in MOOCs. First, regular scheduling of learning activities; and second adaptive scheduling of the learning activities based on the daily work or study schedule. The learners affirming to the first strategy will have higher values for our definitions of regularity than the ones following the later strategy. In the current work we investigate if the regularity is a predictive of performance in MOOCs context.

K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 277–291, 2016.
DOI: 10.1007/978-3-319-45153-4_21

Our study is motivated by previous results on engagement. Behaviours inducing a habit are considered as a key to success of many on-line platforms [4]. Similarly, inducing a habit of participation in an on-line course can indicate a success of the course and of the platform. Second, in our previous studies we found that time management is dependent on employment status [20]. Analysis of regularity can allow us to further understand student's employment needs and opportunities. In this context employment can be seen as an external factor as described in a hypothetical model in Fig. 1.

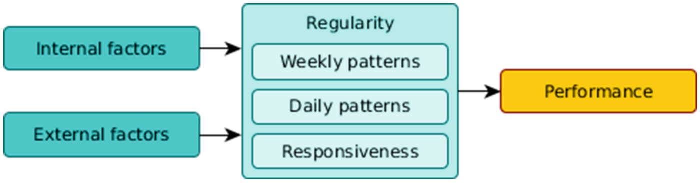


Fig. 1. We analyze regularity as a factor explaining performance, influenced by external and internal variables.

We hypothesize that regularity is one of the key factors related to student's success. In particular, we will answer following research questions:

Question 1. How can we quantify regularity of a student?
Question 2. Is regularity related to performance?

The key contribution of this paper is the definition of different measures of regularity and analysis of their properties. These measures can serve as indicators for quantifying to what extent certain features of a course or platform influences regularity and engagement of participants, or can be used to compare the courses and MOOC platforms regarding their habit inducing properties. Moreover, as we show in Sect. 6.4 the regularity features can be employed to predict users' performance.

2 Related Work

The importance of time management for succeeding in MOOC is highlighted in previous studies [3,15]. Recent studies show that difficulty with keeping up to deadlines is the main obstacle for engaging in a course [8]. In this section, we analyze regularity in the context of consciousness, review measures of regularity which can potentially be used in MOOCs and analyze the link between regularity and performance.

2.1 Conscientiousness and Self-Regulation

Early educational psychologists hypothesized that self-regulation is a key contributor to the academic success of students and it has since been verified [26]. Students' personalities also affect their academic success. The main factor that has been found to be correlated with students' performance is conscientious [16,19,23]. [16] in a review showed that from 33 different studies, examining the relation between the personality factors and academic success (GPA, course grade, average grade, exam score, thesis success), 21 found a significant correlation between conscientiousness and academic success. In two different meta analyses, [23] and [19] showed that the correlation between conscientiousness and academic success is also significant at the university level education.

Procrastination, defined as the tendency to delay of the task completion [11] has also been found to be correlated with the academic success. Klassen and colleagues [9] found that the students with negative procrastination had significantly lower GPA scores. Solomon and Rothblum [22] found that the students who reported higher levels of procrastination attempted significantly lower number of self-paced quizzes. Moreover, Ferrari and Ware [5] found that the task aversiveness was correlated to the self-reported procrastination of the students.

The main feature of both the self-regulative learning strategy and conscientious (in learning context) is organizing and planning learning goals. Time management and regularity are the key constituents for both the aforementioned factors. Thus, we hypothesize that there might exist a correlation between the MOOC performance and students' regularity.

2.2 Time Series Analysis

Time series analysis provides us with technical tools to assess regularity. Our main reference for elementary time series techniques is [2]. We can consider regularity as a seasonal component of a time series and take advantage of tools designed for quantifying seasonality. In classical time series analysis researchers often remove this seasonal pattern and focus on modeling the remaining behaviour of the process. In our case, since the pattern varies between the subjects, it becomes a characteristic of interest discriminating students.

We focus on two key approaches, time domain methods and frequency domain methods. To use time domain methods we slice the time series into segments of the length of interest (e.g. day, week) and compare repeatability of the slices [6,24]. In particular, we use Jensen-Shannon divergence to analyse a histogram of a segmented signal [13]. Frequency domain methods are based on the fact that inner product of a signal with and a periodic function is large if the signal has the same period [21]. Statistical tools have been developed to analyze if the signal on a given frequency is significant [18].

2.3 Performance Prediction

Student's performance is one of the key metrics analyzed in MOOCs. Many studies chose performance as an indicator for showing the value of the categorization methods. Massive datasets allow us to discover relation between performance and even the smallest factors like the number of pauses during watching a MOOC video or ratio of a video replayed [12]. Performance is also a crucial indicator for policy makers and MOOC practitioners. Reports focus on performance of MOOCs as a function of performance of students [14].

In previous studies, measures such as time spent on lecture, homework, forum, quiz and assignments were used to predict students' learning gain [10,25]. Lauria et al. [10] used the amount of content viewed, forum read, number of posts, assignments and quizzes submitted, to predict the performance and the engagement of the students. Other attempts to predict the performance root from the Social Network Analysis of the forum actions of the students. For instance, [17] used the network density, efficiency, individual student's contribution, in- and out- degrees, richness of the content, to find the correlations with engagement and performance. Regarding the analysis of timing patterns, Wolff et al. [25] used the temporal clickstream data to predict students' performance; similarly, Kennedy et al. [7] used number of submissions and active days (submitting days) to predict the final grades of the students in a programming MOOC. Likewise, we focus on the temporal regularity of students' activities, contributing in defining novel measurements for the regularity and showing their link with performance.

3 Methodology

The main steps towards assessing the regularity level of a student are defining what is considered as a regular behaviour and providing methods to capture such behaviour. Regularity in the context of MOOCs can be defined in two domains, actions and time, or a combination of the two. Regularity in actions is evident as repeating patterns in user's actions sequence (e.g. a student who watches the lecture and views the forum before doing an assignment), whereas regularity in time corresponds to repeating patterns in timing of study sessions (e.g. student who studies MOOCs on particular days or times). Regularity in the combined domain on the other hand is reflected by the dependencies between action types and their occurrence time (e.g. student who watches the lecture on Mondays and works on the assignments on Fridays).

In this work, we focus on time regularity. We aim to provide methods for quantifying regularity level of students considering the timing of their activities throughout the course. Regularity in time may emerge in different patterns. We consider six patterns of regularity listed in Table 1 and in Sect. 4 introduce measures to capture these patterns.

Note the difference between P3 and P4 in Table 1, which is the focus on relative (P3) and absolute (P4) amount of participation time on different weekdays. An example for P3 is a student who spends relatively more time on the course on Mondays compared to Tuesdays and Wednesdays, while example of P4 is a student who spends six hours on Mondays, four hours on Tuesdays and two hours on Wednesdays. Therefore P5 and P4 are subsets and more restricted forms of P3.

Table 1. Regularity patterns in time domain

ID	Description
P1	Studying on certain hours of the day
P2	Studying on certain day(s) of the week
P3	Studying on similar weekdays, over weeks of the course
P4	Same distribution of study time among weekdays, over weeks of the course
P5	Particular amount of study time on each weekday, over weeks of the course
P6	Following the schedule of the course

4 Design of Measures

Table 2 provides an overview of our proposed measures and the regularity patterns they reflect. In the following we present problem formulation and detailed description of the measures.

Table 2. Regularity measures and corresponding regularity patterns

Measure	Description	Dimension	Pattern
PDH	Peak on day hour	Intra-day	P1
PWD	Peak on week day	Intra-week	P2
WS1	Weeks similarity measure 1	Intra-week	P3
WS2	Weeks similarity measure 2	Intra-week	P4
WS3	Weeks similarity measure 3	Intra-week	P5
FDH	Periodicity of day hour	Intra-day	P1
FWH	Periodicity of week hour	Intra-week	P1
FWD	Periodicity of week day	Intra-week	P2, P3
DLV	Delay in lecture view	Intra-course	P6

4.1 Problem Formulation

Let n be the number of events by the user and $T = \{t_1, t_2, ..., t_n\}$ be the set of timestamp of events. We assume minutes as a unit of time and set $t = 0$ when the course starts. Let L_m, L_d and L_w be the course length (time from course release till the deadline of the final assignment) in minutes, days and weeks respectively. We can treat user's activity time series as a binary signal defined as (examples in Fig. 4)

$$F_W(x) = \begin{cases} 1 & \text{if } \exists t_i \in T : x = \left\lfloor \frac{t_i}{W} \right\rfloor \\ 0 & \text{otherwise} \end{cases}, \text{ where } x \in \{1, 2, ..., L_m/W, \}$$

where W is the length of a time window in minutes.

Based on this definition, $F_{60}(x) = 1$ implies that user had at least one action at hour x after the course start and $F_{60\times24}(x) = 1$ indicates at least one action at day x of the course.

4.2 Time Based Measures

We define two measures, **PDH** and **PWD**, based on the entropy of the histogram of user's activitiy over time. PDH identifies if user's activities are concentrated around a particular hour of the day and PWD determines if activities are concentrated around a particular day of the week.

We define function $D(h)$ on every hour of a day, and function $W(d)$ on on every day of a week as

$$D(h) = \sum_{i=0}^{L_d-1} F_{60}(24i + h), \text{ where } h \in \{0, 1, ..., 23\}.$$

$$W(d) = \sum_{i=0}^{L_w-1} F_{60\times24}(7i + d), \text{ where } d \in \{0, 1, ..., 6\}.$$

Therefore $D(h)$ corresponds to the number of days in which user was active at hour h of the day, and $W(d)$ represents the number of weeks in which user was active at day d. See examples of these two functions in Fig. 2.

Although resulting histograms are already informative, they still distinguish the time on which regularity appears. In order to define a measure invariant to the time of regularity, we focus on spikes. The popular measure which identifies if given distribution is uniform or has a spike is entropy. Based on its definition, we suggest daily and weekly entropy as

$$E_D = -\sum_{h=0}^{23} \hat{D}(h)\log(\hat{D}(h)), \qquad E_W = -\sum_{d=0}^{6} \hat{W}(d)\log(\hat{W}(d)),$$

where $\hat{D}$ and $\hat{W}$ are normalized histograms.

A small entropy value encodes presence of spikes in the distributions. However, since entropy is computed on the normalized histogram, it does not reflect the magnitude of the spike in the original histogram. To overcome this limitation, we define two regularity measures, PDH and PWD as

$$PDH = (log(24) - E_D)\max_h D(h), \qquad PWD = (log(7) - E_W)\max_d W(d).$$

Therefore PDH is bounded in $[0, log(24).L_d]$ and PWD is bounded in $[0, log(7).L_w]$. A high value of PDH or PWD measure respectively implies a strong spike in $D(h)$ or $W(d)$.

4.3 Profile Similarity

We define three measures **WS1**, **WS2** and **WS3** based on the similarity between weekly profiles of user's activities. WS1 measures if the user works on the same weekdays. WS2 compares the normalized profiles and measures if user has a similar distribution of workload among weekdays, in different weeks of the course. Whereas, WS3 compares the original profiles and reflects if the time spent on each day of the week is similar for different weeks of the course. In the following we describe the construction of weekly profiles and the three similarity functions used to compare them.

We define activity profile of a user during week k as the following vector (examples in Fig. 3).

$$P(k) = [P(1,k), P(2,k), ..., P(7,k)]^T, \text{ where } k \in \{0, 1, ..., L_w\},$$

where $P(d,k)$ represents the number of hours user was active in day d of week k and is defined as

$$P(d,k) = \sum_{i=0}^{23} F_{60}(24(d+7k)+i), \text{ where } d \in \{0,1,...,6\}, \; k \in \{1,2,...,L_w\}.$$

Similartiy Measure 1: Let $Active(k)$ be the set of days in week k, on which the user had some activity. We define the first profile similarity measure as

$$Sim1(P(i), P(j)) = \frac{\|Active(i) \cap Active(j)\|}{max(\|Active(i)\|, \|Active(i)\|)}$$

Therefore for two weeks in which the user is active on exactly same days, this similarity measure returns the maximum value (1).

Similarity Measure 2: The second profile similarity measure compares the normalized profiles ($\hat{P}(k)$) of two weeks based on Jensen-Shannon divergence (JSD) as

$$Sim2(\hat{P}(i), \hat{P}(j)) = 1 - \frac{JSD(\hat{P}(i), \hat{P}(j))}{log(2)}$$

$$\mathcal{JSD}(P_1, P_2, ...P_n) = H\Big(\sum_{i=1}^{n} \pi_i P_i\Big) - \sum_{i=1}^{n} \pi_i H(P_i),$$

where π_i is the selected weight for the probability distributions P_i and $H(P)$ is the entropy for distribution P. We consider uniform weights for all weeks, hence $\pi_i = 1/n$. The value of Sim2 is bounded in $[0,1]$ and high value of this measure reflects similar shapes of activity profiles in the weeks of comparison.

Similartiy Measure 3: In order to capture the similarity in shape and magnitude of weekly profiles, we define the third similarity function, based on χ^2 divergence as

$$Sim3(P(i), P(j)) = 1 - \frac{1}{\|Active(i) \cup Active(j)\|} \sum_{d=1}^{7} \Big(\frac{P(d,i) - P(d,j)}{P(d,i) + P(d,j)}\Big)^2$$

Therefore the highest similarity value (1) is achieved if the two profiles are identical. Finally we define three regularity measures WS1, WS2 and WS3 as the average of pairwise similarity of weekly profiles computed by $Sim1$, $Sim2$ and $Sim3$ respectively.

4.4 Frequency Based Measures

One common approach to detect seasonal components of a signal is to convert the signal ($X(t)$) from its original domain (often time or space) to a representation in the frequency domain ($\mathcal{F}(\theta)$) by applying Fourier transform. Fourier transform of a signal $X(t)$ is defined as

$$\mathcal{F}(\theta) = \sum_{t=-\infty}^{\infty} X(t) e^{(-2\pi i \theta t)}$$

The function $\mathcal{F}(\theta)$ is referred to as spectral density or periodogram, and is used to detect any periodicity in the data, by observing peaks at the frequencies corresponding to these periodicities. For the purpose of detecting weekly or daily regularity, we compute spectral density of user's time signals ($F_{60}(x)$ and $F_{24\times60}$ defined in Sect. 4.1) and in the resulting periodogram, extract values corresponding to daily and weekly periods. We expect a high value for the resulting measures in case there is a daily or hourly repeating pattern in user's activities over time.

We propose three frequency based measures, **FDH**, **FWH** and **FWD** as

$$FDH = \mathcal{F}_h(1/day), \qquad FWH = \mathcal{F}_h(1/week) \qquad FWD = \mathcal{F}_d(1/week)$$

$$\mathcal{F}_h(\theta) = FFT(F_{60}(x)), \qquad \mathcal{F}_d(\theta) = FFT(F_{24\times60}(x))$$

FDH measures the extent to which the hourly pattern of user's activities is repeating over days (e.g. the user is active at 8 h–10 h and 12 h–17 h on every day). FWH identifies if the hourly pattern of activities is repeating over weeks (e.g. in every week, the user is active at 8 h–10 h on Monday, 12 h–17 h on Tuesdays, etc.). FWD captures if the daily pattern of activities is repeating over weeks (e.g. the user is active on Monday and Tuesday in every week).

4.5 Adherence to Course Schedule

Some students watch the lecture right after it is released whereas others postpone watching lectures or submitting assignments. Therefore some users are regular not because of a weekly routine, but they follow the schedule of the course. To capture adherence to the course schedule, we define **DLV** measure as the average delay in viewing video lectures

$$DLV = \frac{1}{m} \sum_{i=1}^{m} (FirstView(i) - Release(i)),$$

where m is the number of video lectures user has watched. We then normalize DLV by the length of the course to get a value in $[0, 1]$.

5 Dataset

Our analysis is based on an undergraduate engineering MOOC offered in Coursera entitled *"Functional Programming Principles in Scala"*. Total duration of the course was 10 weeks and lectures were released on a weekly basis. The final grade was calculated based on six graded assignments and passing grade was 60 out of 100. The initial dataset contained events by a total of 28,002 participants. In the data preparation phase, we removed inactive users, namely those who had less than two weeks with at most four actions of any type (13,102 users). Users who did not submit any assignments were also considered as inactive and hence removed from the dataset (4,644 users). Some participants, never watched a video on the platform, instead they downloaded the lectures and probably watched them offline. Since activity traces for such users is not available, we removed them from the dataset as well (225 users). Therefore, in our analysis we considered all events by remaining 10,031 participants. Their average grade was 55.7 and 51 % scored higher than the passing threshold (60).

6 Results

We computed the proposed regularity measures for participants in the dataset. Table 3 provides an overview of the computed values.

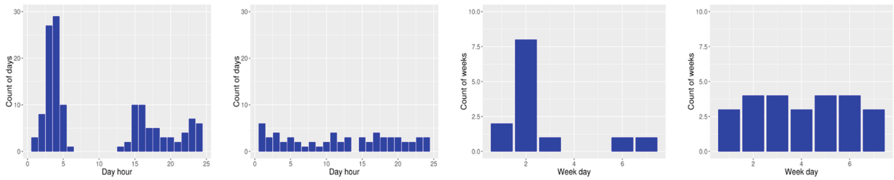
Table 3. Overview of regularity measures in the dataset

Measure	Mean	Max	SD	Measure	Mean	Max	SD
PDH	4.65	49.92	3.65	FDH	0.34	14.65	0.64
PWD	1.12	13.62	1.08	FWH	0.17	4.2	0.25
WS1	0.14	0.90	0.13	FWD	0.36	4.64	0.35
WS2	0.17	0.88	0.15	DLV	0.14	0.95	0.11
WS3	0.11	0.74	0.10				

6.1 Regularity Measures Examples

In the following we present examples of proposed features to verify if they capture the regularity patterns as expected.

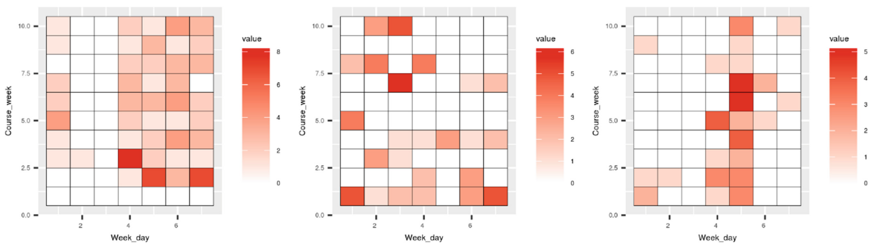
PDH and PWD: Figure 2 illustrates examples of users with high and low value of PDH and PWD measures. Histograms in Fig. 2a and b represent the number of days at which user was active on a particular hour, and Fig. 2c and d show the number of weeks at which user was active on a particular day. Clearly, high value PDH an PWD, represent peak of activity in partciular hour(s) or day(s) and hence they capture regularity patterns P1 and P2 respectively.



(a) PDH=20 (b) PDH=0.7 (c) PWD=6.13 (d) PWD=0.04

Fig. 2. PDH and PWD measures: examples of two users with high and a low values. Clearly a high value reflects a spike in the signal.

WS1, WS2 and WS3: Figure 3 provides examples of weekly activity profiles of three students. In the profile matrix, columns represent weekdays, rows represent week of the course and color intensity encodes amount of study time (hours) on a particular day. As it can be perceived form the profile in Fig. 3a, the activities of first user are clearly concentrated on the second half of the week, whereas no regular pattern is evident in weekly activities of the second user in Fig. 3b. All three profile similarity measures return a high value for the first case (regular) and obtain a low value for the second (not-regular). Figure 3c provides an example highlighting the difference between these three measures. The third user dedicates relatively more time on day five compared to the other days (high value of WS2), but the amount of study hours on this day varies between weeks (relatively lower value of WS3).



(a) 0.68, 0.66, 0.57 (b) 0.17, 0.2, 0.13 (c) 0.5, 0.69, 0.37

Fig. 3. WS1, WS2 and WS3 measures: weekly activity profiles of users with high and low values. Values below each chart correspond to WS1, WS2 and WS3 respectively.

FDH, FWH, FWD: Figure 4 illustrates examples of users with high and low value of FWD measure. As it can be inferred from the time signal (left) in the first row, user's activities follow a periodic weekly pattern which is also reflected by a large value (3.64) at the frequency corresponding to one-week period on the frequency domain chart (right). On the contrary, no seasonal pattern is evident in user's time signal in the second row and consequently FWD obtains a small value (0.04). FDH and FWH measure also follow the same principle.

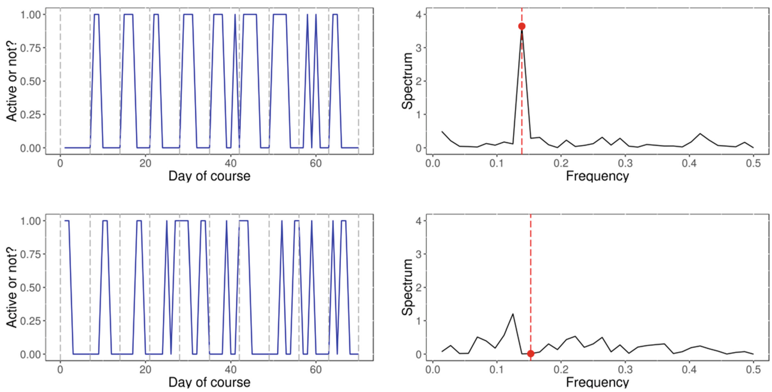


Fig. 4. FWD measure: Examples of activities of two users in time (left) and frequency domain. FWD=3.64 for the first row and FWD=0.04 for the second.

6.2 Correlation Between Measures

The profile similarity measures WS1, WS2 and WS3, although sensitive to different activity profiles (Fig. 3c), result to have strong correlation in pairwise comparison ($r = 0.9$, $p < 0.01$). FWD measure is also moderately correlated with profile similarity measures ($r = 0.57$, $p < 0.001$). The remaining set of measures are not strongly correlated with each other inferring that they capture orthogonal patterns of regularity.

6.3 Clustering Users Based on Regularity Measures

Based on calculated regularity measures, we clustered users into three categories using hierarchical clustering method with euclidean distance metric. Number of clusters was chosen based on the resulting dendogram. Figure 5 presents an overview the three clusters and average grade of users in each group (values were scaled to [0,1] for visualization). The three clusters clearly differ in terms of average grade. Users in the second cluster have the highest regularity according to all measure (except PWD and DLV) and score higher as well. The first and third cluster have very similar regularity values; however users in the third cluster have relatively longer delays in watching video lectures which could explain their lower average grade. Another possible explanation could be that the third cluster contains late-comers in the course who fail to meet the course deadlines. Further investigation of the users activities is required verify these hypothesis.

6.4 Predictive Power of Regularity Measures

In this section we analyze the link between regularity and performance, as presented in Fig. 1. Analysis of correlations between final grade and regularity measures, reveal that final grade is strongly correlated with WS2 ($r = 0.70, p < 0.001$),

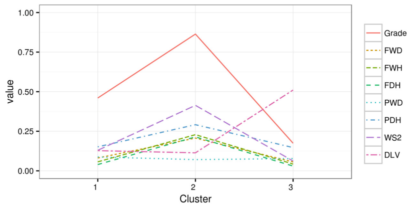


Fig. 5. Average value of regularity measures in each cluster.

FWD ($r = 0.46$, $p < 0.001$), moderately correlated with FWH ($r = 0.37$, $p < 0.001$) and FDH ($r = 0.32, p < 0.001$), slightly correlated with PDH ($r = 0.25$, $p < 0.001$), DLV ($r = -0.25$, $p < 0.001$) and not correlated with PWD measure.

In order to analyze predictive power of the regularity features we build a linear model including all of them and we use penalized regression to improve the model by removing features of low importance. In our dataset, linear model with variables FDH, WS2 and DLV has $R^2 = 0.52$, which assures us about predictive potential of designed variables.

6.5 Other Applications of Regularity Measures

As an example of another application, we investigate the link between regularity and external factors, as presented in Fig. 1. Motivated by our previous results [20], we analyze the employment status. The database contains employment information for about 9.6 % of the participants. Based on these information we extract two categories of users: *full-employed* and *full-students* (559 v.s. 113 users). We assume that users in both categories have a daily or weekly routine imposed by their occupation or school schedule. Considering the time regularity, employed participants have higher regularity in weekly and daily basis. This is reflected by significantly higher value of WS2 measure for employed users ($m = 0.17$ v.s $m = 0.14$, $F[1, 670] = 4.8$, $p = 0.02$), higher value of FWD measure ($m = 0.38$ v.s. $m = 0.3$, $F[1, 670] = 4.2$, $p < 0.05$) and higher values for PDH measure ($m = 0.4.8$ v.s. $m = 3.6$, $F[1, 670] = 9.16$, $p < 0.01$).

7 Conclusions

The key objective of this study was to quantify students' regularity **(Question 1)**. By employing time domain [6,24] and frequency domain [21] techniques, we defined nine measures corresponding to regularity patterns on three dimensions: intra-day, intra-week and intra-course. Investigation of students' activities corresponding to low and high values of these measures illustrates their behaviour. We showed that a subset of the measures are not strongly correlated with each other, providing high predictive power.

We find that regularity is related to performance **(Question 2)**. The predictive power of suggested variables is encouraging for four reasons. First, our proposed measures are general and can be defined outside MOOCs' context. Second, they explain over 50 % of the grade variability, so they can be included in existing performance models. As in previous studies we verify that temporal patterns have significant predictive potential [25]. Third, features are not strongly correlated with each other. Fourth, although our analysis is a posteriori, features which we propose can be estimated throughout the course.

Positive correlation between the defined regularity measures and the performance of the students, supports the hypothesis that students who plan their learning activities in a regular manner have better chances of succeeding in the MOOC [3,15]. There are two plausible explanations for the fact that regularity is predictive of performance in the MOOC. First, regular student follows the structure of the course and therefore attains higher achievement. Second, having high regularity is related to certain factors internal to the students, i.e., motivation, commitment or learning strategies [1,26]. In the future work emerging from this contribution, we will attempt to capture the different factors influencing regularity in the students who have higher values of regularity measures.

Finally, the regularity measures we defined, allowed us to confirm the impact of external factors on regularity patterns [20]. We found that employed learners are more regular both on weekly and daily scales than the unemployed or university students. This application of the measures supports our claim that they can be used in practice to measure effects of interventions on user habits and to compare engagement between courses or platforms.

One limitation of the regularity measure we proposed is that, using our measures one cannot distinguish between the different strategies used by those students who adaptively plan their learning activities. Moreover, as any projections, our measures can only discriminate patterns that they were designed for and should be combined for accurate assessment of regularity. These limitations also enlighten the future work of this contribution.

References

1. Blair, C., Diamond, A.: Biological processes in prevention and intervention: the promotion of self-regulation as a means of preventing school failure. Dev. Psychopathol. **20**(03), 899–911 (2008)
2. Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods. Springer Science & Business Media, New York (2013)
3. Dillenbourg, P., Li, N., Kidziński, Ł.: The complications of the orchestration clock. In: From Books to MOOCs? Emerging Models of Learning and Teaching in Higher Education. Portland Press (2016)
4. Eyal, N.: Hooked: How to Build Habit-Forming Products. Penguin Canada, Toronto (2014)
5. Ferrari, J.R., Ware, C.B.: Academic procrastination: personality. J. Soc. Behav. Pers. **7**(3), 495–502 (1992)
6. Jönsson, P., Eklundh, L.: Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Trans. Geosci. Remote Sens. **40**(8), 1824–1832 (2002)

7. Kennedy, G., Coffrin, C., de Barba, P., Corrin, L.: Predicting success: how learners' prior knowledge, skills and activities predict mooc performance. In: Proceedings of the Fifth International Conference on Learning Analytics and Knowledge, pp. 136–140. ACM (2015)
8. Kizilcec, R.F., Halawa, S.: Attrition and achievement gaps in online learning. In: Proceedings of the Second ACM Conference on Learning@Scale, pp. 57–66. ACM (2015)
9. Klassen, R.M., Krawchuk, L.L., Rajani, S.: Academic procrastination of undergraduates: low self-efficacy to self-regulate predicts higher levels of procrastination. Contemp. Educ. Psychol. **33**(4), 915–931 (2008)
10. Lauría, E.J., Baron, J.D., Devireddy, M., Sundararaju, V., Jayaprakash, S.M.: Mining academic data to improve college student retention: an open source perspective. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, pp. 139–142. ACM (2012)
11. Lay, C.H.: At last, my research article on procrastination. J. Res. Pers. **20**(4), 474–495 (1986)
12. Li, N., Kidziński, Ł., Jermann, P., Dillenbourg, P.: MOOC video interaction patterns: what do they tell us? In: Conole, G., Klobučar, T., Rensing, C., Konert, J., Lavoué, E. (eds.) Design for Teaching and Learning in a Networked World. LNCS, vol. 9307, pp. 197–210. Springer, Heidelberg (2015)
13. Lin, J.: Divergence measures based on the shannon entropy. IEEE Trans. Inf. Theor. **37**(1), 145–151 (1991)
14. McAuley, A., Stewart, B., Siemens, G., Cormier, D.: The MOOC model for digital practice (2010)
15. Nawrot, I., Doucet, A.: Building engagement for MOOC students: introducing support for time management on online learning platforms. In: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, pp. 1077–1082 (2014)
16. OConnor, M.C., Paunonen, S.V.: Big five personality predictors of post-secondary academic performance. Pers. Individ. Differ. **43**(5), 971–990 (2007)
17. Paredes, W.C., Chung, K.S.K.: Modelling learning & performance: a social networks perspective. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, pp. 34–42. ACM (2012)
18. Percival, D.B., Walden, A.T.: Spectral Analysis for Physical Applications. Cambridge University Press, Cambridge (1993)
19. Poropat, A.E.: A meta-analysis of the five-factor model of personality and academic performance. Psychol. Bull. **135**(2), 322 (2009)
20. Boroujeni, M.S., Kidziński, Ł., Dillenbourg, P.: How employment constrains participation in MOOCS? In: Proceedings of the 9th International Conference on Educational Data Mining, pp. 376–377 (2016)
21. Christopher, A.: Sims: seasonality in regression. J. Am. Stat. Assoc. **69**(347), 618–626 (1974)
22. Solomon, L.J., Rothblum, E.D.: Academic procrastination: frequency and cognitive-behavioral correlates. J. Couns. Psychol. **31**(4), 503 (1984)
23. Trapmann, S., Hell, B., Hirn, J.-O.W., Schuler, H.: Meta-analysis of the relationship between the big five and academic success at university. Zeitschrift für Psychologie/J. Psychol. **215**(2), 132–151 (2007)
24. Vetterli, M., Kovačević, J., Goyal, V.K.: Foundations of Signal Processing. Cambridge University Press, Cambridge (2014)

25. Wolff, A., Zdrahal, Z., Nikolov, A., Pantucek, M.: Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In: Proceedings of the Third International Conference on Learning Analytics and Knowledge, pp. 145–149. ACM (2013)
26. Zimmerman, B.J.: Investigating self-regulation and motivation: historical background methodological developments and future prospects. Am. Educ. Res. J. **45**(1), 166–183 (2008)