Students' time allocation and academic performance

AUTHOR 1, AUTHOR 2, AUTHOR 3, AUTHOR 4, and AUTHOR 5

Sociology and computational social sciences have deeply investigated the relation between students' time allocation, i.e., how students spend their time in study activities or working during the university period, and its influence on their academic performance. However, these works rely on approaches that usually consider the aggregation of the time spent of academic activities, thus ignoring the bigger picture of students' everyday behaviour. We analyse how students allocated their time during two academic weeks by leveraging on the combination of time diaries and smartphones, in order to take advantage of both users and sensors as sources of information. Thus, we can obtain a detailed description of how students allocate their time between the different everyday life activities, by taking into account the amount of time spent on them, their duration and their order. We tested this approach on a project whose aim is to understand how different students' daily behaviour may influence their academic performance. Our results show that we can identify different students' daily routines and that, by controlling for gender and field of study, students who are more regular in their time allocation show better performance both in quality and in university career progress.

CCS Concepts: •**Social and professional topics →Student assessment;** •**Human-centered computing →Smartphones;**

Additional Key Words and Phrases: Time allocation; smartphone sensing; data analysis; academic performance; behavioral trends

ACM Reference format:
Author 1, Author 2, Author 3, Author 4, and Author 5. 2018. Students' time allocation and academic performance. *PACM Interact. Mob. Wearable Ubiquitous Technol.* 0, 0, Article 0 (2018), 18 pages.
DOI: 0000001.0000001

1 INTRODUCTION

Empirical evidence has shown how students' time management ability and its consequent translation into time allocation between academic and other daily activities may have an impact on students' performance. In fact, unlike previous education levels like high school, academic activities are usually performed without parental support or teacher supervision leaving some students unprepared to deal with the new challenges proposed by the university system. While this is often the case in different countries, it is especially true in the Italian higher educational system, where the higher number of drop-outs occurs especially in the transition from high school to tertiary education because students may be unable to deal with different tasks and time management skills.

Traditionally, sociologists collect data about time allocation via surveys. There are two main types of surveys: *i)* surveys where individuals are asked to report an estimate of the time spent performing academic activities, e.g. studying or attending lessons, and *ii)* time diaries, i.e., logs where respondents are asked to detail how they allocated their time during a whole day [30]. Nowadays, information about time use can be easily collected through different tools belonging to the ICTs such as smartphones. Given that smartphones are widely used by students, they have been employed both in sociological studies about time use [29] and in the area of computational

DOI: 0000001.0000001

social sciences to discover and analyze students' behaviour, thanks to the collected sensor data. For instance [36] correlate students' behaviour with, among others dimensions of students' life, academic performance.

Currently, there are different types of issues in the literature of sociology and computational science with respect the relation between students' time allocation and students' academic performance. In fact, it has received little interest in sociological research [6, 11] due to the lack of surveys which examine this topic in a systematic way. The majority of students' time use data are obtained through stylized-questions that provide the researchers with aggregated time use information. An aggregate measure of time use usually suffer from a range of biases, i.e., inconsistencies between the answers and the truth, such as the over or under estimation of time spent in some activities [17]. Similarly, computational social sciences rely on estimates of the aggregate time allocated to students' activities by leveraging on smartphone sensors; however, the activities themselves can only be inferred approximately. For instance, in [35], studying is inferred as spending more than 20 minutes in the library. This is also due to the fact that sensor data understand and represent the world of differently from humans [8]. For instance, a person's location can be reduced to (a set of) coordinates, while he or she distinguishes between different types of locations depending on different factors, e.g., functions (house vs workplace).

Our solution is the sequence analysis of students' time allocation based on showing how their different time use strategies significantly affect their academic performance. This analysis relies on two sources of information, i.e., users and sensors, and can extract this information by combining time diaries and smartphones. This allows us to have a system that can leverage on both source of information to exploit their strengths and weaknesses.

The analysis was performed on a project whose main aim is to correlate the behaviour of students in terms of time allocation and their academic performance. Our results show that we can obtain three main clusters of behavioural patterns concerning how students organize their daily routine based on the time diary answers collected via smartphone. We demonstrate how more regularity in time use associated with more time investment in academic activities, i.e., studying and attending lessons, leads students to better overall academic performance.

The paper is organized as follows. Section 2 presents the state of the art in terms of time allocation in sociology and computational social sciences. Section 3 presents our analysis, while Section 4 presents our project. Then, Section 5 illustrates our findings with respect to students' time allocation and academic performance. Finally, Section 6 concludes the paper.

2 RELATED WORK

Time is an invisible element that each person has at his or her disposal and it can assume different economic, social and emotional values. It can be argued that the availability of human time is a limited resource in economy and that how people differently allocate their time to various activities ultimately determines, e.g. the relative prices of goods and services and the distribution of income [16]. Time allocation data serve two main research purposes in socio-economic studies. At a macro level, they have been used in the development and analysis of augmented economic and social systems, e.g., division of household labor [7], while at the micro level, the data have been used to describe and model individual behaviour, e.g., travelling habits [28].

The amount of time that a college student is expected to devote to various activities influences different behavioural patterns. Time management ability, which includes setting goals and priorities, using time wisely and being organized, plays a crucial role in improving students' performance [21]. Consequently, this ability determines effectively students' time allocation. How students allocate their time to different university and extra-university activities, by using their time management skills, is related to their academic achievement [6, 11]. Concerning empirical evidence about students' time use and the relation with academic achievement, several works found that there is a positive correlation between lesson attendance and academic performance [25] and between self-study and academic achievement [4], while [22, 31] analyzed the negative effect of working during university on academic performance. However, there is little evidence in sociological research about time allocation and educational outcomes [6, 11]. The current lack of knowledge between time inputs and students' academic performance is almost certainly a result of the cost and difficulty of collecting appropriate data [32]. There are two main reasons for this:

(1) When examining academic outcomes, e.g., performance or dropping out, they are measured every six months or yearly. Unfortunately, providing accurate measures such as the average number of hours that a person spends for specific activities (e.g., studying) during an academic year or semester is difficult, since researchers interact with students only once a year.
(2) The majority of the researches on this topic usually analyze one or few time variables, especially related to the university sphere and ignore what students do during their free time, thus failing to provide a comprehensive analysis of their time use.

In the sociological literature few studies tried to achieve a more comprehensive view of students' time use. For instance, [19] investigated the significance of learning environment and student's time allocation over study related activities for the acquisition of competencies while [18] used more than one time variable and compared the average amount of time that is spent on different types of students' activities for different countries, finding country specific disparities.

All the cited studies used aggregate time variables for investigate time allocation, usually by asking the amount of time in different activities through stylized-questions within surveys. To the best of our knowledge, only [21] used a time diary approach to explore business and marketing students' time use by suggesting that performance may be a function of a combination of variables that includes not only study time outside class and working but also other time-use variables related to students' free time. To the best of our knowledge, we could not find any sociological work that evaluated not only the impact of the total amount of time spent by students in certain activities but also their order and duration during a specific time window, and the relation with academic performance.

Students are the main focus in computational social sciences due to both their wide adoption of smartphones and their tech-savviness [1]. As an example, a lot of work focuses on extracting behaviours using smartphone data, such as proximity, location, and call logs [5]. These data are combined with surveys, which may be administered via smartphones, for socio-psychological metrics such as personality traits, daily mood, or sleep quality [1]. Among the works in computational social sciences, the Student Life study [35] is the closest one to ours. For this study, smartphones were used to assess the impact of workload on stress, sleep, activity, mood, sociability, mental well-being and academic performance of a class of 48 students (38 males and 10 females) of a computer science class across a 10 week term at Dartmouth College. Moreover, the SmartGPA study [36] used the data from [35] to show that there is evidence of a link between the students' GPA and their behavioral patterns. In these works, regression analyses were used to develop a behavioral slope and behavioral breakpoints. These methods were used to identify changes in a studentfis behavior on a weekly basis. The temporal granularity here and the predictive model does not consider raw data, since the pre-built classifiers which fed into a regression model, such as the accelerometer. The main issue in this area, similarly to sociological studies, is that they can only rely on aggregate time variables data in the prediction of students academic performance that lacks a detailed analysis of students' daily time allocation, since, in the case of [35], they had to use some heuristics to infer human activities such as studying or socializing. Furthermore, relying on classifiers has some limitations [12]. For instance, conversation classifiers, to infer social interactions, may have a hard time distinguishing in-person conversation from conversations occurring on a TV that is around the user. Another example is the proximity classifier using bluetooth devices; in fact, it may lead to a under or over estimation of the number of people around the user. Specifically, it can be difficult to identify how many other people a person is around vs. how many other devices the person is around.

The main issue for sociological tools is that their effectiveness can be impaired by biases and errors in reporting [37]. When reporting using these tools, users may show different response biases causing a lack of congruence between subjects' answers and their true value [33]. Although there are several types of biases, e.g., memory biases or inadequate recalling (i.e., having issues remembering and reporting correctly) [37], conditioning [2] (i.e., reporting socially desirable behaviours) and unwillingness to respond [16] (i.e., failing to report). Moreover, stylized questions methods suffer especially from memory bias, since respondents must not only recall their activities in the recent past, but must also provide an accurate form of averaging [24]. This may lead to overestimation or underestimation of time spent in some activities, together with lack of detail in reporting them [17]. Therefore, sociologists have recently started to explore smartphones in the context of time use surveys. The first pilot study using smartphones as a survey tool [29] developed a diary app where a selected sample of about 150 persons was asked to record their activities for two days, i.e., Wednesday and Saturday, by selecting them from a list of 41 fixed activities from the Harmonized European Time Use Survey (HETUS)[1] from Eurostat; respondents could also record their activities retrospectively in the following day. Smartphone were used to collect the respondents' position via GPS every 10 minutes in addition log-data of the calls and text messages. This pilot study investigates the response rate between paper- and smartphone-based diaries, showing that the number of answers reported using smartphones do not differ substantially from other time use surveys. However, in this work sensors were note exploited, since the main aim was on the efficacy of smartphones for time diaries administration. In fact, collected sensor data, i.e., GPS and communication logs, were only presented visually presented suggesting possible exploitation.

3 METHODOLOGY

Our solution consists in an analysis that allows us to account for how students' activities are distributed throughout the day, including their order and their duration, focusing on activities related to university (study and lesson attendance) and the other activities which characterize the students' life.

We adopt a sequence analysis of time-use, where the aim is to find out patterns of daily behaviours through k-means clustering algorithm, which is a well-known partitional clustering method widely used as an exploratory method within unsupervised learning strategies [13]. These clusters are then analyzed both as dependent variables in order to verify which individual features are most demonstrative of the students' patterns and subsequently as independent variables to estimate the effect on their academic performance.

In order to perform this analysis, we combine two sources of information:

(1) **Users:** We represent the input that users provide when describing their everyday life, i.e., behaviours, surroundings and people. To understand this type of information, we adopt time diaries which represent the most used methods is sociology in understanding time use and adapt them to be administered on smartphones;
(2) **Sensors** We collect the sensor information obtained via ubiquitous devices. To understand this type of information, we use an application that can collect users' data from their smartphones;

From the point of view of handling user information, time diaries are the best candidate as questionnaires because they provide several advantages. Firstly, since respondents have to keep a log of their activities, time diaries allow us to acquire information not only on the average amount of time spent on different activities during a day, but also the duration and frequency of each activity, together with their sequences. Secondly, time diaries provide a systematic tool for also understanding spatial and social relations of users, which enriches and widens our scope of research. Finally, respondents do not need to provide average estimates in time diaries, which lessens the cognitive load while completing them [17], while also reducing the possible mismatch between these answers given and the actual usage of smartphones.

The time diary used in this work was presented in [9], were it was designed after three different dimensions of users' life, i.e., activities, locations and social relations, in accordance with time use surveys methodology [14].

[1]https://www.testh2.scb.se/tus/tus/

The resulting time diary is composed of three sub-questions as shown in Table 1, mirroring the corresponding user dimensions: "What are you doing?" accounts for activities, e.g., "shopping", "Where are you?" accounts for locations, e.g., "home", and "Who is with you?" accounts for social relations, e.g., "relative(s)". The asterisk represents the link between the question "How are you travelling?" and the "En route" activity. When a user selects this option, then, instead of "Where are you?" answers, a list of means of transportation is provided, e.g., "By bike".

What are you doing?	**Where are you?**	**Who is with you?**
Lesson	Class	Alone
Study	Study Hall	Classmate(s)
Eating	Library	Friend(s)
Selfcare	Other University place	Roomate(s)
En route (*)	Canteen	Partner(s)
Social life	Bar/Pub/etc	Colleague(s)
Social media & internet	Relative(s)	Other
Cultural Activity	Home	
Sport	Other Home	**(*) How are you travelling?**
Shopping	Workplace	By Foot
Hobbies	Outdoors	By Bus
Other Free Time	Gym	By Train
Work	Shop	By Car
Housework	Other Place	By Bike
Volunteering		Other
Other		

Table 1. The time diary used in the SmartUnitn project

Smartphones can enhance time diaries by administering them to users, which then are able to answer them in (almost) real time, in addition to performing sensor collection, e.g., GPS, accelerometer, Bluetooth, call logs, and running applications, among others [10]. These two functionalities of smartphones can be exploited to match any given triple of reported activity, location, and social relation with the status of the smartphone as a proxy of the actual user behavior.

In order to manage sensor information, we rely on a mobile application [9, 38] to provide:

- **Data collection:** The mobile application is designed to collect data from multiple sensors simultaneously, both hardware (e.g., GPS, accelerometer, gyroscope, among others) and software (e.g., in/out calls, application running on the device). A dedicated backend infrastructure manages the tasks of synchronizing and storing the streams of data from the smartphones. These streams of data are temporarily stored locally on the device and later synced with a backend infrastructure.
- **Time diaries:** The mobile application can administer the time diary from Table 1 as a question composed of three sub-questions on activities, locations and social relations of users every 30 minutes. Every triple of questions can be answered within 150 minutes from its notification, with a maximum of 5 questions stacked in queue, otherwise it expires and it is treated as null. Questions appear as a silent notifications, shown in Figure 1, in order to avoid bothering students and disrupt their activities too much.

4 STUDY DESIGN

The main goal of the project is to fill the research gap concerning students' time allocation and academic performance by providing a detailed description of how their time management affects their academic achievement. Currently, there is a lack of data about students' time allocation, especially in Italy, which are only available as aggregate data, e.g., [20].

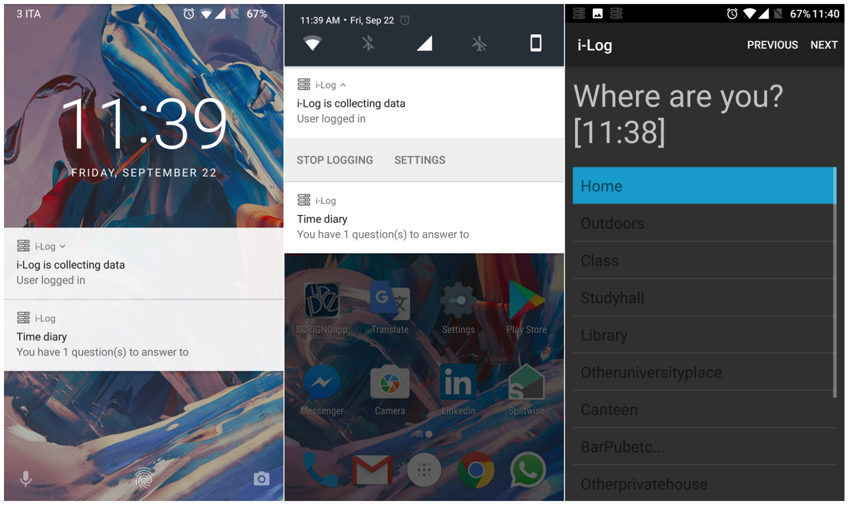

Fig. 1. The mobile application is unobtrusive and does not alter the user experience. It only creates a notification to tell the user that the data collection is running and a second notification when a new question is generated.

In this project, 72 students were selected from the ones enrolled at our university in the academic year 2015-2016 and in particular only those who fulfilled three specific criteria: *i)*, to have filled three university surveys in order to obtain their socio-demographic data, shown in Table 2, and other characteristics, e.g., psychological and time use related; *ii)* to attend lessons during the period of our project in order to describe their daily behavior during the university experience, and *iii)* to have an Android smartphone with an Android version 5.0.2 or higher.

Table 2. Socio-demographics of students from our project

Gender		Departments		Scholarship	
Male	Female	Scientific	Humanities	True	False
61.1%	39.9%	56.9%	43.1%	37.5%	62.5%

The students were asked to attend an introductory presentation where they are presented with the aims of the project and how to use the application. If they wished to participate, after the presentation they signed a consent form, and then installed i-Log on their own smartphones. Users were informed about all aspects of the management of their personal information concerning privacy, from data collection to storage to processing. Furthermore, before starting the data collection, we obtained the approval from the ethical committee of our university.

The project lasted two weeks: during the first one, students were asked to answer a time diary on their smartphone about their time use, while the application was collecting sensor data in the background. During the second week they were only required to have the application running for collecting sensor data.

Students received a fixed money compensation, as an incentive to participate, with additional three final prizes assigned to random users that were considered eligible. Eligibility was based on three parameters: *i)* how much data students' smartphones recorded in via GPS, Bluetooth, and Wi-Fi. We chose these three sensors since they are the only sensors that students could decide to turn off; *ii)* how many questions were answered by students, and *iii)* how long they kept the application running, knowing that they could turn it off at any moment.

It is important to highlight how one whole week of time diaries is considerably more than the usual amount of investigated days in sociological studies, i.e., one or to two days (one weekday and one weekend) [26]. In fact, a larger time provided us with more data to extract patterns from. Furthermore, the amount of participants is in line if not larger than other works in the area of computational social sciences, e.g., almost doubling SmartGPA [36] sample of 48 students.

We collected a total of 110 Gb of data from the 72 students for the whole duration of the project. The resulting dataset is a behavioural dataset that contains both time diaries answers and sensors data, thus exploiting sociological insights from the very beginning. It is also merged both with pre and post project surveys collecting socio-demographic characteristics of students, their time use habits asked through stylized-questions, some psychological traits measured by validated scales (i.e. pure procrastination scale or goal orientation scale) and academic performance data from the administrative office from our university.

Based on this dataset, we obtain the following features to treat as variables for students' behaviour in the analysis:

- **Days of the week:** The reason for considering the days of the week is that time schedule can differ among the different days of the week due, for instance, to students' daily activity-travel patterns [34], their working conditions [3] or academic schedule which are factors that have to be taken into account during the time allocation's analysis.
- **Gender:** The reason for considering gender is that females are more inclined to invest time in academic activities than men because girls and boys are encouraged to prioritize education differently [23]. Furthermore, men and women differ in their social and leisure participation, in childcare time use, in household responsibilities and in time dedicated to sports [15].
- **Faculty:** The reason for considering the faculty attended by students is that humanities faculties (hence, "soft") are structured very differently from the STEM studies (hence, "hard") and, consequently, students may behave differently because of the organization of their study plan.

As for the academic performance, we considered three different measures:

- **Credito Formativo Universitario (CFU):** i.e., credits for each exam taken by a student, whose amount varies depending on the length of a course;
- **Number of exams:** i.e., the total amount of exams successfully taken by a student;
- **GPA:** i.e., the average grade of the student

These measures allow us to capture both the progress of their university career, i.e., CFU and number of exams since they refer to the progress of students' university career, and the qualitative dimension of academic performance, i.e., GPA, since it refers to the quality of students' performance.

In terms of answers, we collected a total number of 27111 answers triples, 9905 were empty because expired, resulting into a final value of 17207 valid answers triples, i.e., 51621 individual answers. A major reason for expired answers is the students were sleeping while they were generated. Furthermore, if we consider that on average people spend 8 hours sleeping, i.e., roughly 33% of a day, this suggests that students answers roughly every available questions in the rest of the day. Table 3 provides a breakdown of all the possible answers' categories divided by their corresponding question, i.e., "What are you doing?" (Table 3a), "Where are you?" (Table 3b), "Who is with you?" (Table 3c), and "How are you travelling" (Table 3d), i.e., the optional location question activated when selecting the "en route" activity.

In the case of activities, we can see that, while studying and attending lessons are common activities as expected for students (12% and 10% respectively), eating (17%) and self care (15%) are the most performed activities. This may be due to the fact that eating could cover also cooking and preparing food in general (which takes more time than actual eating), while self care refers to several activities such as cleaning oneself or indicate sleeping (which will be explained in Section 5.1 as part of our analysis). In the case of locations, home is the most common

Table 3. All answers provided by the students to the time diary questions:

(a) What are you doing?

Answer	Total (%)
Eating	3543 (17,8)
Selfcare	3017 (15.1)
Study	2437 (12.2)
Lesson	2123 (10.6)
Social media & Internet	1957 (9.8)
En route*	1849 (9.3)
Other free time	1679 (8.4)
Social life	1186 (5.9)
Other	419 (2.1)
Housework	379 (1.9)
Work	350 (1.7)
Hobbies	294 (1.4)
Sport	249 (1.2)
Shopping	166 (0.8)
Cultural activity	109 (0.5)
Volunteering	106 (0.5)
Total	19881

(b) Where are you?

Answer	Total (%)
Home	8729 (56.8)
Class	2767 (18.0)
Other private house	1068 (6.9)
Bar/Pub/etc	469 (3.0)
Outdoors	439 (2.8)
Study hall	397 (2.5)
Other place	313 (2.0)
Other university place	305 (1.9)
Workplace	210 (1.3)
Gym	191 (1.2)
Library	165 (1.0)
Shop	162 (1.0)
Canteen	141 (0.9)
Total	15356

(c) Who are you with?

Answer	Total(%)
Alone	6356 (36.9)
Friends	4447 (25.8)
Roommates	1837 (10.6)
Relatives	1579 (9.1)
Partner	1455 (8.4)
Colleagues	1118 (6.4)
Other	413 (2.4)
Total	17205

(d) *How are you travelling?

Answer	Total(%)
By foot	663 (43.1)
By car	529 (34.4)
By bus	278 ()
By train	271 (18.0)
By bike	77 (5.0)
By motorbike	23 (1.4)
Total	1536

location were students spend their time, since they spend there more than half their day (54.8%) and, among the different areas of the university, students spend most of their time in class (18.0%). The smaller amount of time spent in places specifically for studying such as libraries or study halls (2.5% and 1% respectively) may be due to the fact that the project was carried out a couple of months away from finals. In terms of social relations, it seems that students spent more than half of their days (36.9%) alone or with friends (25.8%), which however might also include classmates outside of the university and it may depend on commuters in our sample, since they would have the chance to meet people outside of the university circle. As for the preferred mean of transportation, considering that the university is located in a small to medium sized Italian city, students can easily move around by walking (43%). The fact that car is the second most common answer (34%) may be due to the fact that some students commute daily from neighbouring towns.

5 DATA ANALYSIS

In order to assess the relation between students' time allocation and academic performance, a pre-processing step must be applied to the time diary answers to understand when users where sleeping, described in Section 5.1. Once the sleeping activity was inferred, the overall analysis consists in three main steps:

(1) Applying a clustering method to the answers on activities by students, i.e., k-means, in order to obtain different patterns of time allocation for the whole week, and performing sequence analysis, explained in Section 5.2;
(2) Testing for features such as day(s) of the week, gender and field of study in order to assess their impact on the overall behaviour, explained in Section 5.3.
(3) Moving from days of the week to individual student as unit of analysis, by identifying which modal cluster is associated to each student and then correlate them with their academic performance, explained in Section 5.4

5.1 Inferring students sleeping

There was no way to ask students when they were sleeping, given that they obviously could not answer in that case. Students were told to answer to the last 5 questions with the activity "Selfcare" upon waking up, thus providing an upper bound we could investigate retrospectively.

Since time diaries could not provide us with accurate sleeping information, we decided to leverage on the smartphone sensors to infer the sleeping period. To do so, the following strategies were considered based on the general intuition that students cannot possibly interact with their smartphone while sleeping:

(1) **Screen status:** While students sleep, the screen should always be off. However, depending on the phone settings, whenever a notification is received, the screen lights up, regardless of any actual student interaction. This makes using the screen status not general enough for our sample;
(2) **Smartphone movement:** Another possible strategy is treating the stillness of the smartphone as a proxy for the student sleeping. This can be achieved by considering the accelerometer data. However, this strategy leads to several false positives due to the high sampling rate of 20 values per second. In fact, the accuracy is so high that even the smallest movement can be detected and invalidate the overall inferring process;
(3) **Doze mode:** Android 6 added a new sleeping modality, called Doze Mode[2]. It reduces battery consumption by deferring background CPU and network activity for apps when the device is unused for long periods of time (e.g., unplugged and stationary for a period of time, and with the screen off). Using Doze has the benefit of combining several sensors and it is done automatically. However, there are two main issues with this strategy. Firstly, the system periodically exits Doze for a brief time to let apps complete their deferred activities, which turns on the smartphone. Secondly, this feature is only available from Android 6 and higher, which means that it is limited to only those students having this additional requirement.
(4) **Running applications:** This strategy consists in checking the running applications on the smartphone, treating the sleeping period as the time interval without any application in use. It relies on two temporal windows: *i)* the first hour from when an application is no longer used, since, due to the Android operating system design, any application in the foreground keeps being logged for up to an hour, and *ii)* the actual time when the application is no longer running. Notice the the first threshold resets every time the students uses an application, but considering the empty slot allows a more consistent representation of students' sleeping.

The final strategy used was (4), because it could account for different usage scenarios. For instance, only relying on (2) could have missed the user leaving the smartphone nearby and doing something else before sleeping, e.g., watching a movie. Such a scenario is not uncommon, as a recent survey showed that 77% of smartphone users[3] take up to one hour to actually start sleeping from when they stop using their smartphone. The final output of the running application strategy was to label as "sleeping" 4043 null labels that expired because of students not answering.

5.2 Cluster and sequence analysis

The k-means clustering method was applied on the answers about activities collected during the days of the first week from Monday to Friday of the SmartUnitn One experiment period. Notice that we considered only the days from Monday to Friday, by excluding the weekends, in order to understand how students allocate their time when they have to directly deal with their academic life. The final output of the clustering algorithm produced similar or dissimilar groups in terms of the activities that students performed during those days. When the number of clusters is unknown, as in our case, several k-means solutions with different numbers of groups k (k = 1,...,K) are computed and compared. Figure 2 shows the optimal number of groups k* from the set of K solutions, by observing the kink in the curve generated from the within sum of squares (WSS) or its logarithm log(WSS) for all cluster solutions. In addition, the optimal number of clusters has been also provided by the $\eta 2$ coefficient, which is quite similar to the R2, or the proportional reduction of error (PRE) coefficient.

[2]https://developer.android.com/training/monitoring-device-state/doze-standby.html
[3]http://www.deloitte.co.uk/mobileuk2016/

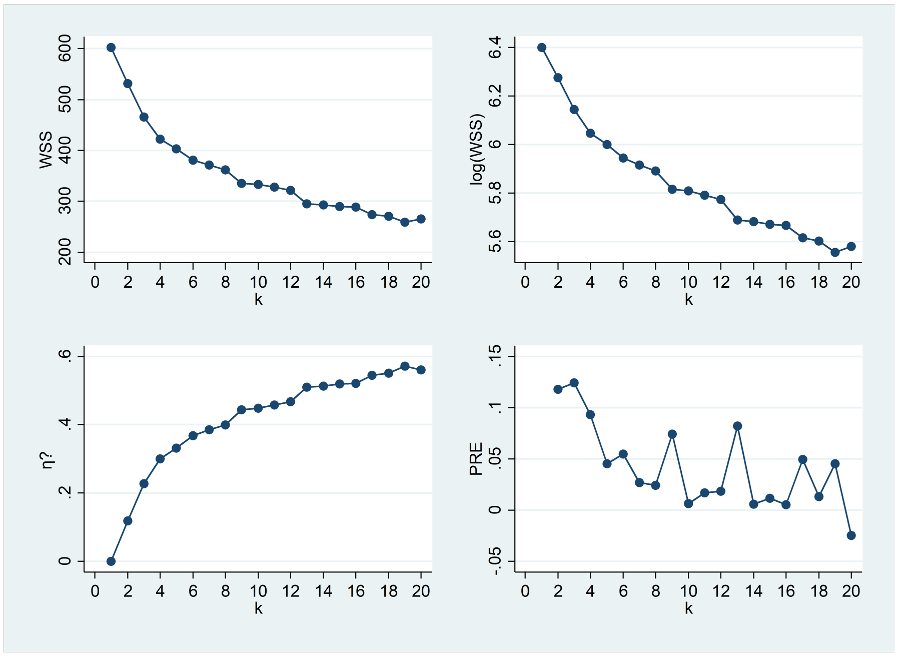

Fig. 2. Optimal number of clusters resulting from the k-mean cluster algorithm

The results indicate clustering with k = 3 to be the optimal solution, which means that the days in which students had to fill the time diary within the project can be divided in three typical groups of days.

The graphical representation of the cluster solution is shown via sequence index plots, which are widely used in social sciences [27]. The three clusters emerged by the previous k-means solution are composed by *n* sequences. Sequence index plots of longitudinal data use line segments (in our case each segment corresponds to an individual student's day during the week) to show how individuals move between a set of conditions or states, in our case represented by students' daily activities, over time (in our case half an hour during the day because of the window of time for answering the time diary). The change of state is represented by changes of color, one per activity type.

The first cluster, shown in Figure 3, is composed by 82 days of our students' sample and it includes those days characterized by an high number of activities related both to the free time (dark yellow) and to other activities (light yellow) not categorized by the time diary survey ("other activity"). It is possible to notice that within these days, the academic activities (i.e. attending lessons and studying, in orange and red), which usually characterize a typical students' day, are rarely presents and they are not structured within specific time slots during the day.

The sleep activity (dark grey) captured by smartphone' sensors is relevant in understanding this first cluster. By checking the last part of the day (after 10PM), there are no days in which the sleep activity is present; this means that students went to bed later in the night. This evidence is confirmed by the sequences where students kept answering to the time diary survey. The most common activities after midnight appear to be those related to free time, social life activities (socializing, going out, etc.) and the use social media. The travelling activity is gathered before and after the academic activities by possibly capturing commuter habits of students. Overall, the first cluster of days does not show any particular regularity in the order or duration or kind of activities carried out by students.

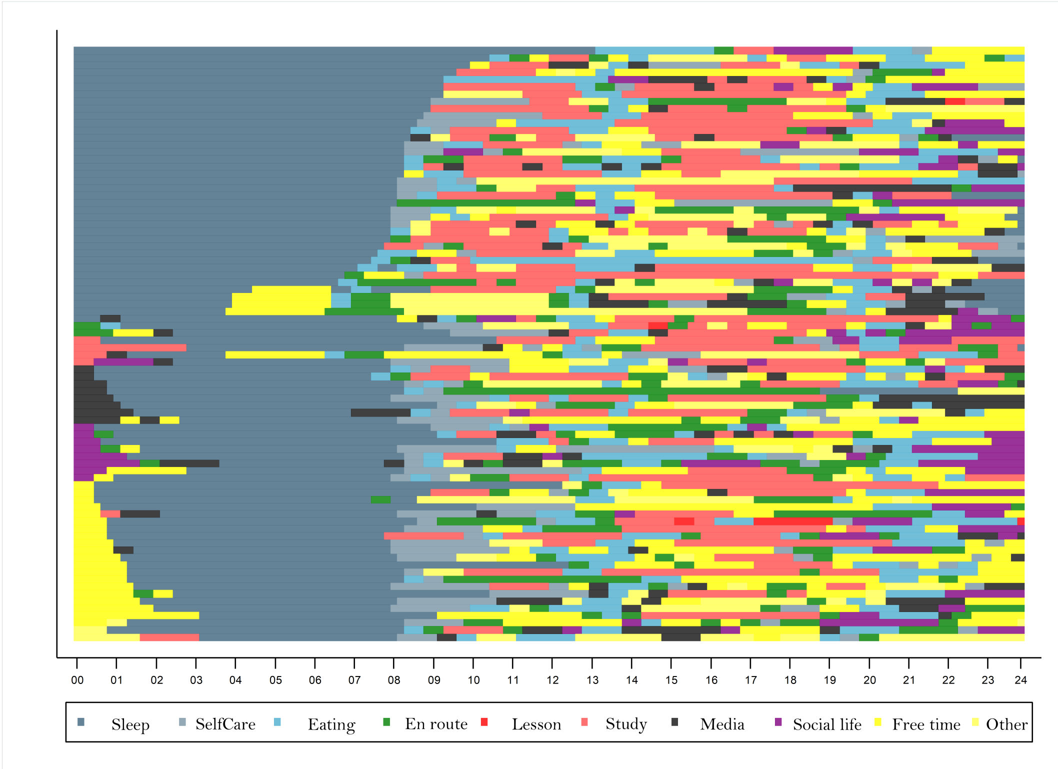

Fig. 3. Sequence index plot of cluster 1.

The second cluster, shown in Figure 4, is composed by 235 days and it is the most recurring. Differently from the first one, it is possible to identify a clear structure that characterizes this group of days. Starting from around 9.00AM until 1.00PM students often indicated they were studying or attending lessons. After a break of about an hour where a stripe of light blue activities, which correspond to eating, can be seen, students resumed their academic activities until 5.00PM or 6.00PM. After the second light-blue stripe, students indicated especially free time activities or activities related to social life (purple). Similarly to the first cluster, the sleep activity suggests that students went to bed after midnight. In half of the days of this plot, students most likely started to sleep around 12.00AM, while in the other half they especially indicated free time activity, social media use activity, social life activity and, in some few cases, study activity until 01.00AM, 02.00AM and even 03.00AM.

The third cluster, shown in Figure 5, covers 87 days and it is possible to see a behavioral pattern similar to the second one in terms of structure. In fact, the two vertical stripes of academic activities during the day are clearly visible, spaced out by the eating activities around 1.00PM. The last part of the day does not a clear pattern as in the second cluster, but all the sequence are characterize by the beginning of the sleep activity around 11.00PM. Therefore, the days in the third cluster are characterized by a more regular sleep cycle in which students went to bed earlier and, consequently, woke up earlier.

Going more into details, the modal sequence index plots in Figure 6, Figure 7 and 8 showshow the modal state of activity within each time point during the day for each cluster. The x axis represents the hours of a day, while the y axis indicates the type of activity. This modality allows us to analyze what the majority of students was

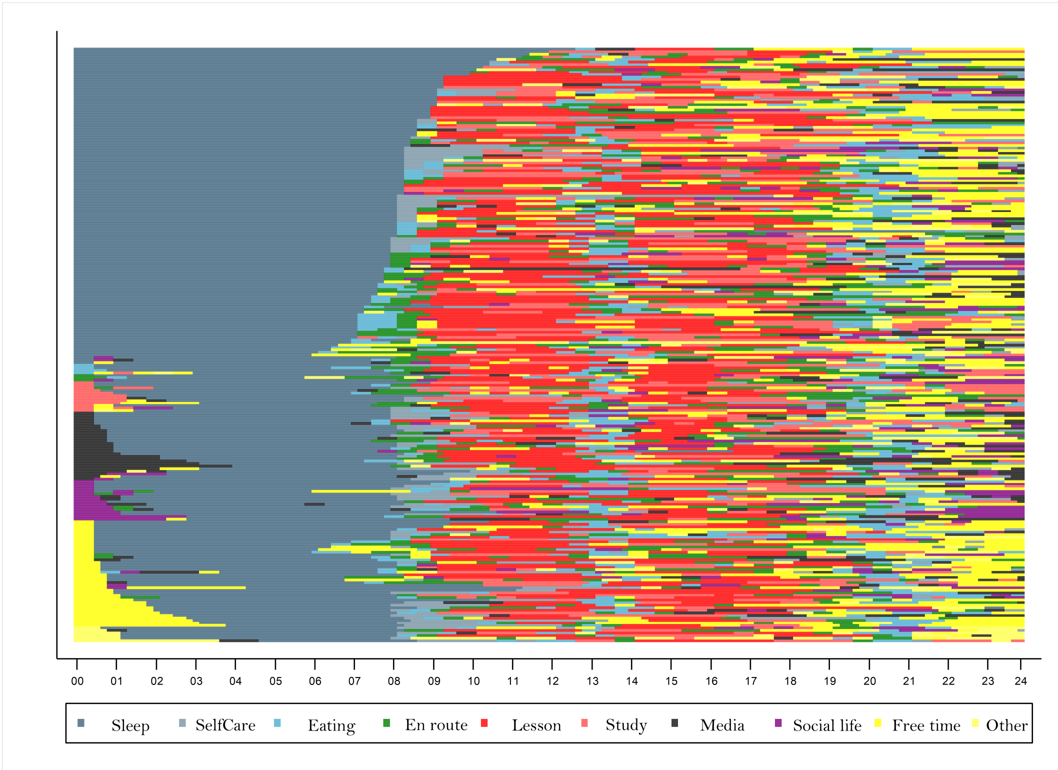

Fig. 4. Sequence index plot of cluster 2.

doing in that specific moment, obtaining a clearer picture of the structure of the days in each cluster. Notice that the light-grey points represent the second most relevant activity conducted by least the 20% of our students in that moment.

The days included in the first cluster, shown in Figure 6, started around 8.00AM, when the majority of students woke up. The lesson activity is not present, while there are two time-slots where the majority of our sample performed study related activities (10.00AM - 12.00PM and 2.30PM -7.00PM). The time for eating approximately lasted two hours both for having lunch and for having dinner. In addition, the other modal activity, which characterized this cluster, is the free time. It occurred especially after and before meals and during the evening. Social life activity seems to be important for at least the 20% of our sample starting from the 10.00PM

Moving to the second cluster, shown in Figure 7, the structure of the day changes. In this case, the attending lesson activity was very frequent both in the mornings and in the afternoons. The lunch break is shorter in comparison to the first cluster (13.00PM-14.00PM) and the free time occurred immediately before the dinner (approximately 7.30PM) and after dinner (starting from 9.00PM). The second more relevant activity during the afternoon is studying. This cluster seems to capture those days dictated by the academic rhythm: wake up, attending lessons for all the days, coming home after 6.00PM, having dinner, and having some free time for themselves.

The last cluster, shown in Figure 8, is similar to the second one but it diverges by some behaviors that clearly emerge. During those days, students woke up a little bit earlier and went to bed around 11.00PM This fact suggests

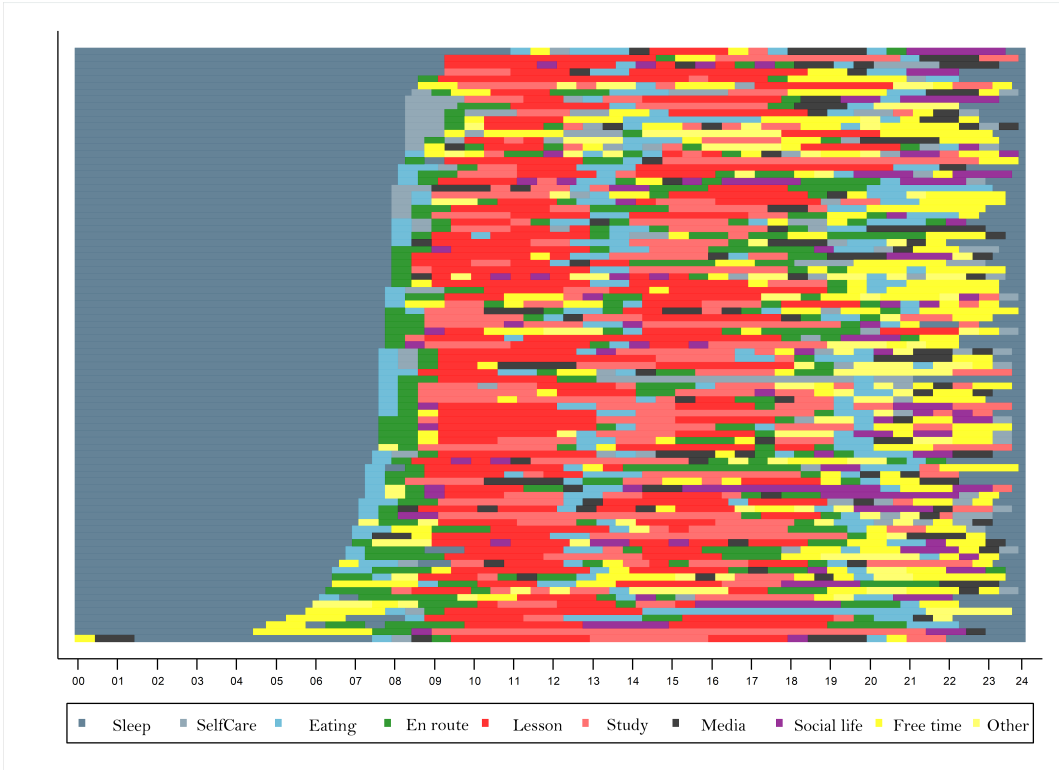

Fig. 5. Sequence index plot of cluster 3.

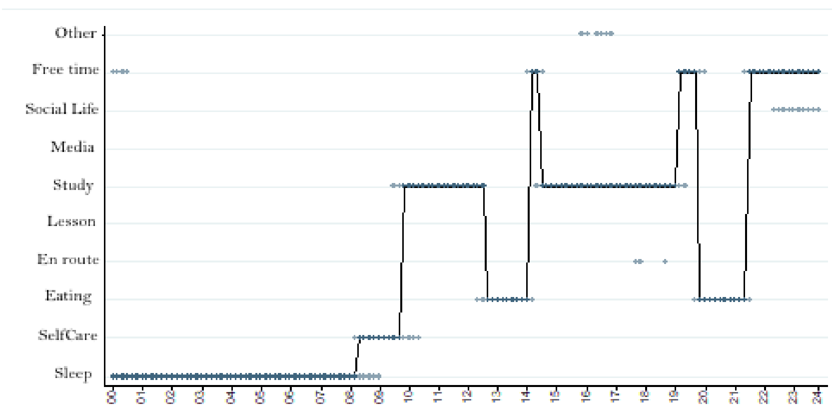

Fig. 6. Modal Sequence index plot of cluster 1.

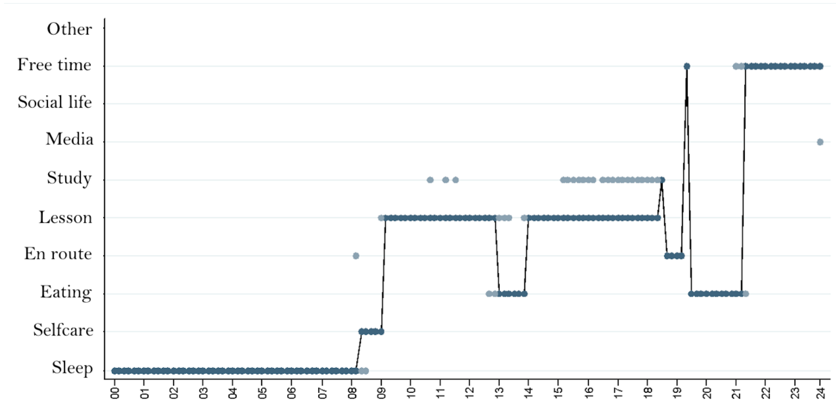

Fig. 7. Modal sequence index plot of cluster 2.

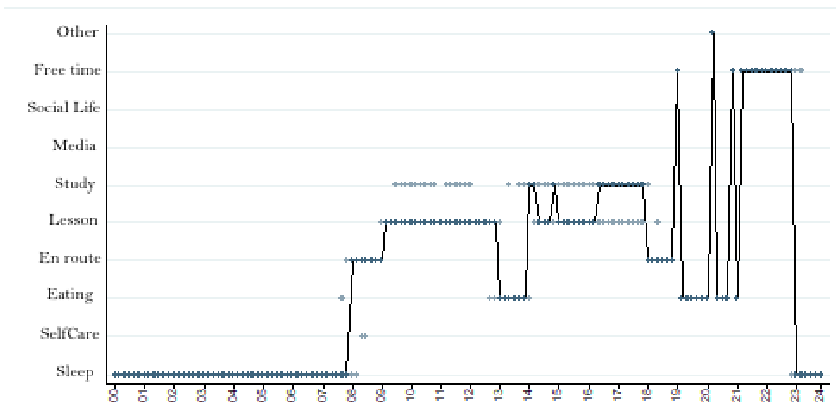

Fig. 8. Modal Sequence index plot of cluster 3.

more regularity in students' life and more time to rest, and it can be likely associated with more concentration during the academic activities of the following day. The time intervals between 8.00AM-9.00am. and between 6.00p.m-7.00PM are those dedicated to travelling. The presence of the "en route" activity within the time-slots suggests a probably high presence of commuter students within our sample. During the day, the most prevalent activity was attending lessons, while the second one was studying.

5.3 Daily pattern and academic performance

Having found the clusters that represent possible behavioural patterns of students, the impact of the variables extracted from our dataset from Section 4, i.e., day of the week, gender and field of study, was evaluated with the Pearson $\chi 2$ test for the independence (whose significance is Pr), as shown in Table 4. In this table, columns show the cluster number, i.e., 1, 2 or 3, and the total number of individual student days considered, while rows provide the chance (in percentage) of how any observed difference between the variables and the cluster appears.

Table 4. Cluster distribution according to the day of the week, department and gender.

	Cluster Number			
Day	**1 (%)**	**2 (%)**	**3 (%)**	**N.**
Monday	31.5	49.3	19.2	73
Tuesday	17.4	65.2	17.4	69
Wednesday	14.8	65.2	22.7	128
Thursday	10.5	61.2	28.4	67
Friday	31.3	49.3	19.4	67
Total	20.3	58.2	21.5	404
$Pearson\chi 2(8) = 19.32, Pr = 0.013$				
Department				
Hard	17.8	58.8	23.4	303
Soft	27.7	56.4	15.8	101
Total	20.3	58.2	21.5	404
$Pearson\chi 2(2) = 5.76, Pr = 0.056$				
Gender				
Female	18.2	52.6	29.2	154
Male	21.6	61.6	16.8	250
Total	20.3	58.2	21.5	404
$Pearson\chi 2(8) = 8.70, Pr = 0.013$				

Cluster 1 is significantly associated with the day of the week (Pr = 0.013) and, in particular, it is more present during Monday (31.5%) and Friday (31.3%). Our university usually schedules less lessons on those days allowing the commuters to coming back home for the weekend. Cluster 2 occurs especially in the middle of the week, on Tuesday (65.2%) and on Wednesday (62.5%). During those days, the majority of lessons take place as well the social events organized by the city for the students' population. This validates the presence of the social life activity as second modal activity during the evenings of those days. The third cluster occurs especially on Thursday (28.4%). A strong difference from the previous cluster is the time when students went to bed. Given the fact that social events are usually organized on Tuesday and on Wednesday, probably this justify the fact that students used to go bed earlier on Thursday after their participation on the night life.

Moving to the association between the clusters and the field of study (Pr = 0.056), soft fields students are more inclined to the first cluster (27.7%) while hard fields students to the third one (23.4%). Lessons schedule validates this evidence; however, no strong difference within our sample emerges for the cluster 2. Additionally, we found a significant association between gender and the way students organize their activities during the day (Pr = 0.013). Women tend to belong more in cluster 3 (29.2%) than men (16.8%), which suggests a stronger inclination towards the academic activities during the week.

5.4 From days to individuals

Until now, our unit of analysis has been the days of the week articulated in different daily activities. In order to identify which cluster belongs to which student, we move to an individual unit of analysis in order to evaluate the impact of different daily time use on their academic performance.

The modal cluster has been assigned to each student and, through an OLS regression model, we evaluate the impact of the students' daily organization on academic performance, by controlling for gender and the field of study. Specifically we used the bootstrap method that is particularly useful for small samples. Indeed, bootstrap can repeatedly draw a sample with replacement (1000 times) in order to produce bias-corrected bootstrap confidence intervals around the coefficients.

Table 5 presents on the columns the academic measures, i.e., CFU, number of exams and GPA, while the rows show clusters, and the gender and faculty variables.The results show that how students organize their daily activities does influence their academic performance. This is particularly relevant for students with modal cluster 3 in comparison of students with modal cluster 1, which is the constant. Indeed, on average cluster 3 reach 18.2 CFU (p value<0.01), 2.5 number of exams (p value<0.01) and 2.5 GPA (p value<0.01), more than cluster 1. There are not statistically significance's difference in terms of performance between students with modal cluster 2 and students with modal cluster 1. The only exception concerns the number of exams: students with modal cluster 2, on average, have 2.1 exams more than students with modal cluster 1. The coefficient for male (b=3.0) is not statistically significantly different at the 0.05 level since the p value is greater than 0.05 while the field of study is associated to the average number of exams: students belonging to soft fields of study on average have -1.8 (p value<0.01) number of exams in comparison of students belonging to the hard fields. This is not necessarily due to the competence of those students but it is probably related to the courses' schedule.

Table 5. OLS regression model for Cfu, N.exams and GPA according to cluster, gender and department

	CFU		**N. Exams**		**GPA**	
Cluster Number	**b**	**s.e.**	**b**	**s.e.**	**b**	**s.e.**
2	13.5	8.49	2.1*	1.14	1.5	1.12
3	18.2**	8.80	2.5**	1.28	2.5**	1.23
Gender						
Male	-3.0	4.84	0.2	0.77	-0.4	0.55
Department						
Soft	-3.7	6.05	-1.8**	0.78	-0.6	0.75
Constant	55.6***	8.70	8.4***	1.23	23.9***	1.16
R-squared	0.11		0.14		0.13	
Adj R-squared	0.05		0.09		0.08	
Bootstrap (1000)						

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

6 CONCLUSIONS

Several works have underlined how the ability of students to manage their time use has relevant consequences on their academic career. In the area of sociology and computational social sciences the relation between time and academic achievements is usually investigated by showing how different prioritization strategies may influence students' academic performance. However, the main trend is to consider an aggregation of the time spent of academic activities, thus ignoring the bigger picture of students' everyday behaviour. We proposed to use sequence analysis that accounts for the time allocation of students and that captures not only the total amount of time spent in some specific activities but also when, for how long, and the order of these activities during all the day. It is based on the combination of time diaries and smartphones, which are tools widely used in sociology and computational social sciences respectively, that can leverage on users and sensors as sources of information.

Our results show that we can discriminate different behaviours of how students organize their daily routine based on the time diary answers collected via smartphone plus the integration of sensor data for the sleep activity. Three daily clusters emerged from the activities that students performed during the project and they are significantly associated to the different days of the week and to students' characteristics such as their department and gender. This suggests different strategies related to time management are due both to contextual and individual factors. Overall, we could provide strong evidence that the more students show regularity in time use, especially with respect to academic activities and to the sleep cycle, the better their performance both in terms of quality of their results and of the progress of their university career.

These results may provide some important implications both for students and administrators. For administrators, it can represent a starting point to assist students in making better decisions in terms of time management through *ad hoc* seminars or tutoring services. Furthermore, it may be necessary for some students to reevaluate how they spend their time especially in terms of academic activities (i.e. studying and attending lessons) because they can have distorted perceptions of available and required time needed to meet course objectives. Future research should take into account that academic performance is not a simple function of study time, as shown in the majority of previous studies, but it a more complex function of how time is allocated between various activities, their order and their sequence during the day. The main limitation of our work is the small sample size and the future direction will consist in iterating our project aiming at a larger sample of students and lasting for a larger window of time.

REFERENCES

[1] Simone Centellegher, Marco De Nadai, Michele Caraviello, Chiara Leonardi, Michele Vescovi, Yusi Ramadian, Nuria Oliver, Fabio Pianesi, Alex Pentland, Fabrizio Antonelli, and others. 2016. The Mobile Territorial Lab: a multilayered and dynamic view on parentsfi daily lives. *EPJ Data Science* 5, 1 (2016), 3.

[2] Louise Corti. 1993. Using diaries in social research. *Social research update* 2, 2 (1993).

[3] Anura De Zoysa and Kathleen M Rudkin. 2007. The effects of employment on academic performance of Australian accounting students. (2007).

[4] Sarah Doumen, Jan Broeckmans, and Chris Masui. 2014. The role of self-study time in freshmenfis achievement. *Educational Psychology* 34, 3 (2014), 385–402.

[5] Nathan Eagle and Alex Sandy Pentland. 2006. Reality mining: sensing complex social systems. *Personal and ubiquitous computing* 10, 4 (2006), 255–268.

[6] Alain Fernex, Laurent Lima, and Erica De Vries. 2015. Exploring time allocation for academic activities by university students in France. *Higher Education* 69, 3 (2015), 399–420.

[7] Makiko Fuwa. 2004. Macro-level gender inequality and the division of household labor in 22 countries. *American Sociological Review* 69, 6 (2004), 751–767.

[8] Fausto Giunchiglia. 1993. Contextual reasoning. *Epistemologia, special issue on I Linguaggi e le Macchine* 16 (1993), 345–364.

[9] Fausto Giunchiglia, Enrico Bignotti, and Mattia Zeni. 2017. Personal context modelling and annotation. In *Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on.* IEEE, 117–122.

[10] Fausto Giunchiglia, Mattia Zeni, Elisa Gobbi, Enrico Bignotti, and Ivano Bison. 2017. Mobile Social Media and Academic Performance. In *International Conference on Social Informatics*. Springer, 3–13.

[11] Barbara S Grave. 2011. The effect of student time allocation on academic achievement. *Education Economics* 19, 3 (2011), 291–310.

[12] Gabriella M Harari, Sandrine R Müller, Min SH Aung, and Peter J Rentfrow. 2017. Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences* 18 (2017), 83–90.

[13] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2002. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Biometrics* (2002).

[14] Mattias Hellgren. 2014. Extracting More Knowledge from Time Diaries? *Social Indicators Research* 119, 3 (2014), 1517–1534.

[15] Sandra L Hofferth and John F Sandberg. 2001. How American children spend their time. *Journal of Marriage and Family* 63, 2 (2001), 295–308.

[16] F Thomas Juster and Frank P Stafford. 1991. The allocation of time: Empirical findings, behavioral models, and problems of measurement. *Journal of Economic literature* 29, 2 (1991), 471–522.

[17] Man Yee Kan and Stephen Pudney. 2008. Measurement error in stylized and diary data on time use. *Sociological Methodology* 38, 1 (2008), 101–132.

[18] Gérard Lassibille. 2011. Student progress in higher education: what we have learned from large-scale studies. *The Open Education Journal* 4 (2011), 1–8.

[19] Christoph Meng and Hans Heijke. 2005. *Student time allocation, the learning environment and the acquisition of competencies.* Technical Report. Maastricht University, Research Centre for Education and the Labour Market (ROA).

[20] Massimo Mucciardi. 2013. Student time allocation and self-rated performance–Evidence from a sample survey in Sicily (Italy). *Electronic International Journal of Time Use Research* (2013).

[21] Sarath A Nonis, Melodie J Philhours, and Gail I Hudson. 2006. Where does the time go? A diary approach to business and marketing studentsfi time use. *Journal of Marketing Education* 28, 2 (2006), 121–134.

[22] Gary R Pike, George D Kuh, and Ryan C Massa-McKinley. 2008. First-year students' employment, engagement, and academic achievement: Untangling the relationship between work and grades. *NASPA journal* 45, 4 (2008), 560–582.

[23] Natasha Yurk Quadlin. 2016. Gender and time use in colleGe: converging or diverging Pathways? *Gender & Society* 30, 2 (2016), 361–385.

[24] John Robinson and Geoffrey Godbey. 2010. *Time for life: The surprising ways Americans use their time.* Penn State Press.

[25] Douglas E Roby. 2004. Research on school attendance and student achievement: A study of Ohio schools. *Educational Research Quarterly* 28, 1 (2004), 3.

[26] MC Romano. 2008. Time use in daily life. A multidisciplinary approach to the Time use's analysis. *Tech. Rep. ISTAT No 35* (2008).

[27] Stefani Scherer. 2001. Early career patterns: A comparison of Great Britain and West Germany. *European sociological review* 17, 2 (2001), 119–144.

[28] Tim Schwanen, Martin Dijst, and Frans M Dieleman. 2002. A microlevel analysis of residential context and travel time. *Environment and Planning A* 34, 8 (2002), 1487–1507.

[29] Nathalie Sonck and Henk Fernee. 2013. Using smartphones in survey research: a multifunctional tool. *Sociaal en Cultureel Planbureau* (2013).

[30] Pitirim Aleksandrovich Sorokin and Clarence Quinn Berger. 1939. *Time-budgets of human behavior.* Vol. 2. Harvard University Press.

[31] Ralph Stinebrickner and Todd R Stinebrickner. 2003. Working during school and academic performance. *Journal of labor Economics* 21, 2 (2003), 473–491.

[32] Ralph Stinebrickner and Todd R Stinebrickner. 2004. Time-use and college outcomes. *Journal of Econometrics* 121, 1 (2004), 243–269.

[33] Roger Tourangeau, Lance J Rips, and Kenneth Rasinski. 2000. *The psychology of survey response.* Cambridge University Press.

[34] Nurul 'Ulyani Mohd Najib, Nor' Aini Yusof, and Nazirah Zainul Abidin. 2011. Student residential satisfaction in research universities. *Journal of Facilities Management* 9, 3 (2011), 200–212.

[35] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* ACM, 3–14.

[36] Rui Wang, Gabriella Harari, Peilin Hao, Xia Zhou, and Andrew T Campbell. 2015. SmartGPA: how smartphones can assess and predict academic performance of college students. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* ACM, 295–306.

[37] Brady T West and Jennifer Sinibaldi. 2013. The quality of paradata: A literature review. *Improving Surveys with Paradata* (2013), 339–359.

[38] Mattia Zeni, Ilya Zaihrayeu, and Fausto Giunchiglia. 2014. Multi-device activity logging. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication.* ACM, 299–302.