



# Utilization Measures in a Learning Management System

Floriana Meluso, Paolo Avogadro<sup>(✉)</sup>, Silvia Calegari, and Matteo Dominoni

Università degli Studi di Milano-Bicocca, 22104 Milan, Italy  
paolo.avogadro@unimib.it, calegari@disco.unimib.it

**Abstract.** Learning Management Systems (LMSs) are becoming more and more popular and incorporate many different functionalities. For this reason, an evaluation of the quantitative utilization of all the parts of a LMS is essential. In this research we propose indicators and techniques which allow to understand in detail how a functionality is accessed by the users. These analytic tools are useful in particular for the administrators of the LMS which are in charge of allocating resources according to the workload and importance of the functionalities. We tested the proposed indicators with the data obtained from the LMS of Università degli Studi di Milano-Bicocca (Milan, Italy) about the messaging functionality. Although the students' messages can potentially be a source of big data, in the present case it is observed that the utilization is limited. With this analysis it has been possible to notice a similarity between the utilization of the message system and the empirical Zipf law. We also introduced the description of the structure of a dashboard which allows to access to the indicators and goes towards the definition of a global tool for students, teachers and administrators.

**Keywords:** Learning analytics · LMS · Indicator · Message system  
Zipf

## 1 Introduction

The term learning analytics (LA) includes many research fields, such as process mining, business intelligence, data processing, information retrieval, technology-enhanced learning, educational data mining and data visualization [13, 14]. Indicators and visualization tools are commonly employed to understand, control and predict [11] the processes related to the learning activities for institutions at different academic levels ranging from primary schools, to universities and including all the learning cases, for example workplaces, etc. LA are expected to provide insights for students allowing them to take full control of their own learning, to give them a better idea of their current performance in real-time and to help them to make informed decisions about the study path [15, 16].

LA are also very useful for teachers to determine, for example expert users, at-risk students, etc. [3, 17]. The learning process is not static and changes with

time, for this reason it requires a constant monitoring, evaluation, and adaption to the requests and needs of the stakeholders to guarantee high quality and ad-hoc outcomes [8]. Among the most important paradigms gathering popularity in education community there is the Flipped Learning (FL) [6], which is considered as an extension of the flipped classroom paradigm where a key role is assumed by the social features within the learning practice. In this context LA becomes fundamental to help and control the process of learning, since this view extends learning beyond the formal boundaries of the classroom and provides a virtual learning environment always available (i.e. anywhere and anytime) for consultation and knowledge sharing with a strong impact on understanding the social dynamics among peers.

The groups of learners within a LMS could form one or more social networks. Once these networks have been correctly identified, it is possible to study their structure with social network analysis (SNA) techniques which allow to uncover non trivial structures [12] and important features.

For the aforementioned reasons the evolution of learning [4] is going toward the definition of a social learning management system (Social LMS) which allows to provide a “complete learning environment” that takes into account the social elements (e.g. collaborating, networking and information sharing capabilities) to improve the practices of learning, for this reason, within these platforms, the social aspects become central for all the activities.

The advantages which can stem from the utilization of a Social LMS are due to the fact that it is possible to provide an easy and uniform academic experience with the help of peers. At present the social features implemented in the majority of the LMSs are very limited and for this reason none of them can be considered as a fully Social LMS; however most of them include messaging systems which can be considered as a embryonic social feature. This research is focused on analyzing the knowledge sharing modules of LMSs in order to gain insights about the communication within the platform among peers. For this reason, we decided to study the utilization of the messaging system of the LMS presently in use at the Università degli Studi di Milano-Bicocca, Italy. This LMS is an instantiation of Moodle, version 3.1. Although LA is mainly directed towards the students, it is also aimed at teachers in order to improve their general vision on how learners are studying, the success of their learning practices, etc. [5, 10].

On the other hand the perspective of this research is (mainly) that of the administrators where the goal is to monitor the several LMS’s functionalities for managing the governance of academic institutions in order to check how learners and academic staff interact with the LMS. The tools here provided could improve the decisional process of the policies (regarding both hardware and software) this in turn should provide to the users a LMS of good quality. In our vision, good quality refers to providing efficient and effective e-learning services with good performances (e.g. high speed) of usability.

The present paper is an extended version of a work presented at the DATA 2017 international conference [1], where the explanation of the results has been

extended, new analysis regarding the received messages have been included and a new paragraph containing the description of the architecture of a dashboard for accessing the indicators has been added.

The methodology followed in this research is the following: we first defined quantities of interest based on the present literature on the subject. Following these needs, a mathematical implementation has been proposed. The formulae have been confronted with data, and the resulting patterns have been modeled with parametric functions in order to summarize the most interesting features. At the end, we introduced new visualization and fruition tools in the form of dashboards. The two utilization indicators here defined are: *specific utilization* and *popularity*, respectively.

The idea behind them is to analyse how widespread the usage a specific functionality of a LMS is. In detail, *specific utilization* is an indicator aimed at verifying how many users accessed to the LMS activities in respect to all the possible users, whereas *popularity* is an indicator aimed at analysing the real usage of the functionality referred to the e-learning community which accessed it.

For a better understanding of this last functionality we defined the *real utilization plot*, which helps in visualizing the distribution of the utilization among the users. The *real utilization plot* allowed to notice a similarity between the observed trends and power laws. For this reason, we fitted the data with analytic functions and compared the parameters thus obtained with the Zipf law (which presents resemblances with our experimental curves).

The paper is organized as follows. A review of the present status of the indicators for LMSs is presented in Sect. 2. Section 3 defines the indicators aimed at analysing the utilization of generic LMS activities, with the objective to tune the policies of governance for better managing the e-learning platform. Section 4 presents a case study where the indicators have been applied to analyse the message activity for the Moodle platform used at the Università degli Studi di Milano-Bicocca, Italy. Section 5 presents the dashboard that was developed to allow an optimal access to the indicators. Finally, in Sect. 6 the conclusions are stated.

## 2 Related Work

Many research projects have been addressing the social activity of the students and teachers on a LMS, this section is going to be devoted to provide a brief report of those works and to highlight the differences and similarities with the present paper.

XRay [11] is an important suite for learning analytics based on Moodle. This suite includes many statistical tools to control and make predictions about the behavior of the students. Our research is focused on the point of view of the administrators rather than the students as in XRay. The particular indicators that we introduce, are not provided with the actual version of XRay.

The main goal of the paper presented in [9] is related to the problem of making predictions about the possible success of students in five classes of an

online course (totaling 26 students) by collecting information of the underlying LMS. In order to better understand the dynamics within the students it is built a sociogram. A logistic regression is performed on top of the sociogram to provide a prediction of the success of the learners. In this respect the study is also aimed at helping the teachers, but at variance with our work it is focused on the single student rather than providing a global utilization view of the features of a LMS.

A survey of the data mining techniques which could be useful for analysing a LMS is provided in [13]. The implementation of some of the techniques to the Moodle suite is also provided.

An interesting work on the importance of social network analysis is presented in [12], the intent is to understand the structures within groups of students. A tool aimed at establishing the educational social networks based on the asynchronous interaction provided by forums is presented and tested. Also in this case the idea is not to obtain a tool for controlling the utilization of the message system in a whole LMS but rather to obtain the structures arising among the students.

An extended method to create social graphs due to the social interaction in both synchronous (chat) and asynchronous (forums) contexts within a LMS is proposed in [2]. This study also provides an approach to take into account the time evolution of the bonds between the students.

In [19] it is introduced a nice experiment where the learning management system for two courses is replaced by Facebook groups. Since Facebook is an extremely popular social tool, it becomes natural to try to understand if it can effectively replace a LMS. The analysis provided in [19] suggests that the features of a carefully crafted LMS are still superior in respect to the utilization of Facebook groups for the same purposes and some students are concerned about their privacy when using Facebook instead of a social media devoted specifically to the learning environment.

### 3 Definition of the Indicators

The number of functionalities/activities in modern LMSs (such as, chat room, messaging system, forum, etc.) is increasing over time. For this reason, one can expect that some of them become more popular than others. In this respect, the range of users which can access to the functionality comprises two main groups: students and academic personnel. The actions performed by the students on a LMS include obtaining new skills, sharing learning material, communicating with peers, etc.; whereas academic personnel includes teachers, university managers, and LMS administrators. Who provide contents for the lessons and manage the functioning of the school activities.

This implies that the e-learning community is very heterogeneous, and thus requires a set of specialized tools for every possible role. It is common in fact that a LMS is provided with monitoring systems which allow to control the different activities happening on the platform. For example, a student might be interested in his/her own grades, a teacher might be interested in the activity

of the single student or a whole class within a single subject. This research proposes an approach to help the governance of an academic institution, where the utilization indicators of a LMS are naturally divided according to different features/parameters (such as courses, academic years, etc.) that allow to correlate the utilization with the structure of the courses. In order to avoid misunderstanding we stress that the term utilization is used in this article as a synonym of the amount of accesses to a given LMS functionality. This quantity can be analysed in detail according to particular needs. In this respect, we propose to consider the amount of accesses divided by the total number of possible users. This quantity has been called *specific utilization* (or, in short, *su*) and it is obtained with the formula:

$$su(a, t, \mathbf{p}) = \frac{\# \text{ of accesses } (a, t, \mathbf{p})}{\# \text{ of users who can access } (a, t)} \quad (1)$$

The *specific utilization* provides a direct insight of the diffusion of a specific LMS functionality among the users within a particular department/area/course of utilization ( $a$ ), at a given time ( $t$ ), according to one or more parameters indicated here with the vector ( $\mathbf{p}$ ). Let's consider, for example, the message system available to the students of the LMS in use at the Università degli Studi di Milano-Bicocca. In this case the parameter ( $\mathbf{p}$ ) of Eq. 1 refers to the fact that we want to distinguish between the sent messages and the received messages. We are also interested in distinguishing subsets of the whole community who can access to the functionality (e.g. males vs females, or particular roles within the university). The *specific utilization* indicator is not limited to social functionalities but can be implemented on all the possible activities that a LMS user can access to. The information required to calculate this quantity is a timestamp related to the access and an identifier of the person who performed the access. Binning the utilization within fixed time spans allows to set the time granularity of the information (the academic year is a very natural choice, but one can decide to follow shorter or longer time frames).

Although the *specific utilization* indicator provides a quick insight about the success of a functionality, it is important to consider that some activities (although accessible from the whole student population) might be aimed specifically to a restricted group of users. In this case it might be more interesting to obtain information about the utilization of those who really accessed to the functionality (while neglecting the information regarding those who could access to the functionality but for some reason did not do it, and would skew the resulting statistical properties of the indicator). A functionality relevant only for a small subset of the whole student population, could be successful but it might obtain a small *specific utilization* score because of the normalizing constant proportional to the whole student population. This is the case of the message system for the present case study (Moodle at Università degli studi di Milano-Bicocca) where all the enrolled students have access to it, but some departments have a very limited implementation of the platform and, as a result, it becomes essentially useless for the student to access message system.

The real usage of an activity in a LMS is thus an interesting quantity, which can be better understood with the help of what we called the *real utilization plot*. The idea of this plot is to display the distribution of the population utilizing a given functionality. For example, on the abscissae there is the number of accesses to the functionality, while on the ordinate is displayed the amount of students which used the activity that particular number of times. The distribution of the population returns valuable information about the success of the activity.

Let us consider a case of a functionality which does not require many accesses within the span of a year and thus which has a low *specific utilization* score; for example it could be the functionality related to the procedure of defining a student's exam plan. This functionality is expected to be subject to a limited number of accesses per student (but most of the students should access it). If the data utilization of this functionality shows that there is a large amount of students who accesses to this functionality tens of times, this might imply that the associated service does not provide clear indications on how to complete the procedure correctly and thus the students need many accesses before solving their problem. The same *specific utilization* could be obtained in a completely different scenario, like a social functionality of a LMS which allows the students to share information with their peers. In this case, if the vast majority of the students accesses this functionality a limited number of times it is reasonable to think that a critical number of users has not yet been reached and for this reason the functionality is not really working as a social binding mechanism. From the *real utilization plot* as detailed above, it becomes natural to extract the weighed average of utilization, which we call *popularity*:

$$popularity(a, t, \mathbf{p}) = \frac{\sum_{n=1}^{\infty} n \cdot U_n(a, t, \mathbf{p})}{\sum_{n=1}^{\infty} U_n(a, t, \mathbf{p})}, \quad (2)$$

where  $U_n(a, t, \mathbf{p})$  is the number of users who accessed  $n$  times to the functionality according to the department ( $a$ ), time ( $t$ ) and (possibly) a set of features denoted as ( $\mathbf{p}$ ). Notice that the sum runs over the number of accesses  $n$ , which ranges from 1 to infinity. This is not a problem since  $U_n$  is different from 0 only on a finite number of values.

We use as a constant to normalize the results the sum of the users who accessed the functionality. The meaning of the *popularity* indicators is to understand how much the functionality has been accessed by the real users and neglect those who had the possibility but did not utilize it.

The administrators of a LMS can exploit the *popularity* indicator to better allocate the resources of the system, in fact the average amount of resources for utilization times the *popularity* times the number of active users provides a good insight of the total amount of resources to be allocated, while the variation of the *popularity* in respect of a given time frame is useful to predict the change in resources which might be needed.

## 4 Implementation of the Indicators

A Social LMS can be considered the natural evolution of a LMS, however today the available data about the social interaction in an e-learning environment is scarce. For this reason, we resorted at applying the indicators defined in the previous sections to the data retrieved from the LMS actually in use at Università degli Studi di Milano-Bicocca, which is Moodle, version 3.1.3.

Moodle (based on social constructivism) is one of the most popular Learning Management Systems in use among universities (there are over 7000 sites in 233 nations which are based on it). Moodle is an open project under the GNU GPL license (which probably allowed it to become one of the most popular LMSs in the world). It is thought as a support tool for the creation and management of online courses. Most of the information collected by Moodle is in the form of a relational database; in this respect, by using MySQL we retrieved the relevant information to extract the indicators so far defined and we used R to analyze the data. One of the modules (functionalities) present in Moodle is the message one, and it allows the students to communicate among themselves, with the administrators and with the teachers. The message module can be thus conceived as a preliminary step in the direction of a Social LMS and for this reason it is the natural point where to start our analysis. This module is at disposal to all of the students of the Università degli Studi di Milano-Bicocca (which comprises about 35000 students per academic year, during the time of our analysis).

In this paragraph we will describe the application of the *specific utilization* indicator detailed in Sect. 3. The available data spans the three academic years 2013/2014, 2014/2015 and 2015/2016. The idea of behind this investigation is to help the governance and control of the university from the point of view of the administrators. The process of retrieving the department of each student was rather cumbersome: from each message, we were able to obtain the internal email of the sender, at this point it was possible to find all the courses where the student was enrolled. In the database, each course is associated with a “department/area”, and thus it was possible to link at least one department to each of the courses. Unfortunately, some of the courses were shared between different departments and this could lead to uncertainty, i.e. whether a student belonged to one or the other department. At the time when this analysis was carried out, this uncertainty could not be avoided, and as a result, some students have been classified in more than one department. Since this problem affects a minority of the population (less than 5%) we included in our analysis all the possible students for each department, allowing for duplicates. This implies that all the results presented in this paper are subject to an error of the order of 5% in the quantification of the indicators.

Although it might not be obvious, it is important to notice that the acts of sending and receiving messages provide different information regarding the use of the message functionality. One important reason for this distinction is that, in Moodle, the students can send one-to-one messages only, however teachers and administrators can send also one-to-many messages (this option has obviously

been introduced in order to create an easy notification system). In this respect we can state that there are two main purposes when utilizing the message system:

1. notification (one-to-many messages)
2. simple interaction (one-to-one messages)

Unfortunately there is no easy way to extract whether a message belongs to the one-to-many or to the one-to-one class; of course a direct analysis of the text, could allow us to make such a distinction but it was beyond the scope of this work to apply data mining techniques to the body of the messages. In the case of the *specific utilization* associated with the sent messages we take into account only those messages which were sent by the student population and remove those which are due to the academic staff/teachers. In this regard we want to use the *specific utilization* to understand the success of the message system among the students as a socialization tool.

At the beginning we will consider the results obtained from the *specific utilization* regarding the time distribution of the sent messages during the different months, in different years. In order to not display a rather complicated 3D picture with both of the messages and departments as free variables, we grouped together all the departments and we considered the time evolution only (Fig. 1). The academic year 2013/2014 was when Moodle has been introduced as the university LMS, and for this reason it was a still rather new tool for all the users. With the passing of the time, the users become more acquainted with the new LMS and in this respect it is no surprise that there has been an increase of *specific utilization* along the years.

When comparing the utilization results related with a single month, it can be noticed that for almost all the months there has been an increase from 2013 to 2016. The variation of utilization from year to year in many cases seems rather constant, however in some months there has been a sort of saturation and the increase has almost stopped, or worse in some cases during the last academic year under consideration there has been a decrease of utilization.

The received messages on the other hand can provide information from both a the social point of view and as notifications (in which they are just another tool to receive technical information). The first obvious difference which can be noticed (Fig. 2) about the sent and received messages by month is that the latter

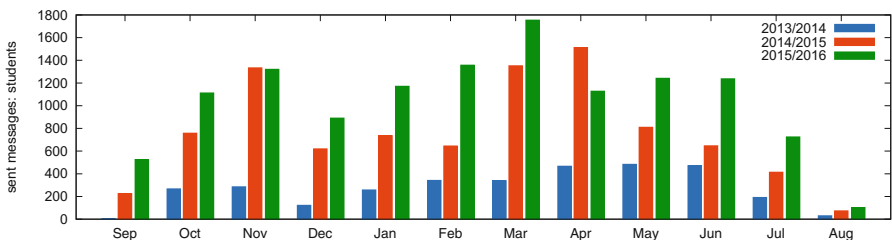


Fig. 1. Total amount of messages sent by month in the three academic years [1].



is about 5–6 times higher. Some features on the other hand are similar, also in this case there is a diffused increase of the number of received messages from one year to the next one, however this feature has many more exceptions in respect to the sent messages. For example, during the month of October 2014 more messages were received than the corresponding month of 2015, this being true also in February and March. There is also a case in which the first year of utilization surpasses the second one (see June of Fig. 2). When comparing the sent and received messages there is another anomaly which is very interesting. In the months of April associated with the academic year 2014/2015 and 2015/2016, in the case of the sent messages there is a decrease of utilization, while the opposite happens for the received messages which experience one of the highest increases over the three years.

The *specific utilization* associated with the sent messages among the different departments is shown in Fig. 3. The bars for each department refer to the three academic years under consideration. The two departments where the value of the indicator is higher are Sciences and Psychology. Almost all the departments are showing a steady increase of the value of *specific utilization*. On the other hand the absolute value of the *specific utilization* is generally very limited and it never exceeds 1. We also notice that Medical Sciences is the department with the lowest score of *specific utilization*.

If we take into account the received messages (Fig. 4), it is interesting to notice that the magnitude of the *specific utilization* is higher than the one of the sent messages (Fig. 3). In the case of the Sciences department (one of the departments with higher usage) the received messages outnumber the sent messages

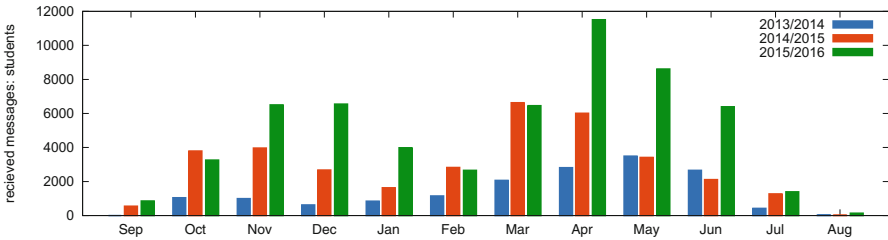


Fig. 2. Total amount of messages received by month in the three academic years.

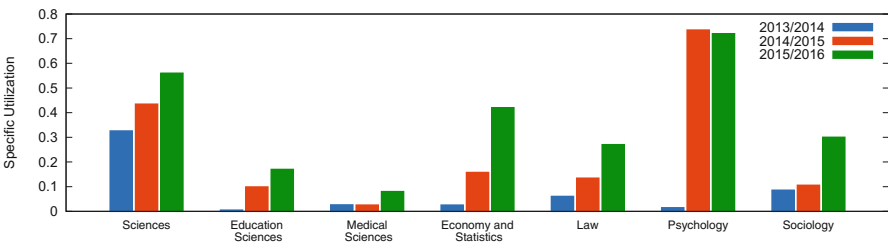
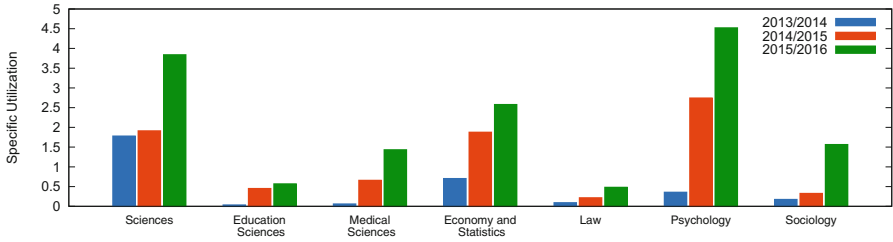


Fig. 3. Specific utilization of the sent messages among the students [1].



**Fig. 4.** Specific utilization of the messages received by the students.

more than 5 times for the academic year 2013/2014 and the gap increases over the years. Another interesting feature of this quantity in respect to the sent messages is that they show a rather poor correlation. For example in the case of Sciences, there has been a rather steady increase in the utilization of the sent message along the three academic years under consideration, while in the case of the received messages the highest increase has been registered passing from 2014/2015 to 2015/2016 when the *specific utilization* of the received messages has almost doubled passing from 2 to 4. This is not confirmed for the department of Economy and Statistics where the *specific utilization* of the sent messages has experienced the highest increase in the last year under consideration, at variance with respect to the received messages which in the corresponding period show the minimum increase. Even more curious is the case of the department of Psychology where there is an inversion. During the last year it has been registered a decrease of *specific utilization* of the sent messages and an increase of the one associated to the received messages.

There is an interesting feature to be noticed regarding the Psychology department since during the first year it had the second lowest score of *specific utilization*, while during the third year it jumped to the top position.

In this case we know that the internal regulations of the department forced all the teachers to migrate their courses over Moodle. In this respect, this is an example of the impact that policy regulations can have on the utilization of a feature. That being said, it is also fair to say that the policy regulation forcing a migration over Moodle had a limited effect if considered in terms of magnitude; in fact the *specific utilization* of the sent messages never exceeded the value of 1, which means that it is still used only by a niche of the total population which can access it.

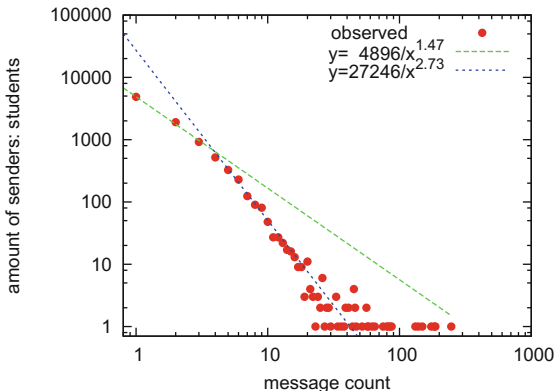
The message functionality of a LMS is of key importance to establish a social network within the learning community. The opinion of the authors is that the social network of a LMS suffers the very strong concurrence of other means of communication [18] which are already well established among the students. In fact, even considering that all of the departments follow this increasing trend it would take many years to reach *specific utilization* values of the order of tens of messages per year. For this reason, the present status of the message system seems to require a qualitative change related to the LMS in order to reach a critical level.

In Figs. 1 and 2 there is a clear indication of a seasonal behavior, however, in order to have a solid statistical indication a few more years would have been needed, and an analysis should be carried out in future investigations. In particular during July, August and September the amount of exchanged messages is lower than the other months (due to the summer breaks), while March and April are typical exam session months which spark the need to exchange information and for this reason the utilization during these months is higher.

#### 4.1 Popularity

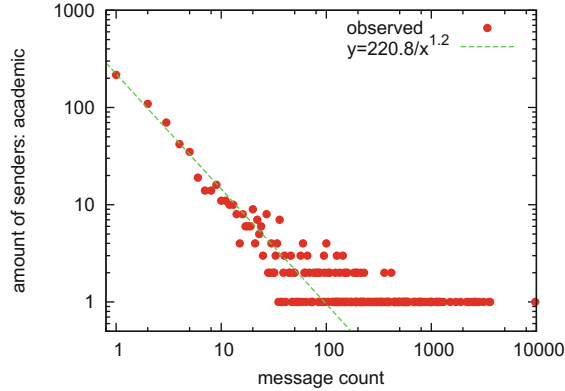
This section considers the *popularity* associated with the sent messages. These messages have been divided in two parts, those sent only by the students and those by the academic personnel (which includes staff and teachers). It should be noted that, in the following analysis, it was not possible to make a distinction between teachers and administrative staff. This, in turn, implies that it is not possible to associate a given department with the senders which are labeled as academic. In order to compare the *popularity* of the students and of the academic personnel, we did not divide them between different departments either.

The *real utilization plots* which result in this case span many orders of magnitude in terms of users and of sent messages. As a result Figs. 5 and 6 have been displayed with a double logarithmic plot. As a contextualization it is important to know that the total number of student senders in the academic years 2013/2014, 2014/2015 and 2015/2016 is 9330, and the number of sent messages amounts to 24881.



**Fig. 5.** Real utilization plot of the sent messages for the students [1].

As shown in the previous paragraph the *specific utilization* of this functionality of Moodle is scarce. Figure 5 shows the *real utilization plot* of the senders as a function of the number of messages sent during the three academic years.



**Fig. 6.** Real utilization plot of the sent messages for the academic personnel [1]

In order to better explain these numbers we decided to fit the data with parametric families of functions. Although there is no obvious parametrization which can constrain all the features of the plot, power laws can catch some important points, in the following we will refer to functions of the form:

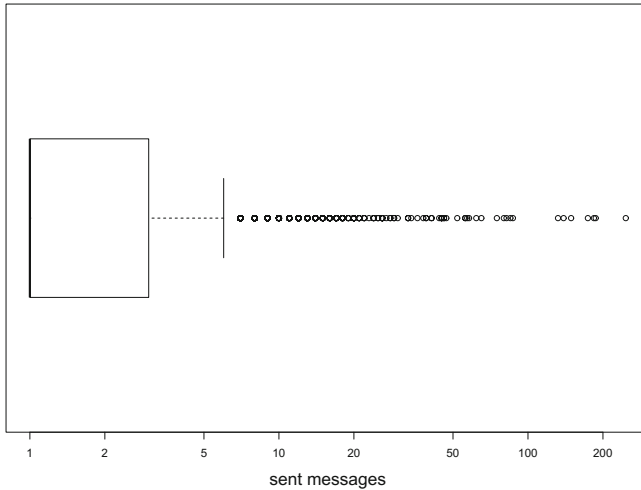
$$U_n = \frac{A}{n^k} \tag{3}$$

Where  $A$  is a constant,  $n$  represents the number of sent messages, while  $k$  involves the steepness of the power law (higher  $k$  implies a steeper descent as a function of  $n$ );  $U_n$  is the number of users who accessed  $n$  times to the functionality.

In Fig. 5 we can observe at least three different patterns. In the first part (between 1 and about 5 sent messages), the points are aligned, there is a kink in the distribution, while the points between 5 and 20 sent messages follow a straight line with a different slope in respect to the first ones. Above 20 messages it becomes difficult to consider the data as being produced by a simple parametrization. It should be taken into account that in this case there are many “exceptions”, i.e. only one student sent 132 messages, while nobody sent 131 or 133 messages, which provides a staggering distribution for the tail.

As a result of a nonlinear least squares fit, the power law which best fits the first part of the graph has a coefficient  $k = 1.5$  while  $A$  is 4896. It is interesting to remark that the value of  $A$  is essentially the number of students which sent just one message during the whole period of time comprising the three academic years. The parametrization which best fits the data between  $n = 5$  and  $n = 20$  is steeper, as a result the value of  $k$  is 2.73, and on the other hand  $A = 27246$  (this would have been the amount of students sending just one message if all the points had followed this parametrization). As noted above the tail is due to many single persons who sent large amounts of messages. In particular in Fig. 7 we show the boxplot related to the messages sent by the students. It is clear that the median is exactly at one sent message, and those messages which are beyond

6 can be considered as being outliers, which means that these points are related with students who sent more than  $\frac{3}{2}(Q_3 - Q_1)$  of the third quartile. In this case it is rather striking the fact that those students who sent 6 or more messages in three academic years can be considered as outliers in terms of sending many more messages than usual! In practice this includes about 8% of the active students.

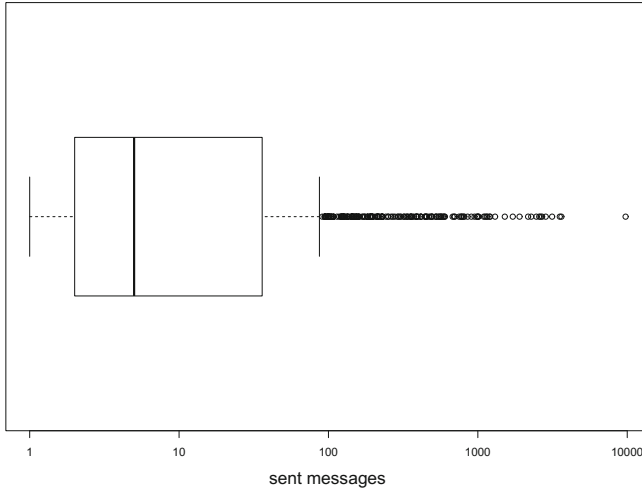


**Fig. 7.** Boxplot of the sent messages by the students.

The value of the *popularity* related to the student population is just 2.8. This confirms that also those who access to the message module do it very sporadically.

We expect that the utilization of the message module by the academic personnel should be rather different, this is due to the different roles of the teachers and the administrators but also because they can access to the one-to-many messages functionality. This possibility allows them to send notices to many people at the same time and thus the social aspect of the messages might not be the more important one, leaving room for a notification function. The amount of users is also very different; there are, in fact, 531 active senders and the total number of sent messages is 73357 (over the whole period of three academic years). In this case the *popularity* of the message system is 137, about 50 times higher than the *popularity* associated with the students. In Fig. 6, it is possible to notice that the data distribution does not show the same features of the student distribution, and a single power law can explain decently the points from  $n = 1$  to  $n \approx 40$ . In this case the exponent  $k$  is equal to 1.2. As the number of sent messages increases the distribution becomes more and more noisy. Also in this case, above a certain number of sent messages there is a very noticeable tail. In this case however, it is responsible for shifting the value of the *popularity* to higher values. In Fig. 8 we show the boxplot corresponding to the academic personnel. In this case it can

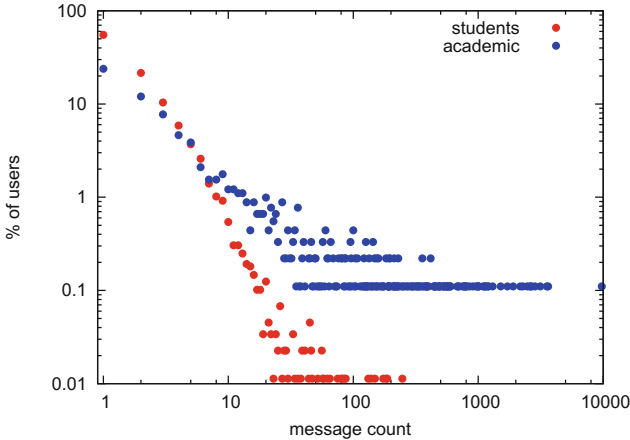
be noticed that the value of the median is 6 and that, in order for a member of the academic personnel to be considered as an outlier, he/she has to send more than 96 messages over the course of the 3 years. Those outlier cover about the 17% of the academic personnel.



**Fig. 8.** Boxplot of the academic personnel sending messages.

In order to better compare the behaviors of the students and the academic personnel we decided to display them on the same graph. On the other hand this would be meaningless when considering the bare amount of users. For this reason we re-normalized the amounts of users by dividing by the total number of users who accessed to the message module (times 100 in order to obtain percentage values). With this re-normalization on the y-axis there is the percentage of students sending a given number of times messages. The two plots combined are shown in Fig. 9.

The usage due to the students shows shorter tails than the academic counterpart and also the number of users drops more quickly as a function of the number of sent messages. It is interesting to notice that around 20% of the academic personnel who accessed the message system did it only once, while this quantity raises to about 45% in the case of the students (this seems a clear indication of the fact that the academic personnel is more involved in the message system of the LMS). A confrontation of the shapes of tails is misleading. In the tails, there are single users who sent many messages but when re-normalized on the total population this returns different percentage values. Nonetheless it is striking that a large percentage of the population of the personnel belongs to the tails while the numbers are much smaller for the student population.



**Fig. 9.** The percentage of message senders as a function of the number of sent messages [1].

## 4.2 Zipf Law

The Moodle message module is thought to enhance communication among the users. In the very same area (communication among human beings), but in a rather different context, i.e. when quantifying the usage of the words in a text, there is a very well known phenomenon called Zipf law [20]. This relation has been discovered by studying the appearance frequency of the words in different texts. A very striking feature, which is present in a large percentage of texts, showing very little dependency on the language or the purpose of the text, is that the first most common term appears two times more frequently than the second most common term, and it is three times more common than the third most used term, and so on...

In detail the frequency of the  $n^{\text{th}}$  more common word is  $1/n$  in respect to the most frequent term. In order to achieve a more general result it is possible to modify this formula where the appearance frequency  $f$  of the words (listed in order from the most common to the least one) read like:

$$\frac{f(1)}{f(n)} = n^k, \quad (4)$$

in detail, those parametrizations where the value of the parameter  $k$  is closer to 1 are more in line with the original formulation of the Zipf law. In this respect, the information gathered from the messages sent by the students seems to be rather distant from a Zipf law, for example because there are different features, a kink, long tails etc. (see Fig. 5)

The utilization of the message system by the academic personnel, however can be parametrized quite well with a single power law, where the coefficient  $k = 1.2$ . (see Fig. 6). In this respect it presents a curious resemblance to the original Zipf law.

The Zipf law has been associated with the principle of least effort [7], according to which humans tend to use the least effort if the result is acceptable for a given purpose. In this respect since there are easier means of communication it is reasonable that the students resorted at using the message module only when other systems were not feasible. The academic personnel, however, which does not have the same level of personal connection with the students was simplified by the features accessible via Moodle. This could be confirmed by the long tails, where, for the teachers/personnel, becomes easier to send one-to-many messages through the LMS rather than via normal email where they should input the name of each receiver.

In a successful message system (e.g. Facebook chat, Whatsapp, etc.) the information is naturally spread and enriched when passing from one person to the other. It is thus conceivable that the real utilization plot of a successful message system does not follow a Zipf-kind law, or at least that the exponent, associated with the descent in number of messages sent per person, should be very small ( $<1$ ).

## 5 Dashboard

The present work is part of a broader project aimed at obtaining better tools for learning environments. As a result in this paper we are going to introduce the tools being developed to access to the indicators so far explained. The most prominent way to access to the indicators is in the form of dashboards. Dashboards, in fact, are very popular Learning Analytic tools for presenting data. Among the main features of dashboards we can find the customizability, i.e. the possibility to insert new indicators and functionalities by importing widgets. Dashboards can be used to record and display all the learning activities in order to promote self-awareness, considerations, and help the students to define goals and track their evolution. Dashboards for learning analytics can be associated with three different groups [23]:

1. The dashboards which support frontal lessons. These dashboards support the teachers to obtain feedback from the lessons they deliver, and allow them to adjust the classes according to the level of the students. For example the dashboard defined by [24] uses a hardware system to keep track of the interest of the students by analyzing their voice and their head movements.
2. Dashboards made to enhance and facilitate group works belong to the second group. For example TinkerBoard [25] tries to characterize the development of the activity of every group by quantifying the commitment.
3. Online-learning support dashboards. In this case the information obtained are designed to be visualized in order to promote discussions within the classroom. Most of the information is gathered from the log files produced by the LMSs.

AAT (Academic Analytics Tool) and X-Ray Analytics are two kinds of dashboards which belong to the third group of the aforementioned classification, since



they extract data from the log files and they process them in order to provide useful information related to the engagement and performance of the students; in particular:

- AAT [21] is a dashboard connected to Moodle Analytics, which is a Moodle plugin. AAT was designed to be an easy access platform which enables to perform complex queries allowing to analyze the behavior of the students in the online courses. The users have a wide array of choices not limited to statistical indicators. The information extracted is mainly aimed at the analysis of single courses but it is possible to create more complex requests spanning more courses at the same time.
- X-Ray Analytics [22] is an application which enables to make predictions based on the information collected through MoodleRooms. It allows to analyze the trend which have an impact on the progresses and final results of the students. It allows to understand the learning behaviors in order to improve school performances and reduce the risk of failure ahead of time.

The dashboard produced for this project is different in respect to the ones just described in that it is focused on analyzing the social interactions of the users. These new characteristics of the users are more focused on soft skills rather than hard skills and for example include indicators of influence (which does not necessarily coincide with school skills). Each indicator is developed following a micro service strategy which allows the dashboard to be modular and adaptable to all the possible Technologies (it is thus not confined just to Moodle).

The project of this dashboard is based on three logical levels as shown in Fig. 10.

1. The first level is related to the storage, where data coming from different sources is analysed. Since the data can be defined in different ways it is required to make a unified representation model obtained with an intermediate level called ETL (Extraction, Transformation, Loading). The output is then saved in a data warehouse.
2. The second level includes data analysis and the definition of the indicators by using R and Python. The HRMS component (Handler Requests Micro Services) manages the requests of the client via restful calls, extracting the data of the requested indicators and later rewritten in JSON format. This last format is then sent to the client which will interpret graphically the data.
3. The third level deals with the front end built on a web application. The framework that we chose is Angular2 which allows to implement web applications usable on all sort of devices, like smartphones, tablets, desktops and laptops. This framework supports the programming language called Typescript. The modern aspect is obtained with material design styles. The library Highcharts is responsible for the interactive graphs within the dashboard.

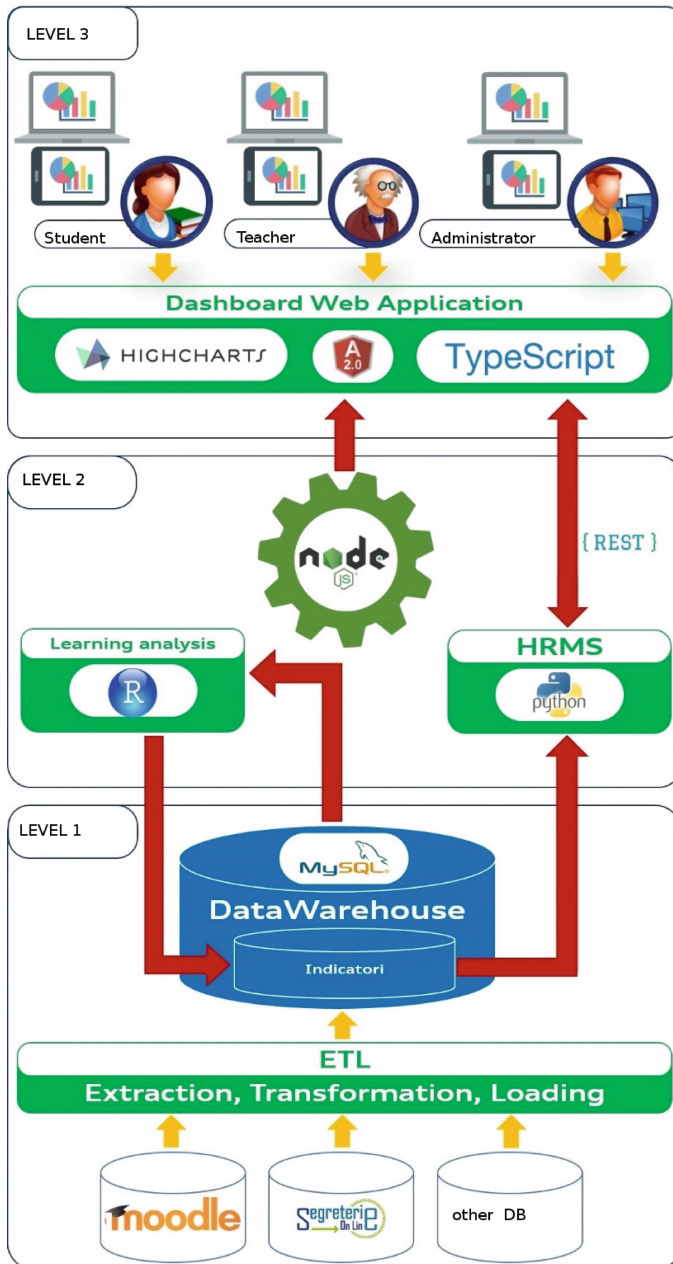


Fig. 10. Architecture of the dashboard.

## 6 Conclusions

The aim of this research is to provide new tools for analysing the amount of accesses to the functionalities of a LMS. In particular we are interested in assessing the utilization of those features which can potentially lead to a Social LMS where the information provided by the students plays a central role in the learning process. Since the social features of such a system might require a careful allocation of resources, this paper is mainly addressed at the administrators of the LMS who need to be in full control of the needs of the system. For this reason, two main indicators have been presented, *specific utilization* and *popularity*. The first one refers to the average utilization of a functionality among all the possible users, while the second one is more specific and it grants an insight about the detail of the real utilization by the users.

After defining the indicators, we tested them by analysing the data obtained by the LMS presently in use at the Università degli studi di Milano-Bicocca, which is an instance of Moodle (version 3.1). This LMS cannot be considered as a Social LMS since it is not centered around the social activity of the students, however there are social functionalities, like the message system. With the help of the newly defined indicators it was possible to confirm that the message functionality has not yet reached a critical stage in which there are active groups which create a self-sustained community. The development over time of the utilization has also been addressed and an increase of the values returned by the indicators has been reported. A similar check, related to the different departments has also been done, indicating that there are some departments which are more active than others, in particular, in one case this was attributed to a change of utilization policy defined at academic level. Even when taking into account the increase of *specific utilization*, in the near-medium term there is no foreseeable intense utilization of the message system, even worse, there are possible hints that the present stage of utilization might start to enter into a saturation phase soon (e.g. a slight decrease of utilization from one year to the next, related to particular months).

After considering the accesses as a whole, it was interesting to check in detail the real utilization of the system, with the help of the second indicator here defined. However, the value of the *popularity* of the message system (senders) among the students is also rather scarce (being around 3). A different perspective can be obtained by looking at the same functionality from the point of view of the academic personnel, where the *popularity* reaches a value of 137. By exploiting the *real utilization plot* it was possible to understand the detail of the noticeable difference between the students and the academic personnel. In the case of this second group of people the tails of the distribution of the users are responsible for this higher value of *popularity*. This is due to the fact that the academic personnel can send one-to-many messages and thus this functionality as a notification tool rather than a social tool.

Although the acts of sending and receiving messages might appear very similar they can carry rather different information. The results for *specific utilization* related to the received messages from the student population are in fact about 4–5 times higher than those related to the sent messages.

The distributions which appear in the *real utilization plot*, span many orders of magnitude and since they are essentially aligned (on double logarithmic plots), it appeared natural to parametrize them with power laws. This behavior is very similar to the empirical Zipf law which accounts for the appearance frequency of the words within a text. This law has been linked to the principle of least effort, where a person tends, provided a result is obtained, to employ the least possible energy in accomplishing it. In this respect one has to take into account that, nowadays, there are plenty of social tools which are extremely popular among the student population, and which are more easily accessible than the message system provided by Moodle. The users of the message system are reasonably more likely employing those tools when they need to communicate. The academic population, on the other hand, is facilitated by the Moodle message system when it is used as notification tool and for this reason the *popularity* of the messages is much higher in this group.

One of the problems related with the creation of a strong educational social network is the fact that the communication medium is limited in time by the enrollment of the students which follows the natural development of their career.

In an approach where a strong community is considered an important asset for obtaining a better education process it seems reasonable to suggest important changes of perspective when designing a Social LMS. In particular, an integration with existing social networks might benefit in terms of allowing the students to access to a familiar social feature.

The general approach of using indicators to control the development of a LMS requires means of visualization and access to the result. For this reason we introduced also the dashboard which is being developed. One of the natural properties of this dashboard is customizability, which in this case is achieved by using a modular approach where a user can define and implement the desired indicators by a simple action with the mouse.

## References

1. Meluso, F., Avogadro, P., Calegari, S., Dominoni, M.: Success of the functionalities of a learning management system. In: DATA 2017 6th International Conference on Data Science, Technology and Applications, 24–26 July 2017, Madrid, Spain (2017)
2. Avogadro, P., Calegari, S., Dominoni, M.: Analyzing social learning management systems for educational environments. In: Fred, A., Dietz, J.L.G., Aveiro, D., Liu, K., Filipe, J. (eds.) IC3K 2015. CCIS, vol. 631, pp. 470–491. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-52758-1\\_25](https://doi.org/10.1007/978-3-319-52758-1_25)
3. Avogadro, P., Calegari, S., Dominoni, M.: Expert students in social learning management systems. *Interact. Technol. Smart Educ.* **13**(3), 202–217 (2016)
4. Dalsgaard, C.: Social software: e-learning beyond learning management systems. *Eur. J. Open Distance E-Learning* **2**(1), 1–9 (2006)
5. Ferguson, R.: Learning analytics don't just measure students progress they can shape it (2014). [theguardian.com](http://theguardian.com)
6. Filiz, O., Kurt, A.: Flipped learning: misunderstanding and the truth. *J. Educ. Res.* **5**(1), 215–229 (2015)

7. Kingsley, G.: *Human Behavior and the Principle of Least Effort*. Addison-Wesley Press, Cambridge (1946)
8. Lukarov, V., Chatti, M.A., Schroeder, U.: Learning analytics evaluation - beyond usability. In: *Proceedings of DeLFI Workshops 2015 Co-located with 13th e-Learning Conference of the German Computer Society (DeLFI 2015)*, Seiten/Artikel-Nr: 123–131 (2015)
9. Macfadyen, L.P., Dawson, S.: Mining LMS data to develop an early warning system for educators: a proof of concept. *Comput. Educ.* **54**, 588–599 (2010)
10. Mödritscher, F., Andergassen, M., Neumann, G.: Dependencies between e-learning usage patterns and learning results. In: Lindstaedt, S., Granitzer, M. (eds.) *Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies (i-Know 2013)*, article 24, 8 pages. ACM, New York (2013)
11. Moodlerooms. XRay learning analytics 2017. <https://www.moodlerooms.com/resource/x-ray-learning-analytics/>. Accessed 26 Apr 2017
12. Rabbany, R., Takaffoli, M., Zaïane, O.R.: Analyzing participation of students in online courses using social network analysis techniques. In: *Proceedings of the 4th International Conference on Educational Data Mining, Eindhoven, The Netherlands, 6–8 July 2011* (2011)
13. Romero, C., Ventura, S., García, E.: Data mining in course management systems: Moodle case study and tutorial. *Comput. Educ.* **51**(1), 368–384 (2007)
14. Scheffel, M.: A framework of quality indicators for learning analytics. *Learn. Anal. Rev.* no. 2 (2015). ISSN 2057-7494
15. Scheffel, M., Drachsler, H., Specht, M.: Developing an evaluation framework of quality indicators for learning analytics. In: *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge LAK 2015*, pp. 16–20. ACM Press, New York (2015)
16. Scheffel, M., Drachsler, H., Stoyanov, S.: Quality indicators for learning analytics. *Educ. Technol. Soc.* **17**(4), 117–132 (2014)
17. Sclater, N., Peasgood, A., Mullan, J.: Learning analytics in higher education. *Jisc* (2016)
18. Susilo, A.: Exploring facebook and whatsapp as supporting social network applications for English learning in higher education. In: *PDE Professional Development in Education Conference 2014, 11–12 June 2014, Park Hotel Bandung* (2014)
19. Wang, Q., Woo, H.L., Quek, C.L., Yang, Y., Liu, M.: Using the facebook group as a learning management system: an exploratory study. *Comput. Educ.* **54**, 588–599 (2010)
20. Zipf, G.K.: *The Psychobiology of Language*. Houghton-Mifflin, Oxford (1935)
21. Graf, S., Ives, C., Rahman, N., Ferri, A.: AAT: a tool for accessing and analysing students' behavior data in learning systems. In: *Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK 2011)*, pp. 174–179. ACM, New York (2011)
22. Website (2016). <https://www.moodlerooms.com/resource/x-ray-learning-analytics/>
23. Verbert, K., Govaerts, S., Duval, E., Santos, J., Van Assche, F., Parra, G., Klerkx, J.: Learning dashboards: an overview and future research opportunities. *Pers. Ubiquit. Comput.* **18**(6), 1499–1514 (2014)
24. Yu, Y.-C., You, S.-C.D., Tsai, D.-R.: Social interaction feedback system for the smart classroom. In: *Proceedings of the 2012 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 500–501. IEEE (2012)
25. Son, L.H.: Supporting reflection and classroom orchestration with tangible tabletops. Ph.D. thesis. [http://biblion.epfl.ch/EPFL/theses/2012/5313/5313\\_abs.pdf](http://biblion.epfl.ch/EPFL/theses/2012/5313/5313_abs.pdf)