

A Proposed Learner Activity Taxonomy and a Framework for Analysing Learner Engagement versus Performance using Big Educational Data

Stathis Th. Konstantinidis^a, Aaron Fecowycz^a, Kirstie Coolin^a, Heather Wharrad^a, George Konstantinidis^b, Panagiotis D. Bamidis^c

^a School of Health Sciences, The University of Nottingham, Nottingham, United Kingdom

^b Department of Computer Science, University of Oxford, United Kingdom

^c Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece

Corresponding author: Stathis.Konstantinidis@nottingham.ac.uk

Abstract— The inclusion of information and communication technologies in Healthcare and Medical Education is a fact nowadays. Furthermore numerous virtual learning environments have been established in order to host both educational material and learner’s online activities. Online modules in a VLE can be designed in very different ways being part of different types of courses, while different models can be used to design the course based on what the creator aims to achieve. Thus, the types and the importance of the different elements of the online course may vary a lot. At the same time the need of a global approach to gather big educational data in order to provide valid meaning to the data through learning analytics and educational data mining is urgent. In order this to be achievable we propose a Learner Activity Taxonomy in which the different elements of the learners activity data can be categorised and a Learner Engagement Framework in which the importance of the different elements is vital in order for an analysis of the big educational data to provide a meaningful result. The initial application to practice of the Taxonomy and the Framework are presented based on data from 3 modules at 2 Universities, while the impact of them along with its limitations are discussed.

Keywords—*Learning Analytics; Big Data; learner engagement; online learning analysis, activity data; paradata;*

I. INTRODUCTION (HEADING 1)

The inclusion of information and communication technologies in Healthcare and Medical Education is a fact nowadays. Further to this numerous learning content management systems (LCMS) or virtual learning environments (VLE) have been established in order to host both educational material and learner’s online activities. The majority of these environments keep logs from user (learner, educator, administrator, etc.) actions accessing content, interacting with other users and performing online tasks. The amount of data produced by these environments is enormous, in a structured format, forming a pool of Educational Big Data.

At the same time multiple studies show increased engagement of users in new technologies [1], [2], and while different approaches have value the measure of students

engagement in learning content management systems [3], [4], more have looked into ways of increasing engagement both in traditional [5] and online or blended learning [6]. Fewer studies though aim to measure the engagement of learners in online environments and study the relationship with student performance.

There are many different and isolated studies of learner generated data from a range of different systems and activities. What we lack however, is a global and consistent approach to organise data into meaningful categories that will enable large datasets to be combined and queried in such a way as to construct new knowledge of how learners’ activity can predict performance, and subsequently provide evidence of how online courses can be constructed for optimum learning and attainment. In this paper we propose a Taxonomy of learners’ engagement activity and a Framework to measure the engagement of learners on the VLE against their performance. Initial results utilising the proposed Taxonomy and Framework are presented.

The remainder of the paper is organised as follows. Initially we set the scene in the background section. In the next sections the proposed Taxonomy and the Framework are described, while initial results utilising the proposed advances are presented. The last section highlights the impact of this framework through discussion, while limitations and future steps are presented.

II. BACKGROUND

A. Learning Analytics

Learning analytics have the potential to be used for a variety of reasons [3] from identifying information for the online content and for providing personalised feedback for learners and educators for evaluation of educational programmes. These data can be used by organisations to make policy decisions regarding return on investment in online content, e-learning platform and organisation’s performance.

Efforts for utilising analytics have started to appear mainly on the Educational layer [3] – a layer of research on analysis of training and learning procedures – and the identification of students learning behaviour and provision of feedback towards a personalised learning approach.

B. Learners Engagement in Online Learning

Student engagement in higher education has been widely researched [7] and four approaches explaining engagement can be identified [7]: the behavioural perspective – emphasis on effective teaching practice; the psychological perspective, emphasis on engagement as an internal individual process; the socio-cultural perspective- emphasis on socio-cultural context; and a holistic perspective – a combination of the previous 3 perspectives.

Researchers have started to set the groundwork of understanding of the type and level of student engagement in online learning based on students views [8], [9], with emphasis placed on cognitive engagement [10] – *integration and utilization of students’ motivations and strategies in the course of their learning* -.

Furthermore, students’ engagement in a course can be also measured through activity data. A couple of models have been proposed to capture activity data or paradata [11].

Several studies [12], [13] correlate student engagement with performance with positive results. The majority of such studies as a measurement of engagement of learners use the visit or hit at a page/activity type, while a few take into consideration the time each user spends on different activity types but only for the visualisation of the results [14].

A Conceptual Framework for Online Learning [15] identified 3 types of learner experience: Expository instruction (technology delivers the content); Active learning (learners actions lead to explore information or address problems); and Interactive learning (technology mediates human interaction) have been proposed. The framework also distinguishes between synchronous and asynchronous learning.

III. THE FRUITS IN THE BASKET

A. A Taxonomy for learners’ activity in a VLE

Online modules in a VLE can be designed in very different ways depending on the type of courses, e.g. an online course, a blended learning course or a content repository. To this extent, different models can be used to design the course based on what the creator aims to achieve. For example a “Flex” model in which the online learning is the main source of learning and educators provide onsite support or a "rotation: model, in which students shift between online learning and traditional learning [16]. Thus the types and the importance of the different elements of the online course may vary a lot.

At the same time the need for a global approach to gather big educational data in order to provide valid meaning to the data through learning analytics and educational data mining is urgent. In order for this to be achievable we propose a taxonomy in which the different elements of the learners’ activity data can be categorised and a framework in which the

importance of the different elements is identified in order for an analysis of the big educational data to provide a meaningful result.

An activity in a VLE can be defined as any action the user performs in a VLE which the system monitors.

The Learner Activity Taxonomy is composed of 6 main categories as depicted in Fig. 1. Each category should be assigned with the activities based on intention of using them and not in one to one matching between activity types of the VLE and taxonomy categories. For example a “forum” in a VLE should not be automatically categorised as a collaborative activity as might contain only information about administrative information for the course. In that case it is an administrative resource.

The “Individual Activity” should include all the action types which engage a type of interaction with the system and not being an assessment activity type. This means that a user visiting a “Page” activity type which includes an interactive learning task goes under this category, while a visit to a “Page” with static text belongs to one of the resources categories.

The “Collaborative Activity” should include all the activities which engage a type of interaction with other users. This could be both synchronous and asynchronous and include not only conversational interactions, but also tasks based collaboration. For example, participation in a Forum dialogue comes under this category, but also the collaborative group work to solve an online Virtual Patient exercise.

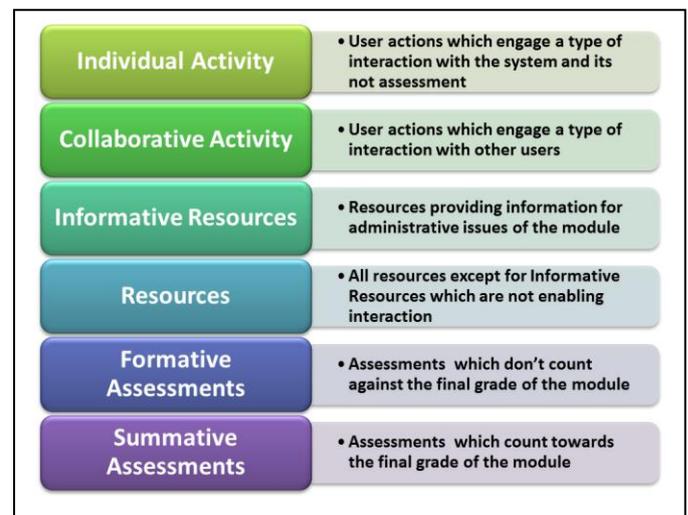


Fig. 1. Learner Activity Taxonomy. (figure caption)

All the activities containing knowledge which are not user interactive will be under either the “Informative Resources” or the “Resources” category. The “Informative Resources” include support materials relating to the administration of the module and “Resources” include all other course materials. For instance, a module handbook goes under Informative Resources while visiting a URL of a static web site is categorised as Resource.

The assessments activities are divided into two categories: the “Formative Assessments” which aim to help the students

identify their strengths and weaknesses and point out areas that need further study which have very small or no value towards the final grade; and the “Summative Assessments” which have the aim to evaluate student learning and count against the final grade of the module.

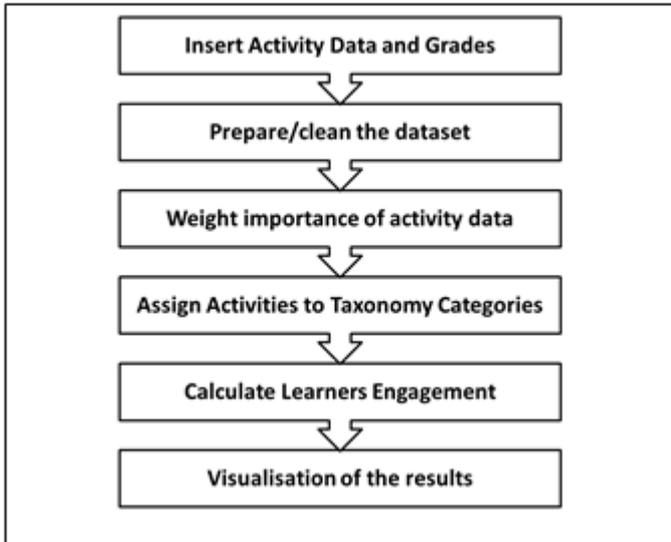


Fig. 2. Abstract process for Analysing and Visualising educational activity data

B. A Framework for analysing user engagement through VLE Big Educational Data

The amount of data that can be analysed for a single module can vary a lot, ranging from a few kilobytes to megabytes. In order for the data to be suitable for processing they should be stored following a structured format. Figure 2 depicts the abstract steps that should be taken in order to provide an effective analysis on the data.

The first step requires the insertion of different datasets of the module into the silo and their unification and correlation. For example these data could be an activity report from a VLE, summative assessments from another system and the final performance on the module (e.g. Final Grade) from the institutional administration system. The second step should be the clearance of any useless data. Useless data can be considered activities from the system or activities for resubmission of assignments, etc. The third step of the process is to assign importance to the different activities of the user in the VLE, while the next one is to assign the different activities to the Taxonomy categories. Then the learners’ engagement can be calculated and visualisation of results can be implemented.

Aligned with this process, the following framework (Figure 3) is proposed in order for the results of different modules to be comparable. Thus each activity type gets a weight of importance which can vary from 1% (for low importance) to 100% for high importance. Then all the activity types are assigned to the different Taxonomy categories. The next step is to calculate each learner’s engagement by multiplying the number of actions for each type with the weight of importance for this weight. The following step is to calculate the

engagement for each different category of the taxonomy. The final engagement is the sum of the engagements of the different categories.

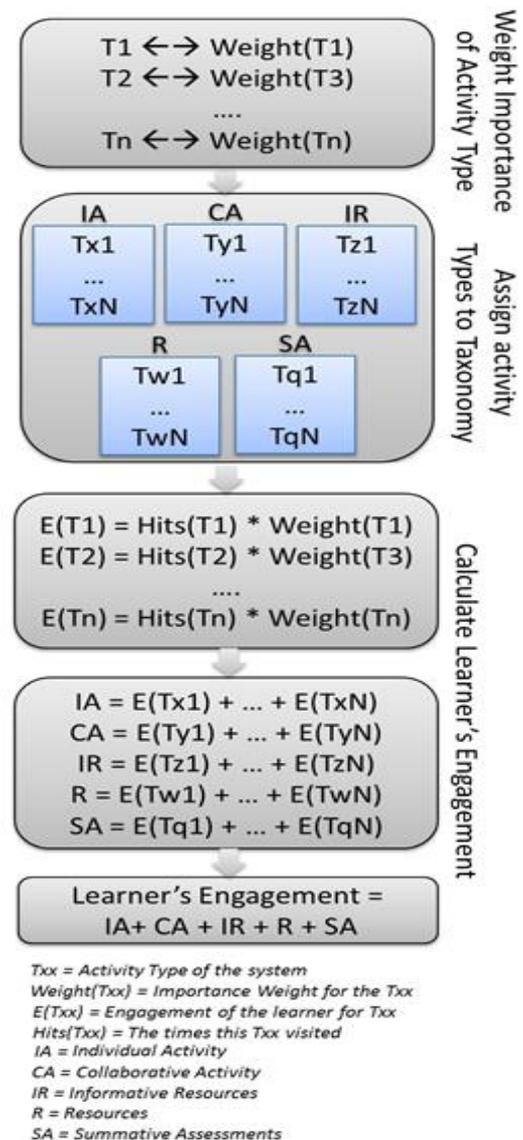


Fig. 3. Learner Engagement Framework

When Visualisation of the results is calculated, it is important to be stressed that the result is relevant to that particular module’s design. For example learner engagement on different modules is affected by the original design and pedagogical intention within a particular module. For this reason, the visualisation results are compared to an average set of engagement results, rather than to another module so as avoid a misinterpretation of the results. A visualisation using absolute numbers could lead to misinterpretation of the results, thus a visualisation comparing learner engagement with the average engagement is proposed.

C. Data Used

We analysed the data from 3 modules belonging to 3 different undergraduate courses from 2 Universities: The

University of Nottingham (UoN) and the Aristotle University of Thessaloniki (AUTH). The data from the AUTH module were from 2 academic years, thus they are treated as 2 separate modules. The following table depicts the number of users and hits per module. The data were exported as CSV files from the institutional VLEs which was Moodle in both cases.

TABLE I. SOURCE OF DATA USED

Module Code	University	Notes	Students No	Hits No
1	UoN		54	483
2	UoN		395	39,386
3	AUTH	1st year	103	13,066
4	AUTH	2nd year	42	1,249

IV. EXAMPLES OF APPLICATION IN PRACTICE

In this section we analyse 3 modules from 2 institutions to showcase the use of the proposed Learner Activity Taxonomy and Learner Engagement Framework in practice and to identify potential strengths and weakness.

The following figure (fig. 4) depicts the AUTH module using the proposed taxonomy. The yellow line depicts the first year module, while the blue dotted line the second year. The differences between the years are small for the “Individual Activity”, “Collaborative Activity” and “Informative Resources” categories, but the second year of the module shows the students visiting the “Summative Assessments” category with more frequency with less emphasis on “Course Resources”.

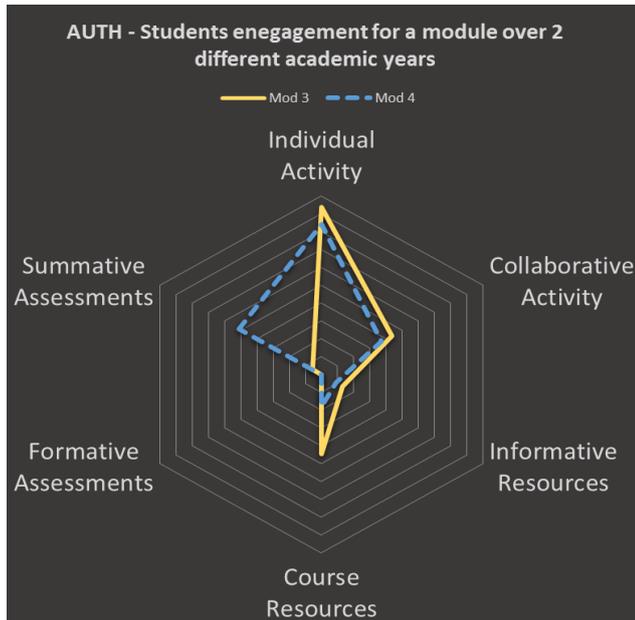


Fig. 4. Students engagement for a module over 2 different academic years depicting according to the proposed Taxonomy.

Figure 5 depicts the engagement of students in the 3 different modules (1, 2 and 3). Module 3 is the only module that has any student engagement in “Collaborative Activity” while module 2 is the only module having “Formative

Assessments” engagement (~3% of the total student engagement in the module). In module 2 the student engagement is more or less balanced between “Individual Activity”, “Informative Resources”, “Course Resources” and “Summative Assessments” 32%, 19%, 22% and 24% respectively. In module 1 “Individual Activity” actions are central to the module since students engagement is over 54% of the total student engagement in the module.

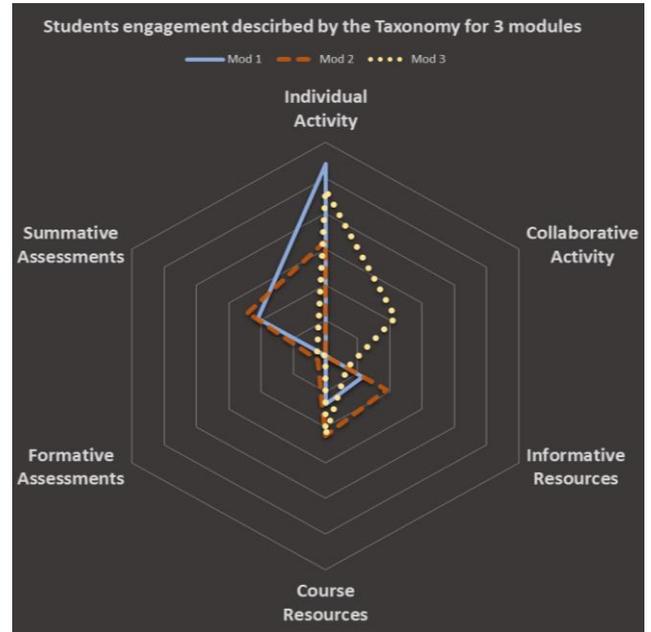


Fig. 5. Students engagement for 3 module comparison depicted according to the proposed Taxonomy. The representation is percentages of average engagement of students of the total engagement of these student in the module.

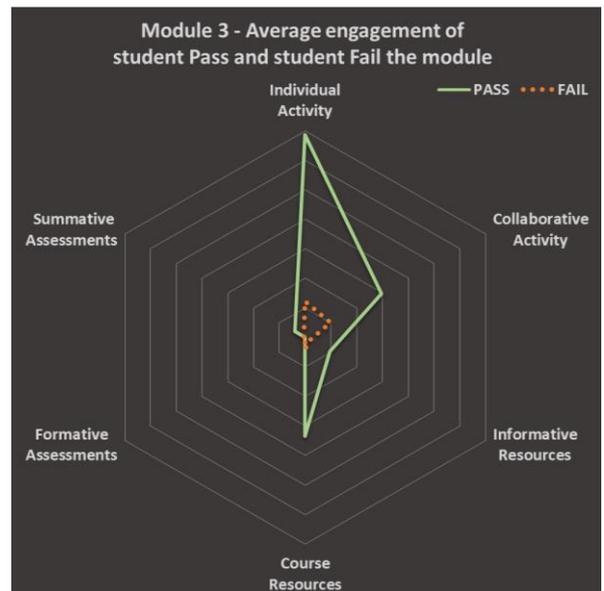


Fig. 6. Average engagement of student Pass and student Fail within the module depicted according to the proposed Taxonomy.

Student Engagement between those who pass or fail the module might be extreme such as the case of module 3 (Fig.6).

In this module thirty six (36) students that failed the module were having significantly less engagement compared to the sixty seven (67) students that passed the course. But, as depicted in figure 6, the engagement preferences against the Taxonomy categories were almost similar. The main differences were in the “Collaborative Activity” category: Students that failed the module had an average of 35% engagement of their total engagement with the module in this category, whereas students who passed the module had 20% engagement. In the “Course Resources” category, students that failed the module had an average of 15% engagement of their total engagement in the module compared to the 23% of the students that passed the module.

Applying the Learner Engagement to this module (module 3) the relevance of engagement with performance increased. The number of students with grades less than the average grade was decreased from 60 to 58, while the number of student with grades more than the average increased from 43 to 45 respectively. The number of students with grades less than the average is 36 while above the average 67. Similar adjustments can be seen in the other modules as well following the proposed framework. Figure 7 depicts the engagement of the students against their performance for module 3. Four different groups can be identified regarding their engagement related to the performance. There is a positive trend in data that represents the relationship between the engagement and performance.

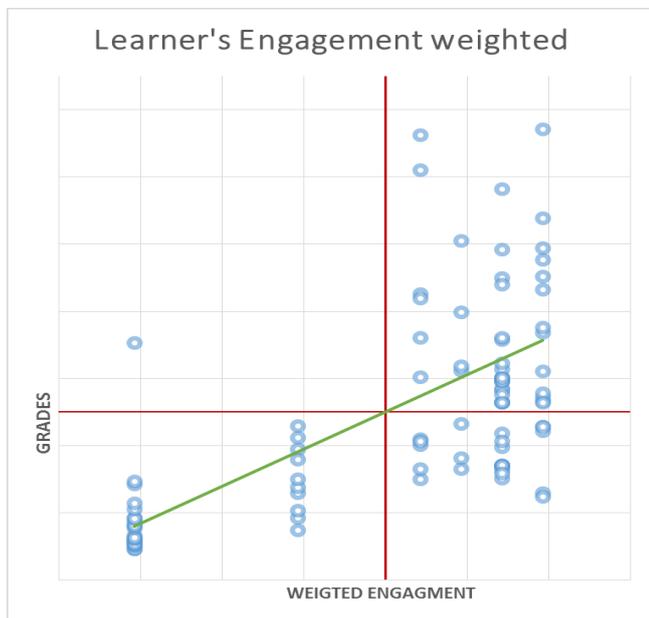


Fig. 7. Students weighted engagement for the module 3. Engagment and grades have been calculated agains their averages.

V. DISCUSSION

Differences in students’ engagement between consecutive years on the same module can reflect changes in the module’s curriculum, or different behaviours of the student cohorts. These differences are difficult to monitor and understand without the use of a common measure. The proposed Taxonomy not only provides this opportunity, but also can be

the basis of meaningful visualisation to influence better design. The differences between the years are small for the Individual Activity, Collaborative Activity and Informative Resources categories, but on the second year of the module the students were visiting more and put less emphasis on “Course Resources”. This can easily be explained when focusing in on the specific module (fig 4), as the second year comprised of more, but smaller sized assignments

The comparison of student engagement in different modules is only achievable if a common representation exists. The proposed Learner Activity Taxonomy provides this and also enables comparison between the different elements of the modules. Thus as Fig 5 depicts, one can argue that modules 1 and 2 do not have collaborative activities, but the educators possibly had put more emphasis on interactive resources that interactions more with the system, such as Reusable Learning Resources (RLOs), therefore produced more data about interaction [17]. Another possible explanation might be that the design of the module followed a flipped learning methodology in which collaboration took place face-to-face. According to [3] these kind of conclusions would help medical educational content providers to understand how their content is used and also medical educational institutions and services to evaluate their modules.

The representation of the engagement as a percentage of the average and not as exact numbers is used for two reasons. First, it provides an extra layer of anonymity in the data, while at the same time making them more meaningful. There are a number of factors that can influence the students’ engagement including the size of the class [18] and the design of the course itself [19]. Thus the comparison of each student compared with the average student of the same module will provide more trustworthy results.

One can argue that Figure 6 is resulting that the collaborative behavior of the student who passed wasn’t significant in helping them pass; or the engagement in this category of activity was low. Further research is needed in cases like this to produce valid conclusions and the use of a classifier algorithm such as regression analysis or J48 [20] might shed some light.

In the example presented in Figure 7, the performance is related really well with the engagement. However there are two groups that need further research. Students with low engagement and high performance and students’ high engagement and low performance. Engagement in terms of time should be also taken into account, while multiple resubmissions of a failed assignment can also be a factor creating an imbalance between engagement and performance. Despite the above limitation the above relationship can act as an indicator to health/medical students who need to track and measure their performance, and identify areas for improvement or to the medical educators who can identify people at risk of failing the module.

An improvement of the relevance of engagement with performance observed. The improvement was relatively small, but it can be influenced by multiple factors including the educator assigned the weights or the initial relation between engagement and performance.

One of the limitations, but at the same time strengths of the proposed Framework is that a person have to assign weights. In order the Framework to produce a meaningful result the weights of the different activities but also the assignment of the activities to the different categories is very important. A person who knows both the content and the educational design of module should be responsible for providing these weights. It could be an experienced educator, the module leader, the educator designed the module or even better a collaboration task of the above.

VI. CONCLUSION AND FUTURE ACTIONS

In this paper we proposed a Learner Activity Taxonomy to categorise and represent the students' actions in an online environment. Furthermore a Learner's Engagement Framework was created which can adjust the engagement according to educator's opinion on the importance of module's different elements. Initial application into practice revealed strengths in term of a common representation of learning engagement for different modules, but also of learning engagement between different groups of students. Further to this, the proposed taxonomy can be the basis for a common framework for students' online activity data representation in order for prediction models to be built and analysis of data to be performed. The Learner Engagement Framework showed a small difference in the engagement data vs performance for our data, however further research is needed to validate it as a tool that improves the measurement of engagement. Further research is also needed to identify the factors that influence the online engagement of the students, while additional information regarding the students might also reveal important relationships between engagement and performance. Nevertheless the proposed Taxonomy and Framework already reveal the usefulness of analysing the learners' activity data and could be the basis for Big Educational Data representation in which analysis may lead to personalised feedback and an enhanced educational experience.

REFERENCES

- [1] R. Junco, G. Heiberger, and E. Loken, "The effect of Twitter on college student engagement and grades," *J. Comput. Assist. Learn.*, vol. 27, no. 2, pp. 119–132, Apr. 2011.
- [2] S. T. Konstantinidis, P. D. Bamidis, and E. Kaldoudi, "Active blended learning in medical education — Combination of WEB 2.0 problem based learning and computer based audience response systems," in *2009 22nd IEEE International Symposium on Computer-Based Medical Systems*, 2009, pp. 1–6.
- [3] S. T. Konstantinidis and P. D. Bamidis, "Why Decision Support Systems are Important for Medical Education," *Heal. Technol. Lett.*, vol. 3, no. 1, pp. 56–60, 2016.
- [4] C. C. Robinson and H. Hullinger, "New Benchmarks in Higher Education: Student Engagement in Online Learning," <http://dx.doi.org/10.3200/JOEB.84.2.101-109>, 2010.
- [5] J. A. Fredricks, "The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments," in *Handbook of Research on Student Engagement*, S. L. Christenson, A. L. Reschly, and C. Wylie, Eds. Springer Science+Business Media, 2012, pp. 763–782.
- [6] P. D. Bamidis, S. T. Konstantinidis, C. Bratsas, and M. S. Iyengar, "Federating learning management systems for medical education: A persuasive technologies perspective," in *2011 24th International Symposium on Computer-Based Medical Systems (CBMS)*, 2011, pp. 1–6.
- [7] E. R. Kahu, "Framing student engagement in higher education," *Stud. High. Educ.*, vol. 38, no. 5, pp. 758–773, Jun. 2013.
- [8] C. C. Robinson and H. Hullinger, "New Benchmarks in Higher Education: Student Engagement in Online Learning," *J. Educ. Bus.*, vol. 84, no. 2, pp. 101–109, 2008.
- [9] M. D. Dixon, "Creating effective student engagement in online courses: What do students find engaging?," *J. Scholarsh. Teach. Learn.*, vol. 10, no. 2, pp. 1–13, 2010.
- [10] J. C. Richardson and T. Newby, "The Role of Students' Cognitive Engagement in Online Learning," *Am. J. Distance Educ.*, vol. 20, no. 1, pp. 23–37, 2006.
- [11] L. M. Campbell and P. Barker, "Activity Data and Paradata," *JiSC Cetus*, p. 8, 2013.
- [12] J. Whitmer, F. Kelley, and W. Allen, "Analytics in Progress: Technology Use, Student Characteristics, and Student Achievement," *Educ. Rev. Online*, 2012.
- [13] K. Clark, C. Beer, and D. Jones, "Academic involvement with the LMS: An exploratory study," in *Curriculum, technology & transformation for an unknown futur. Proceedings ascilite Sydney*, 2010, pp. 487–496.
- [14] P. Walsh, "Evaluating the Data Analytic Features of Blackboard Learn 9.1," *Irish J. Acad. Pract.*, vol. 4, no. 1, 2015.
- [15] U.S. Department of Education, Office of Planning, Evaluation, and Policy Development, "Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies," Washington, D.C., 2010.
- [16] M. B. Horn, H. Staker, and C. M. Christensen, *Blended: using disruptive innovation to improve schools*. Jossey-Bass, 2014.
- [17] R. J. Windle, D. McCormick, J. Dandrea, and H. Wharrad, "The characteristics of reusable learning objects that enhance learning: A case-study in health-science education," *Br. J. Educ. Technol.*, vol. 42, no. 5, pp. 811–823, Sep. 2011.
- [18] J. Hewitt and C. Brett, "The relationship between class size and online activity patterns in asynchronous computer conferencing environments," *Comput. Educ.*, vol. 49, no. 4, pp. 1258–1271, 2007.
- [19] C. Vrasidas and M. S. McIsaac, "Factors influencing interaction in an online course," *Am. J. Distance Educ.*, vol. 13, no. 3, pp. 22–36, Jan. 1999.
- [20] J. Kuzilek, M. Hlosta, D. Herrmannova, Z. Zdrahal, and A. Wolff, "AK15 Case Study 1: OU Analyse: Analysing At-Risk Students at The Open University, Learning Analytics Review, no. LAK15-1," 2015.