

RESEARCH ARTICLE

Open Access



# Impact of using learning analytics in asynchronous online discussions in higher education

Juan Pedro Cerro Martínez<sup>1\*</sup> , Montse Guitert Catasús<sup>2</sup>  and Teresa Romeu Fontanillas<sup>2</sup> 

\* Correspondence: [jcerrom@uoc.edu](mailto:jcerrom@uoc.edu)

<sup>1</sup>Doctoral Programme of Education and ICT (e-Learning), Open University of Catalonia, Barcelona, Spain

Full list of author information is available at the end of the article

## Abstract

Following asynchronous online discussion activities as a complex communication process is a demanding task for teachers. In this paper, the authors have explored the potential in supporting such activity through learning analytics. From the beginning, the authors acknowledged the limitations of technology to support the complexities of a pedagogical activity. Therefore, the methodology used was participatory design-based research (DBR) divided into two main stages. The first design phase dealt with the engagement of teachers and pedagogical experts in defining the data and metrics to be used to support the pedagogical concepts. The second consisted of an implementation phase including pilots with students and with crucial engagement of teachers in commenting their understanding over students' learning processes and the feedback the teachers could offer to them. Overall, the students shown improvements in their performance as monitored through the learning analytics group in contrast with control groups. The discussion over the design and its results could be potentially extrapolated to other educational contexts.

**Keywords:** Collaborative learning, Learning analytics, Online discussions, Higher education, Online learning

## Introduction

The rise of big data has produced a new “hyperbolic” phenomenon related to the datafication of society (Kitchin, 2014). Big data has been presented with great enthusiasm as the new engine of an intensive knowledge economy, in which data mining techniques and artificial intelligence mechanisms are used to generate automated processes tailored to the user whose data is being plotted (Kitchin, 2014). The use of this data has recently come into question with the rise of issues such as the inappropriate, unequal or unethical use of personal data. The datafication phenomenon has also led to emergent practices in higher education and the consequent need to rethink the skills required of academics in dealing with this datafication (Raffaghelli, 2018; Williamson, 2018).

Online learning has also been transformed in recent decades. These changes have impacted students in one way or another depending on how ICTs have been

incorporated into online teaching and how the learning environments have been used by teachers. These teaching and learning environments produce large amounts of data, which are not only generated by students but also by the technological systems themselves, in the form of things such as metadata. This is where the real challenge arises.

In higher education, data mining techniques have spurred on powerful movements, among which it is particularly important to highlight learning analytics (Buckingham & Deakin, 2016; Daniel, 2015; Ferguson et al., 2016). Learning analytics is a tool that can offer us information about interaction processes between students (Caballé & Clarisó, 2016; Gañán, Caballé, Clarisó, Conesa, & Bañeres, 2017). As defined by Siemens and Gasevic (2011): “Learning analytics are the measurement, collection, analysis and reporting of data about students and their contexts, in order to understand and optimize learning and environments in which they occur” (p. 8).

The main opportunities provided by learning analytics as a discipline involve revealing and contextualizing the information so far hidden in educational data and preparing it for the different stakeholders (Greller & Drachsler, 2012). However, analytics has two fundamental objectives: reflection on the evidence of learning (descriptive side) and future prediction based on the data and patterns detected (predictive side). In this paper we will focus on the first one through descriptive reflection.

There are several challenging factors that favor the use of learning analytics (Ferguson, 2012). The first challenge is a technical one and is related to the processes by which we extract information from a large volume of data related to the student. The second challenge is pedagogical, looking for ways to optimize the opportunities offered by online learning. Finally, the third challenge is political/economic, addressing the way to optimize educational results at national and international levels.

This paper details a research project carried out at the Universitat Oberta de Catalunya (Open University of Catalonia, UOC) to design and implement a learning analytics solution for teachers. The aim was to facilitate their access to information for monitoring and assessing asynchronous online discussions, a type of collaborative learning activity in which students interact with each other to jointly construct meanings using dialogue and reflection. In the past, learning analytics has been widely used to evaluate collaborative learning by measuring a cluster of very simple learning processes. Given the complexity of assessing collaborative learning and teachers' information needs in online environments, this research proposes a list of key factors, defined by the teachers themselves, related to the process of communicative interaction and which the designed tool measures. This is a learning analytics-based tool which is useful for assessing the challenges related to students learning needs and how we can personalize this learning by delivering “actionable feedback” (Rienties & Jones, 2019). Allegedly, these types of tools might encompass conflicting effects like behavioral regulation instead of conscious and self-guided learning, as Archer and Prinsloo (2019). In this regard, the authors embrace a third position where the technology can be built over collective reflection and design aiming to the final users appropriation. Technology is conceived as a complex human activity and its definition is participatory and oriented towards the work objectives of the actors involved.

Previous studies to identify the key factors to be considered for the correct monitoring and assessment of collaborative online activities (Cerro, Guitert, & Romeu, 2016) were used as the basis when setting the objective of building a learning analytics tool.

The tool was then made available to teachers to allow them to obtain information on their students' learning processes when interacting in the online discussion spaces. Any teacher using this tool for the first time should be aware of its complexity, and therefore should carefully observe the levels of applicability to the collaborative situation itself. This adapting to and understanding of the data system ultimately allows teachers to determine the teaching function's value in understanding students' performance so that they can offer feedback that is more tailored to the specific work carried out by each individual (Jordan, 2012).

## Methodology

The research carried out had two main objectives, the first of which was to design a learning analytics tool to analyze students' communicative interaction online. The second objective was to have teachers implement the tool to monitor and assess asynchronous online discussions and in turn measure the impact of the tool's use on university students. For this reason, the following research question was proposed: *What impact does the use of learning analytics by teachers for monitoring and assessment have on student and classroom performance within the online higher education context?*

This research question is rather controversial because of the criticisms received by some researchers about whether it is really possible to alter the behavior of students through the use of learning analytics (Ferguson & Clow, 2017). In fact, other studies have shown how the use of analytical tools does not have a significant impact on student learning (Park & Jo, 2015) but does impact on the level of student understanding, their learning process and the change in student behavior that occurs. This is the main reason why this paper presents the impact of the use of a learning analytics tool designed specifically for the UOC campus to monitor and assess asynchronous online discussions held at the university.

Overall, to answer the research question, a design-based research (DBR) methodology was used. Reeves (2007) described three fundamental principles of this research framework which justify the choice of this methodology in our case. These principles can be summarized as follows: the DBR pursues the collaborative search for solutions to complex problems in their real contexts that integrate design principles with technological advances within the framework of a reflexive investigation where innovative learning is carried out iteratively. Amiel and Reeves (2008) pointed out that their ultimate goal was to establish a close connection between educational research and reality following an iterative process that not only evaluates an outcome, but also systematically and simultaneously refines an innovation with the design principles that guide the process.

DBR is used to study learning problems in their natural contexts in order to introduce improvements in the learning process itself (McKenney & Reeves, 2013). These improvements, in the form of educational innovations, are not only at the pedagogical level but also extend to the products that serve as the central axis of these innovations. Such is the case with computer applications that help foster a better understanding of the nature and conditions of the learning taking place.

It is at this point that the research responds to the main objective related to the measurement of the impact of teachers' use of learning analytics on online students. After an initial phase of designing where the pedagogical problem was explored in

search of the metrics and the data supporting it, the tool was proposed to the students. Two classrooms were engaged. The first classroom had an intervention from the teacher with information on student activity through learning analytics (experimental classrooms) and the second classroom had conventional teaching style in which no analytical data on student performance were used (control classrooms). All else was held constant for both groups: the same subject of ICT Competencies was presented to the students through the same pedagogical model, which in time encompassed the same activities resources and evaluation model. Within the pedagogical approach, the student was expected to have an active role. The statistical inference performed hence feasible over the data collected was considered a final step to understand the impact of the process.

The two main phases and a final stage for the conclusions distributed under the methodology of the DBR, is shown in Fig. 1.

### **Phase 1: tool design**

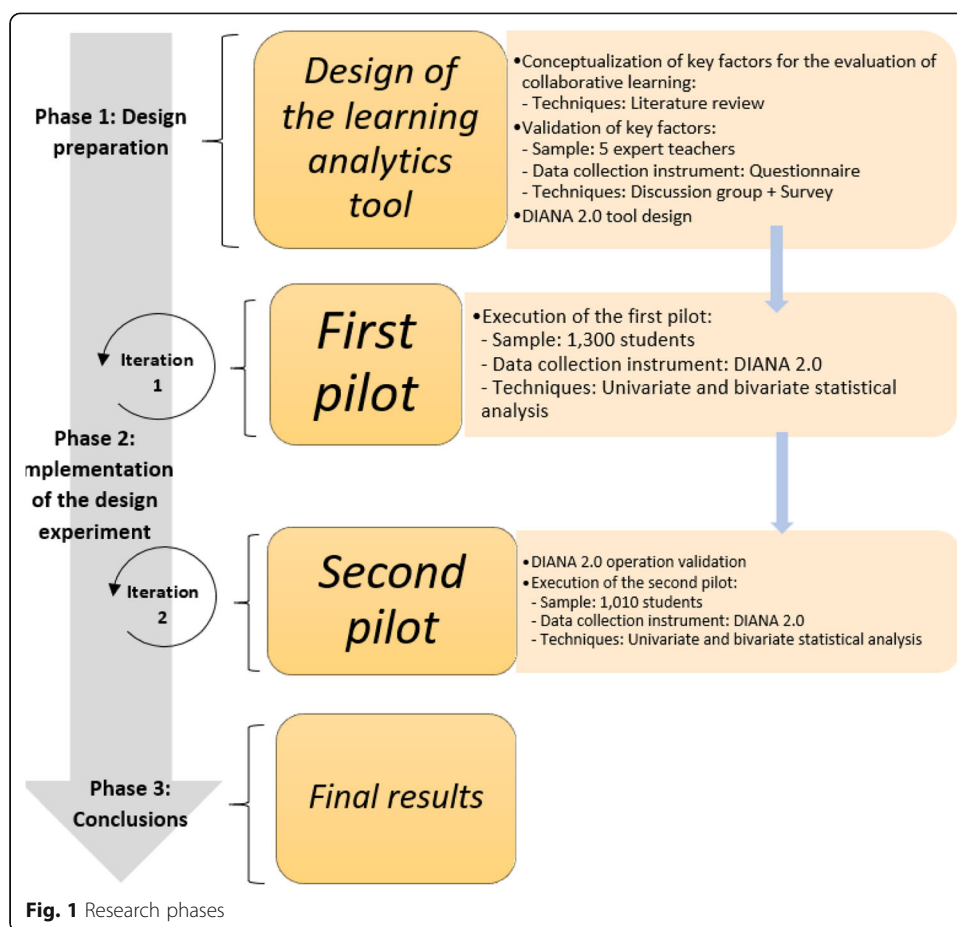
In the first phase of the investigation, the design of the learning analytics tool, a systematized bibliographic review was conducted to find the key factors that favor the evaluation of collaborative learning online. Specifically, these factors were described through a hierarchical model that helped interpret and classify the information related to the various aspects involved in the collaborative learning process so it could be properly evaluated. To define the key factors' model, rubrics related to the evaluation of collaborative activities were analyzed, comparing them with Salmon's (2012) reference framework for structuring an online learning process and with the different dimensions involved in the evaluation of collaborative learning (Iborra & Izquierdo, 2010). This model organized the key factors in three levels: categories, indicators and metrics.

It should be noted that the model needed to be validated against the vision of expert teachers in online collaborative learning methodology, to verify if the key factors' definition and applicability was accurate and in line with teaching experience in this field. This validation and co-definition process was carried out through a discussion group comprising a sample of 5 teaching experts in collaborative learning methodology. A questionnaire was used to collect the experts' proposals regarding the key factors. The answers were then analyzed to adapt the initial proposal and create a second model based on teachers' contributions that would serve as the basis for constructing the learning analytics prototype, DIANA 2.0 (Dialogue ANalysis version 2.0), to be used in the pilots organized in the later phase.

### **Phase 2: experimental pilots**

In the second phase of the design experiment implementation, the same DIANA 2.0 prototype learning analytics tool was used. It was made available for a period of 1 year (2 university semesters) to teachers giving a course called "ICT Competencies". This course is common to all the UOC's degree programs, although it is adapted to the singularities of each one.

Each pilot corresponded to one of the iterations of the DBR. The data for each one were obtained independently of each other. However, the results showed the same trends, and therefore both pilots were considered as a single experiment to increase the



amount of data, lending reliability and solidity to the analysis' conclusions. A survey of the students in the experimental classrooms was also carried during the pilots to ascertain their level of satisfaction with the feedback received on their performance. However, these results could not be compared with those of the control classrooms, since this group had no available data on this matter.

The procedure for analyzing and interpreting the pilots' results was organized based on the following data collection instruments: the DIANA 2.0 tool and the student satisfaction questionnaire. The analysis of the results required the use of mainly quantitative techniques, which are detailed below for each instrument used.

DIANA 2.0 was used as a data collection instrument to measure the impact generated by the use of learning analytics to monitor and assess online discussions. In this case, the following data were taken into account:

- Grades obtained by students in the ICT Competencies course during the current semester.
- Computer files downloaded from the UOC's virtual campus with all messages exchanged by students during the online discussion activity.
- Configuration data for the parameters of the analysis performed by the teachers in each of the classrooms. This information included: start and end date of the online discussion activity, list of keywords used and configuration values of all the

variables defined for the analysis in DIANA 2.0 (minimum and maximum number of messages, maximum degree of dispersion of the online discussion, etc.).

All this information was entered into DIANA 2.0 to obtain the results of each of the metrics implemented in the learning analytics tool. With these metrics, a univariate analysis was performed using descriptive statistics techniques to summarize the calculated values. Secondly, a bivariate analysis was performed by calculating correlations between the results of each pair of metrics in order to discover interpretable relationships between the observed phenomenon. Finally, because part of the analyzed metrics was expressed with heterogeneous values, we proceeded to cross-reference these results using cross-data tables to identify trends in the data.

Finally, in the student satisfaction questionnaire, a univariate analysis was also carried out through the use of descriptive statistics techniques to summarize the values of the questions posed to the students.

During the pilots' development, teachers received specific training on learning analytics through a teaching guide on the academic use of DIANA 2.0 in the classroom, which improved their understanding of the student learning process. Some of the classrooms' teachers used DIANA 2.0, allowing their students to be influenced by the personalized feedback that the teacher gave them based on the metrics that DIANA 2.0 reported on the online discussion activity. These classrooms were considered experimental classrooms. The classrooms whose teachers did not use DIANA 2.0 to monitor and evaluate the online discussion activity were considered control classrooms in order to compare the results with the rest of the online classrooms.

Likewise, it should be noted that prior to using the DIANA 2.0 tool and the metrics reported, teachers had to obtain part of this information manually, which took much more time. Teachers were able to dedicate the time saved to qualitatively superior tasks and to offering information to students about certain indicators linked to their interactions. These indicators allowed teachers to better understand the student learning process.

The pilots involved a sample of 40 classrooms (22 of them were experimental classrooms and the remaining 18, control classrooms) and a total of 2310 students. Table 1 gives a breakdown of the data collected.

### **Phase 3: conclusions**

The last phase of the research involved analyzing the results in order to draw conclusions that allowed us to answer the research question. The conclusions, however, had to take certain critical aspects into consideration related to the ethical use of student data and its manipulation by the teacher, as detailed below.

Informed consent was obtained from all research participants, which mainly included the collaborative learning methodology experts in the discussion group and the teachers involved in the pilots. Prior to the study, the research objectives and how the information collected would be used were explained to the participating teachers.

One of the research's main ethical aspects was to guarantee the security and confidentiality of the data. Both of these criteria were continuously followed by observing the following:



**Table 1** Sample selected for the pilots

Sample selected for the two pilots	Total
<b>Experimental classrooms</b>	<b>22</b>
<b>Students in experimental classrooms</b>	<b>1.367</b>
<b>Messages analyzed by DIANA 2.0</b>	<b>4.029</b>
<b>Control classrooms</b>	<b>18</b>
<b>Students in control classrooms</b>	<b>943</b>

- The confidentiality of the data was guaranteed through the teachers, whose contract with the UOC contains clauses related to the processing of student information. In this contract, which covers the teaching of the ICT Competencies course, the teacher is considered to be a “data processor”, in accordance with the provisions of Article 4.8 of the General Data Protection Regulation (GDPR), which ensures that the data is not used for non-academic purposes.
- Moreover, data security was guaranteed because each of the classrooms in which the DIANA 2.0 learning analytics tool was implemented had a copy of the data (called “instance”); thus, the data hosted in a classroom tool was not visible to the rest and vice versa.

Merriam (1998), adopting a transversal research perspective that goes beyond design, linked research quality with several specific strategies: internal validity, external validity and reliability.

- **Internal validity:** Internal validity is based on the adaptation of the results obtained in the research to the observed reality. In our study, internal validity was achieved using several strategies. The first of these was triangulation, which was not only carried out through the different instruments used but also through access to different types of data and participating agents. Direct observation of the data was also used. This implies that the information collected on students’ activity was not only managed by the teachers in order to redirect the teaching process, but was also developed by the researchers. Student activity analysis required direct access to their messages and their subsequent analysis through the learning analytics tool. Furthermore, a post hoc review of the conclusions was carried out based on the interviews with the collaborative learning methodology experts. This review was carried out by actively listening to the discussion groups’ audio recordings.
- **External validity:** External validity measures the representativeness of the results obtained and the conclusions’ application to other situations. External validity is justified through the changes introduced in the online educational environment of the UOC, since, as mentioned above, the learning analytics tool was introduced in the university classrooms as an available support resource for the teachers in the courses’ virtual spaces. In addition, the application of conclusions to new environments was reflected by comparing and contrasting certain trends observed in the first pilots with respect to other groups of students under the same conditions, but from different academic semesters and degrees.
- **Reliability:** Reliability refers to the way in which the results could be replicated (Pérez-Mateo, Romero, & Romeu-Fontanillas, 2014). This is ensured through

criteria such as data security and confidentiality, criteria which are part of the ethical aspects of the research already mentioned. The reliability of the data was ensured through rigorous monitoring of the research process. An important fact that validates the reliability of the data is found in the learning analytics tool itself, since the information that the tool reported was based on information that the students generated during their activity. This makes it impossible to have falsified the data, since the same source of information is linked directly to the messages exchanged in the communication spaces. The reliability of the data is guaranteed, again, by the triangulating instruments and data, not only by comparing the results from different sources but also by using various instruments for collecting information to draw conclusions. We also ensured reliability by reviewing the analysis process for each phase of the research.

## Results

### First DBR loop

The first loop of the DBR process involved improving the initially proposed tool (through the iterative process of review and improvement). As a result of the first phase of the research, two clearly differentiated products were developed, but they were totally complementary: the list of key factors to be implemented and the learning analytics tool used to gather those key factors.

Initially, a first proposal of key factors was made in the form of indicators and metrics that served as a generic reference for the evaluation of collaborative learning (Cerro, Guitert, & Romeu, 2018). Subsequently, due to the contributions of the experts in the discussion group and the technical limitations of an online environment, the indicators and metrics related only to communicative interaction were taken into account for the tool's design, as shown in Table 2.

DIANA 2.0 was then developed. This learning analytics computer application was based on web technologies developed by the UOC, the objective of which was to provide teachers with information on the analysis of the communicative interactions between students when exchanging messages within the university's online spaces. The application became not only a product of the research but also a data collection instrument for drawing conclusions, which led to the second loop of the DBR.

### Second DBR loop

The results obtained in the second phase of the research not only showed student performance, based on the information available on teachers' use of learning analytics, in the experimental versus the control groups, but also the degree of student satisfaction based on the feedback received from the teacher, which was in turn based on information gathered through the learning analytics tool.

To analyze student performance, the results obtained in the two pilots were grouped together based on the classroom dimension, and an assessment was made of whether teachers' use of analytical tools, along with the deployment of strategies to acquire skills based on the feedback sent, helped to reduce the dropout rate in the experimental classrooms compared to the control classrooms. The experimental and control classrooms were compared based on the statistical results, as shown in Fig. 2.

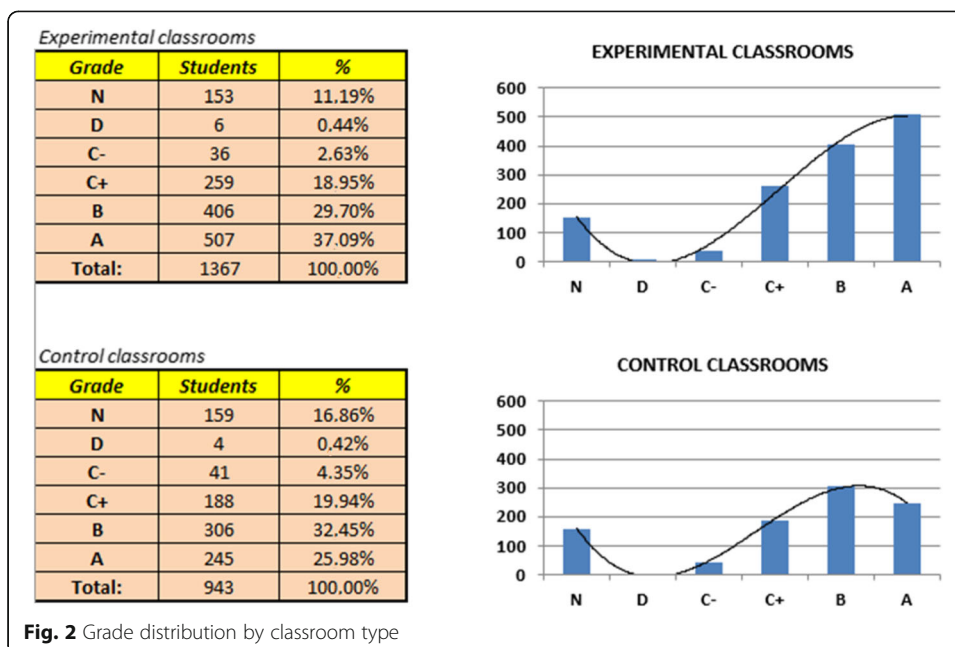


**Table 2** List of the 21 metrics implemented in Diana 2.0

Category	Indicator	Metrics implemented in DIANA 2.0	Metric type	
			Individual	Group
<b>Communicative interaction</b>	Consistency and regularity in group interaction	Temporary distribution of individual messages	<b>X</b>	
		Temporary and group distribution of messages		<b>X</b>
	Participation in communicative interaction	Number of messages posted	<b>X</b>	
		Level of participation in communicative interaction	<b>X</b>	
		Total number of messages posted		<b>X</b>
		Number of users participating in the communicative interaction		<b>X</b>
	Promotion of dialogue and negotiation	Homogeneity in group communicative participation		<b>X</b>
		Answers posted	<b>X</b>	
		Popularity	<b>X</b>	
		Total responses posted		<b>X</b>
	Communicative style and language used	Dialogue level		<b>X</b>
		Average extension in communication	<b>X</b>	
		Global average extension in communication		<b>X</b>
		Tag cloud of conversation spaces		<b>X</b>
	Communication type	Degree of adaptation of the discourse to the semantic field		<b>X</b>
Dispersion / concentration of conversations			<b>X</b>	
Visual representation of asynchronous communicative interaction (graph of nodes - SNA)			<b>X</b>	
<b>Information management and exchange</b>	Exchange of information within the group	Number of files published /attached	<b>X</b>	
		Number of links published / attached	<b>X</b>	
		Number of global files published / attached		<b>X</b>
		Number of global links published / attached		<b>X</b>

Taking into consideration the groups’ distribution according to type of classroom (Fig. 3), significant differences are observed in the variables studied. Indeed, the experimental classrooms experienced a 5.67% reduction in the number of students who did not complete the course (N) compared to the control classrooms. Moreover, 11.11% more students in the experimental classrooms than in the control classrooms who engaged in the online discussion activity obtained the maximum grade (A).

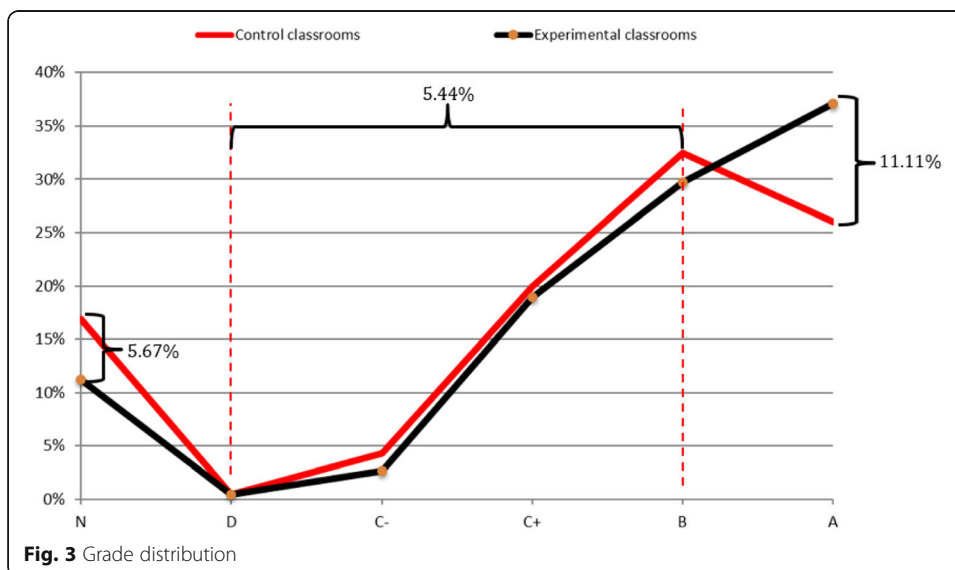
The trend shown in the data shows that, on the one hand, the use of learning analytics reduced the rate of students who dropped out of the activity by 5.67% and, secondly, that the feedback offered to the students based on the information from the learning analytics enhanced student performance. However, to know exactly to what degree student s’ performance increased thanks to the use of learning analytics, the normal distribution (Fig. 4) of the grades of both types of classrooms were calculated. The distribution reveals that the average grade of the students in the experimental classrooms increased by 0.71 points and the standard deviation decreased by 0.32 points. The results thus show that the experimental classrooms increased their grade average

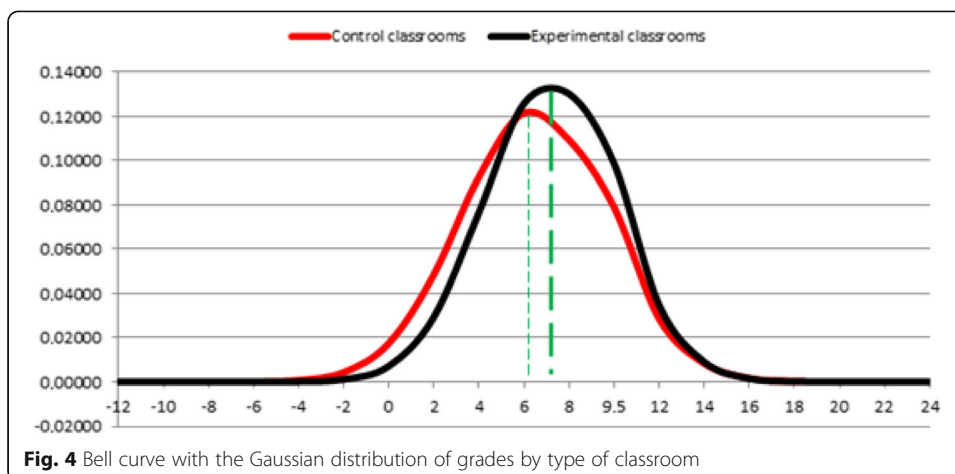


by almost one point while also homogenizing student performance, given that the overall grades were closer to the average than in the case of the control classrooms, in which the learning analytics tool was not used to monitor and evaluate the online discussion activity.

Apart from these results, the interactions produced in the online discussions were also analyzed through the metrics reported by DIANA 2.0. In this respect, we were given an in-depth look into aspects that were not observable at first glance:

- If the activity score is compared with the level of participation, a correlation of 68% is obtained, which indicates that those students who interacted more in the online





discussion activity were more likely to perform better than those with a lesser degree of participation.

- The students who scored best in the online discussion were those who not only exchanged a higher number of messages, and with more extensive arguments, but also those who generated the most impact within the conversation through the number of responses received (level of popularity), in comparison with students with a lower grade.

Finally, it was possible to verify the impact of the use of learning analytics on student satisfaction. The results reveal a high degree of satisfaction on the part of the students, since 88% stated they were satisfied or very satisfied. A high level of agreement was also detected between the grade received from the activity and the student's perception of his / her performance, since 87% of students declared that they agreed or totally agreed with their grade. However, this fact must be qualified since, as in any other quasi-experiment (embedded into the DBR), satisfaction data from the students in the control classrooms was not available, which makes it impossible to know definitely if the high degree of satisfaction detected was attributable to the use of analytics or other factors.

## Discussion

The results obtained in this research heterogeneously compare with those obtained in other studies carried out, differences which we will discuss in the current section.

In online teaching environments, student activity generates a huge amount of information that is disseminated on the platforms within which the activity takes place. Based on the results and the teachers' work, it is important to note that teachers' must understand what happens in the learning activity if they are to assess it, both at the individual as well as the group level. The tools used by researchers to analyze online discourse are inadequate (Law, Yuen, Huang, Li, & Pan, 2007), mainly because these tools have difficulty in managing different information formats, the quantitative indicators insufficiently measure the quality of learning, and participation indicators and content analysis are handled using different tools. In contrast, this research tackles these issues from a different point of view; DIANA 2.0 was integrated into the virtual campus, giving teachers the opportunity to use the same data as that by the LMS. We also

provided teachers with a variety of heterogeneous indicators and metrics, represented not only in text mode but also visually (visual learning analytics): bar charts, tag clouds, gradient meters, etc. Likewise, DIANA 2.0 could export the analyses generated in XML format, making it possible to share information between analytic tools, and combine both interaction metrics and low-level content analysis metrics.

Experts in online learning in higher education predict that learning analytics will be used not only to identify students' behavior patterns but also to improve their learning and retention rates (Avella, Kebritchi, Nunn, & Kanai, 2016). The current study's results support these two last statements, even though the improvements identified were not so extensive. In this vein, Viberg, Hatakka, Bälter, and Mavroudi (2018) classified learning analytics research in higher education in terms of evidence for learning and teaching, pointing out that more than a half of the studies analyzed showed clear evidence of an improvement in learning outcomes and in support for teaching. The results of our contribution back up the meta-analysis carried out by these authors.

During the research's iterative design process, teachers expressed having difficulty in interpreting information. Such information tended to be heterogeneous between variables or difficult to evaluate in comparison with the reference values. This opinion has been reflected in other studies recognizing the limitations to teachers' ability to make quick decisions due to a lack of real-time data analysis and a delay in accessing critical information (Gkontzidis, Kotsiantis, Panagiotakopoulos, & Verykios, 2019). These needs have also been identified by other authors (Mor, Ferguson, & Wasson, 2015), who have stated that the assessment of students' performance is a tiresome and time-consuming process for teachers. This is why previous training for teachers is necessary to help them interpret the information that learning analytics report on student activity. In order to minimize the level of difficulty, the DIANA 2.0 tool's design included different data visualization models, from the most basic, based on icons, to the most complex, based on graphs of nodes in the style of social network analysis. Some of the information produced by learning analytics was used by teachers to understand the learning process carried out by their students, which the teachers then used to provide and improve the feedback sent to students, complemented by other qualitative information.

Another of the relevant elements of this study was that it addressed the way in which teachers access, process and interpret information related to online educational practices. Teachers have to face the challenge of understanding complex phenomena related to learning in educational environments. This is especially important when a multitude of variables and contexts intervene, not so much on how to collect the information they need but rather how they analyze that information to obtain judgments of value that assure correct decision-making. It is here where teachers play a fundamental and essential role, since they decide the action to be undertaken based on the interpretation of the data, no matter how well they are represented. Tió, Estrada, González, and Rodríguez (2011) considered the contribution of the teacher's role to expanding the student's zone of proximal development. In this sense, Gkontzidis et al. (2019) assessed student performance using learning analytics and they concluded, following the trends shown in our research, that the use of data during the teaching process can inform teachers about students at risk, but the authors recognized that advanced prediction is at an early stage.

This research respected the experimental work protocol in full, obtaining favorable impact results by giving visibility to the importance of collaborative learning in

university work. In fact, the results obtained in this research indicate that the use of specific learning analytics instruments by teachers, configured based on a participation process and with teachers being trained in the use of these instruments, has improved specific training in the communicative interaction of students in asynchronous online discussions. We tried to compare these results with others from different studies which involved the use of learning analytics tools, for example Lotsari, Verykios, Panagiotakopoulos, and Kalles (2014), who found no clear correlation between students' participation and their final grade. From our point of view, this lack of correlation was due to the slow process of extracting data from online discussion and its subsequent analysis using statistical software. The authors themselves recognized this limitation, claiming that real-time analysis, which DIANA 2.0 does, would have enriched the results. Kagklis, Karatrantou, Tantoula, and Panagiotakopoulos (2015) had similar results, and the reason is the same lack of real-time analysis as in the previous case. To support this argument, we found a correlation of 68% between grades and level of participation.

Student feedback, along with other qualitative information, is an important element in the application of analytical technology, not as an effective element in itself, but as a mediating tool for what the teacher wants to promote in the pedagogical process and what university students can obtain. In this case, other researchers (Park & Jo, 2015) stated that despite the absence of a significant impact on learning achievement in their study, the pilots organized in their investigation evidenced that learning analytics tools impacted not only on the degree of understanding but also on the students' perceived change of behavior. Gkontzidis et al. (2019) likewise articulated that we have to consider a diversity of indicators, not solely one of them, for predicting students' future achievements and for improving educational outcomes (Avella et al., 2016). Such tasks have to be carried out by teachers, a conclusion which we have also reached in our research.

The observability of the research revealed, through analytics and their visual systems, a slight improvement in individual performance through an improvement in grades and, in general terms, reduced the dropout rate. However, this result would not have been valued by the participants without an understanding of the importance of collaborative learning. In the field of visual learning, other case studies (León, Cobos, Dickens, White, & Davis, 2016) have reported on the advantages of the availability of analytics for teachers and students and, in terms of usability, that teachers consider this information very useful in real-time. This perception was shared by the teachers involved in this papers' research; they felt comfortable using the learning analytics too and were satisfied by its interface. Other works (Tió et al., 2011) have demonstrated similar results, and, going into the topic of student satisfaction in greater depth, Park and Jo (2015) found that satisfaction with using learning analytics dashboards (an example of an analytical tool) correlated with the degree of understanding and student change of behavior.

It may seem obvious that those students who posted the most messages in the online discussions also had a higher average of written words, but what is remarkable is that the "popularity" metric (responses received from the other students) also rose. That is, the average number of words and the total messages are metrics that depend on the individual students themselves, since they directly cause them with their actions, but popularity is based on the number of responses that their messages receive, something they do not control. In other words, it is not an action that a student fosters him- or herself but one that is fostered by the rest of the participants. One possible explanation is that a student's

messages had a significant impact on the group, either due to the quality of the interventions, the notoriety or some other factor that generated a high level of interest and many responses. The teacher's ability to visualize this popularity and give feedback to the student (in the form of a teacher's comment) reinforced the teacher's understanding of the importance of peer interactions in a collaborative process.

## Conclusions

In answering the research question, we first had to define the indicators for measuring student development in online discussions and then develop a learning analytics tool (DIANA 2.0) that would be integrated in the virtual campus of a higher education institution. Finally, we ran pilots, using the tool for gathering information. This, coupled with the literature on the subject and contributions from collaborative online teaching experts, allowed us to identify some of the key factors to be considered when assessing collaborative interactions. However, despite the favorable results obtained when using the information provided by learning analytics to monitor asynchronous online discussions, the results from the pilots must be assessed critically. Although our initial assumption is that the key factors model that was used to apply learning analytics in the institutional case of the UOC could later be transferred to other contexts in the same way, only future research and practices could support or refute this hypothesis. Furthermore, some researchers may criticize some of the key factors and their relevance in generating a real impact on student performance. In this case, however, it must be remembered that the learning analytics data reported by DIANA 2.0 was interpreted exclusively by the teachers monitoring and assessing the learning activity. Therefore, it is not necessary to understand each metric as an isolated and decontextualized value, but as a global interpretation of a set of metrics. These metrics are what allow qualitative elements that help the teacher assess the student's activity to be obtained.

From an empirical point of view, the aim of our research was to design and implement a software tool in the form of a web technology-based computer application. The resulting tool was intended to allow teachers to obtain information on the key factor most prevalent in online collaboration, namely communicative interaction. However, from its inception this tool had a recognized handicap, which was that capturing human interaction of this type is difficult given that at each step it is the stakeholder who conceptually defines the elements they consider to be most effective in communicative interaction within the pedagogical relationship. This is why the experimental pilots were organized to analyze asynchronous online discussions using an analytical tool that would offer information regarding this key factor above all others, as discussed with teachers. As such, the favorable results obtained could be altered if the learning activity in question were one whose central axis was not communicative interaction but another competence for which DIANA 2.0 had not been designed and from which we haven't evidence of learning. The authors commit with the fact that "evidence" is an elusive term somehow and that the data hereby presented might completely change under other contexts and taking into account other educational values (Biesta, 2010). It is to be highlighted that all teachers and teaching coordinators gave particular value to collaborative learning and the asynchronous online communication processes supporting it. This element highlights the non-universality of technological design and the importance of the participation of students and teachers.



It is tempting to think that a greater availability of data on student performance will help us to better predict their future behavior. However, there are some limitations regarding this. One involves the inefficient amount of data on students, such as age, sex, marital status, etc., whose null utility in understanding certain phenomena has been demonstrated (Gullion, 2018). In our research, researchers and teachers collaborated to co-define key elements for assessing collaborative learning, a practice that is in line with other studies (Tió et al., 2011) and which minimizes the inconvenience that comes with using data that is not productive. By using a participatory design process to connect the type of data extracted with the visualizations, we were able to design a tool that produces and supports a good synthesis of collaboration processes. However, our research is not directly transferable, since it was contextualized in a specific situation in which teachers had specific pedagogical needs and advanced technological skills in relation to the disciplinary area.

A critical analysis of the results obtained in the experimental pilots justifies and contextualizes the impact that learning analytics' use had on the students. This critical perspective is particularly important when considering that in this research both the experimental and control classrooms were related to the same curricular conditions, pedagogical model, teaching strategy and instructional design, among others. Thus, the reduction in the dropout rate closely related to the improvement in student performance. A possible argument for this trend is that the teachers who used the learning analytics tool were teachers who were highly committed to using ICT resources to support their teaching. In this regard, practices involving analytics resulted in teachers' increased presence in classrooms and greater efficiency in monitoring students, which impacted positively on the feedback sent to students, something that did not occur in the intervention carried out in the control classrooms. However, in other scenarios, having no change in feedback quality could be due to other factors such as lack of time or a lack of training in the use of the analytical tool. Although the impact of using learning analytics can be clearly understood when looked at from the point of view of reducing the dropout rate, it is perhaps less clear from that of improving student performance, mainly because there was an increase of 0.71 points in the experimental classrooms' average grade as well as a homogenization of the grades, reducing grade dispersion by 0.32 points. These last results cannot be questioned in terms of teachers' motivation in the experimental classrooms, because all the teachers had been trained in the same pedagogical strategies and course procedures. On the contrary, it would be feasible to consider an explanation more focused on the nature of the activity being analyzed, since in asynchronous online discussions, the communicative interaction competence is the main thing to be assessed, and thus the key factor analyzed by learning analytics in this research.

Once the pilots were held with the students, we tried to answer the research question offering a two-axis solution, as identified previously. First of all, and considering the classroom as the object studied (group level), the reduction in drop out rate was significant enough to be highlighted, as well as the increase in the overall grades. These were the main impacts produced in all groups of students monitored using DIANA 2.0. The second axis of the impact is related to student level. In this regard, the research has increased our understanding of students' behavior as a consequence of being monitored using a learning analytics tool. We noticed a significant correlation along the same lines

between grades and participation, and grades and quality in communicative interactions, measuring the latter using a mix of metrics such as number of messages, popularity of the student in the online discussion, extension of the argumentations, and others quantitative metrics that can be combined by teachers for obtaining qualitative information about student development.

One of this study's contributions to the field of online teaching is that it has demonstrated the complexity of building a data system to report on collaborative learning, which is a complex pedagogical problem in itself. In this sense, and likewise related to the research question, the main impact of using learning analytics for teaching is its transformative potential. The results indicate that it allows teachers to find new ways of developing their pedagogical dynamic and to carry out a monitoring and assessment process that is more personalized and better adapted to facilitating students' learning.

Below we provide a detailed list of future lines of research that can be addressed to increase our knowledge in the discipline of learning analytics beyond the limits of the contribution made by this research:

- The first proposal lies in researching how the use of learning analytics impacts the monitoring and evaluation of students from a broader perspective than that of communicative interaction. This would involve other key factors whose source of information allows indicators and metrics belonging to categories such as information management and exchange, planning and organization and so on to be implemented.
- Another important aspect is to research the effect of making the learning analytics available to students when they are involved in collaborative learning. It would also be interesting to research how students self-regulate during this process and what strategies they deploy to successfully achieve their objectives. From a critical perspective and according to Ferguson and Clow (2017), the direct relationship between the use of learning analytics and the change in student behavior has not yet been demonstrated. However, through this research we have tried to obtain information that offers answers that help to better understand that relationship. In this search for answers, students will most likely have to be placed once again at the center of the assessment process for which learning analytics is being used (Broughan & Prinsloo, 2019). This centralization process can be carried out in two different ways: by making learning analytics accessible to the students during the course, and by allowing them to participate in the definition of the criteria to follow in the learning analysis and offering them a critical vision of data use.
- Learning collaboratively in the online environment in the educational field is an analogy of what collaborative online work represents in the professional field, in which a coordinator leads a work team to achieve certain objectives. Learning analytics could be used to provide information to the coordinators of professional teams from different fields about their workers' performance, the way they interact and how they carry out the tasks entrusted to them. It would be interesting to investigate the transfer of the conclusions of this study to the non-academic professional field, and in what way these analytics favor the monitoring of work teams in complex business organizations.

Finally, future prediction models can measure the degree of probability that something will happen but not "why" something is going to happen, and this is where the

biggest challenge is presented. Learning analytics has great potential in the educational context, but it requires training to know how to interpret the information it provides. The information obtained from learning analytics, by itself, is not enough, but it does help teachers to make decisions and will allow them to devote their time to more qualitative or higher-level evaluation tasks.

#### Acknowledgements

Not applicable.

#### Authors' contributions

The author(s) read and approved the final manuscript.

#### Funding

Not applicable.

#### Availability of data and materials

The source code of the prototype version 2.0 (DIANA 2.0) is published openly in the following online repository:

<https://github.com/jcerrom/DIANA-2.0>

This source code is published under the GPL 3.0 license, whose agreements are available from the following address:

<https://github.com/jcerrom/DIANA-2.0/blob/master/LICENSE>

#### Competing interests

Not applicable.

#### Author details

<sup>1</sup>Doctoral Programme of Education and ICT (e-Learning), Open University of Catalonia, Barcelona, Spain. <sup>2</sup>Department of Psychology and Education, Open University of Catalonia, Barcelona, Spain.

Received: 22 January 2020 Accepted: 19 May 2020

Published online: 07 October 2020

#### References

- Amiel, T., & Reeves, T. C. (2008). Design-based research and educational technology: Rethinking technology and the research agenda. *Educational Technology & Society*, 11(4), 29–40. <https://doi.org/10.1590/S0325-00752011000100012>.
- Archer, E., & Prinsloo, P. (2019). Speaking the unspoken in learning analytics: Troubling the defaults. *Assessment & Evaluation in Higher Education*. <https://doi.org/10.1080/02602938.2019.1694863>.
- Avella, J. T., Kebritchi, M., Nunn, S., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning Journal*, 20(2), 13–29.
- Biesta, G. J. (2010). Why 'what works' still won't work: From evidence-based education to value-based education. *Studies in Philosophy and Education*, 29(5), 491–503.
- Broughan, C., & Prinsloo, P. (2019). (Re) centring students in learning analytics: In conversation with Paulo Freire. *Assessment & Evaluation in Higher Education*. <https://doi.org/10.1080/02602938.2019.1679716>.
- Buckingham, S., & Deakin, R. (2016). Learning analytics for 21st century competencies. *Journal of Learning Analytics*, 3(2), 6–21. <https://doi.org/10.18608/jla.2016.32.2>.
- Caballé, S., & Clarisó, R. (2016). Formative assessment, learning data analytics and gamification: In ICT education. Cambridge: Academic Press. In *Book series "Intelligent data-centric systems"*. Amsterdam: Elsevier. <https://doi.org/10.1016/C2015-0-00087-9>.
- Cerro, J. P., Guitert, M., & Romeu, T. (2016). Uso de las analíticas del aprendizaje en entornos colaborativos en línea. In *Revista del Congrés Internacional de Docència Universitària i Innovació (CIDUI)*, 2016, no. 3 Retrieved from <http://www.cidui.org/revistacidui/index.php/cidui/article/view/861/819>.
- Cerro, J. P., Guitert, M., & Romeu, T. (2018). Impacto del uso de las analíticas del aprendizaje sobre el trabajo colaborativo. In *Revista del Congrés Internacional de Docència Universitària i Innovació (CIDUI)*, N. 4.
- Daniel, B. (2015). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904–920.
- Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304–317 Retrieved from [http://oro.open.ac.uk/36374/1/UTEL40501\\_Ferguson%20Jan%202013.pdf](http://oro.open.ac.uk/36374/1/UTEL40501_Ferguson%20Jan%202013.pdf).
- Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., ... Vuorikari, R. (2016). Research evidence on the use of learning analytics – Implications for education policy. In R. Vuorikari, & J. Castaño Muñoz (Eds.). Joint Research Centre Science for Policy Report; EUR 28294 EN. <https://doi.org/10.2791/955210>.
- Ferguson, R., & Clow, D. (2017). Where is the evidence? A call to action for learning analytics. In *Paper presented at the proceedings of the seventh international learning analytics & knowledge conference, Vancouver, BC, Canada, March 13–17*, (pp. 56–65).
- Gañán, D., Caballé, S., Clarisó, R., Conesa, J., & Bañeres, D. (2017). ICT-FLAG: A web-based e-assessment platform featuring learning analytics and gamification. *International Journal of Web Information Systems*, 13(1), 25–54. <https://doi.org/10.1108/IJWIS-12-2016-0074>.
- Gkontzias, A. F., Kotsiantis, S., Panagiotakopoulos, C. T., & Vergyios, V. S. (2019). A predictive analytics framework as a countermeasure for attrition of students. *Interactive Learning Environments*, 1–16. <https://doi.org/10.1080/10494820.2019.1709209>.

- Greller, W., & Drachler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology and Society*, 15(3), 42–57. Retrieved from: <http://hdl.handle.net/1820/4506>.
- Gullion, J. S. (2018). *Diffraction ethnography. Social sciences and the ontological turn*. New York: Routledge.
- Iborra, A., & Izquierdo, M. (2010). ¿Cómo afrontar la evaluación del aprendizaje colaborativo? Una propuesta valorando el proceso, el contenido y el producto de la actividad grupal. *Revista General de Información y Documentación*, 20, 221–241.
- Jordan, S. (2012). Student engagement with assessment and feedback: Some lessons from short-answer free-text e-assessment questions. *Computers and Education*, 58(2), 818–834. <https://doi.org/10.1016/j.compedu.2011.10.007>.
- Kagklis, V., Karatrantou, A., Tantoula, M., Panagiotakopoulos, C. T., & Verykios, V. S. (2015). A learning analytics methodology for detecting sentiment in student fora: A case study in distance education. *European Journal of Open, Distance and E-learning*, 18(2), 74–94. <https://doi.org/10.1515/eurodl-2015-0014>.
- Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1). <https://doi.org/10.1177/2053951714528481>.
- Law, N., Yuen, J., Huang, R., Li, Y., & Pan, N. (2007). A learnable content & participation analysis toolkit for assessing CSCL learning outcomes and processes. In *Proceedings of the 8th international conference on Computer supported collaborative learning*, (pp. 411–420).
- León, M., Cobos, R., Dickens, K., White, S., & Davis, H. (2016). Visualising the MOOC experience: A dynamic MOOC dashboard built through institutional collaboration. *Research Track*, 461–469.
- Lotsari, E., Verykios, V. S., Panagiotakopoulos, C., & Kalles, D. (2014). A learning analytics methodology for student profiling. In A. Likas, K. Blekas, & D. Kalles (Eds.), *Artificial intelligence: Methods and applications. SETN 2014. Lecture notes in computer science*, (vol. 8445). Cham: Springer.
- McKenney, S., & Reeves, T. C. (2013). Systematic review of design-based research progress: Is a little knowledge a dangerous thing? *Educational Researcher*, 42(2), 97–100.
- Merriam, S. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass.
- Mor, Y., Ferguson, R., & Wasson, B. (2015). Learning design, teacher inquiry into student learning and learning analytics: A call for action. *British Journal of Educational Technology*, 46(2), 221–229.
- Park, Y., & Jo, I. H. (2015). Development of the learning analytics dashboard to support students' learning performance. *Journal of Universal Computer Science*, 21(1), 110.
- Pérez-Mateo, M., Romero, M., & Romeu-Fontanillas, T. (2014). La construcción colaborativa de proyectos como metodología para adquirir competencias digitales. *Comunicar*, 21(42), 15–24. <https://doi.org/10.3916/c42-2014-01>.
- Raffaghelli, J. E. (2018). Educators' data literacy supporting critical perspectives in the context of a "datafied" education. In M. Ranieri, L. Menichetti, & M. Kashny-Borges (Eds.), *Teacher education & training on ict between Europe and Latin America*, (pp. 91–109). Roma: Aracné. <https://doi.org/10.4399/97888255210238>.
- Reeves, T. C. (2007). Design research from a technology perspective. *Educational Design Research*, 52–66. New York: Routledge.
- Rienties, B., & Jones, A. (2019). Evidence-based learning: Futures. In *Educational visions: The lessons from 40 years of innovation*, (pp. 109–125). Ubiquity Press. <https://doi.org/10.5334/bcg.g>.
- Salmon, G. (2012). *E-moderating: The key to online teaching and learning*. New York: Routledge.
- Siemens, G., & Gasevic, D. (2011). Learning analytics: A foundation for informed change in higher education. *Educational Technology & Society*, 15(3), 1–2.
- Tiό, L., Estrada, V., González, W., & Rodríguez, R. (2011). Instrumento y herramienta informática para guiar, controlar y evaluar las interacciones de los estudiantes en foros virtuales. *Educación Médica Superior*, 25(2), 59–96.
- Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, 89, 98–110. <https://doi.org/10.1016/j.chb.2018.07.027>.
- Williamson, B. (2018). The hidden architecture of higher education: building a big data infrastructure for the 'smarter university'. *International Journal of Educational Technology in Higher Education*, 15(1), 1–26. <https://doi.org/10.1186/s41239-018-0094-1>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)