Computers in Human Behavior 138 (2023) 107444

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

The Pareto Principle in virtual communities of learning

Enric Serradell-Lopez [a,*], Pablo Lara-Navarra [b], Silvia Martínez-Martínez [b]

[a] Business and Economics Department, Universitat Oberta de Catalunya, Spain
[b] Information and Communication Sciences Department, Universitat Oberta de Catalunya, Spain

ARTICLE INFO

Keywords:
Pareto principle
Educational data mining
Learning analytics
Virtual communities of learning
Collaborative learning
Virtual learning environments

ABSTRACT

In virtual learning communities, becomes evident the need to generate methods of analysing social networks to investigate participation and the exchange of knowledge. Big data, and more specifically educational data mining and learning analytics, allows deepening the study of the learning process in these virtual contexts. In this article, the objective is to analyse the applicability and the validity of the Pareto principle in the field of online education. This classical principle, also called the 20/80 rule, describes the causes and effects of certain variables and proportions. Its fulfilment has been demonstrated in very diverse fields.

Based on reticular analysis for the detection of communication patterns and generation of knowledge in online learning communities, and using the Pareto Principle, the article studies the correlation between academic results and the exchange of knowledge. From a sample of 184 members of the university community (between students and teaching staff) and 1354 messages, the correspondence is established by analysing, on the one hand, the qualifications and, on the other hand, the participation and monitoring of the interactions produced in the communication spaces made available to students in the classroom.

Some of the results obtained are summarized in the following points:

- Difficulty in predicting grades obtained from classroom activity.
- The Forum is shown as an essential element of activity in the classroom.
- The reading of messages in the Forum is concentrated in the first days of the course activity.
- Exists a relationship between the obtained qualification and the monitoring of classroom communication spaces.
- High qualifications do not correspond to the level of participation.

1. Introduction

Vilfredo Pareto, in his study on the distribution of wealth, points to the idea of the unequal distribution of incomes (Basulto et al., 2011). His theory, designed in the field of economics, underwent reformulations and subsequent applications (Kume, 2002, p. 19). It is considered that was Joseph M. Juran who used his name to designate "to this principle of the 'vital few and trivial many'" (Juran, 1975).

Transferred to different areas, this rule of the 20/80 in which a minority causes most of the phenomena (Bonet, 2005), also it has known as the universal law of priorities. For instance, applied to the field of location-based social networks, Cheng et al. (2012: 19) detected the top-20 most visited locations received the 80.5% of all check-ins, "following the Pareto principle".

In particular, our interest is directed to the educational sector, and in this area we can highlight the analysis of academic results made in the work of López (2017) that in his study on LOMCE, based on the results of Hattie (2003) and the conclusions of PISA (2015) where it analyzes "the effective scope of the last Spanish educational reform in light of the Pareto principle" (López, 2017, p. 20).

This article, starting from the analysis of the data generated in controlled environments for the exchange of knowledge between members of virtual learning communities, studies the monitoring of interactions between information resources and the actors that occur in the classroom communication spaces, to later analyse the behaviour of the virtual community. Specifically, it pursues, by applying the Pareto principle to these learning environments, determine the correlation between participation and monitoring (readings) of the interaction and academic performance.

* Corresponding author.
E-mail addresses: eserradell@uoc.edu (E. Serradell-Lopez), plara@uoc.edu (P. Lara-Navarra), smartinezmartinez1@uoc.edu (S. Martínez-Martínez).

https://doi.org/10.1016/j.chb.2022.107444
Received 2 December 2018; Received in revised form 19 July 2022; Accepted 14 August 2022
Available online 18 August 2022

2. Detection of knowledge in virtual learning communities

A network can be defined as a set of nodes interconnected by communication channels or agents. To talk about community, it is necessary an element of cohesion and commitment, and it already raised Rheingold (1993:5) to define the virtual community as a "social aggregations that emerge from the Net when enough people carry on those public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace". Therefore, it would be difficult to think of a static community, where members do not communicate with each other. The most normal would be a community with one or more networks that allow the exchange of many types of things or items. So, the communities "are held together through the mutual production and reciprocal exchange" (Jenkins, 2006). When describing a community, Etienne and Beverly Wenger-Trayner (2015) point to requirements that "members engage in joint activities and discussions, help each other, and share information. They build relationships that enable them to learn from each other; they care about their standing with each other".

The technological evolution has facilitated and has allowed configuring new types of networks and virtual communities also impacting sectors as varied as the social, economic and cultural. The networking and the use of social software platforms, evolving from the initial network of contacts as something instrumental, already become an objective in itself by providing social capital, a useful good to solve eventual future difficulties (Hughes et al., 2014; Lu et al., 2013). There are so-called communities of value, among which are the communities of learning, interest and practice (see Fig. 1) and which are defined by characteristics such as the cohesion factor, size, type of leadership and its temporal limitation (Sanz, 2012, pp. 80–81).

In the so-called learning communities, data, information and knowledge are exchanged through networks. Learning can only flow in one direction or be shared among a set of agents, where all knowledge, and thus constitute a community of learning (Sanz & Reig, 2013). According to Onrubia (2004: 14), this type of community can be defined as: "contexts in which students learn thanks to their participation and involvement, in collaboration with the teacher and with other adults, in genuine processes of research and collective knowledge construction". His study required a clear and real conceptualization of those aspects that are important to analyse as well as a methodology that allows the description and understanding of the environment and the mechanisms that allow the generation of knowledge (González & Andreu, 2013).

3. Big data applied to the study of learning communities

Some of these virtual communities are contacted through the so-called Learning Management System or virtual learning platforms. The different software on the market offer solutions for the administration and management of the learning process. In addition, to facilitating collaborative work, these environments allow overcoming temporal and geographical boundaries while "let you incorporate the benefits of multimedia resources to educational activities and facilitated by different ways the construction of dialogues, either personal or group ones" (Rojas, 2015). In these environments communities of students are contacted and it is possible to get data from messages and content published or shared to own interactions that occur there. There is therefore a great potential for data obtained from multiple sources that are likely to be analysed. "The educational innovation cannot be based only on intuitions, it is necessary to take the step towards experimentation, measurement, rigour and systematicity" (García-Marco, 2013).

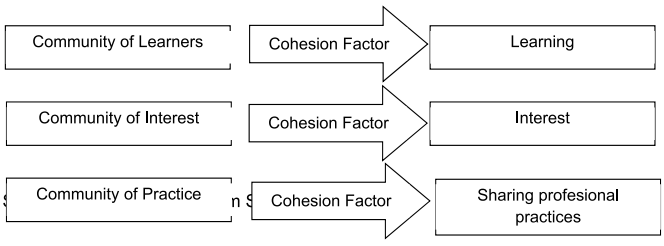

Fig. 1. Topology of value communities according to their cohesion factor. Source: Own elaboration from Sanz (2012: 80–81).

Large collections of structured and non-structured in real-time or deferred (Bowden, 2014) configured what is called big data and which is also characterized by the volume, velocity, variety and value of these data (Russom, 2011; Zaslasyky et al., 2012). These attributes "are presented as an axis and focus of attention for the study, treatment and processing of the data" (Martínez-Martínez & Lara, 2015, p. 577) in different areas. In the case of education, while the study of databases to extract patterns of behaviour of students has already some history, in recent years two lines of study have grown in followers interested "on how big data be exploited to benefit education and the science of learning: Educational Data Mining and Learning Analytics" (Baker & Inventado, 2014). One of the most widely used definitions of Learning Analytics (LA) is provided within the framework of the 1st International Conference on Learning Analytics where the aim is to study the data generated on "learners and their contexts", seeking not only a better understanding of the process of learning but also to be able to improve it. Romero and Ventura (2013), meanwhile point out that educational data mining (EDM) "analyse data generated by any type of information system supporting learning or education (…). These data are not restricted to interactions of individual students with an educational system (…) but might also include data from collaborating students (…), administrative data (…), student affectivity (…), and so forth". According to Siemes (2013), "where LA is more concerned with sensemaking and action, educational data mining (EDM) is more focused toward developing methods".

In this context, from the attributes of entities and existing datasets, this research is interested in the study of social networks for the exchange of knowledge in virtual communities of learning empirically, analysing their structural aspects, that is, identifying the patterns of socio-cultural organization. In this sense, neither the design nor the technical infrastructure can ensure the effectiveness of learning, unless it is dealt with as a problem of social learning (Tarsiero, 2007). Likewise, although they can detect and identify the knowledge, given the enormous complexity of circumstances that affect the generation of content in these environments (Duart; Lara & Saigi, 2003), the focus will be only on some of the dimensions of the learning process.

4. Methodology

This article studies the structure of the interactions between information resources and actors, to later analyse the behaviour of the virtual community. Specifically, it explores, transferring the Pareto principle to this environment, the correlation between academic efficiency and the exchange of knowledge through various tools of communication (forum, debate and Board) made available to the students. It intends, as a general objective of the research, to improve knowledge of virtual communities, the number of established relationships and monitoring of interactions generated during the period of participation and the degree of success in the process of learning to guide the study to results apply to the improvement of processes and systems. From the analysis of the patterns of behaviour, it will obtain essential knowledge to generate new information architectures that allow adapting the content to the different realities of community actors.

The methodology used is based on the study of the big data applied to virtual learning communities and focuses on three aspects: communication channels, information resources, and interaction. To trace the interactions that occur in a virtual community between contents and actors is necessary:

(a) Analyse data necessary and available for the creation of a network model;
(b) Capture the information provided by the virtual community for the generation of the information structures;
(c) Designing models of the network of the virtual community that reflect the relationships and processes of information and communication;
(d) Study the characteristics of the network and relevant information provided by models: topological properties, structures and classification of key elements.

The study is based on a data set that records every interaction between the members of the community, lecturers and students, and data was collected from a teaching semester (from February to July). Interaction means all action on a message, from sending it, reading it, forwarding it, deleting it or answering it. In this article, we focus on the interactions in three spaces of communication (forums, debates, and the board of the virtual classrooms) and more specifically on the reading or the shared messages. Table 1 describes the characteristics and use of the communication spaces in the virtual classrooms.

Table 2 summarizes the details of the record of interactions and the selection of the sample.

This research focuses on two subjects whose interactions have been collected in the data set. The criteria applied to select the subjects' object of study was the academic results of students. So, it is selected a subject with higher academic performance - with the greatest concentration of B and outstanding grade - (community A) and another with lower academic performance - with a higher percentage of approved and failed than the previous - (community B). This facilitates establishing a comparative in assessing the incidence of the readings interactions in the result of the learning process. To carry out a first exploratory and descriptive approach to point to the value of the research and the interest to continue deepening in a statistically significant study that allows establishing generalizations, the article analyses a sample involving a total of 1354 messages and 184 members of the university community between students and teaching staff, all of them members of the Open University of Catalonia and that were distributed in virtual classrooms in selected subjects. The data is generated by the operation of the system in logs and databases that store information automatically. The dependence of data on a net system, where if an individual is removed disappears with him the relationships and associated nodes, leads to ruling out random sampling. Therefore, it is analysed the totality of interactions allowing identify, accurately, the centrality levels of individuals and communication elements.

5. Results

Table 1
Communication spaces.

Name of the space	Characteristics and use
Board	Space from which the lecturer issues teaching content messages addressed to all the students in the classroom.
Forum	Space from which both the teaching team and the students can issue messages. It is used to share questions and doubts related to the subjects and solve them.
Debate	Space in which students and the teaching team can post messages. These are raised around a debate linked to the subject on which to reflect. Participation is usually related to an activity that can be assessed and has an impact on the grade.

Table 2
Record of interactions and selection of the sample.

Transactions		Selections of the sample	
Information obtained	Messages sent by students and teachers	Requirements	The messages studied must added at least one reading by a member of the group
	Readings of the contributions in the forum, debate and board by students and teachers.		The interaction must take place during the six months of the study period
	Final qualifications		The interaction must be done through the communication spaces available in the virtual community.

The comparative analysis of the academic performance of both subjects allows observing similarities and differences between the analysed virtual communities (Fig. 1). The percentage of students who pass the subject is similar in both communities (77.6% in community A and 75.3% in the community B).[1] In community A, the percentage of outstanding grades is almost three points higher than the results seen in community B. In both cases the highest percentage of students is registered among those who obtained a B as a final grade; however, this percentage is also superior in community A. The biggest differences are recorded to study situations of the students that don't pass the subject: while the percentage of failing grades is significantly lower in community A, the figure for non-presented is almost eight points higher than in community B. The abandonment of the learning process is higher, therefore, in community A while the percentage of those who fail despite being presented to the evaluable activities and participating in the learning process is higher in community B (see Fig. 2).

In terms of interaction, we focus on the monitoring of certain communications from the action "reading". In total it is made the monitoring produced by 687 messages of community A and 667 messages of community B. These messages are those that have registered a minimum of one reading and that are published in the board, forum or debate space of the classroom (see Fig. 3).

The correspondence of messages by space is shown in graphic 2 being very similar in both communities.

The total of interactions or readings recorded by members of the community of the messages published in the board, forum or debate spaces during the analysed period is 37099. In community A add a total of 17582 readings and 19517 in community B. In both cases, the highest number of readings is recorded in the forum, an open space for publication by all the members of the community.

The forum spaces' reading represents the 87% of the total. Using this

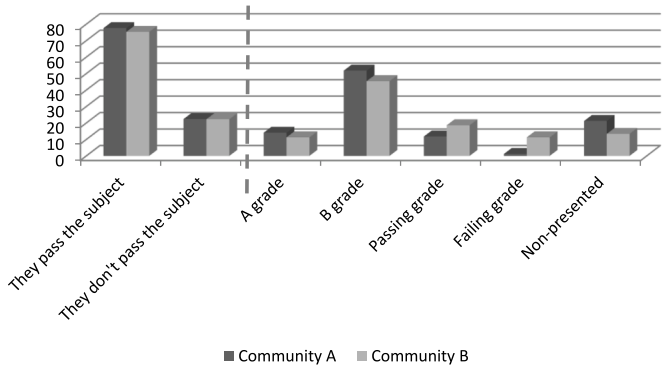

Fig. 2. Academic results in the virtual communities analysed.

[1] Students who fail to obtain a final grade have been accounted as not presented for this purpose.

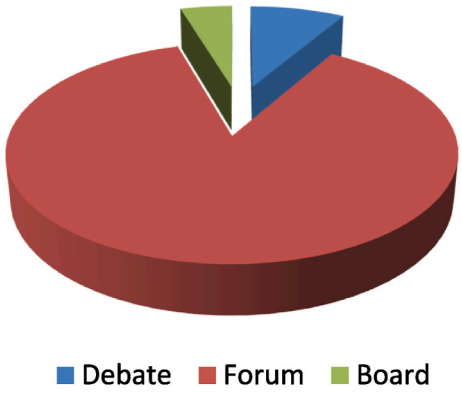

Fig. 3. Percentage of messages with at least one reading by communication space.

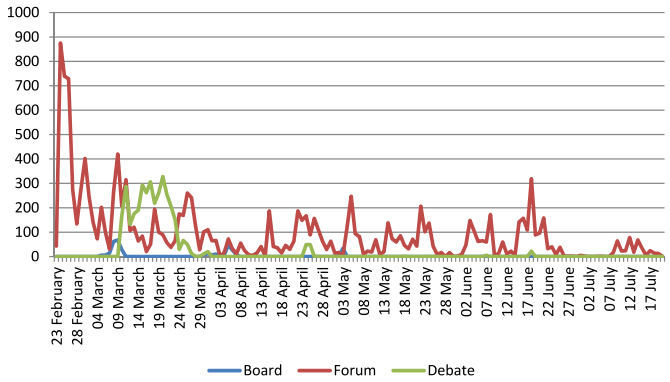

Fig. 4. Readings, per date, of messages published in the board, the forum and the debate space of the community A.

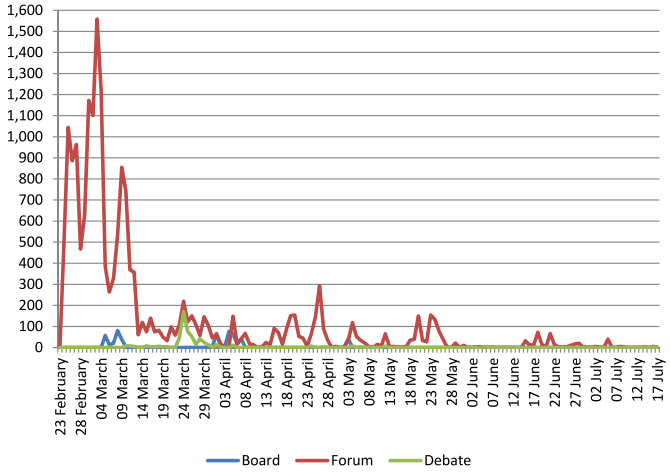

Fig. 5. Readings, per date, of messages published in the board, the forum and the debate space of the community B.

variable, the use of the Forum follows the Pareto principle.

The debate, whose participation is marked by any activity that the lecturer proposes and corrects, is collecting fewer total readings by the members of the community (8.4% of the total), but the average of readings by published messages is much greater in this area than in the rest (62.3 readings per message) in the community A (see Table 3). This result demonstrates the degree of involvement that supposes for the students of this group to follow the conversation that takes place in that space because it is directly related to the evaluation. By contrast, in community B the highest average of readings per message is given in the forum (31.6) and the debate adds the lower average (9.8). In both communities, the board, a space in which fewer messages are accounted for in the study, has fewer total readings (4.6% of the total). The average, however, shows a greater intensity of readings per message in community B (20) than what is happening in community A (9.9).

Considering the temporal evolution of the readings, graphics 3 and 4 allow establishing comparisons between the behaviours of the two communities. In both communities, the readings of messages posted on the board are very temporarily located, which must coincide when the lecturer publishes a notice. In the case of the debate, the readings are also concentrated in a specific period time that starts with a considerable increase in readings to progressively decreases. In the case of the forum, the evolution of the readings is more irregular showing the growths of the activity that come one after another throughout the semester. Especially significant is the record of readings that takes place at the beginning of the period in the forum when the members of the community begin to identify themselves and introduce themselves to the rest coinciding with the interest of the community to know the other participants.

As can be seen in Figs. 4 and 5, and using time as a variable, most of the readings in the Forum take place in the first five days, which seems to confirm Pareto's theory from this point of view.

In order to compare the data with the academic results, the interactions produced by the academic staff are removed from the results to focus on the activity or readings generated by the students.

Average readings are described in Table 4. In community A, the group of students who obtain a higher grade (A) has the highest average reading, 240.6. On the other extreme, students who leave the learning process correspond to the lowest average readings of the communication spaces, 119.1. The same correlation between readings and academic performance is not observed in the case of students who obtain a B (with an average of 222), passing (231) or failing grade (237). By space, it is possible to observe a differential behaviour because, among students who pass the subject, the highest average reading of messages in the forum (189) and the board (5.2) is registered among students who get a passing grade while the highest average reading of the debate (48) is among the students who obtain an A grade in the subject.

In community B, the highest average reading (222.3) is recorded among students who obtain a B grade. The results do not show great differences when observing the rest of the students who pass the subject. Thus, those who obtain an approved have an average of 220.2 readings and those who get an A grade 219.7. The lowest average of readings is found among students who leave the learning process (130) followed by those who failed the subject (162). As it will be observed in community A, when crossing the readings data of each communication space with the results of the academic performance of those students who pass the subject, it is verified that also the students with an A grade are those that

Table 3
Average readings per message.

Space	Community A	Community B
Board	9.9	20
Forum	23	31.6
Debate	62.3	9.8

Table 4
Average readings according to the academic results.

	Average of readings	
Academic results	Community A	Community B
A grade	240.6	219.7
B grade	222	222.3
Passing grade	231	220.2
Failing grade	237	162
Non-presented	119.1	130

show a higher average of readings in the debate space (6.8). In the rest of the spaces, the students that achieve a B grade present the greatest average readings: 207.2 in the case of the forum and 6 in the board.

Analysing the isolated behaviour by individual, 82.6% of the students with A grade surpass the average of readings recorded in the set of the two classrooms. Specifically, in Community A, 83.3% of the students who obtain the highest grade present levels of readings above the average of their group. In community B, this percentage is 72.7%. Among the 16 most active students in terms of reading in community A, 81.25% get an A or B grade. The students who obtain an A or B grade make the 73% of the readings recorded in community A while in community B the percentage is 62.11%.

6. Conclusions

The results show that there is a relationship between the obtained qualification and the monitoring of classroom communication spaces since most of the small group of students with the highest academic results show averages of reading messages higher than the community as a whole. Also, students who drop out and don't present for the evaluation (17%) show the lowest average readings, this group is followed by students who do not pass despite participating in the learning process, whose readings only represents 5.5% of the total of the actions carried out by the group of students. Conversely, it should be noted that among students who pass the subject with high qualification are also cases of low reading of shared messages in the classroom and, among those who do not pass it, there are students with reading figures above average. This confirms the unpredictability of the activity of the members (Barker et al., 2013) and the flexibility of a pedagogical model that allows personal time management and does not penalize the student who presents a low intensity in terms of reading or monitoring the activity of the classroom and the community.

Regarding Pareto principle approach, Forum's reading represents the 87% of the total, we must add the fact that most of the reading take place during the first five days of the course. We have not detected other works that consider these same variables from the point of view of the Pareto principle. Within the field of knowledge of learning, the work of O'Neill (2018) takes this principle into account when identifying critical elements related to learning grammar, or the work of Valkanas and Diamandis (2022) that confirm the Pareto distribution regarding the viewing of online videos linked to learning resources.

In interpreting our study, several important limitations must be kept in mind. The first one is that the research is very specific and focuses on two subjects therefore the results obtained explain only the sample data and should not be extrapolated to other subjects. In any case, we consider that the number of messages analysed helps to fully describe the sample analysed. The second major limitation is that the statistical analysis performed is basically descriptive. Other important variables should be added to future studies, especially those related to the students' profile as gender. Although the presented data correspond to a test of the initial phase of application of the exposed methodology, the obtained results encourage deepening the study, with an analysis statistically significant and establishing longitudinal comparisons from the data collection of successive semesters which can extend the findings to the general population. The triangulation of the results with other variables and sources of information will allow knowing in more detail the behaviour of the learning communities and assessing whether the Pareto principle applied to academic performance is fulfilled when analysing other causes or factors such as writing, or the number of contributions made. The value of the exposed methodology and the analysis of the results is presented as an opportunity to improve the effectiveness of the learning process and model interaction structures that contribute to achieving higher academic performance. This does not imply ignoring the role that psycho-emotive factors (Ghosh et al., 2012) or interactions that occur outside the virtual environment that configures the classroom can perform in the process. Another issue is the fact that the research is based on individual results, so teamwork could not be taken into consideration in this research. In any case, this line of future research is very promising if we want to predict the expected results of the students (Crespo & Antunes, 2015) or if we want to study in depth all the interactions within the virtual classroom (Hernández-Lara et al., 2019).

Finally, the study allows detecting trends in behaviour and, more specifically, in the monitoring or readings of the messages of each classroom communication space analysed. The forum becomes the backbone space with the largest number of reading messages recorded and in which the action of reading remains more active, at a different rhythm, throughout the semester. In addition, it is the fundamental space to put the members of the community in contact and to know each other as observed by the readings of the messages registered coinciding with the beginning of the semester. On the contrary, the debate (linked to an evaluable activity or subject to correction) and the board (where only the teaching team publishes the messages) present reading activities more limited temporarily. In both spaces, therefore, the role of the moderator, in this case, the lecturer, is fundamental and decisive to encourage reading in the group (Barker et al., 2013) when marking, with the publication of notices or through a proposed exercise during activity periods. The knowledge of these rhythms of reading by students can contribute to the positioning of some of its members in the community by getting more readings or followers for each message published in the forum or the debate.

In summary, we would like to highlight the following points:

- Difficulty in predicting grades obtained from classroom activity
- The Forum is shown as an essential element of activity in the classroom
- The reading of messages in the Forum is concentrated in the first days of the course activity
- Exists a relationship between the obtained qualification and the monitoring of classroom communication spaces.
- High qualifications do not correspond to the level of participation

With the available data, it is difficult, if not impossible, to reliably predict the grades obtained by students based on their activity in the classroom. Of course, eLearning research has this issue as one of the central elements where the technology may, over time, offer solutions.

We believe that this work can help teachers to better distribute workloads and use classroom learning tools more efficiently.

Data availability

The authors do not have permission to share data.

References

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learners analytics", 61-75. In J. Larusson, & B.(White (Eds.), *Learning analytics: From research to practice*. New York: Springer.

Barker, D., Quennerstedt, M., & Annerstedt, C. (2013). Inter-student interactions and student learning in health and physical education: A post-vygotskian analysis. *Physical Education and Sport Pedagogy*, 409–426. https://doi.org/10.1080/17408989.2013.868875

Basulto, J., Busto, J., & Sánchez, R. (2011). El concepto de desigualdad en Vilfredo Pareto (1848-1923). In J. M. Riobóo, & I. Riobóo (Eds.), *Historia de la Probabilidad y la Estadística (V). A.H.E.P.E* (pp. 47–76). Santiago de Compostela: Ni-no Centro de Impresión Digital.

Bonet, M. (2005). Ley de Pareto aplicada a la fiabilidad. *Ingeniería Mecánica, 8*(3), 1–9.

Bowden, J. (2014). Reasons to explore big data with social media analytics. SocialMediaToday.com https://www.socialmediatoday.com/content/reasons-explore-big-data-social-media-analytics-videos. January 15.

Cheng, C., Yang, H., King, I., & Lyu, M. R. (2012). Fused matrix factorization with geographical and social influence in location-based social networks. In *Twenty-sixth AAAI conference on artificial intelligence*. AAAI Publications. https://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/view/4748.

Crespo, P. T., & Antunes, C. (2015). Predicting teamwork results from social network analysis. *Expert Systems, 32*(2), 312–325.

Duart, J. M., Lara, P., & Saigí, F. (2003). Gestión de información en el diseño de contenidos educativos online. *Revista iberoamericana de Ciencia, Tecnología, Sociedad e Innovación, 6*, 2003 http://www.oei.es/revistactsi/numero6/articulo05.htm#.

García-Marco, F. J. (2013). Educación y aprendizaje de la información y la documentación: Raíces, desafíos y líneas de acción. *El profesional de la información, noviembre-diciembre, 22*(6), 489–504. https://doi.org/10.3145/epi.2013.nov.01

Ghosh, R., Rude-Parkins, C., & Kerrick, S. A. (2012). Collaborative problem-solving in virtual environments: Effect of social interaction, social presence, and sociability on critical thinking. In *The next generation of distance education*. Boston, MA: Springer.

González, A., & Andreu, C. (2013). Investigación del comportamiento informacional a través del análisis de redes sociales. *El profesional de la información, noviembre-diciembre, 22*(6), 522–528. https://doi.org/10.3145/epi.2013.nov.04

Hattie, J. (2003). Teachers make a difference: What is the research evidence? *Australian Council for Education Research Annual Conference on: Building Teacher Quality*, 1–17.

Hernández-Lara, A. B., Perera-Lluna, A., & Serradell-López, E. (2019). Applying learning analytics to students' interaction in business simulation games. The usefulness of learning analytics to know what students really learn. *Computers in Human Behavior, 92*, 600–612. https://doi.org/10.1016/j.chb.2018.03.001

Hughes, M., Morgan, R. E., Ireland, R. D., & Hughes, P. (2014). Social capital and learning advantages: A problem of absorptive capacity. *Strategic Entrepreneurship Journal, 8*(3), 214–233. https://doi.org/10.1002/sej.1162

Jenkins, H. (2006). *Convergence Culture. Where old and new media collide*. New York and London: New York University Press.

Juran, J. M. (1975). The non-Pareto principle; mea culpa. *Quality Progress, 8*(5), 8–9.

Kume, H. (2002). *Herramientas estadísticas básicas para el mejoramiento de la calidad*. Bogotá: Editorial Norma.

López, F. (2017). Un análisis de la LOMCE a la luz del principio de Pareto. In *Cátedra de Políticas educativas*. Universidad Camilo José Cela. http://hdl.handle.net/20.500.12020/533.

Lu, J., Yang, J., & Yu, C. S. (2013). Is social capital effective for online learning? *Information & Management, 50*(7), 507–522. https://doi.org/10.1016/j.im.2013.07.009

Martínez-Martínez, S., & Lara, P. (2015). El big data transforma la interpretación de los medios sociales. *El Profesional de la Información, 23*(6), 575–581. https://doi.org/10.3145/epi.2014.nov.03

O'Neill, K. S. (2018). Applying the Pareto principle to the analysis of students' errors in grammar, mechanics and style. *Research in Higher Education Journal, 34*.

OECD. (2016). PISA 2015 results. In *Policies and practices for successful schools*. Paris: OECD Publishing.

Onrubia, J. (2004). Las aulas como Comunidades de aprendizaje. *Trabajadores de la Enseñanza, 249*, 14–15.

Rheingold, H. (1993). *The virtual community. Homesteading on the electronic frontier*. Reading, Mass: Addison-Wesley.

Rojas, P. (2015). Paradigmas analíticos en entornos virtuales y de aprendizaje: Una revisión de sus principales puntos de encuentros y diferenciaciones teóricas y de enfoque. *Revista educación y tecnología*, (7), 91–106.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3*(1), 12–27. https://0-doi-org.cataleg.uoc.edu/10.1002/widm.1075.

Russom, P. (2011). *Big data analytics*. TDWI Best practices report. *Fourth Quarter, 19*(4), 1–34. https://tdwi.org/research/2011/09/best-practices-report-q4-big-data-analytics.aspx?tc=page0.

Sanz, S. (2012). Comunidades de práctica. In *El valor de aprender de los pares*. Barcelona: Editorial UOC.

Sanz, S., & Reig, D. (2013). El aprendizaje social y los profesionales de la información. *El Profesional de la Información, 22*(6), 545–553. https://doi.org/10.3145/epi.2013.nov.07

Siemes, G. (2013). Learning analytics, the emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380–1400. https://0-doi-org.cataleg.uoc.edu/10.1177/0002764213498851.

Tarsiero, R. (2007). Facilitating social learning in virtual communities of practice. In *Enhancing learning through human computer interaction. McKay, Elspeth*. London: Idea Group Inc.

Valkanas, K., & Diamandis, P. (2022). Pareto distribution in virtual education: Challenges and opportunities. *Canadian Medical Education Journal, 13*(1), 102–104. https://doi.org/10.36834/cmej.73511

Wenger-Trayner, E., & Wenger-Trayner, B. (2015). *Introduction to communities of practice. A brief overview of the concept and its uses"*. Wenger-trayner.com. April 15, 2015 https://wenger-trayner.com/introduction-to-communities-of-practice/.

Zaslasyky, A., Perera, C., & Georgakopoulos, D. (2012). Sensing-as-a-service and big data. In R. Ranjan, R. Buyya, & A. Basu (Eds.), *Advances in cloud computing; 26-28 july, 2012* (pp. 21–29). Bangalore, India. Hyderabad: Universities Press (India) Pty Ltd, ISBN 978 8173717789, 2012.