

# BIG DATA FOR SOCIAL MEDIA LEARNING ANALYTICS: POTENTIALS AND CHALLENGES

**Stefania Manca<sup>1</sup>**  
**Luca Caviglione<sup>2</sup>**  
**Juliana Elisa Raffaghelli<sup>3</sup>**

<sup>1</sup> Institute of Educational Technology, National Research Council of Italy, stefania.manca@itd.cnr.it

<sup>2</sup> Institute for Intelligent Systems for Automation, National Research Council of Italy, luca.caviglione@ge.issia.cnr.it

<sup>3</sup> Independent researcher, jraffaghelli@gmail.com

**Keywords:** MOOCs, social media, social learning analytics, open datasets, big data, privacy & security, ethics, data anonymization.

Today, the information gathered from massive learning platforms and social media sites allows to derive a very comprehensive set of learning information. To this aim, data mining techniques can surely help to gain proper insights, personalise learning experiences, formative assessments, performance measurements, as well as to develop new learning and instructional design models. Therefore, a core requirement is to classify, mix, filter and process the big data sources involved by means of proper learning and social learning analytics tools. In this perspective, this paper investigates the most promising applications and issues of big data for the design of the next-generation of massive learning platforms and social media sites. Specifically, it addresses the methodological tools and instruments for social learning analytics, pitfalls arising from the usage of open datasets,

for citations:

Manca S., Caviglione L., Raffaghelli J.E. (2016), *Big data for social media learning analytics: potentials and challenges*, Journal of e-Learning and Knowledge Society, v.12, n.2, 27-39. ISSN: 1826-6223, e-ISSN:1971-8829

and privacy and security aspects. The paper also provides future research directions.

## 1 Introduction

Among current and future trends in higher education, serious games, augmented reality, cloud computing and storage, Massive Open Online Courses (MOOCs), social media sites, jointly with the use of big data and analytics, have been identified as the most significant and promising (Johnson *et al.*, 2015). When massive learning through MOOCs, for instance, is specifically considered, learners are increasingly requested to share their learning and social achievements on social media and social network sites to increase engagement and visibility. From this perspective, the learning environment is no longer constrained, but becomes a continuous space “spread” over the Internet (Siemens, 2005).

As of today, learners in online courses typically use a mixed set of Web-based services, including different/specialised SNS, which can interact with smartphones or wearable devices. In fact, the increasing diffusion of SNS accessed from mobile appliances on an *anywhere-anytime* fashion have dramatically boosted the availability of personal data, such as geographical locations, photos and textual entries. Additionally, the growing coupling among specialised SNS, e.g. to share photos or to maintain a list of favourite readings, enables to exchange and merge a vast portion of the data available on the Internet. Therefore, the huge amount of information on learners has to be considered big data, which needs to be properly classified, mixed, filtered and processed, in order to derive significant indications for learning (Merceron, Blikstein & Siemens, 2015).

In general, the most used educational frameworks, such as MOOCs and platforms interacting with third-party social media services, are the opportunity to collect unprecedented volumes of information describing how the users interact with the learning system. This means that data mining techniques can be increasingly applied to gain insights on and potentially personalise human learning (Cooper & Sahami, 2013). Accordingly, there is a growing interest in the adoption of new sources for the personalisation of the various learning experiences, formative assessment of learning, performance measurement, and for new learning and instructional design models (Johnson *et al.*, 2015).

Learning analytics and social learning analytics provide the methodological foundations to gather and analyse large amounts of details about student interactions in online learning (Buckingham Shum & Ferguson, 2012; Siemens, 2013). Currently, the main aims of learning analytics are: *i*) to find out within the bulk information which part is useful to advance in the learning process (see, Coccoli *et al.*, 2014 for the case of smart universities), and *ii*) to reveal

patterns that can be used to improve students' learning (Long & Siemens, 2011). Unfortunately, such a rich volume of personal information also accounts for privacy and ethics issues (Slade & Prinsloo, 2013). As a consequence, several universities have started producing policies on the ethical use of student data, grounded on principles that are linked to particular facets of its collection and analysis<sup>1</sup>.

In this perspective, this paper aims to highlight the most relevant educational potentials and challenges offered by the use of big data in massive online learning environments and social media sites. Emphasis will be put on the following aspects: methodological tools and instruments for social learning analytics, issues of open datasets, and privacy and security issues. For each theme, the main results achieved in the related literature are presented, along with open issues that deserve further investigation and research. Our claim is that this paper is not a systematic and analytical presentation of the literature related to the topic of big data in education. Rather, it aims at pointing out actual and emergent issues related to big data and massive online learning environments and social media sites.

The remainder of the paper is structured as follows. Section 2 deals with big data and social learning analytics, while Section 3 introduces a discussion on open datasets and ethics issues. Section 4 investigates hazards related to privacy and security aspects, and Section 5 showcases future research directions and implications for design. Lastly, Section 6 concludes the paper.

## 2 Implications of big data for social learning analytics

The online interaction of learners in asynchronous learning environments has been traditionally analysed through techniques such as log files and tracking analysis, content analysis, discourse analysis and, more recently, social network analysis (Pozzi *et al.*, 2007). The combination of such approaches, exploiting both textual and visual representation of interactions, has seen a renewed interest in the fields of learning analytics and educational data mining (Buckingham Shum & Ferguson, 2012; Siemens, 2013). Analytics from Learning Management Systems (LMS) offers a source of data to predict the success of learners, e.g. by comparing basic activities related to participation, such as content pages viewed, number of posts, or the duration of participation through time spent viewing discussion pages and content (Ferguson & Clow, 2015; Long & Siemens, 2011). However, even if data collected within LMS closely reflects the interaction of learners within a system, modern distributed networks and the physical world present additional challenges to be addressed. This is

<sup>1</sup> For the UK Open University policy, see <http://www.open.ac.uk/students charter/essential-documents/ethical-use-student-data-learning-analytics-policy>

the case of learning activities on platforms like Facebook, Twitter, blogs, and physical world data, such as those related to library use or access to learning support, that are not captured by LMS analytics (Kop, 2011).

Therefore, to provide meaningful indications for educational purposes, massive User Generated Contents (UGC)s require the capability of collecting, storing, analysing and presenting data gathered from social media. This has recently been addressed by social media analytics, whose purpose is “*evaluating informatics tools and frameworks to collect, monitor, analyse, summarize, and visualize social media data to facilitate conversations and interactions to extract useful patterns and intelligence*” (Fan & Gordon, 2014, p. 74).

Social media analytics has been used primarily in business and marketing studies. In this field, its aim is to collect relevant data by monitoring various sources using different ICT-based techniques, such as the parsing of news feeds, access to public data through proper Application Programming Interface (API) or by means of Web crawlers. The acquired information is then archived by using data modelling, data and record linking from different sources, stemming, part-of-speech tagging, feature extraction, and other syntactic and semantic operations. The selected and filtered information can be assessed for meaningful understanding through statistical methods and other techniques derived from data mining, natural language processing, machine translation, and social network analysis. This stage is the core of the entire process, along with the summary of results, their evaluation and presentation mainly through visualisation techniques (Fan & Gordon, 2014).

For the case of online learning environments exploiting social media sites, the main goal is to understand how to identify effective pedagogies and design techniques to foster deep and meaningful learning (Anshari *et al.*, 2016; Kop, 2011). A recent theoretical model of networked learning relies on quantitative and automated methods, such as social network analysis and quantitative content analysis. Besides, it exploits qualitative techniques of contextual analysis, such as discourse analysis (Joksimović, Hatala & Gašević, 2014). For instance, the aim of the model is to examine the role of course participants within a network of learners engaged in a Twitter conversation, so as to identify influential learners and their contribution to the network knowledge base.

However, as stressed by some scholars (Eynon, 2013), there is not data available for every question that educators can ask. This means that only collectable data can be analysed and studied, thus orienting the questions and research. Indeed, there are significant differences between social media sites and their opportunities for data mining. While Twitter is a ready-made laboratory for data scientists, data mining of services like Facebook is more challenging. In fact, the rich amount of information available on Facebook requires more conservative privacy-preserving policies, thus preventing full-access to user data. Instead,

Twitter only uses short textual entries, resulting into a more “open” design. This leads to many studies focusing on a single platform, which have contributed to overlooking the wider social ecology of interaction and diffusion of methodological issues. If it is true that Twitter has emerged as a “model organism” of big data, the lack of adequate consideration of structural biases of different platforms has led to overrepresentation of models of analysis (Tufekci, 2014).

In addition to these considerations, the tracking of specific activity patterns does not exhaust the behaviours of learners, such as the content of what they share or the meaning people place on connections with others (Eynon, 2013). Adopting a critical lens derived from sociological studies, Selwyn (2015) pointed out further important issues that may arise when digital data are used in education. One area of concern is related to the reproduction of inequalities and social relations. According to Selwyn (*Ibidem*), social power and control dynamics might be reinforced as far as existing power can be made ever more invisible or taken for granted. Moreover, there exists a divide between those who create data for others to process and those who have the expertise to collect and analyse such information.

Another area of concern is related to the role of digital data in reinforcing and intensifying the culture of managerialism within educational settings, as well as their means for “*dataveillance*”. From this perspective, digital data would reinforce the tendency to produce algorithmically-driven thinking systems able to solve complex social problems associated with education and, at the same time, to control educational institutions and their actors.

### 3 Open datasets: from technical to ethical implications

The availability of digital data is generating the illusion of immediate access and usability. However, the very concept of data encompasses a long history of practices in science. The main problem has always been to unravel the arbitrariness in most definitions, based on epistemological, methodological and pragmatic approaches to research problems (Borgman, 2015). Across the domains of hard and social sciences, as well as humanities, digital and open datasets are expected to be interoperable, reusable, in a machine-readable format, with a low degree of granularity, and compliant with the Open Access criteria of being publically available. However, as Borgman (*Ibidem*) stressed, the degree of processing can be varied, according to the type of data, the technologies used and the researchers’ needs.

In the specific case of social learning analytics, data reuse encompasses many drawbacks (Verbert *et al.*, 2012). From late 2000, Technology Enhanced Learning (TEL) environments have provided open datasets, enabling an exponential re-use for research purposes. The most successful cases are the Pittsbur-

gh Science of Learning Center (Koedinger *et al.*, 2010), with more than 270 datasets derived from intelligent tutoring systems; the Mulce Project (Chanier & Reffay, 2011), mostly focused on learners' interactions about math learning; and the Harvard Dataverse project, that contains extensive datasets from edX MOOCs (Daries *et al.*, 2014). In Europe, the dataTEL research network aims at strengthening data sharing culture in TEL research and in the learning analytics field. Its major output is the definition of core elements for effective data-intensive open projects in TEL environments, such as data and dataset properties, together with learning analytics objectives (Verbert *et al.*, 2012).

In spite of the efforts described above in the field of educational data mining and learning analytics, social learning analytics is a research area that requires further understanding. Several datasets are available in projects devoted to social network analysis like UCI Network Data Repository (California University)<sup>2</sup> or SNAP (Stanford University)<sup>3</sup>. However, the challenging issue is to extract data that can be validly applied to the analysis of learning processes. In fact, the original open datasets available were configured for generic social studies purposes by means of social network analysis tools and not for educational analysis.

While the technological front is advancing steadily, ethical issues of digital and open datasets are puzzling and demand careful attention. The questions raised are not new and pertain to the domain of Internet research (Esposito, 2012). In fact, one of the main questions regarding datasets containing sensitive information is the protection of participants' identities against external attacks or misuses.

The issue relates to the process of data de-identification. Data anonymization is a key issue in social research (Jandel, 2014). If one of the main problems is to eliminate "identifiers" (variables such as name, age, and gender), also information that might lead to cross-checks connected to identification, or "quasi-identifiers", have to be hidden (Daries *et al.*, 2014). The case of edX MOOCs open dataset published in 2014 clearly illustrates the point. The two adopted procedures for de-identification, i.e. suppression and generalisation, led to the elimination of data records (e.g. participants enrolled in more than one course), or to the generalisation of data at a level where the record did not provide useful information. As Daries and colleagues (2014) reported, the information lost along the process of anonymization was such as to invalidate some of the correlations between variables vital to answer key questions about the gender or levels of certification in smaller MOOCs. Yet, as The Hague Declaration<sup>4</sup> on open data stressed, the "*ethics around the use of data and content*

<sup>2</sup> <https://networkdata.ics.uci.edu/resources.php>

<sup>3</sup> <http://snap.stanford.edu/data/>

<sup>4</sup> <http://thehaguedeclaration.com/>

*mining continue to evolve in response to changing technology”.*

Another key ethics issue is connected with the participants’ right to give meaning to the data they produce as part of their learning experience (Slade & Prinsloo, 2013). According to a dialogical and iterative approach to ethics in science, participants should be able to negotiate the ways in which data (including themselves) is produced. They are also entitled to ask for explanations about how the information is used for science and socio-technical purposes.

A number of learning analytics techniques have been implemented into many LMS so far, thus providing feedback to learners and teachers a framework to develop skills for lifelong learning. However, at the current state of the art, the ways these scenarios are evolving, the educational questions raised and the several connections with social learning analytics seem to be far beyond end-users’ understanding (Slade & Prinsloo, 2013).

#### **4 Assuring security and privacy issues with big data**

Web 2.0 technologies offer many opportunities to enhance the user experience, as well as to empower social learning analytics. Unfortunately, they also certainly account for privacy and security issues (see Caviglione & Coccoli, 2011) for a comprehensive discussion on new hazards characterising social media services). Besides, the convergence on standard representations, such as the JavaScript Object Notation (JSON), jointly with the offering of powerful sets of API to access information, increase the magnitude of the problem space. We mention among others: *i*) issues in storing the bulk of data to be processed. This can be mitigated by means of cloud-storage services, but at the cost of economical expenditure and loss of control over the data; *ii*) hazards in finding out useful correlations, applying filters and developing models from incomplete datasets. Indeed, privacy filters set-up by users or commercial policies could prevent from having the bits of information needed (e.g. the real name or date of birth), thus leading to noisy datasets; *iii*) challenges in processing the bulk of information in a reasonable amount of time and with an affordable processing power. This can be done through third-party services, but with the same drawbacks as cloud-storage ones (Caviglione, Coccoli & Gianuzzi, 2011).

In this perspective, a trade-off between privacy and security requirements and constraints arising to support social learning analytics must be pursued. This is a fragile balance, especially if recurring to commercially available SNS (e.g. Facebook), which could have conflicting goals, such as forcing the learning platform to use some assets offered on a pay-per-use basis. Therefore, the next-generation of platforms for e-learning wanting to take advantage of big data-like features and social learning analytics tools should be engineered with proper security and privacy requirements. Nevertheless, the “social” part

should be considered as built-in, thus offering efficient abstractions to teachers and students to help in their needs and daily routine. Possible features to be considered are:

- *Support standard formats to store and exchange information*: the use of well-known technologies to encode data could prevent many security issues. In fact, popular formats like eXtensible Markup Language (XML) have been widely studied and their vulnerabilities are largely documented. Besides, ad-hoc or proprietary formats might be prone to attacks using information hiding approaches (Mazurczyk & Caviglione, 2014);
- *Consider proper layers to import/export services*: to be effective, analytics procedures should be able to correlate and dig up a huge variety of data, e.g. pictures, network parameters, and statistics describing lectures attended. Hence, the resulting amount of data could require sophisticated and machine-intensive computations. Therefore, future e-learning platforms should be cloud-aware, as the use of external storage and computation assets might be mandatory. Besides, by providing the support to cloud as a built-in feature, the overall software architecture would be more robust and secure, as it exploits native and strict authentication and security policies, reducing the risk of attacks (Caviglione, 2009);
- *Perform network checks both for security and analytics*: many threats attack the network portion of the software architecture in charge of delivering e-learning services. A good framework should use the “secure” version of protocols, e.g. HTTPs. At the same time, an effective online e-learning platform should perform security checks, such as controlling the number of incoming connections, changes in the IP address used by the user, as well as sudden alterations in the geographic location. Such data could be used to enrich the bulk of information provided by SNS, as network analytics has proven to be effective to infer behaviours of users. Possible examples are the duration of a session, the amount of interaction with peers, and habits with use of learning material (e.g. the usage of a desktop vs a mobile device). This approach has two main benefits: data is available for free and such bits of information are inferred rather than pushing users to put details online in an explicit manner;
- *Reduce the risk of “profile fusion attacks”*: social-based services, including those offering e-learning material, are very specialised. Each one could offer a very narrow, but precise, view of an individual (e.g. preferred books, restaurants and places). In this scenario, learning materials, such as the courses selected, the topics of thesis or mid-term assignments, could be merged with other public data in order to infer

secret or unavailable information. Modern MOOCs could attract more students or craft ad-hoc course materials by exploiting data fusion and social analysis, for instance to offer insights on hot topics of interest. Obviously, the more sensitive information is used, the more secured it must be so as to avoid reducing the overall privacy of a user with public data available on the Internet;

- *Use plug-ins with a proper degree of suspicion*: the urge of porting actual e-learning tools towards a more social environment often imposes exploitation of features provided by third parties, for instance SNS. As a paradigmatic use case, we can consider the feature of log-in into a system through Facebook. On the one hand, this could prevent security hazards due to flawed software implementations or non-updated OSES in the hosting platform. On the other hand, this “lazy” form of delegation would result in less rich and specific data availability and, more importantly, require an external entity to act as data collector and provider.

## 5 Implications for research and design

Big data undoubtedly offer advantages to educational scientists, but there are also a number of implications for research and design that should be taken into consideration.

From an educational perspective, it is necessary to reflect on a series of issues. According to Selwyn (2015), first, it is important to understand what kind of data is available and how it is gathered and analysed by educational institutions and organisations. Second, there is the need to understand who is going to use the collected information and to what outcomes its use will lead. Third, analysts are required to address proper tools and technologies to respond to the questions raised. All these claims demand the development of analytic skills and attributes to engage effectively with the data, as well as the development of methodological competences that could also be aware of the theoretical implications.

In addition, as seen above, the overrepresentation of models of analysis based on certain platforms should be counterbalanced by new “model organisms”, properly identified for different social media platforms and able to contain skewness of mechanism analysis (Tufekci, 2014). Moreover, since social media users usually participate in more than one site at the same time (e.g. Facebook and Twitter), understanding broader patterns of connectivity between platforms becomes imperative if we want to study how people connect with others in social media. The study of social media big data should be complemented, whenever possible, with other methodological approaches, such as qualitative studies, interviews and digital ethnographies.

Also as for the crystallization of big data collected through social learning analytics in open datasets, different challenges are present, especially in the process of collecting, polishing, and curating data, and proper infrastructures are required to support access, reading, and re-usability of data. This encompasses the analysis of workflows, the structure and the type of data as scientific information (Borgman, 2015), as well as the need for the expertise implied in dealing with data as processes and objects. Today, open datasets on social networks abound, although curated collections of open datasets on social learning analytics are rare. Moreover, “data literacy” (Atenas, Havemann & Priego, 2015), such as researchers’ ability not only to create data but also to reuse open data, emerges as a concern for professional development and connected research. With regard to ethical issues, future research should focus on the models of de-identification, privacy and anonymity. A number of projects today are considering “differential privacy” as a means through which to act on data at a level where statistical models can be produced interacting with pre-processed data (Daries *et al.*, 2014). Another trend of research should further explore the legal constraints and the new regulations about privacy, identity protection and data property, as also reported in The Hague Declaration.

As regards security, today reference implementations of social learning analytics mainly delegate the process of collecting and filtering data to SNS. Therefore, the security and privacy layer has to be considered “as-is” rather than reflecting real needs or technological constraints (Caviglione, Coccoli & Merlo, 2014). Moreover, the lack of an open standard and limitations in the access of data often requires the use of “homemade” gathering procedures, such as screen scraping. This, jointly with the heterogeneity of security solutions, prevents from having a unique and coherent layer. Therefore, the advent of next-generation learning platforms with social capabilities must be meticulously engineered, from the phase of data generation, to data collection and the relating process of packaging for dataset exportation and sharing. In fact, they could add important bits of individuals and could also be used as targets to force the student online identity (e.g. by using profile fusion attacks). Nevertheless, exam schedules, selected courses and detailed descriptions of the relationships with classmates are definitely an effective source of information for attacks based on social engineering. For such reasons, the next-generation of users should be properly trained or guided by the software platform used for e-learning to avoid/prevent personal-detail-disclosing pitfalls.

## Conclusion

In this paper, we have presented some of the most relevant applications and

methodological issues of big data for the design of massive learning platforms and social media sites. We have showcased potential advancements of social learning analytics, pitfalls arising from the usage of open datasets, and privacy and security aspects. As discussed, the use of big data can surely represent an important advancement, but there is still the need to address a variety of issues, ranging from ethics to development of mechanisms to enforce users privacy and security.

## REFERENCES

---

- Anshari, M., Alas, Y., Hj Mohd Yunus, N., Pg Hj Sabtu, N., & Sheikh Abdul Hamid, M. (2016), *Online Learning: trends, issues and challenges in the Big Data Era*. Journal of E-Learning and Knowledge Society, 12(1), 121-134.
- Atenas, J., Havemann, L., & Priego, E. (2015), *Open Data as Open Educational Resources: Towards Transversal Skills and Global Citizenship*. Open Praxis, 7(4), 377-389.
- Borgman, C. (2015), *Big data, Little data, No data: Scholarship in the Networked World*. Cambridge, MA: The MIT Press.
- Buckingham Shum, S., & Ferguson, R. (2012), *Social learning analytics*. Educational Technology & Society, 15(3), 3-26.
- Caviglione, L. (2009), *Can satellites face trends? The case of Web 2.0*. In Proceedings of International Workshop on Satellite and Space Communications, 446-450.
- Caviglione, L., & Coccoli, M. (2011), *Privacy problems with Web 2.0*. Computer Fraud and Security, 10, 16-19.
- Caviglione, L., Coccoli, M., & Gianuzzi, V. (2011), *Opportunities, integration and issues of applying new technologies over e-learning platforms*. In Proceedings of the 3rd International Conference on Next Generation Networks and Services (NGNS 2011), 12-17.
- Caviglione, L., Coccoli, M., & Merlo, A. (2014), *A taxonomy-based model of security and privacy in online social networks*. International Journal of Computational Science and Engineering, 9(4), 325-338.
- Chanier, T. & Reffay, C. (2011), *Rapport Final du Projet MULCE*. URL: [http://mulce-doc.univ-bpclermont.fr/IMG/pdf/rapport\\_fin\\_de\\_projet\\_ANR\\_Mulce\\_110317.pdf](http://mulce-doc.univ-bpclermont.fr/IMG/pdf/rapport_fin_de_projet_ANR_Mulce_110317.pdf) (accessed on 4th March 2016).
- Coccoli, M., Guercio, A., Maresca, P., & Stanganelli, L. (2014), *Smarter Universities: a Vision for the Fast Changing Digital Era*. Journal of Visual Languages and Computing, 25(6), 1003-1011.
- Cooper, S., & Sahami, M. (2013), *Reflections on Stanford's MOOCs*. Communications of the ACM, 56(2), 28-30.
- Daries, J. P., Reich, J., Waldo, J., Young, E., Whittinghill, J., Seaton, D. T., Ho, A. D., & Chuang, I. (2014), *Privacy, Anonymity, and Big Data in the Social Sciences*.

- Queue – Practice, 12(7), 1-30.
- Esposito, A. (2012), *Research ethics in emerging forms of online learning: issues arising from a hypothetical study on a MOOC*. The Electronic Journal of e-Learning, 10(3), 315-325.
- Eynon, R. (2013). (2013), *The rise of Big Data: what does it mean for education, technology, and media research?* Learning, Media and Technology, 38(3), 237-240.
- Fan, W., & Gordon, M. D. (2014). The Power of Social Media Analytics. Communications of the ACM, 57(6), 74-81.
- Ferguson, R., & Clow, D. (2015), *Consistent commitment: Patterns of engagement across time in Massive Open Online Courses (MOOCs)*. Journal of Learning Analytics, 2(3), 55–80.
- Jändel, M. (2014), *Decision support for releasing anonymised data*. Computers & Security, 46, 48-61.
- Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2015), *NMC Horizon Report: 2015 Higher Education Edition*. Austin, TX: The New Media Consortium.
- Joksimović, S., Hatala, M., & Gašević, D. (2014), *Learning Analytics for Networked Learning Models*. Journal of Learning Analytics, 1(3), 191–194.
- Koedinger, K.R., Baker, R. S. J. d., Cunningham, K., Skogsholm, A., Leber, B. & Stamper, J. (2010), *A Data Repository for the EDM community: The PSLC DataShop*. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. d. Baker (Eds.). Handbook of Educational Data Mining (pp. 43-56). Boca Raton, FL: CRC Press.
- Kop, R. (2011), *The Challenges to Connectivist Learning on Open Online Networks: Learning Experiences during a Massive Open Online Course*. The International Review of Research in Open and Distance Learning, 12(3), 19-38.
- Long, P., & Siemens, G. (2011), *Penetrating the Analytics in Learning and Education*. EDUCASE review, September/October 2011, 31- 40.
- Mazurczyk, W., & Caviglione, L. (2014), *Steganography in Modern Smartphones and Mitigation Techniques*. IEEE Communications Surveys & Tutorials, 17(1), 334-357.
- Merceron, A., Blikstein, P., & Siemens, G. (2015), *Learning Analytics: From Big Data to Meaningful Data*. Journal of Learning Analytics, 2(3), 4-8.
- Pozzi, F. Manca, S., Persico, D., & Sarti, L. (2007), *A general framework for tracking and analysing learning processes in CSCL environments*. Innovations in Education & Teaching International, 44(2), 169-179.
- Selwyn, N. (2015), *Data entry: towards the critical study of digital data and education*. Learning, Media and Technology, 40(1), 64-82.
- Siemens, G. (2005), *Connectivism: Learning as network creation*. ElearnSpace. URL: <http://www.elearnspace.org/Articles/networks.htm> (accessed on 4th March 2016).
- Siemens, G. (2013), *Learning analytics: The emergence of a discipline*. American Behavioral Scientist, 57(10), 1380-1400.
- Slade, S., & Prinsloo, P. (2013), *Learning Analytics: Ethical Issues and Dilemmas*. American Behavioral Scientist, 57(10), 1510–1529.
- Tufekci, Z. (2014), *Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls*. In Proceedings of the Eight International

AAAI Conference on Weblogs and Social Media, 505-514.  
Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012), *Dataset-Driven Research to Support Learning and Knowledge Analytics*. *Educational Technology & Society*, 15(3), 133-148.