



## Understanding the potential and challenges of big data in schools and education

### Comprendiendo el potencial y los desafíos del Big Data en las escuelas y la educación

Arnon Hershkovitz 

e-mail: arnonhe@tauex.tau.ac.il

Tel Aviv University, Israel

Giora Alexandron 

e-mail: giora.alexandron@weizmann.ac.il

Weizmann Institute of Science, Israel

#### Abstract

In recent years, the world has experienced a huge revolution centered around the gathering and application of big data in various fields. This has affected many aspects of our daily life, including government, manufacturing, commerce, health, communication, entertainment, and many more. So far, education has benefited only a little from the big data revolution. In this article, we review the potential of big data in the context of education systems. Such data may include log files drawn from online learning environments, messages on online discussion forums, answers to open-ended questions, grades on various tasks, demographic and administrative information, speech, handwritten notes, illustrations, gestures and movements, neurophysiologic signals, eye movements, and many more. Analyzing this data, it is possible to calculate a wide range of measurements of the learning process and to support various educational stakeholders with informed decision-making. We offer a framework for better understanding of how big data can be used in education. The framework comprises several elements that need to be addressed in this context: defining the data; formulating data-collecting and storage apparatuses; data analysis and the application of analysis products. We further review some key opportunities and some important challenges of using big data in education.

*Keywords:* decision making; electronic data processing; teaching aid; educational administration.

#### Resumen

En los últimos años, el mundo ha experimentado una gran revolución centrada en la recopilación y aplicación de *big data* en varios campos. Esto ha afectado muchos aspectos de nuestra vida diaria, incluidos el gobierno, la manufactura, el comercio, la salud, la comunicación, el entretenimiento y muchos más. Hasta ahora, la educación se ha beneficiado muy poco de la revolución del *big data*. En este artículo revisamos el potencial de los macrodatos en el contexto de los sistemas educativos. Dichos datos pueden incluir archivos de registro extraídos de entornos de aprendizaje en línea, mensajes en foros de discusión en línea, respuestas a preguntas abiertas, calificaciones en diversas tareas, información demográfica y administrativa, discurso, notas escritas a mano, ilustraciones, gestos y movimientos, señales neurofisiológicas, movimientos oculares y muchos más. Analizando estos datos es posible calcular una amplia gama de mediciones del proceso de aprendizaje y apoyar a diversos interesados educativos con una toma de decisiones informada. Ofrecemos un marco para una mejor comprensión de cómo se puede utilizar el *big data* en la educación. El marco comprende varios elementos que deben abordarse en este contexto: definición de los datos; formulación de aparatos de recolección y almacenamiento de datos; análisis de datos y aplicación de productos de análisis. Además, revisamos algunas oportunidades clave y algunos desafíos importantes del uso de *big data* en la educación.

*Palabras clave:* toma de decisiones; procesamiento electrónico de datos; ayuda a la enseñanza; administración educacional.

*Recibido / Received:* 13-05-2019

*Aceptado / Accepted:* 23-07-2019

*Publicación en línea / Published online:* 20-12-2019

Cómo referenciar este artículo / How to reference this article:

Hershkovitz, A. & Alexandron, G. (2020). Understanding the potential and challenges of Big Data in schools and education. *Tendencias Pedagógicas*, 35, pp. 7-17. doi: 10.15366/tp2020.35.002

## 1. Introduction: Between two classes

At "Fields of Yesterday" school, the teacher tells 8<sup>th</sup> grade students all about separating mixtures with chromatography, showing them a short BrainPOP clip. Some students listen, some don't. Some of those who listen understand, but some don't (though they are ashamed to ask). The teacher gives them home assignments designed for pairs of students that they have chosen. When the learning period is done, the teacher gives the students a test and they write their answers on paper, during class. Two weeks later, when she returns the inspected test sheets, each carries a grade — a digit on the front page — and some feedback hiding inside. Some of those school children no longer remember what the test was about. They now study electric energy.

At the very same time in the very same town, 8th grade students at "Tomorrow's Buds" school work on chromatography in an online unit that is part of an integrative learning system. While using it, the system automatically and continuously collects data that documents the students' activity, and analyzes it in order to recognize the students' level of knowledge. Furthermore, using this data, the system identifies students who feel frustrated, confused or bored during the learning process. While the students are busy with computerized simulations and experiments, their teacher receives visual information on her screen with clear progress reports of their achievements and affective state, as well as information about tasks that were unexpectedly difficult for the majority of the students. Using this information, she decides which students need her attention during the class, with whom she will contact after school hours, and how to adjust her teaching plan. While she walks around in the classroom, some students call the teacher over and ask questions they are ashamed to pose in front of everybody (therefore, it is of no surprise that surveys that were taken in that school showed an improvement in the classroom atmosphere, as well as in teacher-student relationship, in this teacher's classroom). The progress reports (data) help the teacher divide the students into pairs for their home assignments. When the learning period ends, there is no need for an exam because the system computes individual students' activities, offering precise and extensive information, including aspects such as creative thinking and persistence.

If you find the description of the Buds classroom imaginary, think again! The described systems are currently under academic and industrial R&D processes and some are already used in classrooms worldwide; these systems are based on big-data methods for the real-time analysis of large bodies of data, as we will describe in this paper. Information extracted from such data could greatly help various education parties such as students, school staff, parents, directors of all levels, and developers of digital learning environments. Such information could help improving decision-making and learning processes.

## 2. What is big data?

In recent years, the world experienced a technological, social, and cultural revolution centered around the gathering and application of big data in various fields. The revolution is based on the combination of two factors: volumes of data (with unique characteristics, as we explain below) and strong and readily available computation powers; but mainly on the realization that both could be used for various purposes. Big data is the result of extensively used, networked, and online digital systems in almost every field of our modern life — the media, finance, trade, culture, security, transportation, and so on. Once activated, these systems save their users' digital history. For example, every credit-card transaction we do online is saved in the credit company's computers. The companies then use that information to make tailor-made offers to their clients, for example. Clients are pleased with the personalised service they receive, and the companies increase the chances of turning offers into actual deals. The offers are based on advanced algorithms that universities and commercial firms have been developing, turning extensive data into valuable information.

In recent years, this field is being applied in education, mainly due to the extended usage of digital learning environments and learning management. Computerized environments such as Moodle, Courseware, MOOCs (Massive Online Open Courses), and school management systems constantly store extensive and valuable data that can be employed to improve learning and teaching processes. Over the

past decade, multidisciplinary studies in this field have been established and are currently known as Learning Analytics.

By the definition adopted by SoLAR (Society for Learning Analytics Research; <http://solaresearch.org>), learning analysis is the measurement, collection, analysis and reporting of data on learners and their contexts in order to understand and improve learning and their environments (Simon, 2017, p. 200). That general definition directly corresponds with the definition of Learning Sciences as it appears on the internet site of The International Society of the Learning Sciences: "The interdisciplinary empirical investigation of learning as it exists in real-world settings, and how learning may be facilitated both with and without technology" (Packer & Maddox, 2016, p. 131). It is, however, generally agreed that Learning Analytics differ from Learning Sciences (or any other study of learning) by two characteristics: said data is generally collected automatically from (mostly online) digital learning systems; and the amount of data far exceeds the volumes that education researchers usually collect.

This data may be collected by different means and from different sources (Papamitsiou & Economides, 2014). Specifically, it is very common to use log files drawn from online learning environments, in which students' actions (with different levels of granularity) are continuously and automatically stored. These actions characterize students' activity in three dimensions: the action taker (who?), the action itself (what?), and the action time (when?). Analyzing this data, it is possible to calculate a wide range of measurements of the learning process, including identifying students' emotional states (Paquette et al., 2014). In addition to log files, data can be collected and analyzed from messages on online discussion forums, answers to open-ended questions, grades on various tasks, and more. Demographic and administrative information, usually stored in schools' (other) systems, can also enrich this process. In recent years, multimodal data — such as speech, handwritten notes, illustrations, gestures and movements, neurophysiologic signals, and eye movements — have been also used to analyze learning (Blikstein & Worsley, 2016).

It is, therefore, important to understand that the use of big data is but one of numerous tools that could help generate significant insights through empirical studies and, when applied, such insights could help all interested parties without them "getting their hands dirty" with data processing. Below, we discuss both research and application aspects of using data in education, presenting the potential and challenges inherent in using big data in education.

### *2.1. Big data — Characteristics and examples from education*

Big data is generally defined by the characteristics that distinguishes it from "ordinary" data. Initially, they were dubbed "the three Vs": Volume, Velocity, and Variety; Veracity was added as a fourth V, and the fifth, Value, came in later.

**Volume:** As its name indicates, big data is primarily characterized by its extensive volume. For now, let us focus on activity records of online learning systems. Every operation in the system is documented as an "event" in a log file. Several minutes of a student's activity learning analysis is the measurement, collection, analysis and reporting of data on learners and their contexts in order to understand and improve learning and their environments — watching a video clip, running a simulation, or answering several questions — could generate several hundred documented activities. Multiplying that by the number of students active in the system and by the duration of those activities produces huge numbers. For example, a pilot run conducted with 650 4th grade students who used a calculus learning system for 90 minutes a week, for 4–8 weeks, logged some 2 million documented events (Alexandron, Keinan, Levy, & HersHKovitz, 2018). Another popular learning environment is MOOC, where a single course can accumulate millions of operations that document learners' activities (Kahan, Soffer, & Nachmias, 2017; Pardos & Xu, 2016). A log-based study that uses a popular online learning environment, like Khan Academy, can easily recruit dozens of thousands of learners (Gal & HersHKovitz, 2019), a number that is larger by a few orders of magnitude than population sizes that appear in studies that implement traditional (quantitative) research methodologies.

**Velocity:** Another big data characteristic is the speed in which it accumulates. Numerous data is accumulated during a single course or operation and, as learning scenarios become more common, data-

collecting speed increases. For example, the number of MOOCs has been consistently rising recently with thousands of new courses added each year (Almanac 2017, 2017; Shah, 2016). Online learning systems expand too and serve increasingly more learners each year<sup>1</sup>. The ASSISTments System of math learning and practice for school ages, for example, has doubled its users every year for a decade (Heffernan & Heffernan, 2014).

**Variety:** Diversity is the third characteristic. This refers to the fact that the accumulated data is of various kinds. In learning systems, such data can document different system applications such as interaction with clips, answering questions, the correctness of answers, notices published in discussion groups, file downloads, and more. In recent years, various types of data are being gathered, including recorded speech, hand-written texts, drawings, bodily gestures, emotional states, neurophysiological indices, eye movements, and so on (Blikstein & Worsley, 2016). This data can be gathered automatically and analyzed with calculative methods, together forming a significant challenge for researchers.

**Veracity:** Big data is further characterized by its veracity or precision; that is, the extent to which collected data reliably represents what it was designed to measure. This is particularly important in educational contexts. Knowledge is a latent variable that cannot be measured directly. Identifying learning activities by analyzing the interactions between learners and computers is, therefore, a serious challenge. For example, it should be obvious that the fact that a certain file has been downloaded does not mean it was actually read, or the fact that time had elapsed between "play" and "pause" does not mean a video was actually watched; also, a wrong answer does not always attest to a lack of knowledge, and a correct answer is not necessarily an evidence for knowing the material (Alexandron, Ruipérez-Valiente, Chen, Muñoz-Merino, & Pritchard, 2017; Corbett & Anderson, 1995; HersHKovitz, de Baker, Gobert, Wixon, & Pedro, 2013). This is why engagement in learning — a basic and significant variable in educational research — is not necessarily measured by simple data analysis (Fincham et al., 2019).

**Value:** Finally, the data gathered needs to be of value. In our case, value can be measured by the way the data explains or predicts learning behaviors. Namely, value is not an inherent part of the data, but associated with its application. For example, data describing activities in online learning systems could have value as an early predictor of success in a course or earning an academic degree, and even predict higher education attendance years in advance (Manrique, Nunes, Marino, Casanova, & Nurmikko-Fuller, 2019; San Pedro, Baker, Bowers, & Heffernan, 2013; Ye, Biswas, & Biswas, 2016).

## *2.2. A framework for understanding the use of big data in education*

For the use of big data in education to be effective, several complementary elements are required. These steps are not necessarily to be taken by the same individual, but rather they describe a multi-level process that usually requires bringing together of practitioners, policy makers, engineers and researchers, among other stakeholders. We present these components as if they were part of a sequential process, although in practice this is not necessarily the case. Rather, these components operate in a circular process where each stage may give feedback to earlier stages so as to optimize them and thus optimize the entire analysis process.

First, we need to **define the data to be collected** referring to the goal of the relevant interested party and relevant data sources. Let us consider an educational administration that wishes to gain a-priori information on potential dropouts. This organization needs to pinpoint potentially relevant data, which might be: social background and demographic student data; grades and deadlines met by first-year students; class attendance (Adelman, Haimovich, Ham, & Vazquez, 2018; Knowles, 2015). Understandably, such data may be found in a wide variety of databases. In another example, a company that develops learning materials wishes to examine their effectiveness. Such a company may want to collect data on learners' interactions with their systems and their grades in computerized and traditional

---

<sup>1</sup> <https://www.class-central.com/report/mooc-stats-2017>

learning environments (Richards & DeVries, 2012). It should be stressed at this point that although the definition of the collected data should be justified by the literature and studies citing connections between the goal variable and the data collected, collecting data whose end use is not clear when the system is built makes sense too if we wish to allow flexibility in the use of the data for research and creating new knowledge and insights. Finding a balance between these two approaches (collecting "only what is needed" vs. "everything possible") is a challenge with research, engineering, and ethical aspects.

Once the data to be collected is defined, we need to **construct an apparatus for collecting and saving the data**. As noted earlier, data could come from various sources and thus, in various formats. Therefore, it is required to make sure that different representations of the same data are identifiable as one. For example, names could appear in different forms (e.g., "Cristóbal Cobo", "Cristóbal Cobo Romani", "Cobo Cristóbal", "Cristóbal C.", etc.), as well as dates ("1-May-2019", "May 1, 2019", "1/5/19", "5/1/19", etc.) and even gender ("F", "Female", "Woman", or even "2" as an arbitrary numerical representation). Storing this data, one should keep in mind legal and ethical aspects of privacy and of data storing and sharing.

Once they are gathered and stored, **data analysis methods need to be developed**, and there is an extensive variety of such methods (Lang, Siemens, Wise, & Gasevic, 2017; Romero, Ventura, Pechenizkiy, & Baker, 2010). Generally speaking, we have supervised and unsupervised data analysis. In the supervised method, we have the goal variable to which the analysis refers (e.g., analyzing dropout or successful learning when "dropout" and "success" are the variables). There is no such variable in the unsupervised method (e.g., dividing students into learning groups based on the way they completed assignments previously). Such an analysis may work in real time in — for example, an online learning environment that measures updated usage data against past uses; or in other scenarios — for example, in analyzing students' activities at the end of the school year.

Finally, the analyzed data needs to be **used by humans or machines**. For data-based information to be accessible by humans, a mediating system that presents the data in a user-friendly manner is required. This is generally done with the help of an educational dashboard, which makes data about the learners accessible to various stakeholders (e.g., students, teachers, school management, etc.) in an effective way (Verbert, Duval, Klerkx, Govaerts, & Santos, 2013; Yoo, Lee, Jo, & Park, 2015). People who need to make decisions based on data must, of course, understand the data they are presented as well as its usage. If the data should serve decision-making machines, a system must be developed that knows how to do that. In that context, there are three categories of analytics that will be demonstrated here using the dashboard example: (1) Descriptive — the dashboard presents a summary of the classroom status, e.g., using a graph from which it is easy to understand that about third of the students achieve low scores in a certain topic, or using real-time visualization of the students' frustration level; (2) Predictive — the dashboard presents a prediction of future scenarios, for example, indicating students that may find the next topic too difficult, or a group of students that may be quite frustrated while practising it; (3) Prescriptive — the dashboard presents actionable information for the teacher, like pointing out content that may be personally tailored to each student, recommending alternative pedagogy, or suggesting a division of the classroom into groups in a way that will maximize their learning.

A recent intervention scheme for at-risk college students demonstrates this multiple-component process nicely (albeit not in the very order described here). This scheme, called iCLAS (integrated Closed-loop Learning Analytics), aims at recognizing and helping students who are not thriving, as measured by low achievements at an early stage of the course (Syed et al., 2019). While focusing on identifying at-risk students and boosting their success, the authors' main focus was on measuring engagement, which corresponds to our component of defining the data to be collected; the authors operationalized this construct by capturing data about the students' online activity in the learning environment (e.g., logging in and out, clicking on resources, attempting and submitting assignments, amount of time spent on watching course videos, and performance indicators). The authors then detail how the data was captured via multiple systems, such as Sakai (a learning management system), Panopto (a video platform), and Apereo (a learning record warehouse), and was merged using Tableau (data analysis and visualization software), which corresponds to our component of apparatus for collecting and saving the data. The authors report basing their identification of non-thriving students on knowledge-engineered models they

developed (based on statistical analyses of both the captured data and historical data), which corresponds to our component of data analysis methods. Finally, the authors report on the crucial part of their holistic approach — which is using the data by notifying potentially non-thriving students and their advisors — and on the actions these stakeholders had taken, corresponding to our component of using the data. As the authors explicitly mention, their scheme is not linear and is cyclic, with the aim of improving the design of the course and the design of the scheme itself.

### 3. Big data in education — Potential

There is not enough space here to describe the full potential of big data applications in education. Still, below we present a brief review of the four main aspects where we identified an acceleration thanks to data usage.

#### 3.1. Student assessment

A prominent example of the use of data that describe learning is their potential application to student assessment methods. Currently, student assessment is mainly based on exams given when the learning period ends, using only achievement indices. The use of big data, on the other hand, could make assessment part of the learning process, offering students and teachers real-time feedback — giving the former insights into their own learning processes and the latter, updated information on what goes on in their classrooms (Fancsali & Ritter, 2018; Holmes, 2018; Klein & Hess, 2018). Such evaluation methods could consider not only students' knowledge but also information on, for example, the number of attempts made before an answer was found, work frequency and location, cooperation with peers, and more (Berland, Baker, & Blikstein, 2014). Naturally, in addition to using big data for the assessment of cognitive capabilities, they can also help assessing effective and meta-cognitive aspects of the learning process (Paquette et al., 2018).

#### 3.2. Individual adjustment

Personalized learning — based on each student's competencies, needs, emotions and interests — has been discussed in the literature for over four decades (Glaser, 1977). Big data enables such mechanisms in practice, for modifying and optimizing individual learning patterns for each student, using the algorithms that analyze the students' learning history and relying on insights gained from the histories of "like-minded" students. Similar ideas have been successfully implemented in entertainment platforms, like Netflix, Spotify, or YouTube, where the users enjoy personally-tailored, constantly adapted content.

This type of algorithm-based adjustment is known as adaptive learning. In learning methods that include a large element of independent learning with computers, the computer-learner interaction becomes more important. Tools that make use of natural-language processing algorithms could help humanizing that interaction; offer immediate feedback to open-ended questions that thus far was delayed until a teacher reviewed them; identify frustration or inattention, and so on (Hershkovitz et al., 2013). Such personal adaptations would hopefully help increase the students' involvement and learning motivation. Initial evidence of the effectivity of adaptive learning systems is already beginning to accumulate (Graesser, 2016; Herder, Sosnovsky, & Dimitrova, 2017).

Mechanisms for personalized learning are already in use in a number of commercial products, for example, of ETS (<http://www.ets.org>), Fulcrum Labs (<https://www.fulcrumlabs.ai>), or Surgent (<http://www.surgentcpareview.com>). The Summit Learning charter schools (in the US), an organization formed a few years ago with the support of Bill and Melinda Gates Foundation and of Facebook founder, Mark Zuckerberg (among others), relies on an adaptive learning platform. Of course, adapting learning to a student will become optimal when multiple student characteristics are taken into consideration — not just knowledge, but also behavior (e.g., number of failed attempts until a correct answer, or

persisting in improving a correct answer), content preferences, affective state, and more (Chrysafiadi & Virvou, 2013) — just as the teacher refers to a student as a multidimensional individual.

### *3.3. Data-supported decision-making*

The application of data is not reserved only for teachers or students. Principals and directors of educational facilities could use big data to identify drop-out patterns or to evaluate the effectiveness of a curriculum; since collecting and analyzing data is often done at higher levels (mostly due to infrastructure, budget, and professional knowledge-related issues), this issue is of great importance for educational policy makers (Ferguson et al., 2016; Tsai, Moreno-Marcos, Tammets, Kollom, & Gašević, 2018). Additionally, insights from such analyses could be of great use to improve educational content, as usage patterns can indicate on poorly or ineffectively used components (e.g., Richards & DeVries, 2012).

### *3.4. Big data as an education research tool*

The big data age offers wonderful opportunities for education researchers. They can use big data to study traditional learning-process issues with advanced methods. For example, they can build mathematical models that quite accurately outline students' progress on a given issue and suggest ways teachers could use to promote each student. In another example, combining data from various sources — such as math learning courseware, school management systems, or systems where students report how they feel while learning — could predict which of the students might encounter hardships and intervene in time. Big data can help monitor learning processes at very high resolutions and precisely pinpoint the moment of understanding. At the same time, the big data revolution lends itself to studies of new problems that emerge in computer-based learning. For example, using courseware that offers help to allow students to attain correct answers even if they had not made an effort to think the questions through (by first pulling the final clue that offers the answer), which might impair their learning process (Cocca, Hershkovitz, & Baker, 2009); or MOOCs, where the person filling the solutions does not have to be the learning student that the system "believes" he is (Alexandron et al., 2017), solving this by using accumulated data that could reveal types of behavior.

## **4. Big data in education — Challenges**

The revolution of big data in education poses numerous challenges in both exhausting the potential contained in the data, and in ensuring it is used properly. Below, we address several key challenges.

### *4.1. Systemic changes*

Effective learning environments correctly combine learning materials with pedagogic knowledge and the physical environment — which need to be redesigned so as to optimize data-based tools. Specifically, for big data to be used, it must be collected massively; namely, digital learning environments need to be used more frequently. That, in turn, requires some rethinking of pedagogic and physical environments. Thus, teachers will have to acquire new skills, including the ability to interpret data representations and subsequently take learning-promotion steps. That is to say that the fact that computer-based learning systems should be used more frequently if we want to collect numerous data could encourage the rethinking of learning systems and the retraining of teachers in the application of such systems (Collins & Halverson, 2018).

### *4.2. Modeling knowledge structures*

Another challenge relates to what happens "behind the scenes" of big-data-based algorithms. For example, for an algorithm to help advanced students of a computerized algebra course it must "understand" the structure of algebra knowledge so that it could match questions with that structure and

"know" which topics could help advancing other topics. This is the only way in which the system — having identified students' weaknesses and strengths in specific issues — could match students with their optimal learning courses. This challenge is deeply connected with AI, which deals with knowledge modeling and representation, and with developing "smart" decision-making processes based on it (Nkambou, 2010).

#### 4.3. Technological infrastructures

The public, state-sponsored education system in the State of Israel (as in most countries, including developed ones) lacks in technological infrastructures (end-user equipment, wireless broadband networks, etc.). This actually creates a vicious cycle: in the absence of adequate infrastructure, it is hard to opt for technology-based pedagogical novelties; this, in turn, has a cooling effect on the feasibility of developing such advanced learning systems; and since these are absent, there is low motivation to invest in seemingly useless technologies; and on it goes.

On top of infrastructure, the technology required for making the revolution of big data in education possible poses its own challenges. Specifically, the lack of standardization is what makes it quite difficult. Different systems collect different data and save it under a variety of formats, which makes it hard to develop generic tools and methods that could be used with all the systems on the market (del Blanco, Serrano, Freire, Martinez-Ortiz, & Fernandez-Manjon, 2013).

#### 4.4. Economic aspects

Naturally, these challenges carry some serious economic implications because they entail serious financial investments. Given that most of the investment in education and learning is public and that the cost structure is generally rigid, finding applicable economic models for the big data revolution in education is a serious challenge. This is twice as true when dealing with unique linguistic and curricula characteristics (e.g., the Israeli system) because they rarely can benefit from research and technological developments for larger markets.

#### 4.5. Ethical considerations

Finally, the use of big data is associated with significant ethical issues pertaining to the collection and use of information. In education, these issues have certain unique characteristics (Ho, 2017). Two of those are associated with the age of users and the fact that using such systems in schools is mandatory. First, school students are minors, which is significant when we examine their ability to consent to having their data — the topic of this article — collected at all. Second, learning systems are usually used as part of mandatory curricula, which is why we need to ask what to do when students refuse to have their data collected. This calls for a set of clear and applicable rules, backed up by the appropriate technology and methodology, to ascertain that the data collected is used appropriately and serves only for the betterment of education and learning.

## References

- Adelman, M., Haimovich, F., Ham, A., & Vazquez, E. (2018). Predicting school dropout with administrative data: New evidence from Guatemala and Honduras. *Education Economics*, 26(4), 356–372. doi: doi.org/10.1080/09645292.2018.1433127.
- Alexandron, G., Keinan, G., Levy, B., & Hershkovitz, S. (2018). Evaluating the effectiveness of animated cartoons in an intelligent math tutoring system using educational data mining. In *EdMedia + Innovative Learning (Vol. 2018, pp. 719–730)*. Waynesville, NC.
- Alexandron, G., Ruipérez-Valiente, J. A., Chen, Z., Muñoz-Merino, P. J., & Pritchard, D. E. (2017). Copying@scale: Using harvesting accounts for collecting correct answers in a MOOC. *Computers and Education*, 108, 96–114. doi: doi.org/10.1016/j.compedu.2017.01.015.

- Almanac 2017. (2017). Cumulative growth in number of MOOCs 2011–2017. *The Chronicle of Higher Education*.
- Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning*, 19(1–2), 205–220. doi: doi.org/10.1007/s10758-014-9223-7.
- Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: Using Computational Technologies to Measure Complex Learning Tasks. *Journal of Learning Analytics*, 3(2), 220–238. doi: doi.org/10.18608/jla.2016.32.11.
- Chrysafiadi, K., & Virvou, M. (2013). Student modeling approaches: A literature review for the last decade. *Expert Systems with Applications*, 40(11), 4715–4729. doi: doi.org/10.1016/J.ESWA.2013.02.007.
- Cocca, M., Hershkovitz, A., & Baker, R. S. J. D. (2009). The impact of off-task and gaming behaviors on learning: Immediate or aggregate, The International Conference on Artificial Intelligence and Applications. doi: doi.org/10.3233/978-1-60750-028-5-507.
- Collins, A., & Halverson, R. (2018). *Rethinking education in the age of technology: The digital revolution and schooling in America (second edition)*. New York, NY: Teachers College Press.
- Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modelling and User-Adapted Interaction*, 4(4), 253–278. doi: doi.org/10.1007/BF01099821.
- del Blanco, A., Serrano, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2013). E-Learning Standards and Learning Analytics. *IEEE Global Engineering Education Conference*, 1255–1261. doi: doi.org/10.1109/EduCon.2013.6530268.
- Fancsali, S. E., & Ritter, S. (2018). Assessments that care about student learning. In N. Guin & A. Kumar (Eds.), *14th International Conference on Intelligent Tutoring: Workshop Proceedings (pp. 135–140)*. Montreal, QC.
- 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*. Seville, Spain: Publ. Off. of the Europ. Union.
- Fincham, E., Whitelock-Wainwright, A., Kovanović, V., Joksimović, S., van Staaldunin, J. P., & Gašević, D. (2019). Counting clicks is not enough: Validating a theorized model of engagement in learning analytics. In *the 9th International Learning Analytics & Knowledge Conference (pp. 501–510)*. Tempe, AZ. doi: doi.org/10.1145/3303772.3303775.
- Gal, T., & Hershkovitz, A. (2019). Different response-based feedback in mathematics: The case of symbolic and textual messages. In *The 9th International Learning Analytics & Knowledge Conference*. doi: doi.org/10.1145/3303772.3303815.
- Glaser, R. (1977). *Adaptive education: Individual diversity and learning*. Holt, Rinehart and Winston.
- Graesser, A. C. (2016). Conversations with AutoTutor help students learn. *International Journal of Artificial Intelligence in Education*, 26(1), 124–132. doi: doi.org/10.1007/s40593-015-0086-4.
- Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and Teaching. *International Journal of Artificial Intelligence in Education*, 24(4), 470–497. doi: doi.org/10.1007/s40593-014-0024-x.
- Herder, E., Sosnovsky, S., & Dimitrova, V. (2017). Adaptive intelligent learning environments. In D. E., S. M., & S. R. (Eds.), *Technology Enhanced Learning (pp. 109-114)*. Cham, Switzerland: Springer. doi: doi.org/10.1007/978-3-319-02600-8\_10.
- Hershkovitz, A., de Baker, R. S. J. R. S. J., Gobert, J., Wixon, M., & Pedro, M. S. M. S. (2013). Discovery with models: A case study on carelessness in computer-based science inquiry. *American Behavioral Scientist*, 57(10), 1480–1499. doi: doi.org/10.1177/0002764213479365.
- Ho, A. (2017). *Workshop on big data in education: Balancing the benefits of educational research and student privacy*. Washington, DC.
- Holmes, N. (2018). Engaging with assessment: Increasing student engagement through continuous assessment. *Active Learning in Higher Education*, 19(1), 23–34. doi: doi.org/10.1177/1469787417723230.

- Kahan, T., Soffer, T., & Nachmias, R. (2017). Types of participant behavior in a massive open online course. *International Review of Research in Open and Distance Learning*, 18(6), 1–18. doi: doi.org/10.19173/irrodl.v18i6.3087.
- Klein, C., & Hess, R. M. (2018). Using learning analytics to improve student learning outcomes assessment: Benefits, constraints, & possibilities. In A. Lester, C. Klein, A. Johri, & H. Rangwala (Eds.), *Learning Analytics in Higher Education* (pp. 140–159). New York, NY: Routledge. doi: doi.org/10.4324/9780203731864-7.
- Knowles, J. (2015). Building an accurate statewide dropout early warning system in Wisconsin. *Journal of Educational Data Mining*, 7(3), 18–67.
- Lang, C., Siemens, G., Wise, A., & Gasevic, D. (Eds.). (2017). *Handbook of Learning Analytics*. Society for Learning Analytics Research (SoLAR). doi: doi.org/10.18608/hla17.
- Manrique, R., Nunes, B. P., Marino, O., Casanova, M. A., & Nurmikko-Fuller, T. (2019). An analysis of student representation, representative features and classification algorithms to predict degree dropout. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge - LAK19* (pp. 401–410). doi: doi.org/10.1145/3303772.3303800.
- Nkambou, R. (2010). Modeling the domain: An introduction to the expert module. In N. R., B. J., & M. R. (Eds.), *Advances in Intelligent Tutoring Systems* (pp. 15–32). Berlin, Germany: Springer. doi: doi.org/10.1007/978-3-642-14363-2\_2.
- Packer, M. J., & Maddox, C. (2016). Mapping the territory of the learning sciences. In M. A. Evans, M. J. Packer, & R. K. Sawyer (Eds.), *Reflections on the Learning Sciences* (pp. 126–154). New York, NY: Cambridge University Press. doi: doi.org/10.1017/cbo9781107707221.007.
- Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society* (Vol. 17).
- Paquette, L., Baker, R. S. J. D., Sao Pedro, M. A., Gobert, J. D., Rossi, L., Nakama, A., & Kauffman-Rogoff, Z. (2014). Sensor-free affect detection for a simulation-based science inquiry learning environment. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8474 LNCS, pp. 1–10). Springer, Cham. doi: doi.org/10.1007/978-3-319-07221-0\_1.
- Paquette, L., Biswas, G., Baker, R. S., Rajendran, R., Munshi, A., & Ocumpaugh, J. (2018). Modeling learners' cognitive and affective states to scaffold SRL in open-ended learning environments. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization* (pp. 131–138). Singapore. doi: doi.org/10.1145/3209219.3209241.
- Pardos, Z. A., & Xu, Y. (2016). Improving efficacy attribution in a self-directed learning environment using prior knowledge individualization, 435–439. doi: doi.org/10.1145/2883851.2883949.
- Richards, G., & DeVries, I. (2012). Revisiting formative evaluation. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge – LAK '11*, 157. doi: doi.org/10.1145/2090116.2090141.
- Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. J. d. (Eds.). (2010). *Handbook of Educational Data Mining - CRC Press Book*. Boca Raton, FL: CRC Press. doi: doi.org/10.1201/b10274.
- San Pedro, M. O. Z., Baker, R. S. J. d., Bowers, A. J., & Heffernan, N. T. (2013). Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In *The Sixth International Conference on Educational Data Mining* (pp. 177–184). Memphis, TN.
- Shah, D. (2016). *By the numbers: MOOCS in 2016*. Retrieved from <https://www.classcentral.com/report/mooc-stats-2016/>.
- Simon, J. (2017). A priori knowledge in learning analytics. In A. Peña-Ayala (Ed.), *Learning Analytics: Fundamentals, Applications, and Trends* (Vol. 94, pp. 199–227). Cham, Switzerland: Springer. doi: doi.org/10.1007/978-3-319-52977-6.
- Syed, M., Anggara, T., Lanski, A., Duan, X., Ambrose, G. A., & Chawla, N. V. (2019). Integrated closed-loop learning analytics scheme in a First Year Experience course. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 521–530). Tempe, AZ. doi: doi.org/10.1145/3303772.3303803.

- Tsai, Y.-S., Moreno-Marcos, P. M., Tammets, K., Kollom, K., & Gašević, D. (2018). SHEILA policy framework. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge – LAK '18* (pp. 320–329). New York, New York, USA: ACM Press. doi: doi.org/10.1145/3170358.3170367.
- Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10), 1500–1509. doi: doi.org/10.1177/0002764213479363.
- Ye, C., Biswas, G., & Biswas, G. (2016). Early prediction of student dropout and performance in MOOCs using higher granularity temporal information. *Journal of Learning Analytics*, 1(3), 169–172. doi: doi.org/10.18608/jla.2014.13.14.
- Yoo, Y., Lee, H., Jo, I.-H., & Park, Y. (2015). Educational dashboards for smart learning: Review of case studies. In *Emerging Issues in Smart Learning* (pp. 145–155). Berlin: Springer-Verlag. doi: doi.org/10.1007/978-3-662-44188-6\_21.