Introduction to Big Data in Education and Its Contribution to the Quality Improvement Processes

Christos Vaitsis[1]*, Vasilis Hervatis[1] and Nabil Zary[1, 2]

Show details

Abstract

In this chapter, we introduce the readers to the field of big educational data and how big educational data can be analysed to provide insights into different stakeholders and thereby foster data driven actions concerning quality improvement in education. For the analysis and exploitation of big educational data, we present different techniques and popular applied scientific methods for data analysis and manipulation such as analytics and different analytical approaches such as learning, academic and visual analytics, providing examples of how these techniques and methods could be used. The concept of quality improvement in education is presented in relation to two factors: (a) to improvement science and its impact on different processes in education such as the learning, educational and academic processes and (b) as a result of the practical application and realization of the presented analytical concepts. The context of health professions education is used to exemplify the different concepts.

1. Introduction

Higher and professional education is a domain which constantly needs to be evaluated and transformed to follow the fast pace of changing trends in different sectors in the market which in turn creates a variety of needs in workforce. A major factor that has radically altered the way education is conducted is technology. Examples of different types of technologies used in education are mobile devices and apparatuses, teleconference and remote access systems, educational platforms and services and other that students, teachers, academic faculty, evaluation specialists, researchers and decision-makers in education interact with and use in an effort to impact and improve teaching and learning but also to realistically reflect in the learning stage the usage of modern technologies used in real settings. The interaction with these technologies generates large amounts of data that range from an individual access log file to an institutional level activity. Still the educational systems are not yet fully prepared to cope with and exploit them for continuous quality improvement purposes. In particularly, health professions education or health education is a context that these technologies are predominantly used, producing a wide range of educational data. In addition, health education is in constant need of reflecting the growing body of medical knowledge and evidence in order to practically embed it in education and prepare the future health professionals to meet the future challenges of healthcare systems. The need to govern these challenges within health education is now more than ever timely, and therefore, attention has been paid to different approaches such as big data and analytics that could be useful in investigating and exploiting educational data too.

2. Big data and education

2.1. BIG DATA

Big data is extensively used as a term today to describe and define the recent emergence and existence of data sets of high magnitude. It can be found in many sectors. The public, commercial and social sectors receive and produce ceaselessly vast amounts of data from different sources and in different formats. In some cases, the data reach extremely big sizes such as in petabytes exceeding the hardware or human abilities to warehouse, manipulate and process them and therefore is characterized as big data. Nevertheless, this term has been readily given to large sized data, although the size can vary from sector to sector or more specifically between services within a sector [1]. Big data is in fact termed as such given its characteristic of being large in size. Nevertheless, big data is defined by additional characteristics such as the disparate types and formats and different sources the data are collected from but also the speed they are produced, and most importantly, the frequency they are processed, in real time, frequently or occasionally. All these characteristics are summarized as volume (size), variety (sources, formats and types) and velocity (speed and frequency) and add complexity to the data, which is in fact another attribute in concern [2]. Data possessed in a system or a specific domain are considered as big data when simultaneously the volume, the variety and the velocity are high irrespective of whether these three characteristics can be considered "small" to another domain. In this case, this is enough to challenge constrains in manipulating and analysing the data so they can be used for different purposes. Depending on the domain, the size of data can vary from megabytes to petabytes. Thus, big data is context-specific and may refer to different sizes and types from domain to domain but the common challenge that all these domains must cope with is to being able to make sense of the data by processing them in a high analytical level to enable data-driven improvement of processes and procedures [3]. Big data and analytics have added value to data possessed in different contexts and consequently have proven to be an extremely useful approach for investigating its possible impact either in industry in the form of business intelligence and analytics [4] or in academia with educational data mining techniques and learning analytics [5]. Given the limited research on the usage of big data and analytics in the context of health education, we will introduce the reader to the new field of big educational data which places big data in education and how the educational data can be treated in different dimensions and from different perspectives to bring into light insights for different stakeholders such as decision-makers, academic faculty, evaluation specialists, researchers and students in computer science, engineering and informatics courses and encourage accordingly data-driven activities concerning quality improvement in education.

2.2. BIG EDUCATIONAL DATA

One of the domains that volume, variety and velocity coexist in the data is the higher education. Large amounts of educational data are captured and generated on a daily basis from different sources and in different formats in the higher educational ecosystem. The educational data vary from those produced from students' usage and interaction with learning management systems (LMSs) and platforms, to learning activities and courses information consisting a curriculum such as learning objectives, syllabuses, learning material and activities, examination results and courses' evaluation, to other kind of data related to administrative, educational and quality improvement processes and procedures. The limited exploitation of big educational data and the size and type of these data within the context of higher education signifies the need for special techniques to be applied in order to discover new beneficial knowledge that currently is hidden within

data [6]. Such techniques can be derived and adapted from other domains characterized by big data and successfully used to manipulate big educational data. These techniques could be used to enable the development of insights "regarding student performance and learning approaches" and exemplify areas within big educational data—such as students' actual performance according to taught curriculum—that can be positively impacted [7]. Recently, big data and Analytics together have shown promise in promoting different actions in higher education. These actions concern "administrative decision-making and organizational resource allocation", prevention of students at risk to fail by early identify them, development of effective instructional techniques and transform the traditional view of the curriculum to reconsider it as a network of relations and connections between the different entities of data gathered and regularly produced from LMSs, social networks, learning activities and the curriculum [8]. More specifically, one of the identified areas in which big data and Analytics are appropriately applicable for investigation and improvement in higher education is the curriculum and its contents, as a major part of big educational data [9, 10].

2.3. BIG EDUCATIONAL DATA IN HEALTH EDUCATION

Health education is an interesting context since it is complex. Its complexity lies in the constantly increased body of medical knowledge and evidence that continuously needs to be reflected in educational activities in order to match the needs for competent health professionals that meet the demands of the healthcare system and the society as its stakeholder. It produces an enormous amount of educational data considered as big. More specifically, the variety of data encased from teaching, learning and assessment activities, make it an area in which big data and analytics can be very useful to exploit them and sort out the complex information to be found in large diverse data sets [11]. Using big data and analytics techniques as an approach to make sense of the data, representing a health education curriculum and the associations between them, revealed its underlying complexity and the power that these techniques offer in two different cases.

In the first case [12], it was attempted to analyze and visualize the connections between the overall intended learning outcomes (ILO—in red) given in the different courses of an undergraduate medical curriculum and the desired competencies—from both the medical programme (in blue) and the higher education board (in dark and light green)—a medical student should have acquired after graduation from the medical programme. This is considered an attempt to make sense of this data in a small scale but yet, even in this case, the visualizations (**Figures 1** and **2**) reveal and confirm the high levels of complexity of this data. Further, considering as we mentioned before the continuously growing medical evidence that needs to realistically be reflected in the educational activities, the nature of this data is not static and represent only a snapshot of a long-term changeable network on the time it was captured. Yet, meaningful conclusions can be derived in a glance from these visualizations such as which competency is addressed the most with ILOs (connections between light green and red in **Figure 1**), or for example, clusters of ILOs used to address either knowledge or skills while addressing a common competency of the medical programme (connections between red non-clustered and clustered in **Figure 2**), and more.
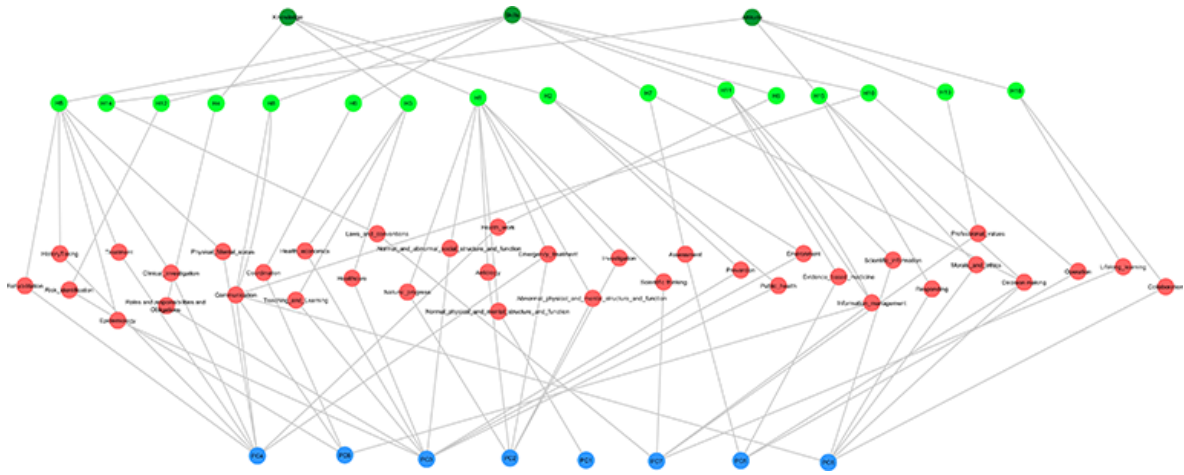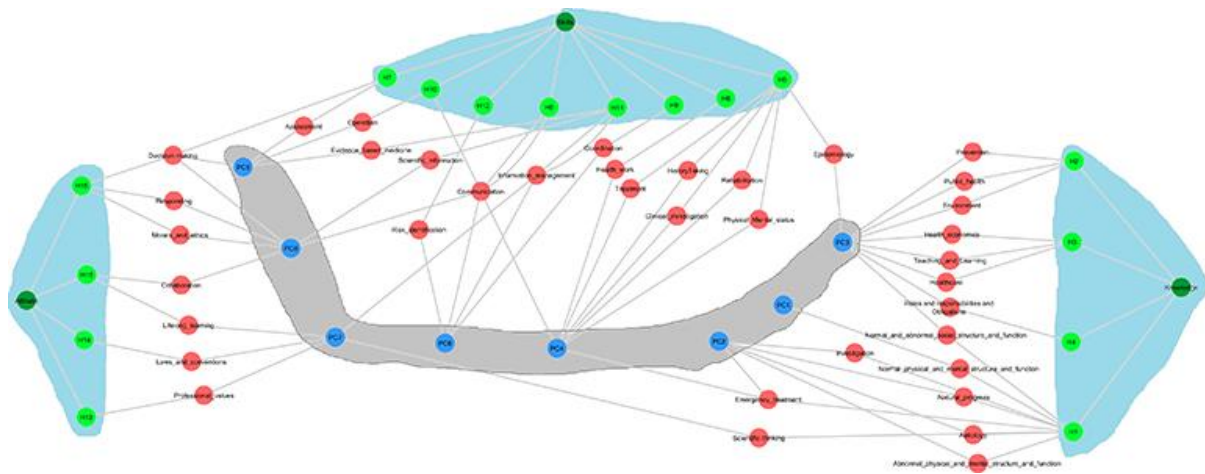
FIGURE 1.

Competencies and ILOs map.

FIGURE 2.

Clusters of competencies and ILOs.

In the second case [13], it was attempted to visualize in a global association map the connections created by the practical incorporation of MeSH terminology in one particular section of a medical curriculum (**Figure 3**). Again, despite the obvious complexity of the MeSH map, conclusions can easily be derived quickly concerning, for example the less often used MeSH terms, here depicted in small clusters and located outside the main big cluster. Of course, this kind of representations require considerable time to be processed by humans due to their high complexity, but definitely they can promote understanding of overview of the situation and facilitate high-level reporting of bulks of information.
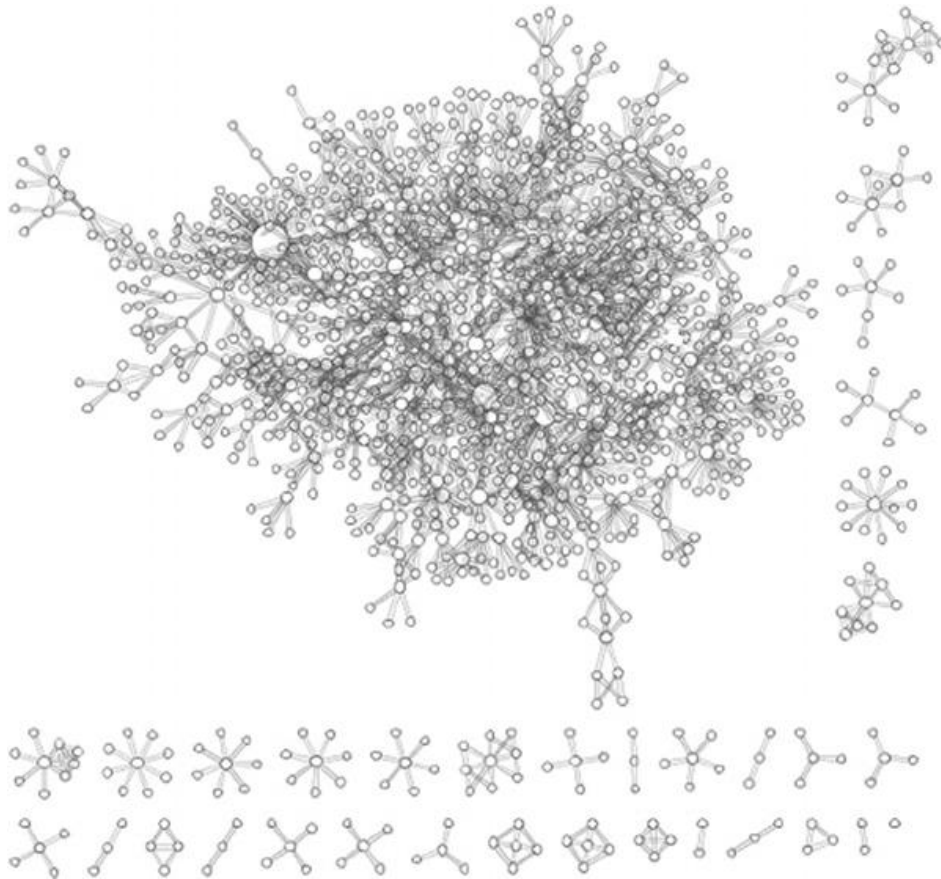
FIGURE 3.

MeSH terms association map of a particular section of a medical curriculum.

3. Analytics

3.1. DIMENSIONS AND OBJECTIVES

From a broad perspective, the development of analytics models has shown promise in transforming big educational data in health education into an Analytics-driven quality management tool. In the world of academic and learning analytics, the sources that big educational data are derived from are distinguished in different levels. This gives a multidisciplinary character to the field of analytics in general, involving various techniques, methods and approaches frequently used in the field. The range of actions that can be taken within the analytics area is wide, and frequently, these actions are classified into different levels and dimensions. For instance, the different actions taken in the field are divided by some practitioners into three different dimensions: time, level and stakeholder. Specific analytical approaches are applied to address respective questions for each of the dimensions. Descriptive analytics, for instance, produces reports, summaries and models in the dimension of time to answer the what, how and why something did happen. It monitors also processes to provide alerts in real time and recommend answers to questions as: What is happening now? In the case of predictive analytics, past actions are evaluated to estimate the future actions outcomes by answering: What are the trends, and what is likely to happen. It also simulates alternative actions outcomes to support decisions. Using analytics, choices are based on evidence rather than assumptions [14].

Analytics has been also classified into five levels: course, department, institution, region and national/international [8]. Other terms attempting to define the different levels more specifically can be applied; "nanolevel" indicates activities in a course; the "microlevel" points an entire course in an education programme; the "mesolevel" includes many courses in a

specific academic year; and finally, the "macrolevel" concerns many study programmes in an educational institution [15]. **Figure 4** shows these four levels and the relation between them.
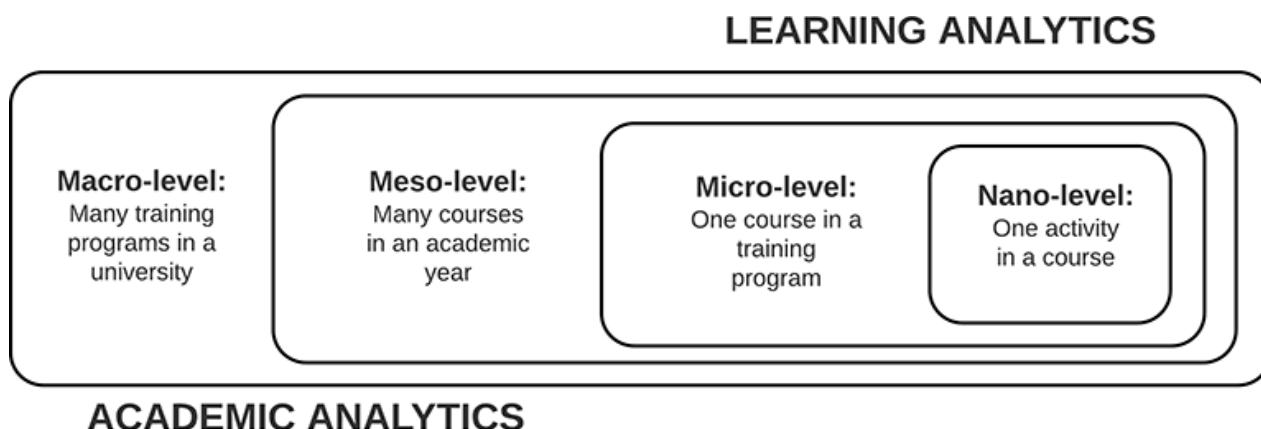

FIGURE 4.

Overlapping of Analytics levels in higher education.

When the focus is on decision-making concerning achievements of specific learning outcomes, then all included actions are governed by "learning analytics" which refers to operations at the microlevel and nanolevel. When the focus is on decision-making regarding procedures, management and matters of operational nature, then it is governed by "academic analytics" which applies to the other two levels, macro and meso [16]. **Figure 4** illustrates how the different levels of analytics in education overlap and complement each other. For example, results of actions taken in the nanolevel can be input to the other levels micro, meso and macro, while it is controlled and monitored by them. The application of analytics in this classification can also be oriented toward different stakeholders, including students, teachers, administrators, institutions, and researchers. They may have different objectives, such as mentoring, monitoring, analysis, prediction, assessment, feedback, personalization, recommendation, and decision support. Despite the categorization of analytics actions in different levels, the data that these levels generate enter the same analytics loop which is defined in five steps in **Table 1** [17].

Steps	Description
Step 1: capture	Data are the foundation of all analytics. These data can be produced by different systems and stored in multiple databases. One great challenge for analytics projects in this step is that necessary data may be missing, stored in multiple formats or hidden in shadow systems
Step 2: report	Dashboards provide an overview of trends or correlations. This step involves creating an overview to scan. Different tools can be used to create queries, examine information and identify trends and patterns. Descriptive statistics and dashboards can be used to graphically visualize eventual correlations
Step 3: predict	Predictions and probabilities can be derived. Different tools can be used to apply predictive models. Typically, these models are based on statistical regression. Different regression techniques are available and each one has limitations
Step 4: act	The goal of analytics is to provide actionable insights through information based on predictions and probabilities that support decision making. Analytics can be used to evaluate past actions and estimate the effects of future actions. In that way, analytics can provide alternative actions and simulate the consequences of different actions

Steps	Description
Step 5: refine	The evaluation feeding back the self-improvement. The monitoring, feedback and evaluation of the project's impact create new data and evidence that can be used to start the loop again with improved performance

TABLE 1.

Steps in analytics loop.

Another type of classification was proposed [18] and provides a division in different dimensions: The environment; what data is available? The stakeholders; who is targeted? The objectives; why do the analysis? And the method; how has the analysis been performed? Finally, analytics can team up with other scientific areas for analysis and high-level communication of actions such as scientific information visualization and data analysis techniques (e.g. data mining and network analysis) elaborated upon later in Section 3.2.4 in the chapter.

3.2. ANALYTICAL APPROACHES

As we saw, there are different components that analytics actions need in order to be effective. These components are the data (type and source) and the context in interest. If these components of analytics are in place, we are able to create different analytics models which can thrive and grow into an analytics engine capable to harness big educational data to ultimately contribute to the quality management and improvement of health education. Based each time on the needs of the health educational ecosystem in question, different approaches can result in building multiple viewpoint analytical models. The analytics approaches presented below are not specifically related to any type of classification in dimensions or levels but rather can work with any type of analytics model which constitutes all necessary components.

3.2.1. DATA-DRIVEN ANALYTICS APPROACH

Reading from the left to the right, **Figure 5** describes the common and traditional data-driven analytics approach, which is quite meaningful to experts in the data analysis area. It starts from the data and ends in the decision. The main focus is on the data and the necessary techniques to collect, store, clean, secure, transfer and process them. According to this approach, the loop starts in the first step by capturing as much data as possible, and then, the data are pushed through the different steps. Into the reporting step, the high volume of data is an asset. The more data we add, the better results we will receive. However, processing massive data sets includes challenges, such as demand for high-level mining techniques and more robust computers, applications, software and skills. To make sense of all this data, estimate the trends and examine all possible associations is a challenging task. Data analysis techniques, necessary to process the data in this step, require expertise usually found in data analysts and most commonly within the educational data mining area. Based on the evidence from previous steps, the engine predicts the trends and suggests actions that might be accurate and precise, but still remain suggestions. Often, the decision makers, frequently because of unknown circumstances, underestimate the recommendations and act differently. The loop finishes with the last step which is to either end the loop or feed the engine with more data in step 1 and run the engine again.
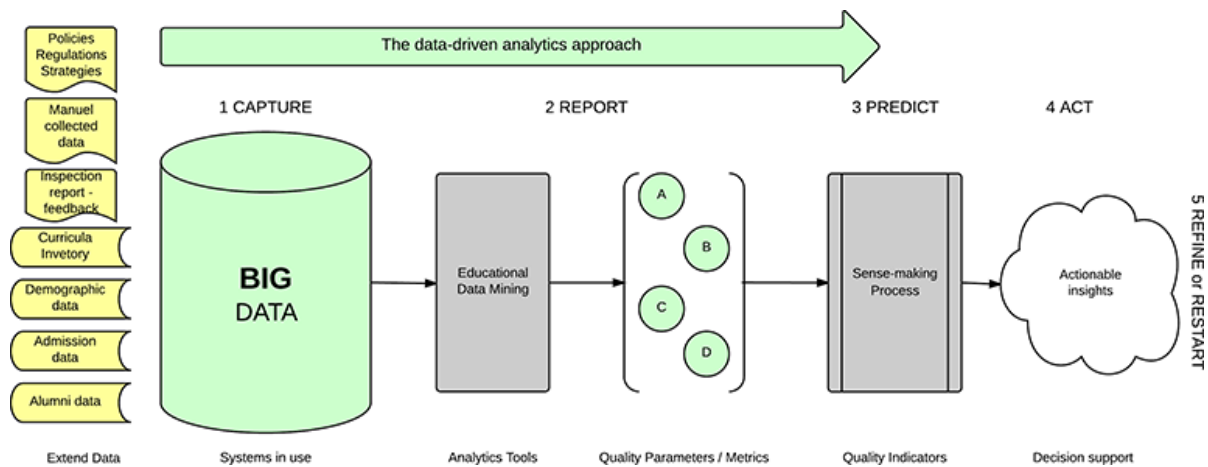
FIGURE 5.

Data-driven Analytics Approach.

3.2.2. CONTEXT- OR NEED-DRIVEN ANALYTICS APPROACH

The model reads also from backwards (steps 1–8 in **Figure 6**). It describes in this way a new analytics approach called context- or need-driven analytics. This approach is more suitable for less qualified group of users in data analysis techniques such as educators and decision-makers. The approach starts from the need for a decision and goes through the analysis of relevant data which could support the decisions. Quality improvements, decisions and actions must be crystal clear. Every detail is important: the stakeholders, the circumstances, particular needs, economic boundaries, accessibility of resources, organizational atmosphere, policies, technological ecosystem, timing and other factors which could influence the decisions. The results of this investigation are the demands of specific information to support a judgment or micro-decisions. This important and particular information emerges from the integration of carefully picked and explicit data. These data are selected, prepared, assessed, compared and produced by analytics tools utilizing particular mining methods. The analytics engine includes additional mechanisms and specific operators to recognize the systems which generate the data or the containers which carry the data. This time, we extract just the necessary data we need. Finally, the analytics loop either filter the data and provide an answer to the primary question or re-enter a new, more precise, question and restart the analytics process [19].
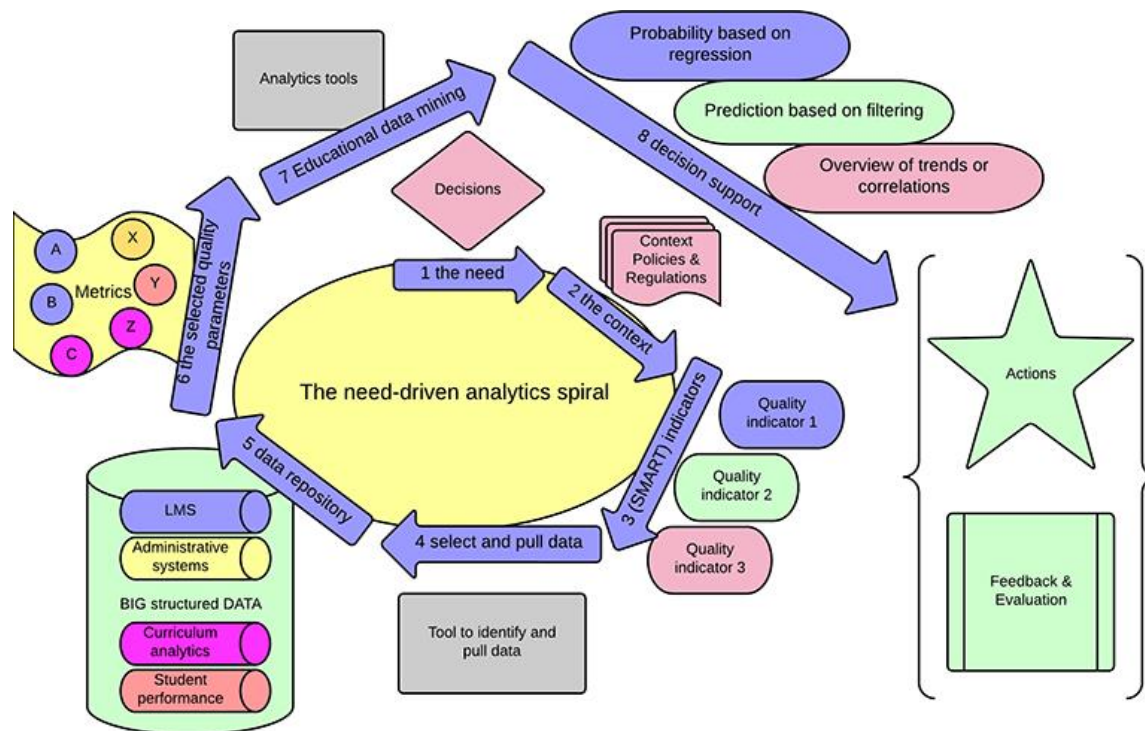
FIGURE 6.

Context- or need-driven Analytics Approach.

3.3. LEARNING ANALYTICS

The term "learning analytics (LA)" is defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" [20] and affects actions and operations at the microlevel and nanolevel in **Figure 4**. Through LA, we can detect similarities in behaviours (e.g. user's satisfaction) or detect anomalous patterns (e.g. cheating). It can function as a bridge between past and future operations by inserting data concerning past events into a LA engine and analyse them to determine the probable future outcomes. It can synthesize thus big educational data and create a set of predictions to suggest different decision options revealing each time the implications of each decision option. LA can be further enhanced through visuals to amplify insight, increase understanding and impact decision-making as we explain further, later in the chapter.

Teachers, usually based on their experience, use their own "gut feeling" to translate students' behaviour and suspect if a student might drop out of a course or even abandon the studies. This can be proven to be either true or false, but without evidence, there is low level of certainty in decisions that are based only on experience. An example demonstrates the LA capacity to use evidence and add confidence to this type of decisions [21]. Here, data mining techniques were applied in big educational data and were utilized as a part of an analytics engine to detect students that perform in high, middle and low levels and notify them accordingly with different types of feedback. Thus, students at risk were identified very early when the institution still had the time to react and take preventive actions.

3.4. ACADEMIC ANALYTICS

The term "academic analytics" is defined as "the intersection of technology, information, management culture and the application of information to manage the academic enterprise" [22] and affects actions and operations at the macro and mesolevel as we saw before in **Figure 4**. The focus of academic analytics includes reporting, modelling, analysis and decision support concerning university and campus services. Examples of this kind of services include, but not limited to

admission, advising, financing, academic counselling, enrolment and administration. Following is a practical use of academic analytics [23], where librarians have used analytics on library usage data as part of the big educational data ecosystem to predict students' grades demonstrating the value that can be provided by the data produced and processed in the library to the hosting institution. In another case [24], it is demonstrated how within the context of health education academic analytics reports extracted from a mapped medical curriculum using data mining techniques, can add transparency to the big educational data consisting the medical curriculum and can be of use to stakeholders to facilitate decisions that need to be taken concerning different kinds of services such as managerial and financial.

3.5. VISUAL ANALYTICS

Methods and techniques have been developed in the recent years that can be used to manipulate complicated data in many different disciplines [25, 26]. Visual analytics (VA) is the science of analytical reasoning supported by interactive visual interfaces as an outgrowth of the fields of information visualization and scientific visualization [27]. VA combines different techniques: information visualization, data analysis and the power of human visual perception (**Figure 7**) [28].
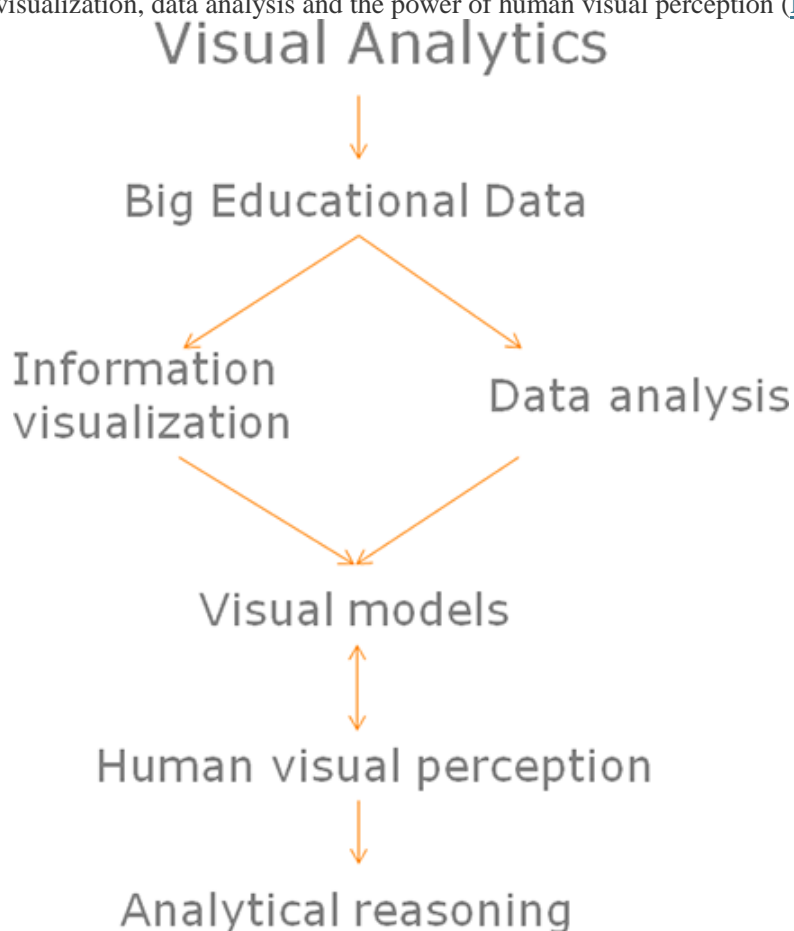

FIGURE 7.

Big educational data are modelled by information visualization and data analysis techniques and represented in visual interfaces with which the human visual perception interacts to impact the analytical reasoning process.

It has the potential to support in the process of manipulating big data and exploit them by creating a holistic view of the data while revealing underlying complex information to the extent possible to positively impact analytical reasoning and decision-making [29–31]. A review of the literature resulted in identifying variables [32, 33] that are able to support

analytical reasoning and decision making through VA and the interaction between human visual perception and visual interfaces as below:

- Increased cognitive resources (V1)

- Decreased need to search for information (V2)

- Enhancement of the recognition of patterns (V3)

- Easier perception of inference of relationships (V4)

- Increased ability to explore and manipulate the data (V5)

The potentials offered by VA making it a promising tool to explore also how big educational data could contribute to the quality improvement of higher education. Different approaches prove the potential of VA to impact quality improvement specifically within the context of health education. It is reported [34] how the analysis and a simple visualization of educational data of a medical programme enabled involved stakeholders to instantly review and preview the effects of implemented changes in a medical curriculum. We will examine how in another case, VA has been practically used to explore its impact on analytical reasoning and decision making using big educational data from a medical programme [35, 36].

In **Figure 8**, we see how the learning outcomes (LO) and the teaching methods (TM) of one course were modelled to visually represent the hidden underlying network of connections and relations between them. The TMs are depicted in percentages in red, to show to what extent each TM is used in the course out of a 100%. Each TM addresses a number of LOs, and these are depicted in light blue. The percentages between an individual TM and its LOs depict the extent in which each TM's content is used to address the specific LO. A number of non-addressed LOs are depicted on the top-right corner to complete the set of predefined LOs (16 in total) that the medical programme should address within the different courses. Here, the LOs and TMs are mapped and represented hierarchically from its 100% of TMs to corresponding percentages of TMs showing to which extent each TM is used in the course. Going further, the percentages between TMs and LOs reveal how much of the learning content of the TM is used to address the specific LO. For instance, the "clinical training" TM is fully addressing LO7 with its learning content while uses only 10% (5% out of 50%) of the learning content to address LO8. Thus, a comparison between learning content usage can instantly show which LO is mostly addressed and reveal the tendency of the TM or even the whole course—when we compare all TMs—towards specific LOs and even further competencies build through the LOs. This approach provides a way of analysing the teaching part of the course in relation to the LOs addressed to support the process of analytical reasoning. In the event of a series of similar comparisons, an instructor can base its decisions concerning the right percentage to address an LO and reform and redesign accordingly if necessary, to be more tailored to the LO's importance. In this way, an instructor evaluates and confirms the correct usage of TMs to address the LOs even if redesigning is not necessary. In parallel, a comparison between addressed and non-addressed LOs and between used and non-used TMs can be performed at any moment, revealing the whole course's map.
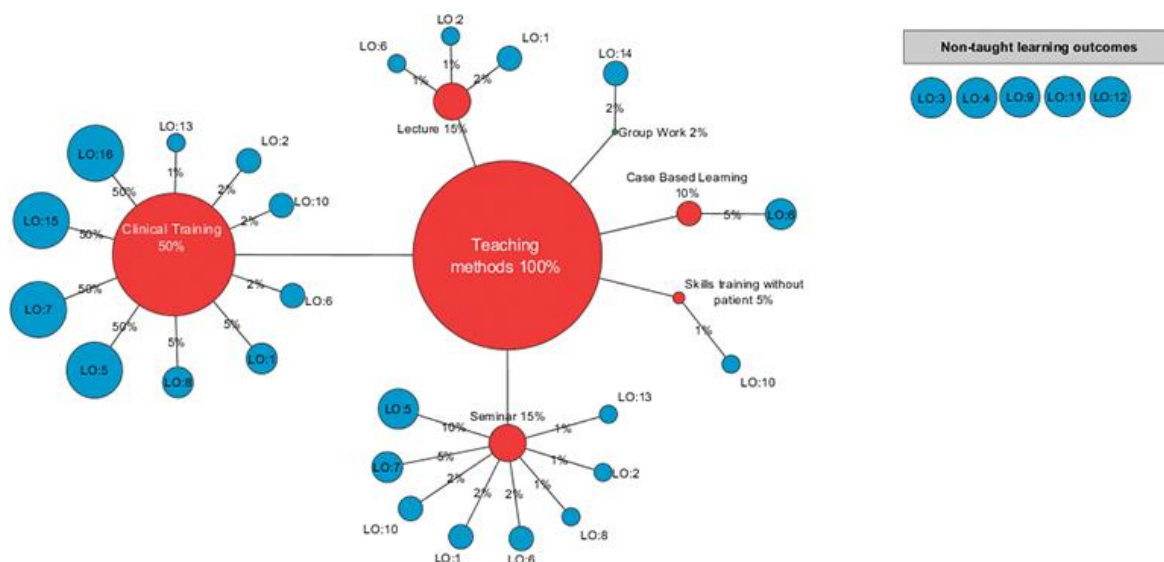
FIGURE 8.

Learning outcomes and teaching methods.

In **Figure 9**, we see how the LOs of the same course were modelled this time against the assessment part and more specifically one part of the assessment, the questions used in the written examination, 34 in total. The percentages on the connections between yellow and red circles depict the proportion (out of 100%) of exam questions used to address the specific LO in red. For instance, eleven questions are used to assess LO5 which corresponds to 32%. Groups of LOs correspond to main outcomes—knowledge, skills and attitude—which are depicted in green. In cases where multiple main outcomes are assessed in groups of questions, the total percentage is divided into single main outcomes as in the case where 30% of the questions are used to assess skills and knowledge corresponding to 15% skills and 15% knowledge. An instant observation is that 83% of the questions on the written examination are used to assess skills, while 16% are used to assess knowledge and 1% attitude. Also, the percentage of questions that assess each of the LOs reveals how the written examination is built around them and which LOs are most heavily assessed. Some LOs are assessed in more than one group of questions, like LO5 in five different cases with corresponding red circles or in combination with other learning outcomes, like LO7 in two cases. The analytical process is supported in this case by instantly evaluating how the LOs of the course are assessed in the written examination. The percentages of questions can be examined against the importance of the assessed LO and thus suggest whether it is the correct percentage of questions, compared to the other percentages of questions used to assess other LOs. Thus, an instructor can decide if these percentages should be adjusted according to the importance of LOs and redesign the questions of the examination or even if it is more appropriate to address these LOs in other types of examination. Finally, this approach can be used to construct a more outcome-oriented written examination by redesigning it to cover identified gaps in addressing important LOs and instantly evaluating it with the updated visual model of the assessment activity.
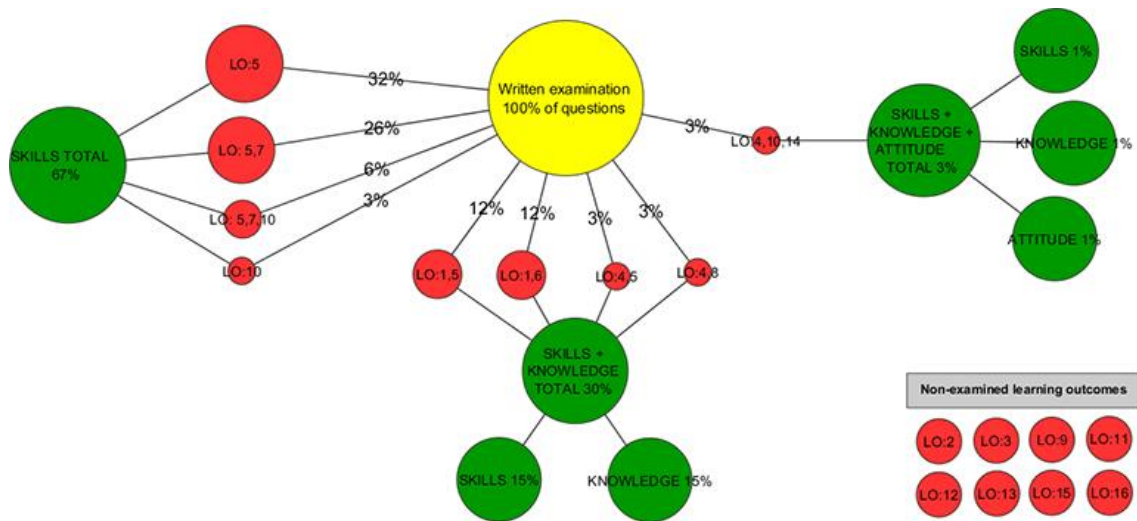
FIGURE 9.

Examination and learning outcomes.

In **Figure 10**, we see an overview of the whole course. The TMs are depicted in red, main outcomes in yellow and LOs in light blue. The total points a student can get from each exam question are depicted inside the orange circles, and the percentages on the connections between these circles to LOs show the average success rate from all student answers on this particular question. The three light blue circles bordered in black (LO4,5, LO4,8 and LO4,10,14) and LO4 in bottom right corner depict the different cases where LO4 it is assessed by exam questions, but it is not taught in any of the TMs. This visualization sums all the information from Figures 9 and 10 providing additionally more information about the course in one place. Here, we can observe and analyse the entire course from different perspectives but also as a whole. Examining this figure from left to right and vice versa, different paths are created to disclose the underlying network in the examined educational data. The most focused and most assessed LOs can be observed instantly, showing the trend of the course towards skills, knowledge and attitude, to what extent these are addressed and if there are any gaps of taught/non-assessed LOs. Finally, the existence or not of the constructive alignment [37] in the course can be verified as a synthesis of possible identified gaps and the utilization of learning activities and LOs in one place presenting the course as a structured network.
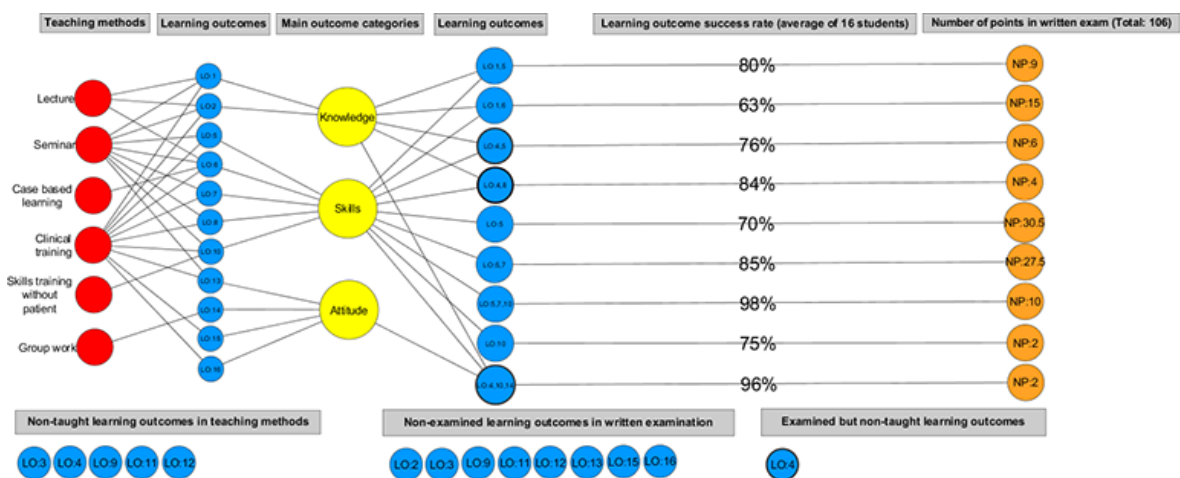

FIGURE 10.

Overview of a course.

The analytical reasoning process is here more enhanced. The entire course can be instantly evaluated for gaps between taught and assessed LOs. For example, the identified gap for LO4 means simply that the written exam questions assess the LO4, but it was never actually taught in any of the TMs. This approach can be used as a tool in the hands of the course stakeholders to analyse it for this type of inconsistencies and possibly redesign it to establish a connection between what it is taught and what it is assessed and verify it again. After the redesigning, a comparison can take place where the different versions of the course will be similarly depicted before applying the desired changes in reality and thus create a more concrete and aligned course without gaps that meets the desired LOs appropriately.

The three presented approaches of using VA on big educational data within the context of health education demonstrate the potentials on impacting analytical reasoning and decision making in connection to the previously identified variables (V1–V5). Specifically, the information depicted is easily recognizable to the stakeholders in interest while making perceptible the different patterns and relations between the data (V1, V3 and V4). Searching for information relevant to the course structure is facilitated to a high extent (V2). The course can be readily analysed for gaps of different kinds while, at any time, the constructive alignment of the course can be verified (V3–V5). Finally, **Figure 10** has been further investigated with the use of augmented reality (AR) technology in an attempt to increase interactivity between the user and the visual and to enrich it with additional information while sustaining the complexity in low levels showing promising results for investigating big educational data by combining VA and AR [38].

4. Quality improvement (QI)

4.1. QUALITY IMPROVEMENT AS AN IMPLICATION OF IMPROVEMENT SCIENCE IN EDUCATION

Quality improvement is defined as "the combined and unceasing efforts of everyone to make the changes that will lead to better outcomes, better system performance and better professional development" [39]. This definition covers all different aspects of health care that inextricably are affected by efforts targeting change. Improvement science instruments all the different ingredients and components necessary to realize this type of efforts that quality improvement requires to be a successful process. Improvement science has been applied in many disciplines such as automobile manufacturing and health care like an alternative approach to bring new knowledge into practice. Projects rooted in improvement science began to show success even within education. The characteristic of the improvement science is the holistic view of the examined context, and the key step is to identify the context (e.g. the organization, the actors and stakeholders, the routines and the workflow) and consider it as a system; deep knowledge of how small changes in a system instance can affect other parts of the system is very important.

Traditionally, improvement science was based on the "plan-do-study-act" cycle [40] attempting to answer fundamental questions such as:

- What are we trying to accomplish with the desired change?

- What changes can we make to achieve an improvement?

- How will we know that a change is also an improvement?

Today, the use of analytics in big educational data can be the "game changer" and can play an undeniably significant role in orchestrating the components of improvement science actions to design changes that successfully lead in improvement in the quality of education. Below is a formula that utilizes big educational data and combines the necessary components along with analytics within the context of education to successfully make a desired change to produce improvement.

4.2. THE FORMULA AND ITS ELEMENTS

The formula illustrates the way in which the different components come together like building blocks to produce improvement and can be used like a guide to design the change.

1	2	3	4	5
Context	+	Actionable Intelligence	→	Improvement

Each of the five elements is driven by a different knowledge area and has its own characteristics and settings.

4.2.1. ELEMENT #1: CONTEXT

Deep knowledge of the particular context is the starting point. Differences on who, when, why, where and what can affect the choices we have or the selections we make. Different stakeholders perceive and use the terms and concepts differently in different occasions, but there are predominantly two ways to describe the context of education and define its quality. Some describe it as the personal development in people focusing on the outcome. They talk about "learning" and consider students like collaborators, or participants. Others describe education as the service of educating people focusing on the process. This group talks about "teaching" and considers the students like stakeholders, receivers, target group or customers/clients. Based on how we describe what education is we use different indicators to define its quality [41].

4.2.2. ELEMENT #2: THE "+" SYMBOL

This element represents the knowledge required about the different modalities for appropriate management of big educational data (analytics and data processing techniques) to properly connect and transform the context knowledge into the next element, the actionable intelligence.

4.2.3. ELEMENT #3: ACTIONABLE INTELLIGENCE

Through analytics, we can transform data to actionable insights and support decisions. As we have demonstrated, different analytics types, approaches and techniques are available (learning analytics, visual analytics, academic analytics, sense-making or predictive analytics, data-driven or need-driven analytics, etc.). Making decisions based on big educational data

collected from complex learning environments may encounter limitations of human cognitive capability. That makes it necessary to expand this field and further investigate how different processes like cognitive artefacts that model human thinking sub-processes (e.g. accommodation, conclusions and categorization) could possibly facilitate the flow of human reasoning and therefore enhance the human cognitive ability [42, 43]. According to multiple analytics reports derived from the same data set, each of which provides a lens that adds more contextual insight will enable, for example the course developers to look for patterns [44, 45]. It is obvious that in our case the used final set of analytical reports as well as the selection between the mass univariate and multidimensional approach will emerge mostly from the available data sources and the technical/ethical possibilities to fuse them. Very often, the measures or parameters presented to the course developers will have to be extracted from the raw data with techniques, such as natural language processing, social network analysis, process mining and other.

4.2.4. ELEMENT #4: THE → SYMBOL

This element represents the knowledge about the execution and management of the change. The knowledge area is based on the Implementation Science and focuses on the methods and techniques required to "make things happen" and drive a successful implementation of an intervention in place.

4.2.5. ELEMENT #5: IMPROVEMENT

Improvement is about changing but not all changes are improvement. This element represents knowledge about the types and methods required to evaluate special types of measurements to show whether improvement has happened and calculate its impact. There are five different approaches depending on how we consider or view the quality [44] summarized in **Table 2**.

Quality is	Approach to measure
Exceptional; quality is something special	We create objectives, checking against standards and try to achieve "high class" or "excellence". This approach allows comparisons or benchmarking
Perfection or consistency; zero defects	In this approach, a service is judged by its consistency and reliability. The focus is on the processes to ensure that faults do not occur
Fitness for purpose; specification/mission and satisfaction	This approach is remote from the others. We accept that quality has meaning only in relation to the purpose and the users/stakeholders. It requires identification of the needs, continues monitoring, periodical re-evaluations and responsive adjustments
Value for money; Performance	This approach uses the terms "efficiency and effectiveness" and focuses on the accountability and linkage of the outcomes to the costs
Transformation; added value and empowering of the user	In this approach, we consider students as participants (not as products, customers, consumers, users or clients). In this case, education is an ongoing process of transformation of the participant and not a service for a customer

TABLE 2.

The different approaches we follow for each one of the views.

4.3. QUALITY IMPROVEMENT OF LEARNING PROCESS

Operations at the microlevel and nanolevel (**Figure 4**) such as teaching or learning activities in a course are referred to LA. Examples of these operations are performed by teachers, course designers, studies and programme directors. The following scenario demonstrates the practical use of LA in the quality improvement circle of a course.

In the preparation phase of a course, the instructors can use curriculum mapping tools to discover actual gaps precisely. They can recognize thus which learning objectives are not properly addressed by teaching or learning activities. They need recommendations for new, more proper and motivational teaching activities to include them into their schedule. With the available Analytics tools, they are able to analyse further the class and predict its needs such as student demographics, performance, different learning approaches, the technology used and the group dynamics. This type of data is processed by a number of algorithms and predictive models that can develop the characteristics of the class [32]. Visualization tools can be used for the following round to give alternative proposals for designing suitable activities fitting this particular class and also illustrate the effects of each of the options. The course director can control the activities and observe students' progress during the ongoing course. They can zoom in and out from the whole class to one working group or one individual student. They can additionally track the flow of the formed social networks. They can judge the overall commitment and identify students at risk. In an extensively used platform, they can also compare particular indicators from other classes, or through to other anonymized data sets within the same program, or from a different department, or even compare against data from related programs in other universities [46]. The results and the produced experiences can be used to build up the knowledge database evidently regarding several pedagogical interventions. This can support in forming new policies in the entire organization and be an important element of the quality development and academic research.

4.4. QUALITY IMPROVEMENT OF EDUCATIONAL PROCESS

We presented how VA could be used to support the analytical reasoning and decision making of stakeholders involved in the quality improvement of the educational process. This is achieved when both visual and analytics factors function as instruments of a harmonized engine that complement and support each other. The analytics factor applied on the big educational data aims at reducing its complexity without losing vital information and critical characteristics; these are kept at the top level of the presented visuals. The other factor is the visualization, which brought pathways and relations into light by taking advantage of the human ability to process and understand visual information more easily. These two factors cannot stand alone without each other and be implemented to data with incoherent structure, which makes Analytics an essential key component to build a strong base for a meaningful VA result. The data analysis preceding visualizations assists in shaping the inchoate big educational data that visuals are then responsible to represent. An important point is the effort needed to apply each of the factors. The effort required for the visual and analytics parts is not comparable, and their roles are totally different. Analytics requires significant effort to shape the data in question and compile all the discrete elements to represent the data adequately. On the contrary, visuals require less effort since the network of connections and relations is already assembled. However, to select and gradually build the appropriate visuals, it requires expertise in order to emphasize in a big picture the essential information existing in the network produced from data analysis and add

scientific value onto it while going beyond simple statistical-based visuals. Of course, the human visual perception is irreplaceable in this chain of actions in order to perceive and interact with the visual interfaces and perform high-level analysis. In summary, VA allows the different stakeholders to easily perceive the structure of the examined data, define how each part coexists as part of a network and reason for its use and importance in the data. It also helps to better understand stakeholders' individual role in the educational process and the consequences of delivering their parts without being able to determine how it can be harmonized with other parts in the data. It supports stakeholders also to decide how to cope with discrepancies and structure anomalies revealed from gap analysis and the existence or not of the constructive alignment in the data. Finally, VA can display currently needed changes for an improved future overall picture in order to deliver health education in pace with healthcare demands [47, 48]. Revealing the underlying network of information in the examined data, identifying gaps, discrepancies and anomalies between the data and being able to verify the appropriateness of the given educational activities promotes the process of analytical reasoning and decision making and transforms the big educational data into an instrument for planning and applying changes in a constant effort for quality improvement in health education.

4.5. QUALITY IMPROVEMENT OF ACADEMIC FUNCTIONS AND CAMPUS SERVICES

Academic analytics has been compared to business intelligence and refers to operations at the macrolevel and mesolevel as we saw in **Figure 4**, including decision support concerning university and campus services. In most of the cases, Academic Analytics have been used to provide actionable insights and support single or isolated decisions [49]. As we demonstrated Academic Analytics is a main part of the quality improvement process and can be beneficial in multiple ways into the steps of the improvement's cycle. Into the early steps of the cycle (the data-driven approach, **Figure 5**), it can support decision makers to identify the gaps and the needs of what is possible or necessary to improve. Into the following steps, academic analytics can support decision about choosing appropriate actions trough predictions and by providing "what if" scenarios using the need-driven approach in **Figure 6**. Academic analytics (through dashboards and reports) can be used to monitor the ongoing processes and support decisions concerning eventual adjustments. At the end of the quality improvement cycle, academic analytics can support in performing evaluations of the intervention's impact demonstrating the hidden connections between actions and events.

5. Conclusion

The goal of this chapter was to introduce the reader to the concept of big educational data and the different forms of analytics as applied scientific areas and go deeper to popular techniques for data manipulation and how they can be transferred within the health education system and used as approaches to exploit big educational data that such systems produce. Apart from the techniques itself, the benefits and potential to use them for quality improvement purposes in health education are provided and discussed in detail.

In the era of technology and its inevitable impact on health education systems, such approaches are proven to be quite utilitarian in order to support the quality improvement process of education and ultimately contribute to health care with highly skilled health professionals.

6. Acknowledgements