Quality Assurance of the Education Process Supported Partly by Big Data Analysis

György Molnár
(molnar.gy@eik.bme.hu)
Budapest University of Technology and Economics, Department of Technical Education

Zoltán Szűts
Budapest University of Technology and Economics, Department of Technical Education

Abstract

The rapid development of ICT is constantly challenges the society, the economy and the way of everyday life. This tendency is becoming more and more intensified nowadays. Besides the transformation of basic ways of life, tasks and roles, changes are certain in some elements of the education system. All these changes can be measured and analyzed – under appropriate conditions – with Big Data analyzes. On the basis of teaching and learning process, this can be of great importance in the analysis of data sets belonging to different higher education courses. Likewise, it has an important role in examining certain indicators of education and economics.

Research presented in this paper supports the idea that online platforms and social media provide information about education that Big Data can use to improve education quality.

The used method is an online questionnaire survey. In the framework of the investigation, the results of the survey will be presented by the authors and further conclusions will be deducted. The target group of the questionnaire survey is the group of students and graduates of the BME specialization training.

The methodological experience and results can be used as a basis for developing a broader, multi-year, Big Data survey using large data volumes, which imaginary use is introduced. The expected results of the research can contribute to the further development of the teacher training courses and to the expansion of the training provision that meets the needs.

Keywords: teacher training, teacher career model, motivation, big data

Introduction

From historical perspective, when looking at the educational system of the given age and what challenges it contains, it can concluded that characteristics of systems based on quality have not changed fundamentally. They always adapt to the given era. When looking at the issue of quality insurance, it must be examined what technological

challenges ought to be met, together with industrial development in education systems? It is also important to ask, what does quality assurance mean today? It has multiple interpretations: innovation, renewal, modern technology and technology.
The bond between the quality of education and the country's economic development is now largely accepted (Hanuschek-Woessmann, 2010). On the other hand, the personality and activity of the teacher is a determining factor in the quality of education (McKinsey Report, 2006). That is why the teacher's training is also important. The basic purpose of the teacher training courses in Hungary was changed by the teacher's lifecycle model introduced in 2011 by the fact that the criteria for further progress between the different degrees of qualification are defined by successful participation in further training.

Big Data

The authors will introduce the possible use of Big Data methods in quality assurance. On order to do so, firstly let us look at the nowadays popular phenomenon of Big Data.

In 21st century information society, the value of information has increased significantly. The term Big Data describes the amount of information and its processing. This information is generated by users and IoT. The processing itself is available to wide variety of researchers and companies. In the discourse on Internet, it is often stated that the global network has changed the way information is collected, processed, stored and shared. It is designed to transmit the data digitally, basically regardless of the nature of the content.

Big Data means capturing, processing, analyzing, sharing data from virtually infinite sources, exploring contexts, and presenting the results transparently with clear visualization. The Big Data trend is increasingly affecting several parts of everyday life. According to Laney (2012), Big Data is defined by three Vs: Volume, Velocity and Variety. The quantity indicates the huge amount of data arriving in real time. In the past, such large amounts of data would have caused storage problems. In the present, due to the cloud technologies, storage space and network speeds have increased, while its cost was reduced. Velocity – or speed – is an essential feature, as the data does not come in large blocks, but flow continuously, with smaller or higher intensity. In terms of efficiency, we need to process these data in real time. The biggest challenge is the variety of data, as data comes in different formats – text, picture, video, sound – and needs to be structured. The goal is to convert data streams to extract valuable information. The variety is characterized by the fact that data comes from text documents, “traditional” databases, video surveillance systems, e-mails, phone calls etc., and analytical systems must recognize patterns (Majkić, 2014). Everything that is born and happens in the context of the network is stored, can be traced and analyzed (Csepeli, 2015: 172–173).
In the framework of Big Data analysis, the following questions have to be answered: what happens if we have to process complex and large amounts of information and can not apply the methods that we have used so far? Do we need to store all the data? Does all data need to be analyzed? How to determine the most important information? How should the information be presented (visualized)? Prior to the method of Big Data, often random sampling had to be applied to reduce the amount of information analyzed or to rely on a large number of analysts. By contrast, Big Data processes all data through machine learning. Therefore, we no longer have to choose between the size and depth of the data (Manovich 2012: 466).

Sentiment analysis in quality assurance

From one side, de huge amount of data and a new method – Big Data – and from the other, sentiment analysis enables us to get valuable insights that can be used in quality insurance. George Siemens points out (https://www.oreilly.com/ideas/education-data-analytics-learning), that schools have been collecting data for a long time. Institutions track grades, attendance, textbook purchases, test scores, cafeteria sales, etc. But little is done with this information - either due to data protection and security issues or they simply lack technical capacity - to enhance the efficiency of learning. With the use of ICT and the open government data, there are clearly many opportunities for data collection and analysis in education (Ujbanyi, Sziladi, Katona, Kovari, 2017).“This data includes any combination of: location, previous learning activities, health concerns (physical and emotional/mental), attendance, grades, socio-economic data (parental income), parental status, and so on. Most institutions will store and aggregate this data under the umbrella of institutional statistics. Education is, today at least, a black box. Society invests significantly in primary, secondary, and higher education. Unfortunately, we don’t really know how our inputs influence or produce outputs.”

Based on this idea, the authors would like to offer a new way of quality assurance. Not only learning efficiency could be enhanced, but efficiency of education too. Big Data could find correlations between new teaching methods, professors evaluation and students’ satisfaction of meals served at the cafeteria. In order, to do this, sentiment analysis should be applied.

Sentiment analysis is method for the analysis of opinions, feelings, subjective manifestations – a sort of opinion mining. Thus, sentiment analysis is a computerized analysis of online postings, comments, criticism, attitudes and emotions that are formulated with a product, service, organization, or, in general, some phenomenon, in our case, the education or some of its elements. The method is closely related to Big Data, as it helps in understanding the collected data better. Schools have benne collection data, but they did not have the necessary tools to analyze it in a complex way. Until now. The sentiment analysis has emerged as a result of the interconnection of language recognition systems and data mining and provides automated analysis of large amount of information. The advantage of this method is that it goes beyond the preconceived preconceptions and prejudices that appear in traditional interpretation of texts (Zhang and Liu, 2014: 1).
However, sentiment analysis is not without an antecedent, since the category of positive negative opinions and the classification of such have long been known. However, there are several new factors in the present network environment. Social media can be mined for opinions, Twitter content can be studied, and it is up to Facebook to keep its promise and allow researchers free access to data. This way, student posting, comments and likes about education can be used to gain important insights that can be latter used in quality assessment.

The description of the empirical inquiry

One of the main features of the information society is that an increasingly large number of social processes along with several aspects of human life take place via on-line networks and the respective data can be stored and analyzed digitally. Consequently, non-formal and informal instruction methods have become popular, more people gain access to knowledge, heretofore muted voices are amplified and public dialogues are

launched in several topics. Our essay focuses on such a pilot research project implying an empirical examination of the practical aspects of social media use described earlier. The respective quantitative and questionnaire-based inquiry utilizes a large sample including N=118 participants. The members of the sample, students of teacher training programs, responded to a survey performed in the spring of 2017. The results of the research process can be integrated into a connectivism-oriented methodological culture whose components can be applied both in micro and meso contexts and the respective learning environments.

Below we introduce the more unique and relevant aspects of the questionnaire by the help of descriptive statistics.
The following pie-chart diagram shows the age distribution of the respondents, who are mostly in-service teachers enrolled in part-time training programs. The chart indicates the prevalence of the 21-30 and 41-50 age groups leading to the conclusion that within the next 5-10 years the pedagogical profession has to undergo a generational renewal.

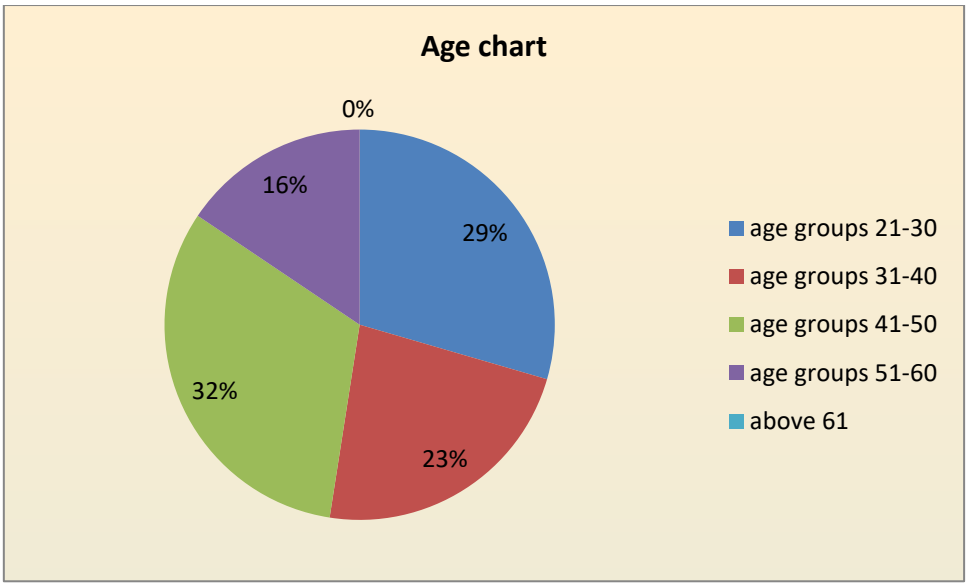

Figure 1 The age distribution of the respondents

The next diagram displays respondent distribution according to place of residence. Accordingly 39% of those responding to the survey live in Budapest, while 35% come from other cities in Hungary.

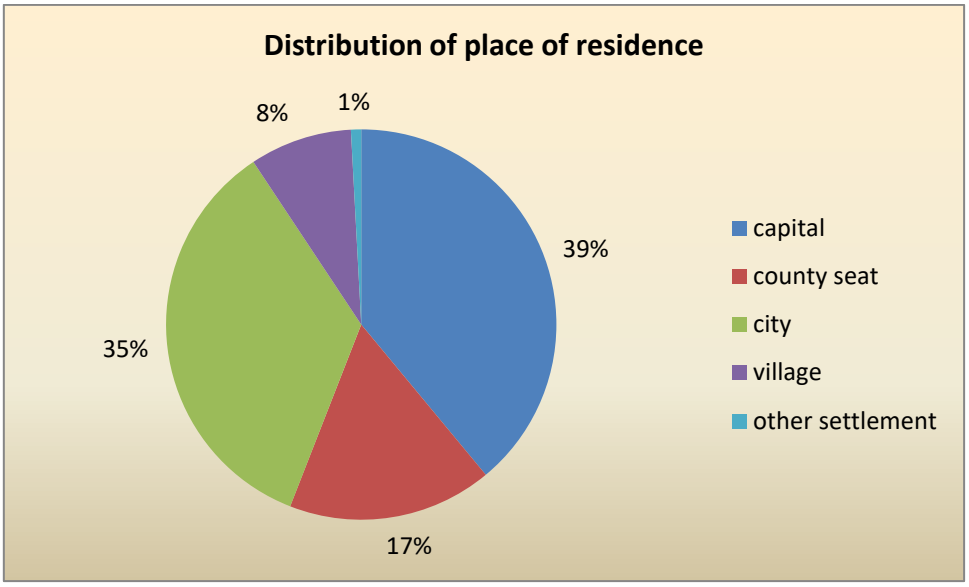

Figure 2 Distribution of respondents by place of residence

The next diagram focuses on the usefulness of the educational application of social media. About 2/3 of the respondents consider social media a useful and appropriate instruction tool.

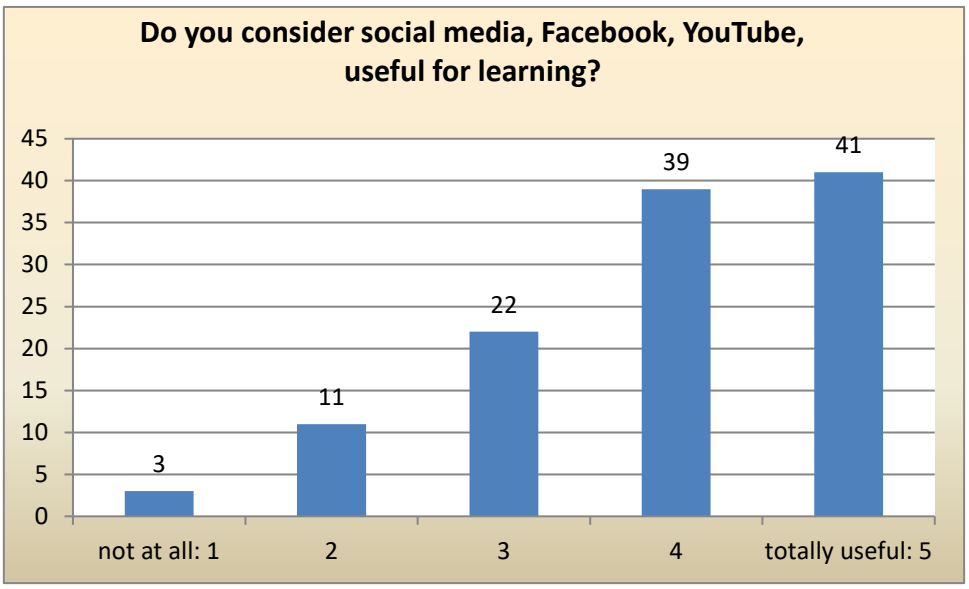

Figure 3 Views concerning the usefulness of social media in education

1/3 of respondents believe that social media offer a moderately suitable site for discussing of and familiarisation with learning-related issues and another 1/3 do not consider social media useful for such a purpose.

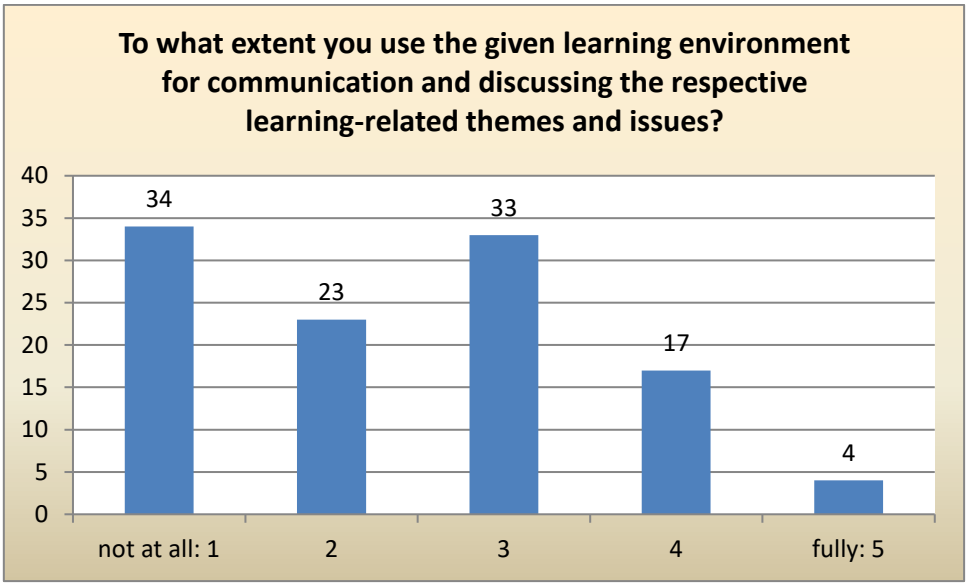

Figure 4 The suitability of social media for the discussion of learning-related issues

Regarding the rate of daily use of social media we can conclude that 60% use them on a daily basis and 1/3 of the respondents turn to social media-related platforms weekly.

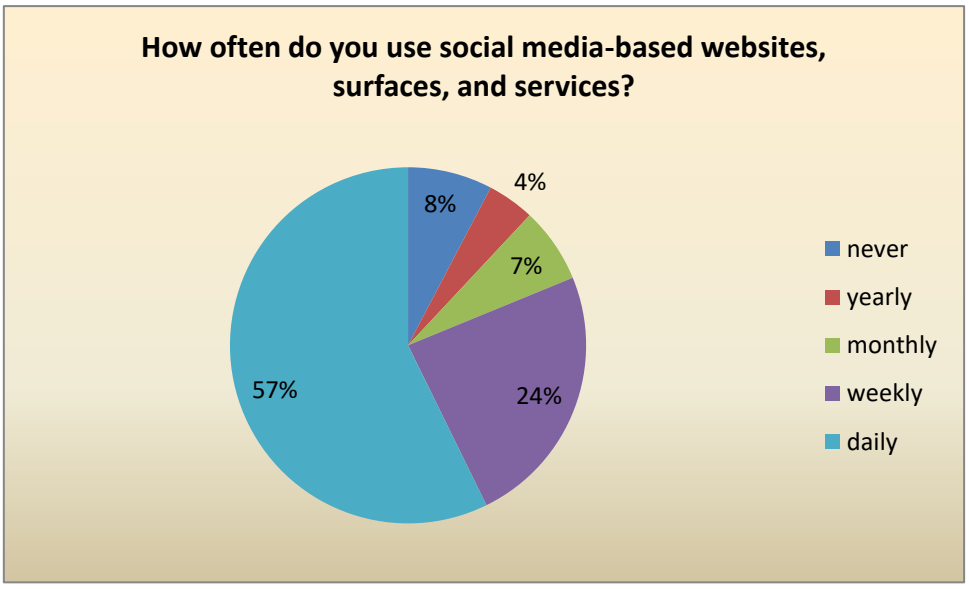

Figure 5 The distribution of the frequency of social media use

The last diagram displays the types of learning-related information ranking online contents supporting the learning process first, followed by the general learning-related news at second place. Information and educational material or text-related feedback, in itself a unique component of new network-based learning theories, is listed third. It is also indicated that ¼ of respondents do not share such type of content.

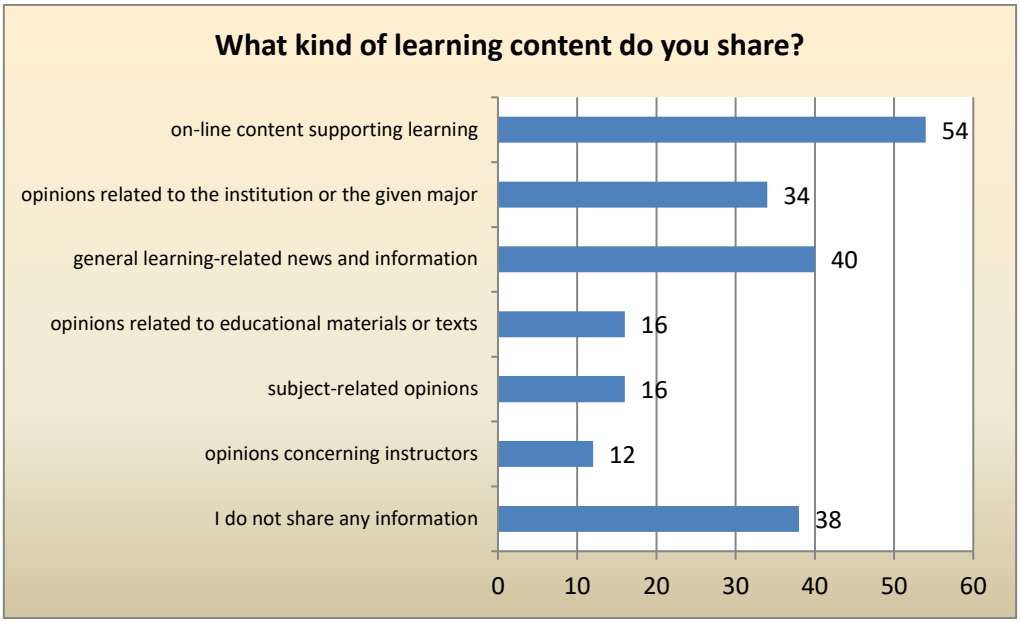

Figure 6 The distribution of sharing learning-related content

Summary and perspective

Big Data and sentiment analysis can provide valuable insights and point out the most important factors that influence the effectiveness of education (time spent on specific tasks, frequency of sign-in to online courses, role played within social networking sites, frequency and intensity of communication with peers or teachers). Peripheral data analysis can include tangible services: access to library resources and learning aids or career counseling services. However, the challenge is that analytics can not grasp the less perceptible elements of learning, such as the motivational motivation of the teacher and the value of informal social interactions.

The use of social media has become part and parcel of the information society. The most frequent users are members of the digital generation characterized by continuous on-line presence and immediate sharing of information. The research program introduced in this essay leads to the conclusion that the role of social media in providing support for the learning process is undisputable and not dependent on age group. Accordingly, besides general and personal information learning related data, views, and opinions are shared most frequently. The required equipment and digital literacy is available and an appropriate level of digital competence can be discerned regardless of age group. Members of the information society can take advantage of networks and the related options. Consequently, informal learning patterns tend to gain legitimacy and an increasingly wider application besides formal and non-formal schemes of knowledge acquisition.

References

A. Hanushek, Ludger Woessmann. (2010), Education and Economic Growth, Economics of Education (Amsterdam: Elsevier, 2010

Audrey Watters, How data and analytics can improve education, (2011), O'Reilly.com George Siemens on the applications and challenges of education data.https://www.oreilly.com/ideas/education-data-analytics-learning

Csepeli György (2015), A szociológia és a Big Data. Replika, 2015/3–4, 171–176.

Laney, Douglas (2012), 3D Data Management: Controlling Data Volume, Velocity and Variety. Gartner. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-ManagementControlling-Data-Volume-Velocity-and-Variety.pdf

Majkić, Zoran (2014), Big Data Integration Theory. Theory and Methods of Database Mappings, Programming Languages, and Semantics. New York, Springer. https://doi.org/10.1007/978-3-319-04156-8

Manovich, Lev (2012), Trending: The promises and the challenges of big social data. In: Gold, Matthew K. (2012 ed.) Debates in the Digital Humanities. Minneapolis, University of Minnesota Press. 460–475. https://doi.org/10.5749/minnesota/9780816677948.003.0047

McKinsey & Company Inc. (2006), Children's health first: McKinsey report, http://www.lenus.ie/hse/bitstream/10147/65281/1/Childrens_Health_First_McKinsey_Report_2006_.pdf

Tibor Ujbanyi, Gergely Sziladi, Jozsef Katona, Attila Kovari (2017), ICT Based Interactive and Smart Technologies in Education - Teaching Difficulties, INTERNATIONAL JOURNAL OF MANAGEMENT AND APPLIED SCIENCE 3.:(10.) pp. 72-77. (2017)

Wesley W. (2014 ed.), Data Mining and Knowledge Discovery for Big Data – Methodologies, Zhang, Lei – Liu, Bing (2014) Aspect and Entity Extraction for Opinion Mining. In: Chu, Challenge and Opportunities. New York, Springer.