ARPHA Proceedings

Proceedings TSNI-2021, 0000-0000
doi:xxx…

Textbook: Focus on Students’ National Identity

Open data in open education: epistemological and didactic aspects

Andrey A. Deryabin* (a), Aleksandr A. Popov (b)

(a) Russian Presidential Academy of National Economy and Public Administration, 119571, Moscow (Russia), 82 Prosp. Vernadskogo.
(b) Moscow City University, 129226, Moscow (Russia), 4-1, 2nd Selskokhoziastvenny Proezd; Russian Presidential Academy of National Economy and Public Administration; Novosibirsk State Technical University; popov-aa@ranepa.ru.

Abstract

The paper suggests a critical view on various formats of didactic materials - from classical and electronic textbooks to digital educational courses. The educational value of a digital educational program is mostly determined not by its media saturation or the abundance of applied online materials, but by an elaborated didactic concept, which development can be a complicated epistemological task. Using the example of the authors’ educational program in Data Science and Machine Learning for secondary school (900 participants) employing social-economic datasets, it is showed how epistemologically complex problem-solving curriculum can be implemented. A possibility to change how working with data and information are handled so that students can reconstruct a school subject, including a design of a customizable digital resources that may replace classic textbooks is discussed.

Keywords: distant learning, data literacy, epistemic thinking, school

Published by Moscow City University and peer-reviewed under responsibility of TSNI-2021 (Textbook: Focus on Students’ National Identity)

Introduction

* Corresponding author. E-mail: deryabin-aa@ranepa.ru

Within the framework of the traditional ‘classroom–lesson’ organization of learning, knowledge is represented by a textbook as a specific model of the subject matter. The structure and content of such model depend on the methodology used by the textbook’s authors, their competence, their principles of selection and construction of texts and visual materials. Under these conditions, any knowledge is subject to a certain reduction, and in the worst case, to substitution of both widely accepted and competing scientific concepts with one of possible approaches to their interpretation.

The standards of e-books that spread more than a decade ago and picked up by publishers have hardly met the hopes that digital textbooks would provide advantages in learning. Today many of them are merely digital wrapping of pre-existed printed textbooks. “Textbooks of the twenty-first century” designed with thoughtful and consistently implemented didactic approaches that employ the potential of digital technologies for learning and teaching still remain a desideratum for educators, students and parents (Behnke, 2018).

The growth of the market for mass online courses that began in the 2010s, the visioning of the AI-driven intelligent tutoring systems and adaptive learning in schools (Luckin et al, 2016) have not yet led to shifts in mass school education. Today, the practical implementation of a personalized learning model based on digital platforms is often implemented as an excessive scope of tasks with different levels of complexity, where a student can choose which tasks to solve at every level. The performing of complex problem-solving or creative tasks is encouraged but is not always obligatory. No matter how media-intensive or visually rich the resources of the online course (hypertext, animation, video, etc.) are, for the students it often remains a sandbox, fenced off by the formal requirements of the school curriculum. The abundance of links to external sources "for self-study" does not guarantee any advantage in terms of academic achievement or personal development of the student.

Purpose and objectives of the study

The aim of the study is to identify and test the possibilities of using open data and methods of their analysis for the design of educational programs addressed to secondary school. Particular attention is paid to the creation of a digital educational environment, rich in cognitive resources and attractive areas of self-determination, which allows to manage the independent educational activities of students in a distance mode.

Literature review

Several studies (Hofer, 2004, p.53) show that if students perceive knowledge on some issue as simple and finite, their search activity on the Internet will be short and superficial, and they will tend not to carry out a profound research with the integration of information from different sources. This and other examples (Kuhn et al, 2000; Smith et al, 2000; Mansfield & Clinchy, 2002) prove that not only complex online learning activities, but a routine Internet search involves student's epistemological thinking. A number of studies have also shown a link between the epistemological thinking of students and their academic achievement (Conley et. Al, 2004; Mason & Scirica, 2006).

In our opinion, whatever the media "wrapping" of the educational course, it is important to use open, design-creative educational tasks, the construction of which poses epistemological questions for developers and for the students. The creation of such didactic tools based on complex epistemological concepts bears significant difficulties, in particular, the need to select high-quality content. On the one hand, it should contain reliable information and methodologically correct ideas about a particular subject area, on the other hand, be perceivable for the student. An example of a solution to this problem can be the intensive educational program “Data Campus”, which integrates training in data analysis with team research and applied projects in the social and humanitarian sphere (Glukhov et al., 2020).

The approach of the authors of the “Data Campus” program is to some extent similar to the initiatives described in the literature for teaching schoolchildren in data science in Europe and the United States. An interesting experience is described in the article by Bryant et al. (2019), who organized a weeklong teen camp focusing on programming and data science “for the public good”. Instead of entertainment (robots, games, coding for Minecraft), epy students explored computational approaches to data science through working on topics of public interest, discovering how computing helps them not only better understand social issues, but also convince others to solve problems. The authors distinguish four levels of activity in their camp, which make it possible to draw the attention of adolescents to socially significant issues: (1) redefining their existing project for the good of society; (2) create a new project for solving a social problem; (3) a real-world case solved as an exercise; (4) not just an exercise, but a stakeholder-driven project to solve a real problem with real benefits.

This example can be attributed to a type of educational initiatives that focus not on mastering technical skills of data analysis, but on the overwhelming development of “critical data literacy” for public good and enhancement of civil rights (Wise, 2019). Practitioners and researchers in this field often address to open government and international data on a wide range of issues of high public interest (Pangrazio & Selwyn, 2019). Data journalism programs (Hewett, 2016; Burns & Matthews, 2018; Graham, 2018), workshops for non-profit organizations and community activists (Fotopoulou, 2020; Carroll et al, 2019) are examples of initiatives aimed at increasing data literacy among community groups.

Methodology

The research method was an ascertaining experiment, which was aimed at testing the educational program "Data Campus" under the conditions of both full-time term and a distance course. Depending on the format, the duration of the program ranged from 40 to 70 academic hours. The main sections of the program were presented by the themes "Socio-economic situation of the regions of Russia" and "Programming for Data Science and Machine Learning", given in the format of lectures, master classes, teamwork and presentation of analytical projects.

When shaping project teams, the students were asked to suggest any topic in one of the following areas (but not limited to them): Science and Education; Law Enforcement System; Healthcare; Culture; Ecology and Natural Resources; Economy, Industry and Trade; Agriculture; Transport and Energy; Finance; Communication and Information Technology.

The educational objectives of the program were as follows:

- to create conditions for independent probe-and-design activity in the field of Data Analysis and computer-aided learning, which includes setting a problem, formulating of analytical hypotheses based on real management problems in the social and economic sphere, analysis of available datasets, their interpretation and development on their basis of management scenarios for solution of the task;

- to create didactic conditions for formalizing the experience gained, the applied thinking patterns and organizing activities, with subsequent self-determination in the areas of data analysis, computer-aided learning as the chosen professional sphere;

- to form and maintain a cognitive and research interest in Data Analysis as a professional field and as a type of practice procedure - throughout the duration of the whole educational program;

- to provide an acquaintance with the mathematical foundations, basic methods, techniques, tasks and problems of modern data analysis and computer-aided learning, as well as with the most likely development trends in this area and points of growth (mainly through the organization of students' own research activities).

900 pupils of 8-11 grades of schools of the Kemerovo region took part in the ascertaining experiment in the period from 2019 to 2020.

Results

“Data Campus” was held both offline and online. In the latter case, the work was organized on the basis of Google Workspace, and the participants worked on their projects as part of geographically spread teams, communicating via video conferencing and chats via desktop and mobile devices. Collaboration of teams on the program code is carried out using Google Colab. All digital content required within the program is placed in Google Classroom, where progress is monitored, and individual and team achievements are assessed.

From the very beginning at "Data Campus", the participants do not immediately begin to program and analyze data arrays. First of all, the student is asked what problems of regional development are of concern to him. Among such problems, there can be completely different ones: social, demographic, medical, environmental, political, and so on. As a start, through discussions and analytical sessions, the teams perceive the problems of life and activity chosen according to their interests, work with a variety of regional analytical materials on that issues. They develop an assessment methodology which will provide them with evaluation of the level of significance of the problem for different territories.

In order to successfully master data literacy, being opposed to Data Science as a field of professionalization, the students are not required to get deep knowledge in mathematical apparatus and specialized software, but to acquire a combination of basic knowledge in this area with the ability to analyze. The dependence of more and more areas of our life on data actualizes such personality characteristics as critical and quantitative thinking in a comprehensively developmental context.

The students must independently set an educational and research task. In this regard, they receive an instruction like follows: “With the data available each group must formulate a grounded hypothesis that reveals some aspect of socio-economic reality. This can be the identification of connection, pattern, trend, or otherwise. You can use any data, include any features in your dataset, find any dependencies, build any models. By means of data analysis, you must confirm or refute your hypothesis, interpret the result of your research, propose a management solution to the problem".

Only after the team has decided what task it will work over, the participants have a need to work with data. And only at this stage, the study of programming tools for working with data begins, since these tools act as a key and a means of analysis within the framework of more voluminous and exciting problems for participants. At this stage, they understand how Data Science will help them in solving specific cases.

The participants of the "Data Campus" are provided with an excessive number of datasets based on open data (‘open data’ is a concept reflecting the idea that certain data should be freely available use and further republishing without restrictions for copyright, patents and other controlling mechanisms (e.g. http://opendatatoolkit.worldbank.org/en/essentials.html). In addition, they can use any data they find to solve their cases. The prepared datasets are, in most cases, "raw", unprocessed in nature, because processing, analyzing and interpreting such data helps students improve critical thinking skills, deepen their subject matter knowledge, and connect their quantitative analytical skills with Data Science and Machine Learning methods (Erickson et al, 2018; Kjelvik & Schultheis, 2019; Hardy et al., 2020).

The result of the program was completed student projects in data analysis and machine learning. We can give as an example the themes of some data projects made by students of the Data Campus:

1. Analysis of correlation between morbidity and environmental conditions in the regions.
2. Prediction of pass rates for universities in the region.
3. Analysis of the relationships between the Index of Happiness and economic variables.
4. Analysis of employment and outflow of young specialists in the regions of Russia.
5. Analysis of poverty and identification of vulnerable groups of the population.
6. Analysis of life expectancy, population size and ecological situation in the regions of Russia.
7. Identification of forest fires by aerial photography, fire hazardous regions and analysis of the causes of fire hazard.
8. Screening for pneumonia by X-ray photos.
9. Screening for skin cancer by photographs.
10. Predictive analytics of the choice of the level of education by school graduates in the region.
11. Classification of architectural styles by photography.
12. Forecasting the average monthly salary in the regions of Russia.
13. Analysis of the factors affecting the popularity of massive online courses.
14. Classification of household waste by pictures.

Discussions

Modern communication technologies and patterns of cultural reproduction over the past two decades have changed the way people communicate, raising important questions about the philosophy of the educational process. The traditional *ontology* of education assumed that it is realized in the mode of the controlling activity of a teacher regarding the goal-setting of students, the sequence and content of their activities. But the rapid digitalization of the educational environment makes it more important to *manage the conditions* of independent educational activities of maturing learners. It is carried out by creating a "programming space" that allows to independently design educational situations in the conditions of maximum availability of the necessary cognitive resources and attractive areas of self-determination (Uvarov et al., 2019. Pp. 181-235).

Data and tools for their analysis as components of a modern, data intensive digital environment are one of the key resources of this process (Cope et al, 2020). They provide an epistemological transition from direct *control* of student’s activity to creating systemic conditions for the learner to define his own transformative actions and master the management over these activities.

However, nowadays the curriculum of the school discipline "Informatics" contains data science and machine learning as just partial components of the basic curriculum. They are taught outside the context of creative problem-solving tasks or projects. Needless to say, they are not considered as factors giving rise to a fundamentally new epistemological paradigm either. At the same time, the amount of available information, open data and the means of their analysis allow a learner to operate with a much larger amount of knowledge than it was traditionally possible twenty years ago. In this regard, the philosophy and sociology of the educational process must change accordingly. These changes should affect not only the educational content given by the teacher, or characteristics of the educational space determined by her, but the principles and tools of working with data and information, which enable students to acquire necessary knowledge and reconstruct their personal ontology.

Open data and the tools for their analysis available today are a powerful resource for organizing learning in the open, activity-based approach. They give students the opportunity to independently select material for the study based on open publications and data in the mode of creative-project activity. In fact, we can talk about the independent work of students in the design of a school subject. Of course, such work should be facilitated by a teacher, and it should be preceded by the building of the supporting content components of the studied subject as both a system of scientific knowledge and a specially organized space for the student's independent activity.

Conclusion

Data analytics, used as an educational tool by the students, can contribute to making education truly *open*, that is, *firstly*, being based on the maximum consideration of all aspects of the personal capabilities and deficits of the maturing person; *secondly*, allowing him to solve the maximum possible range of educational tasks based on the maximum possible range of sources; *thirdly*, independently, without external control, make decisions regarding the strategy and tactics of educational advancement. But this will require both the introduction of new technological solutions and measures to develop students' instrumental competence - the ability to use data analysis and modeling tools for the purposeful solution of the assigned tasks.

The functions of open data and their analysis as a cognitive toolkit that expands the activity-targeted orientation of students' learning opportunities can be summarized as follows:

1. Reconstruction and design of educational material imputed for mastering by educational standards (in the limit - the design of optimal textbooks for each student). This actually makes it possible to make activity-based, not translational forms, basic and decisive in education.

2. The possibility to get basic knowledge in the process of acquisition of the maximum volume of relevant data, but not limited and "prepared" material from textbooks.

3. The ability to successfully conduct educational research, prepare for participation in olympiads and competency contests.

4. The ability to optimally select educational texts and interactive educational materials for the implementation of an individual educational trajectory, first of all, in connection with the preferred professional field and those requirements for subject knowledge that it will necessitate.

References

Behnke, Y. (2018). Textbook Effects and Efficacy. In: Fuchs, E., Annekatrin Bock, A. (Eds). The Palgrave Handbook of Textbook Studies (pp.383-396). https://doi.org/10.1057/978-1-137-53142-1_28

Bryant, C. et al. (2019). A middle-school camp emphasizing data science and computing for social good. In SIGCSE 2019 Proceedings of the 50th ACM Technical Symposium on Computer Science Education, (pp. 358–364). https://doi.org/10.1145/3287324.3287510.

Burns, L. S., Matthews B. J. (2018). First Things First: Teaching Data Journalism as a Core Skill. Asia Pacific Media Educator, 28(1), 91–105.

Carroll, J. M. (Ed.). (2019). Strengthening Community Data: Towards Pervasive Participation. In Proceedings of the 19th Annual International Conference on Digital Government Research, May 30-June 1, 2018, Delft, Netherlands (pp. 358–364). New York: ACM.

Conley, A. M., Pintrich, P. R., Vekiri, I., and Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. Contemporary Educational Psychology, 29, 186–204.

Cope, B., Kalantzis, M., Searsmith, D. (2020). Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. Educational Philosophy and Theory. https://doi.org/10.1080/00131857.2020.1728732

Erickson, T. et al (2018). Data Moves: one key to data science at school level. Proceedings of the International Conference on Teaching Statistics (ICOTS-10), Retrieved from https://iase-web.org/Conference_Proceedings.php?p=ICOTS_10_2018

Fotopoulou, A. (2020). Conceptualising critical data literacies for civil society organisations: agency, care, and social responsibility dilemma. Information Communication and Society, https://doi.org/10.1080/1369118X.2020.1716041

Glukhov P.P., Deryabin A.A., Popov A.A. (2020). Data-gramotnost'. Modnaya tema ili neobkhodimost'? Obrazovatel'naya politika, 3 (83). Retrieved from https://edpolicy.ru/data-science

Graham, C. (2018). A DIY, Project-based Approach to Teaching Data Journalism. Asia Pacific Media Educator, 28(1), 67–77.

Hardy, L., Dixon, C., Hsi, S. (2020). From Data Collectors to Data Producers: Shifting Students’ Relationship to Data. Journal of the Learning Sciences, 29(1), 104–126.

Hewett, J. (2016). Learning to teach data journalism: Innovation, influence and constraints. Journalism, 17(1), 119–137.

Hofer, B. K. (2004). Epistemological understanding as a metacognitive process: Thinking aloud during online searching. Educational Psychologist, 39(1), 43-55.

Kjelvik M. K., Schultheis, E. H. (2019). Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. CBE Life Sciences Education, 18(2), 1-8.

Kuhn, D., Cheney, R., Weinstock, M. (2000). The development of epistemological understanding. Cognitive Development, 15, 309-328. https://doi.org/10.1016/S0885-2014(00)00030-7

Luckin, R., Holmes, W., Griffiths, M., Forcier, L.B. (2016). Intelligence Unleashed. An Argument for AI in Education.

Mansfield, A. F., and Clinchy, B. McV. (2002). Toward the integration of objectivity and subjectivity: epistemological development from 10-16, New Ideas in Psychology, 20, 225-262.

Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial topics by epistemological understanding. Learning and Instruction, 16, 492-509.

Pangrazio, L., Selwyn, N. (2019). ‘Personal data literacies’: A critical literacies approach to enhancing understandings of personal digital data. New Media and Society, 21(2), 419–437.

Smith, C. L., Maclin, D., Houghton, C., Hennessey, M. G., (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological developments. Cognition and Instruction, 18(3), 349-422.

Uvarov, A.Yu., Geibl, E., Dvoretskaya, I.V., Zaslavskii, I.M., Karlov, I.A., Mertsalova, T.A., Sergomanov, P.A., Frumin, I.D. (2019). Trudnosti i perspektivy tsifrovoi transformatsii obrazovaniya. Moscow: HSE.

Wise, A. (2019). Educating Data Scientists and Data Literate Citizens for a New Generation of Data. Journal of the Learning Sciences, vol. 29, no. 1, p. 165–181. https://doi.org/10.1080/10508406.2019.1705678