Cloud technologies and learning analytics: web application for PISA results analysis and visualization

Mariia S. Mazorchuk[1,2], Tetyana S. Vakulenko[2], Anna O. Bychko[2], Olena H. Kuzminska[3] and Oleksandr V. Prokhorov[4]

[1]*V. N. Karazin Kharkiv National University, 4 Svobody Sq., Kharkiv, 61022, Ukraine*
[2]*Ukrainian Center for Educational Quality Assessment, 5 V. Vinnichenko Str., Kyiv, 04053, Ukraine*
[3]*National University of Life and Environmental Sciences of Ukraine, 15 Heroyiv Oborony Str., Kyiv, 03041, Ukraine*
[4]*The National Aerospace University "Kharkiv Aviation Institute", 17 Chkalov Str., Kharkiv, 61070, Ukraine*

Abstract

This article analyzes the ways to apply Learning Analytics, Cloud Technologies, and Big Data in the field of education on the international level. This paper provides examples of international analytical researches and cloud technologies used to process the results of those researches. It considers the PISA research methodology and related tools, including the IDB Analyzer application, free R intsvy environment for processing statistical data, and cloud-based web application PISA Data Explorer. The paper justifies the necessity of creating a stand-alone web application that supports Ukrainian localization and provides Ukrainian researchers with rapid access to well-structured PISA data. In particular, such an application should provide for data across the factorial features and indicators applied at the country level and demonstrate the Ukrainian indicators compared to the other countries' results. This paper includes a description of the application core functionalities, architecture, and technologies used for development. The proposed solution leverages the shiny package available with R environment that allows implementing both the UI and server sides of the application. The technical implementation is a proven solution that allows for simplifying the access to PISA data for Ukrainian researchers and helping them utilize the calculation results on the key features without having to apply tools for processing statistical data.

Keywords

learning analytics, Cloud Technologies, PISA, web application

1. Introduction

Currently, we can observe the rapid growth of the demand in Big Data Analytics and Business Analytics [1]. Big Data and Cloud Computing solutions empower real-time decision-making, identify trends, and allow for creating data models within the most powerful Data Analytics

CTE 2020: 8th Workshop on Cloud Technologies in Education, December 18, 2020, Kryvyi Rih, Ukraine
mazorchuk.mary@gmail.com (M. S. Mazorchuk); vakulenko_tetyana@ukr.net (T. S. Vakulenko); bychko.anya@gmail.com (A. O. Bychko); o.kuzminska@nubip.edu.ua (O. H. Kuzminska); o.prokhorov@khai.edu (O. V. Prokhorov)
https://www.univer.kharkov.ua/en (M. S. Mazorchuk); https://testportal.gov.ua/l (T. S. Vakulenko); https://nubip.edu.ua/en (O. H. Kuzminska); https://khai.edu/en/ (O. V. Prokhorov)
0000-0002-4416-8361 (M. S. Mazorchuk); 0000-0002-7403-1075 (T. S. Vakulenko); 0000-0001-5609-6978 (A. O. Bychko); 0000-0002-8849-9648 (O. H. Kuzminska); 0000-0003-4680-4082 (O. V. Prokhorov)

solutions. Cloud Technologies provide for a brand-new way of data processing to retrieve data and analytics valuable for business and decision-making [2].

The recent research [3], as well as the agenda of the latest international conferences [4, 5, 6, 7], prove that the Big Data processing approaches are applicable and relevant for education as well.

Horizon Report 2020 [8] considers Learning Analytics and Analytics of Student Data to be the most promising areas in educational technologies. Being powered by Cloud Technologies by Google, IBM, Amazon, and Microsoft [9], Learning Analytics can empower effective decision-making on the level of an educational institution, a region, or the whole world. Besides, a vast amount of data is available for the public, so it is easy to utilize for research.

On the local level, introducing the Learning Analytics can facilitate:

- building the educational trajectories for the college undergraduate students [10],
- leveraging data on the students' activities to support them during online learning [11], analyzing the influence of group behavior on the individual student's success [12],
- analyzing the External Independent Evaluation (EIT) data to decide on the school for a pupil [13].

Across the world, the researchers analyze the volumes of data received from the international exams. For instance, the educational community keeps on track of research that studies the influence of the family background (including social, economic, and cultural factors) on the students' success [14, 15].

However, many analysts still cannot take the advantage of the analytics provided by international research such as PISA, TIMSS, PIRLS, PIAAC, NAEP, and TALIS [16].

The results are available for the public, and it's possible to access the databases and technical data, but not every researcher has a full understanding of the research methodology and tools. For instance, to make full use of data, analysts require background knowledge of calculation methodology, data structure and data processing algorithms of the research they work with.

Focusing on the PISA project, which it is currently the most noticeable and comprehensive evaluation project in Ukraine [16]. In 2018, Ukraine participated in the Program for International Student Assessment (PISA) for the first time, yet the question of efficiently utilizing the PISA data is still open to negotiation [17].

The Ukrainian researchers would have leveraged PISA data as a consistent input for analysis and synthesis, indicating and solving problems in our education system. Though, published at the end of 2019, the PISA analytical reports were hardly utilized to analyze the current state of education in Ukraine.

We cannot get away from the fact that PISA reports contain analytical data that reveal how our cultural, social, economic, and educational environment influences the success of Ukrainian students. The data is evidenced by precise calculations and introduces the researchers to the most entire gamut.

However, not every analyst can manage the research results without additional training, which takes them time and effort.

Challenges that arise for them when it comes to processing data provided by PISA or any other international research are the following:

- the analysts have not enough background understanding the research structure and evaluation techniques,

- the analysts receive too much data, that seems to them unstructured and difficult to process,
- the analysts can be confused with the research-specific terminology.

We aim to make the results of any international research more accessible, to allow more analysts to use the analytical results for their independent research and get the most out of the supplied data.

This research suggests developing a web application that would allow the educational community to assess PISA results in the most transparent and user-friendly format.

2. Analysis of PISA tools for results processing

PISA utilizes three main software tools for data processing: the *IDB Analyzer* application, free environment for static data analysis R *intsvy*, and the *PISA Data Explorer* web application powered by cloud technologies.

IDB Analyzer is free software that allows for generating scripts for commercial SPSS (Statistical Package for the Social Sciences) or SAS (Statistical Analysis Software) packages, each of those has a cloud solution. The tool considers the students' selection when calculating the standard deviation and the scheme of the test book rotation (Plausible Value – the probable values of the students' results). IDB Analyzer generates code that allows users to process descriptive statistics and verify the statistical hypothesis without writing the code in SPSS or SAS. Though, utilizing *IDB Analyzer* requires both skills with SPSS and/or SAS, and installing commercial software. Besides, it requires data preprocessing (uploading and cleansing). Also, the researchers require certain skills and expertise with data processing.

intsvy is a free R program package for processing the PISA, TIMSS, PIRLS, PIAAC and ICILS results. This package includes such functions as data import, data analysis and results visualization. The data analysis function considers the complicated selection construct and possible values when calculating grades and standard variance, regression coefficients, correlation coefficients, and frequency array. The visualization tools allow for demonstrating aggregated data in standard graphic form, that can be adjusted via the open function code. Likewise, working with *intsvy* requires users minimal experience with R language and RStudio.

PISA Data Explorer is a cloud-based solution for processing PISA results, available at https://www.oecd.org/pisa/data/. PISA Data Explorer requires from the users at least minimal skills and background with statistical data processing. The main advantage of this tool is that the PISA data should not be uploaded and preprocessed. The whole data is available online. The calculations are conducted online as well. The tool's disadvantage is limited to the fact that a researcher can only conduct the data analysis limited to the built-in PISA Data Explorer functions.

We shall refer to the *Education GPS/Explore Data* service available at https://gpseducation.oecd.org/Home, as an example of an internationally applied application for PISA data display that allows for receiving the PISA estimate indicators.

This service generalizes data on results, publications, and conclusions of the researches held in OECD (PISA, TALIS, PIAAC). *Education GPS/Explore Data* organizes the research data according to the countries, topics and research agendas. Using this service does not require

any specific skills in data analysis. The navigation panel has useful data filters that allow for displaying values for the selected countries. The system stores huge amounts of data, yet due to the structure issues, sometimes it prevents users from finding the data they need.

Thus, we came to a conclusion, that any researcher who is interested in an independent study requires either a considerable experience with data analysis and statistical data processing, or should search for generalized data at the Education GPS resource.

Using international research results, many local researchers face challenges in understanding the data processing methods utilized by particular research centers. These challenges can occur both on the level of statistical data analysis and on the level of understanding the deep context of the research [18].

For instance, the problem with processing statistical data on student's contextual characteristics and learning environments can be caused by terminology.

Also, some PISA conclusions rely on the indexed received from a complicated mathematical model, which bases upon the survey results. These conclusions can be unobvious. To calculate all the values exponents, indexes, or distributions percentage the researcher has to process a big amount of data and understand the whole PISA evaluation strategy.

Thus, we can see how data processing issues make research complicated and prevent many researchers from utilizing Learning Analytics.

This paper suggests designing a web application that would ensure the Ukrainian researchers' quick access to PISA data in a comprehensive, interactive, and user-friendly format. This solution can become a highly effective tool to analyze PISA results for individual research and get the most out of the PISA data.

3. Results

3.1. The challenges addressed by the application

This work aims to design a web application for displaying PISA results for Ukraine and launch it at http://pisa.testportal.gov.ua/ to make the research data available for the public.

The proposed application will allow analysts, researchers, and teachers from various institutions to get the most out of the PISA research experience, including:

- retrieving the data on the evaluation of semantic and contextual assessments from the complicated PISA research structure,
- retrieving the PISA calculation on students' success (shown in PISA scores and the level passing of students' competence),
- retrieving the PISA calculations on contextual characteristics of the learning environment (including students' gender, institution's location, student's social and economic background),
- retrieving the PISA results that demonstrate the research indicators for the national level (e.g. the type of the educational institution can be specified on the national level),
- assessing the interactive data visualizations for Ukraine with descriptions for the PISA indicators,
- accessing open PISA data on Ukraine and the benchmark countries,

- PISA content localized for Ukrainian users.

The detailed descriptions for PISA indicators and results implementation in the format of hypertext guide users through the process of searching and help them understand the strategic evaluation results.

The content in Ukrainian allows more Ukrainian speaking users to utilize PISA results for their research.

3.2. Application structure

The application structure aligns with the structure of PISA research. The survey results allow the researchers to assess various factors that determine the current condition of educational institutions and their effects on the students' success. This allows for analyzing the causes and effects of the education system in a selected country.

The PISA program is a three-year cycle. The evaluation takes place every three years and the results are revealed at the end of each cycle. The results of each cycle are not connected. These results can be compared in a timeline on the countries and key indicators, that provides a thorough grounding for strategic solutions in education.

The research structure consists of two main parts: the results on student's subject competence in math, reading, and natural science and the meaning of various indexes that demonstrate the contextual student's characteristics, educational environment, motivation, and other factors that can influence the students' success.

The data should be displayed both for each cycle separately and in a timeline. Figure 1 demonstrates the main elements of the PISA research results structure.

The questions from the student's survey reflect the main goals and objectives of the PISA research. These questions highlight the key factors required to analyze what influences the students' success. The survey also allows for collecting data on such important for representative analytics factors as:

- gender,
- educational institutions' locations,
- education programs and
- students' social and economic status.

The key indicators shaped on the international level allow for comparing the results and backgrounds both for the same and different categories of students within one country. This allows us to compare results for students from different countries based on their gender, region or social and economic status. The type of educational institution or educational program serves as a national indicator and allows for evaluating results only on the level of a particular country.

The index values serve as self-sufficient research results that can be analyzed on the international level to compare students within the country, or countries on the given parameters, and also as factor values which influence student's success and a range of other indicators.

For instance, the index of students' social, economic, and cultural statuses demonstrates not only the financial situation in a family, but the social and cultural aspects of their lives including their parents' education, availability of books, musical instruments, educational software, etc.

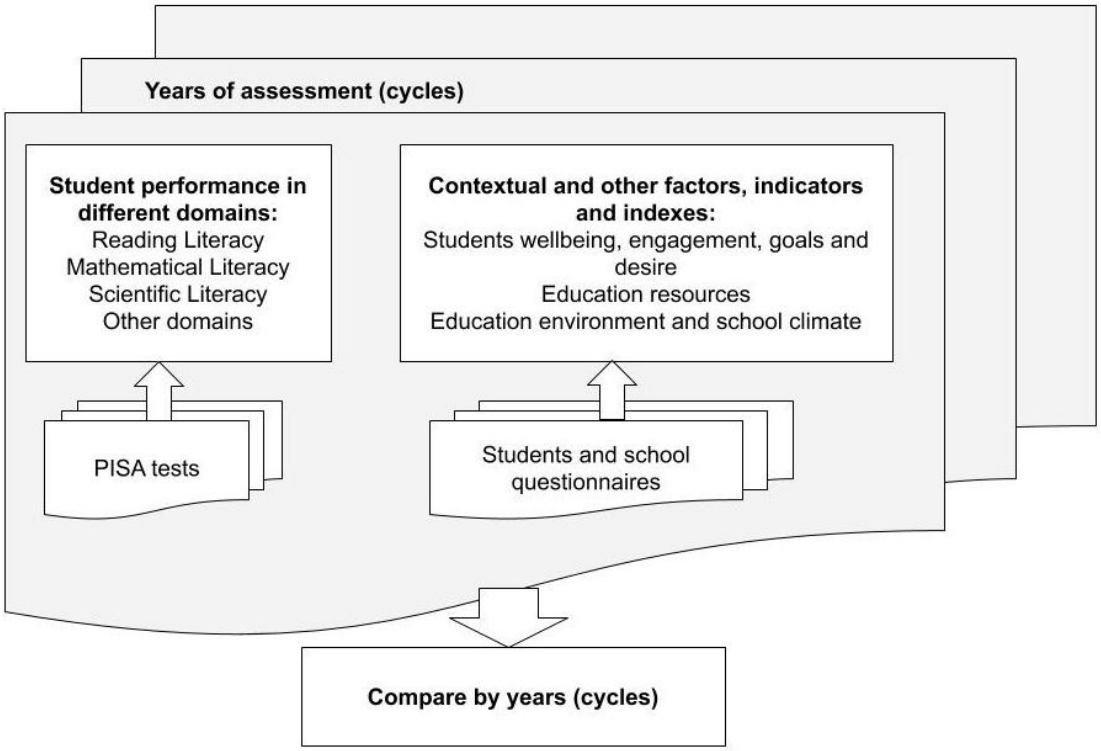

Figure 1: The structure of the PISA study.

The index value is an important indicator of the students' social, economic, and cultural statuses compared to their peers from other countries. Besides, index value influences both students' success and other indexes received during the analysis.

The values of the main indexes, criteria and characteristics can be divided into several categories:

- wellbeing, students' ambitions, expectations, and attitude to education,
- resources invested in education,
- the learning environment and school climate.

Each category has a set of indexes and criterias, received from the questioning students and their educators. Calculated via complicated mathematical models, these indexes allow for comparing values on the international level and separate categories of students. The indexes mostly summarize student' input to the survey and are considered to be imputed relative values that characterize some aspect of the research.

The indexes are compared based on OECD, an average of the countries, that equals to 0. Thus, the indexes mentioned on the graphs or in tables should have explanations to help users interpret the displayed data correctly.

Also, we can use interactive graphs to visualize the real-life condition of Ukrainian education compared to the reference countries with similar cultural, social, or economical conditions or

shared historical background. With a convenient visualization, the user can concentrate on the appropriate for comparing educational systems with similar social and cultural backgrounds.

Studying the experience of countries with similar backgrounds and factors that influence their students' success is more appropriate than comparing Ukrainian experience to the experience of countries with better initial conditions for education.

3.3. Technical implementation

The Ukrainian Center for Educational Quality Assessment experts calculate PISA results with *intsvy package* in R environment and store them in text documents or tables (.csv or .txt). Based on that data, we build diagrams to demonstrate the distribution of indicators. Figure 2 demonstrates the general structure of the results visualization.

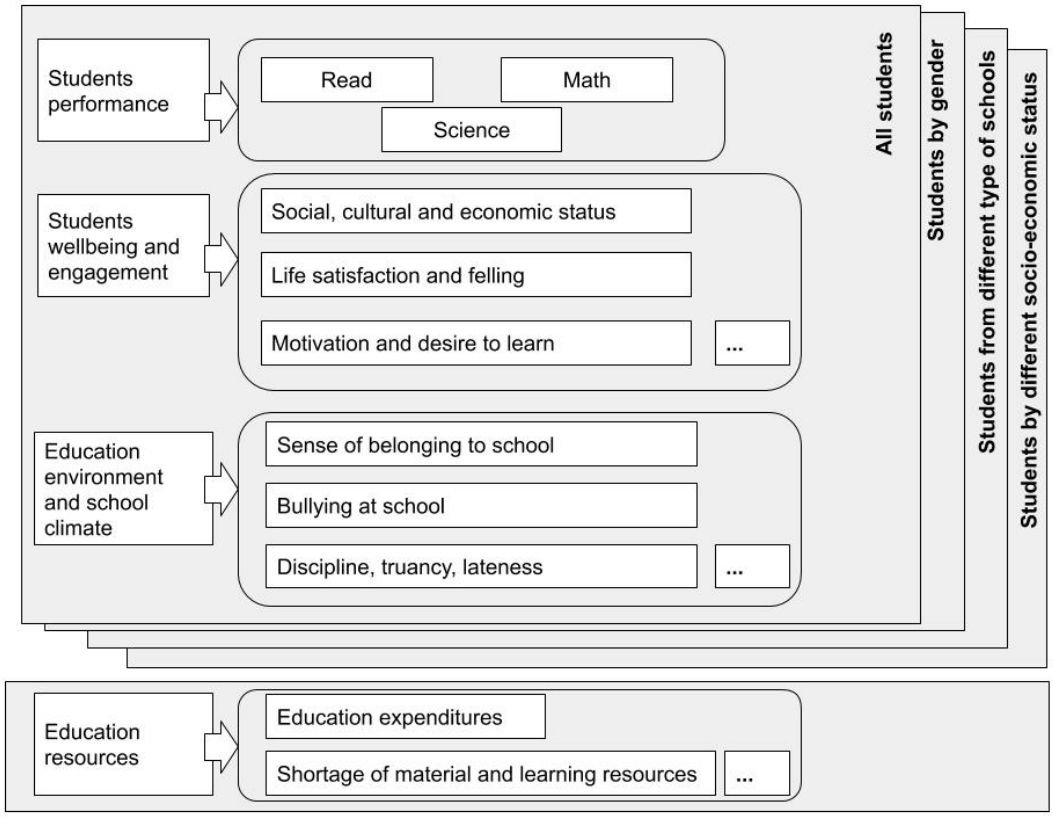

Figure 2: Generalized structure of visualisation of PISA results.

We utilized R Shiny package (https://shiny.rstudio.com/) to design both the application interface and the server sides, and deploy the application to the Cloud [19]. Figure 3 demonstrates the application structure. The application provides for calculating criteria based on PISA results in R via the *intsvy* package. The application stores the calculation results in the text files used to visualize data on the server-side. Thus, the application doesn't conduct any calculations and

doesn't require any time to upload data and process the results.

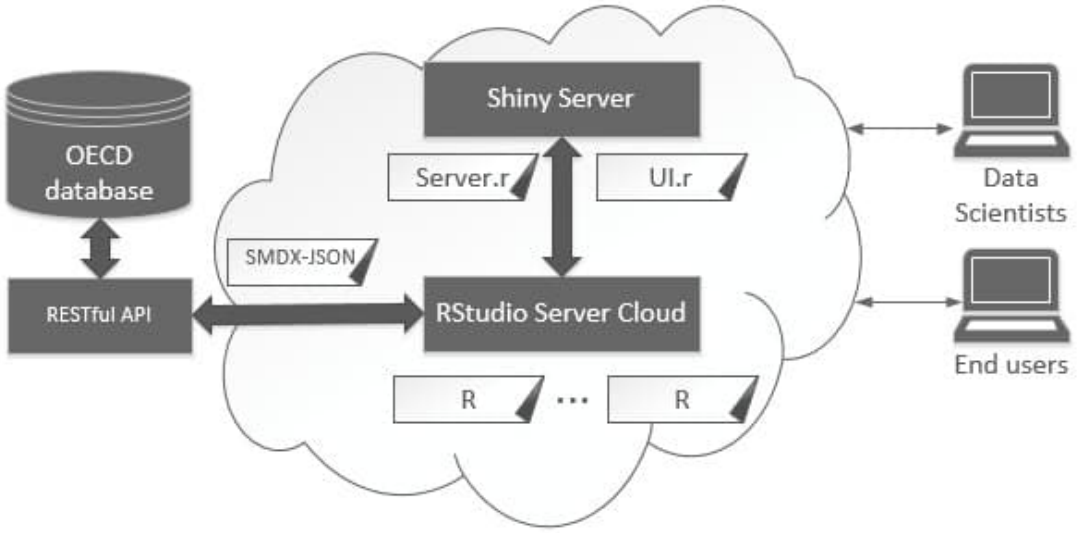

Figure 3: The technical implementation.

The application interface is designed as an interactive dashboard and provides for navigation, interactive menu, graphics, and hypertext links.

Figure 4 demonstrates the application home page and the main navigation elements that allow users to go to the data structure and access the diagrams of PISA scores distribution and students' competence levels.

Figure 5 demonstrates an interactive graphic that displays the data on student's success depending on countries and categories. This graphic allows users to compare different categories of students from different countries and see the factors that influence students' success.

With the Description option, the users can get explanations for the main criteria and indexes. This option is available both from the context menu and from the hypertext prompt message.

4. Conclusion

Working on this project we came to the conclusion that applying Learning Analytics is not only a trend but an effective tool for improving the system of education at all levels. Utilizing the international level results of Learning Analytics can be challenging for the local researchers. These challenges are connected with managing data analysis tools, understanding the research methodology, and the analysed indicators in the local context.

Ukraine participated in PISA in 2018, yet still, only a limited number of researchers work on processing that data. These results are important to understand the problems and make decisions on the different levels of education in Ukraine.

Ukrainian specialists have challenges leveraging the PISA data. The PISA calculations methodology is difficult to understand for researchers without specific background with tooling and a

Figure 4: Home page of application.

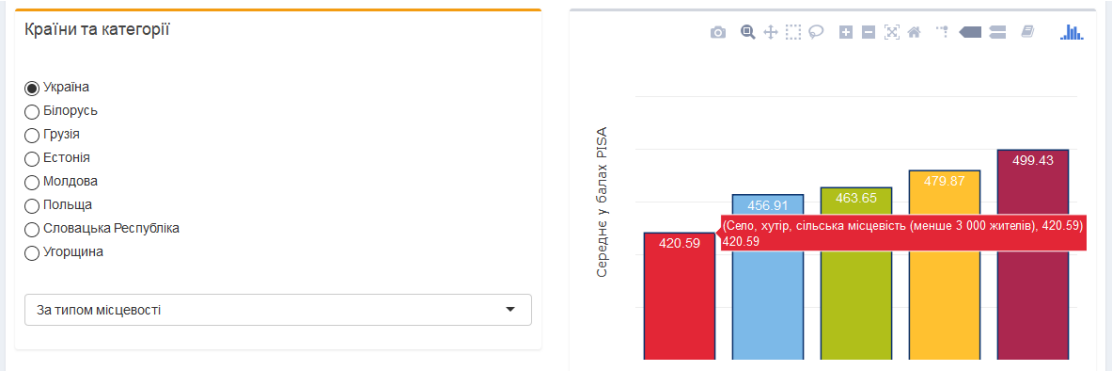

Figure 5: Interactive graphic.

good level of English. Thus, this work proves that having an application that would simplify accessing PISA results is pertinent.

The application we suggest is a convenient tool for a wide range of people working in the education segment who utilize PISA results for their research. This web application will allow for assessing analytical data based on PISA results in the most convenient, comprehensive, and user-friendly format.

This will be an effective tool for accessing PISA data in Ukraine and utilizing the calculations' results on the key indicators. What is more the users should not apply the tools for statistical data processing. Currently, we only implemented a part of the main functionality. Though,

we plan the updates that would allow for uploading the calculated results in table format and utilizing the whole learning analytics on PISA results in Ukraine.

References

[1] The Digital Enterprise: moving from experimentation to transformation, 2018. URL: http://www3.weforum.org/docs/Media/47538_Digital%20Enterprise_Moving_Experimentation_Transformation_report_2018%20-%20final%20(2).pdf.

[2] K. Andriushchenko, V. Rudyk, O. Riabchenko, M. Kachynska, N. Marynenko, L. Shergina, V. Kovtun, M. Tepliuk, A. Zhemba, O. Kuchai, Processes of managing information infrastructure of a digital enterprise in the framework of the «Industry 4.0» concept, Eastern-European Journal of Enterprise Technologies 1 (2019) 60–72. URL: http://journals.uran.ua/eejet/article/view/157765. doi:10.15587/1729-4061.2019.157765.

[3] O. Heyer, From Learning to Data Analytics: Some Implications for IT Strategy and Transformation, EDUCAUSE Review 54 (2019).

[4] A. Kiv, V. Soloviev, S. Semerikov, CTE 2018 – How cloud technologies continues to transform education, CEUR Workshop Proceedings 2433 (2019) 1–19. URL: http://ceur-ws.org/Vol-2433/paper00.pdf.

[5] A. Kiv, M. Shyshkina, S. Semerikov, A. Striuk, M. Striuk, H. Shalatska, CTE 2019 - When cloud technologies ruled the education, CEUR Workshop Proceedings 2643 (2020) 1–59. URL: http://ceur-ws.org/Vol-2643/paper00.pdf, 7th Workshop on Cloud Technologies in Education, CTE 2019 ; Conference Date: 20 December 2019.

[6] A. E. Kiv, V. N. Soloviev, S. O. Semerikov, XII International Conference on Mathematics, Science and Technology Education, Journal of Physics: Conference Series 1840 (2021) 011001. URL: https://doi.org/10.1088/1742-6596/1840/1/011001. doi:10.1088/1742-6596/1840/1/011001.

[7] A. E. Kiv, V. N. Soloviev, S. O. Semerikov, A. M. Striuk, V. V. Osadchyi, T. A. Vakaliuk, P. P. Nechypurenko, O. V. Bondarenko, I. S. Mintii, S. L. Malchenko, XIII International Conference on Mathematics, Science and Technology Education, Journal of Physics: Conference Series (2021).

[8] M. Brown, M. McCormack, J. Reeves, D. C. Brooks, S. Grajek, B. Alexander, M. Bali, S. Bulger, S. Dark, N. Engelbert, K. Gannon, A. Gauthier, D. Gibson, R. Gibson, B. Lundin, G. Veletsianos, N. Weber, 2020 EDUCAUSE Horizon Report: Teaching and Learning Edition Available, EDUCAUSE, Louisville, CO, 2020. URL: https://library.educause.edu/-/media/files/library/2020/3/2020_horizon_report_pdf.

[9] M. S. A. El-Seoud, H. F. El-Ssofany, I. A. T. F. Taj-Eddin, A. Nosseir, M. M. El-Khouly, Implementation of Web-Based Education in Egypt through Cloud Computing Technologies and Its Effect on Higher Education, Higher Education Studies 3 (2013) 62–76. URL: http://www.ccsenet.org/journal/index.php/hes/article/view/27423. doi:10.5539/hes.v3n3p62.

[10] V. Pavlenko, A. Prokhorov, O. Kuzminska, M. Mazorchuk, Competence approach to modeling and control of students' learning pathways in the cloud service, CEUR Workshop Proceedings 1844 (2017) 257–264.

[11] S. Walker, T. Olney, C. Wood, A. Clarke, M. Dunworth, How do tutors use data to support

their students?, Open Learning: The Journal of Open, Distance and e-Learning 34 (2019) 118–133. URL: https://doi.org/10.1080/02680513.2018.1554476. doi:10.1080/02680513.2018.1554476.
[12] J.-H. Zhang, Y.-X. Zhang, Q. Zou, S. Huang, What learning analytics tells us: Group behavior analysis and individual learning diagnosis based on long-term and large-scale data, Journal of Educational Technology & Society 21 (2018) 245–258. URL: http://www.jstor.org/stable/26388404.
[13] EIT 2008-2020: zvity, 2020. URL: https://testportal.gov.ua/ofzvit/.
[14] G. Yang, M. Badri, A. Al Rashedi, K. Almazroui, The role of reading motivation, self-efficacy, and home influence in students' literacy achievement: a preliminary examination of fourth graders in Abu Dhabi, Large-scale Assessments in Education 6 (2018) 10. URL: https://doi.org/10.1186/s40536-018-0063-0. doi:10.1186/s40536-018-0063-0.
[15] H. W. Ko, Y. L. Chan, Family factors and primary students' reading attainment, Chinese Education & Society 42 (2009) 33–48. URL: https://doi.org/10.2753/CED1061-1932420302. doi:10.2753/CED1061-1932420302.
[16] S. Arikan, A regression model with a new tool: IDB analyzer for identifying factors predicting mathematics performance using PISA 2012 indices, US-China Education Review 4 (2014) 716–727.
[17] PISA-2018: zvity, 2020. URL: http://pisa.testportal.gov.ua/pisa-2018-zvity/.
[18] D. V. Vasyl'ieva, M. V. Holovko, Y. O. Zhuk, O. H. Kozlenko, O. I. Liashenko, S. O. Naumenko, V. I. Novos'olova, UROKY PISA-2018: metodychni rekomendatsii, Instytut pedahohiky NAPN Ukrainy, Pedahohichna dumka, Kyiv, 2020, p. 96.
[19] H. Wickham, Mastering Shiny, 2020. URL: https://mastering-shiny.org/index.html.