ПРОГРАМНО-ТЕХНІЧНІ
КОМПЛЕКСИ

UDC 004.9

D. MURATULY
D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan,
e-mail: *dmuratuly@edu.ektu.kz*.

N.F. DENISSOVA
D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan,
e-mail: *ndenisova@edu.ektu.kz*.

Iu.V. KRAK
Taras Shevchenko National University of Kyiv, V.M. Glushkov Institute of Cybernetics
of NAS of Ukraine, Kyiv, Ukraine, e-mail: *Iurii.krak@knu.ua*.

INFORMATION TECHNOLOGY FOR THE PROCTOR TO DETECT VIOLATIONS DURING THE EXAM[1]

Abstract. This article discusses the current problem of cheating students of higher educational institutions during the exam. The subject area is investigated and the results of theoretical and empirical research of cheating as one of the forms of academic fraud of students of higher educational institutions are presented. Statistical data of violations during the exam are shown and the main patterns of violations during the period of distance learning are revealed. The results of the study will help to determine in what period of the final certification students violate the rules of academic honesty more often.

Keywords: distance learning, timing of violations, examination session, clustering, decision making.

INTRODUCTION

The COVID-19 pandemic has led to the largest disruption to education systems in history, affecting nearly 1.6 billion students in more than 200 countries and all continents [1]. The closure of educational institutions affected 94 percent of the world's student population. At the same time, it should be noted that the crisis served as a stimulus for innovation in the field of education. Innovative approaches are used to ensure the continuity of education and training [2]. As indicated in the "Digital Skills of Teachers" section, the skills of teachers should be digitally oriented to find, evaluate and use learning materials in the Internet [3].

In the study [4] it was revealed that after Covid-19, the proctoring system will be very useful for universities that can implement it.

Note that the dialogue around online education has moved to a new level. Various aspects of online courses are reviewed and discussed, such as modality (completely online or hybrid; synchronous or asynchronous), technology platform, assessment and accessibility [5]. The goal of such a dialogue is to develop modern proctoring systems in which examinations are at least as good as traditional examinations [6, 7]. All over the world, the number of educational institutions that

---

[1] The study is carried out within the framework of the project funded by the Ministry of Education and Science of the Republic of Kazakhstan AP08856846 "Methodology for creating a liberal model of On-Line education for higher educational institutions of the Republic of Kazakhstan".

provide education through distance learning using an automated proctoring system for online observation, recording and assessing the behavior of students has recently increased. For many university students, flexibility in time to implement distance learning is seen as an advantage. But flexibility is not only an advantage, but also a disadvantage [8]. The COVID-19 pandemic and the associated economic crisis have presented higher education with unprecedented challenges. For example, the sudden transition to online education in the spring of 2020 in Kazakhstan showed that the transition to distance learning from the technical part will be a difficult test for the universities of the republic: organization of the educational process, prevention of cheating during distance learning, and grading etc.

So, educational institutions faced the question of solving the control problem as quickly as possible using an online proctoring system based on artificial intelligence technologies. In the paper [9] author substantiates the role of proctoring in online testing as a method of reducing the scale of cheating, which means unfair behavior of students (cheating, providing unauthorized assistance to a student from another student or a tutor, up to completing study assignments for the student, etc.).

The definition of the concept of negligence in the examination as a deliberate violation, contrary to the official rules of the examination, aimed at putting the candidate in an unfair advantage or disadvantage is given in the paper [10]. Unfortunately, we do not yet have to rely on the conscious and responsible attitude of all students to the educational process [11]. In turn, authors in their article [12] identify the main problem, which is the perception that academic integrity associated with online tests is compromised by undetected fraud, which leads to artificially higher scores. Academic dishonesty is any behavior aimed at misrepresenting one's academic work as original and can take many forms, including deception, use of unauthorized resources, collusion and plagiarism [13, 14]. On the basis of D. Serikbayev East Kazakhstan Technical University (EKTU), the relevant problem of cheating students of higher educational institutions during the exam was investigated. The need to study this problem lies in the fact that at present there is an acute shortage of qualified specialists who perform their professional activities with high quality, which undoubtedly affects the quality of services provided in such spheres of society as medicine, education and others.

The goal of the given study is information technology creation for analyze of violations in the process of distance learning of students using the example of an exam session and recommendations for proctors to identify violations recommendations develop. The object of the research is the task of increasing confidence in the results of evaluating events carried out in a remote form. The subject of research is violations analysis in the process of distance learning. The study was conducted among students in difference university schools, a total of 12544 records were processed. The proctoring system was introduced at the university in 2019.

METHODOLOGY FOR ORGANIZING AN ON-LINE EXAM

The most of the universities conduct a significant part of the exams online. The bottom line is that through the video conferencing system, the student connects to pass the exam or test from his computer. The examiner (proctor) checks the identity of the student (the student submits his document), and then certification is carried out orally or in the form of computer testing under the supervision of the proctor. While the digitalization process offers more opportunities, it also poses a number of challenges for higher education institutions. Student authentication is recognized as an important issue in online education, as shown in Fig. 1.

In the article [15] emphasized, that students must be authenticated before they can access confidential content such as tests, assignments, or personal notes. Therefore, as online education and e-assessment is evolving, it is critical to improve student authentication. If universities can provide secure and convenient systems for electronic authentication, they can create a more reliable environment in which they will offer a variety of research to all students.

Forms of conducting exams during the midterm period are:

1) computer testing;

2) online oral examination using the video conference systems;

Fig. 1. Stages of passing the exam by the examinee

3) open book, essay/creative exam. The essay attached by the student must be checked for plagiarism.

Computer testing is carried out using the proctoring system developed at EKTU. This system connects to the testing system and records the student's desktop and himself with the help of a webcam into one file, which will be available the next day after the end of the exam and is stored on the servers of the university for 18 months, as shown in Fig. 2.

During exam, proctors are sure to be present, they observe the passing of students' test. For this target a special room have been created with randomly filled with students and proctors.

Before the start of the exam, the proctor identifies the student. After that, the student enters the testing system. To facilitate the proctor's work, the proctoring system warns him if more than one face appears in the frame, the noise level increases, and during the exam, the opening of applications on the test taker's computer is blocked, except for those allowed.

If the student's connection is broken, the test is dimmed and a reconnection to the system occurs. During reconnection, the student cannot continue testing. The proctor can send a comment to the student or complete the test if the student has received more than 5 comments.
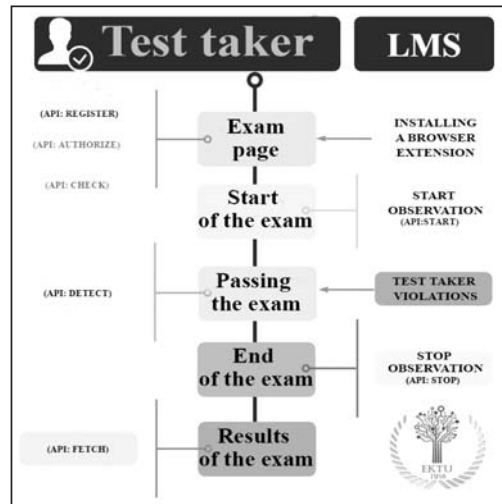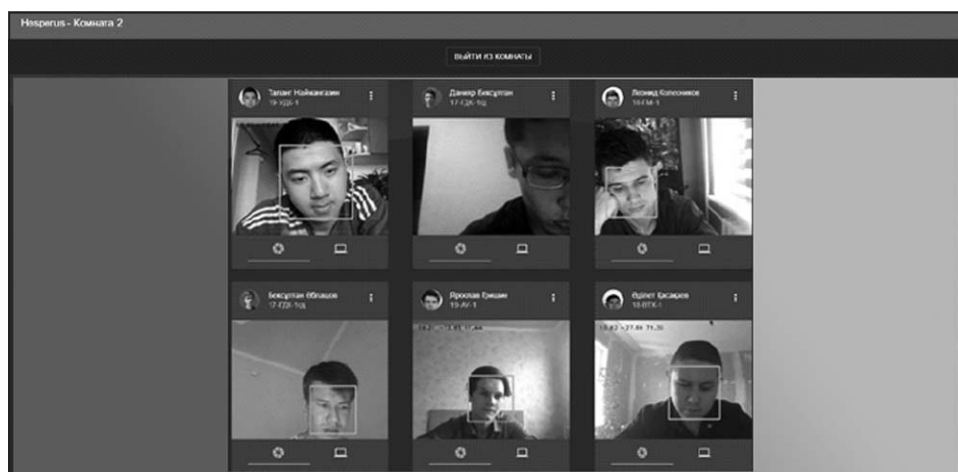

Fig. 2. An example of passing the exam from the proctor's side

Comments	Time of violation
Can you place the camera lower, I see only half of the face	27.10.2020 9:29:00
Can you put the camera in the center	30.12.2020 16:12:14
Can you sit in the center or put your camera in the center	23.10.2020 12:00:57
Can you put the camera in the center	15.12.2020 15:01:44
Can you sit in the center or put your camera in the center	28.12.2020 14:07:54
Can you sit in the center or put your camera in the center	23.10.2020 12:10:04
Can you place the camera lower, I see only half of the face	26.10.2020 10:00:14
Before the start of the exam, please adjust the camera so that you can be seen better	15.12.2020 14:35:13
close unnecessary tabs on the desktop	02.12.2020 15:06:03
close unnecessary tabs on the desktop	15.12.2020 11:04:33
close unnecessary tabs on the desktop	26.10.2020 10:02:24
close unnecessary tabs on the desktop	26.12.2020 15:20:52
Before the start of the exam, please adjust the camera so that you can be seen better	26.10.2020 10:04:48
Can you sit in the center or put your camera in the center	23.10.2020 10:05:53
Before the start of the exam, please adjust the camera so that you can be seen	23.10.2020 10:39:11
Can you place the camera lower, I see only half of the face	23.10.2020 10:07:13
close unnecessary tabs on the desktop	26.10.2020 10:09:10
Before the start of the exam, please adjust the camera so that you can be seen	30.10.2020 12:59:44
close unnecessary tabs on the desktop	28.12.2020 9:14:35
Can you sit in the center or put your camera in the center	24.12.2020 11:03:10
Before the start of the exam, please adjust the camera so that you can be seen	28.12.2020 9:07:41
Can you place the camera lower, I see only half of the face	26.10.2020 10:05:41
close unnecessary tabs on the desktop	31.10.2020 10:09:18
close unnecessary tabs on the desktop	30.10.2020 14:13:00
Can you sit in the center or put your camera in the center	26.10.2020 10:03:54

Fig. 3. Proctor comments and timing of violation

A LARGE AMOUNT OF DATA PROCESSING

Data processing occurs when data is collected and converted into useful information. Typically, data processing is performed by a data scientist or a group of data processing specialists, so it is important that data processing is performed correctly so as not to adversely affect the final output of the data. Data processing starts with the data in its raw form and transforms it into a more readable format, giving it the form and context necessary for interpretation by computers and use by employees of the organization.

During the study of the subject area, 12544 records were processed and all records were divided into classes: "Comments" — 7692 records, "No comments" — 3647 records, "Empty records" — 1205 records. At the second stage, data clustering was carried out, since clustering automatically identifies groups of semantically similar remarks made by proctors, as shown in Fig. 3. Groups are formed only on the basis of pairwise similarity of text notes, and no characteristics of these groups are set in advance, in contrast to the classification of documents, where categories are set in advance.

Clustering can be either for a purely research purpose (to find out the structure of a data) or for a better search and presentation of information, also, clustering of textual information in educational materials is one of the tasks of information retrieval [16, 17]. Text array clustering allows to organize the data and narrow the scope of the documents in question. If necessary, you can carry out clustering of the second level, applying the same partitioning procedure to the clusters obtained at the first stage. Thus a receives a tool for solving problems associated with the analysis of large arrays of textual information, as shown in Fig. 4.

Note that clusters are used not only for a compact representation of existing objects, but also for the recognition of new ones. Each new object belongs to the cluster, joining to which satisfies the clustering quality criterion in the best way. This means that it is possible to predict the behavior of an object, assuming that it will be similar to the behavior of other cluster objects.

Comments ▾
lift your head
lift your head
lift your head higher
lift your head higher
do not lower your head and set the camera so that you can be seen
do not lower your head and set the camera so that you can be seen
raise your head and immediately answer questions
lift your head
lift your head
lift your head higher
lift your head

Fig. 4. Semantic keyword clustering

Algorithms for forming clusters, based on the means method, divide the elements of the set into a known number of clusters, while minimizing the total square deviation of the points of the clusters from their centers:

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} (x_j - \mu_i)^2,$$

where S_i, $i = \overline{1, k}$, are the obtained clusters, k is the number of clusters, μ_i are the centers of mass of the vectors.

Each new division into clusters occurs with the calculation of the center of mass of clusters from the previous iteration. After that, the next division into clusters occurs, and so on until the center of mass does not change.

In the k-median method, instead of the average value, the median of the set is calculated to determine the center of the cluster. The task of partitioning clusters using the k-median method is to find the centers of clusters so that the resulting clusters are compact. Cluster centers are chosen so that the sum of distances between data points to cluster centers was minimal. Also, the k-median method is effectively uses for reduce the dimension of the feature space [18, 19]. The main difference of the method from others is that the algorithm can be trainable or non-trainable. An untrainable cluster search algorithm consists in direct partitioning into k clusters (k-way) or by the recursive bisection method.

Despite the different nature of texts, a lot of methods of their analysis are common. In particular, this concerns the models of the geometric representation of clusters, the choice of metrics and classification methods [20]. For evaluation of the compliance of the obtained clusters with the internal semantic structure of the text corpus, we will use the text corpus obtained in the process of control of students' knowledge and equipped with the correct semantic classifier.

Shannon's measure of mutual information was used to assess the correspondence between the structure of the obtained clusters and semantic classes. Let N is a number of documents in the database, and $m(i, j)$ is a number of texts from the semantic class i, assigned to the class j. Calculate the values $q(i, j) = \dfrac{m(i, j)}{N}$, $q_1(i) = \sum_j q(i, j)$, $q_2 = \sum_i q(i, j)$. Then the value of mutual information (IM) will be found by formulas:

$$IM = \sum_i \sum_j q(i, j) \log q(i, j) - \sum_i q_1 \log q_1(i) - \sum_i q_2 \log q_2(i).$$

Note that semantic clustering of text documents involves dividing text data into sets of texts (clusters), such that texts within the same cluster are as similar as possible in meaning, while texts belonging to different clusters have different meanings [18, 19]. With semantic clustering of texts, it is possible to mark the identified clusters with their thematic descriptors. It is the semantic clustering by keywords that was used in the database.

Results of clustering of 7692 records with comments are:

• 311 (4.04 %) cases were recorded when students resorted to the help of other people during testing;

• 1524 cases (19.81 %) when students used the Internet during testing to find answers to a question;

• 1034 cases (13 %) using gadgets;

• 1390 (18.07 %) cases of remark on non-verbal communication;

• 566 (7.36 %) cases of voice detection;

• 2483 (32.28 %) cases when the examinee's face was poorly visible or not discovered.

All of these actions are usually violations of the rules during a traditional exam. Only 29 % of students did not resort to any of the previously listed actions or did not come into the view of the proctor. The main task of the proctor is to follow the student on the screen so that he does not cheat. The proctor should pay attention to this in time and, if it is repeated, make a comment. If the student receives a certain number of comments, he is removed from the exam. In the studied subject area, there were a maximum of 5 remarks.

TIME ANALYSIS OF VIOLATIONS

Note that each test has an optimal testing time, the decrease or excess of which reduces the quality indicators of the test because the results of such a test will not objectively reflect the level of preparedness of students [21]. In this investigation all examination tests are designed for an optimal 60 minutes. The number of test task for one test taker is 20 or 40 questions depending on the complexity ones. To maximize the effectiveness of testing knowledge in the process of studying the discipline, 85 % of students passed examination tests on 20 questions, since the effect of fatigue can affect the results of the tested. Statistics of violations during the exam (20 questions) are demonstrated on Table 1.

T a b l e 1. Statistics of violations during the exam (20 questions and 40 questions)

Period, min	Quantity of violations	
	20 questions	40 questions
5	2141	404
10	1331	251
15	848	162
20	658	108
25	466	82
30	299	70
35	225	56
40	145	48
45	115	27
50	77	26
55	60	23
60	43	27

Only 15 % of students were tested on 40 questions and statistics of violations during the exam (40 questions) are demonstrated on Table 1 also.

ANALYSIS OF THE RECEIVED DATA

An analysis of practical experience in conducting various online tests showed that students, as a rule, make more violations in the first half of the exam.

As follows from the data given on 40 questions (see Table 1), out of all 1284 comments, 1077 comments (84 %) fall on the first 30 minutes and 207 comments (16 %) for the next 30 minutes. Moreover, in the first 10 minutes, there are 629 comments (49 %). Thus, it is important for proctors to follow the examinees during the first 10 minutes of testing and to eliminate violations with their comments. The data shows that the second part of testing is more relaxed and the number of comments decreases. As for testing on 20 questions (see Table 1), out of 6408 comments, 5743 (90 %) comments fall on the first 30 minutes, and 665 (10 %) comments fall on the second part of the exam. Here, in the first 10 minutes, 2936 comments (46 %) were made. The number of comments decreases by the end of the exam.

Thus, the results of processing proctor's comments show that students receive a significant number of comments in the first 10 minutes of the exam. Types of violations and their quantitative characteristics for the first 10 minutes of testing are shown in Table 2 for a test with 20 questions.

Note that types of violations and their characteristics for first 10 minutes of testing with 40 questions do not significantly differ from the data shown in Table 2.

Based on results of proposed investigation it makes conclude that many learners face significant challenges in taking online exams. Many test takers faced the problem due to technical difficulties and lack of support. For given problems solving need suggest instructors to give trainees the opportunity to rehearse procedures for accessing exam services before the exam and to motivate them and the trainees should be provided with prompt technical support in in real time on any special questions that may arise during the online exam.

Let us note that most countries, teachers has limited digital skills and experience in online teaching and learning [21] and information technology, proposee in this paper, to allow more effective conducting various online tests and results checking. During distance learning, educators could improve their decision-making skills by focusing on the hardships and challenges that students face, especially during Covid-19. Now creation of the information technology, ICT infrastructure, reliable communications remain major to successfully passing online exams in a secure and controlled environment.

T a b l e 2. Statistics of violations for the first 10 minutes (20 questions)

Type of violation	Quantity of violations	Percentage (%)
Application switching	950	27.4
Use of gadgets	517	14.9
Non-verbal communication	449	12.9
Voice detection	211	6.1
Face is not detected	1118	32.2
Several faces in the frame	90	2.6
Other violations, remarks	137	3.9

CONCLUSIONS

In this study, information technology for the analysis of violations in the process of distance learning was carried out on the example of an examination session and the time intervals with the greatest number of violations were identified among the students EKTU.

The transition to distance learning due to the COVID-19 pandemic is accompanied by various types of problems, including technical and organizational ones. Statistics show that many learners still have technical problems that are a stumbling block in online learning. An exam with a proctor is the best alternative to the face-to-face exam, which preserves the transparency and quality of the procedure itself. Such studies can help to develop recommendations for proctors to identify violations and determine the time period for observing, recording and assessing student behavior during the exam.

REFERENCES

1. Pokhrel S., Chhetri R. A literature review on impact of COVID-19 pandemic on teaching and learning. *Higher Education for the Future*. 2021. Vol. 8, Iss. 1. P. 133–141. https://doi.org/ 10.1177/2347631120983481.

2. UNESCO Global Education Monitoring Report. Policy Brief: Education during COVID-19 and beyond. 2020. URL: https://www.un.org/development/desa/dspd/wp-content/uploads/ sites/22/2020/08/sg_policy_brief_covid-19_and_education_august_2020.pdf.

3. Hämäläinen R., Nissinen K., Mannonen J., Lämsä J., Leino K., Taajamo M. Understanding teaching professionals' digital competence: What do PIAAC and TALIS reveal about technology-related skills, attitudes, and knowledge? *Computers in Human Behavior*. 2021. Vol. 117. 106672. https://doi.org/10.1016/j.chb.2020.106672.

4. Raghu R., Sairam B., Veena G., Hardik V., Prema N., Adoption of online proctored examinations by university students during COVID-19: Innovation diffusion study. *Public Health and Emergency*. 2021. Vol. 26. P. 7339–7358. https://doi.org/10.1007/ s10639-021-10581-5

5. Cavanaugh J.K., Jacquemin S.J. A large sample comparison of grade based student learning outcomes in online vs face-to-face courses. *Online Learning*. 2015. Vol. 19, N 2. P. 25–32. https://doi.org/10.24059/OLJ.V19I2.454

6. Dendir S. Performance differences between face-to-face and online students in economics. *Journal of Education for Business*. 2019. Vol. 94, Iss. 3. P. 175–184. https://doi.org/ 10.1080/08832323.2018.1503586.

7. Paul J., Jefferson F. A comparative analysis of student performance in an online vs. face-to-face environmental science course from 2009 to 2016. *Frontiers in Computer Science*. 2019. Vol. 1. 7. P. 1–7. https://doi.org/10.3389/fcomp.2019.00007.

8. Rawashdeh A.Z.A., Mohammed E.Y., Al-Arab A.R., Alara M., Al-Rawashdeh B., Al-Rawashdeh B. Advantages and disadvantages of using E-learning in university education: Analyzing students' perspectives. *Academic Conferences and Publishing International Limited*. 2021. Vol. 19, N 3. P. 107–117. https://doi.org/10.34190/ejel.19.3.2168.

9. Young J.R. Online classes see cheating go high-tech. *The Chronicle of Higher Education*. 2012. Vol. 58, N 6. P. 24–26. URL: www.chronicle.com/article/online-classes-see-cheating -go-high-tech/.

10. Indi C.S., Pritham K. V., Acharya V., Prakasha K. Detection of malpractice in E-exams by head pose and gaze estimation. *International Journal of Emerging Technologies in Learning*. 2021. Vol. 16, N 8. P. 47–60. https://doi.org/10.3991/ijet.v16i08.15995.

11. Prokes C., Lowenthal P.R., Snelson C., Rice K. Faculty views of CBE, self-efficacy, and institutional support: An exploratory study. *The Journal of Competency-Based Education*. 2021. Vol. 6, Iss. 4. P. 233–244. https://doi.org/10.1002/cbe2.1263.

12. Alessio H.M., Malay N.J., Maurer K., Bailer A.J., Rubin B. Examining the effect of proctoring on online test scores. *Online Learning.* 2017. Vol. 21, Iss. 1. P. 146–161. https://doi.org/10.24059/olj.v21i1.885.

13. Norris M. University online cheating — how to mitigate the damage. *Research in Higher Education Journal.* 2019. Vol. 37. P. 1–20. URL: https://files.eric.ed.gov/fulltext/EJ1233121.pdf.

14. Golden J., Kohlbeck M. Addressing cheating when using test bank questions in online classes. *Journal of Accounting Education*. 2020. Vol. 52. 100671. https://doi.org/10.1016/j.jaccedu.2020.100671.

15. Moini A., Madni A.M. Leveraging biometrics for user authentication in online learning: a systems perspective. *IEEE Systems Journal*. 2009. Vol. 3, Iss. 4. P. 469–476. https://doi.org/10.1109/ JSYST.2009.2038957.

16. Anju A., Preeti G. Clustering in Big Data: A review. *International Journal of Computer Applications.* 2016. Vol. 153, N 3. P. 44–47. https://doi.org/10.5120/ijca2016911994.

17. Krak Yu.V., Barmak O.V., Mazurets O.V. The practice implementation of the information technology for automated definition of semantic terms sets in the content of educational materials. *Problems in Programming.* 2018. Vol. 2–3. P. 245–254. https://doi.org/10.15407/pp2018.02.245.

18. Han E-H., Karypis G. Centroid-based document classification: analysis and experimental results. *Proc. 4th European Conference on Principles of Data Mining and Knowledge Discovery* (13–16 September 2000, Lyon, France). Lyon, 2000. *Lecture Notes in Computer Science.* 2000. Vol. 1910. P. 424–431. https://doi.org/10.1007/3-540-45372-5_46.

19. Krak Iu.V., Kudin G.I., Kulias A.I. Multidimensional scaling by means of pseudoinverse operations. *Cybernetics and Systems Analysis*. 2019. Vol. 55, N 1. P. 22–29. https://doi.org/10.1007/s10559-019-00108-9.

20. Krak Iu., Barmak O., Manziuk E. Using visual analytics to develop human and machine-centric models: A review of approaches and proposed information technology. *Computitional Intelligence*. 2020. Vol. 36, Iss. 3. P. 1–26. https://doi.org/10.1111/coin.12289.

21. Kennan S., Bigatel P., Stockdale S., Hoewe J. The (lack of) influence of age and class standing on preferred teaching behaviors for online students. *Online Learn*. 2018. Vol. 22, Iss. 1. P. 163–181. https://doi.org/10.24059/olj.v22i1.1086.

Д. Муратули, Н.Ф. Денісова, Ю.В. Крак
ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ДЛЯ ВИЯВЛЕННЯ ПРОКТОРОМ ПОРУШЕНЬ ПІД ЧАС ІСПИТУ

Анотація. Розглянуто актуальну проблему списування студентами вищих навчальних закладів під час іспиту. Досліджено предметну область та представлено результати теоретичного та емпіричного дослідження списування як однієї з форм академічного шахрайства у вищих навчальних закладах. Наведено статистичні дані порушень під час іспиту та виявлено основні закономірності порушень у період дистанційної форми навчання. Отримані результати дослідження допоможуть визначити, у який період підсумкової атестації студенти частіше порушують правила академічної доброчесності.

Ключові слова: дистанційне навчання, терміни порушення, екзаменаційна сесія, кластеризація, прийняття рішень.