

ORIGINAL RESEARCH ARTICLE

Using information and communication technologies for the assessment of a large number of students

Kasym Baryktabsov, Chinara Jumabaeva* and Ulan Brimkulov

Computer Engineering Department, Kyrgyz-Turkish Manas University, Bishkek, Kyrgyz Republic

(Received: 3 November 2022; Revised: 16 May 2023; Accepted: 25 May 2023;

Published: 20 July 2023)

Many examinations with thousands of participating students are organized world-wide every year. Usually, this large number of students sit the exams simultaneously and answer almost the same set of questions. This method of learning assessment requires tremendous effort and resources to prepare the venues, print question books and organize the whole process. Additional restrictions and obstacles may appear in conditions similar to those during the COVID-19 pandemic. One way to obviate the necessity of having all the students take an exam during the same period of time is to use a computer-assisted assessment with random item selection, so that every student receives an individual set of questions. The objective of this study is to investigate students' perceptions of using random item selection from item banks in order to apply this method in large-scale assessments. An analysis of the responses of more than 1000 surveyed students revealed that most of them agree or completely agree with using the proposed method of assessment. The students from natural science departments showed more tolerance of this method of assessment compared with students from other groups. Based on the findings of this study, the authors concluded that higher-education institutions could benefit from implementing the abovementioned assessment method.

Keywords: computer-assisted assessment; computer-based assessment; e-assessment; learning assessment; education; information and communication technologies

Introduction

'Assessment is an essential component of learning and teaching' (Ferrari et al., 2009). Assessment is critical to student learning and certification (Bennett et al., 2017). It is always present in higher education and influences all stakeholders, including educational institutions, teachers, and students (Stöddberg, 2012). In higher education, it shapes the experiences of students and influences their behavior more than the teaching they receive (Bloxham & Boyd, 2007). Assessment can be summative or formative. Formative assessment usually takes place during the learning process and aims to provide feedback in order to support students' learning. Summative assessment aims to summarize students' accomplishments and usually takes place at the end of the

*Corresponding author: Email: chinara.jumabaeva@manas.edu.kg

course or education process. The most important part of summative assessment is grading and making judgments (Stödberg, 2012).

In the modern world, the implementation of assessment practices through information and communication technologies (ICT) is very important (Boud & Soler, 2016). The use of ICT for learning assessment appears in the literature under such names as computer-assisted assessment (CAA) (Bull & McKenna, 2004), computer-based assessment (CBA) (Thelwall, 2000), technology-based assessment (TBA) (Csapó & Molnár, 2019), and finally e-assessment (Adesemowo et al., 2016).¹ E-assessment works very well for formative assessment, summative assessment (Stödberg, 2012), and self-assessment. E-assessment has many advantages compared with paper-based testing (Alruwais et al., 2018). It allows more complex item types, such as the use of audiovisual materials, and more complex interactions between the learner and the computer (Conole & Warburton, 2005). CAA systems provide richer data about students' performance and rapid feedback (Brimkulov et al., 2017). Different types of CAA tools have already been developed (Contreras-Higuera et al., 2016; Christie et al., 2015). Many CAA systems and tools allow questions to be generated randomly (Kruger et al., 2015). Piaw argued that replacing paper-based testing with computer-based testing does not cause significant differences in test scores (Piaw, 2012).

E-assessment provides an opportunity to decrease lecturers' administrative burden, giving them more time for research and professional self-development. It acquires special significance when it comes to implementing large-scale assessments (Adesemowo et al., 2017). Today, it is clear that systematic large-scale assessments cannot be conducted with traditional instruments (Csapó & Molnár, 2019).

Many large-scale assessments take place every year. In some cases, thousands of students take an exam simultaneously. For example, this occurs in university admission examinations. In 2011, 9.33 million participants took the National Higher Education Entrance Examination (also called Gaokao) in the People's Republic of China (Haifeng, 2012). In China, all students throughout the country sit the above-mentioned exam at the same time (Davey et al., 2007). In South Korea, some 600,000 college applicants took the College Scholastic Ability Test (Suneung in Korean) in 2016 (Kim et al., 2017). In Vietnam, the Ministry of Education and Training administers a national University Entrance Examination, which is undertaken annually by over one million final-year secondary school students (Hayden & Thiep, 2010). In the Russian Federation, 800,000 candidates took the Unified State Exam (a country-wide standardized examination that combines in a single procedure the examination at the end of secondary school with entrance exams for tertiary education) in 2012 (Piattoeva, 2015). The number of participants in the University Entrance Examination in Turkey has exceeded one million in recent years (Içbay, 2005). A number of other countries also organize these kinds of examinations.

The authors of this study also faced the necessity of conducting assessments for large classes. They were teaching 'Introduction to Information Technologies', which is a mandatory course for all first-year students at Kyrgyz-Turkish Manas University (KTMU). The number of students enrolled in the course each semester was about 300–500. During the semester, students had at least two exams: a midterm (formative assessment) and a final (summative assessment). The assessment was conducted in the form of a paper-based test using multiple-choice questions. Organizing this

¹Hereinafter the terms 'e-assessment', 'technology-based assessment', 'computer-assisted assessment', and 'computer-based assessment' will be used interchangeably.

assessment was quite challenging, because it was necessary to bring in all the enrolled students at the same time, prepare venues, print questions on paper, ask additional staff for help (in order to prevent students from cheating), etc. The students were from different departments with different schedules, which made it impossible to find a convenient time on a weekday. It was also challenging to find enough free venues for examinations. This is why the exams were conducted on weekends (usually on Saturdays). In order to avoid all of these inconveniences, we decided to use the benefits of e-assessment. We developed a CAA system, which is described in Brimkulov et al. (2017). Using this system, the students sit exams in small groups at the scheduled lesson time in the computer-equipped classroom where the normal lessons are usually taught. There was a risk of the dissemination of the exam questions by students who had taken an exam earlier to other students. Thus, we decided to use random question selection from an item bank, so that every student received a unique set of questions. The system attracted the attention of other colleagues who also taught courses with large numbers of enrolled students, as they faced the same issues.

After several years of the successful usage of this e-assessment system, we started considering the possibility of using it for the university admission examination. Every year, KTMU organizes a university admission exam in the form of a paper-based test using multiple-choice questions. However, the scale of that examination is much larger. The number of enrollees participating in the exam is about 6000–8000. When we suggested that the university admission examination organizing committee apply the same method that we used in the ‘Introduction to Information Technologies’ course, the committee members expressed concerns about the fairness of the examination. The committee members were afraid that some students might complain that they could get higher scores if they were given another set of questions. It was unknown to what extent such complaints would be received. This fact motivated us to do an empirical study of this issue. As we planned to involve a large number of students (about 1000), performing a survey through a questionnaire was found to be the most appropriate method of collecting data.

The objective of this study is to investigate students’ perceptions of using random item selection from item banks in order to apply this method in large-scale assessments.

The following research questions are studied:

1. To what extent do students agree with the use of random item selection from item banks in formative assessments?
2. To what extent do students agree with the use of random item selection from item banks in summative assessments?
3. To what extent do students agree with the use of random item selection from item banks in university admission examinations?
4. To what extent do students agree with the use of random item selection from item banks in self-assessments?
5. Is there any difference in the perceptions of student groups from different domains of learning?

There are a number of studies on students’ attitudes towards using ICT for learning in general and CAA in particular. Silin and Kwok (2017) examined the factors that support or hinder students’ attitudes towards using ICT in problem-based learning among polytechnic students. Karl et al. (2011) stated that students found CAA to

be equivalent to written multiple-choice tests. Dammas (2016) surveyed chemistry department students and revealed that the majority of respondents had a positive attitude towards computer-based testing. There are also studies that suggest effective algorithms for selecting questions randomly from a database (Liu & Feng, 2009).

Binnahedh (2022) examined the attitudes of students and teachers toward the washback effect of electronic tests. The results of the study show that students' perceptions of electronic testing are more positive than teachers' perceptions of such tests (Binnahedh, 2022).

St-Onge et al. (2021) examined how educators have coped with the challenge of adapting their assessment methods during the pandemic to determine what is required to support and facilitate the future development of quality assessments and the introduction of e-assessment in higher education. The COVID-19 pandemic has provided an unprecedented opportunity to critically evaluate and change assessment methods.

Some studies also suggest paying attention to the 'sustainable assessment' concept. Boud and Soler (2016) consider this term. They believe that sustainable assessment will allow students to regulate and evaluate their own learning and continue their studies outside the period of a course. The development of knowledge assessment is of particular interest (Boud & Soler, 2016).

Stöðberg (2012) studied the major areas of interest in e-assessment. Based on a literature review, he revealed that most of the research was devoted to the study of the learning environment. He also found that the number of research papers on e-assessment was increasing at a high speed (Stöðberg, 2012).

Clarke-Midura and Dede (2010) pointed to the inadequacy of automated versions of item-based paper-and-pencil tests in 21st-century education. In order to use the full power of ICT to innovate via providing richer observations of student learning, the authors explored a virtual assessment of learning achievements (Clarke-Midura & Dede, 2010).

Other authors consider technologically rich environments (TREs) and the pedagogical opportunities they offer to learners and teachers (Shute et al., 2016).

Adesemowo et al. (2016) emphasized the importance of considering security concerns when implementing an e-assessment platform, especially for the assessment of large classes. They also studied the opinions of students concerning the introduction of e-assessment and its impact on the workload of teachers. Based on the results of the study, the use of e-assessment is viable, scalable, and safe, reducing the administrative burden and increasing student productivity.

Adesemowo et al. (2017) revealed that learning management system platforms could provide a TRE for designing innovative text-based assessments for relatively large classes.

A number of studies have examined the features, advantages, and disadvantages of e-assessment.

Jordan (2013) proposed that the online environment makes computer assessments of knowledge more accessible. Jordan (2013) stated the following: 'Questions for computer-marked assessment need to be delivered at low cost and quickly and there is a danger that this will lead to poor quality assessment'. To avoid low-quality assessments of students' knowledge, it is necessary to provide the system with high-quality questions of various types; however, the questions should assess the same learning outcome and they should have the same level of difficulty. Jordan (2013) found that in assessing large numbers of students, the possibility of different students receiving

different sets of questions is very important. In this case, it is necessary to create question banks (Jordan, 2013).

The authors of this study could find only one research paper that discussed students' perceptions regarding e-assessments in which random questions are selected from item banks (Dermo, 2009). Thus, the authors of this study had the impression that there is a lack of sufficient research on students' attitudes towards random item selection in CAA. In order to fill this gap, this study surveyed more than 1000 students, asking them to express their perceptions of the abovementioned assessment method.

The rest of the paper is organized as follows. The next section describes the survey and analysis methodology, and it is followed by a section that describes the collected data and analysis results. Then, there is a section with discussions of this study's findings. The final section concludes the paper.

Methodology

In order to conduct this study, questionnaires in three languages (Kyrgyz, Russian, and Turkish) were developed from scratch by the authors using the Google Forms tool. These questionnaires consisted of two parts. The first part contained general questions about the university, department, and year of the student. The questions in the second part are given in Table 1.

The survey was conducted among university students in the city of Bishkek, Kyrgyz Republic. There is an 'Introduction to Information and Communication Technologies' course that is mandatory for all first-year students of KTMU. During one of the lessons of that course, after the midterm examination in the form of computer-based testing with random item selection, the survey was introduced to the students by one of the authors of the study. Then, the students were asked to fill out the questionnaire right there in the computer classroom. Some third- and fourth-year KTMU students were surveyed in exactly the same manner but during other courses. The students from other universities were provided with a link and asked to fill out the form on their own. The survey was conducted from 2018 to 2019.

The responses to statements #3, #4, and #5, placed in the second part of the questionnaire, are of particular interest, as they are directly related to the aim of the study. The collected data were represented using a Likert scale (Likert, 1932). There has been considerable discussion regarding the statistical methods that are appropriate for the analysis of Likert-scale data. Some authors argue that parametric methods cannot be used to analyze ordinal data like Likert-scale data (Tabachnick & Fidell, 2007). Others argue that 'parametric statistics are robust with respect to violations of these assumptions' (Norman, 2010). There are many papers with recommendations for the analysis of Likert-scale data (Harpe, 2015; Joshi et al., 2015). There was also a computer simulation study in which five-point Likert-scale data were analyzed using the *t*-test and Mann-Whitney-Wilcoxon test. The results of the study 'showed that the two tests had equivalent power for most of the pairs' (de Winter & Dodou, 2010). However, the authors of this study decided to use both parametric (*t*-test, Analysis of variance [ANOVA]) and nonparametric (Mann-Whitney-Wilcoxon test, Kruskal-Wallis H-test) methods and then compare the results in order to ensure the correctness and robustness of the analysis results. For the analysis, the SPSS software was used.

Table 1. The questions in the second part of questionnaire.

Question	Answer options
1 Have you ever taken an examination in the form of computer testing?	• Yes • No
2 What is your level of confidence in the computer testing results?	1 2 3 4 5 I don't trust them at all. I completely trust them.
Example: There are 500 questions in the question bank. The difficulty level is the same across all questions. The computer will randomly select 50 questions from the question bank for the exam. Thus, every student will have a unique set of questions.	
Please provide your level of agreement with the following statements.	
3 It would be good to use the assessment method described in the example above in MIDTERM examinations.	1 2 3 4 5 I completely disagree. I completely agree.
4 It would be good to use the assessment method described in the example above in FINAL examinations.	1 2 3 4 5 I completely disagree. I completely agree.
5 It would be good to use the assessment method described in the example above in UNIVERSITY ADMISSION examinations.	1 2 3 4 5 I completely disagree. I completely agree.
6 It would be good to use the assessment method described in the example above for SELF-ASSESSMENT purposes.	1 2 3 4 5 I completely disagree. I completely agree.
7 Using the assessment method described in the example above will lead to unfair results.	1 2 3 4 5 I completely disagree. I completely agree.
8 It is not necessary to ask the same questions in order to compare students' competence.	1 2 3 4 5 I completely disagree. I completely agree.

Collected data and analysis results

The total number of students who participated in the study is 1034. Most of them (91.7%) are students at KTMU. The rest of the students represent five other higher-education institutions (Kyrgyz-Russian Slavonic University, Kyrgyz State Medical Academy, Kyrgyz National University, International University of Central Asia, and Kyrgyz State Technical University) in different proportions.

As mentioned above, most of the respondents were surveyed during the lesson of one of the mandatory courses for first-year students at KTMU. This is the reason that most of the students are first-year students (87.6%).

The surveyed students represent 42 different departments that have been divided into three groups: natural sciences, economics, and humanities (the number of students in each group can be found in Table 2). Departments such as Computer Engineering, Software Engineering, Chemical Engineering, Environmental Engineering, Food Engineering, Information Technologies, Building Construction, etc., have been included in the natural sciences group. Departments such as Economics, Finance and Credit, Accounting, Management, etc., have been included in the economics group. The rest of the departments have been included in the humanities group (Philosophy, History, Turcology, Linguistics and Translation, etc.).

Table 2. Descriptive statistics for the groups of students.

	N	Mean	SD	SD	95% Confidence interval for mean		Min.	Max.
					Lower bound	Upper bound		
					Natural sc.	243		
Economics	132	10.4394	3.22741	0.28091	9.8837	10.9951	3	15
Humanities	659	10.4841	3.19207	0.12435	10.2399	10.7282	3	15
Total	1034	10.6721	3.20271	0.09960	10.4767	10.8676	3	15

When the survey was administered, most of the students already had experience with computer testing (88.1%).

Seventy percent of the respondents trust or completely trust computer testing results. The percentage of students who do not trust computer testing results was 7%, and 23% were not sure.

The total percentage of students who agree or completely agree with using a random item selection method in midterm examinations was 58.22%. On the other hand, the total percentage of students who disagree or completely disagree with using the proposed assessment method in midterm examinations was 16.05% (see Table 3).

The total percentage of students who agree or completely agree with using a random item selection method in final examinations was 57.7%. On the other hand, the total percentage of students who disagree or completely disagree with using the proposed assessment method in final examinations was 23.0% (see Table 3).

The total percentage of students who agree or completely agree with using a random item selection method in university admission examinations was 56.38%. On the other hand, the total percentage of students who disagree or completely disagree with using the proposed assessment method in university admission examinations was 23.12% (see Table 3).

The total percentage of students who agree or completely agree with using a random item selection method for self-assessment purposes was 77.2%. On the other hand, the total percentage of students who disagree or completely disagree with using the proposed assessment method for self-assessment purposes was 10.7% (see Table 3).

Most of the surveyed students said that using the method of random item selection will not lead to unfair results (see Table 3).

After the survey, it was revealed that the wording of the statement ‘It is not necessary to ask the same questions in order to compare students’ competence’ was unclear to students. This is the reason that, in the opinion of the authors of this study, the responses to that statement were uncertain (see Table 3).

The responses to statements #3, #4, and #5 (in the second part of the questionnaire) are of particular interest. Together, they form the core of the study. The sum of the number representations of a student’s responses to these statements provides information about the student’s general attitude (negative, neutral, or positive). If a student’s response consists of only neutral values (the number representation of the response to each of the given statements is equal to 3), then the sum value will be equal to nine. Thus, a sum value that is equal to 10 indicates a positive attitude, because in this case, a student must have given at least one positive response (3+3+4 or 3+2+5). The sum values of the number representations of the responses to statements #3, #4, and #5 calculated for every respondent represent the aggregated data, which is also

Table 3. The distribution of the responses to the statements from the second part of the questionnaire.

Statement	Responses					Total
	1. I completely disagree	2	3	4	5. I completely agree	
3. 'It would be good to use the assessment method described in the example above in MIDTERM examinations'.	81 7.83%	85 8.22%	266 25.73%	301 29.11%	301 29.11%	1034 100%
4. 'It would be good to use the assessment method described in the example above in FINAL examinations'.	125 12.10%	113 10.90%	199 19.30%	322 31.10%	275 26.60%	1034 100%
5. 'It would be good to use the assessment method described in the example above in UNIVERSITY ADMISSION examinations'.	135 13.06%	104 10.06%	212 20.50%	228 22.05%	355 34.33%	1034 100%
6. 'It would be good to use the assessment method described in the example above for SELF-ASSESSMENT purposes'.	56 5.40%	55 5.30%	125 12.10%	227 22.00%	571 55.20%	1034 100%
7. 'Using the assessment method described in the example above will lead to unfair results'.	311 30.10%	218 21.10%	201 19.40%	169 16.30%	135 13.10%	1034 100%
8. 'It is not necessary to ask the same questions in order to compare students' competence'.	180 17.40%	142 13.70%	252 24.40%	218 21.10%	242 23.40%	1034 100%

referred to as 'Likert-scale data'. The descriptive statistics of the obtained aggregated data are presented in Table 4; the frequencies are presented in Figure 1.

These aggregated data represent the general attitude of the respondents toward the assessment method with random item selection. The values equal to or greater than 10 indicate a positive attitude. The value for Cronbach's alpha for these three statements was $\alpha = 0.754$. According to Field (2013), Cronbach's alpha scores above 0.7 are considered 'acceptable' in most social sciences.

It was interesting to determine whether the attitudes of the students differ among the groups of departments (natural sciences, economics, and humanities) or not. Since the data are not normal (see Table 5) and ordinal, the authors of this study decided to run both a parametric one-way ANOVA and a non-parametric Kruskal-Wallis H-test, and then compare the results. Descriptive statistics for the groups of students are given in Table 6.

Table 4. Descriptive statistics of students' attitudes according to the aggregated data.

Statistic names	Statistic	SE
N	1034	
Mean	10.6721	0.0996
95% Confidence interval for mean		
Lower bound	10.4767	
Upper bound	10.8676	
5% trimmed mean	10.8446	
Median	11.0000	
SD	3.20271	
Variance	10.257	
Minimum	3	
Maximum	15	
Range	12	
Interquartile range	4.00	
Skewness	-0.547	0.076
Kurtosis	-0.357	0.152

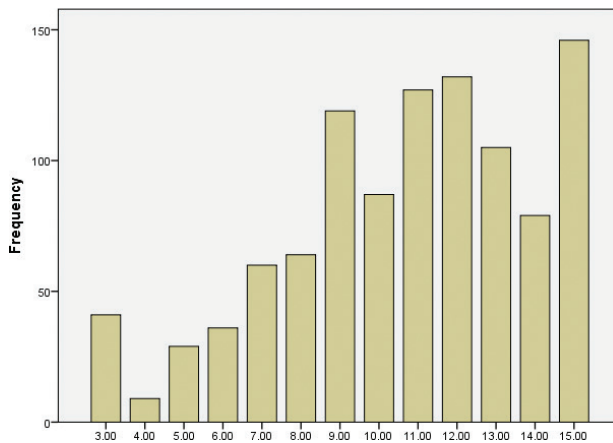


Figure 1. Graphical representation of the frequencies of the aggregated data.

Table 5. Results of the normality tests.

Statistic name	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Students' attitude	0.110	1034	0.000	0.943	1034	0.000

^aLilliefors significance correction.

There was a statistically significant difference between the groups according to the one-way ANOVA ($F(2,1031) = 6.349, p = 0.002$). Levene's test showed that the variances for students' attitudes were equal ($F(2,1031) = 0.383, p = 0.682$). The results of the post hoc test are given in Table 6.

Table 6. Tukey HSD post hoc test results.

(I) Group	(J) Group	Mean difference (I-J)	SD	Sig.	95% Confidence Interval	
					Lower bound	Upper bound
Natural sc.	Economics	0.86925	0.34451	0.032	0.0606	1.6779
	Humanities	0.82458	0.23913	0.002	0.2633	1.3858
Economics	Natural sc.	-0.86925	0.34451	0.032	-1.6779	-0.0606
	Humanities	-0.04467	0.30384	0.988	-0.7578	0.6685
Humanities	Natural sc.	-0.82458	0.23913	0.002	-1.3858	-0.2633
	Economics	0.04467	0.30384	0.988	-0.6685	0.7578

Based on the post hoc test results, it can be stated that there is a statistically significant difference between the attitudes of the students from the natural science group and the students from the other groups.

A Kruskal-Wallis H-test showed that there was a statistically significant difference in students' attitudes between the different groups ($H(2) = 15.127, p = 0.001$), with a mean rank attitude of 582.24 for the natural science group, 493.36 for the economics group, and 498.46 for the humanities group. Then, post hoc tests were conducted to test pairwise comparisons. A statistically significant difference was found between the natural science group and the economics group ($p = 0.006$). There was also a statistically significant difference between the natural science group and the humanities group ($p = 0.000$). The difference between the economics group and humanities group was not statistically significant ($p = 0.857$).

It was also interesting to determine whether there is a difference in attitude between KTMU students ($N = 948$) and students from other universities ($N = 86$). The t -test showed that there was not a statistically significant difference in the attitudes of KTMU students ($M = 10.6445$, standard deviation [SD] = 3.20) and students from other universities ($M = 10.9767$, SD = 3.23) ($t(1032) = -0.921, p = 0.357$). A Mann-Whitney U-test also indicated that there was not a statistically significant difference between the attitudes of students from other universities (mean rank = 547.61) and KTMU students (mean rank = 514.77) ($Z = -0.982, p = 0.326$).

As we can see, the results of the parametric and nonparametric tests lead to the same conclusions.

Discussion

The results of the current study show that the majority of the respondents agree with using random item selection in CAA. However, the more important the examination, the smaller the number of students that support this method. The results mean that the proposed method could be used in different types of examinations. This method has been in use since 2014 at KTMU on the midterm and final examinations of the 'Introduction to Information and Communication Technologies' mandatory course. Thousands of students were tested using this method up to the year 2020. No complaints from students have been received. However, many requests have been received from lecturers to use this method in the other mandatory courses.

Taking into consideration the fact that the students from the natural science group expressed a more positive attitude towards using the method of random item selection

in CAA, this method could be applied even more broadly in the departments and universities that focus on this field of study. This finding correlates with the results obtained by Abdullah et al. (2015). They revealed a ‘difference between Arts and Science students in terms of their attitude towards IT in favor of Science students’ (Abdullah et al., 2015).

The authors of this study also agree with Jordan (2013), who argued that the selection of questions from a question bank or the use of multiple variants of each question can discourage plagiarism. It is worthwhile to invest in the use of e-assessment for the assessment of a large number of students (Jordan, 2013).

It should be noted that Dermo (2009) analyzed students’ perceptions of e-assessments and found that the use of random question selection from item banks was seen by the students as unfair. This result is the complete opposite of the results of our study. This might be due to several factors, one of which is the time that passed between these two studies. The students’ attitudes might have changed. The more e-assessment is used in education, the more students become accustomed to this kind of approach. However, the authors of this study agree with Dermo (2009), who stated that it is necessary to ‘take steps to ensure the quality of these item banks, for example, using item analysis to check the difficulty level of the items in the bank’. The learning outcomes assessed by the items should also be the same (Jordan, 2013). In other words, the focus of the lecturers should change from organizing the examinations, reading students’ responses, and grading to preparing quality item banks, validating and ensuring that items have equal difficulty levels, and monitoring the assessed learning outcomes.

In the opinion of the authors of this study, higher-education institutions should organize an assessment and certification center equipped with the necessary hardware and software. The center would be responsible for performing learning assessments using modern ICT and could be available 24/7 for the convenience of the students. The teachers, lecturers, and other academic staff could be exempt from the duty of supervising exams. They would be responsible for preparing question banks containing high-quality questions.

This kind of assessment center would make it possible to obviate the necessity of having all the students take an exam at the same time and place (which is difficult to organize when there is a large number of participating students or in conditions similar to those during the COVID-19 pandemic). It would reduce paper use, reduce the time that academic staff must spend on evaluation and grading, provide the students with an opportunity to choose convenient times for and a convenient order of exams, and finally allow the use of more complex question types in order to assess competencies such as problem-solving, reflection, creativity, critical thinking, etc.

The same assessment and certification center could be used for university admission examinations. In countries where the national authorities are responsible for holding university admission examinations, the computer-equipped classrooms of schools and universities with available Internet access could be used for CAA.

Conclusion

ICT plays an important role in education today. Many different kinds of systems and tools have been developed to support the educational process, including software for learning assessment. However, there is still much room for improvement in this area.

Many examinations organized worldwide, with a large number of participating students, are still paper-based. Switching to CBAs can potentially provide many advantages. The objective of this study was to explore the attitudes of students towards the method of random item selection that is often used in CBA and that makes it possible to obviate the necessity of having all the students take an exam at the same time and place.

The key findings of this study are the following:

- Most of the surveyed students agree with the use of the method of random item selection in CBA (58.22% for midterm examinations, 57.7% for final examinations, and 56.38% for university admission examinations; the average value is 57.43%).
- The percentage of students who do not agree with the use of this method of assessment was 16.05% for midterm examinations, 23% for final examinations, and 23.12% for university admission examinations (with an average of 20.7%).
- The students from natural science departments showed more tolerance of this method of assessment compared with students from economics and humanities fields.
- There was no difference in the attitudes of students from different universities towards using the abovementioned method of learning assessment.

These findings mean that there should not be many complaints from students regarding the fairness of the examinations when the proposed method is used.

Based on their own experience of using the proposed method of assessment and the analysis of the survey data, the authors of this study assume that the method of random item selection in CAA could be used more broadly in different types of examinations with a large number of participating students at higher-education institutions.

The findings of this study may be useful for academic staff, the decision-makers of higher-education institutions, and policy-makers.

As one of the directions for future research, a study could be conducted in order to understand the attitudes of high school students toward the proposed assessment method and whether there is any difference between high school students and university students. Another study could explore whether there is any difference between the attitudes of students from different countries. The authors of this study would also be interested in doing research to understand why some students do not agree with using the proposed method by interviewing them in the future.

Declarations

Availability of data and materials

The datasets used and/or analyzed in the current study are available from the corresponding author on reasonable request.

Funding

No funds, grants, or other support was received.

Acknowledgements

Not applicable.

References

- Abdullah, Z. D. et al. (2015). Students' attitudes towards information technology and the relationship with their academic achievement. *Contemporary Educational Technology*, 6(4), 338–354. doi: 10.30935/cedtech/6158
- Adesemowo, A. K. et al. (2016). The experience of introducing secure e-assessment in a South African university first-year foundational ICT networking course. *Africa Education Review*, 13(1), 67–86. doi: 10.1080/18146627.2016.1186922
- Adesemowo, A. K., Oyedele, Y. & Oyedele, O. (2017). Text-based sustainable assessment: a case of first-year information and communication technology networking students. *Studies in Educational Evaluation*, 55, 1–8. doi: 10.1016/j.stueduc.2017.04.005
- Alruwais, N., Wills, G. & Wald, M. (2018). Advantages and challenges of using e-assessment. *International Journal of Information and Education Technology*, 8(1), 34–37. doi: 10.18178/ijiet.2018.8.1.1008
- Bennett, S. et al. (2017). How technology shapes assessment design: findings from a study of university teachers. *British Journal of Educational Technology*, 48(2), 672–682. doi: 10.1111/bjet.12439
- Binnahedh, I. A. (2022). E-assessment: Wash-back effects and challenges (examining students' and teachers' attitudes towards E-tests). *Theory and Practice in Language Studies*, 12(1), 203–211. doi: 10.17507/tppls.1201.25
- Bloxham, S. & Boyd, P. (2007). *Developing effective assessment in higher education: a practical guide: a practical guide*. United Kingdom: McGraw-Hill Education.
- Boud, D. & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400–413. doi: 10.1080/02602938.2015.1018133
- Brimkulov, U., Baryktabasov, K. & Jumabaeva, C. (2017). Information technologies in education: the learning assessment tools. *MANAS Journal of Engineering*, 5(2), 27–33.
- Bull, J. & McKenna, C. (2004). *Blueprint for computer-assisted assessment*. London: Routledge Falmer.
- Christie, M. F. et al. (2015). Improving the quality of assessment grading tools in master of education courses: a comparative case study in the scholarship of teaching and learning. *Journal of the Scholarship of Teaching and Learning*, 15(5), 22–35. doi: 10.14434/josotl.v15i5.13783
- Clarke-Midura, J. & Dede, C. (2010). Assessment, technology, and change. *Journal of Research on Technology in Education*, 42(3), 309–328. doi: 10.1080/15391523.2010.10782553
- Conole, G. & Warburton, B. (2005). A review of computer-assisted assessment. *ALT-J*, 13(1), 17–31. doi: 10.3402/rlt.v13i1.10970
- Contreras-Higuera, W. E. et al. (2016). University students' perceptions of E-portfolios and rubrics as combined assessment tools in education courses. *Journal of Educational Computing Research*, 54(1), 85–107. doi: 10.1177/0735633115612784
- Csapó, B. & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: the eDia system. *Frontiers in Psychology*, 10, 1522. doi: 10.3389/fpsyg.2019.01522
- Dammas, A. H. (2016). Investigate students' attitudes toward computer based test (CBT) at chemistry course. *Archives of Business Research*, 4(6), 58–71. doi: 10.14738/abr.46.2325
- Davey, G., De Lian, C. & Higgins, L. (2007). The university entrance examination system in China. *Journal of Further and Higher Education*, 31(4), 385–396. doi: 10.1080/03098770701625761
- Dermo, J. (2009). e-Assessment and the student learning experience: a survey of student perceptions of e-assessment. *British Journal of Educational Technology*, 40(2), 203–214. doi: 10.1111/j.1467-8535.2008.00915.x

- de Winter, J. C. F. & Dodou, D. (2010). Five-Point Likert items: t test versus Mann-Whitney-Wilcoxon. *Practical Assessment, Research, and Evaluation*, 15, Article 11. doi: 10.7275/bj1p-ts64
- Ferrari, A., Cachia, R. & Punie, Y. (2009). *Innovation and creativity in education and training in the EU member states: Fostering creative learning and supporting innovative teaching* [JRC Technical Note 52374], European Commission. Joint Research Centre.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. 4th ed. London: SAGE Publications Limited.
- Haifeng, L. (2012). The college entrance examination in China. *International Higher Education*, 68, 23–25. doi: 10.6017/ihe.2012.68.8617
- Harpe, S. E. (2015). How to fp analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning*, 7(6), 836–850.
- Hayden, M. & Thiep, L. Q. (2010). Vietnam's higher education system. In: *Reforming Higher Education in Vietnam. Higher Education Dynamics*, 29, 15–30. doi: 10.1007/978-90-481-3694-0_2
- Iğbay, M. A. (2005). A SWOT analysis on the university entrance examination in Turkey: a case study. *Mersin University Journal of the Faculty of Education*, 1(1), 126–140. doi: 10.17860/efd.08133
- Jordan, S. (2013). E-assessment: past, present and future. *New Directions*, 9(1), 87–106. doi: 10.29311/ndtps.v0i9.504
- Joshi, A. et al. (2015). Likert scale: explored and explained. *British Journal of Applied Science and Technology*, 7(4), 396–403. doi: 10.9734/BJAST/2015/14975
- Karl, M. et al. (2011). Student attitudes towards computer-aided testing. *European Journal of Dental Education*, 15(2), 69–72. doi: 10.1111/j.1600-0579.2010.00637.x
- Kim, Y., Kang, T.-S. & Rhie, J. (2017). Development and application of a real-time warning system based on a MEMS seismic network and response procedure for the day of the national college entrance examination in South Korea. *Seismological Research Letters*, 88(5), 1322–1326. doi: 10.1785/0220160208
- Kruger, D. et al. (2015). Improving teacher effectiveness: designing better assessment tools in learning management systems. *Future Internet*, 7(4), 484–499. doi: 10.3390/fi7040484
- Likert, R. (1932). A technique for the measurements of attitudes. *Archives of Psychology*, 140(22), 5–55.
- Liu, Q.-J. & Feng, Y.-R. (2009). Research and implementation of random question selection based on genetic and Tabu Algorithm. *Journal of Linyi Normal University*, 31(6), 136–139.
- Norman, G. (2010). Likert scales, levels of measurement and the 'laws' of statistics. *Advances in Health Sciences Education*, 15(5), 625–632. doi: 10.1007/s10459-010-9222-y
- Piattoeva, N. (2015). Elastic numbers: national examinations data as a technology of government. *Journal of Education Policy*, 30(3), 316–334. doi: 10.1080/02680939.2014.937830
- Piaw, C. Y. (2012). Replacing paper-based testing with computer-based testing in assessment: are we doing wrong? *Procedia – Social and Behavioral Sciences*, 64, 655–664. doi: 10.1016/j.sbspro.2012.11.077
- Shute, V. J. et al. (2016). Advances in the science of assessment. *Educational Assessment*, 21(1), 34–59. doi: 10.1080/10627197.2015.1127752
- Silin, Y. & Kwok, D. (2017). A study of students' attitudes towards using ICT in a social constructivist environment. *Australasian Journal of Educational Technology*, 33(5), 50–62. doi: 10.14742/ajet.2890
- Stödberg, U. (2012). A research review of e-assessment. *Assessment & Evaluation in Higher Education*, 37(5), 591–604. doi: 10.1080/02602938.2011.557496
- St-Onge, C. et al. (2021). Covid-19 as the tipping point for integrating e-assessment in higher education practices. *British Journal of Educational Technology*, 53(2), 349–366. doi: 10.1111/bjet.13169
- Tabachnick, B. G. & Fidell, L. S. (2007). *Experimental Designs Using ANOVA* (Vol. 724). Belmont, CA: Thomson/Brooks/Cole.
- Thelwall, M. (2000). Computer-based assessment: a versatile educational tool. *Computers & Education*, 34, 37–49. doi: 10.1016/S0360-1315(99)00037-8