

Automated Assessment, Face to Face

[doi:10.3991/ijet.v5i3.1358](https://doi.org/10.3991/ijet.v5i3.1358)

R. M. H. Al-Sayyed¹, A. Hudaib¹, M. Al-Shboul¹, Y. Majdalawi¹ and M. K. Bataineh²

¹ The University of Jordan, Amman, Jordan

² United Arab Emirates University, Al-Ain, UAE

Abstract—This research paper evaluates the usability of automated exams and compares them with the paper-and-pencil traditional ones. It presents the results of a detailed study conducted at The University of Jordan (UoJ) that comprised students from 15 faculties. A set of 613 students were asked about their opinions concerning automated exams; and their opinions were deeply analyzed. The results indicate that most students reported that they are satisfied with using automated exams but they have suggestions to improve the automated exam deployment.

Index Terms—Automated Assessment, Computer-Based Test (CBT), Computer-Adaptive Test (CAT), Internet-Based Test (IBT).

I. INTRODUCTION

As new technologies emerge, computers have spread to enter all aspects of our lives; one of the major applications that computers are being extensively used in education is assessment; this covers real exams and self assessment performance indicators.

Using computers in these two methods of testing (real and self performance) faced a wide welcoming from experts in the world; they aimed to measure the knowledgeable and the technical skills, and for this purpose, some companies produced special kind of applications to facilitate the assessments process such as Quiz Creator (<http://www.quiz-creator.com>) and QuestionMark (<http://www.questionmark.com>). In some organizations such as UoJ, however, they developed their own assessment software tool.

Automated exams can be networked, computer-based test (CBT), or computer-adaptive test (CAT). A new advancement in networked automated exams is the Internet-Based Test (IBT). IBT can be administered from any place in a university, a ministry or any place in the world. Organizations that administer the exams always have high security precautions to save exams and maintain information. This type of exams has been brought into operation in September 2005 [1] by the Educational Testing Service (ETS) to administer Test of English as Foreign Language (TOEFL) exams. Earlier than that and in 1992, the Graduate Record Examination (GRE) started to be computerized although Scholastic Aptitude Test (SAT) and Law School Admission Test (LSAT) are still following the paper-and-pencil style where the student has to wait some times, few weeks, before getting the results [2]. The content of the CBT exams are similar to the manual paper and pencil one where the exam normally goes in one direction. An advantage of the CBT exam over the manual traditional one is in the exam security and the automated grading. In CAT exams, answering a question will affect the difficulty level of the following questions which are selected by the com-

puter. The difficulty level might go up or down based on the previous questions history until the level reaches a certain stable level. This method might conclude the exam before completing the whole set of questions since the computer will have enough information to judge the level of the person being tested.

Before conducting this study, the researchers of this paper were expecting that all students will be in favor of the automated exams and we based our expectation on the following facts: (1) automated exams have higher test validity. Especially in CAT, it is possible to track the achievement of students' progress and records the needed statistics about the strengths and weaknesses of the test and see the effects of the distracters for the various questions, the time spent to answer each questions, the questions that the student changed his/her answers and the skipped questions before returning to them. These factors help test developers to improve the test validity. (2) More security: In traditional exams, the exams are delivered manually, in person, through various stages to students; where as in automated exams, the exam can be delivered directly to students. Even though, cheating in exams is possible; it is less likely to occur due to the expected randomness in choosing and distributing questions.

However, automated exams have some drawbacks; for instance, some traditional characteristics are missing. For example, the student might not be able to view the text or the questions as a whole; also the student cannot use the pen to highlight some important phrases. Automated exams require some advanced technologies and require a certain level of knowledge to deal with these technologies. Also, additional initial time investment is needed to generate automated exams. Furthermore, the automated exams could not be suitable to all course contents; such as writing algorithms, the steps of formula proof, etc.

In this paper, an empirical study is presented to evaluate the usability of automated exams and to compare the automated exams with the traditional ones. It presents the results of a detailed study conducted at UoJ that comprised students from various faculties. A set of 613 students were asked about their opinions concerning automated exams. The results indicate that the majority of students reported that they are satisfied of using automated exams even though they have suggestions to improve the automated exam deployment.

The rest of this paper is organized as follows: section II reviews the automated assessments issues and some related studies. Section III presents the research method; it describes the purpose, participants and the questionnaire. The results and the analysis are discussed in sections IV. Finally, the conclusion is drawn in section V.

II. LITERATURE REVIEW

Literature about automated assessment is rich but few of it evaluated automated assessment by university students. According to reference [3], the issues that influence the trust after changing from traditional testing to the automated one can be minimized by implementing more checks for plagiarism and cheating. The authors showed also that future developments may increase the trust by standardizing the results and reassessing ambiguous questions. The authors emphasized that the examination board has to trust its employees and associated personnel because information of the exams is stored electronically which makes it possible for editing before or after the exam; this is because computer-based assessment is open to different methods of cheating.

In reference [4], among other factors, the authors discussed the principles, advantages and challenges of online assessment. Security in online assessment, due to authentication difficulties, is a major issue. They indicated that in the absence of the face-to-face interactions, assessment and measurements become more critical. The authors showed that assessment is meant to measure the achievement of the learning goals by two categories of assessment; formative and summative. Formative assessment provides a continuous feedback to both the learner and the instructor; and the summative one is meant to assign a value to what has been learned.

Two different studies covered in [5]. The first study; which is similar to this research study, was to ask 162 students about the format of an accounting exam they prefer. The second study disclosed the reasons behind their selection. Both studies showed how the reasons for selecting a computer against a traditional paper-and-pencil are differing and can be explained. Out of 162 students, 89 preferred the paper-and-pencil format while the remaining 73 students preferred the computer one. For those who preferred the paper-and-pencil exams they indicated greater comfortable with this type of tests while who preferred the computer test indicated that the quick feedback was a good reason for their selection.

In reference [6], the authors conducted a study to examine if the results of online assessment were significantly different from traditional paper-and-pencil style. Among other results, the study revealed that there were no difference between the two methods in terms of age and grade point average. In addition, no differences were reported between gender, ethnicity and educational level between the two methods. The only difference reported is in the longer duration (time) it took the students to complete the paper-and-pencil exam.

In a study examined the attitudes of prehospital undergraduate students undertaking a web-based examination in addition to the traditional paper-and-pencil one [7], results showed high students satisfaction and acceptance of web-based exams. The study indicates that web-based assessment should become an integral component of prehospital higher education. This study concentrated on one category of students while this study covered almost all type of students studying at a big university like UoJ.

III. METHOD

The purpose of this research study is threefold. First, is to assess the satisfaction level of UoJ students as per their faculties about automated exams. Second, is to assess the

satisfaction level of students as per their university level about automated exams. Third, is to assess the students' satisfaction level as per their gender about automated exams. In none of the studies we intended to compare the number of students together; for instance, in the effect of student's gender on students' satisfaction, we did not consider the number of male students to female students a major determining factor; their satisfaction percentage, however, was calculated.

For this purpose, a sample of 650 undergraduate students was asked to answer a set of 29 questions (see Appendix A). The sample size (650) is determined due to the fact that the researchers decided to question the maximum number of students all in one day, who used the automated assessment recently, for that reason, the researchers went to the IT labs in King Abdullah II School for Information Technology (KAISIT) to the classes of the Computer Skills 2 service course that is offered to most university faculties in addition to other IT courses; this explains why the number of IT students is the majority in the sample. The total number of students that were able to be questioned on that day was 650 students. These questions are then grouped into seven categories for analysis purposes. After reviewing all the filled questionnaires, 37 invalid questionnaire rejected for the following reasons: (1) Some important information is missing such as the faculty, student level, gender...etc. (2) Some questionnaires have been filled carelessly by selecting strongly agree or strongly disagree or I don't know for the whole set of the 29 questions. (3) One side of the questionnaire is filled and the other one is left blank.

As shown in Table I, this study included 613 students from 15 faculties of UoJ. The table shows the faculties in ascending order based on the number of students included from that faculty.

For the purpose of this study, the questions were classified into 7 major categories; Table II summarizes those categories and the questions that belonging to them.

It is important to point out that the questions: 10, 15, 17, 21 and 26 (shown inside squares in Table II) have negative indications; therefore, the researchers carefully considered these questions and reversed their meaning when analyzed the data in order to have their meaning go with the students' satisfaction with automated exams.

For each question, the student has to select one of the five answers. Table III shows the answers that were available for students to select from and their weights.

Before conducting this study, the following hypotheses were postulated and set at the 0.05 level of significance:

- Automated exams are good tools for evaluating students at all university levels for theoretical contents
- Automated exams require using other type of questions other than the multiple choice questions
- Students trust the results of automated exams
- Automated exams are a secure tool for evaluating students fairly and quickly
- Students are in favor of using automated exams
- Automated exams require some improvements to satisfy all students' needs.

TABLE I.
FACULTIES AND NUMBER OF STUDENTS

Faculty	Count of Students
Law	5
sport	6
Rehabilitation	7
Agriculture	12
Sharia (Religion)	19
Nursery	23
Dentistry	25
Science	28
Medicine	35
Business	39
Education	39
Pharmacy	53
Engineering	79
Art	81
KAHSIT (IT)	162
Total	613

TABLE II.
CATEGORIES OF QUESTIONS AND THEIR COVERED AREAS

No	Category	Questions	Covers
1	Advantages	9, 11, 12, 13, 15, 17, 19, 27, 28	Flexibility, instructions, multimedia, suitability for contents, identifying mistakes, expressing answers, special need students, answering speediness and unanswered questions reminder
2	Duration	4, 5	Exam durations with respect to the nature and number of questions
3	Environment	26, 29	Cheating and concentration
4	Idea	1, 6, 18, 22, 24	Concept of automated exams, evaluation basis, content coverage style and fairness of evaluation (no human interaction!)
5	Mark	2, 8, 10, 21	Effects of automated exams on marks and its accuracy and displaying results
6	Questions	3, 7, 14, 23	Questions difficulty level, questions types and material coverage (comprehensive!)
7	Trust	16, 20, 25	Exam mark, questions security, and reports of mistakes

TABLE III.
ANSWERS AND THEIR WEIGHTS.

Answer	Weight
Strongly Agree	4
Agree	3
I don't know	2
Disagree	1
Strongly disagree	0

IV. RESULTS AND ANALYSIS

The t-test is used to give indications about UoJ students' satisfaction level of the various categories under study. The significant level is set to 5% (i.e. $p < 0.05$). Since the *average weight* for all answers is 2 (calculated as the summation of all weights and divided by their count $(4+3+2+1+0)/5$), a value of 2 or above (*of course when $p < 0.05$*) indicates a satisfaction level and any value below 2 indicates that students are not satisfied. When $p \geq 0.05$, however, it means that any conclusion cannot be drawn from the survey.

Table IV shows the overall results for UoJ students and their satisfaction level. It is clear to conclude that, in general, students are satisfied with the automated exams; they are satisfied with all its advantages, its duration, its idea, the mark accuracy and that is affected positively and emphasizing on seeing their results at the end of the exam, they trust its marking, security of questions, and when they get a report of their mistakes. Students are not satisfied, however, with exam environment and they believe that they encounter some difficulties in concentrating on the exam and that some students are able to cheat. Questions wise, it was difficult to draw any conclusion.

A. Detailed Results as per Faculties

As an example on how to fill the various satisfaction levels, here are some explanations in a detailed manner about how the faculty of agriculture and the faculty of art satisfaction levels are filled. As can be inferred from Table V, only the *trust* category from the faculty of agriculture has a positive indication (i.e. students are satisfied) and the researchers were unable to draw any conclusion about all other categories because $p \geq 0.05$ for all of them.

TABLE IV.
THE OVERALL RESULTS FOR THE 613 STUDENT (N=613).

Category	p	Mean	Satisfied? (Y/N/X*)
Advantages	.000	2.2696	Y
Duration	.003	2.1446	Y
Environment	.000	1.6803	N
Idea	.000	2.3086	Y
Mark	.000	2.3438	Y
Questions	.346	1.9710	X
Trust	.000	2.5721	Y

*Y means yes, N means no and X means cannot judge or neutral.

TABLE V.
RESULTS FOR THE FACULTY OF AGRICULTURE (N=12).

Category	p	Mean	Satisfaction
Advantages	.112	2.2130	X
Duration	.919	1.9583	X
Environment	.623	1.8750	X
Idea	.686	2.0833	X
Mark	.231	2.3125	X
Questions	1.000	2.0000	X
Trust	.002	2.7778	Y

TABLE VI.
RESULTS FOR THE FACULTY OF ART (N=81).

Category	p	Mean	Satisfaction
Advantages	.000	2.2702	Y
Duration	.032	1.6938	N
Environment	.001	1.6111	N
Idea	.044	2.1877	Y
Mark	.007	2.2438	Y
Questions	.133	1.8673	X
Trust	.000	2.5226	Y

Table VI shows the results for the faculty of art, by excluding the *questions* category; as being not useful to draw any conclusion; the duration and environment have negative satisfaction level. This means that UoJ art students are expecting better exam environment and exam duration. Please refer to Table II for more details on the areas covered in each category. The students, however, are satisfied with the exams *advantages*, *idea*, *mark*, and *trust* categories.

Before analyzing the results, it is important to explain how the Satisfaction Level Percentage (SLP) is calculated. For each category, the researchers add the sample size (n) of that category that is mapped to the satisfaction level (Y, N or X). Mathematically, it is given by formula (1).

$$SLP_i = \frac{\sum_{i=0}^m n_i}{S} \times 100\% \quad (1)$$

Where *S* is the total sample size (613 in the study), *n_i* is the sample size of each faculty, *m* is the number of differ-

ent faculties and *i* = 0, 1, ..., 15. A 0 value for *i* indicates that no faculty is mapped to that satisfaction level and *l* is the satisfaction level (Y, N, or X).

By looking at Table VII, as an example, the values for the SLP_Y, SLP_N and SLP_X for the *advantages* category are calculated as shown in equations: (2), (3) and (4), respectively.

$$SLP_Y = \frac{81+39+25+79+162+5+35+23+53+7+28+19}{613} \times 100\% = 90.7\% \quad (2)$$

$$SLP_N = \frac{0}{613} \times 100\% = 0.0\% \quad (3)$$

$$SLP_X = \frac{12+39+6}{613} \times 100\% = 9.3\% \quad (4)$$

It can be inferred from Table VII that not even a single faculty is satisfied (all faculties are VLS) with the exam environment and that the *advantages* and *trust* categories have high satisfaction level. In addition, the *idea* and *mark* categories have satisfied (S) level.

Also and based on the findings as summarized in Table VII, the researchers produced the graphs shown in Figure 1, Figure 2, and Figure 3, respectively.

TABLE VII.
RESULTS SUMMARY FOR ALL FACULTIES.

Faculty \ Category	Agriculture n=12	Art n=81	Business n=39	Dentistry n=25	Education n=39	Engineering n=79	IT n=162	Law n=5	Medicine n=35	Nursery n=23	Pharmacy n=53	Rehabilitation n=7	Science n=28	Sharia n=19	Sport n=6	Satisfaction level percentage for		
	Y	N	X	Y	N	X	Y	N	X	Y	N	X	Y	N	X	Y	N	X
Advantages	X	Y	Y	Y	X	Y	Y	Y	Y	Y	Y	Y	Y	Y	X	90.7	0.0	9.3
Duration	X	N	X	Y	X	Y	X	X	Y	X	Y	X	X	X	N	31.3	14.2	54.5
Environment	X	N	N	N	N	N	N	X	X	N	N	X	N	X	X	0.0	86.3	13.7
Idea	X	Y	Y	Y	X	Y	Y	X	Y	X	Y	Y	X	X	X	78.5	0.0	21.5
Mark	X	Y	Y	Y	X	Y	Y	Y	Y	X	Y	X	X	Y	X	81.2	0.0	18.8
Questions	X	X	X	X	N	X	X	X	X	X	X	X	X	Y	X	3.0	3.0	94.0
Trust	Y	Y	Y	Y	Y	Y	Y	X	Y	X	Y	Y	Y	Y	X	94.5	0.0	5.5

Note: By knowing any two satisfaction level percentages, the third one is easily calculated because the total percentages must add up to 100%.

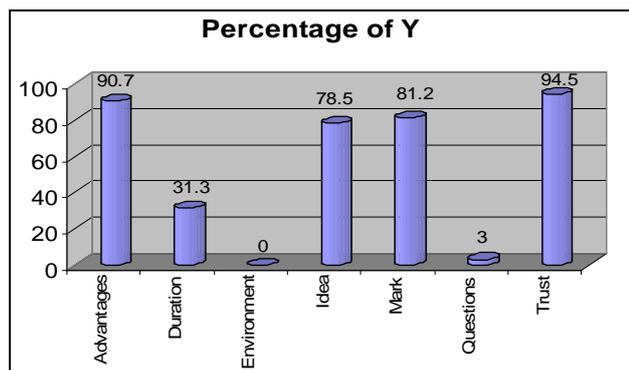


Figure 1. Automated Exams Acceptance Level (students in favor).

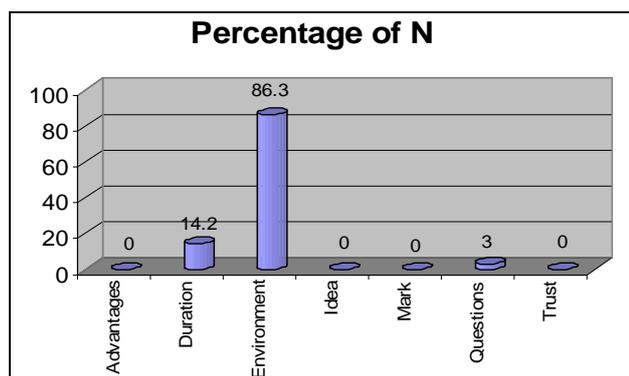


Figure 2. Automated Exams Rejection Level (students against).

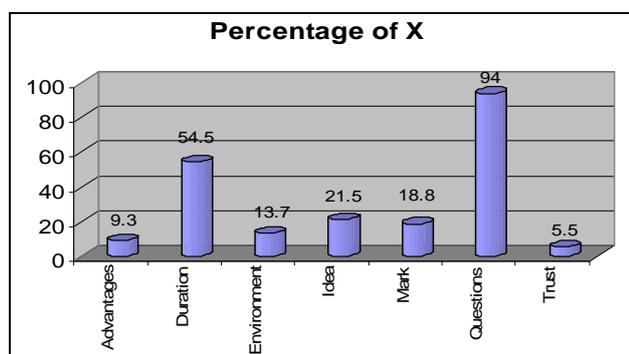


Figure 3. Automated Exams Uncertainty Level (students unsure).

To formally give a *logical* meaning to the meaning of the satisfaction level percentage, the researchers made the assumptions shown in Table VIII.

It is evident from Figure 1 that the students are highly satisfied with the advantages (90.7%) and the trust (94.5%) categories. They are also satisfied with its mark (81.2%) and idea (78.5%) categories. They are very low satisfied with the duration (31.3%) and questions (3%) categories. On the other hand, they are not satisfied with its environment (0%) category. From Figure 2, the rejection level is high for the environment category (86.3%), while there is no rejection (0%) for the advantages, idea, mark, and trust categories. The rejection level for the duration category (14.2%) and the questions (3%) categories is very low. Figure 3, indicates that the majority (high level) of students (94%) cannot make decision regarding the questions category while low percentage of students (54.5%) cannot make decision about the duration cate-

TABLE VIII. SATISFACTION LEVEL PERCENTAGE RANGES.

Percentage Range (v)	Meaning	Abbreviation
$v > 85\%$	Highly satisfied	HS
$65\% \leq v < 85\%$	Satisfied	S
$50\% \leq v < 65\%$	Low satisfied	LS
$v < 50\%$	Very low satisfied	VLS

TABLE IX. RESULTS AS PER UNIVERSITY LEVEL.

Year Level \ Category	First Year=249 (40.6%)	Second Year=114 (18.6%)	Third Year=126 (20.6%)	Fourth Year n=102 (16.6%)	Above 4 Years n=22 (3.6%)	Percentage of Y	Percentage of N	Percentage of X
Advantages	Y	Y	Y	Y	X	96.4	0.0	3.6
Duration	Y	X	X	X	X	40.6	0.0	59.4
Environment	N	N	N	N	X	0.0	96.4	3.6
Idea	Y	Y	Y	Y	X	96.4	0.0	3.6
Mark	Y	Y	Y	Y	X	96.4	0.0	3.6
Questions	X	X	X	X	X	0.0	0.0	100.0
Trust	Y	Y	Y	Y	Y	100.0	0.0	0.0

gory. For the other categories (advantages, environment, idea, mark and trust) in average, $((9.3+13.7+21.5+18.8+5.5)/5=13.8\%)$, about 14% of the participants were unable to make decisions regarding automated exams; although this is a very low level but it gives a good indication about the honesty of results; hence the conclusion was drawn based on the data that were received from students who are 86% sure (100% - 14%) about their opinions.

B. Detailed Results as per Student's University Level

As shown in the Table IX, first year students represents 40.6% of the sample size, second year represents 18.6%, third year represents 20.6%, fourth year represents 16.6, and 5 or more years represents 3.6%. Table IX also shows that the students in all university levels (1st year, 2nd year ...etc) are highly satisfied (100%) with the trust category of the automated exams. This high satisfaction level decreases very little to become 96.4% for the students who stayed in the university for 4 years or less in the advantages, idea and mark categories.

On the other hand, 96.4% of the students who spent not more than 4 years at the university are not satisfied with the exam environment and about 60% of the students who have already finished the first year (2nd year, 3rd year, .. etc) cannot make a clear decision about the exam duration while only 40.6% of students who are in the first year are satisfied with the exam duration category. As almost the case with all the reported results, the students in all university levels are not able to make any decision regarding the exam questions.

C. Detailed Results as per Student's Gender

In this study, the satisfaction level as per student's gender is presented; Table X presents a summary of the data collected for this purpose. The table shows that the percentage of females in the sample size is 65% and the percentage of males is 35%. The table also shows that 100% of male and female students are satisfied with the advantages, idea, mark, and trust categories. On the other hand, neither male nor female students are satisfied (their satisfaction level percentage is 0%) with the exam environment category. The satisfaction level percentage for male students drops to become very low (35.0%) for the duration and questions categories. Furthermore, Table X shows that all female students (65% of the sample size) are not satisfied with the questions category and this same percentage (for female students) cannot make decision about the duration category.

D. Comparison Among the Threefold

The results recorded about students were classified into three groups: results per faculties, results per university level, and results per gender. Then, these results were compared together. The results collected as per student's faculty, the results collected as per student's university level, the results collected as student's gender were compared and summarized in Table XI. To simplify analyzing the data, three graphs were produced and shown in Figure 4, Figure 5, and Figure 6, respectively.

When compared the Y satisfaction level, Figure 4 indicates that the difference between the three groups is minimal for the advantages, the duration and the trust categories. It also can be noticed that the students' opinions agree for the university level and gender for all categories except for the questions category.

As shown in Figure 5 and when comparing the unsatisfaction level (N) for the three groups, it can be noticed that there is agreement for all categories except for the questions and the duration categories. In Figure 6 and when comparing the "I don't know" level (X), however, the researchers found the agreement in the duration category. To wrap this out, the researchers found little variation between students' opinions about their satisfaction level.

TABLE XI. COMPREHENSIVE SUMMARY OF RESULTS.

Category	Y			N			X		
	All	Univ. Level	Gender	All	Univ. Level	Gender	All	Univ. Level	Gender
Advantages	91	96.4	100	0	0	0	9.3	3.6	0
Duration	31	40.6	35	14	0	0	55	59.4	65
Environment	0	0	0	86	96.4	100	14	3.6	0
Idea	79	96.4	100	0	0	0	22	3.6	0
Mark	81	96.4	100	0	0	0	19	3.6	0
Questions	3	0	35	3	0	65	94	100	0
Trust	95	100	100	0	0	0	5.5	0	0

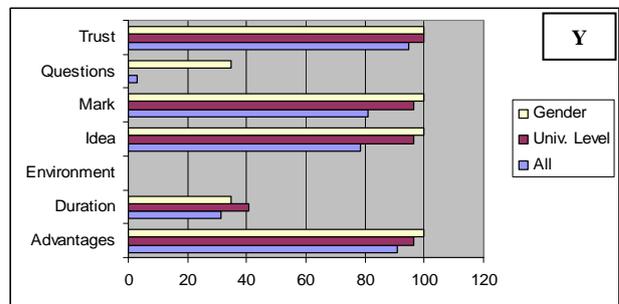


Figure 4. Y Satisfaction among the three groups.

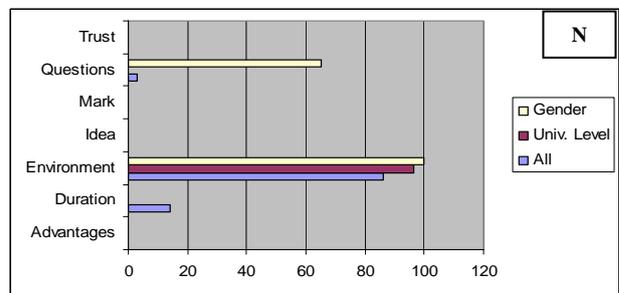


Figure 5. N Satisfaction among the three groups.

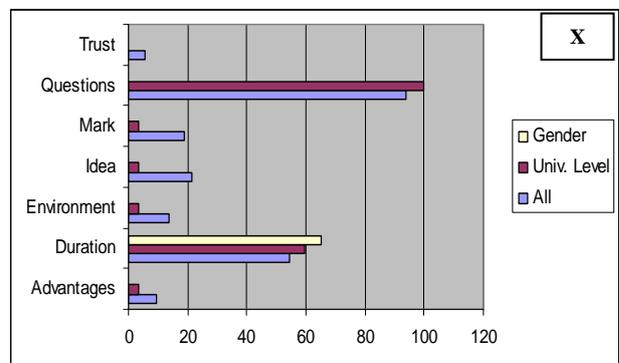


Figure 6. X Satisfaction among the three groups.

TABLE X. RESULTS AS PER STUDENT'S GENDER.

Gender \ Category	Gender		Percentage of Y	Percentage of N	Percentage of X
	Male=214 (35%)	Female=399 (65%)			
Advantages	Y	Y	100.0	0.0	0.0
Duration	Y	X	35.0	0.0	65.0
Environment	N	N	0.0	100.0	0.0
Idea	Y	Y	100.0	0.0	0.0
Mark	Y	Y	100.0	0.0	0.0
Questions	Y	N	35.0	65.0	0.0
Trust	Y	Y	100.0	0.0	0.0

E. Major Comments from Participants

In the questionnaire and in addition to the 29 questions, a space was left blank for any extra comments the students would like to add. Out of the 613 students, 105 comments were recorded by 89 students; i.e. some students wrote more than one comment. The major comments as added by student are summarized in Table XII.

It evident from Table XII that about 25% of the comments suggest that automated exams are good for some types of contents but not for all course contents, for instance, physics and C++ courses are preferred to follow the traditional style than being automated. Also, about 24% of the comments are asking to supply the students with a detailed report detailing his/her mistakes.

Also from the comment shown in Table XII, the students emphasize the importance of allowing students to navigate through the questions in both directions; not only forward; about 11% of the comments ask for allowing the students to make changes to their previous answers after answering and leaving that question. Regarding the functionality of computers, about 10% of the comments raise this issue as an important one for having the exams running smoothly without affecting students' achievements. Issues such as the ability to collect previous questions, the exam environment, cheating, and exams errors have about 3% of the comments weights. About 2% of the comments raise points about the number of the questions, partial credit for some questions, the fear from the remaining time and showing the final exam results. Because the final exam result is linked to the overall students' results and the curve of that course, results require processing before the students can be told about their final grades. The weight for each comment of the remaining 12 comments as shown in Table XII is 1%. The researchers believe that among these 12 comments, warning students about the remaining time, fixing cameras and separating between students in the exam hall are good comments that need to be addresses; the remaining 9 comments are either directly or indirectly covered in the 29 questionnaire questions.

V. CONCLUSION

This research was threefold. The first to assess the satisfaction level of all students about automated exams, the second to see the effect of students' university level on their satisfaction level about automated exams, and the third to assess the effect of students' gender on the satisfaction level of automated exams.

The study questioned a sample of 650 students each a set of 29 questions. Because some questionnaires missed important information, some questionnaires have been filled carelessly by selecting "strongly agree" or "strongly disagree" or "I don't know" for the whole set of the 29 questions, or one side of some questionnaires is left unanswered, 37 participants out of 650 were rejected and dropped out from the sample size. This research paper evaluates the usability of automated exams and compares them with the paper-and-pencil traditional ones. It presents the results of a detailed study conducted at UoJ that comprised students from 15 faculties; the opinions of the questioned students were deeply analyzed. The overall results of satisfaction as per students faculty, student university level and student gender indicate that the students are in favor of continuing using automated exams but they are looking for some improvements such as automated

TABLE XII.
TRANSLATION OF STUDENTS' EXTRA COMMENTS.

Item	Frequency	%	Comments
1	26	24.8	Automated exams are good for some types of contents not for all course contents
2	25	23.8	Students must see mistakes after the exams
3	12	11.4	Students must navigate in both directions not only forward (Some exams has one direction)
4	10	9.5	Be sure that the computers are functioning properly
5	3	2.9	Students can collect questions from previous students
6	3	2.9	Environment is noisy because of the proctors movements
7	3	2.9	Control various types of cheating
8	3	2.9	Automated exams might have entry errors
9	2	1.9	Reduce number of questions
10	2	1.9	Give partial credit to some multiple choice questions (MCQ) answers that require solution steps
11	2	1.9	Some students tend to fear looking at the time remaining
12	2	1.9	Show final exam results
13	1	1.0	Warn students about the remaining time
14	1	1.0	Transfer all IT courses to be automated
15	1	1.0	Add partition (or any kind of separation) between students in the exam hall to avoid watching each other screens
16	1	1.0	Install cameras for proctoring purposes
17	1	1.0	Give extra credit questions
18	1	1.0	Extend exam duration
19	1	1.0	Use different types of questions in addition to the MCQ
20	1	1.0	In case of failure, the student has to redo the whole exam in the remaining time
21	1	1.0	For each course, one exam should not be automated
22	1	1.0	Automated exams (AE) cause vision problems and concentration distraction
23	1	1.0	AE causes problems between the student and the instructor
24	1	1.0	Questions that deal with numbers only must be avoided
Total	105	100.0	

exams are good for some types of contents but not for all course contents, the need to supply the students with a detailed report detailing their mistakes after the exam, and allowing students to navigate through the questions in both directions; not only forward.

REFERENCES

[1] Y. Zhang, "Repeater Analyses for TOEFL® iBT (ETS RM-08-05)", Educational Testing Service research memorandum. 2008.

[2] K. Hayes, and A. Huckstadt, "Developing Interactive Continuing Education on the Web", Journal of Continuing Education in Nursing, 2000, Vol. 31, no. 5, pp. 199-203.

[3] R.D. Dowsing and S. Long, "Trust and the automated assessment of IT Skills", Special Interest Group on Computer Personnel Research Annual Conference, Kristiansand, Norway, 2002, Session: 3.1 IT Careers, pp.90-95.

[4] S. Kerka and E. W. M. E. Wonacott. "Assessing Learners Online: Practitioner File", ERIC Clearinghouse on Adult, Career, and Vo-

- cational Education (ERIC/ACVE), 2000, Retrieved June 10, 2010, from <http://www.ericacve.org/docs/pfile03.htm>.
- [5] K. Lightstonem and S. M. Smith, "Student Choice between Computer and Traditional Paper-and-Pencil University Tests: What Predicts Preference and Performance?", *Revue internationale des technologies en pédagogie universitaire / International Journal of Technologies in Higher Education*, 2009, Vol. 6, no. 1, pp. 30-45.
- [6] J. Bartlett, M. Alexander and K. Ouwenga, "A comparison of online and traditional testing methods in an undergraduate business information technology course", *Delta Epsilon National Conference Book of Readings*, 2000.
- [7] B. Williams, "Students' perceptions of prehospital web-based examinations", *International Journal of Education and Development using Information and Communication Technology (IJEDICT)*, 2007, Vol. 3, Issue 1, pp. 54-63.
- [8] ECDL, European Computer Driving License, 2010, Retrieved June 8, 2010, from <http://www.ecdl.co.uk>.
- [9] S. Hazari and D. Schnorr, "Leveraging Student Feedback to Improve Teaching in Web-Based Courses", *T.H.E. Journal (Technological Horizons in Education)*, 2007, Vol. 26, no. 11 pp. 30-38.
- [10] S. Long, "Computer-based assessment of authentic processing skills", Thesis, School of Information Systems, University of East Anglia, Norwich NR4 7TJ, UK, 2001.
- [11] L. W. Miller, "Computer integration by vocational teacher educators", *Journal of Vocational and Technical Education*, 2000, Vol. 14, no. 1, pp. 33-42.
- [12] M. W. Peterson, J. Gordon J, S. Elliott and C. Kreiter "Computer-based testing: initial report of extensive use in a medical school curriculum. *Teaching and Learning Medicine*". 2004, Vol. 16, no. 1, pp. 51-59. [doi:10.1207/s15328015tlm1601_11](https://doi.org/10.1207/s15328015tlm1601_11)
- [13] D. Potosky, and P. Bobko, "Selection testing via the internet: Practical considerations and exploratory empirical findings", *Personnel Psychology*, 2004, Vol. 57, no. 4, pp. 1003-1034. [doi:10.1111/j.1744-6570.2004.00013.x](https://doi.org/10.1111/j.1744-6570.2004.00013.x)
- [14] B. Smith and P. Caputi, "The development of the attitude towards computerized assessment scale", *Journal of Educational Computing Research*, 2004, Vol. 31, no. 4, pp. 407-422. [doi:10.2190/R6RV-PAQY-5RG8-2XGP](https://doi.org/10.2190/R6RV-PAQY-5RG8-2XGP)
- [15] M. Vrabel, "Computerized versus paper-and-pencil testing methods for a nursing certification examination: a review of the literature", *Computers, Informatics, Nursing*, 2004, Vol. 22, no. 2, pp. 94-8. [doi:10.1097/00024665-200403000-00010](https://doi.org/10.1097/00024665-200403000-00010)

AUTHORS

R. M. H. Al-Sayyed, A. Hudaib, M. Al-Shbool and Y. Majdalawi are with The University of Jordan, Amman, Jordan.

M. K. Bataineh is with the United Arab Emirates University, Al-Ain, UAE.

Manuscript received June 22nd, 2010. Published as resubmitted by the authors August 4th, 2010.

APPENDIX A

	Questions	Scale				
		4	3	2	1	0
1.	I prefer employing automated exams					
2.	Automated exams affect my mark positively					
3.	Questions in automated exams are easier than those in traditional exams					
4.	Exam duration of automated exams is fairly enough with respect to the nature of questions even if extra files are needed					
5.	The time duration of the automated exam is enough with respect to the number of questions					
6.	Automated exams evaluate the students properly					
7.	The type of questions in the automated exam is fair to evaluate the student					
8.	I prefer showing the results at the end of the automated exam					
9.	Automated exams are highly flexible (in terms of navigation through questions and making answers) when compared with traditional exams					
10.	Traditional exams marks are fair more than the marks of the automated exams					
11.	Instructions on how to use automated exams are fairly enough					
12.	I prefer using multimedia (voice, video, ...etc) in presenting questions because it helps me to better understand the questions					
13.	Automated exams are suitable for all courses and contents					
14.	Automated exams are comprehensive in covering all parts of the educational material when compared with traditional exams					
15.	One of the drawbacks of automated exams is that you cannot know your mistakes after knowing your mark					
16.	I trust my mark in the automated exam more than that in the traditional one in terms accuracy and error free					
17.	I believe that one of the drawbacks of the automated exams is that I cannot express my answers in details					
18.	I prefer to abandon using traditional exams and replace it completely with the automated exams					
19.	Automated exams consider special need students and is handy and easy to them					
20.	I trust that automated exam questions are always kept secure and not revealed to students					
21.	My mark is affected negatively in the automated exam when compared to the traditional one					
22.	Automated exams should cover only the theoretical parts and those parts that require steps to solve the questions shouldn't be automated (i.e. must be given using the traditional method on paper-based)					
23.	Automated exams must include other types of questions such as drag/drop, matching etc and not only multiple choice questions					
24.	The automated exam is a fair computer tool for evaluation since there is no human interacting (who might make mistakes)					
25.	My trust in automated exams will increase once I get a detailed report showing my mistakes					
26.	In automated exams environment, some students may be able to cheat without being noticed more than the environment of the traditional exams					
27.	I can answer automated exams questions more quickly and easily					
28.	An extra advantage of automated exams is that the ability to be reminded with those questions you have skipped or left unanswered					
29.	Automated exam environment helps me to concentrate more on the exam content since it has more equipment					