

## TWO APPROACHES IN STUDENT ACHIEVEMENTS ASSESSMENT – PRO ET CONTRA

**Erika Nagle**

**Aija Strode**

Rīga Stradiņš University, Latvia

**Abstract.** *Student achievement depends very much on reliable and valid assessment methods. In this paper assessment of reflecting writing and computer based on - line tests will be discussed. The main challenge in evaluation reflecting writing tests is objectivity. For this purpose test papers were evaluated mutually by different lecturers and grading compared by Pearson`s correlation coefficient. Results showed that strong correlation exists in all compared assessor pairs. In computer based on-line test in Moodle platform evaluation is done by computer programme and grading is announced just after student finishes the test. In purpose to maintain student`s reflective skills clinical case and/or problem were included in computer based tests. These questions were evaluated manually. Two types of assessment – manual and computer based have their advantages and disadvantages. Our conclusion is that different tests and assessment methods should be used for comprehensive and objective evaluation of student`s outcomes.*

**Keywords:** *Cohen`s kappa; Computer based tests (CBT); Paper-based tests (PBT); Student outcomes.*

### Introduction

The aim of medical education is to train students to become medical doctors with comprehensive knowledge and perfect practical skills in the field which will have a positive impact on the recovery and well-being of patients. Student outcomes depend very much on testing system, and reliable and valid assessment approaches. Two types of the tests and their assessment were analysed - paper based tests (PBT) assessment and computer based test (CBT) assessment. Paper based tests are widely used in medical universities. These tests mainly check student reflective writing skills. Reflective writing is a popular tool to support the growth of reflective capacity of undergraduate medical learners (Wald et al., 2012). It is well known that reflective writing impacts medical student empathy levels (Chen & Forbes, 2014), develops deepened understanding of patients` experience of illness, promotes practitioners` well-being (Wald & Reis, 2010), leads to improvement in skills such as communication, collaboration,

professionalisms, and serves as a tool for student assessment (Mamede et al., 2008). Computer based tests with the introduction of technologies in the teaching/learning environment can enhance student learning. CBT and their assessment have advantages because it is possible to provide students with the feedback just after the test, therefore assessment should be integrated in the learning process (Stobart, 2008). Additional benefits of CBT and their assessment include cost and time saving because of automated delivery, enhanced levels of student engagement because of relative novelty, and enhanced validity because of automation of the marking process that can reduce the scope for human error (Hewson et al., 2012).

### **Methodology**

This study describes student assessment by paper-based test and computer-based test in the two semesters of the academic years 2015/2016 and 2016/2017 at Riga Stradiņš University. Students' outcomes in PBT were evaluated in 23 randomly chosen tests. To secure reliability of the results chosen tests were anonymous and graders involved in assessment were labelled by the letters A, B, C, D and E. To ascertain reliability Cohen's kappa method was used. Each test was graded by 3 examiners, and Pearson's correlation coefficient was calculated. In paper based tests students have the complete set of the questions included throughout the time given for the exam. Common strategies observed in PBT include making notes on the margins of the test paper, marking key words by underlining them, ticking off answered questions. This study includes also CBT assessment for the student outcomes evaluation. Student outcomes evaluations by CBT were realized via the University online learning platform *Moodle*. Test results of 164 students were evaluated in this research. In data bank tree types of the questions were included: multiple choice questions (MCQ), matching questions (MQ), and short answer questions (SAQ). In addition to these questions clinical case and problem in genetics were included in the test. These two tasks were evaluated manually. In CBT students could skip, review and change answers. Tests taking mode shows disparity between PBT and CBT. To avoid possible confusion about tests taking mechanism in CBT tests pilot test for each topic was placed on e-studies. The e-studies administrator takes care of all administrative tasks: generate variants from the question data bank for each student individually, system provides a link to the students at a definite time and date who take the test online in the classroom, register time spent on each question and on the whole test. System provides grading of the test automatically and feedback to the student just student has finished the test. Before starting the test all tests takers are required to enter following information: name, surname, students ID number, and password provided just before the test. Administrator

provides individualized time for each student allowing students to start the test when they are ready to do it.

### Results

Total grading marks in both PBT and CBT are shown in figure 1 and figure 2.

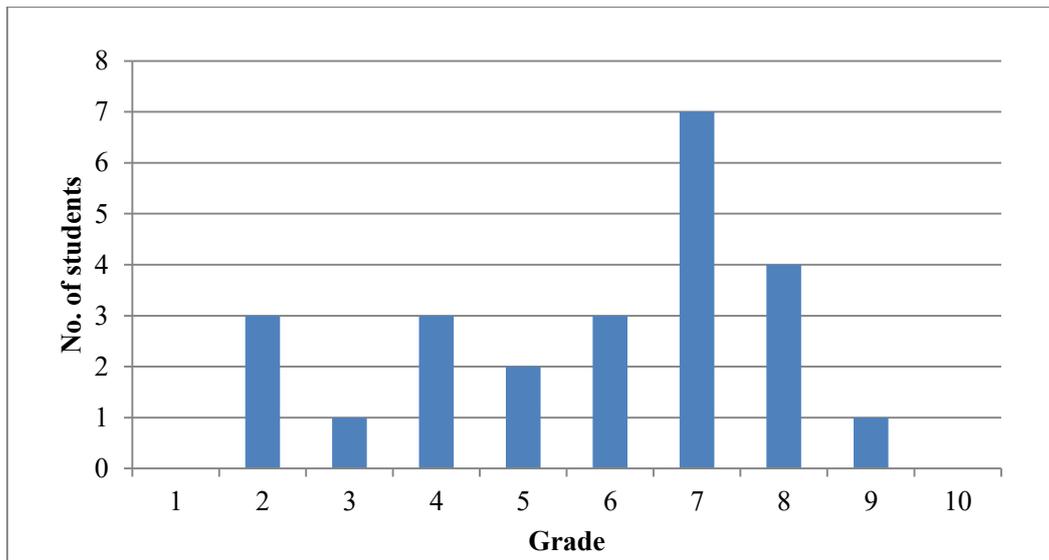


Figure 1 Overall test scores obtained in paper-based tests

Figure 1 shows that most of the test-takers received grade 7, only a few students failed, and no one received grade 10.

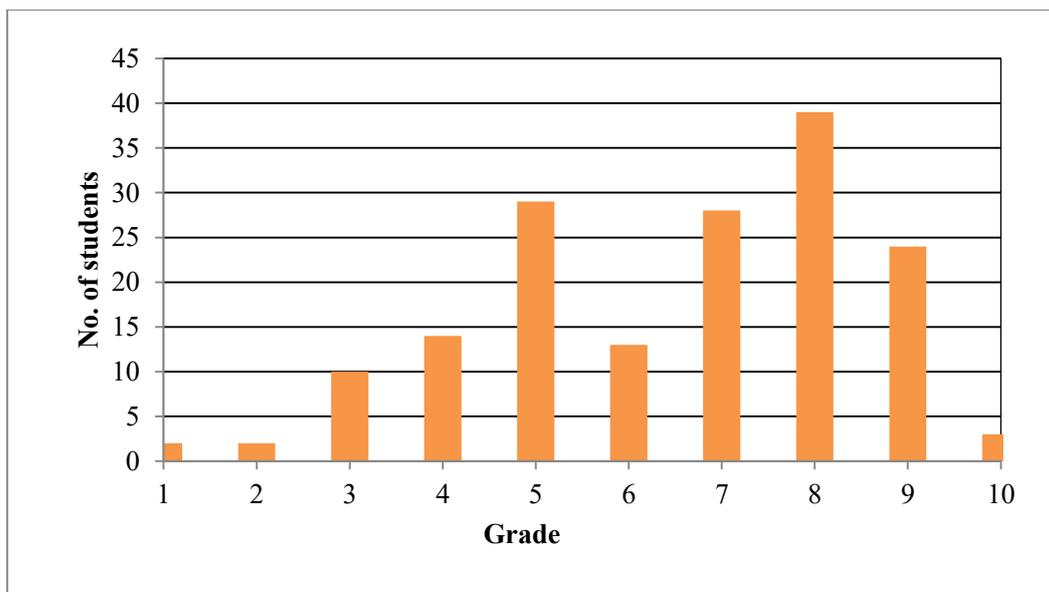


Figure 2 Overall test scores obtained in computer-based tests

In figure 2 it could be seen that most of the students received grade 8, a few students failed, and some received even grade 10.

**Table 1 Grades in PBT and CBT and number of students assessed**

Grade	No. of students	
	<i>CBT</i>	<i>PBT</i>
1	2	0
2	2	3
3	10	1
4	14	2
5	29	2
6	13	3
7	28	7
8	39	4
9	24	1
10	3	0

Results represented in table 1 show that overall test grades are higher in CBT in comparison with PBT. Grade 7 and upper in CBT received 57 percent of the students, but in PBT – 52 percent of the students. Furthermore in CBT three students received grade 10 – with distinction, but in PBT such grading was not received at all. Grade lower than 4 was received by 8.5 % of the CBT test takers and 17 % - PBT test takers.

For the purpose of verifying that all examiners had the same acquirements in PBT assessment correlation coefficient was calculated between pairs of examiners. Results are shown in table 2.

**Table 2 Correlation coefficients between pairs of graders**

Graders	A vs. B	A vs. C	A vs.D	A vs. E
<b>Correlation coefficient</b>	0.76	0.81	0.91	0.62
<b>P value</b>	0.0012	0.0004	0.0001	0.0271

As is shown in table 2, correlation is strong and statistically significant in all compared pairs of graders.

Study outcomes depending on the modes of the question were evaluated in CBT.

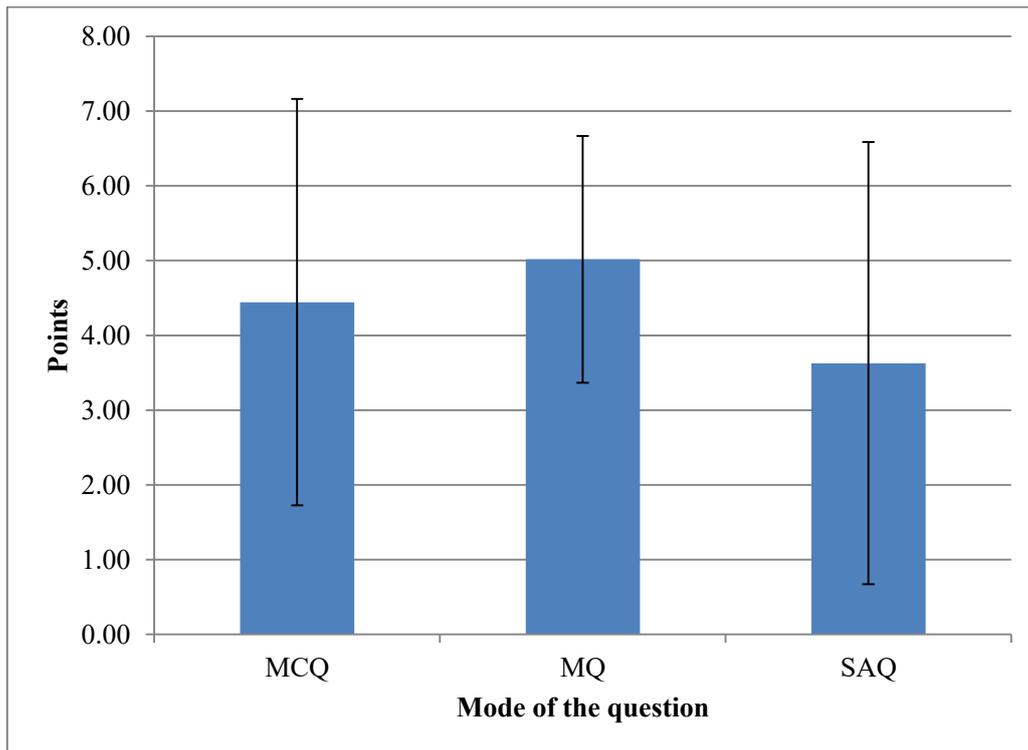


Figure 3 Grading depending on the modes of the question

Figure 3 shows grading results depending on the mode of the question and standard deviation. The highest grades examinees received in matching questions, but the lowest - in short answer questions, however our study did not show statistically significant differences between the modes of the question.

To strengthen our observation that the modes of the questions do not influence students' outcomes, median and mode for all three types of questions were calculated. Results are presented in figure 4.

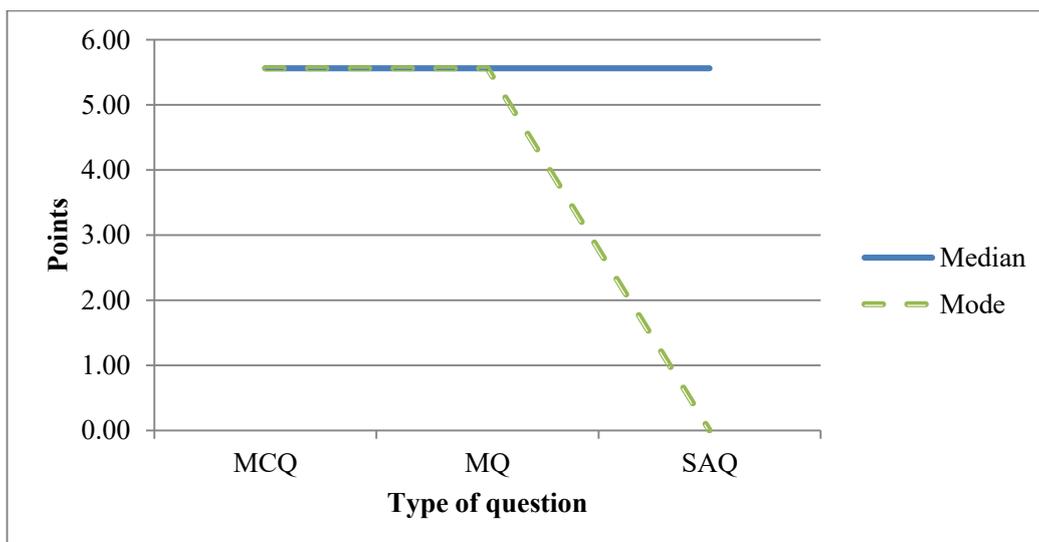


Figure 4 Median and mode values in three types of questions

As shown in figure 4, median values for all types of questions are shown by blue line in the graph. Median values were equal for all types of questions, which confirm that student outcomes do not depend on the mode of the question, and all three types of questions used in our tests should be used in student assessment to obtain reliable results. On the contrary, mode values (green dashed line in the graph) show that grades for MCQ and MQ questions were higher than for SAQ. This observation may indicate that SAQ were more difficult and in these answers students made more mistakes.

In paper-based tests students receive the feedback from the grader after the test is checked. Grader gives his/her opinion orally. Such feedback includes indication of mistakes, unclear formulation seen in the test, and teachers' conclusion about the test. CBT feedback is received immediately after student finishes the test and it gives information not only of final sum of the points, but also grading received in each question giving information about each topic separately.

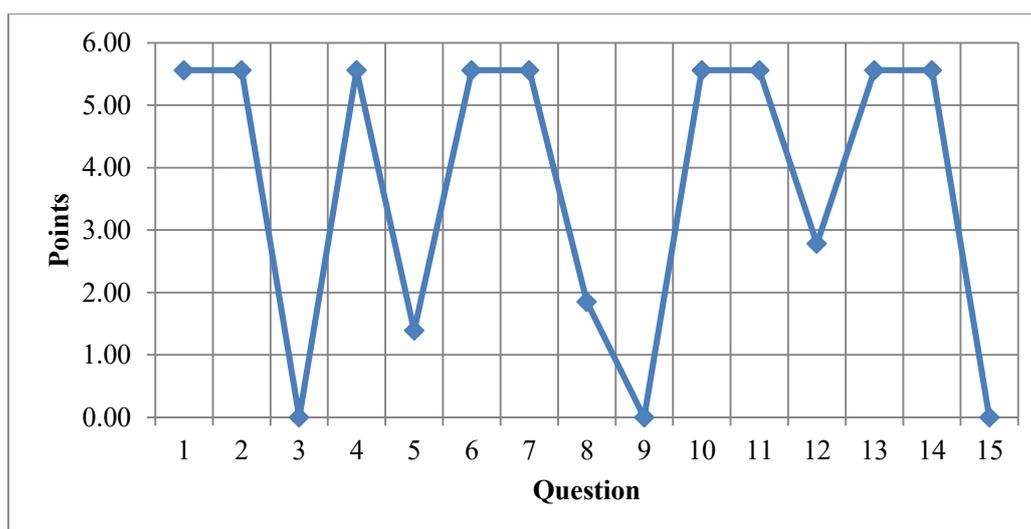


Figure 5 Example of the feedback received by a student after the CBT

Scores for each question are seen in points from maximum 5.5 points for each question. Feedback for the student “X” in the colloquium shows that the maximum point’s student received in questions 1, 2, 4, 6, 7, 10, 11, 13, and 14, several other questions, e.g. 5, 12 were partly correct, but questions 3, 9 and 15 were not answered at all. This allows student to understand their strength and weakness and to prepare better for the examination.

Additional points for the test were added after clinical case or problem in genetics was checked by lecturer. Final result student received within few hours after the test. Grading in accordance with RSU study regulations appear automatically on e-studies.

## Discussion

Paper based tests and computer based tests are widely used in students outcomes assessment in medical universities. The aim of our study was to analyse successes and failures of these tests and assessment methods. Feedback analysis allows students to understand their strong and weak topics and to prepare better for examination. In a given example student X failed in question 3 and in question 9. Student can analyse mistakes, ask the teacher to explain his/her mistakes, and read additional material if needed for better understanding of the weak topics.

Differences in mode values depending on the type of question were similar in MCQ and MQ, but very different and much lower in SAQ. This observation may be a topic for the future considerations and in SAQ formulations and predictive answers should be improved.

Regardless of the common belief that paper-based tests results depend very much on the perception of a grader, statistically different grading results between different graders were not observed in our study.

PBT and CBT tests were used to evaluate student achievements and advantages and disadvantages of each type of test were analysed. Main items are shown in table 3 and table 4. Table 3 provides the perception of teachers on various aspects of paper-based tests. It could be seen that if time is saved on test creating than it is wasted in addition for correction of the tests. Essay type questions could be considered as an advantage because students can show their creativity, logical thinking, ability to express their thoughts, however this advantages is shaded by possible bias in evaluation depending on grader.

Table 3 **Advantages and disadvantages of paper-based tests**

<b>Advantages</b>	<b>Disadvantages</b>
Shorter time is required for preparing the test (Ward & Murray-Ward, 1999)	Time consuming correction A. (Ward & Murray-Ward, 1999)
Essay type questions can be used (Ward & Murray-Ward, 1999)	Subjectivity of the examiner (Hoic-Bozic et al., 2008)

Table 4 shows views expressed by CBT by graders. It should be considered first that not all advantages are really advantages in all situations and vice versa. For instance, immediate feedback what could be evaluated as advantage in most cases, for some examinees may be a disadvantage, because it may cause test takers anxiety (Wise & Plake, 1989). It was recommended that feedback could not be used until its effects are better understood (Marla, 2006). As a disadvantage limited question format could be considered. Recently only three types of

questions could be included in data bank. As a time goes on this problem could be eliminated by developing more diverse question format. In addition CBT can be supplemented by questions which are not evaluated automatically by computers, but manually by examiner and scoring added to final grade.

As with any teaching/learning process, there are acquisitions, problems, and challenges to be met. Some of these problems have been considered to be improved, some have been ignored, but some other should be developed to perfect assessment process. Several disadvantages could be dismissed. For instance, some additional questions which are corrected manually can be added. Such approach was used in our CBT where calculation was added to automatically tested questions on/about clinical cases and also problems about recurrence risk prognosis.

Table 4 **Advantages and disadvantages of computer-based online tests**

<b>Advantages</b>	<b>Disadvantages</b>
Efficiency; less processing time (Karay et al., 2015). It takes less time for teachers to grade online tests if to compare with paper based tests.	Time consuming question bank preparation.
Automatic scoring with rapid feedback (Cantilon et al., 2004; Csapo et al., 2012; Yorke, 2005; Bartram & Bailey, 2010). The computer can automatically score the assessment and provide students with immediate feedback about their answers.	Limited question/answer format. Format is established by administrator and this includes MCQ, MQ, SAQ.
Cheating controls, increased test security (Kuzmina, 2010). Administrator has a variety of tools that prevent cheating.	Students cannot show their “extra” knowledge about the question. Answers to the questions are strongly restricted.
More objective assessment (Conole & Warburton, 2005). Computers make fewer errors than humans do.	Assessment is less creative, individual approach to each test is missed.

The hope is that, the advantages will become greater as the disadvantages will eventually disappear in both PBT and CBT evaluation.

### **Conclusions**

This study provides information about advantages and disadvantages of paper based and computer based grading. Several conclusions could be challenging for the improvement of learning/teaching process at the university.

1. Overall grading in CBT is higher than in PBT.
2. Mode of the question did not influence the students' outcomes.
3. Feedback information received immediately after the test can help in student learning process.

### References

- Bartram, B., & Bailey, C. (2010). Assessment preferences: A comparison of UK/international students at an English university. *Research in Post-Compulsory Education*, 15 (2), 177–87.
- Cantillon, P., Irish, B., & Sales, D.(2004). Using computers for assessment in medicine. *BMJ*, 329, 7466.
- Chen, I., & Forbes, C. (2014). Reflective writing and its impact on empathy in medical education: systemic review. *Journal of Educational Evaluation for Health Professions*, 11, 20.
- Conole, G., & Warburton, B. (2005). A Review of Computer Assisted Assessment. *ALT-J, Research in Learning Technology*, 13 (1), 17-31.
- Csapo, B., Ainley, J., Bennett, R. E., Latour, T., & Law, N. (2012). Technological issues for computer-based assessment. *Chapter in Assessment and Teaching of 21st Century Skills*, pp 143-230
- Griffin, P., McGaw., B., & Care, E. (2012). Assessment and teaching of 21st century skills. *Netherlands: Springer*: pp. 143.
- Hewson, C., & Charlton, J., & Brosnan, M. (2007). Comparing online and offline administration of multiple choice question assessments to psychology undergraduates: do assessment modality or computer attitudes influence performance? *Psychology Learning and Teaching*, 6, 37–46.
- Hoic-Bozic, N., Mornar, V., & Boticki, I. (2008). Introducing adaptivity and collaborative support into a Web-based LMS. *Computing and informatics*, 27 (4), 639-669.
- Kuzmina, I. P. (2010). Computer-based testing: advantages and disadvantages. *Вісник НТУУ «КПІ». Філософія. Психологія. Педагогіка: збірник наукових праць*. 1(28), 192–196.
- Mamede, S., Schmidt, H. G., & Penaforte, J. C. (2008). Effects of reflective practice on the accuracy of medical diagnoses. *Medical Education*, 42, 468–475.
- Marla, L., & Domino (2006). *Psychological testing: an introduction*. Cambridge University Press, pp. 475-485.
- Wald, H., Borkan. J., Taylor, J., Anthony, D., & Reis, S. (2012). Fostering and evaluating reflective capacity in medical education: developing the REFLECT rubric for assessing reflective writing. *Academic Medicine*, 87 (1), 41– 50.
- Wald, H. S., & Reis, S. P. (2010). Beyond the Margins: Reflective Writing and Development of Reflective Capacity in Medical Education. *Journal of General Internal Medicine*, 25, 746.
- Stobart, G. (2008). Testing times: The uses and abuses of assessment. *London: Routledge*.
- Ward, W., & Murray-Ward, M. (1999). Assessment in the Classroom. *Belmont, CA: Wadsworth Publishing Co.*
- Wise, S. L., & Plake, B. S. (1989). Research on the effects of administering tests via computers. *Educational Measurement: Issues&Practice*, 8 (3), 5-10.

*Erika Nagle, Aija Strode. Two Approaches in Student Achievements Assessment – Pro Et Contra*

Yassin Karay, Stefan, K., Schauber, Christoph, Stosch, & Katrin Schüttpelz-Brauns (2015). Computer Versus Paper—Does It Make Any Difference in Test Performance? *Teaching and Learning in Medicine*, 27 (1), 57-62.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45 (4), 477–501.