

Case Study on Performance and Acceptance of Computer-Aided Assessment

Bianca Schoen-Phelan, Brian Keegan
Dublin Institute of Technology, Ireland

Abstract

Computer aided assessment (CAA) comprises a set of assessment techniques that are conducted on a computer. These forms of assessment may be purely online or network based or single, stand-alone device based [1]. The device may be a desktop computer or include the use of mobile devices. This paper investigates if CAA leads to improved performance and satisfaction [2]. The student cohort that participated in this pilot study came from a degree in computer science. A preliminary investigation was performed by executing two continuous assessments with one cohort of four students in an advanced databases class. One of the assessments was purely paper-based (PBA), the other assessment was fully computer-aided. Both assessments were conducted in an open-book manner. Additionally, both contained an element of replication tasks and applied knowledge task [3]. It was anticipated that computing students would excel in computer-aided assessment, and that this form of assessment would significantly improve both satisfaction and task performance on the side of the students. However, detailed analysis revealed that there was no statistically significant increase using CAA versus traditional PBA. Furthermore, contrary to initial hopes, students did not appear to have gained a higher level of satisfaction conducting tasks on a computer. One of the reasons given was that most exams at the host institution are PBA; consequently, assessments should prepare for the related exams, and should also be paper-based. As the research reported here is a pilot study on a small cohort, follow up studies on larger and more diverse cohorts will further inform the body of knowledge in this area.

1. Introduction

In recent years there has been a trend towards including more technology in the classroom for student learning and assessment [1]. The inclusion of technology for student assessment Computer-Aided Assessment (CAA) refers to a multitude of approaches for testing student knowledge, all of them incorporating some degree of computerization of the assessment process [4]. In literature, we find

the following terms frequently used in this context [5];

- Computer Assisted Assessment
- Computerized Assessment
- Computer Based Assessment
- Computer-Based Testing
- Online Computer Based Assessment

In the context of this paper the term CAA is going to be used in a broader sense: a method of assessment that the student completes on a computer only. CAA can refer to tasks that also exist in paper versions, for example Multiple-Choice-Questions (MCQs), and methods that are strictly computerized, for example programming software in a bespoke programming language.

Among the often cited advantages of CAA we mainly find the following [1][6];

1. Ease of scale to accommodate large classrooms,
2. Immediate feedback to learners,
3. Reusability of tests

Within this context, we see a strong argument for the use of personal mobile phones for learning [7]. Most of the available studies have been conducted in third world countries where we see a distinct lack of infrastructure that schools are able to provide to their students [8]. Additionally, increased pressure is put on learning facilitators to include a myriad of techniques in first world education [9].

From previous experience of teaching in the School of Computing, there had been an assumption that particularly for this set of students, it would be natural to strive for strong technology inclusion for teaching and assessment. Students in computing are expected to embrace the opportunity to include technology in every aspect of their learning and assessment. This paper evaluates if computer-aided assessment actually does increase performance as well as satisfaction for computer science students.

Noteworthy in this context is also that the majority of assessment conducted in the host institution relies on traditional paper-and-pen methods. Discussions on this topic among staff revealed that the main reason was typically “ease of use” and “clarity of expectations” (on both sides).

In the following, this paper highlights the current debate over the usefulness of CAA in relevant literature. This is followed by a presentation of the study conducted, its findings and a subsequent critical discussion on the issues presented here.

2. Background on CAA

This section examines the current state-of-the-art of CAA. We do see a clear trend towards the inclusion of blended learning, most notably due to the natural inclusion of information technology learning for student-centred learning [10]. The use of virtual learning environments has even been seen to strengthen emotional intelligence in students [11]. It is often argued that the main reason for under-use of technology in education is a distinct lack of education of the educators [12]. Despite the availability of repositories listing interesting apps for technology-based assessment [13], educators often perceive these as a challenge to navigate, and consequently understand the plethora of apps available to find one suitable for the purpose of a technology course. Furthermore, in view of student inclusivity, it is sometimes not obvious which apps are available for Android and iPhone technology, or alternatively offer an online frontend that is mobile accessible. In many circumstances pricing of these apps are not immediately exposed.

Critical voices are already emerging. For example, Davis argues that Multiple-Choice-Questionnaires (MCQs), although the most used form of computerised assessment, cannot be the only form of technology assisted assessment in order to be academically credible [6]. Others already argue that the high expectations in CAA could not be fulfilled [14]. Furthermore, McKenna et al. discuss the importance of an institution wide strategy for quality assurance for CAA [15]. Furthermore, while general e-learning trends develop and adapt rapidly, assessment of students within this space largely still relies on the same principles used in traditional class room settings [16]. The biggest concern from a lecturer view is certainly what to do when things go wrong. Harwood recommends a list of recovery actions, mostly resolving around having alternative forms of assessment prepared in advance [17].

On the other hand, many benefits can be seen in CAA, such as consistency and objectivity of standards, and often facilitate immediate, automatic and detailed feedback [2]. On the other hand, not all forms of assessment are in fact lifting the burden on the educator but increase the burden on the educator. This may be one of the reasons for a rather low uptake of opportunities.

An increasing amount of mobile apps that have recently emerged for use in education are geared towards concept understanding. In fact, there is a large body of work that argues that concept maps or

mind maps aid students in understanding new concepts and relationships in a more meaningful manner compared to traditional forms of lectures [18][19], and also encourage critical thinking [20]. These can be created by any presentation or drawing tool and does not require domain-specific software. It is thus quite easy and straightforward to include this aspect into assessment. Consequently, this form of assessment is part of the applied knowledge evaluation reported on in this paper.

3. Research Methodology

In order to investigate the potential power of CAA in assessing computer science students two lab tests were designed for students of a Database class. Each lab test was worth 5% of their overall grade. The first lab test was conducted on a purely paper-and-pen basis. It consisted of four questions, of which the first two were simply replication tasks of previously learnt content. The latter two aimed at demonstrating that students can independently apply previously learnt material on a different problem (applied knowledge).

The second lab test was structured in a similar manner: the first set of questions were aimed at replication and the latter two at application and abstraction. However, this test was conducted on a computer. Blackboard [21] was used for the first part, and then students were asked to use a presentation tool of their choice in order to construct firstly a mind-map and secondly a presentation of the new material. This student cohort would be used to regularly using Office programs, either online or desktop based. Both tests awarded a maximum mark of 20 out of 100 for the replication part and 80 out of 100 for applied knowledge. There were two weeks between the tests. Both tests were open-book examinations.

Analysis was carried out on the basis of students' achievements with regards to their grade. In order to assess student satisfaction they were asked to complete a satisfaction questionnaire two weeks after completing the last lab test. The following set of questions were presented to the students with a mix of MCQ, Likert scale (LS) and free text field (FTF);

1. Which type of assessment to you prefer? [MCQ]
2. Please provide a brief explanation of your answer to Q1. [FTF]
3. Do you prefer closed-book or open-book assessment? [MCQ]
4. Please provide a brief explanation of your answer to Q3. [FTF]
5. On a scale from 1-5, where 1=not at all satisfied and 5=extremely satisfied, how satisfied were you with your achievements in the lab test 1 (paper based)? [LS]

6. Do you think that you would have achieved a higher score in lab test 1 (paper based) if it had been conducted on a computer? [MCQ]
7. Please provide a brief explanation of your answer in Q6. [FTF]
8. Have you ever used mind maps as part of your study? [MCQ]
9. What is your opinion on using a range of computer-aided resources for learning and assessment. [FTF]
10. On a scale from 1-5, where 1=not at all satisfied and 5=extremely satisfied, how satisfied were you with your achievements in the lab test 2 (computer-aided)? [LS]
11. Please provide any additional feedback in relation to lab test 2 (computer-aided) [FTF]

The lab tests themselves ran smoothly and everything went according to plan. Especially in the computer aided case there was no need for any intervention. The cohort itself consisted of four students. However, from a brief and informal interview session with the students after the computer-aided session it was surprising to find that they seemed irritated with the format. This is further evaluated and discussed in the following sections.

3. Analysis

3.1. Academic Achievements

Table 1 illustrates a broad overview of results marked out of 100. We see an improvement from paper-based to computer-aided assessment where minimum and average achievements are concerned. Interestingly, the maximum achievement is nearly constant. However, the Wilcoxon Rank sum test for small sample sizes reveals that this is not a statistically significant result.

Table 1. Overview of Achievements, n=4

Achievement	Paper-Based	Computer-Aided
Min	41	62
Max	76	79
Average	56.88	73.75

Table 2 depicts a more detailed breakdown of the results presented in Table 1, which identifies if students performed better in replication or application type tasks.

Table 2. Results by Test Type, n=4

	Paper-Based	Computer-Aided
Average replication (out of 20)	11.13	13.75
Average application (out of 80)	45.75	60

This breakdown highlights how different styles of tests are affected by the method of examination. It appears that applied knowledge tests specifically benefit from CAA. Interestingly, the Wilcoxon ranked sum test calculated for both types of tests individually that the improvement is indeed statistically significant.

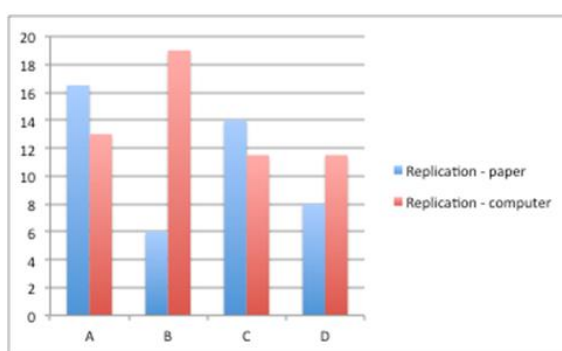


Figure 1. Replication Task by Individual Student and Assessment Method

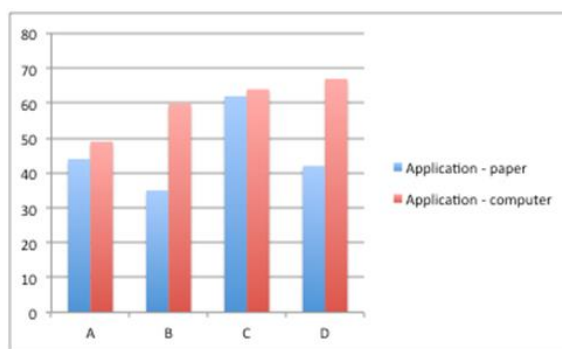


Figure 2. Application Task by Individual Student and Assessment Method

Figures 1 and 2 provide a breakdown by individual student in order to illustrate some interesting differences among subjects. For example, subject B appears to generally benefit from CAA. In contrast to this, subjects A and C appear to generally benefit from a paper-based approach. Furthermore, the gaps between paper and computer-based tests are much narrower in the case of these individuals, compared to the other subjects. In the case of applied knowledge tasks Figure 2 shows us that subject C is

not affected by the method of assignment, while subject D excels in this case.

3.2. Satisfaction Questionnaire Results

Section 2 of this paper presented the 11 questions that formed the post-test satisfaction questionnaire. This section analyses and critically discusses findings. Most of the questions contained a lead question asking for the students' opinion on a Likert scale, followed by an opportunity for the student to further support their opinion with a free text field.

Table 3. Assessment Preferences

Assessment Type	Amount
Computer-Aided	1
Paper-Based	3
I don't mind	0

Surprisingly, there is a strong preference for paper-based assessment among this student cohort, which needs to be further investigated. The most striking response came from one student who stated that it would be more useful to have paper-based assessment due to the fact that most exams in the host institution are paper-based and that the assessment should prepare for the exam. However, other responses indicated that the host institution's infrastructure does not support these tasks very well, while other students did not want to commit to an opinion.

Table 4. Open-Book vs Closed-Book Examination

Assessment Type	Amount
Open-Book	2
Closed-Book	0
I don't mind	2

Interestingly, most students preferred open-book examinations or did not have a preference. However, none of the students preferred closed-book examinations. This is an interesting result with respect to informing any module redesign tasks. Students explained their preference by acknowledging that they can see a benefit in both forms of assessment, or explained that they preferred a mix of both assessment types.

Regarding student satisfaction with their results, students were requested to reflect on both tests, paper-based and computerized using a Likert scale from 1-5, where 1 indicated a strong dissatisfaction and 5 a strong satisfaction.



Figure 3. Satisfaction with Lab Test 1 (paper-based)

Considering that most students preferred paper-based assessment to CAA, it is surprising to find that they were not more satisfied with their achievements. Furthermore, none of the students thought that they would have achieved a better result had the test been conducted on a computer. When asked to explain their evaluation of the situation, students provided the probably most diverse set of responses, ranging from the self-awareness of having not studied enough to being surprised about the actual form of the assessment although students had been prepared in class prior to the examination about what format to expect. Most likely, these results should caution to avoid over-generalization from studies such as these, as in education there exists a strong level of individual differences among students.

Contrary to satisfaction on the paper-based examination, opinions on the computerized case were much more varied, as illustrated in Figure 4. This is a more varied response in terms of satisfaction and is therefore inconclusive.

Question 8 checked whether the mind-map component of the computer-based assessment was a familiar tool to students or if an additional difficulty had been introduced.

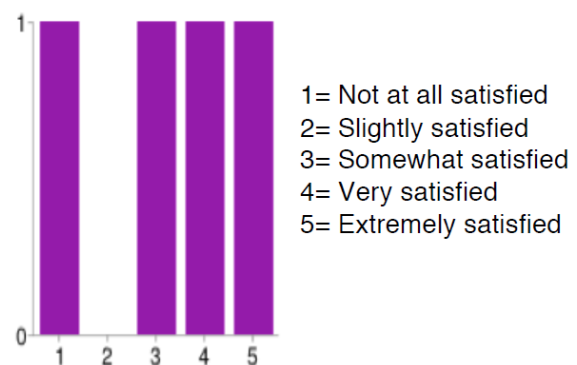


Figure 4. Satisfaction with Lab Test 2 (computerized)

Table 5. Previous Experience with Mind Maps

Previous Experience	Amount
Used Before	2
Not used before	1
Used either mind maps or other techniques at times	1
Not sure	0

The majority of the cohort has used these methods before. Unfortunately, this also revealed that there was one individual who was not familiar with the techniques. This case had not been foreseen due to the fact that no specialty tools were required. Google docs and Microsoft Office are commonly used for all assignments and the students' project. However, in future, this needs to be handled in advance of the actual test. Students need to be questioned in class and well in advance of the test regarding their proficiency with these tools.

Furthermore, opinions on the usefulness of these methods are varied, as **question 9** revealed. One student insisted that there is no relationship at all between creative thinking and being tests and that it is not appropriate to combine these elements. However, the majority of students found this format interesting and appropriate in the context of a course in computing. This is a more varied response in terms of satisfaction and is therefore inconclusive.

Finally, **question 11** gave students an opportunity to voice any final opinions regarding CAA. A particularly interesting response came from a student who felt that he did not deserve the grade he received in the applied knowledge task. This student might not have fully understood the rationale behind this type of test and is more used to traditional forms of assessment that are more directed. However, the level of responses throughout the questionnaire revealed the maturity and deep insight of the students in this cohort.

4. Conclusions

The initial hope of this study had been that CAA yields better results for students and would also engage them creatively on a stronger level and therefore increase satisfaction. To some extent the study revealed that CAA does indeed produce better results, but did not provide an overall statistically significant improvement. Strong individual differences influenced student performance and satisfaction. Overall, CAA did not provide the anticipated increase in satisfaction.

Student responses clearly showed a heightened degree of confusion rather than satisfaction, which may be mainly due to this type of testing being largely not used within this school at the host institution.

Shortcomings of this study certainly lie in the reduced number of students, which means it can be seen more as an initial pilot study rather than a full conclusion on the matter. Qualitative methods, such as an informal guided discussion immediately after the assessment in order to gain a general understanding have been helpful and were complimented by a structured questionnaire, that balanced Likert scales and MCQs with text fields in order to provided greatest possible exploitation of the opportunity.

It could be argued, that the fact that the second assessment was CAA, that students now knew to study better, and also that the material as well as the style of the lecture are more familiar and would naturally produce better results. However, if this were the case, an overall statistically significant increase in performance across all types of tests should have been determined, which was not the case. Specifically, the replication tasks should have resulted in a stronger performance improvement had students studied more.

This study will need to be re-run on a more diverse and larger set of students in order to gain more data on larger student groups by and getting other groups involved. Furthermore, given that one of the main advantages of CAA could be rapid feedback, as mentioned in introduction section, the assessment methodology could be improved in order to provide a solution to students immediately, which may increase satisfaction with this method.

5. References

- [1] G. Conole and B. Warburton, "A review of computer-assisted assessment," *Research in Learning Technology*, vol. 13, no. 1, pp. 17–31, 2005.
- [2] A. Tshibalo, "The potential impact of computer-aided assessment technology in higher education," *South African Journal of Higher Education: NADEOSA*, vol. 6, no. 21, pp. 684–693, 2007.
- [3] G. Brown, J. Bull, and M. Pendlebury, *Assessing student learning in higher education*. 2013.
- [4] W. Müller, C. Bescherer, U. Kortenkamp, and C. Spannagel, "Intelligent Computer-aided Assessment in Math Classrooms: State-of-the-Art and Perspectives," in *Conference on Imagining the future for ICT and Education*, 2006.
- [5] C. O'Farrell, "Enhancing Student Learning Through Assessment," 2014.
- [6] P. Davies, "Computer Aided Assessment MUST be more than multiple-choice tests for it to be academically credible?," in *Proceedings of the 5th International CAA Conference*, 2001, pp. 145–150.
- [7] T. Ekanayake and J. Wishart, "Investigating the possibility of using mobile phones for science Teaching and Learning: is it a viable option for Srilanka,"

International Journal for Cross-Disciplinary Subjects in Education (IJCDSE), vol. 2, no. 2, pp. 372–380, 2011.

[8] S. Ekanayake and J. Wishart, “Identifying the Potential of Mobile Phone Cameras in Science Teaching and Learning,” *International Journal of Mobile and Blended Learning*, vol. 3, no. 2, pp. 16–30, Jan. 2011.

[9] C. Nerantzi, A. Middleton, and S. Beckingham, “Facilitators as co-learners in a collaborative open course for teachers and students in higher education | Open Education Europa,” *eLearning Papers*, vol. 39, 2014.

[10] B. Aguti, R. J. Walters, and G. B. Wills, “Effective Use of E-Learning Technologies to Promote Student-Centered Learning Paradigms within Higher Education Institutions,” *International Journal for e-Learning Security (IJeLS)*, vol. 4, no. 2, pp. 391–398, 2014.

[11] F. Donkor, “Towards Emotional Intelligence behaviour within Virtual Learning Environments: Perceptions of Secondary School Teachers,” *International Journal for e-Learning Security (IJeLS)*, vol. 4, no. 1, pp. 359–365, 2014.

[12] H. Fuller, “First teach their teachers: Technology support and computer use in academic subjects,” *Journal of research on computing in education*, pp. 511–537, 2000.

[13] E. Inc., “edTechTeacher”, 2014. [Online]. Available: <http://edtechtteacher.org/apps/>. [Accessed: 01-Jan-2016].

[14] M. Jenkins, “Unfulfilled Promise: formative assessment using computer-aided assessment,” *Learning and Teaching in Higher Education*, pp. 67–80, 2004.

[15] C. McKenna and J. Bull, “Quality assurance of computer-assisted assessment: practical and strategic issues,” *Quality Assurance in Education*, pp. 24–32, 2000.

[16] E. G. Morales-Martinez, E. O. Lopez-Ramirez, and A. E. Lopez-Gonzalez, “New Approaches to E-Cognitive Assessment of E-Learning,” *International Journal for e-Learning Security (IJeLS)*, vol. 5, no. 2, pp. 449–453, 2015.

[17] I. Harwood, “When summative computer-aided assessments go wrong: disaster recovery after a major failure,” *British journal of educational technology*, pp. 587–597, 2005.

[18] J. Novak, “Concept mapping: A useful tool for science education,” *Journal of research in science teaching*, vol. 27, no. 10, pp. 937–949, 1990.

[19] J. Novak and A. Cañas, “The theory underlying concept maps and how to construct and use them,” *Florida Institute for Human and Machine Cognition*, vol. 284, 2008.

[20] D. Willingham, “Critical thinking: Why is it so hard to teach?”, *Arts Education Policy Review*, Taylor & Francis, vol. 109, no. 4, pp. 21–32, 2008.

[21] B. Inc, “Blackboard.” [Online]. Available: <http://www.blackboard.com>. [Accessed: 31-Dec-2015].