

Evaluation of Digital Assessment

Kurt Gammelgaard Nielsen¹, Lise Petersen², Birgitta Wallstedt³, Poul Basse⁴, Pernille Stenkil Hansen⁵, Søren Sten Hansen⁶, Dorthe Majlund Sørensen⁷

¹ IT-service, kgn@sdu.dk (corresponding author); ² IT-service, lise@sdu.dk; ³ SDU Centre for Teaching and Learning, bwallstedt@sdu.dk; ⁴ Faculty of Social Sciences, pba@sdu.dk; ⁵ SDU Centre for Teaching and Learning, pha@sdu.dk; ⁶ Faculty of Science, shan@sdu.dk; ⁷ Faculty of Health Sciences, dksorensen@health.sdu.dk;

University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

1. ABSTRACT

University of Southern Denmark (SDU) decided to start a project to implement digital assessment in May 2010. By the exam term of January 2014 the formal goal of the implementation project - all written exams should be digital - was reached by all 5 faculties and 5 campuses. Paper is no longer used at exams, and the students handed in a total of 75.000 digital assignments in 2013 (SDU Assignment Stats, 2014).

In this paper we will evaluate the project according to generally accepted criteria for quality of assessment:

Validity - the assessment measures the whole testable domain (knowledge, skills, competences) that it is supposed to measure.

Reliability - the assessment is reliable - i.e. there is a consistent discrimination between good and poor performance.

Acceptability - the assessment is widely accepted as being 'good' by students and faculty.

Feasibility - the resources, which are required to conduct the assessment, are available

According to these criteria the implementation of digital assessment at SDU has been partially successful:

- A conscious approach to assessment and ongoing discussions has brought us a long way in increasing validity and reliability.
- The acceptability of digital assessment is high among students.
- The resources have been present, and the goal of feasibility of digital assessment has been partly achieved, due to automation of workflow and due to students bringing their own computer (BYOD).

The formal goals of implementing digital assessment at Danish Universities are described in the Danish eGovernment strategy (2011). The overall goals of digitization are to strengthen the teaching and increase operational efficiency. By using our set of four criteria in evaluating digital

assessment we attempt to impose more aspects to the evaluation, and to address the quality aspect in a more concrete and direct way.

1. Overview - The Digital Assessment Project at SDU

SDU offers 222 programs (Bachelor and Master). The overall goal of the Digital Assessment Project is that digital assessment is implemented in all courses using written exams. Digitization of teaching and assessment is part of the foundation of a modern university, and the society expects that graduates are competent in working digitally.

The project period formally was 2010-2013. It was setup by the vice-chancellor with a steering committee and 5 working groups, in accordance with the Danish eGovernment Strategy (2011). The steering committee has had 17 meetings, and in September 2012 the Business Case was extended to include digital workflow before, during and after the exam.

The 5 working groups were:

1. Legal advice group, implementation of ministerial order, SDUs obligations to students etc.
2. Technical learning management system group, handling and administrating the possibilities in the LMS.
3. IT-technical group, developing and implementing technical requirements.
4. Invigilator group, education and coordination of invigilators.
5. Workgroup for digital feedback and anti plagiarism.

Many different assessment methods are in use at SDU, incl. oral examination, and several IT-tools have been developed and implemented and are under continuous development to meet the demands of a multifaceted university. At an early stage in the project it was decided that the BYOD principle (Bring Your Own Device) was to be used in written exams, to ensure flexibility, smoother logistics and a higher service level for the students.

Initially, the focus of the project was primarily on the following methods of assessment (exams):

- Tests with multiple choice and multiple answer questions.
- Essay exams with and without access to the Internet and to other sources of information (books, notes etc.). Some of them with access to special applications like SDUs digital pen tool, SDU Scribble, various math software or similar.
- Combinations of the above.

In the later phases of the Digital Assessment Project, more informative assessment methods, incl. portfolio, have received more attention.

In addition, at the Faculties of Science and Social Sciences there is a tendency towards more digital tests, either to replace or to supplement other exam formats (essay exam, take home assignment). Whereas the Faculty of Health Sciences was well under way with a reorganization of teaching and exam when the Digital Assessment Project started, and tests was already a widely used exam format.

During the project period exam formats have received increased attention, and been discussed more in the new digital context - and even though the project formally ended, this process is still ongoing.

A series of workshops on assessment methods are offered by the SDU Centre for Teaching and Learning. Hopefully, this will support the development towards increasingly valid and reliable assessment strategies in the future.

2. Validity and reliability

Validity means that the assessment measures the content (knowledge, skills, competences) it is supposed to measure - that the entire subject area is covered, and the exact outcomes of the course are tested. *Reliability* means that the assessment is reliable - i.e. there is a consistent distinction between good and poor performance.

At an early stage of the project we spotted the potential of digitizing tests, and digital tests have become widely used at SDU. Digitization allows you to ask and mark numerous questions. Validity and reliability is a function of the number of questions asked in the test - a high number of questions ensure that a substantial amount of the content is tested (leading to a high *content validity*). This also means that the result is not determined by chance - all students answer the same questions, and there are no inter-item reliability problems. Furthermore, it is easier to create question databases across institutions and nations.

At the same time, *reliability* is affected by the quality of the individual question - a question no one is able to answer, or one that every student answers correctly, does not discriminate between good and bad performance. Digitization facilitates the (statistical) analysis of the test results so that the inferior questions (with low discrimination) can be identified and removed, in the process of marking. Thereby *reliability* as well as *validity* is very well supported by digitization.

SDU and The Faculty of Health have done much work developing assessment with digital tests, primarily with multiple choice questions. In the period of March 2012 to February 2014 The Faculty of Health conducted 85 digital test exam events, with a total of 10.000 participating student (app.). Typically, the tests consist of a high number of questions (50 questions or more per hour).

According to the Faculty of Health Sciences, digital assessment has not been a precondition for using multiple choice tests as an exam format. The implementation of tests began before the digital assessment project, and was done without digitization, but digitization allowed a feasible way of handling the tests and test exams, and allowed for a safer and faster statistical analysis and subsequent processing of exam results.

At the Faculties of Science and of Social Sciences the use of digital tests for assessment is increasing. However, a slightly different approach is used, where each question requires time consuming calculations to be made, and the result thereof enables the student to pick the right answer. This assessment form has proven fully able to discriminate between good and poor performance (high *reliability*), and has

gained acceptance from faculty as well as students. Indeed, one exam in Biochemistry with 93 students, consisting of 50% essay and 50% test with multiple choice questions, showed good correlation between individual students performance in the two parts of the exam. The two parts were marked by different persons. This result further supports the general view that digital test exams are reliable.

One aspect of *reliability* is the risk of students cheating and plagiarizing. In general, it is wise to assume that a certain fraction of students will try to cheat. At SDU this problem is being addressed in several different ways.

Digital assessment allows for some measures that make cheating far more difficult. Among them are anti-plagiarism software, and software that partly lock the computer, and block Internet access.

With open-internet exams, anti plagiarism software is a very important tool in hindering cheating and plagiarism - and we have experienced that when the students are informed that their assignment will be checked they are less likely to cheat. It is easier to cheat when the content of an assessment is presented in the same way to all students, but in a digital test it is possible to randomly select test questions from a databank, or randomize the order of questions, or the order of answer possibilities.

Special courses have been designed for our corps of invigilators to make them able to spot when students cheat. Furthermore, hands-on inspections of student computer during exam have been done at random, by IT-staff (app. 1000 spot-checks).

In 2013 175 students was caught cheating/plagiarizing, by copying text from the internet, from fellow students or in other ways breaking the rules of academic integrity. During the period of implementing digital assessment the number of students caught cheating has been increasing.

When SDU compare these numbers with other Danish Universities SDU has a higher relative score. However, the baseline is not the same, in that SDU, due to digitization and anti-plagiarism software, has a better chance of spotting the cheaters. The other universities have digitized their assessment to a lesser extent, and therefore they are not able to benefit similarly from using anti-plagiarism software.

It is our experience that not only does the use of anti-plagiarism software help in discovering cheating, it certainly also is a great help in investigating and documenting cheating cases. It should be noted that suspicion of cheating at SDU is always followed up by interviewing the student under suspicion. Furthermore, it is our experience that administration and follow-up on the output from the anti plagiarism software (reports from every assignment) can be time consuming and tedious. A robust digital flow is necessary to be able to take full advantage of anti-plagiarism software.

3. Acceptability

Acceptability means that students as well as faculty and external stakeholders have confidence in the assessment, and that it is generally accepted as a good assessment method.

In the original Business Case an important argument was that the students wanted digital assessment.

In the evaluation of the Study and Educational Environment 2013, all full time and part time students have been asked to respond to a questionnaire. 22% of all students responded (Studie og Undervisningsmiljøvurdering, 2013).

"The students' overall evaluation of the virtual study and educational environment is above average, and they have predominantly given a positive response regarding IT-help at the digital assessment" (Evaluation of the Study and Educational Environment, 2013).

It is noteworthy that the student satisfaction with IT-help during exam ranged from 3,49 to 3,72 (based on Faculty) on a scale from 1 to 5. From the survey there is no significant difference between faculty, gender, age or campus.

Furthermore, a subgroup of the students expressed a positive attitude towards digital assessment in the 'comments' field of the survey, whereas there were no negative comments.

Alltogether, the survey indicate that acceptability among students towards digital assessment is high. It is an interesting observation that even though the approach among faculty members to digital assessment is rather diverse, the student satisfaction is at a consistently high level.

At Aarhus University, Denmark a similar workflow based system was introduced. Here a survey was sent to 369 faculty members of which 98 responded. 75% of the faculty members were positive towards the digital assessment solution (Projekt Digital Eksamen, Aarhus University, 2012).

There has been no formal evaluation of faculty acceptability towards digital assessment at SDU. However, it is evident that in the first phases of the project a fraction of faculty members were worried about acceptability and security/technical issues. Among the issues were formulas and sketches drawn by hand, and other things that are not easily digitized. Digital pen, scanners etc. are now in use, and aid in solving these problems. Basically, doing digital assessment encourages a different way of thinking, which may in turn lead to innovation. Technical problems have proven to be much less significant than expected, also leading to a higher acceptability among faculty.

4. Feasibility

Feasibility means that the resources, which are required to conduct the assessment, are present - e.g. staff, examination rooms, IT equipment and time. In the assessment strategy you have to consider the balance between the significance and the costs of each test.

SDUs LMS (Blackboard with supplementary modules) is under continuous development, to meet the demands from the different methods of digital assessment. In January 2014, 21108 students logged on to the LMS, and 22946 assignments were handed in during the month¹. These numbers illustrate how widely used the LMS is for teaching as well as for assessment.

Almost all assignmentst at SDU are registered via the LMS, and in 2013 a total of 75.000 hand-ins (written exams) were registered.

In order to do a cost-benefit analysis, we first need to estimate the direct cost:

Six full-time IT-supporters: 6 * 50.000:	300.000 Euro.
Student employees (TA):	66.000 Euro.
Technical solutions and IT-staff for education software:	730.000 Euro.
Invigilators/supervisors:	500.000 Euro. ²

There are no other direct costs because the exams are conducted in ordinary classroom and the students bring their own device. The estimated cost for the 75.000 assignments is therefore app. 1,596 mill. Euro per year, or 21 Euro per assignment.

It should be noted that the Faculties have additional administrative workload with digital assessment. The cost of this is not estimated here, and is expected to decrease with time.

In addition to the benefits described in the three chapters above, digital assessment gives us the possibility of digital workflow and automation, and a reduction in printed pages (estimated 500.000 pages).

In an article comparing digital assessment at University of Tasmania, Australia and University of Edinburgh, Scotland (Fluck & Mogey, 2013) the described solutions do not incorporate digital workflow, hence the benefits of digital assessment are smaller.

In a project on digital assessment at Aarhus University the cost and benefit of digital assessment with workflow was estimated for app. 100.000 assignments (Lindhardt, Blok & Hansen, 2012).

¹ SDU had in 2013 a total student population of 28729: app. 73% of all student logged on

² According to definitions used in Bencheit, international benchmarking for HEI

In the pre-project 1671 users were involved with 327 censors and they concluded that the cost for 100.000 assignments would be 0,82 mill. Euro per year. The conclusion was that the main benefit was gained from automation of the workload of administrative faculty members. The Digital Assessment Project at Aarhus University incorporates workflow, and is therefore comparable to the one at SDU.

At SDU we have not estimated the time spent on administration before and after digital assessment, but because the 2 universities are similar in organization, technical setup and competence, we find the estimate to be realistic. Although SDU conduct only 75.000 assignments, and therefore only $\frac{3}{4}$ of the numbers estimated by Aarhus University.

In comparison, Aarhus University cost per assignment was estimated to 8,2 Euro, and SDU cost is estimated to 21 Euro.

The difference is partly due to differences in how and where to allocate the cost. For instance, the cost for invigilators is at SDU allocated to digital assessment, but not at Aarhus University. This alone would reduce the SDU cost by 32%. The setup for both SDU and Aarhus University requires invigilators to be present in the exam rooms.

From the benchmarking we can only conclude that implementing digital assessment has not yet reached return of investment at SDU. At best it has re-allocated the cost when conducting digital assessment

5. Conclusion

In conclusion, we may assume that digital assessment have provided improved opportunities in terms of:

- Digital assessment helps us to test what we want to test by offering new opportunities and tools in search of increased validity and reliability. First and foremost, the students work with their computer when they study, and now they can do so at assessment as well. It is therefore potentially possible to obtain a larger degree of alignment between teaching and assessment. However, difficulties with digitizing certain handwritten things, the open Internet problem a.o. have not yet been solved. The potential of digital assessment can still be developed and exploited better, and will be in the coming years.
- The acceptability of digital assessment is high among students, but possibly lower among faculty members. It can be assumed that faculty member acceptability will increase as digital assessment is developed more.
- Improved possibilities for testing the reliability of a digital test exam, and thereby improved possibilities for developing exams towards better reliability
- The resources have been present, and the goal of feasibility of digital assessment has been partly achieved, due to automation of workflow and due to students bringing their own computer (BYOD). We expect that we in the coming years will be able to diversify and re-allocate the costs, and thereby achieve higher feasibility.

6. REFERENCES

Evaluation of the Study and Educational Environment (English summary) (2013). Retrieved March 13, 2014 from <http://static.sdu.dk/mediafiles//F/9/B/%7BF9B7F963-D501-4E38-9911->

[7979A1D0607A%7DEnglish%20summary%20-%20The%20Evaluation%20of%20the%20Study%20and%20Educational%20Environments%202013.pdf](http://www.digst.dk/~media/Afsluttede%20projekter/100%20procent%20digitalisering%20af%20skriftlige%20opgaver%20ved%20universitetet/Moderniseringsstyrelsen_final_m_bilag.pdf)

Fluck & Mogej, Comparison of Institutional Innovation: Two Universities' Nurturing of Computer-based Examinations, 2013. X World Conference on Computers in Education, Torun, Poland, 2013.

Lindhardt, M.W.; Blok, R. and Hansen, H.J. (2012) Overgang til Digital Aflevering og bedømmelse af skriftlige opgaver (in Danish), p. 33. Retrieved March 13, 2014 from http://www.digst.dk/~media/Afsluttede%20projekter/100%20procent%20digitalisering%20af%20skriftlige%20opgaver%20ved%20universitetet/Moderniseringsstyrelsen_final_m_bilag.pdf

Projekt Digital Eksamen, Aarhus University (2012). Retrieved March 13, 2014 from <http://digitaleksamen.au.dk/projektdigitaleksamen/evalueringer/pilot-4/>

SDU Assignment Stats (2014). Retrieved March 13, 2014 from <http://webhotel.sdu.dk/prod/assignmentstats/>

Studie og Undervisningsmiljøvurdering (in Danish) (2013). Retrieved March 13, 2014 from <http://static.sdu.dk/mediafiles//6/9/C/%7B69CD8934-7E79-45CB-B967-7EDC08A555C2%7DStudie-%20og%20Undervisningsmilj%C3%B8vurdering%202013.pdf>

The Danish eGovernment Strategy 2011-2015 - The Digital Path to Future Welfare (2011). Retrieved March 13, 2014 from http://www.digst.dk/~media/Files/Digitaliseringsstrategi/Engelsk_strategi_tilgaengelig.pdf

7. AUTHORS' BIOGRAPHIES



Kurt Gammelgaard Nielsen

http://www.linkedin.com/profile/view?id=5351271&locale=en_US&trk=tyah2&trkInfo=tas%3AKurt%20Gammelgaard%20Nielsen%3A1-1-1

Lise Petersen

http://www.linkedin.com/profile/view?id=2455863&trk=nav_responsive_tab_profile

Birgitta Wallstedt

http://www.linkedin.com/profile/view?id=21512541&locale=en_US&trk=tyah2&trkInfo=tas%3ABirgitta%20Wallstedt%3A1-1-1

Poul Basse

http://www.linkedin.com/profile/view?id=5359218&locale=en_US&trk=tyah2&trkInfo=tas%3APoul%20Basse%3A1-1-1

Pernille Stenkil Hansen

<http://www.linkedin.com/inbox/#forwardProfileMsg?displayCreate=&profileID=5520661&profileName=Pernille+Stenkil+Hansen&network=I>

Søren Sten Hansen

<http://www.linkedin.com/in/sorenfyn>

Dorthe Majlund Sørensen

Education:

University of Southern Denmark, Master of Arts (MA) In French and Organizational and Cultural Studies, September 2003- May 2009

Employments:

April 2009-december 2011: Copenhagen Business School, Technical Services, Academic Administrative Officer. Resource administration, and Quality Assurance of exam administration.

January 2012 - present: University of Southern Denmark, Faculty of Health Sciences, Academic Administrative Officer. Project manager for digital exams. Best practice in relation to teaching- and exam administration, and study board administration. Consultant on LMS system, BlackBoard.