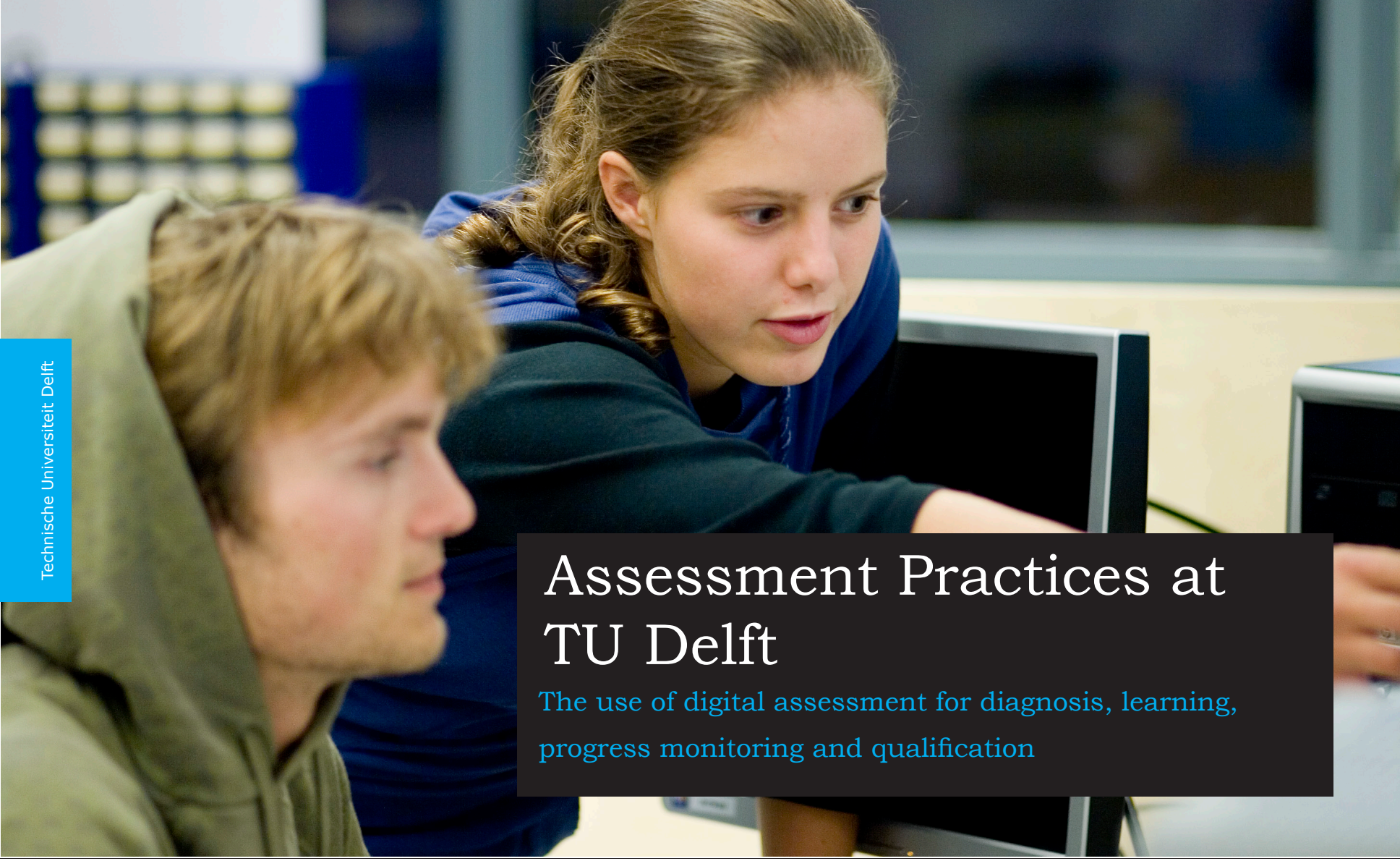

Technische Universiteit Delft

Assessment Practices at TU Delft

The use of digital assessment for diagnosis, learning, progress monitoring and qualification

TU Delft Delft University of Technology

Challenge the future

Commisioned by:
Director SSC - ICT

Editors:
C.G. Huizer
Drs. G.M. Ouwehand
Dr. Ing. A.H.W. van der Zanden

In close cooperation with:
G.N. Bruens
Dr. W.T.M. Caspers
Dr. J.P.R.B. Daudt
Dr. M.C. Cuellar Soares
Ir. P.C. Roling
Drs. P.M.M. Rullmann
Dr. ir. G.N. Saunders-Smits
ir. H.R. Schipper
Drs. M.A.W. Swennen
Dr. ir. J.F.M. Tonino
Drs. M. Weltje-Poldervaart

Author:
Ir. M.A. Keijzer-de Ruijter
Delft University of Technology 2011

INHOUDSOPGAVE

Foreword - by Hans Tonino

As you read this – probably sometime in the autumn of 2011 or perhaps in the spring of 2012 – TU Delft is working to set its teaching on an entirely new footing. The reason for these innovations is the extremely low pass rate of our degree programmes. Given the current economic climate, we can no longer justify the length of time that students take to complete their studies. For this reason, we have recently launched Project Study Success, which is aimed at achieving drastic improvements in the pass rates of our educational programmes.

A successful curriculum includes a good assessment plan

Dr ir Hans Tonino
Director Education
EEMCS Faculty

Assessment is the heart of the eductional process. Digital testing can play an important role.

The question is, 'How can we adapt our teaching in order to achieve a significant increase in the pass rates?' After having had the opportunity to be one of the first to read this booklet, I concluded that it contains valuable and useful ideas for addressing this question. Although its topic is digital testing, it addresses the full breadth of the subject of 'testing' – a subject of great importance. As educators would agree without hesitation, if we wish to achieve true success in our curriculum, we must include a good assessment plan. In short, assessment is the heart of the educational process. Digital testing can play an important role within this process. One example would involve providing timely feedback to students regarding their learning progress.

What could be better for a university of technology than a technological challenge?

In addition to the ambition of addressing the low pass rates, we have at least two other reasons for wanting to focus on digital testing. First, student enrolment is increasing each year, even as our funding from the government is decreasing. Digital testing can help us organise our testing more efficiently. As demonstrated by the pilot projects described in this booklet, we continue to face a number of hurdles. That having been said, what could be better for a university of technology than a technological challenge? I believe that technological universities should take a leading role with regard to these technologies..

TU Delft should take a leading role in these technologies

The second reason for our focus on digital testing involves TU Delft's ambition to offer fully digitalised degree programmes. Within the context of OpenCourseWare, such a programme would include the ability to take tests digitally, even from remote locations. This is another challenge with which TU Delft can strengthen its international position.

What does this booklet ultimately have to offer?

First, it provides a nice overview of the various functions of testing: diagnosis, learning tool, monitoring progress and qualification. This theoretical framework provides the reader with the appropriate tools for classifying the presented cases or pilot projects within the broader landscape of testing. Following this framework, the various cases are presented in a stylised and playful manner. For example, the text is interspersed with quotations from the lecturers involved. It is a fun way to excite the reader. The cases show that TU Delft possesses great potential in its lecturers, who are willing to experiment with new teaching and testing techniques in order to improve their own teaching. We must obviously not forget that all of this has been made possible by OC Focus, the technical support provided by the ICT department and the practical support provided by E-Learning Support (ELS).

I advise you to start reading and get inspired.

Dr. ir. J.F.M. (Hans) Tonino

1 Challenges in our teaching

The number of students at TU Delft has increased greatly. Scenario calculations show that these numbers will continue to increase in the coming years (Graafland, 2010). This poses organisational challenges, including overcrowded classrooms, high teaching workloads and tight teaching schedules. In addition to the increased student enrolment, the long average duration of studies for TU Delft students is placing a heavy burden on available resources. Promoting academic success is an important priority. One of the ways in which our university is meeting this challenge is through the effective use of information technology (ICT). For example, the digital learning environment Blackboard is being used extensively for communication between lecturers and students, clickers are being used to hold the attention of students in the classroom and recording and publishing lectures in Collegerama (http://collegerama.nl) is offering the opportunity to attend lectures outside of the classroom.

Overcrowded classrooms and lengthy duration of studies are posing important challenges to our university

Drs. Paul Rullmann
Vice president Education
Executive Board

Assessment is crucial. As learning materials are becoming freely available, assessment will be a critical factor for success.

TU Delft is considering a package of measures supplementary to the 'Bachelor-before-Master' and the 'binding recommendation on continuation of studies' (Brakels, 2011). These measures are intended to ensure that:

- **The right** student enrols in the right place.
- **The teaching** encourages the student to keep up with the required reading.
- **The lecturer** is encouraged to use ICT in order to optimise the learning process.

An improved study-load distribution, more frequent tests and proper coordination between courses are likely to prevent procrastination amongst the students.

The ICT office and the Education and Student Affairs (E&SA) office have learned from experience that digital testing can contribute in several ways to meeting the previously mentioned challenges. In addition to the use of various testing systems, the Shared Service Centre-ICT (SSC-ICT) offers extensive technical assistance, educational support and hands on support in the field of digital testing. In addition, E&SA and the various faculties have begun to prepare testing policies, in which various aspects of digital testing will be included.

The effective use of ICT is important when addressing the practical challenges posed by the large number of students and promoting academic success

The E&SA and ICT offices are cooperating with the faculties to develop testing policies in which digital testing will play a prominent role. This booklet is a part of this process. It is about both testing and digitisation. It illustrates the fact that we at TU Delft take the digitisation of education seriously. The Executive Board is therefore fully convinced that this is where our future lies. Good lecturers

will always be necessary, and face-to-face contact will always be important. This is the greatest strength of education. These elements must nevertheless be embedded within a context of projects, activities, assignments and tests, of which digital resources will always form an explicit component.

Lecturers will be more explicitly confronted with digital resources in their teaching

With regard to digitisation, TU Delft is considering educational practices within the university, as well as in the broader environment outside of the university. We are working to digitise education through such efforts as OpenCourseWare, refresher and deficiency courses, platforms for international students, modules on iTunes U, virtual lectures with other universities around the world and e-coaching. This is the best way to make our knowledge accessible to many, and it is the best way to anticipate the world of tomorrow. Technology is evolving rapidly, and the social adaptation of new ICT resources is proceeding quickly – especially in the Netherlands. Each new generation of students is more skilled than the previous generations of students or the current lecturers. The sooner we orient our teaching in this direction and become proficient in it, the better equipped we will be to train new generations of students.

Several TU Delft lecturers have now gained experience with digital systems for both testing and peer evaluation. In this booklet, you will be treated to educational practices in which teachers from various faculties have used these systems to address the previously mentioned challenges.

2 Learning differently, assessing differently

The current high-tech society is characterised by information and innovation (Castells, 1997 and Toffler, 1980). This society is placing different demands on the knowledge and skills of students. In addition to being asked to reproduce the requested knowledge, they are also expected to develop skills, including problem solving, leadership, analysis, synthesis, coaching and presentation. Students must be able to work together in multidisciplinary teams. They should be able to provide constructive feedback on the work of others and consider their own work critically. This requires a different way of learning, for which new educational strategies have been developed and implemented. These strategies assume an active attitude on the part of students, in order to stimulate the development of the skills mentioned above.

2.1 New teaching strategies

Dochy, Segers and De Rijdt (2002) argue that, in order to exploit the benefits of educational innovation to the fullest, changes in educational approach should be accompanied by different methods of assessment. Research by van Rossum, Deijkers and Hamer (1985) endorses this proposition and shows that students do indeed learn differently when they are assessed in a different manner. A high-quality testing structure that is coordinated with all learning objectives is essential to the success of educational innovation.

Students will not start to learn differently until they are assessed in a different manner

Dr. ing Piet van der Zanden
Consultant IT in Education
Shared Service Centre - ICT Education Technology

A mix of learning strategies requires a palette of complementary assessment methods.

Van der Zanden (2008) describes five teaching strategies that are regularly applied within the teaching practices at TU Delft:

- **Project-oriented learning** involves the use of methods and techniques for specific assignments. Following a theoretical explanation of specific methods for design and development, students apply these methods in practice within the context of a practical project. It is important to test both components – theory and practice;
- **Productive learning** involves the production of design activities for authentic situations. Self-reflection and feedback from supervisors and fellow students are valuable testing instruments in this regard;
- **Active learning** involves learning by doing and the application of the student's own thinking capacity. The required reading is offered in such a manner that the student must start working with it. Formative testing provides students with an idea of 'where they stand';
- **Cooperative learning** involves the joint execution of complex assignments. In addition to the substantive assessment of the assignment, it is important to evaluate the group process, including the role and effort of the individual student. Peer evaluation is an important source of information for assessing the group process.

- **Supervised self-study** involves the independent execution of the assignment under the supervision of a coach. Formative testing can be used as a learning tool within this context. The results of the test, accompanied by feedback and references to the required reading, can help the student to learn.

Proper assessment of student performance requires a combination of assessment methods

Van Berkel and Bax (2006) indicate that a combination of assessment methods is necessary in order to assess students in their development as competent professionals. They describe various types of testing that are currently applied in higher education, including written tests, design assignments, essays, group assignments and internships.

2.2 Various functions of testing

Van Berkel en Bax (2006) argue that tests generally include all activities that are aimed at making a decision about students. These decisions may be either summative (pass or fail) or formative (guiding). In this regard, they distinguish three periods within a course during which decisions are made. Assessment has a different purpose within each of these periods, as reflected in the table below.

The various means of testing are explained in more detail further on in this section.

	Means	Timing	Purpose
A	Diagnosis	Pre course	Categorisation
B	Learning tool	During the course	Guiding the learning process
C	Monitoring	During the course	Adjusting the teaching process
D	Qualification	Post course	Certification

Table 1:The purpose of testing varies throughout the course

A Assessment as a Diagnostic Tool

The introduction of the Bachelor-Master structure makes it easier for students to take courses in other departments or institutions. Considering the increasing enrolment of international students, it is becoming more difficult to take the prior knowledge of students as given.

It is becoming more difficult to take the prior knowledge of students as given, due to the wide variation in the backgrounds of students who enrol

Starting-level diagnoses provide lecturers with insight into the prior knowledge of their students. A test at the beginning of the course can immediately involve students with the required reading, provided that the test is not considered as inconsequential. For example, the results of the test can be used to form groups or to offer remedial course materials. Mastery of the prerequisite knowledge can even be specified as a condition for the final evaluation in the course.

Drs. Martijn Ouwehand
Consultant IT in Education
SSC - ICT Education Technology

Tools like the Learning Management System, E-portfolio, open access learning materials, mobile apps and digital assessment offer a great potential to design effective and efficient ways of education.

B Assessment as a Learning Tool

One important factor in promoting study success involves ensuring that students remain actively engaged with the required reading throughout the course. Research by Karpicke and Blunt (2011) shows that the process of taking short tests (including self-tests) when studying the required reading is a highly effective method of learning. Test-taking requires students to retrieve the material from memory, thus increasing the long-term learning effect. Recalling the reading material can best be achieved through open questions for which short answers must be formulated. Feedback to the student consists of providing the correct answers.

Regular and repeated testing produces the best learning results

Cumulative testing increases learning results considerably. Roediger and Karpicke (2006) show that the regular testing of all required reading that has been required up to a given point increases the retention of previous material, thus allowing students to focus on the material that they have not yet sufficiently mastered. When testing is used as a learning tool, extensive feedback is not necessary. Reference to the readings that require further study is sufficient. It is important for all topics to be covered in each test, so students know after each test which topics they have mastered (and retained) and which topics they need to put more effort in. To encourage students to take these tests seriously, it is wise to develop a way of including these test scores in the final mark for the course.

The effect of repeated testing on the score on a final test.

In the experiment by Roediger and Karpicke (2006), 4 groups of 40 students studied 40 word pairs, followed by taking a test. This cycle was repeated until all questions were answered correctly. However, the repeat-conditions varied amongst the groups:

The first group (BaTa) studied all 40 word pairs again and took a test on all word pairs. The second group (BnTa) only had to study the word pairs they didn't recall in the test, but were tested over all word pairs. The third group (BaTn) studied all, but were only tested on the ones they didn't recall in previous tests. The final group (BnTn) only had to restudy the word pairs they couldn't recall in the tests en they only got tested on these 'forgotten' word pairs. All students, regardless their group enrollment (and thus study method) recalled all word pairs within the same time for study. All students (regardless their study method) predicted they would score 50% on the final test, taken one week after.

One week after the abovementioned study period, all students took the final test. The first two groups (BaTa en BnTa) scored 80%. While the final two groups (BaTn en BnTn) scored 36% en 32% respectively.

The results demonstrate the critical role of retrieval practice in consolidating learning and show that even university students seem unaware of this fact. Repeated testing on all study materials is a more effient way of studying than repeated studying all materials.

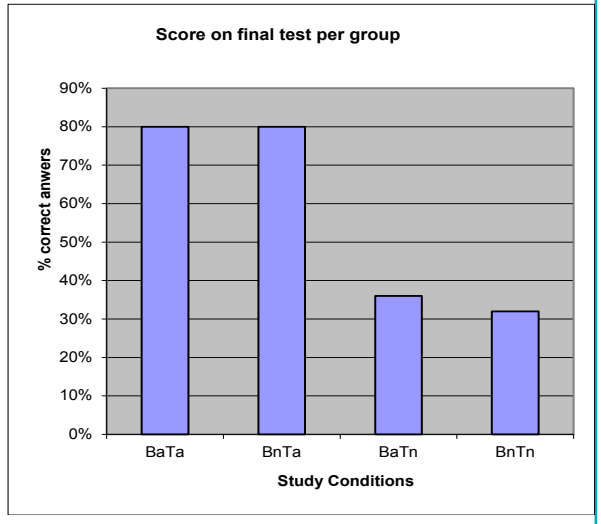

Image 1: The Roediger & Karpicke experiment demonstrates that repeated testing of all the study materials has a significant impact on the score on the final test. (Starreveld, 2011)

C Assessment for Monitoring Progress

By requiring regular homework assignments and continually analysing the results, lecturers gain better insight into the progress of students and the subjects that are perceived as difficult. Lecturers can use this insight in deciding whether to repeat particular topics or to offer them in an alternate manner. It is important to understand the reasoning that students have used in answering the questions in order to identify the point at which they have made mistakes.

Drs Marike Weltje-Poldervaart
Educational consultant
E&SA - OC Focus

Peer evaluation quality depends on the way in which teaching staff introduces this topic to students. Students need to be well informed on the purpose of the evaluation and privacy regulations that are in place.

Regular progress monitoring offers the possibility of making adjustments throughout the course

Feedback on individual questions as well as on the entire test is essential. The manner in which feedback is provided can vary. For example, lecturers could address common mistakes during class or provide automated feedback.

By providing each student with an individual set of problems (parameterisation), lecturers can reduce the likelihood that answers will be copied. Allowing students to discuss amongst themselves how to approach the problems can create a situation in which students explain difficult subjects to each other. Parameterisation is easily achieved through digital testing

Progress in the group process

When students work on assignments in groups, the course of the group process and the individual effort of the group members is the subject of progress monitoring. Peer evaluation is often used for this purpose. Peer evaluation is understood as follows: The provision of process-oriented feedback by students with regard to each other's functioning in an education-related partnership, with the goal of reflecting on their own behaviour and that of others, in addition to learning from the process of collaboration (Weltje-Poldervaart, 2011).

- **Peer evaluation** can be used with the goal of improving (interim evaluation) or reflecting on (final evaluation) collaboration. In all cases, the evaluation should be used as input for discussing the collaborative process (process evaluation). Peer evaluation is thus not intended as a means of final assessment of a student's performance.
- **Peer feedback** We speak of peer feedback when students are required to evaluate and/or provide feedback on the designs, essays or other assignments of fellow students (product evaluation). This instrument is applied with the goal of practising the following learning objective: 'the student provides constructive feedback on the work of fellow students'.

In terms of logistics, peer evaluation and peer feedback are time-consuming instruments. The ICT applications that are available at TU Delft allow lecturers to save considerable time in this regard.

Peer evaluation and peer feedback are time-consuming instruments. The use of ICT saves considerable time

The peer-evaluation tool known as Scorion takes the group structure from Blackboard and links the group to a questionnaire. The evaluation process begins with the press of a button, after which all students receive a questionnaire by e-mail. They complete the

questionnaire online. Reports are generated automatically at the close of the evaluation process: an individual report for the student and group reports for the instructor and/or supervisor.

The peer feedback process makes use of the Self and Peer Assessment Module in Blackboard. This module facilitates the entire process, from the preparation of the assignment (including the model answer and/or the evaluation criteria) through the submission of the project, the distribution of the submitted work across peers, the registration of the feedback or evaluations down to the final evaluation by the lecturer.

D Assessment for Qualification

Most assignments in higher education serve a qualification purpose: instructors use assessment results to determine which students have sufficiently mastered the required readings and which have not. It is essential that the assessment results provide an objective and fair statement about the student's level of knowledge and skills.

Van Berkel and Bax (2006) describe three levels at which aspects of quality should be addressed::

- **Test question** The construction of test questions should conform to requirements in terms of content and form (Teelen, 2004). This is ensured by assessing the test questions before the test.
- **The test** It should be valid and reliable. A test is valid if the questions are relevant and representative, and if the test allows students to see which knowledge and skills they have sufficiently mastered. A test is reliable if the test results accurately reflect the student's knowledge and skills. The use of a good testing matrix provides support in test construction..
- **Test results** A psychometric analysis is conducted after the test. This analysis includes the indicators level of difficulty (P value), discrimination capacity (Rit value) and reliability (Cronbach's α), which are used to identify questions that require further investigation.

Aspects that define test reliability

- ***Objectivity:*** *The test outcome is independent of the person grading the test, the assessment method or test conditions.*
- ***Specificity:*** *Only students that master the learning materials can answer the questions.*
- ***Efficiency:*** *The question texts are clear and its lay-out facilitates easy reading. Students have ample time to finish the test.*
- ***Transparancy:*** *It is clear to the students what topics are being tested and on what level. A short introduction to the test states the amount of questions presented in the time given and the scoring rules.*

The following table displays the various indicators that are calculated in a psychometric analysis. The table shows the range of values, along with the action that should be taken in response to the calculated result..

Indicator	*Range*	*Result*	*Repair*
P-value	*0 to 1*	*Less than or equal to the likelihood of guessing*	*Does the system score the item correctly?* *No, adjust the question or the answer* *Yes, omit the item from the test* *Rerun the analysis*
		Higher than the likelihood of guessing, but clearly less than the desired value	*If R_{it} >0: retain the item* *If R_{it} <0: omit the item from the test and rerun the analysis*
		Close or equal to 1	*Retain the item*
Rit –value	*-1 to 1*	*Lower or equal to 0*	*Does the system score the item correctly?* *No, adjust the question or the answer* *Yes, omit the item from the test* *Rerun the analysis*
Cronbachs α	*0 to 1*	*Lower than 0,70**	*Regard the test results purely as formative*
		Between 0,70 and 0,80	*If the test is part of a series of tests, the outcome is not disturbing.*

Table 2: Summary on the means of repair actions to be taken considering the outcome of the psychometric analysis.

** In practice, values of 0.60 or higher are considered as valid, reliable tests, as difficult questions are purposely included in some cases, in order to distinguish good students from excellent students. This has a negative impact on the Cronbach's α value.*

Increasing the quality of tests by using a testing matrix

The testing matrix is a blueprint for the test in which important characteristics of a test are established, including:

- **Learning objective** the learning objectives upon which the test will be based,
- **Behavioral level** the level at which the learning objectives should be worked out in questions,
- Question s**election** the number of questions include in the test and the distribution of the questions across the learning objectives

The contents of the testing matrix are determined by the learning objective and the behavioural levels to be tested. For example, behavioural levels are determined according to the taxonomy developed by Bloom (1971). In the summary presented below, sample sentences are used to illustrate what is involved in the various behavioural levels:

- **Knowing -** 'I know how a house is built'
- **Understanding -** 'I understand the logical steps that must be taken when building a house''
- **Application -** 'I am capable of building a house on my own'
- **Analysis -** 'I can analyse the way in which a house is built'.'
- **Synthesis -** 'I can provide suggestions to improve this house'
- **Evaluation -** 'I can evaluate whether this house has been well built'.'

If each question in the item bank is tagged with a code for the testing term and the behavioural level, a new test can be compiled easily.

Test Name		*Behavioural level acc. to Bloom*			
Testing term	*Number of questions*	*Know*	*Understand*	*Apply*	*Analyse, Synthetise, Evaluate*
1.1	*6 (10%)*	*4*	*2*	*-*	*-*
1.2	*3 (5%)*	*1*	*1*	*1*	*-*
1.3	*3 (5%)*	*-*	*2*	*1*	*-*
1.4	*3 (5%)*	*2*	*1*	*-*	*-*
	25%				
2.1	*2 (3%)*	*1*	*-*	*1*	*-*
2.2	*3 (5%)*	*2*	*1*	*-*	*-*
2.3	*6 (10%)*	*2*	*3*	*-*	*1*
2.4	*4 (7%)*	*2*	*2*	*-*	*-*
2.5	*6 (10%)*	*1*	*3*	*1*	*1*
	35%				
3.1	*6 (10%)*	*1*	*2*	*2*	*1*
3.2	*12 (20%)*	*4*	*6*	*1*	*1*
3.3	*6 (10%)*	*-*	*2*	*3*	*1*
	40%				
TOTAL	*60*	*20*	*25*	*10*	*5*

Table 3: Sample test matrix

Pitfall

One pitfall associated with using tests to monitor progress or as a learning tool is that students may come to perceive them as inconsequential. If no consequences are attached to taking interim tests, students will have little motivation to take them. Although bonus points are often granted, in practice, many students withdraw before completing the course. Should circumstances force them to miss a portion of the course, such that it is no longer possible for them to earn the bonus point, they are likely to lose their motivation to participate in the other tests.

One method that has been applied by several institutions of higher education in their approach to study success is 'to register = to participate = to pass (in Dutch, Inschrijven = Meedoen = Halen; Van Der Vaart, 2011). This has proven to be an effective method for keeping students involved in the required reading material and encouraging them to participate in tests. Each component of the course (including tests) carries a certain weight in the final mark

The table below provides an example based on the course Language Development offered by the Faculty of Humanities at the University of Amsterdam (Weerman, 2011):

Metha Kamminga
Math teacher &
Trainer MapleTA
CAN-diensten

Students should be able to finisch a digital homework assignment in 15 minutes. This way the work can be planned better with the other study activities.

A series of short quizzes is more effective than one long assignment.

Type of assignment	*Amount*	*Evaluation*	*Weight (in course points)*	*Toelichting*
Homework assignments	*5*	*Individual*	*20*	*Each assignment is worth a maximum of 4 points.*
Research assignment	*1*	*Group*	*15*	
Presentation	*1*	*Group*	*5*	
Interim tests	*4*	*Individual*	*20*	*Each test is worth a maximum of 5 points*
Exam	*1*	*Individual*	*40*	
TOTAL			*100*	

Table 4: Example of a combination of evaluation moments, including (Weerman, 2011)

This construction of the final mark is announced to students at the beginning of the course. Incorporating participation in the interim tests into the final mark in this way has led to an extremely high level of participation (90%, Starreveld, 2011). Success rates for the course have increased as well.

3 Support for Digital Testing

When testing takes place digitally, questions can be easily reused, feedback can be automated and the results can be analysed more quickly. Digital testing has been conducted on a small scale at TU Delft for more than 15 years. Individual lecturers have experimented with various applications. They have selected their own applications and perfected their own working methods. These solutions have not been shared widely, however, and in many cases, departing lecturers do not arrange for transferring these methods to their successors. In addition, experience has shown that disseminating the use of a digital testing system does not happen automatically. For example, the proprietary testing system Etude has few users.

Central support promotes continuity and helps to improve the organisation of technology and the necessary support

In order to encourage digital testing, the offices of ICT and E&SA have launched the project 'Digital Testing Across the Campus'. In collaboration with a group of lecturers, a package evaluation was carried out in order to identify a testing package that would be widely applicable. In this process, MapleTA was selected as the centrally supported package. In addition, central support was provided for the use of the testing modules in Blackboard. Clear choices made at the central level make it possible to improve the organisation of the technology and necessary support, in addition to promoting continuity.

MapleTA and Blackboard: supported by Shared Service Center ICT and E&SA

Though MapleTA is specialized in mathematics, evaluating calculations and formula's, it also supports more common question types like multiple choice, multiple selection, fill-in-the-blanks and hot spot. The extensive ways to randomise questions, built-in test analysis, and the building block to connect MapleTA to Blackboard makes it accessible to teaching staff.

The testing module in Blackboard offers teaching staff an excellent way to gain some experience in the field of digital assessment. This software offers less question types than MapleTA, randomisation options are limited and unfortunately the use of (mathematical) formulas is restricted to question design (a student cannot enter a formula when answering a question). It also lacks an extensive test analysis module. On the other hand, teaching staff and students are accustomed to using Blackboard and its grade Center.

Both MapleTA and Blackboard are available for summative testing in a secure environment to prevent fraud.

Now that the technological infrastructure for digital testing has been realised, various activities have been initiated in order to encourage the use of the systems.

Workshops for using the centrally supported testing systems are organised several times each year. Within various faculty sub-projects, lecturers receive assistance with converting their open questions from 'hard-copies' to digital versions that can be answered and marked online. In collaboration with several other educational institutions, supplementary modules in the area of digital testing are being developed in OpenCourseWare, so that teachers can inform themselves about digital testing anytime and anywhere..

Support for lecturers consists of educational advice by OC Focus, technical support from ICT and hands on support by E-learning Support (ELS)

In the area of peer evaluation, various systems have also been used within our university. In 2009, the decision was made to develop Scorion, an application for the entire university. For a number of courses, this application has since been used with a fixed questionnaire and synchronisation with the group structure from Blackboard. In 2011, the functionality of the system was enhanced with features including the customisation of questionnaires, more compact reporting and a more user-friendly system for peer evaluation with large numbers of groups within a single course. The implementation of this application is supported by E-Learning Support

Campus wide support on peer evaluation: Scorion

Scorion is a webbased application. It accommodates the peer evaluation process of sending a questionnaire to team members regarding student performance within the team. The software monitors the progress of filling out the forms and creates customised reports for both student and teaching staff.

To show how these services provided by the SSC-ICT and E&SA contribute to the promotion of study success, the following chapters use practical situations to illustrate how lecturers have made the transition to digital testing or to an automated workflow for peer review.

4 Testing as a Diagnostic Tool

As is becomes more common that students take courses in other faculties or institutions, their prior knowledge can no longer be assumed equal. A diagnostic test at the beginning of a course helps to activate prior knowledge and to identify gaps. This chapter discusses three different practices, in which testing is used as a diagnostic tool. For example, the entry test in mathematics for first-year undergraduates is intended to reactivate mathematical knowledge acquired through secondary education. The diagnostic test in the Master's degree course LM3751 Transport and Separation identifies deficiencies in students' prior knowledge, thus allowing the programme to be modified where possible. The English Placement Test is administered in order to assign students to various course groups according to their starting levels.

4.1 Quizzes at the click of a mouse

Math entry test and the remedial course Math1000

In 2005 the math entry test was developed in a joint effort of the three technical universities in the Netherlands. At that time there was great concern about the level of math skills gained in secondairy education. Too many students received insufficient math skills.

At TU Delft this test is obligatory to all first-year students as its assesses the required level of math. On a yearly basis 1500 to 1700 students take the test. Students that fail the test are offered a remedial course in math, called Math1000.

Dr Wim Caspers
Lecturer
EEMCS Faculty

By sharing the question bank amongst different lecturers, the creation of weekly assignments in MapleTA is quite simple and efficient. Lecturers can monitor their students work and students tend to keep up more often.

Most faculties consider the entry test in mathematics as a diagnostic tool for the student: 'Which topics have you mastered, and in which topics do you still need considerable practice?' In the faculties of Aerospace Engineering (AE), Applied Science (AS) and Electrical Engineering, Mathematical and Computer Science (EEMCS), passing the entry test is required in order to pass the Calculus course in the first semester. The Math1000 convergence track was developed for this purpose. Each year, about 500 students follow this track. Math1000 offers a large collection of exercises and references to the required reading. The faculties in which passing the entry test is required offer sessions in which students receive explanations and complete exercises, under the supervision of student assistants. These exercises are offered both on paper and through the Maple TA digital testing program.

'Ongoing developments in secondary education have resulted in identifying the gaps in mathematical knowledge act accordingly ', observes Wim Caspers, course coordinator for Math1000. 'For this reason, the convergence track has shifted from teaching essential mathematical knowledge that students are lacking, towards activating the necessary prior knowledge. This has shifted the responsibility for the remedial track more towards the lecturers for the Calculus course.

The use of MapleTA facilitates this process without placing an additional burden on the teacher. When the practice tests and the associated feedback are placed in a shared item bank in MapleTA, the calculus teachers are able to offer additional practice materials to students with little effort.

The practice tests from the convergence track are posted in the shared question bank in MapleTA. Teachers simply activate the tests within their own courses in Blackboard

'Within the EEMCS faculty, several teachers from the calculus course have begun collaborating with the SSC-ICT to develop a shared question bank in MapleTA', relates Wim. 'I was able to join and add all of the exercises and practice tests from Mathematics 1000 to the MapleTA question bank, as well as the old entry tests. The only thing that teachers have to do now is to activate the exercises for their students. The results of the tests are then made available within their own courses in Blackboard'.

The teachers are enthusiastic about offering the practice tests in this automated manner. Because the practise tests are also made available to lecturers who are not yet participating in the project, more lecturers have been exposed to the testing application in MapleTA. Several of them have subsequently started offering their own practise materials in MapleTA.

It is still too early to discontinue the practise sessions supervised by student assistants in Mathematics 1000 completely, because the students are not yet accustomed to being responsible for brushing up on the necessary knowledge all by themselves. It will remain necessary for the facilities to be offered and for the students to be held accountable for them by their teachers.

The entry test consists of multiple-choice questions. Students take this test on paper using response forms. The review task is automated. The response forms are scanned, checked and analysed in the centrally supported application Sonate. MapleTA offers the possibility of digitising the testing process and moving away from multiple-choice questions. Although we would like to do this eventually, it is currently not feasible. Due to the large number of participants computer capacity is insufficient. Entering formulas into the computer also requires some experience. This would be too great a step for a first test (Kamminga, 2011). 'But we certainly would like to do this eventually', adds Wim Caspers.

4.2 Insight into the student's level of knowledge

LM3751 Transport and Separation

This is a mandatory course for all Biochemical Engineering first-year students at the Life Science and Technology faculty. The learning goals of this course are first refreshing prior knowledge on transport phenomena, thermodynamicd and basic concepts of separation technology. :second learning subjects like scaling of process equipment and different aspects of separation of multi component systems.

Dr Maria Cuellar Soares
Lecturer
Faculty of Applied Sciences

Since the testing tool is webbased I can have my students from abroad take a test before they travel to Delft. Hence they know what is expected from them in terms of prior knowledge required. They come better prepared

The intake of students in the course LM3751 Transport and Separation is highly diverse. Students from Leiden and Delft, as well as from other countries participate in this first-year Master's course. Because the course begins by refreshing the knowledge that students are expected to have mastered, a test is given at the beginning of the course in order to provide an indication of the actual level of knowledge amongst the participating students. Based on these test results, certain topics are treated in more or less detail within the course.

Maria Cuellar Soares, lecturer for the course: 'I had my first experience with digital assessment and the MapleTA software during a grassroots project. We scheduled the test during lab time, primarily in order to acquire insight into whether students encountered any problems when taking the test. This was not the case. We also wanted to make sure that everyone took the test. After all, in a pilot, you want to see results'

"Because the test is made available through Blackboard, it is easy to incorporate into the course, nobody has to log in to an additional assessment system. It was also very pleasant to be able to use a variety of question types (numerical, multiple choice, multiple answer and fill-in-the-blank), thus creating a varied test. The possibility of entering extensive feedback was also extremely attractive', Maria adds enthusiastically. 'I currently have a "demo" version of the test, but I'd like to expand it with more topics and a stricter test of the necessary knowledge. I would also like to make better use of the randomisation capabilities.'"

"I definitely see potential providing international students with a better picture of our Master's degree programme. They can take the test in their own countries and receive feedback on the gaps in their knowledge, as well as on the level of English that is expected of them."

4.3 Logistics of the Placement Test in English

Institute for Technology and Communication (TPM)

The English department within the Technology, Policy and Management faculty's Institute for Technology and Communication organises four different courses in English for both PhD's and students.:

- ***English** for Academic Purposes (4 levels)*
- ***Written** English for Technologists (2 levels)*
- ***Spoken** English for Technologists (2 levels)*
- ***Scientific** Writing (1 level)*

Drs Martine Swennen
Lecturer
Faculty of Technology, Policy and Management (TPM)

Converting the placement test into a digitalised version diminished time for review extensively. Furthermore, the integration of MapleTA into Blackboard facilitates logistics in an excellent way.

In addition to courses for staff members, the Institute for Technology and Communication within the Faculty of Technology, Policy and Management (TPM) organises four different English courses for students and PhD candidates. To ensure that they are placed in the appropriate level group, students take a diagnostic test – the English Placement Test – each semester, before the courses begin. Each semester, the test is taken by approximately 200 students and 50 PhD candidates. Participants can take the test at 3 or 4 moments prior to the start of the English courses.

The reason

In the past, the placement test was done on paper. This test consisted of one hundred 'fill-in-the-blank' (semi-open) questions and two short essay questions, in which student were required to formulate their reasons for their programme choices or research topics. The burden associated with marking these tests was enormous. It took four lecturers working full-time for two days in order to mark the tests and report the results to participants in a timely manner. Moreover, the number of participants was increasing.

'From my previous experience with the QuestionMark package, I thought that this test could easily be taken in digital form', shares lecturer Martine Swennen. 'A pilot project of the SSC-ICT offered the possibility of converting the test to a digital version in MapleTA'.

The benefits

'The digitisation of the Placement Test saves us considerable time', reports Martine with satisfaction. 'The hundred fill-in-the-blank questions are now marked automatically, and we no longer have to decipher handwriting for the open questions. It is no longer necessary to divide the task of reviewing student motivation texts amongst the lecturers in advance. As soon as you log in, you can see which tests

have been reviewed and which have not. It is also no longer necessary for lecturers to be present for the entire testing session. We have a very good student assistant who supervises the testing. A lecturer is present to answer questions from participants at the beginning of each session, but once everyone is busy with the test, the student assistant can manage just fine.

Converting the fill-in-the-blank texts into digital form did require some effort. In order to preserve the character of open questions, it was necessary to enter all possible alternative answers for one hundred 'blanks' into the testing system. We went through many old tests in order to collect alternatives'.

Entering questions and answers takes quite some time, but we have already recovered this time

The time needed for marking has been shortened considerably. Now, only the essay questions must be assessed by the lecturers. In addition, it is no longer necessary to divide piles of tests amongst the correctors. Each lecturer can log in to the MapleTA gradebook, evaluate the answers to the essay questions and assign the final score. These final scores are then automatically converted into course codes that are imported into the Grade Centre. A single announcement in Blackboard allows the participants to know that they can find their course codes in the 'My Grades' section in Blackboard

New insights

The digitisation of the test has also introduced another factor: logistics. How and when will participants enrol? Who will gain access to the test, at which point and at what time? How will the results of the test be reported in a simple manner to participants? This logistical process proved an enormous task for the secretary's office at first. Fortunately, it was possible to automate all of these matters using Blackboard.

In combination with Blackboard, MapleTA can manage the entire logistical process of the test: from registration and testing to reporting and communication with students

Participants register themselves for the Placement Test course through 'self-enrolment' in Blackboard. In this area, they can use 'group enrolment' to indicate the testing session in which they will take the Placement Test. Announcements and any changes regarding the test can easily be passed along to the registered participants.

Testing sessions are scheduled in advance in Blackboard using the 'Adaptive Release' function, which opens the MapleTA test at the appropriate time for the appropriate group. Due to the diagnostic nature of the test, no additional security measures are taken. Students have no reason to cheat, given that the test is intended to assign them to the appropriate level.

Areas of concern

Because the testing systems were new for everyone involved in the project (including those from the support departments of ICT and E&SA), start-up mistakes were made and resolved. This cost additional time at first. Other lecturers will benefit when they make the transition to digital testing. E-Learning Support (ELS) has provided regular and patient assistance to this project. This has definitely contributed to the success of this project.

The support of E-Learning Support (ELS) has certainly contributed to the success of this project

'Finally, I would like to compliment the participants, who were extremely patient in waiting for solutions when problems arose' adds Martine Swennen. 'I also have a suggestion: When you are working with colleagues on this type of project, take time to work through the various steps in the system together and to write a manual where necessary, in order to clarify processes and prevent errors. Even if the project is implemented by only one person, it is important to keep colleagues informed of progress. If you wish to implement a new system like MapleTA or a new module in Blackboard, conduct a test in which you try out different options. In some cases, applications may react somewhat differently from what you expect. If you would like to know how a particular process will work for the student, ask E-Learning Support to test it for you'.

5 Testing as a Learning Tool

In our education, project-oriented learning is currently applied regularly as a teaching strategy. In this way, various courses are bundled together under a single topic and treated intensively within a short time. These courses are usually tested with a group assignment, in which a design is created and presented jointly by a group. The summative test is therefore eliminated. It is nevertheless important for students to have mastered the theory. Teachers can encourage students to study the theory by offering interim tests, a portion of which can count towards the final mark.

This chapter tells how this type of block instruction and the associated use of testing as a learning tool has been realised in the course CT4215 Façade Design Plus. The questions in the various tests emphasise important components from the required reading, and they force students to interpret the required reading in practice, such that they keep up with the course and are able to provide a better theoretical foundation for their designs.

5.1 Improved distribution of study load

CT4215 Façade Design Plus

CCT4215 Façade Design Plus is mandatory course for Master students in Building Engineering. Each year approximately 30 students take this course. About 75% of the students have graduated from the preceding Bachelor at TU Delft. The remaining 25% did their Bachelor studies elsewhere both in the Netherlands or abroad.

Two guest lecturers give a series of lectures. The course materials are spread into five modules, each resulting into a test. Furthermore, students need to design a façade. The final grade on this course is calculated based on the grades of the five tests and the design assignment.

The total studyload fort his course is set at 3 ECTS. It is a supplementary course to AR2AE035 Building Design & Engineering (12 ECTS), Which is scheduled in the same period as CT4215. The AR2AE035 course assigns students a major design to be finished by the end of the period. This means study load is heavy near the end.

Lecturer Roel Schipper explains: ‘In order to improve the distribution of the study load across the block, I have chosen to concentrate all of the theory in my course in the first half of the period. All of the theory is treated in the first six weeks, thus leaving plenty of time for the façade design assignment. In addition, the theory that is addressed is directly related to the Architecture course that runs at the same time. This allows the students to apply the theory that has been addressed directly into the building designs that they are required to create in that course’.

Roel Schipper considers it essential to test the theory that is offered: 'If the course is tested only according to the design project, students are likely to start designing with both a blank sheet and a blank mind'.

'In order to prevent procrastination on the part of the students, I have chosen an active working method for my course', explains Roel Schipper. In addition to the lectures, the students must study a reader that includes a variety of scientific articles. Every two weeks, they conclude a module by taking a test that consists of open questions about the articles that have been presented. They are allowed to consult their readers during the test. By asking questions about the texts, students learn to read and comprehend scientific articles.

Ir Roel Schipper - Lecturer Faculty of Civil Engineering and Geosciences (CEG)

The Self and Peer Assessment module in Blackboard allows me to have an open question test without the workload of revising all tests by myself

We have adopted an activating working method in order to prevent procrastination on the part of students

The special feature of the tests is that they are based on open questions that are marked by fellow students instead of by a lecturer. 'For this purpose, I use the Self and Peer Assessment tool in Blackboard', says Roel. 'I have a model answer, evaluation criteria and a scoring model for each question'.

A computer room is reserved for taking and reviewing the test. In the first hour, the students take the test. In the second hour, each student receives two tests taken by fellow students to review. The dual evaluation makes the assessment more reliable. 'If the scores of the two assessments are too far apart, it is a signal that the test in question should be reviewed more closely'. The average score on these five tests counts for 25% of the final mark for the course. This ensures that the students take the tests seriously.

'For the most part, the peer assessment module in Blackboard was quite usable. An initial test with colleagues went fine', relates Roel with enthusiasm. 'After entering the questions, the model answers and the evaluation criteria, you need to specify the phasing of the operations – the period during which the test will be open and when the evaluation by fellow students will take place. You also specify how many "peers" should evaluate the test. Blackboard takes care of the logistics. Blackboard provides me (as a lecturer) with a fine overview of the tests that have been submitted and the status of the "reviews", including the scores'

'Given the answers that students provide to the open questions in the tests, they appear to take the tests seriously. About 90 % of the students participated in all of the tests. That was my goal. I wanted to ensure that the students kept up with the reading while learning something as well. This appears to have been successful!'

Students take the task of reviewing their fellow students very seriously

One major question was whether the students would take the task of assessing their fellow students seriously. 'This was not an issue', continues Roel, somewhat relieved. 'It obviously helps that they cannot see who took or reviewed the test. In addition, we're talking about graduate students. In general, they are genuinely interested in the subject matter. Most of the reviewers made use of the possibility of providing feedback as well as assigning a score. As a lecturer, it obviously remains my responsibility to conduct random checks of the assessments, but I do not have to go through all of the tests'

The Peer Assessment module did, however, have one important disadvantage: Blackboard assumes that every student who is 'enrolled' in the course (whether active or inactive) will participate in the test and the review of the tests. For this reason, several students received blank tests to review in the first testing session. Even worse, a number of completed tests were submitted to inactive students for review. For this reason, a number of tests could not be reviewed, and it was necessary for the lecturer to evaluate them retrospectively, as these students were not actively involved. 'The workaround for this was difficult'.

'This issue was reported to Blackboard through E-Learning Support, and it has since been remedied, such that we can use the same arrangement next academic year, but without the workaround. It is actually very easy to transfer this arrangement to the new academic year', explains Roel.

Finally, Roel notes, 'Next year, I will be teaching a Bachelor's degree course to about 300 first-year students. I would also like to use interim tests to keep these students on track, but I think that the use of peer assessment would be asking too much in this context. Interim digital tests with closed questions and built-in feedback with Maple TA could offer a solution in this regard'.

6 Testing to monitor progress

By offering regular homework assignments and analysing the results, lecturers gain better insight into the progress of students and the topics that the students perceived as difficult. Lecturers can use this insight to decide whether to repeat particular materials or to offer them in another manner.

In this case, it is important to follow the reasoning that the student used when answering questions, thereby identifying the step in which the student made a mistake. In this case, it is essential to provide feedback on individual questions as well on the entire test. The manner in which feedback is provided can vary. For example, the lecturer can address common mistakes during contact hours, while also providing automated feedback.

Good monitoring takes time. This section tells how a number of lecturers in the EEMCS Faculty collaborate intensively in order to distribute tasks. Together, they are filling an item bank with mathematics questions and using it to compile interim tests that may be used by other lecturers as well. The AE Faculty uses peer assessment consistently for the group assignments in all years of the Bachelor's degree programme. The questionnaire is completed twice during the academic session. The automation of the peer review process saves lecturers and supervisors considerable time, and it provides good insight into the progress of the evaluation.

6.1 A shared question bank

Calculus is a first-year course in all of the Bachelor's degree programmes at TU Delft. This course is taught by various lecturers from the EEMCS Faculty and is provided to the other faculties. The required level varies by faculty, but the subjects overlap. In the calculus course, it is important for students to practice a lot. Preparing and reviewing practice tests comprise an important part of the tasks of lecturers. The overlap in curriculum provides opportunities for collaboration. The introduction of MapleTA ensured that this cooperation was actually realised.

Mathematics education has a long history of collaboration between lecturers, even at national level with the National Knowledge Bank of Basic Skills in Mathematics (in Dutch, Nationale Kennisbank Basisvaardigheden Wiskunde or NKBW). The focus of this collaboration was often on the level of mathematics in secondary education in transition to higher education. 'In their own teaching, however, our lecturers had never collaborated in order to establish an item bank to which everyone contributes items and in which they reuse each other's items', relates Joanna Daudt, educational consultant in the EEMCS Faculty and initiator of the project.

Mathematics is an excellent example of a course in which practice is important in order to process the required reading. Lecturers invest considerable time in encouraging their students to practice with the required reading and not to postpone this until immediately before the examination. Preparing and checking homework exercises is time-consuming, especially given the increasing number of students.

It was easy to collaborate in order to create a shared question bank in MapleTA

'When MapleTA became available campus-wide, it soon became apparent that it would be very simple to collaborate in order to create an item bank that could be called up within each course in Blackboard. This offered a fine opportunity to start collaborating efficiently as a faculty, in addition to exchanging assignments digitally and compiling and re-using tests'.

'In order to make it possible to exchange questions between various national and international projects (for example Mumie and Telmme), we have chosen to organise the item bank according to the prevailing international mathematics taxonomy', remarks Joanna. 'In addition, users soon started requesting the ability to search for exercises according to the chapter structure of the widely used calculus books: Adams, Edwards and Stewart. This was also not a problem'

Dr Joanna Daudt
Educational consultant & Coördinator IT in Education Faculty of Electrical Engineering, Mathematics and Computer Science

Let's use IT to stimulate students to study more and relieve the workload of teaching staff.

Test quality and editing

Lecturers have many off the shelf exercises. These could obviously be entered first. The most important aspect is to make good agreements regarding the quality of questions to achieve uniformity in the formulation of questions. At present, two groups of lecturers who are teaching similar courses in different faculties are engaged in collecting and classifying questions. Eventually, a general editorial board will grant approval to add exercises to the item bank. The actual task of entering the exercises into the item bank (Math Question Bank) is performed by student assistants.

It is nice to see that this project is encouraging lecturers to discuss their courses with each other

Joanna remarks, 'It is nice to see that this project is encouraging lecturers to discuss their courses with one another: the order of the required reading, the difficulty level of exercises, the structure of interim tests and their experiences with encouraging students to keep up with the required reading. In addition to preparing individual exercises, they are preparing interim tests that can easily be used by everyone. They are also making agreements with regard to experimenting with various forms of interim tests with the goal of arriving at a best practice. The practice tests for the first weeks of the new academic year have already been prepared. All that is left to do is to activate them at the beginning of the course period. This leaves lecturers with plenty of time to focus on analysing the test results. They can use the students' results to identify the topics that should be explained in more detail'.

MapleTA versus Mumie and Telmme

Mumie en Telmme are projects in which teaching staff work together to create math questions, but serving a different purpose: In the Mumie project a self study course is developed in an open source tool: It combines instruction materials, simulations and exercises provided with extensive feedback.

In the Telmme project adaptive tests are developed in Mathdocs. It contains refences to course materials and exercises.

Entering self-study materials into either Mumie or Telmme requires considerable development and programming work. The nice thing about MapleTA is that it is relatively simple to enter exercises. Feedback can be limited to the correct answer, a reference to the required reading or even the step by step solution method. This makes MapleTA particularly well suited for offering practice materials to students.

6.2 Insight into the group process with peer evaluation

Dr. ir. ir. Gillian Saunders-Smits, Lecturer
Faculty of Aerospace Engineering

Peer evaluation can be the lubricant to good teamwork in project assignments.

The Faculty of Aerospace Engineering uses peer evaluation consistently for five major project assignments in all years of the Bachelor's degree programme. During the peer evaluation, students assess each other's contributions to the group assignment according to a questionnaire. The completed questionnaires provide input to the lecturer or group mentor with which they can open the group process or the efforts of individual group members to discussion (Van den Bogaard & Saunders-Smits, 2007). The questionnaire is completed twice during the project period. This offers students the possibility of modifying their behaviour according to the feedback from their peers and possible intervention by the lecturer or supervisor.

Peer evaluation offers lecturers an important tool for discussing the group process with students

According to lecturer Gillian Saunders-Smith, 'We see peer evaluation as an important part of preparation for professional practice, in which colleagues regularly evaluate and are evaluated by each other. In the course of the Bachelor's degree programme, we see that the students are completing the evaluations in increasing detail and adding more comments in order to explain their evaluations'.

'In my course, I have at least 40 project groups with from 8 to 10 students in each group' continues Gillian. 'It is therefore impossible to have a good view of the individual efforts of all students. Peer evaluation helped the mentors and me to compare our own perceptions of the students to the images that students have of each other. That is very valuable, because we are able to discuss the group process according to the peer evaluations'.

Structure of the peer evaluation

Gillian observes, 'I use rubrics in order to make the results of an evaluation comparable to each other (see Box). These rubrics are based on the learning objectives of group work. Within a rubric, an assessment (excellent, good, satisfactory etc.) is linked to a concrete description of the behaviour exhibited by the student. This makes the evaluation more standardised. The literature contains enough good rubrics so that it is not necessary to create them ourselves. For example, the rubrics we use are from the US Air Force Academy.

Rubrics

Rubrics are used to quantify subjective assessments. Rubrics contain a set of criteria and standards connected to the learning goals. It is very helpful in assessing an essay or design.

The assessment process becomes more transparent when using rubrics.

Notes on use

Marike Weltje-Poldervaart, peer evaluation specialist at the Educational Expert Centre Focus: 'The students' motivation to take the questionnaire seriously is crucial. If they are not motivated, there is no value in using this instrument'. As a lecturer, you play an important role in ensuring the quality of peer evaluation that is to be performed. It is essential to introduce the peer evaluation to students in an appropriate manner. The purpose of the evaluation and the method of reporting should be clear to the students. In the projects, the evaluation points and group conferences are scheduled in advance, so that the supervisor can discuss the results of the evaluations with the group in a timely manner. 'We therefore recommend providing the students with information in advance with regard to the process of cooperation and group dynamics, including their relation to professional practice. The structural effort that is made in AE is obviously very nice, as it allows students to see their own development throughout the various projects'

The quality of peer evaluation is only as good as the lecturer who works with it

Peer evaluation can encourage good cooperation as long as the signals that are obtained through the questionnaires are handled and discussed in the group conference. In exceptional cases, a conversation can be held with an individual student. Because the results of peer evaluations are not directly expressed in a mark, students usually do not perceive them as threatening.

The quality of the peer evaluation is only as good as the lecturer who works with it. The lecturer should take it seriously, for example, by reminding students that the evaluation must be completed or especially by providing good and timely feedback on the results. Only then will the students take it seriously.

Gillian continues, ‘In my situation, the use of a software application is essential in the implementation of the peer evaluation. I started with a hard copy. Fortunately, however, my student assistant at that time did not want to deal with it and drafted a small program for the purpose. This application has been adapted by successive student assistants, and we have been using it for a long time, but we will soon be switching to software that will be used and supported throughout the campus. This will make it possible to link the peer evaluation tool to Blackboard in order to generate accounts and enter groups, as well as to use ‘Single Sign On’, which allows students and lecturers to log in with their NetID’

7 Testing as qualification

Tests have traditionally been used for qualifying students: have they achieved the stated learning objectives? Within the teaching strategies that have been adopted, tests are often one component of a combination of assignments that together provide an impression of the student's learning performance.

The quality of test questions and the reliability of the test are essential to the application of testing as a means of qualification. The test should provide a demonstrable distinction between students who perform well and those who perform poorly. This distinction is ultimately used to determine the extent to which students are competent enough to serve as engineers in their fields

The use of digital testing for qualification poses a number of challenges in comparison to tests on paper

ir. Meta Keijzer- de Ruijter
Consultant IC in Education
SSC ICT - Education Technology

We do not use digital assessment to its full potential yet. In order to gain more efficiency, we should be willing to redesign existing questions and tag them correctly to simplify reuse.

Since the assessment of the student's ability is so heavily dependent upon qualifying tests, it is extremely important for the test to be completed under controlled conditions. One must be able to reasonably assume that the test results represent the student's actual ability. Fraud must be excluded. In case of regular paper examinations, procedures exist for this purpose, including the use of examination slips as proof that the test has been submitted, the preparation of two versions of the test (requiring additional effort when reviewing), testing under the supervision of invigilators and identity verification. The digital testing process offers a number of significant efficiency benefits. For example, it makes it extremely simple to generate multiple versions of the test or test questions without increasing the burden of review, as the test scores are revealed immediately. In addition, the Education and Examination Regulations (EER) specify that students should be able to view their results after the examination, with or without feedback. With digital tests, this is easily arranged online. The lecturer decides which information in the system will be released to the student and at what time.

Digital testing, however, also poses a number of challenges. In addition to monitoring the affairs of the examination room, it is necessary to control the computer environment. For example, access to applications, network drives and communication tools is limited. The following two case studies to describe how lecturers have experienced these opportunities and challenges in their teaching practice

Digital testing for qualification does indeed provide efficiency

Case 1 describes the use of digital testing in the course IDE1041 Design Experience. This course is assessed according to a combination of three interim examinations and four design assignments. Giving interim examinations in digital form reduces the burden of marking for lecturers to the evaluation of design projects. The interim examinations are taken using Blackboard, in a secure environment, and the results are known immediately.

Case 2 describes how the AE Faculty conducts digital examinations using MapleTA. The AE Faculty had already gained experience with digital testing using Etude. Before the advent of MapleTA, it was still necessary to make several adjustments with regard to security.

For example, during the examination, the computers in the examination hall are equipped in such a way that only MapleTA and related software (for example a calculator or a statistical program) are available. Other software and communication tools (for example MSN, Twitter and Internet browsers) are blocked.

Before providing further details on these cases, it is important to explain several key aspects of secure testing

7.1 Security for digital tests

It is important to prevent fraud during summative testing. Although the Education and Examination Regulations (EER) obviously include a variety of measures in this regard, digital tests require specific, supplementary measures, as compared to tests taken on paper. This section provides an overview of these measures.

Student verification

In order to prevent identity switching during a test, the invigilator ensures that the person at the computer and the campus card that is presented correspond to the name and student number displayed on the computer screen.

Test monitoring

The test is taken under the supervision of invigilators in order to ensure that only authorised tools are used and that no communication takes place between students. Both the room and the computers on which the examination is taken are equipped for this purpose. Tests can not be available to students outside the hall. This is currently achieved by allowing access to the test only to students who are seated in the examination hall. This strategy, however, does require some time at the beginning of the examination. For this reason, it will soon be possible to take the test in MapleTA only from specific computers according to the IP address.

Allowable tools

For many examinations, tools are allowed, including calculators, software or formula sheets. Offering these tools within the controlled environment allows stricter verification that these are not being used inappropriately. A good example is the calculator. Offering this tool on the computer offers a solution to such problems as programmable calculators. The Respondus LockDown Browser allows only the use of a calculator, a formula sheet in pdf format or an Excel sheet. The Lockdown Browser ensures that the test is presented in full screen and that keyboard shortcuts and Alt-Tab (window switch) or other key combinations are disabled. Students may only leave the key screen by closing the test, after which they are no longer able to return to the same test.

MapleTA also allows other software to be offered as a tool during an examination. The lecturer is responsible for specifying the tools that are to be made available, so that the SSC-ICT can arrange the environment accordingly.

Communication between students

Even if the use of multiple applications is allowed, it is necessary to prevent the use of unauthorised communication possibilities. The secure environment contains provisions in this regard.

When large numbers of students must takea digital test, it is often impossible for all of the students to take the examination at the same time. Consecutive sessions are organised for these situations. In order to prevent test questions and answers from the first session from being passed to subsequent sessions, it is important to ensure that tests cannot be copied. In addition, digital testing systems offer extensive capabilities for generating many different versions of a test. This minimises fraud by copying from neighbours or by passing along questions and answers to participants in a subsequent session.

Technical problems

Both Blackboard and MapleTA are web applications. Taking the test thus requires a connection to the internet. If the testing session is interrupted unexpectedly, the invigilator can provide the student with access to the test once again. In MapleTA, students can then continue working from the point at which they were interrupted. This is because answers are always saved once they have been entered. Although this is also possible in Blackboard, the process is more tedious: the invigilator saves the score from the first session manually and then deletes the session, after which a new attempt is permitted.

In some cases, Blackboard and MapleTA use slightly different security methods, as described in Sections 7.2 and 7.3

	Paper	Digital
Student verification	*Verification of campus card during testing*	*Verification of campus card, student name and student number in the testing screen*
Admission to the examination	*Examination is distributed only to those present*	*Test is available only to students who are present and only from computers in the examination room*
Tools	*Student's own calculator, formula sheet etc*	*Stricter control of tools, because they are provided by the institution. Multiple tools (software) can be made available*
Copying	*Distribute various versions of the test (increases the burden of marking)*	*The randomisation options make it possible for each student to receive a different version of the test without increase the burden of marking*
Communication between students	*Surveillance in the hall*	*Surveillance in the hall and blocking communication functions on the computer*

Table 5: Comparison of fraud measures for examinations on paper and digital testing.

7.2 Case 1: Secure testing with Blackboard

IO1041 Design Experience

IO1041 Design Experience first-year course for bachelor studies in Industrial Design Engineering. It's curriculum is set up around several project assignments each focussing on a different aspect of product design. During the course students take three tests and hand in four project assignments.

On a yearly basis about 400 student attend this course given by three different lecturers.

In the latest BSc review within the IDE Faculty, the courses Modelling Studies, Design History and various guest lectures in the area of emotion and aesthetics were merged into a single course: IDE1041 Design Experience. The new design was aimed to reduce study load by improving the coordination of the various assignments and tests in the individual subjects.

Ger Bruens
Lecturer
Faculty of Industrial Design Engineering

I still don't feel comfortable around the start of a digital exam. Will everything go well?

The course incorporates various assessment methods including design projects and knowledge tests, in order to do justice to the various topics covered in this course and to assess the students with regard to the skills that are to be developed. 'This means that there is a tight schedule. Within a period of 10 weeks, the students take three knowledge tests and create four designs,' reports the course coordinator Ger Bruens. 'With enrolments of four to five hundred students, this places a considerable burden on the lecturers involved. The three knowledge tests are therefore conducted digitally and marked automatically using the testing module in Blackboard. Digital testing also offers us more options for using high-quality colour images'

Digital testing is indispensable for large student enrolments in order to maintain the feasibility of the teaching process

Ger relates, 'The preparation of multiple-choice tests in Blackboard proved relatively simple, but it was necessary for the testing to take place in a shielded environment'. I brought this question to the IT Department. They had implemented a solution for precisely this problem: the Respondus LockDown Browser. This software ensures that students have no access to other applications or internet communication channels during the test. Moreover, the test cannot be started if the LockDown Browser has not been activated by the student.

'In our course, we were the first to work with this type of security; we therefore encountered all of the start-up problems. With the arrival of Windows 7 and now Blackboard 9, these issues have become a thing of the past. Preparing a test nonetheless remains a painstaking task: Blackboard is an American system, in which the date and time format is different. Blackboard also uses the decimal point instead of the decimal comma. Because of this, it is always necessary to have your settings checked by someone else'.

Ger remarks, 'The beginning of a exam session on the computer will always be a "sweaty hands" moment" for me. 'Fortunately, problems are identified quickly and resolved on site by the ICT support personnel. Students don't seem to to be bothered much, even when things did not proceed quite the way they should have. They particularly liked seeing the results of the test immediately'.

The testing process

For each examination, four one-hour sessions are scheduled. Each hour, a group of about 100 students starts the examination. Upon entering the computer room, the identity of the student is verified, and the student is added to the group of 'attendees' in Blackboard. Only these 'attendees' receive one-time access to the test at the pre-set time.

The student follows the login instruction on paper: Start the LockDown Browser and open the test, which is password protected. The student cannot exit the test until all of the questions have been answered. If a question has been left unanswered, Blackboard provides a notification. The student then receives another opportunity to answer this question. As soon as the test has been completed, the student sees the mark.

'This method saves a lot of time, especially for my fellow lecturers. They no longer have to review open questions on paper. Every testing period, they provide me with a new set of questions, including answers and feedback texts. I enter these questions into Blackboard and arrange the testing session. By using the smart setting for scores by question, the test score can be immediately converted to a mark. Digital testing saves us considerable time, which we can now devote to evaluating the design assignments. Because each testing period new questions are added to the "pool" in Blackboard, the number of available questions is steadily increasing. By allowing Blackboard to draw questions from the pool 'at random', we ensure that each student receives a different set of questions. This eliminates the benefits of copying your neighbour's answers'.

Areas for improvement

Blackboard is not a particularly intuitive system: 'I am satisfied with the functionality that Blackboard offers – drawing questions at random, '"adaptive release'" and, of course, automatic marking and posting of marks in the Grade Centre'.

Although it is possible to analyse the quality of digital questions, Blackboard unfortunately does not yet offer this functionality. It would certainly be desirable to be able to carry out such an analysis in order to monitor the quality of a test. The registration of students as they enter could probably be automated, such that scanning the campus card and a visual inspection by the invigilator would be sufficient to grant access to the test'.

7.3 Case 2: Secure testing in Maple TA

AE1100 Exploring Aerospace Engineering en AE1200 Design and Construction

Both courses are mandatory first-year undergraduate courses. Both courses are assessed by a test and a project assignment and several lecturers are involved. Each year about 500 students participate.

AE2205 Experimental Research and Data Analysis

This is a second-year undergraduate course. Three digital tests, consisting of both numerical and multiple choice questions, add up to define the total grade on this subject. Each year about 400 students participate.

AE4415 Airport Design and Operations

This is an elective course in a Masters study within the Control and Operations track. Assessment is done by means of two group assignments and one digital test. Each year about 80 students enroll.

ir Paul Roling
Lecturer & coordinator IT in Education
Faculty of Aerospace Engineering

Taking exams on the computer puts an emphasis on security. Not only should fraude prevention during the exam be watertight, but securing the question bank is also an issue. You don't want all your items to be out in the open. Since you want to reuse them in the future.

For the courses AE1100 Exploring Aerospace Engineering, AE1200 Design and Construction, AE2205 Experimental Research and Data Analysis and AE4415 Airport Design and Operations, examinations are taken in a secure environment using MapleTA. Paul Roling, lecturer and ICTO coordinator, explains, 'In AE, a few lecturers were already involved with digital testing using Etude, the previous campus-wide testing system. We migrated to MapleTA soon after its introduction, and we have collaborated with ICT to invest considerable effort into designing a secure environment for testing'.

Two servers are available for MapleTA. One server is linked to Blackboard and is used for interim tests. The other server is used for taking digital examinations.

A well-secured environment for test taking requires considerable effort

One advantage of having a separate examination server is that students do not have to navigate their ways through Blackboard in order to start the exam. The first question appears immediately after the student logs in. It also reduces the load on the server: Only the students participating in the examination are working on this server. Students who are practising or staff members who are preparing tests work in MapleTA's 'formative' testing system, which is linked to Blackboard.

The examination server is made available in the examination hall through a secure environment that provides access only to the testing system and the allowable tools.

The examination process

In preparation for the examination, the list of students who have registered for the examination is imported into MapleTA. The computer room is set in examination mode by ICT. To take the test, students log on to a computer with a special examination account. It is not possible to take the examination from another location, because only students who have reported to one of the invigilators upon entry receive access to the testing system.

Due to the scarcity of large computer rooms, consecutive examination sessions are scheduled, but randomisation makes the tests so different from one another that passing along answers is of no use.

8 The next step

TU Delft has sufficient facilities that are equipped for taking both summative and formative tests, and these facilities are used in practice. The practices described above show what is currently possible with digital testing. One initial conclusion is that switching to digital testing requires a considerable investment of time on the part of the lecturer. Not only caused by the considerable learning curve, but this also requires converting questions into a digital variant enabling the automation of the marking process. It is also necessary to become familiar with the new working methods and software applications. With proper support from the SSC-ICT, this often goes well. It ultimately results in considerable time savings and efficiency.

Obviously there are many wishes for the future, and these are being intensively addressed at different levels and with various partners.

Functionalities in Blackboard

The item analysis in Blackboard is quite limited. For multiple-choice questions, it is possible to establish the quality of distracters. If a block of questions is used for drawing questions at random (Random Block), however, Blackboard can no longer generate analyses for these questions. We should determine whether we should try to fill this gap in Blackboard or migrate the task to MapleTA.

Security in MapleTA

A specially equipped environment is currently being used for secure testing in MapleTA. For tests that do not require other applications, it would be nice to be able to use some sort of LockDown Browser like with Blackboard. The IT department is discussing this with the supplier. This feature will be added.

With regard to technology, the IT department is very much aware of the areas in which improvement is needed. These insights are definitely attributable to the willingness of lecturers to participate in various pilot projects. We are in close contact with the suppliers of our software and are thus able to bring our wishes and demands to their attention. Within the context of improving study success, digital testing has high priority. In the period 2011-2014, various technological and organisational measures will be taken to optimise and facilitate digital testing.

The exam service

Logistically speaking, digital testing is quite a chore. All stakeholders (E&SA, ICT and Facility Management) must be well informed; the security aspects for both the test and the server must be set correctly and, in many cases, multiple examination sessions must be arranged. It is also necessary to ensure compliance with the Education and Examination Regulations. For this reason, Faculty for Mechanical, Maritime and Materials Engineering has launched a pilot project for establishing an exam service, featuring a test coordinator.

The test coordinator will assist teaching staff in the process of digital testing. The lecturer will always be responsible for preparing questions of good quality and for determining the distinction between those who pass and those who do not. The test coordinator will be responsible for the proper input of the questions into the testing system, the logistics of the testing process and the quality analysis of the results. Although the exact tasks and responsibilities are still being defined in the pilot project, the first reactions to this initiative have been positive. At the end of this pilot project, the possibility of introducing the testing-service concept throughout the campus will be considered, as well as the manner in which it could be realised.

Cock Huizer
Manager
Education Technology

De juiste ondersteuning van het Shared Service Centre - ICT levert op de langere termijn veel tijdswinst en efficientie voor onze docenten. Samen met hen zijn we op weg naar de optimale invulling daarvan.

Examination halls

To date, there is sufficient capacity in the computer rooms to schedule all of the requested digital examinations. However, we expect digital examinations are likely to become much more common. In this case, capacity could soon become a problem. This issue is being addressed along several different routes. In cooperation with ICT and Facilities Management, E&SA is investigating opportunities for establishing an examination hall with a larger capacity, as well as the possibility of including additional computer rooms in the pool of examination halls. Within the consortium of the three Dutch Universities of Technology, 3TU, we are also investigating the possibilities of secure testing on student owned laptops. This would make more large teaching rooms available that currently cannot accommodate digital testing.

Digital testing is not only a current topic of attention at our university; it also occupies a prominent position on the agendas of many institutions

Digital testing is a topic of considerable attention, and not only at our university. SURF, the national foundation in which institutions of higher education collaborate on innovations in ICT for education, is organising professional development workshops and networking events. It supports projects in the area of digital assessment, and it has recently taken the initiative to establish a Special Interest Group on Digital Assessment. Within this group, various institutions of higher education will be able to exchange experiences.

The E-merge consortium, a partnership between TU Delft, Leiden University and The Hague and Leiden universities of applied sciences in the area of innovative ICTO projects, is also devoting attention to digital testing. This consortium is currently developing and implementing a process of professionalisation that will allow lecturers to start working with digital testing in their own teaching. The quality of tests and test questions, item analysis, test security and the conversion of open questions into questions that can be marked digitally are all being addressed. In this process, lecturers are supervised by educators, with practical support provided by student assistants. The supporting materials are available online to all lecturers.

The IT department is intensively involved in the developments and experiences both within and outside the institution, and it is working hard to improve and expand digital testing within our university. In this way, the IT department is providing an efficient working environment for lecturers. We are pleased to work together to build the quality of our education.

The Shared Service Centre-ICT is always available to provide additional information about digital testing.

ir. Meta Keijzer - de Ruijter,
Project Manager for digital assessment

SSC-ICT Education Technology TU Delft

Referenties:

Berkel, H.J.M. van & Bax, A.E (2006) *Toetsen in Hoger Onderwijs*. Bohn Stafleu Van Lochum

Bloom, B.S., Hastings, J.T & Madaus, G.F. (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill

Boogaard, M.E.D., & Saunders-Smits, G.N. (2007), *Peer & Self Evaluations as Means to Improve the Assessment of Project Based Learning*, 37th ASEE/IEEE Frontiers in Education Conference, Milwaukee

Brakels, J. (2011) *Naar een bachelor diploma in vier jaar*. TU Delft, Interne adviesnota. Januari 2011

Castells, M. (1997, second edition, 2004). *The Power of Identity, The Information Age: Economy, Society and Culture Vol. II*. Cambridge, MA; Oxford, UK: Blackwell

Dochy, F., Segers, M. & De Rijdt, C. (2002). *Nieuwe ontwikkelingen: de assessmentcultuur in Assessment in Onderwijs*. LEMMA BV, Utrecht

Graafland, N. (2010) *Capaciteitsplanning TU Delft 2010 - 2020.* Interne notitie.

Kamminga, M. (2011), *De Tien Valkuilen van Digitaal Toetsen*. Euclides 86-5, p. 207

Karpicke, J.D. & Blunt, J.R. (2011), Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. *Science, 331, 772-775*

Karpicke, J.D. & Roediger, H.L. III (2008), *The Critical Importance of Retrieval for Learning*. Science, 319, 966-968

Ouwehand, G.M. en Van Der Zanden A.W., *Blackboard Portfolio in uw Onderwijspraktijk.* TU Delft 2009

Rossum, E.J. van, Deijkers, R. & Hamer, R. (1985). *Student's learning conceptions and their interpretation of significant educational concepts.* Higher *Education*, 19, 617-641

Starreveld, P. (2011). *Presentatie 'Digitaal Toetsen en Eindtermen' tijdens symposium 'Studiesucces'* UvA, januari 2011 (link: http://www.communities.uva.nl/)

Teelen, *Toetsontwikkeling in de Praktijk. Hoe maak ik goede vragen en toetsen.* (2004). Teelen

Toffler, A. (1980). *The Third Wave.* Bantam Books

Vaart, R. van der. (2011). *Presentatie "Het Utrechtse Onderwijsmodel'* tijdens symposium 'Studiesucces'. *UvA, januari 2011* (link: http://webcolleges.uva.nl/mediasite/Viewer/?peid=41c7e19b241e4c5a904f1320a1a275fa1d

Weerman, F. (2011). *Presentatie 'Toetsen en Motiverende Werkvormen'* tijdens symposium 'Studiesucces' *UvA, januari 2011* (link: http://webcolleges.uva.nl/mediasite/Viewer/?peid=72f00dd4548b489ba924887a4e23e2bd1d)

Weltje-Poldervaart, M. (2011). *Do's and don'ts bij het inzetten van Peer Evaluation voor groepen studenten*. Intern document. TU Delft

Zanden, A. H. W. Van Der, (2008). *ICT in het Onderwijs 2008. Onderzoek naar opbrengsten van ICT voor de onderwijspraktijk bij de TU Delft op basis van de mening van docenten*. TU Delft., Delft.

ICT in Education	http://icto.tudelft.nl
OC Focus	http://ocfocus.tudelft.nl
E-Learning Support	http://els.tudelft.nl
Educational consulting	http://soc.tudelft.nl

SSC – Education & Student Affairs
SSC – ICT

Landbergstraat 15
2628 CE Delft

www.OenS.tudelft.nl
www.ssc-ict.tudelft.nl