

Testing and Assessment Manual

Established by the Faculty Board on 21 March 2006, having consulted the Programme Committees and the Joint Assembly.



Content

Introduction	2
1. The basic requirements of testing	3
1.1 General quality requirements.....	3
Validity	3
Reliability	3
Transparency	3
Usefulness.....	4
1.2 Forms of testing	4
1.3 Content-related testing.....	5
Validity	5
Reliability	5
Transparency	6
Usefulness.....	6
Written examinations.....	6
Oral examinations.....	8
1.4 Skills-based testing	9
Validity	9
Reliability	9
Transparency	9
Usefulness.....	9
Written assignments	10
Presentations.....	10
Theses	10
2. The basic requirements of assessment.....	11
2.1 General quality requirements.....	11
Determining the standard.....	11
Peer testing	11
Academic misconduct.....	11
The Examination Board.....	12
2.2 Content-related testing.....	12
Written examinations.....	12
Oral examinations.....	13
2.3 Skills-based testing	13
Written assignments	13
Presentations	14
Theses	14
Bibliography.....	15

Introduction

The aim of the *Testing and Assessment Manual* is to give an insight into the quality requirements set for testing and assessment within the Faculty of Social Sciences. This document contains the rules as they are set out in the faculty regulations, in addition to recommendations for designing various types of tests and for correcting and assessing tests. The document deals with specific problems relating to the testing and assessment of knowledge and skills, and gives practical tips and recommendations for various types of tests, as well as examining practical issues with regard to testing.

This document forms part of the Quality Assurance Handbook (<http://www.intranet.vu.nl>, see: other/reference works/regulations) which forms part of the system of internal quality control in education throughout VU Amsterdam. In the internal quality control system, both internal and external incentives have an important part to play. External incentives include the Dutch-Flemish Accreditation Organization NVAO, the Ministry, the Executive Board and student expectations. Internal incentives include the opportunity to recycle knowledge, traceable working methods (accountability) and transparency. In this regard, everything is geared towards ensuring that an organization displays predictable and structured behaviour.

The *Testing and Assessment Manual* is primarily intended as a practical guide for the teaching staff at the Faculty of Social Sciences. It is important for students to be aware of the rules with regard to testing and assessment. Chapter 1 deals with the basic requirements of testing. The second chapter deals with the basic requirements of assessment. Both chapters start by addressing the general quality requirements. They then go on to make a distinction between various forms of testing and the various ways in which the quality requirements apply to them.

Much of what is described is not new, but an account of existing practice. This document places the rules as they apply at the Faculty of Social Sciences within a framework, with a reference to the regulations in which the rule can be found (see also: <http://student.fsw.vu.nl/faculteit/regelingen/>). The terms 'quality requirements' or 'basic requirements' refer to the content-related requirements which underlie a sound approach to testing and assessment. They do not constitute fixed rules. Practical tips and recommendations from the Quality Assurance Handbook and other sources have been incorporated in the text, often in the form of a summary.

1. The basic requirements of testing

1.1 General quality requirements

In general, the concept of testing encompasses the entire process of designing, administering and marking tests as part of the curriculum. The aim of testing is to contribute to the student's learning process. This contribution can be selective, in that it assesses whether the student has mastered what he or she is supposed to master as stated in the final attainment levels and learning objectives, or diagnostic, in that it points out the areas in which a student is weaker and in need of additional guidance. Lastly, a test can have a predictive function with regard to the continuation of the student's learning process and (study)career. The curriculum as a whole constitutes a mix of these functions. In this text, the term 'test' is not synonymous with 'examination' but also takes in other forms of testing, such as written assignments and oral presentations. Whatever its function, a test is an instrument of measurement which can only provide useful information if a number of basic quality requirements are met:

- *validity*: the instrument must achieve its purpose and measure what it is designed to measure;
- *reliability*: it must be possible to base decisions on the student's score;
- *transparency*: the student should know what is expected of him or her;
- *usefulness*: the type of test should be appropriate to the course material, the learning objectives and the size of the group. The test should be realistic and relevant with regard to the financial and/or material framework, but also in terms of the time available.

The aim is to achieve the best possible results in relation to the above requirements so that the measuring instrument can say something genuine about the student's progress and his/her current situation as regards the learning process. A good measuring instrument ensures that students do not fail or pass without good reason, but also ensures that students are given a realistic impression of their learning process.

Validity

The value of a test result is based on the premise that the test is a relevant and representative spot check and inspection in relation to the learning objectives. Equipped with a valid test or form of testing, the teacher should be capable of making a well-founded statement about the extent to which the educational objectives have been met by the student.

Reliability

The marking of a test must be defensible and thorough and should do justice to the person being assessed. Doubts may arise regarding the reliability of a test if the mark is based on too few measurements, if too few components are tested or if the test questions are not formulated precisely enough.

Transparency

Transparency means that it should be clear to the student what is expected of him or her when being tested. This means that clear information should be given about the material to be tested, the timing of the test and the type of test. Relevant aspects include clear communication about the literature and the lecture material that form the basis of the

test, when an exam is to be held, the type of questions to be asked (open questions or multiple choice) and the extent to which any written assignments must be passed or count towards the final mark.

Usefulness

Different forms of testing vary in terms of their usefulness, that is to say in the extent to which they fulfil intended functions. Organizational and procedural aspects are particularly important where usefulness is concerned. The primary issue in this respect is whether the effort required of both the lecturer and the student for a given test is in proportion to the amount of information it generates.

1.2 Forms of testing

It is possible to distinguish between various forms of testing. Ideally, when selecting a certain form of testing, the first concern should be the criterion of *validity*, followed by *reliability* and finally *usefulness*, not the other way round. The aim of the test should also be clear. In this respect a general distinction can be made between content-related testing on the one hand, which mainly tests knowledge, and skills-related testing on the other hand, which is concerned with skills such as writing, associative thinking and conducting analyses. If the aim is to test a student's knowledge, it would be more appropriate to use a written or oral exam rather than a written assignment such as an essay question. By the same token, a written or oral exam is a less appropriate way of testing writing skills, while getting students to write a paper would be far more effective in this respect. There are a number of forms of testing available:

- written examinations
 - multiple choice / closed questions
 - questions inviting a short answer
 - essay questions / issue-related questions
- oral examinations
- written assignments
 - essays
 - papers
 - reviews
 - paraphrasing
- presentations
- theses

Written and oral exams are suitable for testing knowledge. Certain skills can be tested by means of written assignments, presentations and theses. Both types of test call for different sets of quality requirements.

A great many misconceptions exist with regard to forms of testing. One such misconception is that knowledge can only be tested using closed questions. Research has shown that several forms of testing are suited to testing the same cognitive skill. A student who scores well in an exam with open questions also scores well in an exam with closed questions. Students often feel that a test with open questions does more justice to their preparations for a test. Various experiments have been conducted to study this. Students who had prepared for a test with closed questions were given a test with open questions and vice versa. The results were then compared to the results of

students who sat the type of test they had prepared for. No difference in the scores was found (Van Berkel, 1999: p.21 onwards).

The reason behind choosing a certain form of testing is not just a matter of general quality requirements and the aim of the test. Arguments regarding the efficiency, representativeness and objectivity of a certain form of testing are also important. A given form of testing is often selected based on arguments of efficiency: the number of students taking the test and the amount of correction work generated, or the amount of time available, for example to administer oral exams. With a view to the representativeness of a test, the number of questions plays an important part: the more questions you ask, the more representative a check it is. The only way of asking a large number of questions in a short time is to ask closed questions. The number of questions that can be asked in an exam is necessarily limited by the time available. Lastly, objectivity when marking open questions or written assignments can be enhanced by using multiple assessors and by working with a list of predetermined criteria that the assignment has to meet.

Whatever the form of testing used, the test has to establish the extent to which the student has mastered the final attainment levels/learning objectives of the subject or educational component. These final attainment levels form part of the subject descriptions in the digital study guide, but they often need to be described in greater detail in order to develop test questions. In order to make learning objectives more tangible, inspiration can be drawn from content-related aspects such as the course material, the likely field of future employment (e.g. formulating questions relating to application) and class discussions between students and teaching staff.

1.3 Content-related testing

Validity

In a written or oral examination, it is important that the questions accurately reflect the material taught, that all important topics are equally represented in the exam and that the exam measures the intended level of ability. Exams should therefore be valid both in terms of the content of the teaching (*content validity*) as the command of skills (*comprehension validity*). The following recommendations apply when compiling an exam:

- the exam should be a relevant and representative check of the material covered by the exam
- the questions should be distributed across all areas of the set material
- the number of questions should be determined on the basis of the number of topics to be tested and the time available.

This means that if the aim of the subject is to provide a broad introduction to a certain discipline (e.g. an introduction to the social sciences), it is probable that a wide range of topics will be covered. Taking into account the representativeness and the time available for a written exam, the obvious form of testing would be a multiple choice exam.

Reliability

With regard to open questions, it is difficult to determine whether a sufficient level of reliability has been achieved, particularly in comparison with closed forms of testing, such as multiple choice questions. A test can be valid in the sense that it asks questions

about all the important subject matter, but may still be unreliable if the questions are not formulated precisely enough. The latter, of course, also applies to multiple choice questions. Reliability can only be guaranteed if the questions are phrased carefully and unambiguously.

Transparency

In order to achieve transparency, it is important that several exam sittings for one and the same subject (in the same academic year) are comparable, for example by making use of parallel questions (differently worded questions that measure the same thing). It is also important to provide students with clear information on the type of exam and questions they can expect, so that they can prepare properly.

Usefulness

An exam that consists of essay questions is usually limited in scope given the amount of time available. If the aim is to measure the level of knowledge, an exam of this kind does not usually meet the criterion of content validity. Another important factor in this regard is the size of the group: oral exams are not the preferred form of testing for large groups of students. However, it should be noted that the extra time needed to formulate closed exam questions only starts to pay off for groups of 50 or more students.

When it comes to feasibility, it is worthwhile asking whether the exam is realistic and appropriate within the financial and/or material frameworks, but also within the time available. These demands are often at odds with one another. In the interests of efficiency and cost-savings, a less valid and reliable form of testing is sometimes selected in preference to others. In order to prevent a situation in which selecting a certain form of testing will have a negative effect in relation to one of the criteria, a test can be compiled from several different forms, with component scores that compensate each other. Students generally respond positively to such a form of testing.

Written examinations

The rules that apply before, during and after a written examination can be found in the documents *Regulations for Written Examinations*, *Guidelines for Written Examinations* and *Instructions for Invigilators at Written Examinations*. These documents deal with matters such as registering for an examination, publication of the exam timetable and location, the deployment of invigilators and their role during the exam, academic misconduct (fraud), and periods for correction and inspection rights once the exam is over (right to lodge an appeal and opportunity for feedback).

The general rules are set out in the above-mentioned regulations, while the content-related and subject-specific information is provided in the study guide, on the teaching website and Blackboard. At the very least, information should be provided on the following points:

- the learning objectives;
- the amount of material;
- the form of testing;
- the relationship between the various components, if applicable;
- the standard for a pass;
- the member of teaching staff responsible;
- the scheduled dates for testing and resits;
- the opportunity for inspecting the corrected exam paper.

In addition to providing students with clear information prior to the examination, for example in the digital study guide and on the teaching website and Blackboard, just as much attention should be paid to the way in which exam questions are formulated.

When producing exam questions it is important to consider the following:

- set questions on key topics that are considered important, not on details;
- preclude the possibility of a question being interpreted in several ways by phrasing it carefully;
- preferably work with other teachers to draw up questions;
- state all the information needed to answer the question, for example the literature to which the question relates (e.g. if several articles by one author have been used, it would clarify matters to indicate which of the author's articles the question refers to);
- phrase the questions as concisely and straightforwardly as possible;
- avoid double negatives;
- ensure that the question can measure the final attainment level;
- ensure that the wording of the question matches the course material.

With specific regard to multiple choice questions:

- preferably give 3 or 4 alternative answers in a multiple choice question: the key (the correct answer) and 2 or 3 red herrings;
- make sure that the key is clearly correct or the best answer;
- do not give any hints (e.g. grammatically) in the question;
- make the alternatives more or less the same length to avoid a situation in which one alternative contains more information or details than another;
- avoid repeating the same words in the alternatives;
- make sure the alternatives seem reasonable and plausible (do not include nonsense answers);
- avoid negative questions or emphasize any negative formulations in order to prevent confusion, for example by underlining them;
- avoid vague, relative specifications such as 'usually' and 'often';
- avoid absolute statements: propositions with absolute formulations such as 'always' and 'never' are seldom always correct or always incorrect;
- avoid the alternative 'none of the above': this option usually indicates a lack of inspiration and does not constitute a real alternative;
- choose alternatives which are mutually exclusive;
- organize the answers logically.

With specific regard to open questions:

- make it clear to the student how comprehensive the answer should be (minimum and maximum);
- use a larger number of relatively short open questions rather than one or two long open questions in order to increase representativeness;
- compose a model answer when formulating the question;
- only use the phrase 'give your opinion' if argumentation is required; use a phrase such as 'give reasons for your answer' instead.

In terms of layout and structure, it is important to consider the following:

- the questions should be presented in the same sequence as the teaching material as much as possible;
- the pages should preferably be numbered;
- use a clear layout in which different fonts have a clear purpose.

The information on the front page of the exam paper should preferably include:

- the name of the subject
- date of the exam
- the time the exam starts and finishes
- sitting (i.e. first or second opportunity to sit the exam)
- responsible teacher
- number of pages
- number of questions
- types of question
- instructions regarding the scope and structure of the answer
- the maximum mark per assignment and the pass mark
- information regarding the opportunity for the student to inspect the marked exam paper.

The number of questions to be asked in a written exam is determined by the amount of time available for the exam and the scope of the course material. An exam has less integrity as a measuring instrument if factors such as speed, exam know-how or writing skills have too big an influence, since these factors say nothing about the extent to which a student has mastered the course material. Yet the influence of such factors cannot be ruled out completely and indeed there is nothing wrong with expecting a minimum of skill and competence from students with regard to such matters. After all, a student's language and writing skills also provide information about his or her insight into the material and his or her ability to make links. In order to ensure that students do not write in telegram style and to prevent language errors, it is advisable to state on the front page of the exam paper that answers should be given in legible, correct Dutch and in complete sentences. This makes students aware that language and stylistic errors can affect their mark.

Oral examinations

Oral exams do not form a standard part of the educational programme, which means that extra care must be taken with regard to this type of testing. As a rule, no more than one student at a time can sit an oral exam and the exam is public unless the Examination Board decides otherwise or the student objects to the public nature of the exam. The student is allowed to invite someone of his or her own choice to be present at the exam, if this helps him or her create a safe environment. Requirements with regard to validity and reliability also apply in the case of oral exams, of course, but are more difficult to safeguard, as the nature of this type of testing is often more informal. For this reason it is all the more important to formalise the situation where possible. For example, it is advisable to have a second teacher present at the exam as standard procedure.

1.4 Skills-based testing

The testing of skills is a complex matter, especially since this often involves a combination of different skills. The writing of a thesis, for example, is a complicated interaction of writing skills and cognitive skills such as analysis, synthesis and structuring. Skills are usually tested by getting the student to produce a piece of work, such as a thesis, a detailed assignment or an individual or group presentation. Several questions should be asked when marking a student's work:

- Is the student familiar with the rules for writing a paper/giving a presentation?
- What can be said about the student's ability to function in a future work situation?
- Is there evidence of sufficient capacity for abstraction?
- Is there evidence of sufficient creativity?
- Is the student's ability to apply knowledge sufficient?

When testing skills, the following general recommendations apply:

- ensure feedback and opportunities for feedback; without feedback the learning effect is limited
- do not give assignments if they cannot be corrected
- students can give each other feedback but the teacher remains the person who ultimately assesses each student and gives them a mark
- limit the number of assessment criteria
- allow the degree of difficulty of the criteria to increase as the curriculum progresses (the demands made of a first year Bachelor's student should differ from those made of a Master's student).

Validity

In the case of a written assignment or an oral presentation, the assignment should form a logical extension of the subject. It must be possible to write the assignment or prepare the assignment within the time available. In order to ensure that all the relevant aspects are reflected in the piece of work that the student produces, it is of great importance that the criteria for assessment and feedback are clearly formulated.

Reliability

In skills-based testing, the reliability of the test is strongly influenced by such aspects as the number of assessors (subjectivity, interpretation), the number of criteria for assessment and the scope of the test. For example, should the assessor only focus on the final product or also look at the period in which the student worked on the test and therefore how he or she arrived at the final product? These are aspects that should be made clear to the student and teacher in advance.

Transparency

For written products or other skills-based tests, it is important to make clear to the student what is expected of him or her. In this regard, information on the scope of the product, the exact nature of the assignment and the time available is essential. It is preferable to use publicly available assessment and feedback criteria for this purpose.

Usefulness

Skills-based testing usually assesses a piece of work produced by a student. This form of testing is only worthwhile if it is feasible to provide feedback on the products. When

giving feedback, the focus should be on giving reasons why something is good or is not good. If organizational factors preclude this (e.g. if too many students are being taught by a limited number of teachers), it is questionable whether the student's learning process will benefit from a skills-based test.

Written assignments

Written assignments can serve a range of purposes:

- to inspire students to start work as quickly as possible;
- to test the knowledge acquired in practice;
- to ascertain whether students have really understood the course material;
- to practise skills.

Whatever the aim, assignments miss their target if they do not at least have a feedback function. It therefore does not serve any purpose to set assignments if there is no opportunity for correction.

Presentations

Presentations can be tricky in that they are both process and final product at the same time. More than in other forms of testing, the result will be influenced not only by knowledge of the subject but also by personal factors such as fear of speaking in public, mood swings and interest in the subject. It is therefore important to create a safe climate for assessment.

Theses

The four requirements of validity, reliability, transparency and usefulness should also be taken into account when testing takes the form of a thesis. Content validity can be enhanced by giving the student the freedom to choose his or her own thesis topic. Transparency is achieved by making it clear in advance what the objective of the thesis is or which skills are being tested. A thesis generally tests not only knowledge of a certain subject but also cognitive, editorial and creative skills, analytical capacity and, for example in the case of the Master's thesis, insight into and understanding of the methods and techniques of research in the social sciences. In terms of the usefulness of the thesis as a form of testing, it should be noted that getting a student write a thesis is only worthwhile if sufficient supervision is available.

Every programme has its own research training guide and/or thesis guide which states the requirements pertaining to the content and the form of the research training and/or the thesis. (<i>Master's Thesis Regulations</i>)

2. The basic requirements of assessment

2.1 General quality requirements

Determining the standard

There are various methods for drawing up a standard or provisional standard for a pass mark. The main distinction to be made is whether the method is *absolute* or *relative*. The standard for an absolute method is defined in terms of the course material and is therefore fixed in advance. The standard for a relative method depends on the performance of the group. In such a case, the standard is determined after the test. Since relative methods depend on the group (the same score in a relatively good group is worth more in a group that scores relatively badly), they are controversial to say the least.

It is therefore advisable to work with a standard that is determined in advance. The exact nature of a standard will always depend on specific factors but the broad outlines should be established for each curriculum. If the percentage of students who pass is below 50%, the teacher in question should take action to find out why the score was so low. Of course, this does not only involve looking at the course material and the test. Contextual issues of timetabling, coordination with subjects running in parallel and other non-content factors should all be taken into account. It is advisable to compare tests from the same subject in different years with one another in order to determine the degree of difficulty and preferably to obtain the opinion of several teachers in this process.

Peer testing

It is very important that teachers meet on a regular basis to present and discuss both tests and test results. The teaching managers monitor this process and are charged with the task of initiating meetings of this kind. Staff can exchange experiences, problems and best practices at these meetings and novice teachers can receive coaching from more experienced staff.

Academic misconduct

Academic misconduct can be discovered while it is being committed or later on during the assessment process. In written exams, copying the work of other students or taking illegitimate aids into the exam is regarded as academic misconduct. In written assignments, plagiarism is regarded as academic misconduct. Teaching staff in the Faculty of Social Sciences have access to a scanner which they can use to check for plagiarism in cases where testing takes the form of written work.

- The examiner (teacher) will report all instances of academic misconduct to the Examination Board.
- The Examination Board is empowered to impose a penalty of exclusion from one or more examinations during one calendar year on the perpetrator of academic misconduct.

The Examination Board

By law, the formal responsibility for administering tests and exams lies with the Examination Board. The following rules apply to the assessment of tests:

- A final mark will be given for every component completed.
- The assessment will be given in half or whole points, with the exception of marks between five and six. If the final mark is above five but less than six, the result given to the student will be rounded off to a whole number. Final results from 5.01 to 5.49 will be rounded down to a five. Final results from 5.5 to 5.99 will be rounded up to a six.
- In the case of a testamur, consisting of component examinations or several components, all of which have to be completed successfully, the mark for each component will be rounded off and then the final mark will be rounded off.
- In the case of a testamur, which consists of component examinations or several components with constituent marks, the final mark will be rounded off.

(Regulations for the Examination Board)

2.2 Content-related testing

Written examinations

As stated above, there are various methods available to determine the line between a pass and a fail in an exam. In the case of multiple choice questions, it is advisable to take the probability of guessing the right answer into account when setting the pass mark in advance¹. In exams with open questions, too, the absolute score is looked at first before taking the relative context into account. Of course, it is not possible to apply a guesswork percentage to this type of test, but it is possible to determine whether students perform relatively well in a test or in a particular question in a test. Such a conclusion might lead to all of the students' marks being rounded up.

Before students receive their marks, i.e. before they are assessed, the mark first has to be determined. Determining a student's mark, the correction, can cause problems, particularly in the case of open questions. When *correcting*, the test result is determined and a conclusion is reached as to what percentage of the course material the student has mastered. In the *assessment*, the question at hand is whether this result is sufficient. It is up to the teacher to determine what mark has to be achieved in order to achieve a pass. The level of this mark depends on the type of subject or the type of knowledge being tested. In order to prevent assessment effects interfering at the correction stage as much as possible, it is important to draw up as detailed a model answer as possible beforehand accompanied by a value in terms of points. Trial corrections or correcting per question instead of per student can be useful resources in this respect.

When giving a mark, the final step in the assessment process, it is important to have obtained as much certainty as possible about the quality of the measuring instrument used. When processing multiple choice exams, a statistical exam analysis is automatically carried out by the Centre for Educational Training, Assessment and

¹ For example, in a multiple choice exam that consists of 60 questions with four possible answers and a fixed pass mark of 55% (set by the teacher), a student needs to get 55% of 45 questions right (60 minus 15 answers which could have been guesses) in order to achieve a mark of 6. This means that the student has to answer 25 + 15 of the 60 questions correctly.

Research (CETAR). This is seldom the case for exams with open questions. After an exam with open questions, the only points of reference are the comments made by the students or questions which were clearly wrongly interpreted or too difficult (because the mark deviated strongly from the statistical norm). For open questions it is very important to check the quality of the questions in advance (for example by asking a colleague to take a look at the questions). If the pass rate for a test is considerably lower than expected (and in any case if the pass rate is below 50%), it is advisable to carry out both a statistical and content analysis or to have these analyses carried out. As regards content analysis, CETAR can, on request, give indications as to why a certain test item performed particularly poorly from a statistical point of view.

Lastly, feedback is also important in written exams. Before or during the exam, it is advisable to let students know how they can act on their right to inspect their marked exam paper. For example, this can be done by stating the time of a joint exam discussion on the front page of the exam paper or the times at which students can come along to inspect their marked exam paper or to discuss the exam.

Prior to the second opportunity to sit the interim examination, candidates can inspect the questions and assignments given at the first opportunity, as well as the marking criteria used. A written request to this effect should be submitted to the examiner in question. On behalf of the second opportunity to sit the interim examination, a stipulated period for inspection of twenty working days after the final marks are published is applicable.

(Guidelines for Written Examinations)

Oral examinations

When marking oral exams it is important to establish the assessment criteria. An important tool in this respect is a clear assessment form which clearly states the elements the students are expected to include in their answers.

The examiner will inform the student of the result of the oral interim examination immediately after the examination has taken place.

(Regulations for the Examination Board)

2.3 Skills-based testing

Assessing skills-based tests is more complicated than assessing content-related tests, since the assessment is more subjective. It is important that both teacher and student are aware of the assessment criteria for a skills-based test. What standards does a project or a thesis have to meet? What criteria form the basis for assessment? Is equal weight given to all the criteria? What exactly is it that is being assessed: the final product alone or also the process leading up to it? For skills-based testing it is always advisable to make an assessment form available to students beforehand so that they know which aspects of their performance are going to be assessed.

Written assignments

Written assignments can count towards the final assessment in a given subject and students should be made aware of this beforehand. Here, too, it is essential that assessment and feedback criteria are made known to the students in advance.

Presentations

Assessing a presentation makes heavy demands on the assessor's short-term memory, unless the presentation is recorded on video. With this in mind, it is important to ensure that the assessment form does become an endless list of criteria. The form should be practical and easy to use. Once again it is important that the criteria for assessment are made known to the students beforehand. If several people are involved in assessing a presentation, it is advisable for each assessor to focus on different aspects.

In a group assessment it is also necessary, although more difficult, to identify individual contributions. Mutual feedback and self-reflection by the students can be useful resources for teachers when focusing on the contribution made by individual students within a group. These can bring to light the contributions made by separate students and can result in individual assessments (possibly as components of a final assessment).

Theses

There are two factors which can influence *reliability* when assessing theses: the subjectivity of the assessor and the unreliability of the assessment instrument. The first factor can be limited by working with more than one assessor. This is standard procedure for Master's theses. The Bachelor's theses are assessed by a single teacher. In cases where there are doubts as to whether a student should be awarded a pass, it is therefore advisable to consult a second assessor in order to safeguard the minimum level of acceptability for the Bachelor's thesis. However, this alone is not enough to guarantee the effectiveness of individual guidance and feedback during the process. Progress reports can be used to gain insight into this supervision. The reliability of the assessment instrument can be increased by working with an assessment plan for the thesis (not necessarily a single fixed plan), which is made known to the student beforehand (*transparency*).

- Master's theses are assessed independently by two members of teaching staff to increase the reliability of the assessment. (*Master's Thesis Regulations*)
- To ensure progress with regard to the thesis, the thesis supervisor and the student keep electronically track of the schedule, agreements made and the dates of meetings and/or discussions, either electronically or in written form. (*Master's Thesis Regulations*)

Bibliography

- **Toetsen en beoordelen** [Testing and assessment]. Gids voor docenten van de Faculteit der Economische Wetenschappen en Econometrie. [A Guide for Teachers at the Faculty of Economics and Econometrics.] 1996. Amsterdam: Vrije Universiteit.
- **Handleiding Toetsen en Beoordelen** [Teaching and Assessment Manual]. 1998. Amsterdam: Vrije Universiteit, Faculty of Law.
- M. Jaspers and M. Schade. **Toets & Beleid** [Testing and Policy]. Toetsbeleid en geautomatiseerde toetsing [Testing Policy and Automated Testing]. Fontys
- Milius, J., Oost, H. and Holleman, W. 2001. **Werken aan academische vorming** [Working towards Academic Development]. Ideeën voor actief leren in de onderwijspraktijk. [Ideas for Active Learning in Educational Practice]. Utrecht: Utrecht University.
- Berkel, H. van. 1999. **Zicht op toetsen** [Testing in focus]. Toetsconstructies in het hoger onderwijs [Testing Constructions in Higher Education]. Assen: Van Gorcum.
- Dousma, T., Hortsen, A. and Brants, J. 1997. **Tentamineren** [Examinations]. Groningen: Wolters-Noordhoff.

Readers who would like to explore the subject of testing and assessment in greater depth can contact the Centre for Educational Training, Assessment and Research (CETAR) at VU Amsterdam. <http://www.onderwijscentrum.vu.nl/>. CETAR organizes a range of courses within the framework of professionalization in education.