

# Contents

0	Summary & the Test Development Cycle.....	2
0.1	Summary .....	2
0.2	Test Development Cycle Overview.....	3
1	Phase 1 – Test Design .....	6
1.1	Learning outcomes .....	6
1.2	Types of test.....	11
1.3	Taxonomies.....	12
1.4	Electronic assessment.....	14
1.5	Assessment plans.....	16
2	Phase 2 – Constructing an assessment matrix .....	17
2.1	Why develop assessment matrices? .....	17
2.2	Creating an assessment matrix .....	17
2.3	Assessment matrix and resits .....	19
2.4	Pre-test transparency.....	19
3	Phase 3 – Test Construction and Setting Standards .....	21
3.1	Reliability, validity and transparency of tests .....	21
3.2	Constructing a test.....	22
3.3	Drawing up marking instructions (to improve rater reliability).....	22
3.4	Group assignments.....	23
4	Phase 4 – Test Administration .....	23
5	Phase 5 – Assessment, Processing and Analysis.....	26
5.1	Calculating grades (and setting the cut-off score).....	26
5.2	Psychometric analysis of question-based tests.....	29
5.3	Detecting plagiarism.....	31
5.4	Inter-rater reliability.....	33
5.5	Feedback on assessment .....	35
6	Phase 6 – Recording and Communicating Results .....	37
7	Phase 7 – Evaluation and improvements .....	38
8	Appendices .....	40
8.1	Bloom’s Taxonomy .....	40
8.2	Aid for Assessing Learning Outcomes .....	44
8.3	Types of Test.....	45
8.4	Assessment Matrix Template for the Cognitive Domain .....	50
8.5	Pre-Test Transparency Checklist .....	51
8.6	Test Construction Directions .....	52
8.6.1	Test cover sheet .....	52
8.6.2	Question-based tests .....	53
8.6.3	Task-based Tests .....	61
8.6.4	Competence-based assessments.....	63
8.6.5	Assessment Form Checklist.....	66
8.6.6	Sample Examination Instructions.....	67
8.7	Protocol for Recording Grades.....	70
8.8	Protocol for Viewing Exam Scripts in the Speech Therapy Programme.....	72
8.9	Checklist for the Evaluation of Thematic Block Tests with a Block Team .....	73

## 0 Summary & the Test Development Cycle

### 0.1 Summary

#### **Intended for: examiners and assessment committees**

The purpose of this workbook is to provide examiners and assessment committees with guidance and tools for test development. At Hanze University, the roles of lecturer and examiner are closely interwoven. The examiner's role in testing and assessment – in other words, the test development cycle – ranges from designing examinations, administering them and assessing examination scripts, right up to recording the grade achieved by the student. All the rest, then, forms part of the lecturer's duties. Examiners are appointed by examining boards.<sup>1</sup> The role of assessment committees is to oversee the quality of (final) examinations acting under the responsibility of an examining board.<sup>2</sup>

#### **The workbook is organised in accordance with the Hanze University Test Development Cycle**

The test development cycle is described in Hanze University's student assessment policy.<sup>3</sup> Each chapter of the workbook covers a phase, or step, in the test development cycle and includes suggestions and materials for ways in which examiners can deal with that particular phase in the cycle. Staff are free to develop new materials during any phase of the cycle, so the chapters can be read independently of each other.

Although this workbook is by no means comprehensive, it provides a solid grounding in all the phases of the test development cycle. A huge amount of material about testing and assessment is available from a variety of other sources: the HG intranet, the Internet, books, etcetera. In this workbook, however, we have tried to explain the basics of testing and assessment as it is conducted at Hanze University. Another important source of information which supplements this workbook is the [website Toetsing HG](#).

#### **Examinations and test types**

The types of test described in this workbook fall under the definition of an (interim) examination (Du: *tentamen*) as formulated in the Student Charter (Art. 4.1.4): “an investigation of the student's knowledge, understanding and/or skills. An examination can be in the form of a written, oral or computer test and includes practicals, practical examinations, in-term assessments, (project) assignments, group assignments or any other form of assessment approved by the examining board. Students are always assessed individually; this includes work that is carried out as a group assignment.”

#### **Changes in comparison to the previous version of the workbook**

The first edition of this workbook was published in February 2012 in the context of the project, Promoting a Culture of Quality. The present version is one of the products that have resulted from the Testing and Assessment Quality project. Contributions have been made to both versions by teaching and learning advisors from the Teaching and Research Department, lecturers and examiners.

The main changes compared to the first version are:

- Phases of the test development cycle: names and descriptions have been adapted (see Chapter 0);
- Additions: products in the various phases of the test development cycle (see Chapter 0);
- Individual phases: the phases have been aligned with the new test development cycle, examples have been added, content has been clarified and supplemented (see Chapters 1-7);
- Additions: standardisation of test types, checklists per test type.

---

<sup>1</sup> Section 7.12c of the Higher Education and Research Act (“WHW Act”).

<sup>2</sup> Executive Board (2012). Hanze University Articles of Government.

<sup>3</sup> Executive Board (2004). *Toetsnotitie: Implementatie toetsbeleid en kwaliteitszorg* (Student Assessment Policy Document). Educational Affairs Department, Hanze University Groningen.

If you have any questions, the teaching and learning advisor in your school can help. They are able to provide more details and information about the latest changes or they can refer you to relevant literature.

## 0.2 Test Development Cycle Overview

The test development cycle is a PDCA (Plan, Do, Check, Act) cycle for testing which is made up of seven phases, or steps. The cycle is illustrated on the cover of this workbook; the figure below (Figure 0.1) shows the relevant products for each step.



*Figure 0.1: Products in each phase of the Hanze University test development cycle*

### Phase 1 (Plan): Test Design

Before a test can be constructed, the context of the test and its relationship to the educational programme must be established. After all, a test must match what is taught and be aligned to the intended learning outcomes, which implies that the learning outcomes must be defined beforehand. Otherwise they must be formulated during the design phase. Learning outcomes can be written for a curriculum, year, block or unit of study.

The learning outcomes, the school's vision on teaching and learning and its teaching methods form the basis for the assessment plan that can then be drawn up. An assessment plan at the block level is requisite. This plan is the basic design of all assessment in the block – it offers a first impression of the content and design of the tests in the block and shows what the examiners want the assessments to achieve.

### Phase 2 (Plan): Constructing an assessment matrix

Once the learning outcomes have been defined, a blueprint (a detailed plan) of the test is made so that the test can be constructed in a methodical way. The best known type of blueprint is the assessment matrix, which indicates the distribution of tasks/assignments across the learning outcomes and shows the extent to which the learning outcomes are assessed, and in what manner.

### Phase 3 (Do): Test construction and standard setting

As soon as the assessment matrix is available, construction of the actual test can begin. If a new unit of study is being developed, the first step can be to develop the test components. This could range from designing a complete case study with multimedia to writing a multiple-choice question about a particular topic. When writing the test components, it may also be necessary to write model answers and feedback comments about correct and wrong answers, collect relevant support materials or describe the case study. Good test instructions are also required ('what must the student do'). Finally, the test can be assembled from the several components that have been prepared, and joined to the examination cover sheet which lists all the test information.

When the test is ready, appropriate (provisional) standards have to be set. The (provisional) cut-off score must be determined and marking instructions provided. For group assignments, it has to be clear how individual assessment will take place. If no criteria have been set in Phases 1 or 2, this should be done now. The test and the standards should be screened by a second examiner (and, in some cases, the assessment committee). Completed tests are added to the test bank.

#### **Phase 4 (Do): Test Administration**

When the test is administered, the actual student comes into the picture. The student's identity must be established whether the test is written, electronic or oral. Digital identification is not easier, in itself, than photo recognition. Electronic tests must also be administered according to a fixed procedure.

#### **Phase 5 (Do): Assessment, processing and analysis**

After the test has been administered, it is assessed against the standards set in the test construction phase (Phase 3). Analysis of the assessment process and the results will provide an insight into the degree of difficulty, the (inter-rater) reliability and the validity of the test. This may prompt a review of the standards and the cut-off score and, possibly, items in the test bank. When the definitive standards and cut-off score have been established, each student receives a final score; this may be after they have viewed their script or the exam has been discussed in class and feedback has been given.

#### **Phase 6 (Do): Recording and communicating test results**

The examiner records the test results by entering them in ProgRESS, the software programme used by Hanze University for the registration of students and for recording their assessment and examination results. The examining committee formally determines the definitive test results. Students may include their test results in their individual portfolios if they wish.

#### **Phase 7 (Check and Act): Evaluation and improvements**

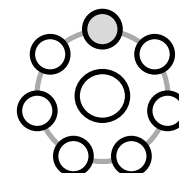
The purpose of evaluation is to assess the quality and coherence of a set of tests, for example, all the tests of a block. The analyses of the individual tests (with regard to the degree of difficulty, reliability, validity, etc.) can be compared, and related to the assessment matrices and the assessment plan. In this way, weaknesses can be identified and amended. The insights gleaned from the evaluation can serve as a basis for decisions to modify the learning outcomes, the assessment plan, assessment matrices and tests or test components.

#### **Test banks and assessment committees**

This workbook does not concern itself with test banks or assessment committees (indicated in the circle in the middle of the test development cycle). Assessment committees investigate the quality of examinations, acting under the responsibility of an examining board (HG Articles of Government, 2012), and are also responsible for assuring the quality of the test development cycle and that it is being applied within programmes or programme clusters (HG Teaching and Learning Framework, 2012). Their organisation and exact duties may vary with the school or programme at the discretion of the examining board. The duties and responsibilities of all the actors involved in testing and assessment are dealt with extensively in the 'Actors Interrelations Guidance' (only available in Dutch, probably), [handreiking samenspel actoren](#).

Test banks are regulated by the Higher Education and Research Act [*Wet op het hoger onderwijs en wetenschappelijk onderzoek*], the NVAO assessment frameworks and the retention periods set by Hanze University. The way in which test banks are maintained depends on the school or programme.

The Test Bank and Examination Files Manual may be of assistance here (except it is only available in Dutch); [handreiking toetsarchief en examendossiers](#).



# 1 Phase 1 – Test Design

## 1.1 Learning outcomes

One way of describing study programmes, or parts of them, is to look at what the programme wants a unit of study to achieve (lecturer perspective). Alternatively, one can consider what knowledge, understanding and skills students are expected to master upon completion of a programme or a part of it (student perspective). At the moment, in the Netherlands and internationally, a shift from the lecturer to the student perspective can be observed with regard to the use of intended learning outcomes. Some programmes have made more advances in this area than others.

### Why write learning outcomes?

There is no single, unambiguous way to describe competencies – they are often described in general and abstract terms. In some cases competencies may already have been described in learning outcomes. An example of such a general description is:

*The student is able to formulate and solve legal questions on the basis of legally relevant facts and legal sources.*

Competencies can be translated into concrete learning outcomes observing the guidelines in the Guide to Formulating Degree Programme Profiles.<sup>4</sup> An example of a concrete learning outcome:

*The student will demonstrate that, acting in a professional legal context, after qualifying and analysing a stated case he is able to give sound legal advice based on legal literature, statutory law and case law.*

All programmes at Hanze University have to formulate learning outcomes which students must have achieved on completion of their training, or which reflect the exit level requirements of the programme. This is in harmony with agreements made in Europe, and is important because external agencies including the NVAO (the Accreditation Organisation of the Netherlands and Flanders) and the Education and Training Inspectorate are exercising closer supervision of the intended exit levels of programmes and whether these levels are actually achieved.

### Translating competencies

Hanze University imposes no requirements on the way in which study programmes concretise competencies for any year, block, unit of study or strand of learning. As a general principle, the programmes are accountable themselves for their teaching and assessment in relation to the required exit level.

Teaching and assessment must also be transparent for lecturers, examiners and students. Each programme uses a competency matrix which states what competencies are covered at what levels in which (thematic) block. There are three levels: introductory (1), intermediate/deepening (2) and advanced (3). In addition to a competency matrix the programme can also draw up a table which lists the learning outcomes for each block (see the example in Figure 1.1) or it can choose to integrate the learning outcomes in the competency matrix.

Block 1.1 Learning outcome A1 Learning outcome C1	Block 1.2 Learning outcome A2 Learning outcome B1	Block 1.3 Learning outcome B2	Block 1.4 ...
Block 2.1	Block 2.2	Block 2.3	Block 2.4
Block 3.1	Block 3.2	Block 3.3	Block 3.4
Block 4.1	Block 4.2	Block 4.3	Block 4.4

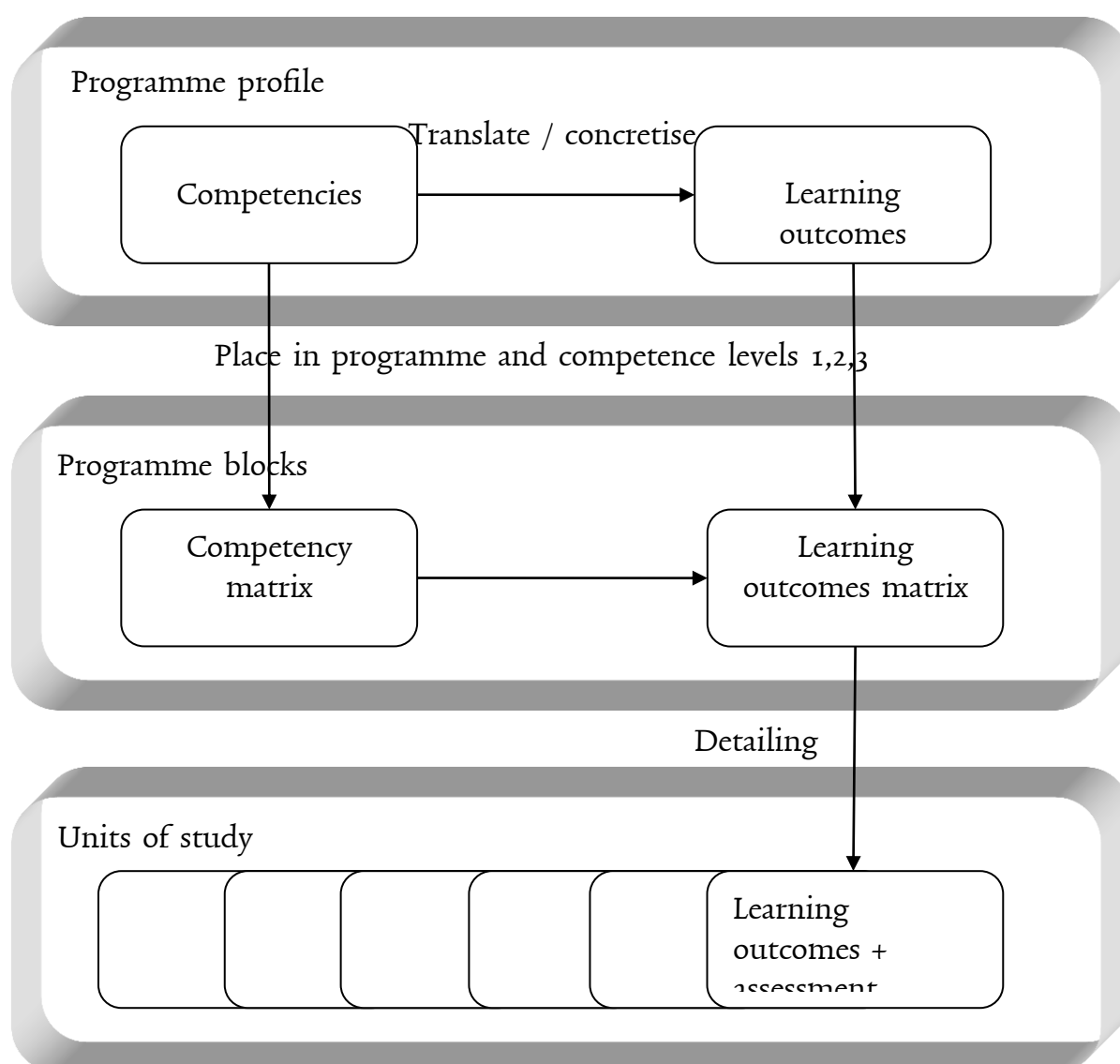
Figure 1.1 Learning outcomes table

<sup>4</sup> Tuning, Educational Structures in Europe: A Guide to Formulating Degree Programme Profiles. Published by, e.g., Deusto/University of Groningen/Nuffic/NVAO (p. 45).

Learning outcome A1 is the translation/concretisation of competency A at level 1, learning outcome A2 is the translation at level 2. The programme can indicate in this table which blocks/learning outcomes define the exit level of the programme.

The learning outcomes can then be detailed by block so that examiners can develop relevant units of study and appropriate tests based on the learning outcomes. In practice, most programmes at the very least formulate learning outcomes at the level of units of study/ProgRESS codes (assessment level). This makes it clear to students what they must do to pass the test and it guides the learning process. For examiners, the learning outcomes serve as a starting point for creating the assessment matrix, as an assessment matrix indicates how the learning outcomes are assessed and what weight is assigned to each learning outcome. Chapter 2 deals with this in detail.

With the diagram below, a programme can clarify the relationship between teaching/assessment and programme competencies.



*Figure 1.2 Relationship between programme competencies and teaching/assessment*

Learning outcomes can also be used to describe strands of learning. For example, all the learning outcomes for the Learning Dutch strand that have been formulated at the level of the units of study, together form a helpful starting point for describing the whole strand.

### **The importance of writing learning outcomes**



Learning outcomes form the basis for developing teaching materials and tests. Developing these materials and test items on the basis of interrelated learning outcomes will create a strong relationship between teaching and testing.

If learning outcomes are phrased in too abstract terms, there is a risk that the curriculum and the tests will not be properly aligned. So, only develop materials and tests once good learning outcomes have been formulated. If not, they need to be developed first.

Developing teaching materials and tests on the basis of learning outcomes will produce a transparent relationship between teaching and testing. If the teaching materials and the assessment matrix are designed in the context of a uniform teaching methodology, this will further strengthen the relationship. In the diagram below, learning outcomes have been taken as the starting point.

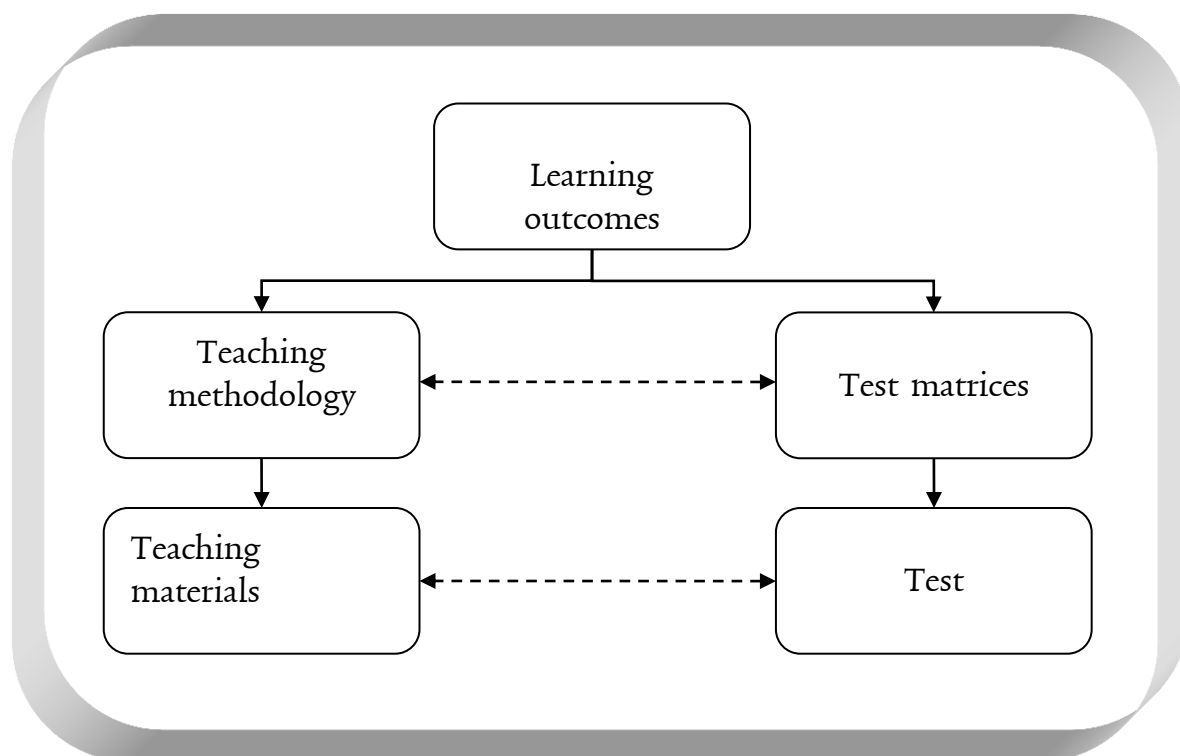


Figure 1.3: Learning outcomes as the starting point for developing teaching materials and tests

### Writing learning outcomes

In this workbook we focus on writing learning outcomes (LOs) for the purpose of developing tests. A learning outcome is what the student ‘is able to do’ as a result of what they have learned. Learning outcomes should be written in concrete and active terms using specific verbs (‘action verbs’) and should contain the following five elements:

- An action verb (according to Bloom’s taxonomy)
- An indication of the type of learning outcome
- The subject matter covered by the learning outcome
- An indication of the standard or level established by the learning outcome
- The scope and/or (professional) context of the learning outcome

Example:

*Based on a given research question, the student will be able to carry out independent desktop research by finding and analysing the relevant sources and writing a critical paper of approximately sixty pages based on their research.*

The five core elements of a learning outcome:

Carry out	Skill	Desktop research	Critical paper of	Independently,
-----------	-------	------------------	-------------------	----------------

			some sixty pages	based on a given research question
<b>A. Verb</b>	<b>B. Type</b>	<b>C. Subject</b>	<b>D. Standard</b>	<b>E. Scope/context</b>

Figure 1.4: Example of the elements of a learning outcome

*Re A.*

Use an action verb which is appropriate to the intended level. If, for example, the intended level relates to explaining certain phenomena, this must be stated in the learning outcome. A concrete example: If the student, applying Ohm's laws, must be able to explain why a cyclist cycling on a level road will come to a halt if he stops pedalling, a learning outcome at the knowledge level (e.g., 'The student is expected to be able to define Ohm's laws') will not do. To choose a suitable action verb you can refer to Bloom's taxonomy (see Appendix 8.1). Avoid using vague verbs like know, understand, learn, be familiar with or be aware of, because such terms are difficult to measure.

*Re B.*

It must be clear whether the learning outcome relates to knowledge, attitude or skills. This can usually be expressed by using a suitable verb (see 'A' above).

*Re C.*

The subject area that a learning outcome pertains to can either be specific or general. An example of a very specific description is, 'the action of the ossicle muscles'. Less specifically, one could have 'carrying out qualitative research methods'. If the latter learning outcome can be specified more narrowly, for example, if only in-depth interviews are assessed, this must be expressed in the learning outcome. So, if students only have to be able to conduct in-depth interviews with a certain target group upon completing a certain unit of study, the learning outcome can be articulated even more specifically as, e.g., 'conducting in-depth interviews with immigrants'.

Discussions can arise about what exactly falls under 'subject area' and what under 'context' (e.g., only a certain target group may be relevant in the professional context) or is actually related to 'level' (some target groups are easier to interview than others). Something that relates to the level in one programme may be a matter of context in another programme.

*Re D.*

The standard or level can be expressed in different ways in the learning outcome: in the description of the result (the expected result of the action), the success rate (e.g., the student is allowed to get 10% of the items wrong) or as 'present/absent' (the student demonstrates the action or not).

Often, the level of the learning outcome can be inferred from the suggestions, or a list of do's and don'ts that an examiner or lecturer gives to students when they start out on a task. It concerns things that people would or wouldn't do if they were experts (above standard versus beginner's mistakes). This information enables one to determine what requirements the student's results or behaviour must meet. Clues as to the level will also be evident from the *preconditions*: the materials (say a calculator) and information (say a given formula) provided and the *minimum result* expected (say identifying three differences). A learning outcome which is as specific as this is almost as good as an assessment criterion.

*Re E.*

Describe the context in terms that are also used for describing the competence levels (1, 2 and 3) and make sure that the learning outcome is aligned with the professional context and/or the relevant national professional or programme profile. Level 3 concerns the exit qualifications of the programme, after all, the newly qualified student must be competent to start working as a professional practitioner. You could use the description of the complexity (a pre-structured assignment or, conversely, a very open one) or the degree of guidance provided (working independently or under close supervision). For units of study that take place in actual practice settings (such as placements or final projects) one can state the actual professional context. For other units of study it may be more difficult to denote the context. Some examples of context features:

- in a group/individually;
- for a fictitious client/for an SME client;
- in an international or national company;
- in a legal context;
- in the role of someone handling a case or giving treatment in a role-playing game;
- in the role of chairperson or member of a project group;
- for a customer in China/for a customer in Belgium;
- in a hospital or an old people's home as a newly qualified social worker.

### Learning outcomes and competence levels

A question that is often asked is whether the levels of Bloom's taxonomy can be linked to the three competence levels which most programmes distinguish. The answer is that there is no one-on-one relationship. Bloom's higher orders can be assessed even in year 1 at competence level 1, as long as it is in the form of a simple, pre-structured assignment. In such a case, the learning outcome for year 1 should be described in such a way that it cannot also apply to year 4, the exit level.

In practice, it is mainly the lower orders (remembering, understanding, applying) that will be assessed in year 1. In the course of the programme greater emphasis will be placed on the higher orders (analysing, evaluating, creating). To get an idea of the relationship between the competence levels and Bloom's levels, compare the levels in the figure below but remember that there is no one-on-one relationship.

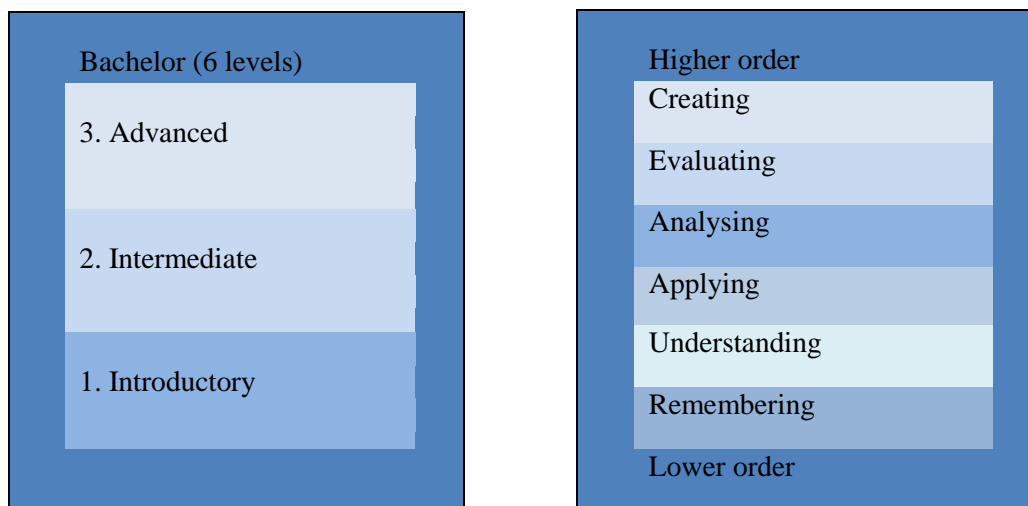


Figure 1.5: Competence levels and Bloom's levels

### Suggestions for formulating learning outcomes

- Avoid complex sentences. If necessary, use more sentences to ensure clarity.
- Learning outcomes must be assessable. If they are expressed too broadly, it may be difficult to assess the outcome whilst, if they are expressed too narrowly the list of learning outcomes could become too long and detailed. As a result, it could happen that not all the learning outcomes are actually assessed (this will become evident when the assessment matrix is created). If the subject matter deals with sixteen statistical tests but only two or three of them are assessed, it is not advisable to specify all the tests in the learning outcome; better use a collective name. All the same, it must be apparent from the description of the subject matter that the students are expected to be able to perform all sixteen tests. It is recommended to write five to ten learning outcomes (at most) for any component.
- When writing learning outcomes, keep in mind the period/time limit within which students should achieve the learning outcome, and consider whether it is realistic for the outcome to be achieved within the time and resources available. Units of study are not normally taught in isolation but, rather, form part of a strand of learning and it is important that the strand units articulate with each other. For example, Unit II may continue from the learning outcomes of Unit I.

- Check whether the learning outcomes you have written make sense to colleagues, former students and industry professionals.
- Writing learning outcomes in terms of the five core elements is not always easy (see the example given under ‘C’). For one thing, the level of a learning outcome appears from the combination of the elements, so it’s important not to set too rigid criteria for any learning outcome. Rather, writing learning outcomes should be regarded as a useful aid in discussions with colleagues in order to get a clear focus on what students must be able to demonstrate on completion of a programme, block or unit of study. Working together, any overlaps or hiatuses in the curriculum can be more easily identified and cleared up.

A method for assessing learning outcomes can be found in Appendix 8.2.

## 1.2 Types of test

The type of test to choose depends much on the teaching and learning methods used by the programme. In vocational education, subject-based teaching and learning has given way to theme-based, project-based and/or problem-based teaching, or combinations of these, and the intended learning outcomes also influence the choice of test type.

Each type of test has advantages and disadvantages. To make the right choice, one should consider the teaching methods, the learning outcomes, the purpose of the test (formative assessment, summative assessment, prediction, other), its quality (validity, reliability, transparency) and the resources available. As to resources: what resources are available for the type of test, such as a test bank or a computer room, but also the amount of time available to the examiner. Choosing a suitable test type is even more difficult if the learning outcomes relate to professional competencies or complex skills, as these involve all three of the domains of knowledge, attitude and skills. Often, it will be best to use a combination of test types to assess whether a student has achieved all the learning outcomes. It can also be interesting to observe how assessment of comparable assignments/products is conducted in professional practice.<sup>5</sup>

To encourage examiners and students at Hanze University to use the same names for the various types of test as much as possible, the test types have been standardised. The recommended names are printed in bold in the table below. For an explanation of the test types, see Appendix 8.3.

<b>Overview of Test Types</b> (for details about test types, see Appendix 8.3)	
<b>Question-based tests</b>	
<b>Multiple-choice test</b>	Other names used: <ul style="list-style-type: none"> <li>• <i>MC test, written examination</i></li> </ul>
<b>Open-ended questions</b>	Other names used: <ul style="list-style-type: none"> <li>• <i>Written examination, oral examination</i></li> </ul>
<b>Case study</b>	Other names used: <ul style="list-style-type: none"> <li>• <i>OverAll Test, problem-based learning task (PBL task), applied knowledge test</i></li> </ul>
<b>OverAll Test</b>	Other names used: <ul style="list-style-type: none"> <li>• <i>Case study, applied knowledge test</i></li> </ul>
<b>Task-based tests (assignments)</b>	
<b>Assignment</b>	Other names used: <ul style="list-style-type: none"> <li>• <i>Placement, study abroad, final project</i></li> </ul>

<sup>5</sup> Text adapted from:

[http://www.utwente.nl/mb/onderwijs/themas/toetsen\\_en\\_becoorden/Job%20Aid/keuze\\_toetsvorm.doc/](http://www.utwente.nl/mb/onderwijs/themas/toetsen_en_becoorden/Job%20Aid/keuze_toetsvorm.doc/) This website also offers materials about the purposes of testing, quality requirements and possible applications.

<b>Reflection assignment</b> Other names used: <ul style="list-style-type: none"> <li>• <i>Process report, reflective account, (self) evaluation, peer coaching, peer supervision, (keeping a) logbook, 360° feedback</i></li> </ul>
<b>Skills test</b> Other names used: <ul style="list-style-type: none"> <li>• <i>Practical assignment, professional skills test</i></li> </ul>
<b>Project assignment</b> Other names used: <ul style="list-style-type: none"> <li>• <i>Project, external/internal assignment, 'co-makship', professional product, virtual business case</i></li> </ul>
<b>Integrated assessments</b>
<b>Competence-based assessment</b> Other names used: <ul style="list-style-type: none"> <li>• <i>Behavioural assessment, peer assessment, performance assessment, expert assessment</i></li> </ul>
<b>Portfolio</b> Other names used: <ul style="list-style-type: none"> <li>• <i>Portfolio interview, portfolio assessment</i></li> </ul>
<b>Criterion-based interview (CBI)</b> Other names used: <ul style="list-style-type: none"> <li>• <i>Competence-based interview, final interview</i></li> </ul>

Figure 1.6: Types of test

### 1.3 Taxonomies

The level of learning outcomes and assessment criteria can be included in a taxonomy, that is, a hierarchical classification. In education, taxonomies have been developed for assessment criteria. The best known ones are:

- Bloom's taxonomy;
- Romiszowski's taxonomy.

Other taxonomies include those developed by De Block and Merrill.

When comparing Bloom's and Romiszowski's taxonomies, the main differences are:

- Bloom's taxonomy is less specific than Romiszowski's. Which is the most appropriate one to use depends much on the learning outcomes and the assessment criteria. If the assessment criteria relate to cognitive skills, both taxonomies are equally suitable. If the assessment criteria also address other skills, Romiszowski's taxonomy would be more appropriate. Bloom's taxonomy is more often used in higher vocational education, whereas Romiszowski's taxonomy is generally preferred in secondary vocational education.
- Romiszowski's taxonomy is much more labour-intensive than Bloom's, but generates more information. Bloom's taxonomy is generally preferred, which may be for this reason and also because it is valuable even if it is applied less accurately, whereas Romiszowski's taxonomy is more demanding.

In this workbook, we have chosen to go with Bloom's taxonomy.

### Test Types and Bloom's Taxonomy

In the table below, test types are linked to Bloom's taxonomy. If a learning outcome relates to the taxonomy, you can use this table to find appropriate types of assessment.

Learning Outcome Levels  Type		Appropriate Types of Test				
		Cognitive knowledge (Bloom: lower order cognitive domain)	Cognitive skills (Bloom: higher order cognitive domain)	Physical skills (Dave: psychomotor domain)	Attitude (Bloom: affective domain)	(Maximum) integration of knowledge, attitude and skills
Question-based test	Multiple-choice	XX	X			
	Open-ended questions	X	X			
	Case study		X			
	OverAll Test		X			
Task-based test	Assignment	X	X	X		
	Reflection assignment		X			
	Skills test		X	X		
	Project assignment		X	X	X	X
Integrated assessment	Competence-based assessment		X	XX	XX	XX
	Portfolio		X	X	X	XX
	Criterion-based interview (CBI)		X		X	XX

Figure 1.7: Types of test in relation to learning outcomes

## **1.4 Electronic assessment**

Electronic assessment systems can be helpful for the construction, administration or processing of a variety of test types. When the type of test has been determined, it is a good idea to check whether electronic support might be useful. In the beginning, electronic testing will cost time as you learn how to use the system, organise your folders, design test items, formulate feedback and assemble the test, but in the end it will save time and, not least, benefit student learning and improve the quality of testing.

Electronic testing and processing is particularly helpful for use with question-based tests. Electronic testing is also suitable for tests whose purpose is to assess student learning (summative assessment), for diagnostic tests and for practice tests (formative assessment).

### **Reasons for administering a summative paper-and-pencil test electronically**

- Saving time. In the end, e-assessment saves time. Closed-ended questions can be marked automatically.
- Benefit to student learning because of swift feedback. Students receive their test results as well as feedback immediately after they have completed the test. Students appreciate this quick feedback and it can benefit their learning process.
- Improved quality. Many electronic assessment systems have a range of features for assessing the quality of test items and the test as a whole, and improve them accordingly. This increases the validity of the test.
- Using different types of questions. E-assessment offers the option of introducing types of questions that may be problematic to use in a paper-and-pencil test such as questions that involve audio or video materials (questions about audio or video fragments).
- Random order of items/alternatives. To prevent students from cheating, items and alternatives can be presented in random order which makes copying or consulting more difficult. This improves the reliability of the test.
- The quality of answers to open-ended questions. It has become increasingly clear that electronic assessment of tests with open-ended questions offers significant benefits. It makes it easier for students to phrase and structure their answers well and this, along with good legibility, also benefits the examiners.
- Building an item bank. An electronic item bank can be set up with fellow examiners which can be used to assemble tests. If the item bank is large enough and is regularly refreshed and supplemented, items can be used repeatedly in both summative and formative tests.

### **Reasons for digitalising formative tests**

- Flexibility. Students can do digital formative tests at any time or place, as often as they want and at their own pace.
- Immediate feedback and study advice. In formative tests, students can receive immediate feedback about why particular answers they gave were wrong or correct, and what they need to work on. It is an active form of instruction which can support student learning.
- Random test assembly. The examiner can assemble a unique test for each student, for example, to provide students with several different practice tests. As soon as the practice test is ready for delivery, the examiner has nothing left to do in the way of assembly. In some systems the lecturer and the examiner can see which student has done the practice test and what effect this has had on the outcome of the 'real' test; in other words, they can monitor the student. Random test assembly is less suitable with regard to summative tests in connection with test quality analysis.

### **Other types of test**

Other types of test can also be supported by digital resources. For example, there are systems which support the development, administration and processing of assessments. In these systems no item banks are created but assessment criteria are recorded. The examiner, an external expert or the peers awards scores on the basis of the criteria and the system registers, combines and processes the score lists. To be able to use these systems, it is a precondition that clearly defined learning outcomes or assessment criteria are in place.

### **Electronic assessment systems**

Electronic testing is being applied more and more at Hanze University. The following systems are available.

#### Question Mark Perception

Question Mark Perception (QMP) enables examiners to write ('author') and administer ('deliver') questionnaires, tests and assessments online. It includes a reporting feature for viewing and analysing results.

QMP cannot only make testing more efficient, it can also be a powerful teaching and learning tool if it is used for practice tests with extensive feedback. This enables students to test their knowledge (and skills) at any time wherever they are.

#### Connecticut

Connecticut is primarily a system for managing competencies, learning outcomes and assessment criteria. It generates assessment lists based on assessment criteria which form the basis for assessment by examiners or experts. In addition, students can assess themselves and each other (360 degree feedback, peer assessment). Students' progress can be monitored via the competency file. For more information, see **Fout! De hyperlinkverwijzing is ongeldig.**

#### Blackboard Learn

Blackboard Learn incorporates several modes for testing and evaluation:

- (Practice) test;
- Assignment: optionally with plagiarism detection through SafeAssign;
- Self and peer assessment;
- Survey;
- Discussion board (graded);
- Wiki (graded);
- Blog (graded);
- Journal (graded).

Blackboard Learn also includes an item bank, the Pool Manager, and questions can be imported or exported to a file.



## 1.5 Assessment plans

An assessment plan is an overview of all the assessments within a curriculum. What units of study is the curriculum comprised of, what types of assessment do the units use, what competencies are assessed, and at what level? An assessment plan at block level is required.

The assessment plan shows the learning outcomes (or competencies) and the types of test used to assess them. It is also a useful instrument for checking whether the learning outcomes for the thematic block are being assessed appropriately, whether the tests/assessments are sufficiently varied and whether the intended level is achieved. Below is an example of an assessment plan:

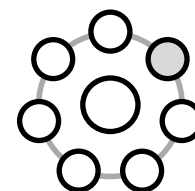
<b>ASSESSMENT PLAN FOR BLOCK X:</b>							
Unit of study	Learning outcome (LO) assessed	Time of test/handin g in	Purpose of the test (formative / summative)	Form of assessment	EC TS	Competenc e level	Assessment (mark, pass/fail, weighting)
Unit 1	LO a, b, c, d	Week 11	Summative	Report (advice)	6	2	Group mark 50%
		Week 10		Letter			Mark 30%
		Week 11		Presentatio n			Mark 20%
	LO b	Week 5	Formative	Assignmen t			Feedback
Unit 2	LO c	Week 11	Summative	Examinati on (MC)	2	2	Mark
Unit 3	LO d	Week 11	Summative	Examinati on (open- ended questions)	3	2	Mark
Unit 4	LO d, e	Week 9	Summative	Assignmen ts (3)	4	2	Group mark 50%, peer assessment 50%
		Week 11	Summative	Reflective account			Pass/Fail

Figure 1.8: Assessment plan at block level

All the assessment plans at block level together provide insight into the coherence and variation of the types of tests/assessments. They can also tell you whether certain competencies or learning outcomes are being assessed too little or too much at this or that level.

<b>ASSESSMENT PLAN FOR PROGRAMME X</b>			
Assessment plan for block 1.1	Assessment plan for block 1.2	Assessment plan for block 1.3	Assessment plan for block 1.4
Assessment plan for block 2.1	Assessment plan for block 2.2	Assessment plan for block 2.3	Assessment plan for block 2.4
Assessment plan for block 3.1	Assessment plan for block 3.2	Assessment plan for block 3.3	Assessment plan for block 3.4
Assessment plan for block 4.1	Assessment plan for block 4.2	Assessment plan for block 4.3	Assessment plan for block 4.4

Figure 1.9: Assessment plan at programme level



## 2 Phase 2 – Constructing an assessment matrix

### 2.1 Why develop assessment matrices?

Based on the learning outcomes a blueprint (a detailed plan) of the test is developed so that the test can be constructed in a methodical way. The best known type of blueprint is the assessment matrix. An assessment matrix helps assure that the subjects assessed in the test are represented and correctly weighted. It provides an overview of the test components and how they are weighted (how many points each of them attracts). In addition, the weight of the level of mastery can be specified (using Bloom's taxonomy); see the example below.

Advantages of using an assessment matrix:

- It prevents the writing or inclusion of superfluous questions or tasks addressing the same subject matter or skills.
- It is more likely that the test is a representative sample of the learning outcomes to be assessed.
- When two tests are constructed that deal with the same subject matter, the tests are more likely to be equivalent if they are both based on the same assessment matrix.
- An assessment matrix can serve to account for the test contents to colleagues, review teams, etc.
- Copies of the matrix can be handed out to students to promote transparency and ensure that students concentrate on the core subject matter. Used in this way, an assessment matrix can be a powerful tool for guiding student learning.

### 2.2 Creating an assessment matrix

Assessment matrices can be used for all sorts of test types. Project assignments and skills tests can also easily be divided into components. Essentially, an assessment matrix lists the learning outcomes for a test and the level and weighting of the learning outcomes in the test as a whole. Examples of assessment matrices can be found on the website, 'toetsing HG'.

#### Assessment matrix for the cognitive domain

For tests which only address cognitive learning objectives, the assessment matrix can be limited to the cognitive domain. A blank template for the cognitive domain is provided in Appendix 8.4.

Learning outcome	Knowledge questions / Reproduction	Comprehension questions / Understanding	Application questions	Analytical questions	Evaluation questions	Creation / design questions	Points
	Lower order			Higher order			
Outcome A	3 questions x 5 pts		2 questions x 5 pts				25 pts
Outcome B	2 questions x 5 pts	10 pts		10 pts			30 pts
Outcome C		5 pts	2 questions x 5 pts		2 questions x 5 pts		25 pts
... etc. ...							
Total	30 pts	15 pts	20 pts	10 pts	10 pts		80 pts

Figure 2.1: Assessment matrix for the cognitive domain

The shaded row shows the various levels of mastery in the cognitive domain. The rows below it show the learning outcomes. For each learning outcome, the level of mastery is indicated (see Chapter 1), Bloom's taxonomy) and the number of points the various categories attract. For example, mastery of Learning Outcome A at the knowledge level is worth 25 points out of a total of 80. As you can see, in this test Learning Outcome B carries more weight than the other two LOs (30 points against 25). In short, the assessment matrix conveniently indicates the priorities of the subject and, thus, the test as a whole.

### Assessment matrix for the three domains

If not just cognitive learning outcomes are assessed but psychomotor and affective learning outcomes as well, one can simply adapt the assessment matrix (template). In the example below, learning outcomes in the cognitive and the affective domains are assessed.

	Cognitive domain (knowledge)			Affective domain (attitude)			Psychomotor domain (physical skills)	
Learning outcome	Knowledge questions / Reproduction	Comprehension questions / Understanding	Application questions	Receiving	Responding	Valuing	...	Points
Outcome A	2 x 5 pts		5 pts		10 pts			25 pts
Outcome B	2 x 5 pts	10 pts		10 pts				30 pts
Outcome C		5 pts	2 x 5 pts		2 x 5 pts			25 pts
... etc. ...								
<b>Total</b>	20 pts	15 pts	15 pts	10 pts	20 pts			80 pts

Figure 2.2: Assessment matrix for all three domains

The shaded row shows the various levels of mastery in the cognitive, affective and psychomotor domains. The rows below it show the learning outcomes. For each learning outcome, the level of mastery is indicated (see Chapter 1, Bloom's taxonomy) and the number of points the various categories attract. For example, mastery of Learning Outcome A at the knowledge level is worth 10 points out of a total of 80. As you can see, this test focuses mainly on the cognitive domain (50 points out of 80).

### Two ways of describing learning outcomes in an assessment matrix

How to 'fill in' an assessment matrix depends on the agreements made about the formulation of learning outcomes. There are two ways of doing this.

Example:

The student must be able to describe and apply a certain formula.

Method 1:

You could decide to write two learning outcomes for this task:

*Learning outcome A: The student is able to describe the formula.*

*Learning outcome B: The student is able to apply the formula.*

Method 2:

The other option is to write only one learning outcome, namely, for the highest level that the student must achieve, based on the idea that it includes the lower levels:

*Learning outcome A: The student is able to apply the formula.*

Applying the first method, only one cell in each row in the assessment matrix is filled in, as the levels are reflected by specific learning outcomes.

### Method 1

Learning outcomes	Knowledge questions / Reproduction	Comprehension questions / Understanding	Application questions	Analytical questions	Evaluation questions	Creation / design questions	Points
Outcome A	3 questions x 5 pts						25 pts
Outcome B			10 pts				30 pts
Outcome C			2 questions x 5 pts				25 pts
Outcome D				2 questions x 5 pts			
<b>Total</b>	30 pts	15 pts	20 pts	10 pts	10 pts		80 pts

Figure 2.3: Assessment matrix with learning outcomes according to Method 1

The second method is demonstrated in the assessment matrix below. Learning outcome A ('the student is able to apply the formula') is assessed at two levels. Students can see in the assessment matrix that the test includes a reproduction-based question.

### Method 2

Learning outcomes	Knowledge questions / Reproduction	Comprehension questions / Understanding	Application questions	Analytical questions	Evaluation questions	Creation / design questions	Points
Outcome A	3 questions x 5 pts		2 questions x 5 pts				25 pts
Outcome B	2 questions x 5 pts	10 pts		10 pts			30 pts
Outcome C		5 pts	2 questions x 5 pts		2 questions x 5 pts		25 pts
... etc. ...							
<b>Total</b>	30 pts	15 pts	20 pts	10 pts	10 pts		80 pts

Figure 2.4: Assessment matrix with learning outcomes according to Method 2

### Stating the number of questions in the assessment matrix

In the assessment matrices above, the number of questions is stated but you don't have to include this information. In that case the weighting will have been established but it is still open how many questions will be attached to the weightings. This allows scope for variation in the number of questions for each learning outcome and level per test.

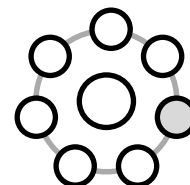
## 2.3 Assessment matrix and resits

All the learning outcomes that have been defined for a unit of study must also be assessed in resits. This may sound logical, but it doesn't always work out this way in practice. For instance, take the case of a student who has to do an individual make-up assignment for a project he failed. If the fail was a consequence of poor co-operation with the other group members, the underlying learning outcome cannot be (re-)assessed by an individual assignment. By comparing the assessment matrix and the project result to establish which components have to be made up for, the examiner can ascertain what the resit assignment should look like. Individual assessment of group work is dealt with in detail in Chapter 3.4.

## 2.4 Pre-test transparency

The products from Phase 1 (learning outcomes, assessment plan) and Phase 2 (assessment matrix) contribute to the transparency of assessment. Transparency means that the assessment procedure must be clear to all the people involved. We distinguish transparency before a test is administered ('pre-test

transparency’), transparency of the test itself (see Chapter 3.1) and transparency of the assessment procedures (see Chapter 5). A checklist of the requirements regarding pre-test transparency is provided in Appendix 8.5.



### 3 Phase 3 – Test Construction and Setting Standards

The next step in the test development cycle consists of constructing the test using the assessment matrix drawn up in the previous phase, and setting standards. Important concepts in this context are reliability, validity and transparency, which will be explained first. Some tools are then offered for constructing tests and setting the standards.

#### 3.1 Reliability, validity and transparency of tests<sup>6</sup>

This section deals with the reliability, validity and transparency of tests. Phase 5 of the test development cycle deals with the reliability and validity of test results.

##### Reliability of tests

A test or assessment is considered to be reliable if the same results are produced when it is repeated under the same conditions. A reliable test is:

- Of sufficient length  
Reliability is affected by the number of items in a test. If a test has few items/questions, the weight of each individual item is much greater than if the test has many items. Using an assessment matrix, guidelines and checklists helps ensure that the test is long enough. Be aware that the length of a test must always be commensurate to the time available for doing it.
- Of an adequate level/degree of difficulty  
The items/questions must have the right degree of difficulty in relation to the intended level of learning required at that specific stage of the programme. To ensure this, the questions must be based on the learning outcomes and an assessment matrix should be used.
- Specific (addresses the subject matter)  
The test questions must be written in such a manner that only students who have mastered the subject matter can answer them correctly. The questions/items must mirror the content of the learning outcomes.
- Selective (differentiates)  
The test as a whole and each individual item should differentiate between weak and strong candidates. The difficult items must be answered correctly by the strong candidates in particular.

##### Validity of tests

A test is valid if it measures what the designer wants it to measure. In other words, a valid test measures what it is intended to measure. A valid test is:

- Representative  
A test should reflect the subject matter taught in the block as well as the teaching and learning methods used. A good test is representative of the teaching content and the learning outcomes (content-related validity). Using a test matrix helps to assure content validity. The test type, the questions/items and the aids that students are allowed to use in the test (calculators, formula books, law codes, etc.) must be based on, and relate to the learning objectives (construct validity).
- Balanced  
Tests should include both easy and difficult questions, starting with simple questions and going on to more complex ones. The assessment matrix indicates the type(s) of questions and the level of complexity required in the test, and helps ensure that the test is well-balanced.
- Sensitive  
The questions/items must be suitable for the whole target group. They must relate equally to everyone in the target group and must not advantage or disadvantage any section of the group.
- Unambiguous  
Questions/items must be worded clearly without any ambiguity. There must not be any doubt as to the answer that is expected from the candidate. Extremely difficult or easy questions should be avoided.

---

<sup>6</sup> Berkel, H van, en Bax, A. (red). *Toetsen in het Hoger Onderwijs*. Houten: Bohn Stafleu van Loghum, 2006

### Transparency of tests

A test is transparent if the students have been provided with clearly presented instructions.

Presentation/layout of the test:

- Questions are clearly separated;
- The numbering of the questions is correct and transparent;
- References to text or figures are correct;
- Questions, their components and options (in MC tests) are clearly separate.

Instructions for students:

In written tests, provide clear instructions and information (you could use the test cover sheet) about:

- the test procedure;
- the use of aids;
- the criteria and scoring information;
- the distribution of points over the questions or tasks;
- the cut-off score;
- the time available for completing the test.

Sample test instructions are provided in Appendix 8.6.6.

Instructions for invigilators:

Make sure that invigilators are fully informed about the exam procedures. Specifically, it is important that the cover sheet has been filled in correctly by the examiner so that it is clear what aids the students are allowed to use.

## 3.2 Constructing a test

Different rules of construction apply to different types of test. Directions for constructing various kinds of test are provided in Appendix 8.6. In addition, sample test instructions are included (8.6.6) along with a standard cover sheet for question-based tests (8.6.1) and a list of points for attention for the assessment forms of task-based tests and integrated assessments (8.6.5). If a test is designed by a single examiner it must be reviewed by another examiner before it is administered, in accordance with the ‘four eyes principle’.

## 3.3 Drawing up marking instructions (to improve rater reliability)

Marking instructions (*Du: correctievoorschrift*) serve as a manual for examiners. They are intended to ensure a high degree of agreement between the examiners (rater agreement) and include guidelines and instructions for interpreting and assessing answers to test items. Rater reliability or rater agreement alone, however, cannot guarantee the reliability of a test. A high degree of rater agreement is a necessary but not a sufficient condition for achieving high test reliability.

Rater reliability refers to either:

- intra-rater reliability: the consistency of test scores awarded by one rater on separate occasions; or
- inter-rater reliability: the consistency of scores awarded by two or more raters for the same test.

Rater reliability is important to enable equivalent and consistent statements to be made about test results, and assurances must be built in to exclude rater effects as much as possible. Marking instructions serve this purpose. Using them effectively contributes to a good quality of interpretation of test results and helps assure that the test results are reliable.

Marking instructions are composed of:

1. Guidelines for Assessors (*Du: beoordelaarinstructie*)

Guidelines for Assessors give general rules for assessing test takers' performances. They describe the assessment process and the marking requirements that are relevant to the nature and form of a specific test. For example, they describe how to deal with:

- partially correct answers and answers that are not included in the model answers;
- consequential errors (where an error in one item affects the answers to related questions);
- identical errors;
- grammatical errors.

They also describe the order of marking: item by item starting with the first item of the scripts, or script by script.

2. Model answers or marking guidelines (Du: *antwoordmodel; beoordelingsschema*)

Model answers and marking guidelines list the correct – or partly correct – and wrong answers, and include assessment criteria to guide the examiner.

3. Marking scheme (Du: *scoringsvoorschrift*)

A marking scheme sets out clearly which answers merit the maximum score and which deserve a partial score, and what the (partial) score will be. A marking scheme also explains how partial scores are combined into a final score and what the cut-off score is. At this stage, the marking scheme and the cut-off score are provisional. The definitive marking scheme and cut-off score are not established until the work has been marked and moderated.

### 3.4 Group assignments

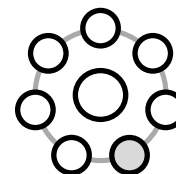
With the advent of competence-based teaching and learning, group assignments became more popular with teachers. However, when it became clear that group assessment often didn't do justice to individual students, individual assessment was made compulsory. Using group assignments has a number of advantages, notably efficiency, authenticity and the stimulus provided by working in a group. In addition, certain learning outcomes (such as collaboration) can only be demonstrated in a group.

Group assignments are subject to certain preconditions, one of them being that individual assessment is also required. No specific requirements apply to individual assessment, but every student must (be able to) demonstrate the learning outcomes.

Assessing students individually in group assignments can be done in various ways. Two methods are distinguished for establishing the final grade:

- group assessment by the examiner in combination with individual assessment by the group members, the peers.
- group assessment by the examiner (which may be supplemented by peer assessment) in combination with individual assessment by the examiner.

The various types of group assignments and their advantages and disadvantages are described in the report, Individual Assessment in Group Assignments: [Individuele beoordeling bij groepsopdrachten](#).



## 4 Phase 4 – Test Administration

Tests should be administered under the best possible conditions and must match the learning outcomes as well as the teaching method used during the learning process. Identical tests should be administered under identical conditions. During a test session, several things can influence the reliability and validity of the test results, such as the test instructions, the test itself, the aids used by the students, the environment or the people involved in the test. This chapter concerns tests that are scheduled at a specific time, which will usually be question-based tests and integrated assessments. Task-based tests usually are not scheduled at a specific time, as students work on the assignments during a certain pre-agreed period and the examiner gives the students deadlines for handing in their work.

Depending on the type of test, the following factors are of lesser or greater importance.



### Tests and test instructions

Tests must be identical for all the students assessed. In question-based tests this will normally be automatic, as each student gets an identical copy of the test paper and the test instructions. Make sure that all the copies are of the same (good) quality and that enough copies are made so that each student has one.

For test types that are assessed orally (tasks, assignments, questions) or where a combination of oral and written assessment is applied (e.g., an oral explanation to a written task), it is important that all the students are assessed in the same way (e.g., in a behavioural assessment). Student A should not be given different or more detailed directions for solving a problem or performing an action than student B; the conditions should be the same for all the students as much as possible. Similarly, role-play actors must act in the same way in identical situations. This is important to keep in mind, especially if different examiners and role-play actors are involved in administering a test.

### Test aids

In phases 1 and 2 of the test development cycle it is decided what aids students may use during a test (such as formula books or calculators). Students must also be capable of using these aids. If the programme makes materials available, such as practical equipment, there has to be enough for all the students and it must be in good working condition. Students may not have any aids at their disposal that they are not permitted to use. Invigilators must keep a watchful eye on this.

### Test environment

The (practical) room must be suitable for the test. The test conditions must be as favourable as possible for the students so, for example, there must not be any construction work going on in the building, lighting and ventilation must be good and there should be enough room (unless the learning outcomes require otherwise). Students must not be able to see each other's work, so the space must be large enough and the tables should be arranged in such a way that students are not tempted to cheat.

### Persons involved in administering the test

It will be clear from the above that examiners have only so much control over the test conditions. Besides, the degree of their control varies with the type of test. In question-based tests, in particular, examiners have no influence at all during the test. Invigilators must observe the invigilation instructions and the students' test directions in fulfilling their task of ensuring that all goes smoothly. The clearer the instructions, the less confusion during the test.

In integrated assessments, examiners do have an important role to play. Integrated assessments are also time-consuming to administer, as students are assessed individually or in small groups and one or more examiners as well as role-play actors are present at each administration.

### Location

Most tests are held in examination rooms but they may be held in a variety of places. Student Administration (STAD) reserves the locations for tests they schedule. If the test is an electronic one, a computer room must be booked.

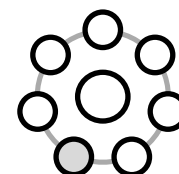
Occasionally, it can happen that a student does a test in a lecturer's room and then no invigilator will normally be present. In such cases the examiner or an assistant must ensure that the student can do the test without being disturbed, that the student doesn't use any unauthorised materials and that the answer script is collected on time.

Some programmes offer the opportunity to take distance tests, for example, to students who are doing a placement abroad or part-time students who are temporarily abroad for work. Programmes must have definite rules in place for the administration of tests abroad, as it is impossible for them to check whether the student is using any unauthorised aids. Also, not all types of test lend themselves for remote testing; in an open-book exam, for example, the chance that a student uses unauthorised aids is smaller than if the exam has many knowledge questions.

### **Electronic delivery of question-based tests**

Question-based tests can be administered electronically as well as on paper, but the logistics of electronic delivery place different demands than those of a paper-and-pencil test, notably:

- The test must be held in a computer room. These are smaller than ordinary rooms and fewer in number. In addition, planning should take into account that some computers may be out of order. This means that large classes may have to take the test in shifts or be divided over two or more rooms.
- Invigilators have more extensive duties: they must be computer-literate so that they can assist students who, for example, have problems logging in. They must also ensure that students don't do the test outside the computer room; in electronic summative tests, monitor passwords are used to prevent this. The invigilators are responsible for preventing or checking that monitor passwords are not passed on to students outside the exam room.
- Students who have to take the test must be known to the electronic assessment system. The list of students must be updated regularly. For security reasons, the students are linked to the test in the electronic assessment system so that only those students who are linked to the test can log in. This entails that the relevant procedures must be carried out strictly.
- Specific procedures have to be followed to prevent risks such as the leaking of exam questions, fraud and test delivery problems. To organise this properly, every school must have an electronic assessment coordinator (DTC, for *digitaal toetscoördinator*) who coordinates the preparation and delivery of electronic tests in their school.



## 5 Phase 5 – Assessment, Processing and Analysis

The choices made in the previous phase will determine what is important in the assessment phase. In formative tests, giving feedback is of particular importance whereas in question-based tests it is determining the test result: calculating the mark, or grade.<sup>7</sup> The following sections deal with those aspects of the various types of test that require particular attention, namely:

- Assessment of summative question-based tests: calculating grades (5.1);
- Psychometric analysis of question-based tests (5.2);
- Assessment of task-based tests: detecting plagiarism (5.3);
- Points for attention when assessing students: pitfalls (5.4);
- Inter-rater reliability (5.5);
- Feedback on formative and summative tests (5.6);

### 5.1 Calculating grades (and setting the cut-off score)<sup>8</sup>

Marking requires that criteria are set against which test scores can be compared so that student work can be awarded a grade. The ultimate criterion is whether a student about whom a final pass/fail decision has to be made will be a competent professional or not but this, of course, is not something the examiners know at the time they are assessing the student; so criteria, standards, have to be set. The examiners can choose between two fundamentally different approaches to standard setting, relative or absolute (*Toetsen in het hoger onderwijs*, 2006). Both methods are described in this section, and the steps the examiner must follow to arrive at a definitive grade.

- Step 1. Calculate the **score** for each script.
- Step 2. Establish the **cut-off score** (the score to be achieved to obtain a mark of 5.5).
- Step 3. Set the scores for the marks of **10** and **1** using a **conversion table**.

#### Step 1. Scoring scripts

Make a list of the students' names and record the item scores and the total score of each student. This scoring list will show which questions were answered well or not so well, how the students performed in particular subjects and how the scores are distributed among the students (is it a normal distribution, a bell-shaped curve, or are there many exceptionally high and/or low scores).

#### Step 2. Determining the cut-off score

Next, the examiner must determine the pass mark, the **cut-off score**. There are no objective criteria for this, but there is no lack of silly notions, like:

- 'One should have 25% fails' – as if it makes no difference whether the student has been working or learning well or not, or as if teaching can never be successful or, conversely, ineffective;
- '60% correct is a six' – as if you couldn't set higher or lower requirements, or as if each test has exactly the same level of difficulty.

Standards may be **absolute (criterion-referenced)** or **relative (norm-referenced)**, or a **compromise method** can be used to set the standard. Whichever method one chooses, there is no golden rule for determining 'true' competence, as any type of standard contains an element of arbitrariness (Cizek & Bunch, 2007).

#### Absolute standards

<sup>7</sup> The terms 'mark' and 'grade' have the same meaning in this Workbook.

<sup>8</sup> Couzijn, M. (2005). *Cijfers berekenen voor een toets*. Graduate School of Teaching and Learning, University of Amsterdam.

Using an absolute method to set the cut-off score (the score a student must obtain to pass the test) is determined before the test is administered. This is the method most commonly used in the Netherlands. In a test with open-ended questions the cut-off score is usually set at 55% to 60% of the maximum score. In multiple-choice tests one should first subtract a certain percentage from the total score to adjust for guessing and then calculate 55% or 60% (see Appendix 8.6.2).

Advantages of using an absolute standard:

- An absolute standard reflects the required level of knowledge/comprehension;
- It is simple and transparent; students know the requirements in advance.

Drawbacks of using an absolute standard:

- The (incidental) difficulty of a test is not taken into account;
- The pass rates for different tests may fluctuate;
- High pass or fail rates may occur;
- Passing a test may sometimes be a matter of chance which can adversely affect students' learning behaviour. Discouragement and a negative attitude to learning can lead to lower success rates and less progression.

When setting the cut-off score, one should take into consideration the required level of knowledge and the reasonable efforts that students must make to achieve this level. Making these choices requires expertise and experience. To set the cut-off score, choose a percentage which matches the importance of answering the questions correctly (to illustrate: in a particular test, a prospective surgeon may make fewer mistakes than a student training to be a nurse). In disciplines in which, for example, safety is not a key issue, one could take 50% as the cut-off score. By analysing the test (see Chapter 5.2) one can adjust for the degree of difficulty by setting the cut-off score higher for an easy test or lower for a more difficult one.

#### Relative standards

If a relative method is used to set the cut-off score, this score is determined after the test has been administered: for instance, the average performance of the whole group could be taken as the starting point. Needless to say, a relative standard also takes the degree of difficulty of the test into account. It is a precondition that all the students have prepared well for the test and that the test is not a resit. As it often happens that not all the students have prepared well, using a relative standard is considered less acceptable than using an absolute standard.

#### Compromise method

Compromise methods of standard-setting seek to combine the advantages of the absolute and relative methods. One advantage of compromise methods is that they take the degree of difficulty into account. An example is Cohen's method (Cohen-Schotanus, Van der Vleuten & Bender, 1996) which uses a fixed cut-off score and takes the highest score achieved as a reference score; the highest score being the average score of the five per cent of highest scoring test-takers. The five per cent of highest scores will be lower in a difficult test than in an easy one. A precondition for using Cohen's method is that at least one hundred students participate in the test. Also, the test must not be a resit, as 'resitters' are not representative of their year group (the best students don't participate). The same cut-off score is used in the resit as applied in the first sitting.

### **Step 3. Linking scores to marks using a conversion table**

Establish the scores that correspond to the marks of **10** and **1**. A number of arbitrary decisions have to be made when converting scores to marks. The fact that they are arbitrary means that the responsibility for these decisions rests on the examiner's shoulders and that he/she has to be aware of the nature and significance of his/her choices, and is accountable for them. As stated above, there are no objective criteria for making these choices. Another irrelevant opinion:

- 'Only a maximum score merits a 10' – as if only a perfect test result merits the qualification 'outstanding'.

Two examples of a test with twenty questions and a possible score of 0 to 20 points are illustrated below. They are borrowed from Michel Couzijn of the Graduate School of Teaching and Learning, University of Amsterdam (November 2005). In these examples a mark of 6 represents the pass/fail boundary, or the cut-off score. At Hanze University, a cut-off score of 5.5 is normally applied.

In the second example, the cut-off score is higher – 14 points required for a mark of 6 – but there is a slightly greater chance of getting a mark of 10. As you can see, there is a break in the line; the slope is less acute from mark 6 upwards. A higher mark is awarded for a good result (few errors) in this example than in the first one, and a lower mark for a poor result (many errors).

Points obtainable in the test	Example 1: Cut-off score: 12 points Mark of 10: 20 points Mark of 1: 2 points	Example 2: Cut-off score: 14 points Mark of 10: 19 points Mark of 1: 4 points
20	10	10
19	9.5	10
18	9	9.2
17	8.5	8.4
16	8	7.6
15	7.5	6.8
14	7	6
13	6.5	5.5
12	6	5
11	5.5	4.5
10	5	4
9	4.5	3.5
8	4	3
7	3.5	2.5
6	3	2
5	2.5	1.5
4	2	1
3	1.5	1
2	1	1
1	1	1
0	1	1

Figure 3.1 Converting scores to marks

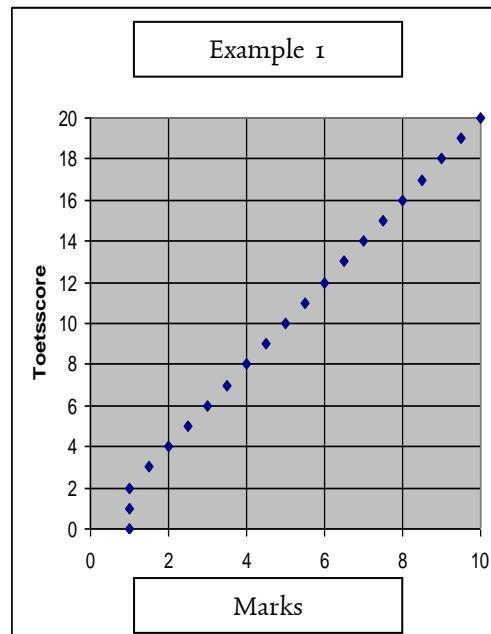


Figure 5.2  
Two errors =  
one mark off

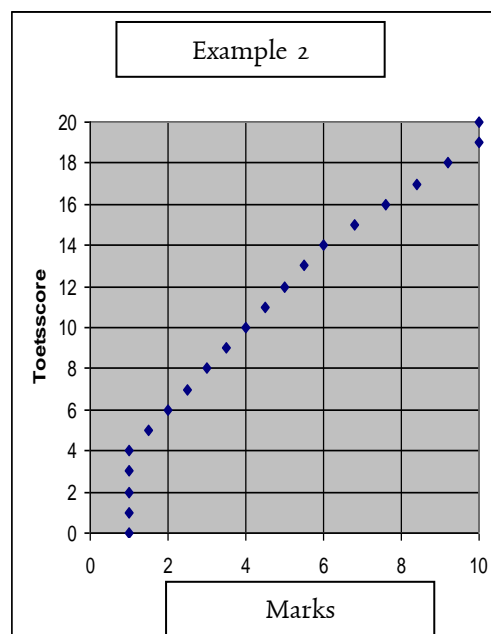


Figure 5.3  
Point range 14  
to 19: One error  
= 0.8 mark off  
Point range 4 to  
14: One error =  
0.5 mark off

One could also  
say:  
- 1<sup>st</sup> error: no  
deduction  
- 2<sup>nd</sup> to 6<sup>th</sup>  
errors: 0.8  
mark off  
- 7<sup>th</sup> error and  
onwards: 0.5  
mark off

## 5.2 Psychometric analysis of question-based tests

After a test has been administered and the raw scores for the individual items have been calculated, the test can be subjected to psychometric analysis. This will give the examiners insight into the quality of the test items and the test as a whole. Currently, psychometric analysis techniques only exist for tests with multiple choice questions and open-ended questions.

### Target values for P, *Rit* and Cronbach's Alpha <sup>9</sup>

To get an insight into the quality of a test, three quality indicators are calculated, namely:

1. The difficulty of the test item (p-value);
2. The discrimination value of the test item between students who master the subject matter and those who don't (*Rit* value);
3. The reliability of the test (coefficient  $\alpha$ ).

#### Item difficulty (p-value, or P)

The difficulty level of a closed-ended question is determined by the proportion of students who answered it correctly. This proportion, expressed as a percentage, is called the p-value. If a question is answered correctly by 36 out of 50 students, then the p-value is:  $P = 36/50 = 0.72$ .

A good target value for item difficulty is the maximum contribution of an item of average difficulty to the test as a whole. This average is the mid-point between the maximum and minimum item difficulty.

Number of options	Desirable p-value
2	0.75
3	0.67
4	0.63
5	0.60

Figure 5.4: Desirable p-values in a summative test with closed-ended questions

Very high p-values are often an indication that the questions are too easy which, of course, does not contribute to measuring differences. Likewise, very low p-values often indicate a misleading question or a question with an incorrect key. The p-values of the distracters can sometimes also provide insight into what has gone wrong, particularly if the p-values are close to or lower than the guess rate (the probability of the test-taker guessing the correct answer). If the p-values are near the guess rate, this may indicate that it was not possible for the students to answer the question correctly. Low p-values, however, that are above the guess rate may indicate difficult questions which, for example, are answered correctly by very good students but not by most of the others. The related *Rit* value (see below) may confirm this if it is positive. All the same, examiners must be careful not to include too many difficult questions in a test, because it can lead to high fail rates which is not always defensible.

#### Discrimination value (*Rit*)

Each test item should discriminate, as much as possible, between students with high or low total scores. In other words, it should have discriminatory power. To make discrimination possible, it is a necessary condition that not all the students give the same answer to the same question.

Discriminatory power is determined by correlating the score for the individual item with the total score for the test. Test questions that are more frequently answered correctly by high-scoring students than by low-scoring students clearly have discriminatory power, while questions that are answered correctly by an equal proportion of both groups don't. Items that are answered correctly more often by the low-scoring group than the high-scoring group may be discriminatory, but in precisely the wrong way. The resulting *Rit* values (*Rit* = item-total correlation; also known as point-biserial correlation, PBS) in these cases will be positive, zero or negative, respectively. The correlation coefficient has a range from -1 (minimum) to +1 (maximum).

---

<sup>9</sup> From: Analyses voor toetsen; Jan-Carel Vierbergen.

<b>Rit value</b>	<b>Interpretation</b>
> 0.35	good / very good
0.25 – 0.35	fair / good
0.15 – 0.25	questionable / fair
< 0.15	poor / questionable

*Figure 5.5: Interpretation of Rit values*

Items that discriminate well have a positive *Rit* value. If the *Rit* value is negative, this means that the high-scoring students in particular have answered the question wrong. The most obvious reason would be that the item key is incorrect or, for example, that one of the distracters is also correct. Imprecise wording or other flaws may also have put the students on the wrong track. In any case, questions with low *Rit* values deserve the examiner's attention. If necessary, questions can be revised and the scores adjusted (in the subsequent analysis).

#### Reliability co-efficient (co-efficient alpha)

The reliability of a test is determined by the stability of the test scores. Special methods, namely, KR-20 (Kuder-Richardson Formula 20) and Cronbach's Alpha (aka coefficient alpha) enable reliability to be expressed as a number. The length of the test plays an important role. As a rule, the more items there are, the more reliable the test. High reliability can usually be achieved in a test that has sixty items.

The value of alpha ranges from 0 (unreliable) to 1 (maximum reliability). If a test is unreliable, the test results have no significance. If a test is 100% reliable, this gives assurance that the test results are statistically significant. Whether the test results are also meaningful is not related to the reliability of the test, but must appear from analysis of the content validity of the test. However, reliability is a precondition for content validity.

<b><math>\alpha</math></b>	<b>Interpretation</b>
> 0.90	good / very good
0.80 – 0.90	fair / good
0.70 – 0.80	questionable / fair
< 0.70	poor / questionable

*Figure 5.6: Interpretation of coefficient alpha in summative tests (reliability)*

In some cases, a lower reliability than 0.80 is acceptable, for example, if the test score can be compensated by a score from another test (in which case 0.70 or higher is acceptable) or if the test is a formative assessment (in which case 0.60 or higher is acceptable).

The reliability of a test has great influence on the percentage of students who pass or fail a test unjustly. This percentage increases as test reliability decreases. When we refer to students who have passed or failed a test unjustly, we mean students whose results might have been different if they had taken a similar test on another occasion. The percentage of inconsistent decisions increases if test reliability decreases, and also if the percentage of failed students increases, where the percentage of failed students is determined by the cut-off score. Figure 5.6 illustrates the percentages of inconsistent decisions as a function of the percentage of failed students and reliability. It should be noted that using this table is only meaningful if the test scores are more or less normally distributed.

### Revision options / measures for improving qualitatively weak items

If distracters in a test are chosen in less than 2% of cases, these distracters can be considered as ‘dead’ and should be revised.

Characteristic	Revision
The p-value is lower or equal to the guess rate.	Is the key correct? If P is not lower or equal, amend the key and re-analyse. If P is lower or equal, remove the item from the test and re-analyse.
The p-value is higher than the guess rate, but considerably lower than the desirable value.	If the <i>Rit</i> is positive, keep the question in. If the <i>Rit</i> is negative, remove the item from the test and re-analyse.
The p-value is (almost) 1	Keep the question in.

<i>Rit</i> is 0 or lower.	Is the key correct? If not, amend the key and re-analyse. If yes, remove the item from the test and re-analyse.
<i>Rit</i> is positive but lower than 0.15	If the p-value is also low, remove the item from the test and re-analyse.
Alpha is lower than 0.70	The test must be considered as a formative one.
Alpha is higher than 0.70 but lower than 0.80	If the test can be compensated by other tests, this outcome is no cause for alarm.

Figure 5.7: Revision options based on statistical analysis (van Berkel, Bax 2002, p. 49 ff.)

### 5.3 Detecting plagiarism<sup>10</sup>

Plagiarism, as defined by the Van Dale dictionary, is the act of “copying another person’s work and passing it off as one’s own.” Even though one can loosely define the concept, it can still be difficult to judge where (acceptable) reference to common knowledge ends and (unacceptable) plagiarism begins. People in education know as well as anyone that it is impossible to write a text which is not largely based on other people’s knowledge, and that it is impossible to reference older works for every idea or line one puts on paper.

Interviews held with examiners at Hanze University revealed that they had very different opinions about what constitutes plagiarism, especially in relation to the use of researched materials. Some examiners regard it as proof of a student’s creativity and research skills if they find a lot of material on the Internet and uses it (perhaps even cuts and pastes it to) an essay or paper, while other examiners demand that a piece of work must show that the student has internalised the subject matter and that it should, therefore, contain only a minimum of quotes.

Generally in the academic world, it is a requirement that a text must evidence that the author has mastered the subject matter. Quotes should be used sparingly (and the sources accurately cited).<sup>11</sup>

Plagiarism can be misleading to different degrees, as the following examples illustrate.

- Parts of a student’s piece of work have been copied from other sources, whether paper or electronic, in particular from the Internet, and the student presents the work as their own work without acknowledgement of the sources. Whether the source texts have been copied verbatim or paraphrased, the student has not produced an original piece of work.

<sup>10</sup> Omgaan met plagiaat: Plagiaatbeleid en plagiaatdetectie bij de HG, 2005

<sup>11</sup> Oefening baart kunst: Onderzoeken, argumenteren en presenteren voor juristen / M.G. IJzermans & G.A.F.M. van Schaaijk.- Den Haag, Boom Juridische uitgevers, 2003



- A piece of work consists of (more or less literal) quotes from other sources. The sources are identified in a list of sources but the text itself doesn't clearly indicate (large) sections that have been more or less copied word for word.
- A piece of work written by someone else is copied as a whole and submitted by the student as if it is their own work, with a number of modifications; or a student has someone else write a paper and submits it as their own work.

Hanze University's plagiarism policy, [plagiaatbeleid van de Hanze](#)<sup>12</sup> contains guidelines for how to use other people's work, gives rules for how to act if plagiarism is suspected and compares several plagiarism detection systems.

Detailed information about prevention and detection of plagiarism can be found on the website: Toetsing HG, [Good Practices](#). It features a procedure for dealing with plagiarism that examiners must follow and a flow chart which explains what to do if plagiarism is suspected. If an examiner suspects a student of committing plagiarism or fraud, they must report this to the examining board which will take appropriate measures.

---

<sup>12</sup> <https://intranet.hanze.nl/organisatie/onderwijsenonderzoek/onderwijs/Pages/Formats%20oen%20formulieren.aspx>

## 5.4 Inter-rater reliability

Inter-rater reliability concerns the consistency of assessments given by different markers marking the same test. This has already been referred to above in the context of marking instructions (Phase 3). In this fifth phase, assurances for inter-rater reliability must be built into the cycle in the form of double marking. Double marking is a control measure which involves two examiners each verifying the same data separately to make sure that assessment is objective. Double marking is used especially in integrated assessments and also, in certain cases, in task-based tests such as project assignments or final assignments (placements, final projects). Programmes are free to make their own choices as long as they can guarantee that quality assurance measures are in place for assessment at the curriculum level. Inevitably, (inter-rater) reliability will not always be 100% on each test, this is inherent in competence-based teaching.

In the cases referred to above, the purpose of having a second examiner mark a test is to increase the reliability of the marking, and the examiners are interchangeable. Less frequently, double marking takes the expertise of the examiners into account. In this case, examiners with different backgrounds are sought out who bring their particular expertise to the assessment procedure so that a paper or other piece of work can be assessed from different perspectives. This will increase the intrinsic value of the assessment for the student.<sup>13</sup>

Assessment by two markers will also increase inter-rater reliability in question-based tests. It may seem to conflict with the principle that assessment should (also) be efficient, but there are smart ways of double marking (question-based) tests efficiently. For example, examiner A could mark the first half of the test items and examiner B the second half and then, when they are halfway through the scripts, do it the other way around. The examiners then discuss the borderline cases, which may result in adjustment of the marks.

Inter-rater reliability will also be higher if the examiner is alert to certain typical forms of assessment bias. In question-based tests, bias relating to a student's personality (sympathy, antipathy, stereotyping, discrimination) can be (partly) avoided by covering the name of the student when marking their script. A more delicate matter is, avoiding being influenced by one's conscious or unconscious opinions or predilections (see the Table below).

It is certainly more difficult to avoid such assessment pitfalls in task-based tests and integrated assessments, as the assessment criteria are often more difficult to measure in these types of assessment than in question-based tests, and the examiner knows the student they are assessing. For this reason, double marking is particularly important in these types of assessment.

---

<sup>13</sup> Moerkerke & Douchy (2000). *De toetsen van complexe vaardigheden*. Uit Dam, Van Hout, Terlouw & Willems (2000). *Onderwijskunde hoger onderwijs, handboek voor docenten*. Van Gorcum, Assen (p. 264).

<b>Assessment pitfalls related to personal characteristics of the student</b>	
Discrimination / stereotyping	Giving a lower mark than deserved because of unfair prejudices (race, sex, religious beliefs, disability). - “Girls are no good at abstract thinking.” - “She’s the only girl among all these male techies, she must be talented.”
Sympathy	Giving a higher mark than deserved because the examiner is sympathetic to the student rather than considering him or her competent.
Antipathy	Giving a lower mark than deserved because the examiner is unsympathetic to the student rather than considering him or her to be incompetent.
Halo effect	This occurs when an examiner assumes, on the basis of one piece of work, that the student’s other work will be equally as good or bad. In other words, a favourable impression based on a number of criteria is carried forward and positively influences the assessment of other criteria. - “This young man has been working so hard, he’ll make it.” - “Wow, this is such a good start to the presentation, the content is sure to be good as well.”
Horn effect	When an initially unfavourable impression based on a number of criteria is carried forward and negatively influences the assessment of other criteria. - “Those girls did nothing but talk through all the lectures, they’re bound to fail.” - “For heaven’s sake, why didn’t the fellow use any of my feedback about literature references?”
Prejudice	The assessment of the student is based on presuppositions about his or her characteristics founded on evidence from earlier experiences (a self-fulfilling prophecy).
Assessing development rather than capabilities	The examiner focuses too much on the student’s efforts or progress rather than their actual capabilities.

<b>Assessment pitfalls related to conscious or unconscious opinions about assessment</b>	
Projection	Attributing one’s own motives, characteristics or flaws to others. “This student enjoys writing as much as I do, you can see it at a glance. This paper will be a good one.” “My focus was on innovation, but there’s nothing in this work to evidence that.”
Leniency	The tendency to give overly high marks. - “The students had to complete a lot of assignments for this subject, this should be rewarded in some way.”
Severity	The tendency to give low marks. - “What this student has produced doesn’t reflect the required level and work attitude at all.”
Central tendency	The tendency to give marks in the middle range of the scale. - “Tens don’t exist, it’s all about sevens and eights.”
Contrast effect	This arises when the results of an assessment are determined by comparing a candidate with others, whether these performed well or not, instead of against standards.

## 5.5 Feedback on assessment

### Feedback in relation to the purpose of the assessment

Feedback and assessment are often mentioned in the same breath. Ideally, assessments not only provide information about a student's performance at a certain time but also yield feedback about what they need to do to develop to a higher level. Research consistently shows that feedback is a major influence on student learning (Hattie & Timperley, 2007; Kluger & DeNisi, 1996).

How and when to give feedback depends on the purpose of the test: is it formative (diagnostic, a 'practice test') or is it summative (discriminatory):<sup>14</sup>

- **Formative assessment:**  
The purpose of formative assessment is to provide feedback on student learning. Feedback supports students' professional development and is used to draw up or modify personal learning or development plans, improve students' performance and strengthen their intrinsic motivation. The key question is: what is the student able to do at this point, and what are they not yet able to?
- **Summative assessment:**  
Summative assessment means giving feedback of a different kind, namely, feeding the marks/grades achieved by students into the student records system. This feedback is used to inform formal decisions about the start, progress and end of a student's learning career at the university. Learning achievements that are judged to be satisfactory are rewarded in the form of certificates and diplomas. The key question here is: Is the student competent or not?

The feedback described above relating to summative assessment is actually too limited, because summative tests, besides being discriminatory, also have a diagnostic purpose. Moreover, assessment in a competence-based curriculum implies that examiners and lecturers regard tests as part of the learning process and that testing and assessment contribute to the development of a 'feedback culture' (HG Student Assessment Policy Document, 2004). Last but not least, the Student Charter provides that students have the right to view their exam scripts.

### Giving feedback

Examiners and lecturers have a number of options for how to provide feedback. First, the timing and level of the feedback should be distinguished:<sup>15</sup>

The time of giving feedback:

- After the test;
- During the test, immediately after a question has been answered.

The learning effect of feedback is greater if the students get it as soon as they have answered a question. This feedback doesn't necessarily mean giving the correct answer, but may consist of pointing the students in the right direction. Interim feedback is only provided in formative assessments because, in summative assessment, it would influence the test result. Feedback on summative tests is always provided afterwards, usually in the next phase of the test development cycle (Phase 6) when the students get the opportunity to view their scripts.

Levels of feedback:

- Feedback at test level: *Congratulations, you have passed this test.*
- Feedback at item level: *The correct answer to this question is.....*
- Feedback at answer level: *Alas, your answer was incorrect. The correct answer to this question is...*

---

<sup>14</sup> Edited text based on: Kwaliteit van toetsing in het Hoger Onderwijs, Fontys

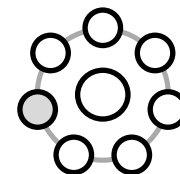
<sup>15</sup> Test Development Made Practical, Teelen Kennismanagement, 2011, Wilp

- Feedback at alternatives level: *Alas, the answer to ... is incorrect, because ... The correct answer to this question is ...*

The more detailed the feedback, the greater the learning effect on the student. However, formulating specific feedback at, say, the alternatives level is also a lot of work. In summative tests, it is advisable to give feedback at the item or answer level. In formative assessments, as it is essential that the student gains insight into their strengths and weaknesses, it is recommended to provide feedback at the answer or alternatives level, as much as possible.

Some suggestions for formulating feedback:

- Formulate the feedback in a positive, non-threatening way;
- Explain why the answer is correct, even if the candidate gave the correct answer;
- Don't include any new information in the feedback;
- Use illustrative images in the feedback if this seems expedient;
- Vary your choice of words in the feedback;
- For feedback at the alternatives level, state the correct answer if the student chose a wrong answer;
- Give feedback often for individual components ;
- Give specific, detailed feedback;
- Concentrate on the work, not the person;
- Give feedback in a timely manner; it's better to give it on time and be less accurate or comprehensive than to give it (too) late;
- Give feedback that is readily understandable.



## 6 Phase 6 – Recording and Communicating Results

In the sixth phase of the test development cycle, the test results established in the preceding phase are recorded in the student records system and announced to the students, and students are given the opportunity to view their scripts.

### Entering test results in ProgRESS

Test results must be entered in ProgRESS, the software programme used by Hanze University for student registration and records. This applies to all tests that have a ProgRESS code. Note that, when you enter results:

- The date the test was administered must be entered (not the scheduled date); this applies to all types of test apart. For task-based tests, the submission date applies;
- Only tests that have been marked may be recorded. Enter all the grades in ProgRESS: passes as well as fails (in connection with first sittings and resits);

After the results have been recorded in ProgRESS, the final grade lists are delivered to Student Administration (STAD). This process varies per school (for example, the examiner may send the grades to support staff (OBP) which then forwards them to STAD). Grade lists must be signed by the examiner.

### Announcement of results

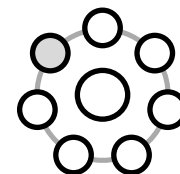
Test results must be announced to the students as soon as possible. The Student Charter prescribes a term of fifteen working days from the end of the test and a term of five working days before any resit. The results of oral tests should be announced on the same day the test is held, if possible (Student Charter, Art. 4.9.2).

### Viewing and checking

After the test results have been recorded, students have the right to view their scripts. This is important because it allows students to see what mistakes they made and what the correct answers are (feedback at the answer level). It will also help them when they prepare for the resit. Question-based tests are usually discussed in plenary sessions where, preferably, the examiner should be present. Depending on the size of the group and the students' requirements, the examiner or the lecturer will provide explanations and discuss commonly made errors and individual cases.

The viewing/feedback session is also an opportunity for the students to verify whether they received the right grade (have the points been added up correctly?). Make sure that students can't make any changes to their scripts during the session and that they don't take the scripts home. The programme is obliged to keep scripts for a certain period of time (refer to the HG retention periods). The Student Charter provides that students be given the opportunity to view their scripts within twenty (20) working days from the end of the relevant period and no later than five (5) working days before the resit (Student Charter, Art. 2.10).

An example of a protocol for the entry of grades is provided in Appendix 8.7. This protocol covers the whole process from collecting the scripts and marking them, right up to recording the final result in ProgRESS. The test viewing protocol of the Speech and Language Therapy programme is included in Appendix 8.8 as an example.



## 7 Phase 7 – Evaluation and improvements

Determination of the test results in Phase 6 concludes the programme component. In most cases, the component won't be offered again until the next academic year so there is enough time to evaluate and, if necessary, revise the curriculum and the assessment procedures. It will be wise not to delay this evaluation for too long, as it will be most valuable immediately after the end of the component or block, when the teaching and assessment are still fresh in the minds of all concerned – students, lecturers, examiners and block coordinators. Also, certain changes including changes to the types of test used may have to be recorded in the DOP (the digital course planner). This is particularly relevant for components in blocks 1 and 2, as DOP changes must be published ultimately in May of the preceding academic year. Curriculum overviews must be submitted in March.

Evaluation of assessment means that the examiner (or a group of examiners in a block team) review the phases of the test development cycle guided by a range of empirical data. The quality assurance policy and/or the assessment policy of a programme or school lays down what data must be available. The school or programme has latitude as to what data they collect and retain: data that has to be available for each test (e.g., pass rates) or data that has to be provided periodically (e.g., administering a block questionnaire every two years). The assessment committee often has a role in supplying data as well.

Evaluation of assessment is an essential part of the test development cycle and should always be conducted to complete the PDCA cycle. A 'good' pass rate is not a guarantee that the quality of assessment was good. Low or very high pass rates are always a reason for evaluating the assessment procedures thoroughly; this also applies to pass rates that exhibit a small standard deviation.

Some examples are given below of data that is commonly used for evaluating the quality of assessment. Using a variety of quantitative and qualitative data relating to or furnished by lecturers, examiners and students will increase the reliability of the evaluation.

### Measurement data of students (qualitative)

Students can give their opinions about the assessment and teaching in several ways (solicited or unsolicited):

- Evaluation form immediately after a test  
Some programmes require students to complete an evaluation form immediately after a test.
- Block questionnaire  
A standard questionnaire is available which includes questions about testing and assessment. One can always add extra questions. The block questionnaire is merely an 'alert tool' to expose any dissatisfaction among students. Students can be asked to take part in panel discussions, for instance, to discover the underlying reasons.
- Panel discussions  
Panel discussions are usually held after the block questionnaire has been administered. The moderator can ask the students why they rate this or that poorly, and ask them if they have any suggestions for improvement.
- Complaints to the examining board  
Students can file complaints about assessments with the examining board.

#### Measurement data relating to tests (quantitative and qualitative)

- Test analyses  
Test analysis has been dealt with in detail in Phase 5 (see Chapter 5). The outcomes can be used in the evaluation phase and for measures for improvement.
- Pass rates  
Pass rates can be retrieved from ProgRESS.
- Test participation rates  
Test participation rates can also be found in ProgRESS. Caveat: In some cases the pass rate may be good but only half the students may have taken the test.

#### Measurement data relating to teaching (quantitative)

- Cancellation of classes  
Cancellation of classes may be a contributory cause of low pass rates.
- Attendance of classes  
By keeping an attendance list, the examiner or lecturer can see whether a correlation exists between students' rate of attendance in classes and the test results they achieved.
- Active participation, e.g., by doing homework assignments.

#### Measurement data relating to examiners and lecturers (qualitative)

- Peer consultation  
Comparing and discussing the marks given for the same test by different examiners can lead to a common standpoint on the way a test should be assessed. 'What is the cut-off point? What is the difference between the marks of 5, 5.5, 6 and 6.5?'
- Evaluation by the block team  
Teaching at Hanze University being competence-based, the various components within a block are closely interrelated and so are the assessment methods. For this reason it is important to analyse whether all the tests together form a coherent set.  
A checklist for the evaluation of assessment in a block by a block team is provided in Appendix 8.9.

Evaluation may result in the improvement of test components and identification of weaknesses. It is advisable to keep a record of the evaluation process stating what has been measured, how it was measured and what conclusions were drawn. Based on these conclusions, an improvement plan can be drawn up which may engender new ideas for the design of tests or test items.



## 8 Appendices

### 8.1 Bloom's Taxonomy

#### *Three domains*

In the nineteen fifties the American educationalist Benjamin Bloom<sup>16</sup> developed taxonomies for three domains: the cognitive, the affective and the psychomotor. Of these, the taxonomy for the cognitive domain (Table 1) is the best known and is the one that is used most often. The taxonomy for the affective domain (Table 2) is applied when the learning outcomes concern matters such as attitude, feelings and values. The psychomotor domain (Table 3) relates to physical skills and is applied mainly in areas such as laboratory science subjects, health sciences, art, music, engineering, drama and physical education.

Bloom's work on the cognitive and affective domains was later revised by Anderson and Krathwohl, who changed the category names from nouns to verbs and switched/changed the Synthesis and Evaluation levels in the cognitive hierarchy to Evaluate and Create. Table 1 below is modelled on their revision.

Bloom did not work out the psychomotor domain in detail, and Table 3 below is based on the work of R.H. Dave.<sup>17</sup>

In each of the three tables, a number of verbs are linked to each level. These verbs can be used to formulate learning outcomes, assessment criteria and examination questions and assignments.

---

<sup>16</sup> Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: the classification of educational goals; Handbook I: Cognitive Domain* New York, Longmans, Green, 1956

<sup>17</sup> Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956) *Taxonomy of Educational Objectives, the Classification of Educational Goals. Handbook I: Cognitive Domain* New York: Longman.

Krathwohl, D.R., Bloom, B.S., and Masia, B.B. (1964) *Taxonomy of educational objectives: The classification of educational goals, Handbook II: The affective domain*. New York: David McKay.

Anderson, L. W. and Krathwohl, D. R. (eds.) (2001) *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Dave, R.H. (1970) "Psychomotor levels." In *Developing and Writing Educational Objectives*, R.J. Armstrong (ed.). Tucson AZ: Educational Innovators Press.

Table 1.1: Bloom's Taxonomy for the Cognitive Domain (as revised by Anderson/Krathwohl)

Cumulative development of knowledge according to Bloom's organisation into six categories			
	Category	Example	Illustrative verbs
1	<b>Remembering</b>	Ability to remember and recall (factual) information from a theory or model.	<b>Key words:</b> define, describe, enumerate, formulate, identify, list, look up, name, outline, recognise, reproduce
2	<b>Understanding</b>	Ability to understand something, describe a phenomenon in technical terms, follow instructions, state a problem in own words, explain a model or proposition from a theory.	<b>Key words:</b> convert, deduce, distinguish, estimate, explain, finish, give examples, indicate the difference or correspondence, interpret, predict consequences, summarise, understand
3	<b>Applying</b>	Ability to link a theory or model to a new, concrete (real-life) situation; doing something one has learned. E.g., applying a method or procedure correctly.	<b>Key words:</b> apply the theory to a case study, calculate, carry out, conclude, demonstrate, modify, predict, prepare, research, show, solve, use
4	<b>Analysing</b>	Ability to break down a problem or case into components or constituent elements, and, after critical reflection, determine the structure or coherence of these components or elements. Ability to distinguish facts from conclusions.	<b>Key words:</b> analyse, categorise, compare, contrast, deduce, differentiate, distinguish, explain, identify, illustrate, outline, relate
5	<b>Evaluating</b> <sup>18</sup>	Developing a reasoned opinion about, e.g., the significance of a certain idea, phenomenon, method or instrument. Ability to form and defend an opinion. Ability to judge the value of something in relation to a certain goal. The opinion is based on criteria (which may be formulated by the student him/herself).	<b>Key words:</b> account for, compare, criticise, conclude, convince, defend, explain, judge, justify, recommend, reflect, value
6	<b>Creating</b>	Ability to combine knowledge and understanding into something new. Starting from facts and models, or components, the ability to form a new whole, e.g., a new insight, a novel application or an advanced model.	<b>Key words:</b> combine, compose, create, design, develop, draft, make, modify, organise, plan, reconstruct, replace, revise, what if?, write a paper

<sup>18</sup> Evaluating one's own functioning (personal development) is a different kind of evaluation which is suitable, in particular, for all kinds of reflection assignments.

Table 1.2: Bloom's Taxonomy for the Affective Domain (as revised by Anderson/Krathwohl)

Cumulative development of attitude according to Bloom's organisation into five categories			
	Category	Example	Illustrative verbs
1	<b>Receiving</b>	Being receptive to information, willingness to hear, listening to others with respect, showing sensitivity to social problems.	<b>Key words:</b> ask, choose, give, follow, identify, listen, name, use
2	<b>Responding</b>	Active participation in the learning process, for example, by showing interest in a particular subject; being willing to do a presentation, taking part in discussions, helping others.	<b>Key words:</b> act, answer, comply, co-operate, discuss, organise, participate, report, support
3	<b>Valuing</b>	Having certain values and demonstrating involvement, for example, by believing in democratic procedures, recognising the value of science in daily life, being concerned for the well-being of others, being tolerant to individual and cultural differences.	<b>Key words:</b> accept, acknowledge, demonstrate, differentiate, discuss, explain, invite, involve, share, value
4	<b>Organisation</b>	Ability to bring values together, resolve conflicts between them, internalise certain standards and values, e.g., by taking responsibility for one's own behaviour.	<b>Key words:</b> act, arrange, assess, decide, explain, give meaning, initiate, judge, organise, reflect, relate
5	<b>Characterisation (internalising a value system)</b>	Having a set of standards and values, in terms of own beliefs, ideas and attitudes, which control one's behaviour (consistency, predictability), e.g., showing self-reliance when working independently, displaying professional commitment to ethical issues, ability to adapt to personal, social and emotional circumstances.	<b>Key words:</b> act, adhere, affirm, challenge, convince, demonstrate (a belief in), display, fulfil, influence, integrate, judge, perform, persevere, praise, propose, revise, stick to, value

Table 1.3: Taxonomy for the Psychomotor Domain (after Dave)

Cumulative development of physical (psychomotor) skills			
	Category	Example	Illustrative verbs
1	<b>Imitation</b>	Observing and imitating someone else's behaviour as a first step towards learning a more complex skill.	<b>Key words:</b> copy, follow, mimic, observe, repeat
2	<b>Manipulation</b>	Ability to perform certain actions and apply certain skills by following instructions.	<b>Key words:</b> administer, attach, build, construct, deliver, disassemble, dissect, gesture, heat, make, mix, calibrate, outline, stitch, use, weigh
3	<b>Precision</b>	Ability to perform a task with few errors, and refining it without needing the original source or instructions.	<b>Key words:</b> adjust, demonstrate, design, execute, fix, make, refine, specify
4	<b>Articulation</b>	Ability to perform a series of actions by combining two or more skills. Being able to modify patterns to fit specific circumstances or to solve specific problems.	<b>Key words:</b> alternate, apply, combine, differentiate, fix, handle, integrate, modify, treat
5	<b>Naturalisation</b>	Ability to perform at a high level naturally ('without thinking'). Performing a combination of skills in a logical sequence consistently with relative ease.	<b>Key words:</b> examine, grasp, handle, check, master (fully), perform skilfully, treat

## 8.2 Aid for Assessing Learning Outcomes

The checklist below can be used for giving feedback about learning outcomes to colleagues/peers.

	Points for attention	Feedback to colleagues, suggestions
	General	
1	For each study component, five to no more than eight learning outcomes (LOs) have been set.	
2	The LO is clearly related to the intended competencies in the block of study.	
	Individual learning outcomes: description	
3	The LO is concrete and specific as to what the student must be able to do at the end of the programme.	
4	The LO <u>describes the behaviour</u> clearly (an action verb and a noun)	
5	The LO describes the scope or context.	
6	The LO is related to Bloom's taxonomy correctly (the verb corresponds with the intended level).	
7	The LO clearly indicates an intended standard or level (degree of supervision, complexity of the assignment, etc.). In other words, it is evident whether the LO has been achieved or not.	
8	The LO is formulated clearly.	
	Individual learning outcomes: applicability	
9	The LO is assessable (measurable in an 'academic situation').	
10	The LO is plain to industry professionals.	
11	The LO is attainable in terms of the timeframe and the resources available.	

### 8.3 Types of Test

Question-Based Tests	
<p><b><u>Multiple-choice test</u></b></p> <p>Also referred to as:</p> <ul style="list-style-type: none"> <li>• <i>MC test</i></li> <li>• <i>MCT</i></li> <li>• <i>Written examination</i></li> </ul>	<p>A multiple-choice test is a test with closed questions and several answer options.</p> <p>MC tests are particularly suitable for assessing the lower order levels of cognition according to Bloom's taxonomy: remembering, understanding and applying knowledge.</p> <p>MC tests are an efficient tool for assessing knowledge as students can answer many questions in a relatively short time and marking can be done quickly. On the other hand, constructing an MC test – formulating the questions and the answers/alternatives – does take a lot of time.</p>
<p><b><u>Open-ended questions</u></b></p> <p>Also referred to as:</p> <ul style="list-style-type: none"> <li>• <i>Written examination</i></li> <li>• <i>Oral examination</i></li> </ul>	<p>Tests with open-ended questions are particularly suitable for assessing Bloom's higher order levels of knowledge and comprehension: applying, analysing, evaluating and creating. There are several types of open-ended questions.</p> <p><u>Fill-in-the-blanks and completion questions</u>: Completing an incomplete sentence, calculation or drawing. Reading and answering a question of this sort takes little time, no more than ten seconds.</p> <p><u>Short-answer questions</u>: The answer could be a quote, a phrase, a simple sentence (25 words max.), a number, a (simple) drawing or a formula. The time needed to read and answer the question normally doesn't exceed one minute.</p> <p><u>Long-answer questions</u>: These usually require a structured response which consists of several sentences. Alternatively, the answer required could be an explanation, a complex calculation, a drawing or a mathematical proof consisting of several steps. Reading and answering a long-answer question normally takes from five to ten minutes.</p> <p><u>Essay/argumentation questions</u>: The answer to an essay-type question typically involves writing a coherent story with a clear structure (introduction, body, conclusion) but may also consist of a very detailed drawing or calculation. Reading and answering an essay or argumentation question is time-consuming; fifteen to thirty minutes is not unusual.</p>
<p><b><u>Case study</u></b></p> <p>Also known as:</p> <ul style="list-style-type: none"> <li>• <i>OverAll Test</i></li> <li>• <i>Problem-based learning task, PBL task</i></li> <li>• <i>Applied knowledge</i></li> </ul>	<p>A case study presents a practice-based situation or a description of a real-world problem which is accompanied by a set of questions. Case studies are suitable for testing the higher cognitive skills of applying, analysing, evaluating and creating. Short and long cases studies are distinguished.</p>

<i>test</i>	
<b><u>OverAll Test</u></b>  Also known as: <ul style="list-style-type: none"> <li>• <i>Case study</i></li> <li>• <i>Applied knowledge test</i></li> </ul>	<p>This is a particular type of case study which tests the application of cross-curricular knowledge. It is good for testing the higher cognitive skills such as application, analysis, evaluating and creating, especially in practice-based problems that involve several areas of learning. OverAll Tests typically test integration of knowledge from several disciplines.</p> <p>Assessment is carried out on the basis of marking instructions in order to achieve a high degree of rater reliability.</p>
<b>Task-Based Tests (Assignments)</b>  An assignment used as an assessment tool is a structured instruction which leads to a professional product. Based on the product criteria it is determined whether, or to what extent the intended learning outcomes have been achieved. Assignments are appropriate for testing (integration of) knowledge, skills and attitude.	
<b><u>Assignment</u></b>  Other names: <ul style="list-style-type: none"> <li>• <i>Placement</i></li> <li>• <i>Final project</i></li> </ul>	<p>Assignments test (partial) professional competencies in an integrated way. The outcome of an assignment is a (written) report or a professional product. The product is chosen on the basis of the assessment plan and the learning objectives of the unit of study. A professional product is defined as a product which is used in actual professional practice.</p>
<b><u>Reflection assignment</u></b>  Also known as: <ul style="list-style-type: none"> <li>• <i>Process report</i></li> <li>• <i>Reflective account</i></li> <li>• <i>(Self) evaluation</i></li> <li>• <i>Peer coaching</i></li> <li>• <i>(Keeping a) logbook</i></li> <li>• <i>360° feedback</i></li> </ul>	<p>Reflection assignments test a student's ability to reflect on and (re)interpret their experiences and knowledge. They are appropriate for testing to what extent a student is able to stand back and view their actions in both academic and practice settings, and base choices for the future on their reflection.</p>
<b><u>Skills test</u></b>  Also known as: <ul style="list-style-type: none"> <li>• <i>Practical assignment</i></li> <li>• <i>Professional skills test</i></li> </ul>	<p>Skills tests assess skills that a student must possess to be able to perform well as a professional practitioner. Different types of skills are distinguished:</p> <ul style="list-style-type: none"> <li>- Thinking: cognitive skills;</li> <li>- Actions: motor or psychomotor skills;</li> <li>- Responding to people or situations: reactive skills;</li> <li>- Communicating with people: interactive skills.</li> </ul>
<b><u>Project assignment</u></b>  Also known as: <ul style="list-style-type: none"> <li>• <i>Project</i></li> <li>• <i>External or internal assignment</i></li> <li>• <i>'Co-makship'</i></li> <li>• <i>Professional</i></li> </ul>	<p>Project assignments are associated with project-based learning. On the basis of a completed product or service, a project assignment assesses integration of knowledge, skills and attitude in relation to pre-determined professional competencies.</p> <p>Assessment is carried out on the basis of marking instructions in order to achieve a high degree of rater reliability.</p>

<i>product</i> <ul style="list-style-type: none"> <li>• <i>Virtual business case</i></li> </ul>	
--	--



Integrated Assessments	
<p><b><u>Competence-based assessment</u></b></p> <p>Also known as:</p> <ul style="list-style-type: none"> <li>• <i>Behavioural assessment</i></li> <li>• <i>Peer assessment</i></li> <li>• <i>Performance assessment</i></li> <li>• <i>Expert assessment</i></li> </ul>	<p>An integrated assessment is a multi-dimensional test in which the student performs tasks in a critical setting and demonstrates their level of competence. An integrated assessment tests the extent to which the student has mastered professional competencies.</p> <p>Integrated assessments may consist of several complex tasks that are representative of the practice of a professional. They are concluded by an interview in which the student is asked about the choices they made when they performed the tasks. Integrated assessments increase in complexity stage by stage. They are different from portfolio assessments (see below) in that the student is put in a certain situation whereas, in a portfolio assessment, the student determines the setting. Assessment is carried out on the basis of marking instructions and double marking to assure a high degree of rater reliability.</p> <p>A variation is an assessment where a list of questions or assessment criteria are used to assess the student. Depending on who carries out the assessment, it may be referred to as self-assessment (the student assesses him/herself), co-assessment (the student and the examiner do), expert assessment (an external expert) or peer assessment (fellow students).</p> <p>Peer assessments can be used as an instrument/test type to assess an individual's performance in a group assignment. Various kinds of peer assessment and their advantages and disadvantages are described in the Good Practice Guide, 'Individual assessment in group assignments' (<i>Individuele beoordeling bij een groepsopdracht</i>).</p>
<p><b><u>Portfolio</u></b></p> <p>Also known as:</p> <ul style="list-style-type: none"> <li>• <i>Portfolio interview</i></li> <li>• <i>Portfolio assessment</i></li> </ul>	<p>A portfolio assessment is an oral evaluation of the work (evidence) collected by a student in a portfolio. Two types of portfolio are the <u>showcase portfolio</u> and the <u>progress portfolio</u>.</p> <p>A showcase portfolio demonstrates the student's (level of) mastery of competencies on the basis of the evidence in the portfolio.</p> <p>A progress portfolio demonstrates the student's progress over time and is substantiated by methodological and theoretical knowledge which is presented in the form of a reflective account.</p> <p>The portfolio interview is based on pre-determined criteria which are derived from the learning objectives. Preferably, the interview should be structured along the lines of a criterion-based interview (see below). Assessment and the interview are conducted on the basis of marking instructions and the four eyes principle in order to assure a high degree of rater reliability.</p>

<p><b><u>Competence-based Interview (CBI)</u></b></p> <p>Also known as:</p> <ul style="list-style-type: none"> <li>• <i>Criterion-based interview</i></li> <li>• <i>Final interview</i></li> <li>• <i>Behavioural interview</i></li> </ul>	<p>A competence-based interview (CBI) is a structured oral assessment based on pre-decided criteria whose purpose is to establish the extent to which a student masters competencies. Because the interview has a definite structure and clear assessment criteria, each student is assessed in the same way.</p> <p>CBIs are mainly used for assessing and predicting behaviour and mastery of competencies. The same kind of questions are asked for each competency, always in the same order in accordance with the STARR(T) model (see Appendix 8.6.4). Assessment and administration are carried out on the basis of the marking instructions and the four eyes principle in order to assure a high degree of rater reliability.</p>
--	--

## 8.4 Assessment Matrix Template for the Cognitive Domain

Learning outcome	Knowledge questions/ Reproduction	Comprehension questions / understanding	Application questions	Analytical questions	Evaluation questions	Creation / design questions	Points
<b>Total</b>							

## 8.5 Pre-Test Transparency Checklist

<b>Pre-Test Transparency Checklist</b>	<b>Achieved / outstanding</b>
Competencies including the level of the unit of study/ block	
Learning outcomes of the unit of study (or block)	
Assessment criteria for the unit of study	
Assessment plan for the block	
Assessment matrix for the unit of study	
Assessment form for the test (as regards assignments and integrated assessments)	
Aids allowed during the test (calculator, books, etc.)	
Examples of past examination papers: <ul style="list-style-type: none"> <li>• Sample test (complete test; sample questions are no indication of the test as a whole)</li> <li>• Video of example presentations and integrated assessments</li> <li>• Elaboration of answers</li> </ul>	
Description of the subject matter to be assessed, list of compulsory and recommended literature	
Description of the resit regulations; this is particularly important for (group) assignments	
The deadline for handing in the work (with regard to assignments).	
The information above can be found easily (DOP, Blackboard Learn).	
The information is available to the students on time (at the start of the block).	

## 8.6 Test Construction Directions

### 8.6.1 Test cover sheet

The form below is the standard cover sheet used in English-language examinations.

(Source: [Financieel Economische Zaken > Studentenadministratie > Voor medewerkers \(regels en documenten\)](#)  
[> Documenten tentamenorganisatie](#))

<b>Code:</b> Code	<b>Test:</b> Test											
<b>Date:</b> Date	<b>Time:</b> Time	<b>School:</b> School										
<b>Room:</b> Room	<b>Class:</b> Class	<b>Duration:</b> Hours										
<b>Lecturer :</b> Lecturer  <b>Can be contacted during the test on tel. no.:</b> Number		<b>Number of pages:</b> Number										
<table border="0"> <tr> <td><b>Aids:</b></td> <td><b>Further aids:</b></td> </tr> <tr> <td>Aid 1</td> <td>Further aids</td> </tr> <tr> <td>Aid 2</td> <td></td> </tr> <tr> <td>Aid 3</td> <td></td> </tr> <tr> <td>Aid 4</td> <td></td> </tr> </table>			<b>Aids:</b>	<b>Further aids:</b>	Aid 1	Further aids	Aid 2		Aid 3		Aid 4	
<b>Aids:</b>	<b>Further aids:</b>											
Aid 1	Further aids											
Aid 2												
Aid 3												
Aid 4												
<b>Handing in the test paper:</b> Test paper  <b>Handing in scribble-paper:</b> Scribble-paper												
<b>Particulars:</b> Particulars    												
<b>Student's name:</b>		<b>Student number:</b>										
<b>Class:</b>												

Cover sheet from STAD: essential information for test administration

### 8.6.2 Question-based tests

<b>Checklist for multiple-choice tests<sup>19</sup></b>	<b>Achieved / outstanding</b>
<b>Level</b>	
Does the test address all the intended learning outcomes? (Knowledge reproduction, comprehension and application)	
Have verbs been used that invite an answer at the intended cognitive level?	
<b>Language</b>	
The question is grammatically correct.	
The sentences are easy to understand (no subordinate clauses).	
The question does not contain any double negatives.	
Plain wording is used.	
The question is not phrased negatively unless with good reason.	
The question is phrased without ambiguity.	
There is no risk that the meaning of a question changes because of different stress placement.	
<b>Multiple-choice questions (stem)</b>	
The question contains enough information for it to be answerable.	
The information and the problem are distinct.	
The item format is appropriate for the learning outcome assessed.	
The question is not a trick question.	
The does not contain any unintended clues that point to the right answer.	
The does not contain any unintended clues that point to the right answer of another question.	
All the information provided is relevant for solving the question.	
The stem (the question) does not invite the student to give their opinion.	
The question does not contain any double negatives.	
The stem does not contain any superfluous information.	
No words like 'never', 'only', 'always' or 'all' are used.	
<b>Answer alternatives</b>	
The correct option is no longer or shorter than the alternative options.	
The options are in alphabetical order.	
All the options are consistent with the stem (logically or grammatically)	
No words that occur in the stem are repeated in the correct answer.	
The options are not mutually exclusive.	
The stem and the options are on the same page.	
<b>Use of contextual materials</b>	
Notes relating to figures are clear and succinct.	
All figures, diagrams, etc. are functional.	
Figures are clear and legible and have brief legends.	
References top figures/appendices are correct.	

<sup>19</sup> Based partly on: Berkel, H. & A. Bax, 2002. *Toetsen in het hoger onderwijs*. Bohn Stafleu Van Loghum (ISBN 9031336394).

<b>Presentation/layout</b>	
Questions and question components are clearly separated.	
The numbering of the questions is correct and transparent.	
The general conventions regarding the use of symbols, punctuation and spelling are observed.	
<b>The test as a whole</b>	
A multiple-choice test should contain at least 40 four-option questions, 60 three-option questions or 80 two-option questions. Don't mix two, three and four-option questions. If you are using case studies, devise several questions for each of them.	
The correct answers are distributed randomly between positions a, b, c and d.	
The items are independent of each other: are there any items that contain clues to other items?	
All parts of each time are on the same page.	
The degree of difficulty of the test as a whole meets the requirements.	
The test is a fair reflection of the learning objectives to be assessed.	
<b>Marking Instructions</b>	
<p>A (provisional) cut-off score has been set which adjusts for guessing.</p> <p>The guess score is the score which a student would on average get if they chose an answer for each item at random. It is important to adjust the cut-off score for the random guess score because, otherwise, there may be students who pass the test without having mastered the subject matter. This adjustment will result in a higher cut-off score. The formula used for determining the cut-off score is:</p> <p><i>Cut-off score based on the guess score = (maximum score – guess score) x cut-off percentage + guess score</i></p> <p>Example:</p> <p>The guess score in a test with 40 four-option questions is 10 (<math>40 \times 0.25 = 10</math>). If the cut-off score is set at 55% and is not adjusted for guessing (the questions carrying equal weight) the student has to get 22 questions right to pass. After adjusting the cut-off score for guessing, the candidate has to get 27 questions right to pass the test.</p> <p>Applying the formula, the calculation is: <math>(40-10) \times 0.55 + 10 = 16.5 + 10 = 26.5</math> (27, rounded off) and the cut-off score is <math>26.5 / 40 \times 100\% = 66.25\%</math> (66%, rounded off).</p>	
A score conversion table is provided.	
Guidelines for Assessors are provided.	
A correct answer is provided for each question.	
The correct answer is unequivocally correct.	
The correct answer is the only correct option.	

<b>Checklist for tests with open questions<sup>20</sup></b>	<b>Achieved / outstanding</b>
<b>Language</b>	
The question is grammatically correct.	
The sentences are easy to understand (no subordinate clauses).	
The question does not contain any double negatives.	
Plain wording is used.	
The question is not phrased in negative terms without good reason.	
The question is phrased without ambiguity.	
There is no risk that the meaning of a question changes because of different stress placement.	
The question does not contain any extraneous information.	
The way that the question is formulated cannot give rise to any misunderstandings.	
<b>Information</b>	
The question contains enough information for it to be answerable.	
The question indicates the desired length of the answer.	
The question indicates the desired form of the answer (drawing, description, etc.).	
It is evident from the question that the answer must be substantiated.	
The information and the problem are clearly separated.	
<b>Relevance and content validity</b>	
It is evident from the question what subject matter and/or skill is being assessed.	
The level of difficulty of the question is reasonable.	
The level of difficulty is not needlessly increased by irrelevant information.	
The question format is appropriate for the learning objective assessed.	
The question is not a trick question.	
The question does not contain any unintended clues that point to the right answer.	
All the information given is relevant for solving the question.	
<b>Use of contextual materials</b>	
Notes to figures are clear and succinct.	
All figures, diagrams, etc. are functional.	
Figures are clear and legible and have brief legends.	
References to figures/appendices are correct.	
<b>Presentation / layout</b>	
Questions and question components are clearly separated.	
The numbering of the questions is correct and transparent.	
Tables, diagrams, formulas, etc. have been checked for mistakes.	
At each question, the maximum number of points is stated that can be earned.	
The general conventions regarding the use of symbols, punctuation and spelling are observed.	

<sup>20</sup> Based partly on: Berkel, H. & A. Bax, 2002. *Toetsen in het hoger onderwijs*. Bohn Stafleu Van Loghum (ISBN 9031336394).



<b>The test as a whole</b>	
The level of difficulty of the test as a whole meets the requirements.	
The test is a fair reflection of the learning objectives to be assessed.	
Questions do not contain any unintended clues that point to the right answer of other questions.	
The test is a fair reflection of the learning objectives described in the assessment plan.	
<b>Marking Instructions</b>	
Model answers have been written for all the questions.	
The model answers are plausible.	
Partially correct and wrong answers are clearly specified.	
Clear general guidelines are included.	
Clear Guidelines for Assessors are included.	
The correct answer to each question is given.	
The marking instructions are not too general.	
The marking instructions are not so detailed that they are difficult to apply.	
The marking instructions are laid out so that the marker can quickly understand his/her tasks.	
An answer has been written for each question.	
This answer is correct.	
It is clear which elements must be mentioned, at minimum, for the answer to be marked correct.	
Answers about which doubts may exist as to whether they should be marked right or wrong, are specified.	
It is clear how many points must be awarded for each element of the answer.	

### Case studies

A case study consists of the description of a practical case and a number of associated questions which are used to assess the student's ability to make crucial decisions. Case studies can be long or short. Long case studies can be useful for diagnostic and formative purposes, such as analysing a complete case history and devising a plan of action. Generally speaking, long case studies will focus on the process of problem-solving. In formal examination settings, short case studies are preferable because of their greater reliability (more measurements) and validity (greater content specificity). In a short case study, the emphasis is usually on the outcome, the solution to the problem and the arguments given rather than on the problem-solving process (many paths lead to Rome). Different types of questions, both open and closed ones, can be used in case studies. The quality requirements appropriate to the type of question apply and the questions must follow logically from the case text.

Pointers for the construction of short case studies:<sup>21</sup>

#### *General*

- When constructing a case study, use an assessment matrix to ensure that the whole set of cases is representative of the teaching objectives.
- Gather realistic and authentic case materials by consulting people with hands-on experience.
- Submit the case to colleagues/experts, get their feedback and modify the case, if necessary, before you set it to the students.
- Keep the questions – and so, the marking scheme – as simple as possible.
- Be clear about the objective: is it the solution that matters or the problem-solving process?

#### *Constructing the case study*

- Use representations of real-life situations and avoid artificial situations.
- Make the description of the case as clear and concrete as possible.
- Ensure that there is a good balance between the quantity and relevance of (contextual) information.
- Present the information without any interpretation (raw).

#### *Writing the questions*

- Link the questions directly to the case.
- Make the questions focus on making critical decisions.
- Limit the number of decisions that must be made in a case.
- Phrase the question as precisely as possible without any ambiguity.
- The content of the question should determine the form, not the other way round.
- Provide unambiguous model answers which are defensibly correct.
- Focus the questions on the essentials and, where possible, limit the length of the answer to a minimum, for example, by pre-formatting the answer space.

Pointers for the construction of long case studies:

#### *Preparation*

- Determine the learning outcome: Are the students to analyse what the case is about or what the meaning of a certain concept is, or do they have to apply certain principles to information from the case, or are they required to solve a problem using the information from the case description?
- Decide the amount of time that the students have for the task, taking into consideration the importance of the intended learning outcome for the case study.
- Choose the subject or central problem and gather information. Brainstorm situations or examples of scenarios (real or fictional) which illustrate the concepts, principles or processes.
- Decide whether the students may work in groups or not.

---

<sup>21</sup> Schuwirth, L.W.T., 2002. In: *Toetsen in het hoger onderwijs*, van Berkel et.al. Bohn Stafleu van Loghum.

- Enhance realism. Gather documentation from, e.g., placement reports, annual reports, statistics, research reports, interviews or professional journals.

#### *Selection and organisation*

- Choose a single situation which seems promising.
- Select facts and events.
- Take the students' initial status into account (complexity, level of difficulty).

#### *Writing*

- Optionally, you could write role descriptions/create characters with feelings, attitudes, habits, expectations and viewpoints about the situation/problem.
- Don't use comical or stereotypical names.
- Develop the events, background information (time, place, setting) and actions.
- Present events in chronological order.
- Short dialogues can be useful so that the reader can hear what the characters themselves have to say.
- Adapt the length of the case to the task.
- Write plainly and concisely. Be concrete and realistic. Describe things objectively (facts only).

#### *Finally*

- Give a descriptive title to the case study.
- Ask a colleague to proof-read the draft case study (without supplying any notes) and ask them whether the situation described is comprehensible and clear.
- Use the comments to fine-tune the case study.
- Develop tasks which test students' problem-solving capabilities (process/product).
- Make sure that the presentation and layout is attractive and neat (instructions, task and case information are easy to distinguish).

Long case studies can also be used as take-home examinations. This could be in the form of an OverAll Test where the student can study the case text at home in preparation for the actual examination.

<b>Checklist for case studies</b>	<b>Achieved / outstanding</b>
Case text and questions are clearly separated from each other.	
The case information is to the point. It is:	
• unambiguous	
• only contains relevant information	
• worded clearly	
• well-organised and has a transparent layout	
• is attractive	
Valid and specific questions	
• There can be no misunderstanding about the nature or length of the answer.	
• Source references are sufficiently specific.	
• The problem is evident (or follows logically from the case).	
• The questions are formulated clearly and succinctly.	
• They are divided into sub-questions if appropriate (one question at a time).	
• The assessment criteria are indicated.	
• The model answer follows logically from the question.	
The question can be assessed objectively without ambiguity.	
Composition	
• The time available for the test is sufficient.	
• The questions are not interdependent (can be answered separately).	
• The level of difficulty is reasonable (borderline students can achieve a score of 55%).	

<b>Checklist for OverAll Tests</b>	<b>Achieved / outstanding</b>
<b>Articles / problem situations</b>	
The test is based on different problem situations (presented in articles).	
The problem situations are authentic, which means that they represent or resemble real-world situations.	
The problem situations are new to the students.	
The problem situations are approached from the viewpoints of several disciplines.	
The articles are written in language that is accessible to the students.	
The problem situations described match the core objectives of the unit of study.	
The article presents a practical application of core teaching content of the unit of study, or is a theoretical analysis which to some extent contradicts or supplements the content of the unit of study.	
The author presents his ideas lucidly so that students cannot interpret them in any way other than intended.	
The problem situations provide sufficient leads for formulating a set of questions.	
Students can study and analyse the problem situations within the time allotted for the component.	
<b>Studying guidelines</b>	
The studying guidelines advise the students about what to focus on in their preparation.	
<b>Test questions</b>	
The test questions address the core aspects of the problem situation.	
All the questions can be answered within the time allotted for the test.	
The question CANNOT be answered without reading or referring to the article.	
The question CANNOT be answered by reference to the article alone.	
The answer CANNOT be found directly in the article or the compulsory literature of the unit of study.	
The questions correspond to the studying guidelines.	
The set of questions for each unit of study is representative of the core content of the unit.	
<b>Model answers</b>	
For each correct / ? / incorrect question there is/are:	
an answer key	
literature references	
in the case of an incorrect answer, an explanation (at minimum)	
A model answer is provided for each open-ended question.	
The model answer covers all the elements of the question.	
If several correct answers are possible to a question, the assessment criteria are stated (what distinguishes a perfect answer from a partially correct or wrong answer?).	

### 8.6.3 Task-based Tests

#### Reflection assignments

You may find using the ABC model devised by Fred Korthagen (University of Utrecht, IVLOS) helpful when designing reflection assignments. It consists of three questions which capture the main elements of reflection:

- A. What was it **A**bout, from your point of view?
- B. What is its **B**earing on you, what was important about this?
- C. What **C**onclusions can you draw from this and what does this mean in terms of your intentions?

This may be complemented by the addition of a question between B and C, namely:  
How would a professional have tackled that problem?

#### Project assignments

Project assignments are composed of four basic elements:<sup>22</sup>

- the case;
- the client;
- the task/assignment;
- the product to be produced and the related specifications.

These four elements together must be of sufficient quality to be purposeful.

---

<sup>22</sup> Bie, D. de en J.J. Gerritse. *Onderwijs als opdracht*. Houten, Bohn Stafleu van Loghum, 1999

<b>Checklist for Project Assignments</b>	<b>Achieved / outstanding</b>
<b>General</b>	
The task(s) are clearly formulated.	
The tasks reflect the learning outcomes.	
It is clear what the students have to do, the final product is explicitly defined.	
<b>Assignment content</b>	
The case: The case is tangled and rich in context, the story line is as realistic as possible. The problem is indicated in a more or less casual way.	
The client and the assignment: The client regards a specific professional product as the best solution to the case problem.  The students have well-defined roles, also in a simulation (e.g., employee of an engineering firm, a senior member of staff, a self-employed professional, an employee of a consultancy).	
The product and its specifications: The client provides realistic specifications (quality requirements, product requirements, design requirements, a schedule of requirements).	
<b>Assessment Form</b>	
The assessment form clearly indicates what the final grade is composed from.	
The individual component is specified in the final grade.	
The assessors are identified.	
Assessment is based on the four eyes principle.	
The submission dates for the first sitting and the resit are stated.	
The resit regulations are described in clear terms.	
The pass criteria are evident from the assessment form.	

## 8.6.4 Competence-based assessments

### (Behavioural) assessment

<b>Checklist for (behavioural) assessments</b>	<b>Achieved / outstanding</b>
<b>Developing the assessment (ready at the start of the study component)</b>	
Clear guidelines for students are provided before the assessment which inform them of the objectives, learning outcomes, method, assessment criteria and procedure. For some components it will be sensible to advise the students about how to prepare and practice.	
(Optional) Experts from outside the programme are involved in developing the assessment and advise about suitable assessment methods and tools.	
The assessment components are based on critical situations and professional roles.	
All the learning outcomes of the assessment are assessed.	
To increase reliability, training sessions are organised for examiners in order to fine-tune the use of the assessment instruments and criteria through inter-subjective assessment.	
<b>Organisation</b>	
Administration of the assessment is properly planned and organised. Points for attention:	
• roster	
• timetable	
• booking and setting up video equipment	
• hiring actors	
• reception room for examiners	
• availability of assessment forms	
• first and second assessors are known	
• roles of the first and second assessors have been discussed	
<b>Assessment / Assessment form</b>	
The assessment instruments and criteria have been discussed with the fellow examiners and any experts from outside the programme to ensure the reliability of the assessment.	
The weighting of the various components has been agreed.	
The 'good', 'pass' and 'fail' criteria have been agreed.	
It has been agreed whether examiners may intervene during the assessment, in other words, may they give students a second chance to demonstrate the desired behaviour or not.	
The assessment instruments (assessment forms) are filled in during the assessment session.	
It is clear when the result of the assessment will be released: immediately on the spot or after the assessors' meeting (to allow adjustment of the cut-off score).	



## Portfolio

<b>Portfolio Checklist</b>	Achieved / outstanding
Clear guidelines are given to the students explaining the purpose, the assessment criteria, the resit procedure, the organisational method and the conditions for participation in the portfolio assessment.	
Working with portfolios is firmly embedded in the teaching methodology (usually, portfolios are used across subjects).	
In a portfolio assessment, all the learning outcomes for a particular study component are assessed. If students are able to formulate learning outcomes themselves, it is clear how they came about and how they are recorded.	
The students are clear about what types of evidence they can include in their portfolio and what freedom they have in this respect (the degree of standardisation is evident).	
It is clear whether the portfolio is a private or public document.	
It is clear what supervision students receive during the compilation of their portfolios.	
The students are given an assignment which is unambiguous and states the preconditions, objectives and assessment criteria.	

### Assessment criteria and procedures for portfolios

One difficulty with assessing portfolios is that the guidelines that are available for evaluating students' work in standardised assessment situations cannot be applied as readily. Portfolios can vary widely in content, which makes it difficult for examiners to compare them. Standardising portfolios in terms of content, size and structure will lead to greater reliability of the assessment but, at the same time, can also discourage students from reflecting on their development.

The following factors contribute to making assessment of portfolios reliable:

<b>Training</b> The assessors should be well-prepared and receive training so that they have a common understanding of the criteria and the procedures to be followed.
<b>Number of examiners</b> Employ only a small group of examiners – fewer than ten – to assess portfolios. This allows them to discuss the criteria with one another and decide how to apply them during their assessments.
<b>Standardise the criteria</b> Set generic criteria to make it possible to compare performances across different tasks. Standardisation according to subject to make portfolios comparable is incompatible with the student-oriented nature of portfolios.
<b>Assessment rubrics</b> Use assessment rubrics to describe the levels to be reached in each competency. The students and the examiner can draw up the rubrics together. The information provided in the rubric is helpful for the students while they prepare their portfolios and also for the assessor who can use it afterwards to inform the students about the level achieved for each competency.
<b>Focus on outcomes, not activities</b> Ensure that portfolios focus on objectives – learning outcomes and competencies – and not on activities and tasks, as is often the case in practice. At the same time, the portfolio must be consistent with the programme situation. In this way, it is easier to assess and also clearer for the student to identify whether their portfolio contains sufficient evidence and whether the evidence is of good quality.
<b>'Hawks' and 'doves'</b> Neither 'hawks' (severe assessors) nor 'doves' (lenient assessors) should be charged with assessing portfolios. Inter-rater agreement increases sharply if assessors who are consistently extreme in their

judgements are excluded.
<b>Peer and co-assessment</b> Involve fellow students in the assessment so that students can learn from each other's mistakes and achievements.
<b>Authenticity</b> The assessor must be able to check that the evidence in the portfolio is authentic, that is, produced by the student him/herself.
<b>Analytical or holistic assessment</b> It is advisable to choose a middle path between holistic and analytical assessment. Practically, this means you should give a single total score but that you should break it down into assessments of the various partial competencies which you then use to substantiate the final assessment given; the (weighted) sum of the individual partial competencies will thus not be required.

### Criterion-based interview (CBI)

Checklist for Criterion-based Interviews (CBI)	Achieved / outstanding
The interview is structured.	
What competencies will be tested at what level is decided in advance.	
The same kinds of questions are asked for each competency (criterion), always in the same order.	
The questions are structured, for example, by using the <b>STARR(T) model</b> <b>S</b> = Situation – in which the student has demonstrated the competency <b>T</b> = Task – the student's input in the given situation <b>A</b> = Action – which the student performed in the situation <b>R</b> = Result – the effect of the action <b>R</b> = Reflection – on the actions in the situation <b>(T</b> = Transfer – application in different kinds of situations)	
A more or less identical questionnaire is used each time.	
The time frame is strict, because all the aspects have to be addressed.	
Tasks are divided within the team of assessors (asking the questions, writing down the answers, keeping time).	

### Sources:

- Berkel, H. van, en Bax, A. (red). *Toetsen in het Hoger Onderwijs*. Houten: Bohn Stafleu van Loghum, 2006
- *Handboek toetsplanontwikkeling* Digitale Universiteit, 2004
- *Toetsnotitie - implementatie toetsbeleid en kwaliteitszorg*. Hanzehogeschool Groningen, 2004
- brochure *Anders opleiden, anders toetsen* (2001), deel 1 in de [BVE-brochurereeks 'Perspectief op Assessment'](#), een uitgave van [Cito unit BVE](#).
- *Toetsvormen*. Toetscommissie SIRE, Hanzehogeschool Groningen, 2012
- *Toetsvormenboek*. Fontys, 2001

### 8.6.5 Assessment Form Checklist

An assessment form is required for task-based tests and integrated assessments.

	Points for attention	Feedback from colleagues
	Correspondence with the unit of study / the learning outcomes	
1	The criteria on the assessment form are representative of the learning outcomes and the intended level.	
2	The criteria are related to professional practice in a recognisable way, if possible.	
3	The criteria are distinct from the criteria applicable to other units of study in the same block.	
	Usage of the assessment form	
4	The form includes a description of the assessment process and how the form has to be filled in (marking instructions).	
5	The names of the markers are written on the assessment form (or there is a field for their names).	
6	There is a field on the form for the name(s) of the student(s).	
7	All the criteria listed on the assessment form are concrete and specific (in other words, there can be no mistake about whether the student has fulfilled the criterion).	
8	For each criterion, the number of points that can be earned are stated (or, at any rate, how they contribute to the total mark).	
9	The assessment form includes an explanation of how the mark is calculated (or whether pass/fail is applied).	
10	The assessment form clearly states the student's <u>individual</u> final mark.	
11	The assessment form is well organised and has a clear layout.	
12	The assessment form is/seems easy to use (also by less experienced examiners).	
	Clarity of the marking scheme (this can be included in the form, but is sometimes separate if that is more convenient)	
13	If necessary, minimum requirements are formulated which the student must comply with to achieve a pass mark (e.g., critical or conditional criteria / what qualifies as a six?).	
14	If several assessors are involved, the method for determining the final mark is set out clearly.	
15	The marking method for group assessments which contain an individual component is explained clearly.	

### 8.6.6 Sample Examination Instructions<sup>23</sup>

#### Examination: Introduction to the Theory of Education, June 1998

##### Instructions

In a moment, the exam paper will be handed out to you. Before you start to answer questions, read the instructions below carefully. You have 2½ hours to complete the examination. You will receive three warnings towards the end of the exam period advising you that the time to hand in your script is approaching.

##### Answering instructions

- For the fifty multiple-choice questions, circle the option (A, B or C) which you feel best answers the question on the question sheet. Then transfer the answer you have selected onto the OMR sheet. As you can see, there is a column on the OMR sheet with numbers that correspond to the numbers of the multiple-choice questions, and three other columns where you write your answer. If you have chosen answer 'A', place a horizontal line between the vertical lines in the first column, if it is 'B' or 'C', then place a line in the appropriate (second or third) column.
- Answer the question in any case. Wrong guesses are not penalised.
- Use an HB pencil to place the line. If you want to change your answer, completely erase the line you first placed. If you have forgotten to take a pencil and rubber (eraser), get them NOW from the invigilator.
- The second part of the examination consists of twelve open-ended questions. Write your answer in the response field immediately below the question. The questions have been formulated in such a way that there is sufficient space. DON'T write in the margins or on the back of the sheet.

##### Scoring and result

- One point is awarded for each correct answer to a multiple-choice question. The twelve open-ended questions are worth five points each, so sixty points in all. The maximum score achievable, therefore, is 110 points.
- In assessing your results the results achieved by your fellow students are taken into account, but you can rest assured that if you have earned a score of sixty points you have achieved at least a pass.
- Please write your answers inside the designated spaces. Anything written outside these spaces cannot be read by the marker and will not be considered to be part of your answer.
- Points will be deducted for errors in your answer even if you have also given the correct answer, but the minimum score for this answer will never be negative.

##### General

- You are NOT allowed to use any aids during the examination such as calculators, dictionaries, notes, etc.
- You are NOT allowed to communicate with fellow students in any way; this includes exchanging pencils, paper, rubbers or pens.
- The use of a mobile phone during the examination is not permitted.
- Write your name and student number on EACH page of the question paper and the OMR sheet.
- Check the page numbering of the examination booklet. If any of the pages are missing, raise your hand now and ask for a new copy.
- The Examining Board wishes to inform you that the rules on cheating contained in the Examination Regulations apply to this examination.
- If you want to ask a question, raise your hand.
- You are not allowed to leave the room for the first thirty minutes of the examination; this includes toilet visits.
- If you need to visit the toilet, raise your hand and ask the invigilator for permission.
- When you have finished the exam, hand in your script to the invigilator and leave the room.

---

<sup>23</sup> Taken from, almost integrally: Berkel, H. van, en Bax, A. (red). *Toetsen in het Hoger Onderwijs*. Houten: Bohn Stafleu van Loghum, 2006

- After the exam, you will be asked to hand in a short evaluation form outside the examination room to the invigilator.



## 8.7 Protocol for Recording Grades

Day	What	Who	Particulars
1	The Front Office collects the exam scripts from STAD.	Front Office	<ul style="list-style-type: none"> <li>Empty the cabinet at STAD daily.</li> </ul>
1	Exam scripts completed by students in special circumstances are added. Scripts relating to the same subject which are received later are handed over in person to the examiner named on the exam timetable. Any irregularities are reported to a member of the Examining Board	Front Office	
1, 2	Receipt of the script is ticked off on the exam timetable.	Front Office	<ul style="list-style-type: none"> <li>Check for late scripts.</li> <li>Report irregularities to the Team Manager.</li> </ul>
1, 2	Examiners are notified by email that the scripts are available for collection.	Front Office	<ul style="list-style-type: none"> <li>Tick off: forwarding emails to the subject coordinator and the examiner.</li> </ul>
1, 2	Collection of the scripts. The date of receipt and the examiner's name are ticked off on the exam timetable.	Examiners, Front Office	<ul style="list-style-type: none"> <li>Check that exam scripts are collected.</li> <li>Report irregularities to the Team Manager.</li> </ul>
2 to 11	Marking by examiners.	Examiners	<ul style="list-style-type: none"> <li>Examiners mark the scripts.</li> <li>Examiners submit the class scores / an error analysis to the Department.</li> </ul>
12, 13	Analysis of the results by the subject coordinator; consultation with examiners, if necessary.	Subject coordinator / examiners	<ul style="list-style-type: none"> <li>The Department consult about adjusting the standards and determine the definitive standards.</li> </ul>
14	The final grades are determined by the examiners.	Examiners	<ul style="list-style-type: none"> <li>The examiner adjusts the grade in accordance with the adjusted standards, and enters it in ProgRESS WWW.</li> <li>The examiner sends the grade sheets to the Front Office within fifteen (15) working days.</li> </ul>
Up to day 15	Signed ProgRESS list (grade sheet) is copied, archived and taken to STAD.	Front Office	<ul style="list-style-type: none"> <li>Grade sheets submitted are taken to STAD at the end of the day.</li> <li>Processing time by STAD is one (1) working day maximum.</li> <li>Comments about additional / non-curricular components (<i>Du: NIP-opmerkingen</i>) are dealt with within three (3) working days (STAD)</li> </ul>

16	Check whether the grades have been entered into ProgRESS in time.	Programme support staff	<ul style="list-style-type: none"> <li>• The test result is determined and announced to the student as soon as possible, but no later than fifteen working days after the examination date and no later than five working days before any resit. Unless the Examining Board determines otherwise, the result of an oral examination is published on the day that the examination was held (Art. 4.9.2 of the Student Charter).</li> <li>• If an examiner is late, he/she is sent an email by the Team Manager.</li> <li>• STAD must process grade sheets in time to enable them to be checked via ProgRESS WWW.</li> </ul>
17	Request unprocessed grade sheets from STAD.	Examining Board member	

NB: Checking/submission of the grade sheets:

- The Front Office registers the name of **each examiner** who collects an exam script.
- The Front Office registers the name of **each examiner** who hands in a grade sheet, and when it is handed in.
- If an examiner still hasn't submitted a grade sheet of the scripts which he/she collected by the fifteenth working day after the examination, then the Front Office will inform the Team Manager of this by email.
- The Team Manager then requires the examiner to hand in the grade sheet and the grade that same day. The Front Office takes the grade sheet to STAD so that it can be processed, all on the same day. As this must all be done in a single day, it must be explicitly agreed with STAD.



## **8.8 Protocol for Viewing Exam Scripts in the Speech Therapy Programme**

15 March 2012

The following procedures should be observed when students get the opportunity to view their examination scripts, to prevent fraudulent practices.

What are the risks?

- Large numbers of students
- Just one lecturer to oversee everything
- Photographing of exam scripts
- Removal of scripts
- Copying exam questions
- Writing questions and answers on the same scripts (open-ended questions)

Before the viewing (planning):

- Scheduling a central viewing may not be the best way to organise viewings. It could also be done through tutors. Another possibility is not to schedule any viewings but give students the opportunity to view their scripts during office hours or by appointment.
- Plan the viewing at the beginning or end of the day so that only students who are genuinely interested in the content show up, and to discourage students who are only out to get a higher grade.
- Book a suitable room; not a computer room with workstation partitions as this will make supervision difficult.
- Have two lecturers attend: an expert (the programme owner) and a supervisor (this could also be a support staff member).

During the viewing session (execution):

- Bags and mobile phones must be left in a corner of the room.
- The students may not take notes.
- There are several options for discussing the questions:
  1. Digital projection of the exam questions and dealing with them one by one (without handing all the questions out on paper).
  2. Discussing only those questions that the students have indicated.
  3. Signing up. The lecturer only takes along materials for the students who have signed up for the viewing. The question paper is given a number and handed out together with the UOCG form. The lecturer shows the answer on a screen and the students view the answers in silence. Students can then ask questions individually. All the materials must be returned. Checking is easy because the numbers have been recorded.
- Agree the division of tasks between the two people involved: the programme owner discusses the questions, the other person supervises the session.

After the session:

- Collect and return any materials used.
- Shut down the computer or close the file.

Miscellaneous:

- Don't let students note any answers on the examination paper but on separate exam sheets. This is part of the examination production process.
- Follow-up steps in the event of irregularities: Any irregularities must be reported to the Examining Board which will decide whether follow-up steps should be taken.

Students who didn't sit the examination are not allowed to attend the review session.

## 8.9 Checklist for the Evaluation of Thematic Block Tests with a Block Team

The checklist below from the School of Financial and Economic Management (FEM) supports the analysis of the whole assessment procedure.

**Block:** [block number and name]

**Block coordinator:** [name]

**Date:** 7-10-2013

**Completed by:** [name of block team member]

### Instructions for completing the questionnaire

Please complete the questionnaire in digital form as follows. For the Yes/Partly/No questions, leave the appropriate option untouched and delete the other options. Please type the answers to open-ended questions into the designated space.

The checklist is based on the report, Effective Assessment at FEM (*Goed toetsen bij FEM*). All the questions in the checklist are derived from the assessment policy which the FEM has formulated in this report. The block coordinator asks all the lecturers to complete the checklist for their specialist subjects and then compiles a single completed checklist for the block. Differences between subjects can be noted in the columns by specifying the subjects below the 'Yes/Partly/No' cells. Such information may also be included under 'Notes by the block coordinator'.

Example:

x. Assessment is efficient, tests with pre-set answers are used where possible (e.g., multiple choice).	Yes	Partly	No
---	-----	--------	----

**On behalf of the FEM Assessment Committee, we thank you in advance for completing this checklist.**

### ASSESSMENT PROCEDURE IN GENERAL

<b>Organisation</b>			
1. The assessment planning has been announced to students clearly and fully (feasibility). (Blackboard: examination timetable, submission dates, referral regulations, contact person for questions)	Yes	Partly	No
2. The assessment content has been announced to students clearly and fully (transparency). (Blackboard, per test: assessment frameworks, assessment criteria, workload, sample examination/trial test and answers)	Yes	Partly	No
3. The students know in advance what assessment methods will be used.	Yes	Partly	No
4. The assessment methods are in compliance with the Study Guide/Teaching and Examination Regulations and state the correct ProgRESS codes.	Yes	Partly	No
5. The assessment planning takes into consideration the proportion of time an examiner spends on assessment and other educational activities such as preparation, contact hours, giving feedback between times, etc. (Also consider matters such as keeping 'shadow records' of student work, organising resits, etc.)	Yes	Partly	No
Notes by the block coordinator:			
Comments by the Assessment Committee:			
<b>Efficiency</b>			
6. The assessment is efficient, tests with pre-set answers are used where possible (e.g., multiple choice).	Yes	Partly	No

7. The ratio of time spent on assessment to time spent on other activities is satisfactory in this block.	Yes	Partly	No
8. An bonus scheme is in place for one or more subjects.	Yes	Partly	No
If so, please state the amount of the bonus:			
Notes by the block coordinator:			
Comments by the Assessment Committee:			
<b>Content</b>			
9. All the competencies and/or learning outcomes that are (or should be) addressed in this block are assessed. (There are overviews or matrices to prove this.)	Yes	Partly	No
10. A student who has successfully completed all the units of study of the block has mastered all the competencies or achieved all the learning objectives that were assessed.	Yes	Partly	No
11. Assessment criteria have been set for each assessment/test and they are linked to competencies/learning objectives.	Yes	Partly	No
12. The assessment criteria are clear and unambiguous.	Yes	Partly	No
Notes by the block coordinator:			
Comments by the Assessment Committee:			
<b>Types of tests</b>			
13. The test types used are varied. (At least three different types of test are used in each block.)	Yes	Partly	No
14. The test types are consistent with the various strands of learning (skills, theory, integrated assignment, academic advising).	Yes	Partly	No
15. Each study component includes an element of individual assessment.	Yes	Partly	No
16. Where possible, diagnostic assessment is applied so that students gain an insight during the block into where they stand and what more they need to do to achieve the required level of competence.	Yes	Partly	No
Notes by the block coordinator:			
Comments by the Assessment Committee:			

## THE TESTS

<b>Reliability</b>			
17. The block team uses standard marking guidelines.	Yes	Partly	No
18. The block team always uses a plagiarism detection system in open-ended tests.	Yes	Partly	No
19. In multiple choice tests, at minimum the guidelines included with this Workbook are applied; see the Appendix to <i>Goed toetsen bij FEM</i> (Effective Assessment at FEM).	Yes	Partly	No
Notes by the block coordinator:			
Comments by the Assessment Committee:			
<b>Validity</b>			
20. Content and form of the tests match the teaching content.	Yes	Partly	No
21. The objectives formulated are (also) described in terms of behaviour. This means that phrases like 'The student can name', 'The student can apply', 'The student is able to demonstrate using	Yes	Partly	No

... ' are used instead of 'The student knows', 'The student is familiar with', 'The student has insight into'.			
22. Assessment is relevant to the students' future professional practice. (As much as possible, assessment takes place in realistic contexts and tests relate to real-world problems.)	Yes	Partly	No
23. The assessments discriminate between good and poor performances.	Yes	Partly	No
Notes by the block coordinator:			
Comments by the Assessment Committee:			

#### TEST DEVELOPMENT AND QUALITY ASSURANCE

<b>Teamwork</b>			
24. Written tests including the marking instructions are always devised by at least two examiners working together.	Yes	Partly	No
25. Oral tests are always assessed by two examiners.	Yes	Partly	No
26. The block team employs a standard method for developing or modifying tests.	Yes	Partly	No
27. When developing or modifying tests, the block team begins by (re-)establishing block objectives and assessment criteria, and then determines the number, nature and timetable of the tests.	Yes	Partly	No
28. Student input is used when developing or modifying tests through analysis of block questionnaire outcomes, the assessment committee's report on assessment, and information gained from meetings with class representatives.	Yes	Partly	No
29. The block team systematically evaluates whether the assessment procedures as a whole are still adequate, using test analyses.	Yes	Partly	No
Notes by the block coordinator:			
Comments by the Assessment Committee:			
<b>Overview</b>			
30. The block coordinator consistently coordinates the assessment in the block with other block coordinators.	Yes	Partly	No
31. The level of the assessments/tests is consistent with the course year.	Yes	Partly	No
32. Block questionnaire outcomes are analysed to assess whether the planned workload corresponds to the actual workload.	Yes	Partly	No
33. Every two years, the content and form of the assessment applied in the block as well as the types of assessment used are discussed with industry representatives.			
Notes by the block coordinator:			
Comments by the Assessment Committee:			