TIPS FOR TESTS

Universiteit Leiden
The Netherlands

Acknowledgements

Contents

Introduction

'Tips for Tests' is a translation of 'Tips bij Toetsen', which was written for and by the Faculty of Humanities (in association with ICLON).

'Tips for Tests' is intended to help teachers to set and assess tests. Consultation with teachers has shown that many find it difficult to set tests that adequately assess the content covered during the course. Teachers also report that they find it difficult to mark tests in the right way. This handbook contains tips for teachers, with references to examples and extra explanation.

'Tips for Tests' covers the five steps in the testing cycle given below: Designing tests (Chapter 1), Devising tests (Chapter 2), Marking tests (Chapter 3), Giving tests (Chapter 4) and Evaluating and improving tests (Chapter 5). Finally, Chapter 6 contains some examples of test types, test questions and assessment standards used in the Faculty of Humanities.
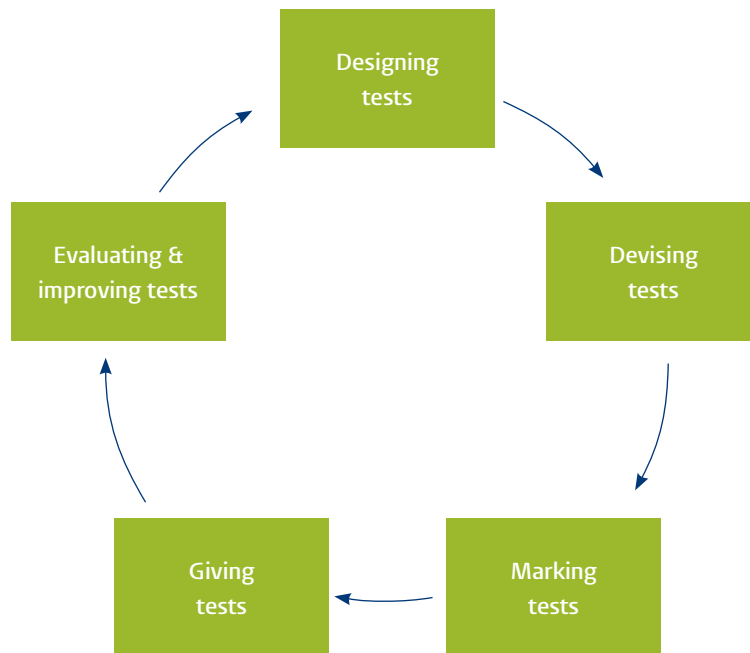
Figure 1. Test cycle

1. Designing tests

What is the design of a test?

You can use tests to gain insight into the students' level of knowledge. This can be formative, for example to help students to structure their learning process, or summative, for example to determine whether a student has mastered the course content. Whichever function a test has, the learning objectives are central to its design. The first step is to construct an overview of the learning objectives and their associated test designs. Then you decide on the test format. You also consider the point in time at which the test will be given, and its organisation.

How do you design a test?

There are three central questions when designing a test:

1. Why are the students being tested?
2. What is being tested?
3. How is it being tested?

1. Test function

When designing a test, it is important to know why you are testing the students. If you design a test because you want to collect feedback during the course about students' learning progress ('formative' testing), you will probably use a different test from the one you give at the end of that section of the course, when you assess the students' achievement level ('summative' testing).

2. Learning objectives

Learning objectives, content and testing are inextricably linked. Clearly-formulated learning objectives are essential when designing a test. When you begin to design a test, bear the learning objectives in mind; specific and measurable learning objectives direct the choice of an appropriate test format and appropriate test questions.[1] This works best when learning objectives contain specific, measurable action verbs, for example: *describe*, *explain*, *interpret*, *compare*, etc. Words such as *knowledge*, *insight* and *understand* are neither specific nor measurable. One example of a learning objective is: 'The student describes three causes of the Second World War'. During the course, students gain the necessary knowledge and insight in order to describe 'three causes of the Second World War', and this is tested at the end of the course.

3. Test format

The best test format depends on a number of factors. The most important are:[2]

▷ Learning objectives: the test format must allow you to concretely measure whether the learning objectives have been achieved;
▷ Necessary and available construction time: constructing multiple-choice questions takes more time than setting open questions or assignments;
▷ Necessary and available assessment time: multiple-choice tests take less time to assess than essays.
▷ Necessary and available time to take the test: students must be able to complete the test within the time allowed.

You can use different test formats to test different cognitive levels (such as memory, comprehension, application or analysis). One test format may be more appropriate (and easier to use) for a particular cognitive level than another. For example, do you want the student to be able to compare two perspectives on Object X? If so, it would be better to ask open questions, or to set an essay task. Does the student have to name Aspects X, Y and Z? If so, you should ask closed questions. Appendix I contains an overview of test formats that are appropriate for various cognitive levels.

Before you begin to devise the test, it can be helpful to make a test matrix (also known as a specification table). This can be very comprehensive, but it can also be less comprehensive. In the simplest variant, you list the learning objectives that you want the test to assess (topics which have also been covered during the course), and you decide how difficult/important each topic is by assigning a number of points to the topics. You set at least one question per topic. In a more extensive variant, you would also indicate the cognitive level on which the questions must be formulated; this is helpful in evenly distributing the test questions (for example if you want to avoid setting only 'knowledge regurgitation questions'). You can also use a test matrix to record which test format(s) you use; that is only possible if you use several test formats separately or in combination. For more information about test matrices, see Appendix II.

---

1   See Appendix I.
2   Source: Kallenberg et al. (2014).

Why is it important to design a test?

Designing a test gives you insight into the connection between learning objectives and test questions/formats, and helps you to include all the relevant learning objectives in your test. In addition, the learning objectives can help you to decide the best way of teaching, so consider the test design from the start of the course development process. Finally, during the lessons students have to practise achieving learning objectives and receiving feedback on their progress. It is therefore important to start working on developing a test design as soon as you start to plan a course.

More information...

▷ About setting specific and measurable learning objectives and appropriate test formats: Appendix I
▷ About making a test matrix (incl. examples): Appendix II
▷ About types of learning opjectives: Knevel, R. (2013). Taxonomieën zijn hot en handig. *Bureau ICE.* (In Dutch).

2.^Devising tests

When devising a test, you can choose between various test formats. This chapter will deal with the following test formats:

▷ closed test questions
▷ open test questions
▷ written assignments
▷ oral presentations
▷ group projects

2.1^ Closed test questions

What?

Closed questions are test questions that offer the student two or more options to choose from, one of which is the correct or best answer. Closed questions allow students to show that they have remembered or understood something well, or that they can apply it in a new situation. There are various types of closed test questions: multiple-choice questions, true/false questions (usually with reference to a statement or proposition), yes/no questions or case study questions with a number of possible answers. You can use closed questions in a final exam (summative), but also during the course, for example by setting quizzes during a tutorial or lecture (diagnostic or formative).

With closed questions you can test not only the reproduction of knowledge but also higher cognitive levels, such as insight and application. This requires you to pay a great deal of attention to the construction of both the question and the possible answers. Writing 'application questions' demands comparatively more attention; these questions are usually introduced with a description of a situation (a case study), on the basis of which the student must be able to work out which of the possible answers is correct.

How?

A closed test question consists of two parts: the *stem* and the *alternatives*. The stem contains the context of the question and the question itself, and the alternatives are the possible answers. The incorrect alternatives are called *distractors* and the correct alternative is known as the *key*.[3]

3   *Source: Van Berkel, Bax & Joosten-ten Brinke (2014).*

Tips for formulating closed test questions

Writing the stem (the context and the question itself)

1. Give clear references in the stem. The student should be able to answer the question even without the multiple-choice answers.
2. Describe the context and the central idea of the question.
3. Keep the stem short and concise.
4. Use positive wording, and wherever possible avoid words like not, no or except. If it is necessary to use negative words, make sure that these words are noticeable, for example by using capital letters.

Writing the alternatives (possible answers)

1. Make sure that only one answer is correct.
2. Vary the place of the correct answer, or always list answers in the same logical order (for example alphabetically).
3. Make the answer options unique; there should be no overlap between options.
4. Make sure that the answer options have similar levels of content and grammar.
5. Ensure that the answer options are roughly the same length.
6. Be sparing in the use of wording such as 'none of the above...' or 'all the above...'.
7. Formulate answer options positively; avoid negative formulations such as not.
8. Do not give clues to the correct answer, such as:
   a. Specific clues (always, never, completely, absolutely);
   b. Associations with stem text; words that are similar or identical to words in the stem;
   c. Grammatical inconsistencies between possible answers;
   d. Several related answer options;
   e. Noticeably different answer options;
   f. Absurd answer options.
9. Make sure that all distractors are plausible.
10. Use typical student errors to write distractors (you can collect these during the course).

Why?

Closed questions are quick to mark, which makes them useful in large groups of students, and because you have a clear answer key you assess each student in the same way. Open test formats are often assessed less objectively.

ICLON's test and examination service can automatically mark questions using your answer key for reference. ICLON also implements a quality assessment (see also: Chapter 5) and calculates the test's pass threshold (pass/fail boundary). This analysis and threshold calculation can also be implemented with open questions. The testing and examination service is a paid service.

More information?

▷ Van Berkel, H., Bax, A. & Joosten-ten Brinke, D. (2014). *Toetsen in het hoger onderwijs*. Houten: Bohn Stafleu van Loghum.
▷ Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education, 15*(3), 309-333. doi: 10.1207/S15324818AME1503_5
▷ Information about testing via Blackboard (and Blackboard's other possibilities): http://hum.leidenuniv.nl/medewerkers/ict-voorzieningen/ontdek-de-tools/blackboard2.html
▷ ICLON Test and examination service: http://www.universiteitleiden.nl/iclon/ho/toets--en-tentamenservice

2.2^ Open test questions

What?

Open questions are test questions to which students must actively formulate an answer. Open questions allow you to test not just factual knowledge, understanding and the application of knowledge (as with closed questions), but also higher cognitive levels such as critical analysis and the formulation of a well-argued opinion.

How?

Tips for formulating open questions[4]

▷ Decide how many test questions you want to write. To help with this, you could estimate beforehand how much time students will need to answer the questions. See table 1 for an indication of answer times for open questions.

Table 1: Indication of answer times for open questions.

Length of open question	Answer time
1 word or phrase	1 minute
¼ A4 page	5 minutes
½ A4 page	10 minutes
1 A4 page	25 minutes
2 A4 pages	60 minutes

▷ Use a test matrix to determine the required number of points and levels of the questions.2 You should also specify the maximum number of points students can achieve for each question.[5]

▷ Formulate the model answer first, then the question.

▷ If possible, write the questions together with colleagues.

▷ If necessary, divide the question into two parts: an introduction containing background information, and then the question itself.

▷ Be specific about what you expect in the answer, for example: 'Give three reasons for…'.

▷ Ask the question in the form of an assignment and use action verbs, as you ideally did when outlining the learning objectives.

▷ Use standard formulations for the questions to make them more consistent, such as:
  o Show that …
  o Demonstrate …
  o Explain …

Why?

In contrast with closed questions, open questions require students to actively think of an answer. By including open questions, you almost completely eliminate 'guessing'. This increases the validity (you measure what you want to know), as long as the questions are well formulated.

More information?

▷ Jensma, E. & Hubregtse, M. (2014). *Vragen beoordelen: Richtlijnen voor het opstellen van theoretische toetsen en antwoordmodellen*. KCH Examens.

▷ Van Berkel, H., Bax, A. & Joosten-ten Brinke, D. (2014). *Toetsen in het hoger onderwijs*. Houten: Bohn Stafleu van Loghum.

▷ Checklist for devising, assessing and editing open questions: see Appendix III

---

4   *Source: Van Berkel, Bax & Joosten-ten Brinke (2014).*
5   *See Chapter 1 and Appendix II.*

2.3^ Written assignments: essays, reports and articles

What?

The test format of the essay, report or article is mainly intended for short written assignments, to differentiate them from a final project or thesis. A written assignment offers the possibility to gain insight into the level to which students have mastered the higher cognitive levels, and also often concerns skills such as critical evaluation, demonstration of cohesion or reasoned argument. For example, a written assignment may consist of a reasoned stance in relation to a given statement, a critical discussion of an article, or a discourse about the social relevance of a particular research result. Students usually produce these written assignments individually, at a time and place of their own choosing. The test conditions are thus difficult or impossible to control.

How?

When setting a clear test task for a written assignment, it is important to clarify the format and the criteria you will use in connection with the content (related to the learning objectives). You can use these criteria to formulate the task well.

Tips for setting a task for a written assignment[6]

▷ The question
Indicate in the task what you expect from the student: for example, a clear introduction to the questions, simple formulation, an application to a given context or a link to the current situation.

▷ The method and/or the analysis techniques
For example, you might indicate in the task that you expect an adequate description of the student's method, and perhaps also their reasoning for selecting that method. This also applies to any analysis techniques expected from the assignment.

▷ Presentation of findings
Include a specific instruction in the task if you expect the student, for example, to apply a consistent ordering system, or to make good use of graphs and figures.

▷ Conclusions
You might include an instruction, for example, to ensure that conclusions follow logically from the preceding analysis, and that the conclusions and the answer must be relevant to the question.

▷ Form
For example, set criteria for the scope (number of words, text size, line spacing), correct use of language, readability, references, paragraphing, etc. A general rule of teaching is that an assignment that does not adhere to the form requirements automatically fails, regardless of the quality of the content (also known as 'knock-out criteria'). If this is applicable to the assignment, state this clearly in the task brief.

▷ Good examples
You might direct the student to some good examples of comparable written assignments to illustrate the instructions in the task.

Why?

The students have great freedom to show what they know and can do. Completing an assignment requires autonomy, well-developed intellectual capacities and (self-)reflection. Fellow students may be engaged in the process by giving feedback, which enables the exchange of knowledge and experience between students, making the written assignment a valuable part of their education. Consistency in marking written assignments is an important point. It seems to be especially difficult to set clear assessment criteria for higher cognitive levels. As a result, sometimes in practice the format requirements are made explicit while the content requirements are less clear. The effect of this is that the feedback function of the assessment is limited, and the students cannot contribute significantly to their learning process. In addition, where clear content assessment criteria are lacking there is a danger that students will work mainly towards the form requirements and focus less on the higher learning objectives.

---

6   *Source: Van Berkel, Bax & Joosten-ten Brinke (2014).*

More information:

▷ Tips for setting written tasks: https://www.uantwerpen.be/nl/faculteiten/antwerp-school-of-education/deelentiteiten/expertisecentrum-hoger-onderwijs/didactische-tips/onderwijstips/archief-onderwijstips/tip-34--drie-kwalite/

▷ SCALA - Scaffolding Assessment for Learning (2012). Tools voor digitaal feedback geven en beoordelen van schriftelijke werkstukken: trends en ontwikkelingen. https://www.surf.nl/binaries/content/assets/surf/nl/kennisbank/2014/literatuurstudie-tools-voor-digitaal-feedback-geven-en-beoordelen-van-schriftelijke-werkstukken.pdf

▷ Van Berkel, H., Bax, A. & Joosten-ten Brinke, D. (2014). *Toetsen in het hoger onderwijs*. Houten: Bohn Stafleu van Loghum.

▷ ExpertiseCentrum Hoger Onderwijs (2013). *Vijftig onderwijstips*. Antwerpen-Apeldoorn: Garant.

2.4^ Oral tests: presentations

What?

Oral presentations generally test both the content of the presentation and the student's presentation skills. With an oral presentation you test whether the student has mastered the content, but also, for example, whether the student can argue a point of view, think critically and hold a debate.

This paragraph will only discuss oral presentations, because other oral tests (such as a situation in which the teacher asks a student questions) are less widely used in situations in which students are assessed.

How?

There is often a specific reason to choose an oral presentation as a test format: perhaps you want to test something that cannot be assessed in another way, or it may fit the 'academic skills learning pathway' students practise (and on which they are assessed) when they give oral presentations.

Tips for setting a task for an oral presentation[7]

▷ Plan ahead. Oral presentations take time to prepare and give. Depending on the students' level and experience, you may have to teach presentation techniques and give students time to prepare and practise.
Be clear beforehand exactly what you want to test with the oral presentations, and how you can best structure them. Decide whether students will give individual or group presentations, how long the presentations should last in order to achieve the required aims, and the time period in which the presentations should take place (all on the same day, or spread out over a longer period).

▷ Formulate a clear and complete task for students, so they know what is expected of them and on which points they will be assessed.

▷ Clarify the aims of the presentation.
- Indicate how long the presentation should last (perhaps with a margin, for example 10-15 minutes) and the sections it should contain (for example introduction, description of an investigation, conclusion).
- Describe all parts of the task, including how students should prepare (for example by submitting a draft or bibliography beforehand). You should also indicate whether you want students to submit the presentation slides. You might choose to record the presentations, to allow students to reflect on their performance later.
- Specify important deadlines (for example for the submission of sections of the task, and for giving the presentation itself).
- Specify the aspects of the presentation you will mark.

Why?

Oral presentations are often used in learning pathways for academic skills. By giving presentations, students learn to express themselves well and they can show that they not only have knowledge and understanding, but that they can also generate their own ideas. In addition, presenting skills are likely to serve students well in their later career.

More information:

▷ De Grez, L., Valcke, M., & Roozen, I. (2013). The differential impact of observational learning and practice-based learning on the development of oral presentation skills in higher education. Higher Education Research & Development, 33(2), 256-271. doi: 10.1080/07294360.2013.832155

▷ De Grez, L., Valcke, M., & Roozen, I. (2009). The impact of an innovative instructional intervention on the acquisition of oral presentation skills in higher education. Computers & Education, 53(1), 112-120. doi:http://dx.doiorg/10.1016/j.compedu.2009.01.005

▷ Tips for devising and marking oral presentation tasks: http://www.speaking.pitt.edu/instructor/oral-assignments.html

2.5^ Assessing group work: written assignments or oral presentations

What?

When testing group work, you will usually mark a written assignment or oral presentation on which a group of students has worked collectively. In theory, you can also observe and assess a group while they are working together, but in higher education that is rarely the case. That being the case, this section discusses setting a written or oral task for a group.

How?

You can also use the tips for setting written assignments (2.3) and oral presentations (2.4) when setting written or oral group tasks. When dealing with group projects, as well as instructions for the written or oral 'product', you also have to describe what you expect from students' teamwork (the 'process') and how you will ensure that assessment is fair. In addition to assessment criteria for the content, when assessing group projects you will pay extra attention to *how* a product came about. As such, you will also use criteria related to the group process; ideally each group member will play an equal role in that process. Clarity from the start about the group process, and about the assessment of the group and of individuals, contributes greatly to the effectiveness of students' (group) learning experience.

Tips for setting group projects

▷ State what the aim of the group project is. Describe not only the content aims, but also – if necessary – aims in terms of teamwork and communication.

▷ Decide how groups will be formed: by the students themselves or by the teacher.

▷ Decide whether students must make their own arrangements for the group process, or whether you as their teacher will give guidelines about organising this (for example, how to assign group roles, what is expected of each group member, that students give each other feedback).

▷ Give students guidelines on how to fulfil various group roles and responsibilities. The roles may also rotate between students, so that all students benefit from the same experience. Depending on the task, important factors may include coordinating the project, keeping minutes of group discussions, finding literature on various topics, summarising articles, implementing (sections of) an investigation, or giving (sections of) a presentation.

▷ Indicate how much time students should allow for a group project.

▷ Describe the points on which you will mark the group project. It is important to indicate whether you will assess the group as a whole or group members' individual contributions (see also: Chapter 3.2.4).

▷ Consider requiring students to submit drafts or keep logbooks of their progress (and each member's contribution), to track the group process.

▷ Consider allowing time at set points during the course to discuss group projects.

Why?

Teamwork – collectively working on a product – is a competence often asked of graduates, so it makes sense to spend time on these skills during the course and also to assess students on them. Students can learn a lot from each other during a group project. It gives some students stronger motivation to get started; other students may wait until it is their turn to contribute. On the other hand, group projects can also encourage 'free-riding', so it is important to make clear arrangements for each group member's contribution and to stimulate students to give each other feedback on the work they have submitted (or, indeed, not submitted).

More information:

▷ On assessing group projects: James, R., McInnis, C. & Devlin, M. (2002). Assessing Learning in Australian Universities. *Centre for the Study of Higher Education, Australian Universities Teaching Committee.*

▷ On problems with and solutions for marking group work: Davies, W.M. (2009). Groupwork as a form of assessment: common problems and recommended solutions. *Higher Education, 58, 563-584.*

2.6^ Rules of thumb for setting tests

▷ As a measuring tool, a test can only provide useful information if the following quality requirements are met:

- validity: the test measures what it is intended to measure; it is representative of what it was intended to investigate;

- reliability: the test assesses the same things at different times and with different students and assessors;

- transparency: the students knows what is expected of them;

- suitability: the test format is appropriate to the material studied, the learning objectives and the group size. The test fits the financial and/or material framework, but can also be completed within the time available.

Aim to meet as many of these quality requirements as possible so the measuring tool can give a real indication of students' progress and where they are in the learning process. A good measuring tool ensures that students do not pass or fail undeservedly, but also that they gain a realistic idea of their learning process.

▷ Make sure that your assessment (which may be made up of several tests) is relevant to all the learning objectives of the course. At the end of the course you want to be able to assess the achievement (or lack thereof) of all the learning objectives.

▷ Choose an appropriate test format for the level of the learning objectives. For example, do not use multiple-choice questions if students are expected to argue a viewpoint.

▷ Ensure that there is a balance between the study load of a test and the degree to which the test contributes to the final grade. For example, it would not be reasonable for a 20-page assignment to be worth only 10% of the final grade.

▷ Formulate clear, unambiguous questions or tasks. Avoid complex sentence constructions, double negatives, unnecessary extra phrases, etc. (see also: Appendix III).

3.^Devising assessment models

Assessment models can take various forms, from an answer key for closed questions to *rubrics* for assignments or presentations. In closed test formats the correct answers are clear, which makes answer keys easy to write. Therefore, this chapter only concerns assessment models for open test formats, followed by specific tips and attention points for testing:

▷ written assignments (such as essays and papers);
▷ final projects (theses);
▷ oral presentations;
▷ group projects.

The chapter will also discuss how to decide the pass threshold after you have created an assessment model.

3.1^ Assessment models for open test formats

What?

An assessment model is a set of guidelines for marking open test formats. Whereas closed test formats allow for a simple answer key, open test formats require a broader model for checking and marking the test. That assessment model can take several forms, from a short description of the correct answer to a comprehensive assessment form.

How?

An assessment model is made up of three parts: a model answer (or *rubric*), a score sheet and instructions for assessors. The following information will help you to compose these three sections.

1. The *model answer* is an overview of correct, partially correct and incorrect answers. A model answer includes the following:
   ▷ A description of the correct or ideal answer. This can also consist of the *assessment criteria* if the question has no clear correct or incorrect answer. These may be questions for which students must give their own opinion or solution.
   ▷ The maximum number of points per question (section).
   ▷ Examples of incorrect answers and/or frequent errors. It is useful to include these in the model answer, to let the assessor know which errors might appear.
   A rubric is a common type of model for assessing and giving feedback on open test formats. A rubric organises your assessment criteria into a scale, with a clear explanation of each level (known as 'achievement indicators'). This will help you to show whether a student has mastered a criterion well, satisfactorily or unsatisfactorily. The scale can of course also be used for a numerical grade. If you will ultimately have to translate qualitative scores into a numerical grade, it is important to decide beforehand which grade (or grade range) you will link to each qualitative judgement (for example: 'unsatisfactory' means ≤5, 'satisfactory' means 6 or 7, and 'good' means ≤8).

Table 2. An example of a rubric for (a section of) an oral presentation.

Assessment criterion	Unsatisfactory	Satisfactory	Good
Matching the content to the aim of the presentation	The content of the presentation does not match the aim and the information is incomplete.	The content of the presentation matches the aim and contains the necessary information.	The presentation's content and aim form a coherent whole and the presentation contains the necessary information.
Use of sources	The student has not cited any sources.	The student has cited some relevant sources.	The student has cited several relevant sources.
Use of examples	The student has not used any examples.	The student has given one or a few examples.	The student has given various relevant examples.

Adapted from: Van Berkel et al, 2014

It is important to set out the achievement indicators in as concrete a way as possible, as they make up the feedback you give the student. A rubric is thus not only an assessment tool, but also an efficient and unambiguous way to give students feedback. In addition, students can also use a rubric to assess themselves or each other (*self-* and *peer assessment*). Once a rubric has been developed, you can use it in multiple years for the same test, which makes the time investment worthwhile. It is also important to adapt the rubrics on the basis of experience. It may be a good idea to develop a rubric together with one or more colleagues.

Step-by-step plan to develop a rubric[8]

1. *Determine the scope.* Is the achievement you are assessing specific to your course, or more of a general skill? Especially in the latter case, it is useful to check whether colleagues have already developed suitable rubrics. Rubrics for presentations or assignments can be (partially) adapted to various disciplines.
2. *Link to learning objectives.* Specify which learning objectives the rubric is testing.
3. *Set assessment criteria.* What knowledge and/or which skills do you want to test? Do you also want to test the student's attitude, for example their active contribution to group work? Link the criteria to the learning objectives and try to achieve an appropriate number of criteria; between five and eight criteria are generally achievable in practice. If you have too many, the detailed assessment criteria give a false sense of accuracy, which reduces their value for feedback and assessment.
4. *Determine the number of levels.* Three, four or five levels are often suitable; using more levels makes it difficult to think of distinct achievement indicators.
5. *Write down the achievement indicators.* Make sure that the achievement indicators are distinct from each other, and use measurable, observable formulas. Start with the highest level: what is the most a student can achieve? Then think of the lowest level: what is the least a student can achieve? Then fill in the levels in between.
6. *Decide on standards.* Assessment criteria can be weighted differently if some criteria are more valuable to you than others. If you expect students to be devious and compensate for unsatisfactory answers with answers graded as satisfactory, you can include a requirement that all the criteria must be met. You can use previously marked products as examples (reference points) for each level. Another tip is to go over some trial assessments with colleagues, to test a rubric.

2. A *score sheet* contains a detailed description of how points should be awarded to (partially) correct and (partially) incorrect answers. Some instructions that can appear in a score sheet are:
   ▷ Scores rounded to whole/half points.
   ▷ No answer = ... points.
   ▷ Correct = … points awarded; incorrect = … points deducted.
   ▷ Errors will/will not influence follow-up tasks (teachers usually decide against letting errors affect later questions).
   ▷ Spelling and grammatical errors = ... points deducted.
   ▷ Only count the first two answers (if this was a restriction in the question!).

3. *Assessors' instructions are general guidelines for the assessor. These instructions should prevent pitfalls for assessors. Examples of assessors'* instructions
   ▷ Mark exams per question, not per student (also known as 'horizontal marking').
   ▷ Change the order of exams from time to time.
   ▷ Only give a grade once all the questions have been marked.
   ▷ Add any newly discovered errors to the model answer.
   ▷ Add any newly discovered good answers to the model answer (check whether there are references for this).
   ▷ Get a colleague to independently mark some individual questions, and check how far you agree with each other. Only continue with marking if you both agree.

---

8   *Adapted from: Van den Bos, Burghout & Joosten-ten Brinke (2014).*

As well as using an assessment model, you can take the following measures to improve the reliability of the assessment:

▷ Mark the answers per question (instead of marking an entire test per student).

▷ Change the order of the answer sheets before beginning to mark the next question.

▷ Anonymise the answer sheets.

▷ Consider using several assessors: the greater the consequences of the test result for the student, the more reliable the assessor needs to be. If several assessors are involved:

    ▷ Distribute the questions (not the students) among the assessors.

    ▷ In case of doubt, ask a colleague to re-mark the question.

Why?

When assessing open test formats it can be difficult to remain consistent and therefore reliable. The following *assessor effects* can appear:[9]

▷ *Shifting standards*: the assessor's tendency to adapt the strictness or leniency of their assessment to the average achievement level of the group. During the assessment process the examiner may also become more lenient, for example from fatigue.

▷ *Sequence effect*: the over- or undervaluing of an answer under the influence of the quality of previous answers from other students.

▷ *Halo effect*: the assessor's knowledge about the student's previous achievements partly influences the assessment.

▷ *Assessor-specific effect*: the use of different criteria when several assessors mark the same question.

▷ *Contamination effect*: the freedom in the assessment is used (deliberately or not) for other purposes than that of an unbiased assessment. This can also be a problem if the assessors give slightly lower scores to indicate that their discipline is difficult.

▷ *Restriction of range:* one assessor's personal tendency may be to use all possible values of the assessment scale (scores from 1 to 10), while another assessor assigns scores in a narrower range (for example scores from 5 to 8).

Working with a good assessment model limits this kind of negative effect. The assessors' instructions combat assessor effects and ensure as much as possible that you mark more consistently (and therefore more reliably). A good model answer ensures that you can mark more quickly, which makes the test more useful.

In addition, it is also important for the student to know how and why a certain mark was awarded. The use of model answers, score sheets and assessors' instructions invites greater transparency of the assessment, because you are explicit about what and how you are marking.

If you decide not to provide an assessment model or form, bring in a second assessor to increase the test's reliability and validity, especially concerning work you assess as 'mediocre'.

More information:

▷ Jensma, E. & Hubregtse, M. (2014). *Vragen beoordelen: Richtlijnen voor het opstellen van theoretische toetsen en antwoordmodellen.* KCH Examens.

    ▷ Van Berkel, H., Bax, A. & Joosten-ten Brinke, D. (2014). *Toetsen in het hoger onderwijs*. Houten: Bohn Stafleu van Loghum.

    ▷ On assessment models (with examples of assessment models for language proficiency): http://www.toets.nl/uploads/Artikelen/Het-belang-van-een-goed-beoordelingsmodel.pdf

---

9   *Kallenberg et al. (2014); Van Berkel et al. (2014).*

3.1.1 Marking written assignments (such as essays and papers)

When marking written assignments, a certain amount of subjectivity is unavoidable. The assessment can be made somewhat more objective by working from a *standardised assessment tool*, and by bringing in a second assessor if necessary.

Tips for assessing written assignments[10]

▷ Write down the assessment criteria beforehand, regarding both content and form. Make as much use as possible of standardised assessment tools.

▷ If necessary, discuss the assessment criteria with colleagues.

▷ Make sure that the assessment tool and the weighting of the assessment criteria do not overemphasise the importance of the lower cognitive achievements (such as the format of the assignment) at the cost of the higher cognitive achievements (such as critical evaluation or reasoned argument).

▷ Good assessment tools have *specific* criteria. That is to say, each criterion is clearly different from or does not overlap with other criteria.

▷ Start the marking process by reading (some of) the written assignments from start to finish; this can give you an impression of the standard to use.

▷ Select a few 'models' of very good, good, satisfactory and unsatisfactory answers, to provide a basis for assessing the rest of the written assignments. This is especially important if various assessors are involved, or when the assessment is spread over a longer time period.

▷ If possible, avoid long intervals between marking; this helps keep the assessment uniform.

▷ In addition to a more *analytical assessment* on the basis of an assessment form or scheme, you can also (especially as an experienced teacher) implement an *overall assessment* based on the entire impression the assignment makes. In practice, it is not uncommon to make small adjustments to a draft completed assessment form in order to bring the final grade, based on that form, into line with the overall grade.

▷ Write feedback for students, linked to the assessment criteria, to make follow-up conversations easier and to make the assessment more comprehensible for students.

▷ Consider giving students a role in assessing their fellow students' work (peer assessment). Clear assessment criteria are also important here. When students work on a collective product as a group, peer assessment can be used to correct the group mark for each individual student's contribution to the group project. Students can assess each other on teamwork within the group project, or they can give each other feedback on content. In any case, it is important to train/supervise students in giving feedback. In addition to training students, they (or their products) can also be assessed by two or more fellow students, in combination with the teacher's assessment (or sample).

3.1.2 Marking final projects (theses)

The final project carries considerable weight within the complete testing schedule of a course. As the final study component, most of the attainment targets that count towards the final course grade are tested in the final project. For rules and guidelines for (assessing) final projects, see 'Kwaliteitsborging Toetsing. Een Handleiding voor examencommissies' (Exam Board Handbook) available from the Faculty of Humanities.

---

10 *Partly derived from 'Onderwijstips', from the Expertise Centre for Higher Education, Antwerp University.*
    www.uantwerpen.be/nl/faculteiten/antwerp-school-of-education/deelentiteiten/expertisecentrum-hoger-onderwijs/didactische-tips/onderwijstips/archief-onderwijstips/tip-43--een-schrifte

Tips for assessing final projects

▷ When marking final projects, theses or research reports in the final phase of the course, ALWAYS use the faculty's standardised assessment form.

▷ In general terms, the same comments apply to the assessment of final projects as to the assessment of other written assignments. Assessor subjectivity is unavoidable, but experienced assessors who have already marked many final projects are aware that the *overall assessment*, on the basis of the overall impression the final project makes, is an important indicator for the assessment of the student's overall achievement.

▷ It is ALWAYS prescribed that a final project be assessed by a second reader. In addition, the two assessors should make their assessments *demonstrably independently* of each other. In practice, that means that you should preferably mark the assignment independently first, without any knowledge of the other assessor's conclusion. The two independent assessments then lead to a discussion between the assessors, who agree on a final grade plus relevant feedback for the student. When re-marking final projects, exam boards, midterm boards and review committees may want to see both the final assessment and the original independent sections.

▷ As a supervisor who is directly concerned in the student's research and/or writing process, you not only assess the final project on the quality of the final version, you also take into account the process the student followed to reach the final product. The second assessor is usually in a far weaker position to form an opinion of that creative process. This disparity can have an impact on the initial independent assessments, but it also contributes to the objectivity of the assessment, and following the discussion between the assessors it does not usually constitute an obstacle to agreeing on a final grade.
Ideally, the assessment form should make a clear distinction between assessment criteria focused on the implementation of research that led to the final project, or the writing process itself, and criteria focused on the quality of the definitive final product.

3.1.3   Marking oral presentations

A certain degree of subjectivity and assessor effects also play a role in marking oral presentations. As when marking written assignments, it is preferable to mark oral presentations using a standardised assessment tool.

Tips for assessing oral presentations

▷ With assessment tools for oral presentations you can usually assess both the student's higher cognitive levels (for example analysis of the study material, argument, cohesion, comparison and solution) and their communicative and presentation skills. Make a clear distinction between these two aspects in the assessment tool.

▷ A rubric for presentation skills might consist of three content criteria (introduction, structure and conclusion), five criteria related to presentation style (eye contact, use of the voice, enthusiasm, interaction with the audience and body language), and a general (final) mark.

▷ A student's oral presentation performance may be negatively influenced by personal traits, such as shyness or stage fright. Of course this student must also be assessed and receive feedback on their presentation skills, but it is especially important in that case to make a clear distinction between the different aspects of the assessment.

▷ Adjust the weight of the various assessment criteria based on the importance assigned to each learning objective you are assessing.

▷ Inform the student beforehand of the nature, content and manner of assessment.

▷ If possible, work with two assessors to compare assessments and, if applicable, to produce an average. This contributes to the reliability of the assessment and the quality of the feedback.

▷ Consider giving fellow students a role in supporting the assessment and enriching the feedback the student receives.

3.1.4 Marking group projects

One concern many students have is that their personal achievement will not be fairly assessed in group project marks. Be clear with students about the goal of the group project, what will be assessed and how, and consider the following:

▷ Will you assess the final product, the group work process, or both?

▷ If you will assess both the process and the product, how much weight will each aspect have in the final grade?

▷ Which criteria will you use to assess the group project?

▷ Will you decide these yourself, let the students decide, or will you decide together?

▷ Who will assess the process and/or the product? The teacher, the student(s), or both?

▷ How will you assign a mark? One collective group mark; a group mark made up of the average of the different group members' marks; individual marks; or a combination of a group mark and an individual mark? (This last option is preferable.)

Tips for assessing group projects

▷ Give students a role in setting the criteria for assessing the group process. This gives them insight into the factors that are important in group work, and helps them to take more ownership of the group process.

▷ Share the assessment tools with students beforehand.

▷ Take the students' feedback and assessment into account when deciding their final mark.

▷ Do not assess only end products, but also intermediate products such as logbooks, minutes of meetings or action plans.

▷ Give interim feedback, so students can finetune their approach if necessary.

▷ Make the assessment procedure clear.

▷ When you involve students in the assessment: make sure that you, as the teacher, take the lead and the ultimate responsibility. Students can give feedback or advice, without giving a formal mark.

More information:

▷ Bax, A., Perrenet, J. & Berkel, H. van (2014). Toetsen van groepswerk. In: Berkel, H van, en Bax, A. (eds.); *Toetsen in het Hoger Onderwijs.* Houten: Bohn Stafleu van Loghum

▷ On ways to assess group work: http://tauu.uu.nl/wp-content/uploads/2015/01/Zeven-manieren-om-tot-een-individueel-cijfer-te-komen-bij-groepswerk.pdf

3.2^ Setting pass thresholds

What?

For all test formats you must decide when students' achievement is satisfactory. The threshold is the pass/fail boundary when marking a test. This can be determined in three ways: *absolutely*, *relatively* or a mixture of the two. In practice, these methods are often combined.

How?

▷ The *absolute* method does not take other students' scores into account: you stick to a fixed standard. That standard is usually a certain minimum percentage of the questions that must be answered correctly. For example, students who answer 75% of the questions satisfactorily may receive a mark of 7.5. The threshold, or the pass/fail boundary, is often determined on the basis of 55% of the maximum points available. If points are equally distributed between the learning objectives in the test matrix, this is also a logical boundary. The disadvantage of this method is that any weaknesses in the test itself are not corrected in the marking.

▷ In the *relative* method, you determine a student's mark by comparing their work with the other test candidates. The mark a student achieves is thus dependent on what the other candidates achieved. For example, if a student gets 90 out of 100 points on a test, and their fellow students have all achieved fewer points, that student will receive a high score. Van Berkel et al. (2014) use the metaphor of a ladder: the highest-scoring student stands on the top rung, the lowest-scoring student stands on the lowest rung, and the rest of the students are distributed evenly in between. The relative method is rarely applied in its purest form, because it is open to fraud (students can prepare for the test together) and because it gives unreliable results when there are a small number of candidates.

▷ In practice, a *mixed form* is usually used: an absolute standard is used, with the possibility of later correction on the basis of relative scores. That is to say: if a group's overall test scores deviate significantly from what is usually achieved in a similar test, you can correct the marks afterwards. For example: if more than 40% of the test participants have failed, and the test is new and as such the difficulty has not yet become clear, you might consider moving the pass threshold.

More information:

▷ Van Berkel, H., Bax, A. & Joosten-ten Brinke, D. (2014). *Toetsen in het hoger onderwijs*. Houten: Bohn Stafleu van Loghum.

▷ On the process of threshold determination (with sample calculations): http://www.toets.nl/uploads/Downloads%20van%20Bureau%20ICE/VO/Artikel%20toets%20wanneer%20geef%20ik%20een%205,5.pdf

4. Giving tests

What?

When invigilating a test, it is important to prevent and recognise fraud. 'Fraud' is understood as everything that prevents the result of a test being traceable to the results of a student or group of students; from cheating and plagiarism to 'free-riding' in group projects.

How?

Tips on preventing cheating:

▷ Be explicit beforehand about which resources are and are not allowed (for example the use of a dictionary). Mobile phone use is always forbidden.

▷ Ensure that there are enough active invigilators during an exam. The number of invigilators is determined by the number of registered students and the number of rooms in which the exam is held. Consider bringing in two invigilators if there are more than 50 students.

▷ Ensure that students cannot look at each other's work, by regulating the distance between them (at least two metres left/right and in front/behind), and perhaps by using a variety of versions of the tests (the same questions, but in a different order) and alternating them when giving them out.

Tips to avoid plagiarism:

▷ Inform students about how they should use sources and citations.

▷ Discuss plagiarism with students; not all students recognise it.

▷ Make students aware of the faculty's policy on plagiarism.

▷ Make assignments less sensitive to plagiarism, for example by:
- giving specific and clearly outlined assignments;
- making a list of subjects for students;
- making clear which specific elements the work must include;
- having students write and submit their assignments in phases. In this way the teacher has greater oversight over the creative process.

Tips for recognising plagiarism:

▷ Sometimes features of the text point to plagiarism, for example:
- different citation and referencing conventions within a text;
- the complete absence of citations and references;
- conspicuous typesetting, for example with lines that end halfway through;
- outdated information.

▷ Use a tool such as Turnitin to uncover plagiarism

Tips for preventing free-riding:

▷ Ask students to make a clear plan, including the distribution of tasks. Have them take minutes so it is clear who is responsible for which part and how the group process has progressed.

▷ Ask active supervisory questions, alternating between asking the whole group and asking individuals.

▷ If students are required to give an oral presentation, only let them know on the day of the presentation which group members are going to present, ensuring that every member prepares a presentation (and thus studies the material).

▷ If there are problems and if not everyone in the group contributes equally, make sure students know who to alert in good time.

▷ Use a 'peer evaluation' system, whereby students judge each other's presentations.

Why?

You want to be sure that the result of a test really is the result of the individual student's effort, or in the case of group projects that it is the result of the group's effort. Informing students about fraud and taking steps to check it will contribute to the test's reliability and transparency.

More information:

▷ There is advice in the Board of Examiners' rules and guidelines about how to proceed in the event of fraud.

▷ Information about preventing and recognising fraud: http://hum.leidenuniv.nl/medewerkers/onderwijs/onderwijsbeleid/plagiaat-docenten.html

▷ https://www.utwente.nl/ces/vop/onderwijs_links/toetsing_en_assesment_fraude_e/index.html#voorkomen_van_fraude_en_plagiaat

▷ Information about plagiarism: http://hum.leidenuniv.nl/studenten/reglementen/plagiaat.html

5. Evaluating and improving tests

What?

A test evaluation is an evaluation of the quality of the test. Whilst marking and checking tests, you can tell which questions were answered poorly, and the assessment criteria for which students received low scores. When you notice those things, you are already evaluating the test (and the course). Here are some structural ways to evaluate a test.

How?

There are various ways to evaluate and improve a test:

▷ Make a list of common student errors in order to adapt your course, devise new test questions, set new assessment criteria or hone existing criteria. In this way you help students to work more on things that they demonstrably find difficult; in other words, things they have to learn.

▷ Ask students to give feedback on the test, for example using a short (online) questionnaire or by adding a few questions at the end of the test or assignment. For example, ask students which questions they found clear or unclear, whether the test was a good reflection of the course, and whether they have any tips for improvement.

▷ You can also use *peer assessment* as an evaluation tool: getting students to assess each other using the assessment model you made will test the model's clarity and suitability. In this way you can devise the assessment model together with students, making the students 'co-owners' of the assessment.

▷ On the basis of your entire student population's test results you can determine the eventual *pass percentage* and the *average score of the best 5% of the students*. The best 5% of the students can be expected to answer practically all the questions well. If this is not the case, the test was probably too difficult. In that case, you can decide to move the threshold based on the result of the best 5% (see also: Chapter 3.2).

In various test questions or test sections, you can look out for:

▷ Sections that were answered well: they may be too easy. You might consider reducing the weight of those sections, or testing them at a higher level. Sections that were answered poorly may be too difficult, or may not have been explained well enough during the course. It is of course also possible that you added those difficult sections deliberately, to give the best students an extra challenge.

▷ The distinctive character of a question or section: if students who got high marks on the test scored poorly on a certain question or section, there may be something wrong with that part.

For tests with open and closed questions, the ICLON can run a psychometric test analysis to analyse the quality of the test and the individual test questions.[11]

More information:

▷ On test analyses: http://media.leidenuniv.nl/legacy/toetsing-en-toetsanalyse.pdf

▷ On psychometric test analysis: Appendix IV

---

11   *See Appendix IV for an explanation of psychometric test analysis.*

6.^Practical examples from the Faculty of Humanities

Sample learning objectives

A student who has successfully completed the course is able to:
▷ describe and discuss epistemology as a philosophical and academic discipline;
▷ explain and discuss the above-mentioned key concepts in epistemology;
▷ describe the above-mentioned debates in epistemology;
▷ describe and assess viewpoints and trends within the framework of these debates;
▷ form and support their own view in relation to these debates;
▷ describe the above-mentioned problems in modern epistemology, and make and assess suggestions to resolve these problems.

Sample test questions

Sample open test questions

QUESTION 10 (with an image of a work of art)
10a. Give the first and last names of the artist, the period and the genre.
10b. Explain the name of the genre, using three important features of this artwork.
10c. Explain the features you mentioned in 10b on the basis of the cultural context in which this genre appeared.
10d. On the basis of this artwork and the class on this genre during the course, briefly explain the problem of the definition of the terms 'abstract' and 'realistic'.
10e. Choose three approaches from those named by Charles Harrison to interpret works like this example, and briefly summarise what the approaches contain.

Sample open test question with model answer

QUESTION
Based on your reading of Hans Bloemsma's Carl Friedrich von Rumohr and the early Italians; why did Carl Friedrich von Rumohr (1785-1843) value the works of Cimabue and Duccio more highly than those of Giotto?

MODEL ANSWER
(Hans Bloemsma) According to Rumohr, whereas the work of Cimabue and Duccio focused on the sublime imagining of sacred ideas, this was a secondary concern for Giotto. In contrast, he focused mainly on the worldly aspects of his subjects, emphasising the burlesque and the pathetic. According to Rumohr, for this reason Giotto's depictions did not satisfy the (Romantic) ideal of Christian art, as Friedrich Schlegel articulated: 'It is the destiny of art to exalt religion and to make her mysteries even more beautiful and clear.' In other words, Giotto's work is too worldly and lacks glorification and religious feeling.
[It is of course unnecessary for a student to cite Schlegel in order to achieve the correct answer!]

Sample task for a written assignment
Written assignment

This class examines Tony Bennett's concept of the museum as 'exhibitionary complex', and locates, comments on and updates his influential 1988 article 'The Exhibitionary Complex'.

Lit.: Tony Bennett, 'The Exhibitionary Complex'. First appeared as an article in new formations, Number 4, Spring 1988. The text is available to download from: www.columbia.edu/itc/anthropology/.../bennett.pdf

Also available in: Greenberg, Reesa, Bruce W. Ferguson & Sandy Nairne (eds.), Thinking about Exhibitions, Routledge, 1996, pp. 81-112.

TASK
Write a critical analysis of Tony Bennett's work, with reference to a museum of your choice.

Essay Score Form: English Linguistics

	Grade

student: …………………………..…………………………...…
course: ……………………………………………… …………
graded by: ……………………………..…………………………..

linguistic analysis [well-argued, detailed (observation, description), formal (using the appropriate terms and concepts), and precise analysis of linguistic data]	○ excellent ○ good ○ acceptable ○ insufficient
structure [information presentation, logical build-up of text and argumentation]	○ excellent ○ good ○ acceptable ○ insufficient
use of scholarly sources [informed, critical and creative use of relevant secondary material]	○ excellent ○ good ○ acceptable ○ insufficient
language [correct and formal academic prose]	○ excellent ○ good ○ acceptable ○ insufficient
presentation [use of stylesheet and layout]	○ excellent ○ good ○ acceptable ○ insufficient

Comments

Frequently Made Mistakes

Your tutor can use this list to indicate what common mistakes, if any, occur frequently in your essay.

linguistic analysis

insufficient empirical evidence for analytical claims	
overview of linguistic data and/or existing literature without offering a critical evaluation or analysis	
inaccurate or imprecise use of linguistic terminology	
imprecise and overgeneralized claims	

use of scholarly sources

secondary sources are of insufficient academic quality (*Wikipedia* is an especially egregious example)	
insufficient references or empirical evidence to support claims	
use of secondary sources not effectively integrated into the argument	

structure

the essay lacks a clear, precise thesis statement	
body paragraphs lack a clear topic sentence	
cohesion between sentences and/or paragraphs is unclear	

language

grammatical mistakes (e.g. dangling infinitives and participles: 'Having processed the questionnaire, the results were presented in a table'; subject/verb disagreement: 'the verbs in this sentence is …'; incorrect verb tenses: 'In *The Canterbury Tales*, Chaucer is not using *do*-support'; etc.)	
spelling and punctuation mistakes (e.g. it's/its, belief/believe, there/their, traveled/travelled, etc.; comma splice, incorrect punctuation of relative clauses, etc.)	

presentation

inconsistent bibliographical references (no italics for titles of books and journals, no publisher or year of publication, no page range for articles, etc.)	
layout problems (no double-spacing, no tabs at beginning of paragraph, no page numbers, etc.)	

Essay Assessor

assignment: TITLE OF THE COURSE

personalia	
mark	
TOPIC, QUESTION, RESULT ▷ *relevance* ▷ *validity* ▷ *significance* ▷ *originality*	
ARGUMENT ▷ *cogency* ▷ *clarity & structure* ▷ *originality* ▷ *theoretical orientation* ▷ *methodology*	
RESEARCH ▷ *sophistication* ▷ *extent of the effort*	
LANGUAGE USAGE ▷ *correctness* ▷ *style*	
PRESENTATION ▷ *scholarly (citation, bibliography etc)* ▷ *general (formatting & neatness)*	
OTHER	

Assessment form for Bachelor Tutorial (BA-WC) Leiden University – History

Instructions for filling in the form:

▷ The completed form must be returned to the secretariat.

▷ Grades will not be added to USIS if the form is missing.

▷ The student receives a completed version of the form.

▷ Both Commentary fields must be completed for <u>Written assignment</u>.

▷ The teacher may add the learning objectives from the e-Prospectus to the last field.

▷ The teachers themselves must add weighting factors to sections. Those factors must be the same as those contained in the course descriptions in the e-Prospectus. Weighting factors must be chosen so as to ensure that an assignment may only be marked Satisfactory when it is of satisfactory quality.

▷ '0' may be filled in as the weighting factor.

BA tutorial		Teacher's name	
Date		Teacher's signature	
Student's name		Student number	
Final Grade			

Presentation to fellow students (structure, content, form)	**Participation** (attendance, participation in discussions, feedback given, teamwork)	**Assignment(s)** (planning, receiving feedback, teamwork, bibliography)
☐ Outstanding ☐ Good ☐ Satisfactory ☐ Acceptable ☐ Unsatisfactory **Grade:** **Weighting:**	☐ Outstanding ☐ Good ☐ Satisfactory ☐ Acceptable ☐ Unsatisfactory **Grade:** **Weighting:**	☐ Outstanding ☐ Good ☐ Satisfactory ☐ Acceptable ☐ Unsatisfactory **Grade:** **Weighting:**

Comments (if any):

<u>Written assignment</u> (aspects of content/analysis) **- Problem definition and sub-questions, setting in historiography, justification of method, use of theory and/or hypotheses, analysis and interpretation of research data.** **- Composition and consistency of argument, conclusion, critical analysis, originality.**	**Weighting:**
Comments:	**Assessment**

<u>Written assignment</u> (form and technical aspects) **- Technical aspects: heuristic aspect, organisation of material, background knowledge and historical awareness.** **- Form aspects: command of language and style, layout of paragraphs and chapters, annotations, illustrations, tables and graphs.**	**Weighting:**
Comments:	**Assessment**

Assessment form for <u>Bachelor's Thesis</u> Leiden University – Faculty of Humanities – History

Programme	History	Specialisation	
Date		EC	15
Student's name		Student number	
Thesis title			
First assessor		Second assessor	

1. In your opinion, is the thesis free from plagiarism?

o **Yes**   o **No**

2. BA theses in the repository are ordinarily kept confidential. If the student and the teacher see good reason to make a thesis publicly available, indicate the following.

o **This thesis may be made public immediately via the repository.**

o **This thesis may only be made public from __-__-20__ via the repository.**

Criteria	Assessment
Knowledge and insight (content, relevance to field of study)	**Optional: weighting: %**
Comments:	o Very good o Good o Satisfactory o Unsatisfactory
Application of knowledge and insight (methodology)	**Optional: weighting: %**
Comments:	
Assessment factors (interpretation, argument, conclusion)	**Optional: weighting: %**
Comments:	
Communication (writing proficiency, structure)	**Optional: weighting: %**
Comments:	
Study skills (process)	**Optional: weighting: %**
Comments:	
Formal requirements	
Ca. 35 pages, max. 15,000 words (incl. footnotes, excl. literature list) Comments:	
Programme-specific criteria for History	
1. Clearly formulated historical problem 2. Embedding in historiographical and/or academic discussion 3. Justification for the research method 4. Based on critical research in primary sources 5. Sufficiently broad study of literature 6. Well-structured argument 7. Academically-supported conclusion 8. Correct use of language NB Only theses which satisfy these eight criteria may be assessed as Satisfactory Comments:	
Summarised grade/commentary	

First assessor's signature	Final grade Second assessor's signature

NB OSZ Staff

If the final grade is a 6–6.5, do not enter the grade in USIS until a third reader has signed it. The official secretary to the Exam Board will issue this instruction at the request of the first reader.

Name of third reader	Third reader's assessmentThird reader's signature

Sub-criteria for the assessment of the **Bachelor's thesis** (to be completed with programme-specific requirements related to the programme-specific attainment goals):

Knowledge and insight (content, relevance to field of study)

Dublin descriptor: Has demonstrated knowledge and understanding in a field of study that builds and improves upon general secondary education; works at a level which implies some aspects requiring knowledge of the latest developments in the field of study, supported by specialised textbooks.

Including:
· *the research question is based on a problem which shows understanding of the central debates and methods in the field of study;*
· *clarity, relevance and demarcation of the problem;*
· *basis in existing literature.*

Application of knowledge and insight (methodology)

Dublin descriptor: Can apply his/her knowledge and understanding in such a way as to demonstrate a professional approach to his/her work or profession, and shows competence in proposing and supporting arguments and in solving problems in the field of study.

Including:
· *critical analysis of the material and sources (quality of the analysis);*
· *operationalisation and use of terminology;*
· *use of suitable research methods;*
· *description of and justification for the chosen method.*

Assessment factors (interpretation, argument, conclusion)

Dublin descriptor: Can collect and interpret relevant data (generally in the chosen field of study) with the aim of forming an opinion partially based on weighing relevant social, academic or ethical aspects against each other.

Including:
· *logical and consistent argument; conclusions follow logically and are supported by the material presented;*
· *the degree to which the research question is answered in practice;*
· *if applicable: weighing social and ethical aspects when forming an opinion.*

Communication (writing proficiency, structure)

Dublin descriptor: Can communicate information, ideas and solutions to an audience of specialists or non-specialists.

Including:
· *use of language (working language or target language of the programme: level of language proficiency, readability, style, spelling, grammar, use and explanation of correct specialist terminology);*
· *structure and composition of the thesis (chapter and paragraph distribution, contents page, use of illustrations);*
· *demonstration of knowledge of annotation (correct use of referencing system, full references, bibliography, etc.);*

Study skills (process)

Dublin descriptor: Has the necessary study skills to undertake further studies requiring a high level of autonomy.

Including:
· *degree of autonomy;*
· *planning and time management;*
· *dealing with supervisors' feedback;*
· *if applicable: participation in a thesis support group.*

Formal requirements

· *Ca. 35 pages, max. 15,000 words (incl. footnotes, excl. literature list)*

Sources consulted

Literature

Berkel, H.J.M. van, A. Bax & D. Joosten-ten Brinke (eds.) (2014). *Toetsen in het hoger onderwijs*. Houten: Bohn Stafleu van Loghum.

Biggs, J.B. (2003). *Teaching for quality learning at university.* Buckingham: Open University Press/Society for Research into Higher Education. (Second edition)

Breukers, H. et al. (2011). *Handboek toetsen en tentamens bij de Open Universiteit*. Heerlen: Open Universiteit.

Bruijns, V. en Kok, M. (2013). *Leidraad toetsen en beoordelen.* Amsterdam: Hogeschool van Amsterdam.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education, 15*(3), 309-333. doi: 10.1207/S15324818AME1503_5

Kallenberg T., van der Grijspaarde, L., Ter Braak, A. & Baars, G. (2014). *Leren (en) doceren in het hoger onderwijs.* Den Haag: Boom Lemma.

Universiteit van Amsterdam (2012). *Kader Toetsbeleid*. Amsterdam: UvA. (Universitaire Beleidsnotities).

Websites

Utrecht University, Faculty of Humanities (consulted July 2015). *Testing – Modules.* ► http://toetsing.hum.uu.nl/modules/

Hogeschool van Amsterdam (consulted July 2015). *Score – Testing and assessment.*

► https://score.hva.nl/

Appendix I: Setting learning objectives and appropriate test formats

Learning objectives consist of the intended level of knowledge, skills and/or behaviour at the end of a meeting, course or programme. Ideally, a learning objective is made up of four components:

1. Content: the content the student must master;
2. Behaviour: what students must be able to do with the study material;
3. Requirements: under what conditions and with which sources or resources the student must demonstrate the behaviour;
4. Standard: how much a student must do, within what time period (especially important when setting tests).

Specific and measurable learning objectives are SMART, which stands for:

Specific: The learning objective is described in concrete behavioural terms;

Measurable: The behaviour may be assessed on the basis of previously-determined criteria;

Acceptable: The behaviour is relevant, and motivational for students and teachers;

Realistic: The aim is achievable and relevant to the student's 'zone of proximal development';

Time-bound: The aim is achievable within the time allowed.

Tips for formulating learning objectives:

▷ Do not record any learning activities in the learning objectives (such as 'by doing group work, students must learn to…'). Record only the end result.

▷ Describe only the highest level of mastery (*apply* implies *know*). Write 'Students use various methods for X' and NOT 'students know various methods for X and apply them'.

▷ Try to use simple action verbs. Choose action verbs which apply to the activity you expect of students (see the table below).

Level of Mastery	Behaviour characteristic	Examples of action verbs for formulating learning objectives	Examples of appropriate test formats
Remember	reproduction of facts, recognition, memory	classify, recognise, identify, put in order, locate, paraphrase, name, summarise, reproduce, categorise, select, differentiate, reflect *(For example: 'Recognises important events in US history')*	Written: closed (multiple-choice) questions and open questions
Understand	comprehension, interpretation, logical reproduction	clarify, show, formulate, illustrate, characterise, instruct, outline, indicate, sort, paraphrase, represent *(For example: 'Paraphrases a section of a given text')*	Written: closed and open questions, essay questions, papers Oral: presentation, interview
Apply	use elements of 'recall' and 'understanding' in new situations; choose between the correct laws, rules, schedules, concepts	contribute, interpret, describe, ensure, contribute, define, demonstrate, use, employ, resolve, set out, give an overview, choose and follow a procedure, report, explain, predict, design, formulate questions *(For example: 'Employs various strategies to select field-specific academic literature')*	Written: closed and open questions, essay questions, papers, assignments Oral: presentation, assessment
Analyse/ synthesise	order by content, form, function, etc.; assemble elements into a unique and original whole	reject, infer, weigh up, suggest alternatives, argue, comment on, weigh interests, conclude, construct, check, discuss, determine, reformulate, take command, model, motivate, negotiate, design, criticise, build, solve problems, order, organise, prioritise, relate, compile, summarise, implement, establish *(For example: 'Determines the essay writer's viewpoint concerning his or her political vision')*	Written: closed and open questions, essay questions, papers, assignments, group projects Oral: presentation, assessment, debate
Evaluate, create	evaluate, assess, apply outside a student's own discipline	advise, assess, comment on, critically review, evaluate, explain, formulate, support, judge, defend, test, show independently *(For example: 'Formulates a hypothesis to explain an observed phenomenon')*	Written: essay, paper, literary review, assignment, project, research proposal Oral: advice, debate, presentation

Appendix II: Devising a test matrix

1. Draw a table with columns for: *learning objectives*, *knowledge questions*, *comprehension questions*, *application questions*,[12] *points* and, if necessary, *test formats*.
2. Determine the weighting of each learning objective (some learning objectives are more important than others).
3. For each learning objective, indicate how many questions you want to set and how many points each question will be worth.
4. Calculate the number of points per level, and check whether the ratio matches what you intended (for example, the ratio may have too many knowledge questions and too few application questions).

Examples of types of test matrices:

Learning objectives	Knowledge questions	Comprehension questions	Application questions	Points
Learning objective A	2 x 5 pt			10 pt
Learning objective B		1 x 10 pt	1 x 15 pt	25 pt
Learning objective C	2 x 5 pt	1 x 10 pt		20 pt
Learning objective D		1 x 10 pt	1 x 15 pt	25 pt
Learning objective E			1 x 20 pt	20 pt
Total	20 pt	30 pt	50 pt	100 pt

Content	Cognitive level being tested						Practical organisation			
Learning objective/ topic	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation	Question type	Number	Weight on the test	Who is answering the question?

---

12 *You might also choose other levels, for example: reproduction, comprehension, application, problem-solving, motor skills, interactive skills, etc.*

Appendix III: Checklist for devising, marking and editing open questions

Source: CITO – Het construeren van open vragen (Constructing open questions)

http://www.toetswijzer.nl/html/toetsenopschool/hoofdstuk7_herzien.pdf

Checklist for devising, marking and editing open questions

Use of language
▷ Is the question formulated in a grammatically correct way?
▷ Does the question contain complex sentence constructions?
▷ Does the question contain any double negatives?
▷ Does the question contain unnecessary phrases?
▷ Is the question asked in an unnecessarily negative way?
▷ Might the formulation of the question cause misunderstanding?
▷ Could shifting the word stress give the question a different meaning?

Information
▷ Does the question contain enough information to make an answer possible?
▷ Does the question give enough information about the desired length and form of the answer?
▷ Is it clear that an answer must be explained/clarified?
▷ Is the information clearly separated from the problem?

Relevance
▷ Can the question be answered using different knowledge from that intended?
▷ Does the question suggest an unrelated problem?
▷ Does the question contain inadvertent hints at the answer?
▷ Is the difficulty of the question acceptable in the context of the whole test?
▷ Is the difficulty of the question unnecessarily increased by irrelevant details?

Context
▷ Is the use of context (drawings, graphs, texts, images) practical?
▷ Is the context clearly and correctly rendered?
▷ Does the context contain irrelevant information?

Presentation
▷ Are the questions and the question sections clearly distinguishable from each other?
▷ Is the numbering of the questions logical and organised?
▷ Have the conventions relating to spelling, use of symbols, punctuation, etc. been respected?
▷ Are all references to texts, drawings, etc. correct?

Appendix IV: Explanation of psychometric test analysis

A psychometric test analysis measures the following:

▷ The *reliability* of the test, measured with Cronbach's alpha, in which alpha is located between 0 and 1. An alpha of 0 means that there is absolutely no relationship between the scores for the various test questions, in which case you would have to ask yourself whether the test is internally coherent. An alpha of 1 represents the maximum relationship between the various test questions, which might make you wonder whether the test is really only testing one question. In that case, the scoring of each test question is exactly the same. In practice, the alpha is never 0 or 1, but lies somewhere between 0.5 and 0.9. A value of about 0.8 is considered good (Van Berkel, Bax & Joosten-ten Brinke, 2014).

▷ The *difficulty* of each test question, measured with the P value, which is located between 0 (all students achieved 0 points for the question) and 1 (all students achieved the maximum number of points for the question). In other words, the higher the P value, the easier the question.

▷ The *distinctive capacity* of the question, measured with the item-rest correlation (Rir), which is a correlation between the scores for that question and the scores for all the other questions. A positive correlation means that students who answered that question well also did well in the other questions, and students who answered that question poorly also did poorly in the other questions. This means that the question clearly separates good students from poor students. If the correlation is negative, students who answered that question well did poorly on the other questions, and students who answered that question poorly did well on the other questions. In that case, the question's distinctive capacity appears to be poor, and there may be something wrong with the question. You, as the teacher, must determine what is wrong. For example, it is possible that a question is open to several interpretations, or that a closed question's answer key contains errors.

▷ Multiple-choice questions also allow you to ascertain whether a lot of students chose a certain distractor. If so, you can investigate whether this option might also be seen as correct. If that turns out to be the case, then you might consider marking this so-called 'attractive distractor' correct, as well as your preferred answer.

▷ Most multiple-choice questions have distractors which students never or hardly ever choose. In this case, you might consider whether these are plausible distractors, or whether for example you should use a two-choice question instead of a four-choice question. The less plausible the distractors, the easier it is for the student to answer the question correctly. If a distractor turns out to be implausible, try to come up with a better one, for example by noticing the mistakes students often make during the course. You can then use these mistakes as distractors in multiple-choice questions.