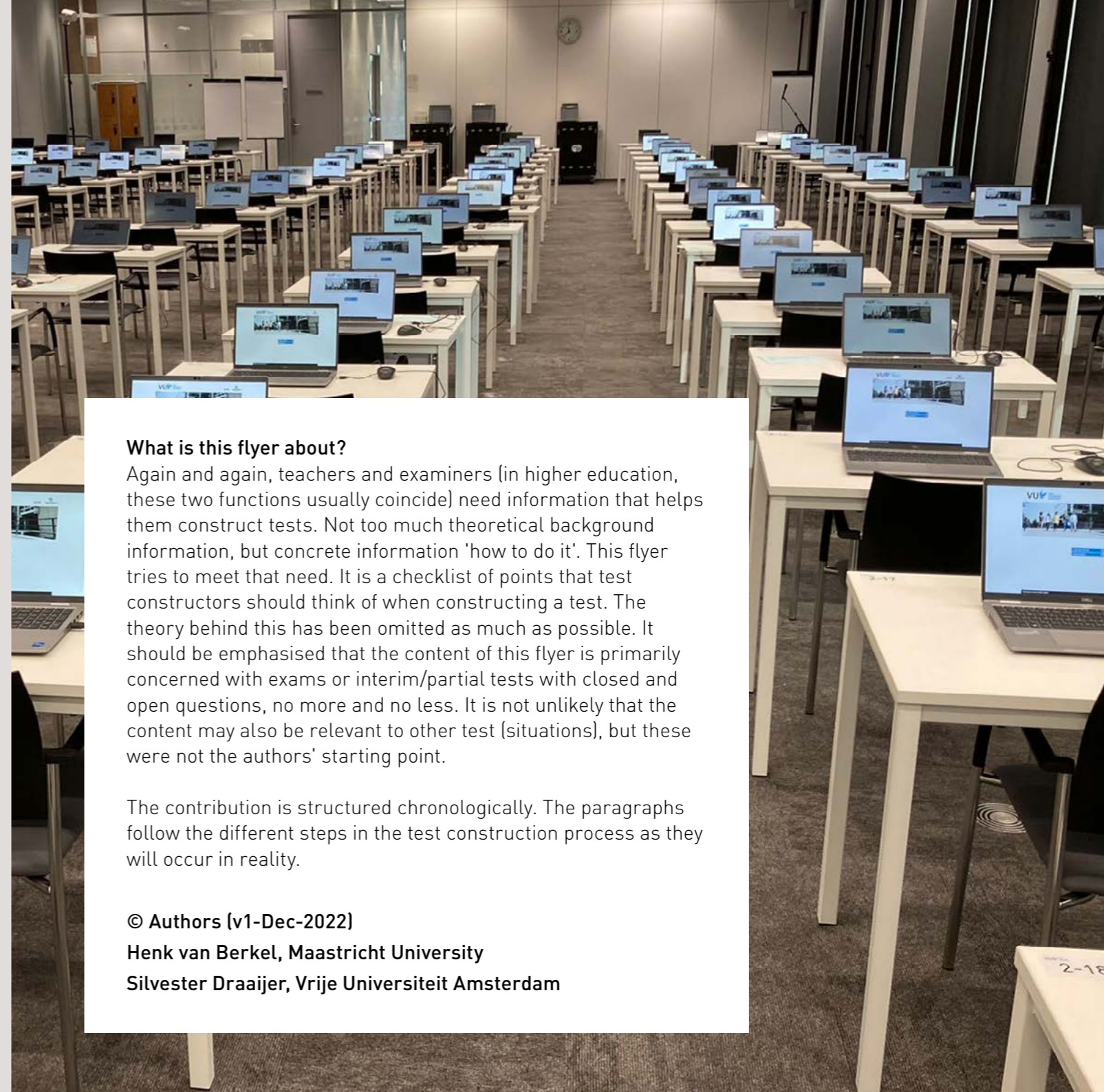
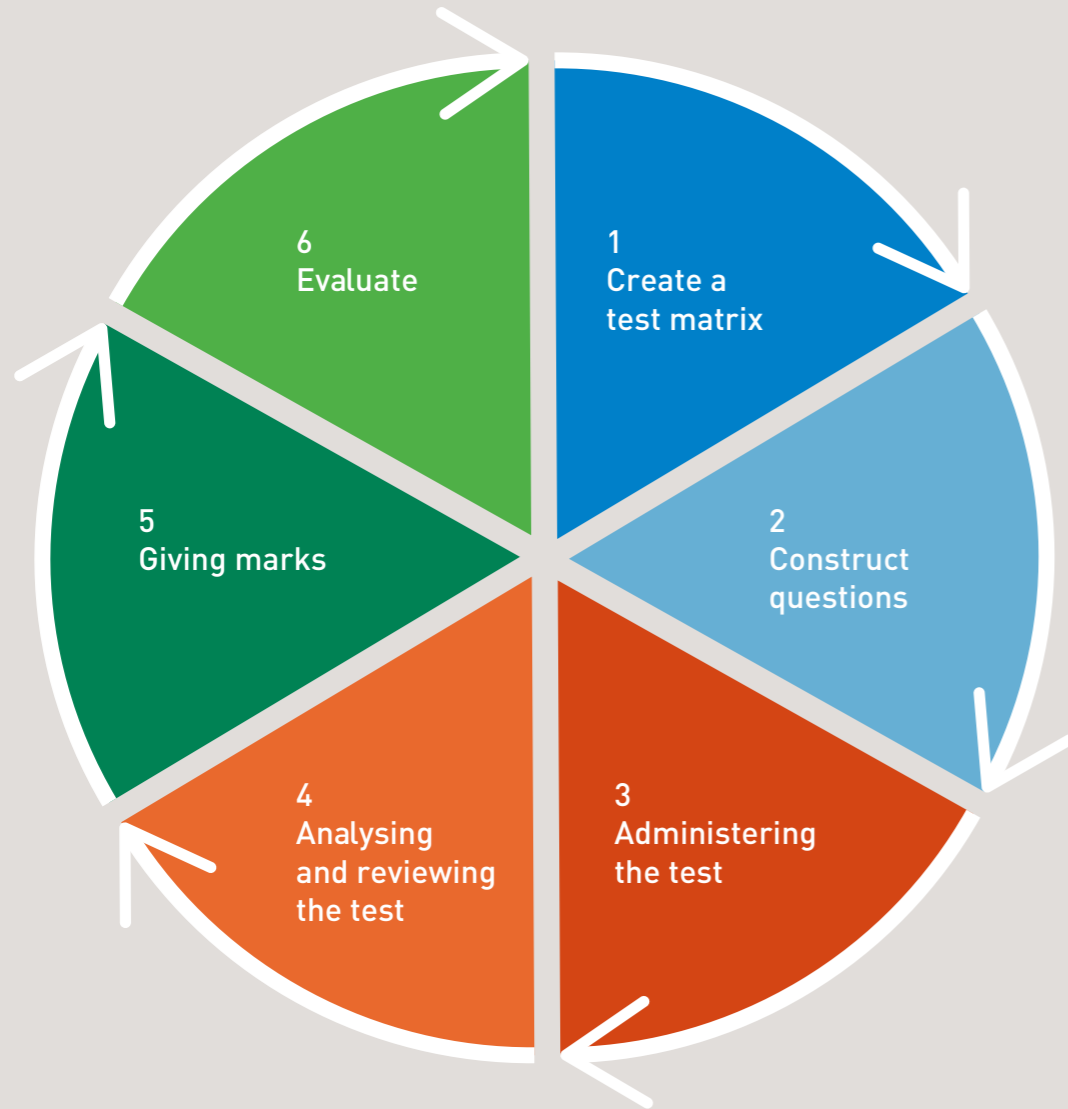


# Quickstart Test Development

Design and analysis of tests and test questions



# Quickstart Test Development



## What is this flyer about?

Again and again, teachers and examiners (in higher education, these two functions usually coincide) need information that helps them construct tests. Not too much theoretical background information, but concrete information 'how to do it'. This flyer tries to meet that need. It is a checklist of points that test constructors should think of when constructing a test. The theory behind this has been omitted as much as possible. It should be emphasised that the content of this flyer is primarily concerned with exams or interim/partial tests with closed and open questions, no more and no less. It is not unlikely that the content may also be relevant to other test (situations), but these were not the authors' starting point.

The contribution is structured chronologically. The paragraphs follow the different steps in the test construction process as they will occur in reality.

© Authors (v1-Dec-2022)

Henk van Berkel, Maastricht University

Silvester Draaijer, Vrije Universiteit Amsterdam



# 1 | Create a test matrix, choose a question type and test length

A test matrix (specification table or blueprint) consists, on the one hand, of a number of important topics from the course you want to ask questions about and, on the other, of the cognitive level you want to target with the questions.

Topics	Knowing - Reproduction Test questions at the level of reproducing knowledge appeal to memory (factual knowledge). Facts need to be recognised or facts need to be related.	Application In application questions, students are presented with a problem that is new to them. The problem is solved by relating different knowledge elements to each other or to a formula, law or solution principle or procedure.	Insight Insight questions require students to have a high degree of reorganisation of the problem and independent thinking. The questions call on intellectual skills such as interpretation, analysis, synthesis and evaluation.
1			
2			
3			
Etc.			

## 1.1 Subject and number of questions

- ☑ List the main topics from the course, usually around 15 suffices.
- ☑ The number of questions is important for the representativeness (content validity) of the test. Proper distribution of questions across the cells in the specification table will ensure that the test is representative.
- ☑ The number of questions also matters for the reliability of the test (the more questions, the better).
- ☑ The more important you think the topic is, the more questions you should ask about it.
- ☑ Indicate in the cells the number of questions you want to ask on the relevant topic at the relevant level.
- ☑ Make sure at least one question is asked on each topic.

How many questions can be asked depends on the format and test time.

## 1.2 Open question

Type of question	Response time	Sample question
Insertion and completion question	½ minute	What is the part of the inner ear in which Corti's organ is located?  That is the/the _____.
Short answer question	one minute	According to the New Testament, what did Pilate do with his hands after condemning Christ to death?
Long answer question	Ten minutes	What were the long-term effects (name three) of the Crusades? Reason your answer.
Essay question / argument question	25 minutes	Reason the trends over the 10-year period of unemployment if all wages are halved in one fell swoop. Assume no other changes in the business cycle.
Calculation question	Depending on the complexity of the issue	

## 1.3 Closed question

Type	Response time	Number of questions normally needed for a reliable test
Correct/incorrect question or two alternatives	50 seconds	80
Three options	60 seconds	60
Four options	75 seconds	40



Some other considerations:

- It is a persistent misconception that closed questions can only be used to question factual knowledge.
- It is a persistent misconception that questions on factual knowledge automatically result in high scores and questions on application or understanding in low scores.
- Long open test questions and essay questions take you a very long time to check. Moreover, it has been proven that all sorts of things go wrong when reviewing open-ended questions that affect the reliability (and hence fairness) of the assessment.
- Closed questions are more effective than open questions if more than 50 students participate in the test.
- Long answer test questions take a lot of time to answer, allowing fewer different topics to be questioned in the test time. This can quickly compromise the representativeness of the learning objectives in the test.



## 2 | Constructing and checking questions

Sources of inspiration:

Sources of inspiration for constructing questions are:

- The learning outcomes the problems/tasks in textbooks, syllabi, etc.
- Main points from literature and lectures.
- The future profession.
- Skills and practicals during the course.
- Discussions in lectures, working lectures, etc.
- Questions from tests already taken (especially questions that are not too difficult or easy and distinguish well between students who have mastered the material and those who have not).
- Graphs, tables, diagrams and other images of relevant features or processes can serve to develop multiple questions.

If you want to follow the common format know/reproduce - apply - understand, use certain **verbs**, for example, to develop questions.

<b>Knowing -Reproduction</b>	Name, Describe, Cite, Define, Extrapolate, Identify, Interpret, Distinguish, Enumerate, Paraphrase, Summarise, Estimate, Select, Explain, Translate, Tell in your own words
<b>Application</b>	Calculate, Demonstrate, Use, Make, Develop, Solve, Organise, Produce, Relate, Transfer, Change, Prepare, Modify
<b>Insight</b>	Criticise, Categorise, Compose, Conclude, Contrast, Deduce, Formulate, Rewrite, Illustrate, Interpret, Make, Distinguish, Support, Unravel, Design, Justify, Relate, Summarise, Outline, Explain, Validate, Defend, Compare, Explain, Appreciate

The **starting sentences** below also provide good guidance when formulating questions, especially those that aim to measure a high cognitive level.

<b>Knowing:</b> Checking the student's knowledge of topics	What is the best definition for ....? What is (not) characteristic of ....? What parts does the problem consist of? What is the history of the problem? What are the different categories in the problem?
<b>Critical thinking (evaluation):</b> Checking whether the student can use the properties of facts, procedures, principles or theories.	What is most effective (appropriate) for ....? Which is better (worse) ....? What is most effective for ....? What is the most critical step in to procedure? If you know that X is true, what is also true about Y? What is (not) necessary in a procedure? What is the importance of the problem ?
<b>Critical thinking (predicting):</b> Checking whether the student can indicate consequences, consequences, etc. based on facts, procedures, principles or theories.zijn.	What would happen if ....? If this happens, what would you do? Based on what ....., what would you do? Given ... what is main reason that ....
<b>Problem solving (given a scenario):</b> Assess whether the student can provide solutions or assess solutions based on given problems.	What is the nature of the problem? What do you need to solve this problem? What is a possible solution? What is the most effective (efficient) solution? Why is .... the most effective (efficient) solution?



#### Tips:

- Construct at least 25% more questions than you eventually want to include in the test. You will always deem questions incorrect after the first round of development.
- Think ahead, if you also already need to develop a resit exam or tests for consecutive years: immediately develop multiple variant questions of the same type.
- Use a question bank programme to store your questions.

## 2.1 After the construction phase

- ☑ Devote at least one discussion with others to the draft questions (four-eye principle) in which the content, the answer model or the correct alternative and the form are up for discussion.

## 2.2 Response models

An answer model is a tool for the reviewer of open-ended questions that serves to increase the reliability of reviewing. It includes the following items:

- ☑ Keyword listing of desirable answers, and their point allocations
- ☑ Procedure when the answer does not appear in the listed keyword list
- ☑ What to do with language errors?
- ☑ For long answers: content, technical design and argumentation, and a point award on parts

## 2.3 Checklists

Use the checklist below **after you have** constructed the questions.

### Content

- ☑ Is the chosen test format (open or closed) the most appropriate?
- ☑ Does the test consist of a sufficient number of questions?
- ☑ Does the question contain only one obvious problem?
- ☑ Does the question not contain subjective statements?
- ☑ When a quote is included in the question, is there context?
- ☑ Does the question not contain unnecessary information?
- ☑ Doesn't the question unnecessarily relate to a detail?
- ☑ Is the question not a trick question?
- ☑ In a statement question, is the statement actually 100% correct or 100% incorrect?
- ☑ Does the question contain sufficient information for answering?

- ☑ Does the question not contain complicated sentence structures?
- ☑ Can the wording of the question not give rise to misunderstandings?
- ☑ Are negations like not underlined or italicised?
- ☑ Does the question not contain double negations?
- ☑ Does the question not contain words like always, never, usually, definitely?
- ☑ Have the conventions regarding spelling, symbol use, punctuation, and the like been observed?
- ☑ Is the question divisible into a data and a question part?

### Process

- ☑ Has a test matrix been created?
- ☑ Did any peer review take place?

### Form

- ☑ Does the question not contain any vague indications?
- ☑ Isn't the question formulated in an unnecessarily negative way?

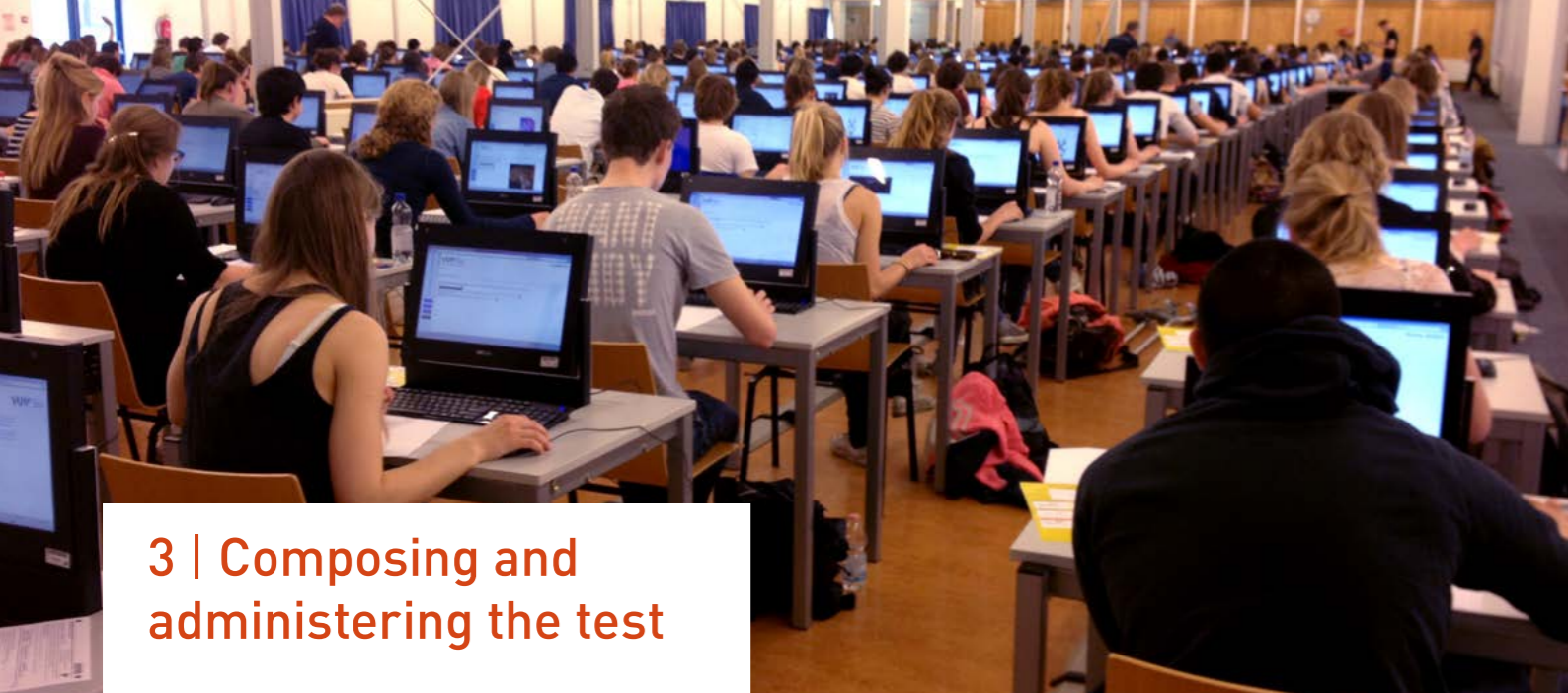
### Open questions

- ☑ Does the question give enough information about the desired length and shape of the answer?
- ☑ Is it clear that an answer needs to be explained/explained?
- ☑ Has an answer model been created?

### Closed questions

- ☑ Are there no overlaps in the options (answers that are subsets of each other cause a lot of confusion)?
- ☑ Are all the options about the same length?
- ☑ Are the options placed in ascending/alphabetical order?
- ☑ Are all the options plausible?
- ☑ Is the form 'none of the above' or 'all of the above' really necessary?





## 3 | Composing and administering the test

### 3.1 Compose

A test is more than a collection of individual test questions.

- ☑ From the approved questions, compile the test using the test matrix.
- ☑ Cluster the questions by topic in the order in which they came up in teaching.

### 3.2 The intake

Student test instructions for pen-and-paper tests should include information on the following topics:

- ☑ Duration
- ☑ Closed questions: fill-in instructions (e.g. tick the dash or selection round)

- ☑ Weight of the questions (for open questions, the weight may vary per question; for closed questions, each question counts equally in principle)
- ☑ Method to set the pass/fail score (cut-of score)
- ☑ Announcement of results: date
- ☑ Consulting resources: which ones are/are not allowed?
- ☑ Ban on use of mobile phones and other devices
- ☑ For paper-based tests: the number of pages and the number of questions in the test; students are responsible for any misprints
- ☑ Reference to fraud regulations
- ☑ What to do in case of problems, e.g. raise hand
- ☑ Regulation of toilet visits
- ☑ Submission instruction
- ☑ Possibility to make comments



Tip:

A test with both multiple-choice and open-ended questions can have advantages.

- Multiple-choice questions ensure that in relatively little time students can be quizzed on a wide scale of topics that are also automatically scored by the assessment system.
- The (short) open-ended questions lend themselves well to in-depth questioning and prevent the more superficial studying that occurs more quickly in tests with only multiple-choice questions.

## 4 | Analysing and reviewing the test

### Open questions

1. Use the pre-made answer template.
2. Check the test 'crosswise'. That is, by question and not by student.
3. From time to time, change the review order and, if possible, use the anonymise function of the review programme.
4. In practice, it will often not be possible to have two assessors review all questions, but at least review the test of students who score around the fail/pass threshold. If necessary, average both assessments.
5. Round up the scores only at the end.
6. You can also easily make a test and item analysis of an open-ended question test.

7. Collect the student comments, put them alongside the test analysis if necessary and discuss them with the question constructor.
8. Make a decision on the answer templates: maintain, change or add to them
9. If warranted, review the questions again using the updated answer model.

### Closed questions

1. In the digital examination system, you can perform the analysis yourself. When using paper answer forms that can be scanned, the exam service will carry out the analysis for you.
2. Collect the student comments and put them alongside the test and item analysis and discuss them with the question constructor.
3. Make a decision on the questions: keep, remove or key change. Realise that this change will generally produce only minor changes in the outcome.
4. Get another analysis done.

## 5 | Giving marks

There are many methods of grading which fall under three categories: absolute, relative and the compromise method.


Absolutely	With the absolute method, you determine the pass/fail score (cut-of score) in advance. This limit is dictated by the principle of checking whether students meet the requirements that can be derived from the objectives of the education.
Relatively	The relative method is based on the idea that the test should be tailored to students who are entitled to be educated - and in principle, they should be considered majority capable of passing a test. Since it is not known in advance what students are capable of in the <i>given</i> education and test, the standard <i>cannot</i> be set in advance. The result on the test must be awaited to set the standard.
Compromise	The compromise method tries to bridge the principle differences between absolute and relative norms and is preferred in educational practice. The compromise method generally assumes an absolute standard and specifies the circumstances under which this absolute standard will be deviated from. One of the compromise methods is described below.

### Open questions

1. Start by defining the pass/fail boundary. It is not unusual to place it at 55% or 60% of the maximum obtainable score.
  2. If you feel that too many students are failing your test, the following adjustments to the procedure are possible (after consultation with the examination board).
    - a. Instead of 60%, set the limit at 50-55% of the maximum score and/or
    - b. You do not take the maximum score as your starting point, but the average of the five highest scores.
  3. You now have the raw score yielding 0 and the raw score yielding 10: you determine the remaining marks by dividing the raw scores into 10 equal units with a mark attached.
    - The probability score for multiple-choice questions is the number of valid questions in the test divided by the number of alternatives per question. For example, the probability score with 105 correct/incorrect questions is  $105/2=52.5$ .
    - In the said test, if the highest score is 96%, the pass/fail boundary comes to 60% of the score between 52.5 and 96 = 78.6.
2. Students with the highest score (96) get a 10, those with the probability score a 0. Linear regression is used to calculate the remaining grades.
  3. If you feel that too many students fail your test: the following adjustments to the procedure are possible after consultation with the examination board:
    - a. Instead of 60%, set the limit at 50-55% of the maximum score and/or
    - b. You don't take the maximum score as the starting point, but the average of the five highest scores.

### Closed questions

1. Start by determining the pass/fail boundary. In principle, you set this at 60% of the highest score, taking into account the probability score.



"A fair test gives students who have studied sufficiently a good chance of passing"





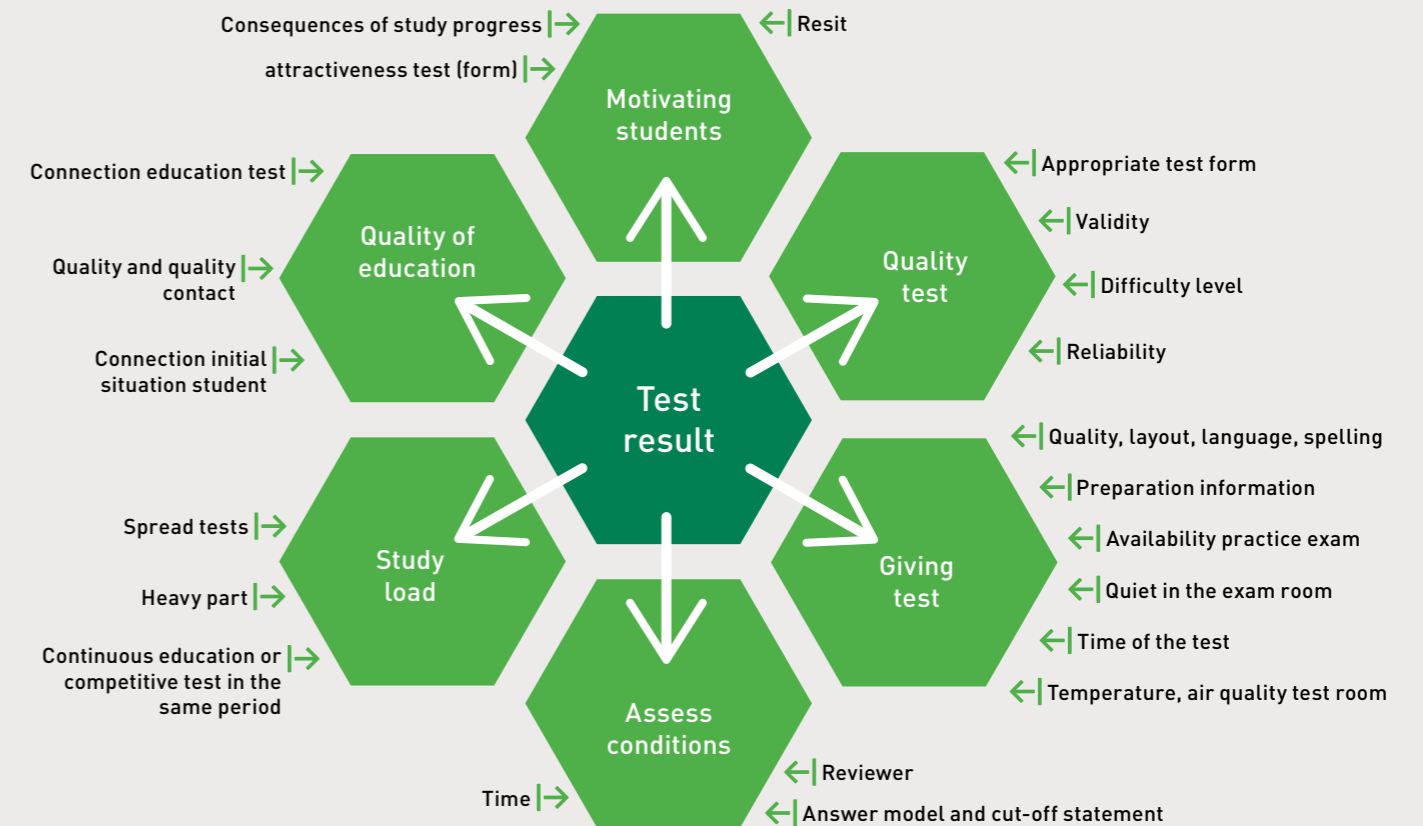
## 6 | Evaluating the test

The (online) report contains quality indices covering the test as a whole and individual questions. A study of report provides information to make improvements in the test next year:

1. Look at the reliability of the test. If it is lower than 0.70, then it is recommended to include more questions in the test next year.
2. Study the very difficult and easy questions. Find causes for these so you can avoid such questions next time.
3. Examine psychometrically very good questions (that is, questions that distinguish well between competent and non-competent students). This is shown by the so-called Rit value (or discrimination value) of the question. Such questions are worth including in the next test as well.

Consider that the result on a test is influenced by many aspects.

Study the figure below which lists a number of factors that can influence scores and test reliability in both positive and negative directions.





Commissioned by VU Network for Teaching and Learning and VU SOZ Procesregie

[vu.nl/en/education/vu-ntl](https://vu.nl/en/education/vu-ntl)



Want to learn more?

[vu.nl/en/employee/didactics/active-blended-learning-in-practice-at-vu-amsterdam](https://vu.nl/en/employee/didactics/active-blended-learning-in-practice-at-vu-amsterdam)



Learn how the assessment process at VU Amsterdam is designed and what the responsibilities of different people are:  
[vu.nl/en/employee/testing-and-assessing/exam-organisation](https://vu.nl/en/employee/testing-and-assessing/exam-organisation)



Learn how the VU Amsterdam computer based exam system TestVision works:  
[canvas.vu.nl/courses/31787](https://canvas.vu.nl/courses/31787)