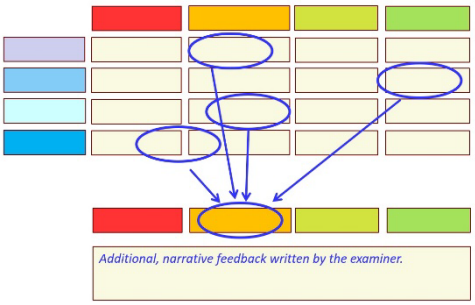
Aan:	**Education developers**
Van:	Willibrord Huisman, educational consultant
Ter:	Consideration, use, reaction, discussion, referencing, completion, correction
Status:	This is an 'inspiration document' written for the revision of the medical and biomedical programs of Radboud University; is has been used elsewhere as well since
Datum:	Versie 1.00, 27 februari 2015, translation 4 April 2022
	You'll find the latest version of this document www.willibrordhuisman.nl/onderwijsadvies/65

Aligning and Assessing with Rubrics



Wat this is

This 'inspiration document' is intended for education developers who are developing new teaching or wish to improve existing teaching and are considering using rubrics for that purpose.

Inhoud

1. About rubrics 2
1.1 What is a rubric? 2
1.2 Why are rubrics so 'hot'? 3
1.3 Who uses the rubric? 3
1.4 How to develop a rubric? 6

2. What does a rubric look like? 8
2.1 What does it say from top to bottom? 8
2.2 What is written from left to right? 8
2.3 What is written in the cells? 9
2.4 What is the writing space for? 9
2.5 How do you calculate the final assessment? 10
Calculated total score 10
From total score to final assessment 11
Compensate, but don't forget 11
Rare borderline cases 11

3. Further study 12

1. About rubrics

1.1 What is a rubric?

A rubric, also known as a scoring rubric, is a *specified* instruction for the assessment of a performance (usually involving a behavior or product).

The rubric usually takes the form of a pre-filled table. In the table, each row refers to one *criterion* and each column to a *score (rating)*, for instance, a grade, or an indication such as 'fail' or 'very good', or an advancement level, such as 'beginner' or 'expert'. Each cell contains a *descriptor* that describes the quality of the performance that leads to the relevant score for that criterion.

Score Criterion	Fail	Pass	Good
Visual	A PPT or other visual aid is missing, or is present but distracts from or contributes little.	Use of PPT, board or other media generally aids the presentation.	Use of PPT, board or other media makes a good contribution to the presentation.
Structure	The presentation has no recognizable structure.	The presentation has a clear structure and generally works.	The structure of the presentation is well thought out and is consistently maintained.
Contents	The speaker has no clear point to make or the point is incoherent.	The gist of the argument is largely clear and coherent.	The argument is coherent and contains new and enlightening points of view.
Speaker	The speaker comes across as uninspired and/or doesn't gauge his/her audience well.	The speaker is intelligible and clear and gauges his/her audience relatively well.	The speaker is convincing and interacts well with the audience.

Quick – thought up in half an hour – example of a rubric for four criteria at three rating levels and thus twelve descriptors. You will probably see all kinds of flaws and points for discussion right away. Improving such a rubric with colleagues, students, and external stakeholders not only creates a tool that is useful for everyone, but also leads to greater understanding of mutual expectations and gives clearer direction to learning, teaching, and assessment.

For each criterion, the assessor decides which descriptor best fits the student's performance. In essence, this amounts to circling or clicking on the table cells that apply. This then leads to a score for each criterion. If necessary, a further weighting can be applied to the criteria themselves. The sum is the score of the performance as a whole.

Such a rubric can be used for appraisals, but also (and above all) for educational assessment, possibly by the students themselves or by students among themselves, as well as for orientation towards the objectives. The major advantage over empty assessment tables is that the text in the cells, if properly worded, provides a significant part of the justification for the decision and the narrative feedback.

The word 'rubric' originally means 'liturgical prescription'. This is written in red (*ruber*) letters. It's confusing because the Dutch word 'rubriek' also means 'category'. This may be the reason why 'rubric' is usually not translated in the Netherlands.

1.2 Why are rubrics so 'hot'?

Rubrics render explicit what was previously only implicit.

In traditional education, students follow a course, subconsciously distill the subject matter from the lectures and books, study diligently, take an exam, have an interview and receive a grade. It may have been clear to the lecturers what it was all about and how it was assessed, but it was not written down and thus not explicit. For the students, it was generally vague. If they got a failing grade, students' conclusions might be 'I was unlucky with this examiner' or 'I didn't study hard enough', or worse, 'I'm not smart enough for this'.

Meanwhile, to lecturers it is usually clear what it is all about. They nowadays formulate final qualifications and course objectives and derive the assessment and teaching from these. However, for students, this only becomes meaningful once they have completed the course. What they need is a *prior* description of the quality of the expected performance. They can gauge themselves against this standard and from there they can choose their own study path.

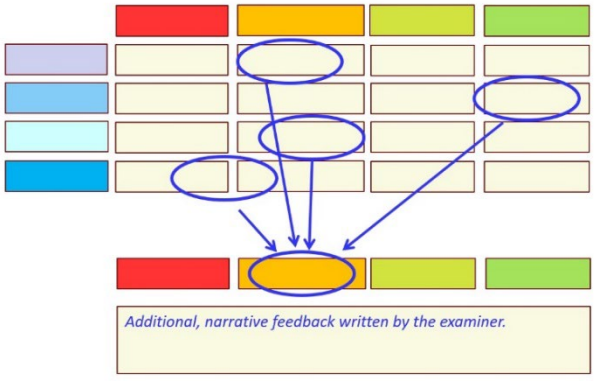
The development and subsequent use of rubrics for educational and assessment purposes ties all this together because students, lecturers, and assessors use *the same* rubric.

1.3 Who uses the rubric?

a) **Education developers** (lecturers, students, professionals in the field) develop the rubric, often in an intensive and iterative process of development, testing, evaluation, adjustment, etc. This makes it clear to everyone what the actual learning objectives, the relevant assessment criteria, and the expected performance levels are.

b) **Examiners** use the rubric to be able to give as well-founded, valid and reliable a decision as possible about a student's performance, and be able to underpin that decision to the student.

- "Here look. From the circled cells, you can see how I rated your performance on those criteria and how I arrived at this decision."
- "On top of that, in the field below the rubric, I've written a specific point for improvement for you that you might incorporate into your next courses."



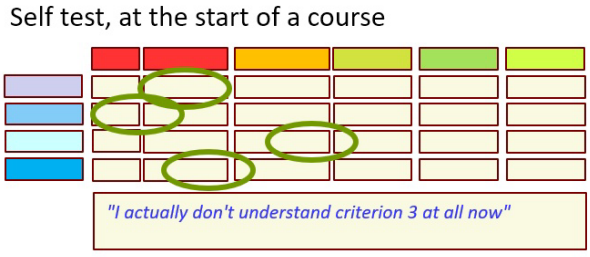
c) **Examiners** use the rubric to determine how to administer the assessment.

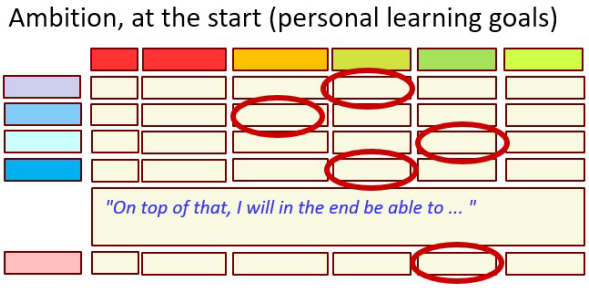
- "I have now interviewed this student carefully, but in the end I still don't have enough information for two criteria to be able to give a rating there. I have to ask the questions differently. Maybe I'll just look how another examiner does it."

d) **Supervisors** (mentors, content experts) use the rubric to examine how the student or students are performing on some or all of the criteria. They adjust their supervision and other educational activities to accommodate this.

e) **Students** use the rubric to focus and adjust their study behaviors to be (more) in line with the objectives. Independently or in mutual interaction, they use it to estimate how they are performing on the criteria, where they need to continue improving, and what they are happy with.

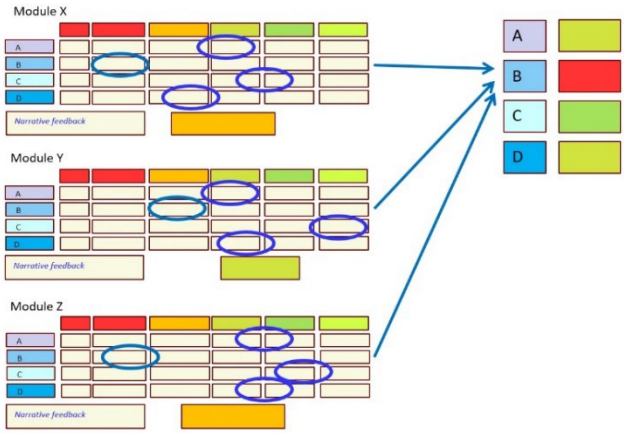
f) **Students** use the rubric at the beginning of a course (or other unit of study) to determine their starting position and personal learning objectives (expectations, ambitions) within the framework of the objectives set by the course (and objectives to be assessed).





- "I think I'm currently underperforming on this criterion; so I want to work hard on that."
- "On this criterion, I want to perform at a high level because that's what I really want to pursue."
- "I'm satisfied with a pass on these points because I'm not going to pursue this path later anyway."
- "Moreover, by doing this course, I want to achieve that and that, although that is not part of the learning objectives. I'm going to add that as an additional criterion to the rubric."

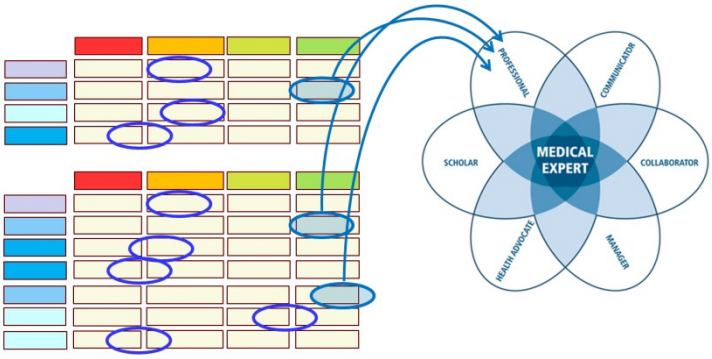
g) After an appraisal, **students** receive not just a grade, but the rubric with the ratings, the final assessment, and any further personal feedback. In doing so, they evaluate their progress and draw conclusions about their subsequent activities.



- "The final assessments I receive are always pass or good, but on criteria related to manual skills I always score just a pass and sometimes even a fail. I need more practice with that and should really ask for guidance."

h) **Students** can link the partial ratings of certain types of rubrics to competencies and thus demonstrate how those competencies are developing.

- "The rating I received for the 'patient handling' criterion from this internship shows how far along I am in the 'professional conduct' competency."

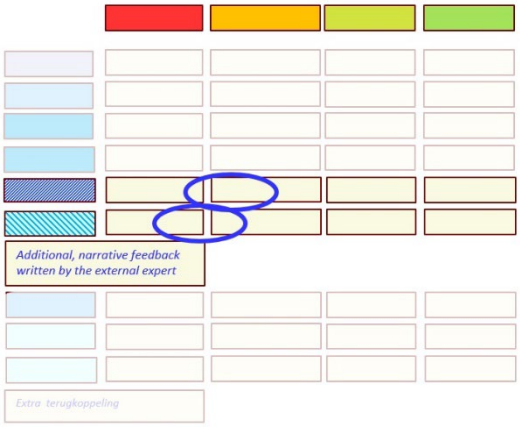


i) **Education developers** use the detailed scores from all rubrics to draw conclusions about the effectiveness of the study program.

- "On criteria related to independent writing, three-quarters of the students generally scored low. We need to pay more attention to that."

j) **External experts** assess student performance on certain criteria without having to immerse themselves in the other criteria.

- "As a patient, I notice that although this student is a little clumsy with his words, he is able to instill confidence."
- "As a speech therapist, I hear that this student will develop vocal issues if he continues like this. I now rate his 'use of voice' as a fail."



k) **Testing experts** analyze the ratings statistically and can thus make statements about the inter-assessor reliability at the criterion level.

- "On this criterion, examiner A's ratings are, on average, considerably lower than examiner B's. Apparently something is not clear about the descriptors."

l) **Assessment committees** and other representatives from society or the professional field can use the rubrics to gain a quick overview of the relevant criteria, their significance, and the expected levels.

- "They pay a lot of attention to this aspect and the students also have to perform quite well to get a passing grade on this."

1.4 How to develop a rubric?

Developing a rubric is similar to developing a test with a grading scheme. Only there's nothing secret about it: students have an active role.

- Form a development team of stakeholders. This could include:
  - education developers,
  - testing experts and examiners,
  - supervisors,
  - students,
  - representatives from the professional field.
- Look around you in your close and extended surroundings. There is a good chance that rubrics have already been created elsewhere that you can adopt or use as a source of inspiration. See ↓ 3 Further reading.
- Find out what the program's policy is regarding rubrics. Certain rubric design variables may be better established at the program level. For example: the grading scale, the way in which the final assessment is calculated from the total score, and the definition of criteria that repeat in other rubrics.
- Examine the objectives of the unit of study.
  - Look at the relationship with the final study program qualifications.
  - Look at the formulations of the objectives: are they sufficiently concrete?
  - Are there any criteria and/or pre-existing grading schemes?
- Set criteria.
  - How can one tell the extent to which the student has achieved the objectives?
  - Aggregate them into a manageable number.
  - Try to describe them in keywords; see also ↓ 2.1
- Set the standard.
  - For each criterion, what does a performance look like that we consider to be just above a passing grade?
  - Describe this in one or two sentences. This should generate a column of descriptors.
- Determine what 'fail' and 'good' constitute (or other or additional ratings, see ↓ 2.2)
  - Same as setting the standard. Avoid relative references to the other descriptors (i.e. not: better than...)
  - Describe the descriptors and compare them to those of the other ratings. Are they similar in nature? Are they distinctive enough?
- Describe borderline cases
  - Examine when a performance on a criterion is so bad that the indication 'fail' is not an adequate description, and examine what makes a performance excellent. It often makes no sense to create separate columns with descriptors to this end. Describe or refer to separate regulations (see ↓ 2.5).

- Determine weighting and assessment method
  - Consider whether the criteria are equally important; if not, classify them differently (see ↓ 2.1), or determine a weighting method (see ↓ 2.5).
  - Determine how the total score will be calculated.
  - Determine how the final assessment is determined (see ↓ 2.5).
- Iterate
  - Jump back to earlier steps in this development process. It often only becomes clear during the formulation of the descriptors what the actual objectives are.
  - If necessary, revise the objectives or their formulation so they are more in line with what is being tested and assessed.
- Whittle it down
  - There is a tendency in rubric discussions to make ever more stipulations and refinements. In the end, all that detailing is counterproductive. Summarize, aggregate, leave some room for interpretation. Aim for a single A4 sheet.
- Test
  - Present the rubric to people outside the development team and let them practice with it. Observe what happens, ask for reactions, perhaps analyze how the rubric is applied.
  - Discuss the findings and adjust where necessary.
- A rubric is never really finished.
  - Record the rubric as it will be used this year.
  - During the year, collect material that clearly fits a particular descriptor and material that has raised questions.
  - Evaluate at the end of the year and make adjustments.

2. What does a rubric look like?

2.1 What does it say from top to bottom?

In a rubric, the assessment is broken down into a number of **aspects** of the performance; preferably a limited number. Ideally, these should not overlap. How this is broken down depends on the nature of the assessment.

- **Criteria** are the benchmarks by which measurements are made. A presentation is judged for structure, logical consistency, originality, language use, presentation, etc.
- If the rubric describes a skill that is assessed on several performances, a classification by the **test items** may be practical. If necessary, this may be followed by criteria that go beyond these elements.
- **Requirements** do not belong in a rubric. If a requirement is not met, there will be no assessment. For instance, a requirement is that the summary is handed in one week in advance.

Investigate what criteria are used elsewhere in the study program and how they are named. Use the same as far as possible; this will make it easier to aggregate partial ratings across rubrics. (See ↑ 1.3 g and i)

Sometimes certain criteria are more important than others.

- You can address this with a **weighting**. Give each row a weighting factor, or state the number of points it is worth in each cell. This weighting is taken into account in the final calculation.
- You can also **split** an important criterion into two, or **merge** two not very important ones if that is sensible in terms of content. In this way, the rubric remains comprehensible and there is no need for numerical weighting.

2.2 What is written from left to right?

Performance is scored for each criterion on a scale. This is not necessarily the same as the scale on which the final assessment is expressed.

Broadly speaking, three types of ratings occur:

- Expression of the *quality* of performance on the criterion in question. For example:
  - fail – pass – good
  - idem, for grades: 4 – 6 – 8
  - poor – fail – just pass – comfortable pass – good – very good - excellent
    (= in Dutch: slecht – onvoldoende – juist voldoende – ruim voldoende – goed – zeer goed - uitmuntend, generally graded as 4 to 10)
  - *4 and less* – 5 – 6 – 7 – 8 – *9 and better*

- Expression of the level of *achievement* from the one who delivers such a performance in that curriculum. For example:
  - emerging - developing - proficient - advanced
  - beginner - advanced - expert
  - level 1 - level 2 - level 3
- Neutral scale: more points = better quality or more advanced
  - 1 – 2 – 3 – 4

2.3 What is written in the cells?

In principle, each descriptor consists of one or more sentences describing the quality of the performance on the relevant criterion. The assessment of the student's performance could then consist of the collected circled descriptors; it should be understandable even without a rubric.

In practice, this can lead to a lot of text and a lot of repetition in adjacent cells, especially if the rubric has numerous columns. It becomes unwieldy and means a lot of dead weight for the assessors (who work with the rubric a lot). There are three possible ways to cope with this.

- In the following cells, only add additional information not included in the cells to the left. This works for rubrics that describe levels of advancement. For example, the following message could appear above the rubric: "Each level encapsulates the lower level; only new competencies are listed."
- Use concise terms and employ footnotes to explain what can and cannot be understood by those terms. Usually, reading those footnotes once should be enough.
- Rephrase each criterion so it becomes a half sentence ending with "...". The descriptors in the cells then form variants of the second half of that sentence.

  Example: the criterion 'scientific basis for your diagnosis', tested in an oral exam.
  Your reference to medical theory...
  cell with value 5: *... is often incorrect* or irrelevant*.*
  cell with value 7: *... is largely correct and relevant.*
  cell with value 9: *... is correct, critical and relevant and has been arrived at independently.*
  (* make sure to clarify what *incorrect* and *relevant* mean outside of the rubric).

Furthermore, it is advisable to provide descriptors with characteristic and unambiguous **example material** in the form of text, images or video clips. For the assessors, this should lead to greater reliability in their ratings. While for the students, it is good orientation and practice material; for instance, let them first decide among themselves which example fits which descriptor.

Gathering such material is not easy. It could be an extra focus point for assessors: identify examples – could be bad examples, just make sure they are clear – so they can be used in the future.

2.4 What is the writing space for?

The rubric does not cover everything; and even if it did, it is inconvenient for users to be allowed to circle or tap only cells and nothing else. On paper, you can still write in the margins, but in a digital rubric you can only do so if there is a text field available.

Some reasons to create a text field.

- The text field invites **assessors** to provide additional feedback or justification in addition to the partial ratings and the final assessment.
- It invites **students** to look beyond the items displayed in the rubric. In the orientation phase, they can describe additional ambitions there. (In ↑ 1.3 f, a student even drew an additional row).
- For **fellow students** and **supervisors** who use a rubric in the learning phase to assess a student's performance, the text field provides an opportunity to give additional feedback.
- For the **development team**, the existence of the text field is a good argument for not wanting to encapsulate everything in the rubric. The team may give assessors guidance on assessment and feedback issues that can be addressed in the text box.
- For **evaluation purposes**, a second text field could be considered in which assessors make notes of problems they encountered during the appraisal.

The development team must carefully consider and specify what happens to the text in the various text fields: who can write, who can read, how is it stored, etc.

2.5 How do you calculate the final assessment?

Because a rubric is a table into which cells are selected by the assessor, an overall score can be calculated from it. The examiner determines the final assessment on this basis.

Please note: For the sake of clarity, we will try to make a distinction between the final assessment *process* (the 'final appraisal' - in Dutch 'eindbeoordeling') and its *result*, the 'final assessment', in Dutch 'eindoordeel'. We avoid the term 'final grade' because the final assessment is not always expressed as a number. Many find 'assessment' (eindoordeel) sounds too harsh and also use 'appraisal' for the assessment result.

Calculated total score

- If the ratings are expressed in words, they must also be translated into numbers in order to be able to calculate them. Fail – pass – good then usually becomes 4 - 6 - 8.
- In order to grade the results, the examiner adds the values of the selected cells and divides them by the number of rows. This produces the total score.
- A weighting may be applied to specific criteria (= rows). Especially when doing manual calculations, a clear system should be chosen that reduces the risk of errors. Two examples:
  - Don't work with 4 - 6 - 8 but with points, e.g. 1 – 2 – 3. For a criterion that is weighted double, the points are then 2 - 4 - 6. The points should then be listed *in* the cells. Create a table that translates the number of points achieved (the total score) to the scale used by the examiner (e.g. 32 points = 6, 40 points = 8, etc.)
  - State a weighting factor per criterion (= per row). Assume that the examiner correctly applies this factor when calculating the total score.

  The disadvantage of weighting is that it makes the rubric 'optically' incorrect: a circle around a cell of an extra weighted criterion looks the same as one around a less critical criterion. Consider splitting or merging criteria (see ↑ 2.1).

From total score to final assessment

There are several ways to get from the calculated total score to the final assessment.

- The final assessment is by definition *equal* to the calculated score.
- The examiner gives the final assessment after having *seen* the calculated score. In this way, there is still room for assessment of aspects that are not well covered by the rubric. If the examiner deviates from the calculated score, he/she shall write an explanation.
- Finally, it is possible *not to* calculate *a total score* at all. The examiner gives a *holistic* final assessment as usual, but only after having circled the partial ratings in the rubric. In this way, the rubric is more of a tool than a prescription.

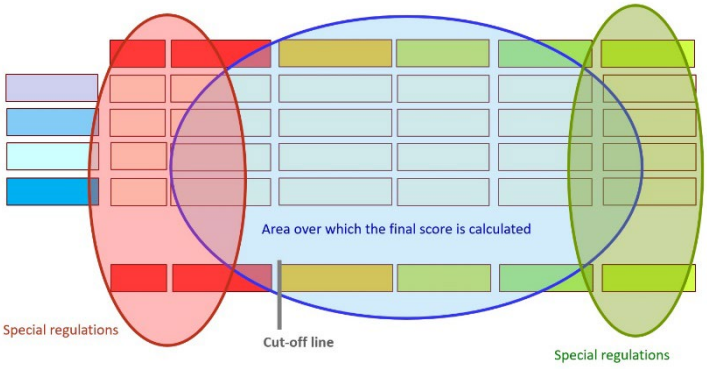
Compensate, but don't forget

A rubric shows that performance on certain criteria can be a fail while the total score is still a pass. One criterion is then compensating for the other. Sometimes assessors feel it is wrong to accept fail grades in this way. There are three things to note about this.

- In fact, this compensation also takes place in assessments without a rubric, only it is less in-your-face. "On this criterion you underperformed, but because those and those were very decent you still get a pass."
- Precisely because the fail criteria become visible by using the rubric, they can also be addressed. The student and the supervisor can do a lot with this.
- If the partial ratings of several rubric assessments are stored centrally, it may well come to light that a student consistently fails to score well on certain criteria. This can be used in the learning phase (see ↑ 1.3 g) and it could even lead to an additional assessment.

Rare borderline cases

The descriptors in a rubric mainly describe the most common cases. It is not worth writing out the extreme cases separately for each criterion: the exceptional performances and the very inadequate performances. Separate criteria, regulations and possibly measures are usually available for both.



It is important, however, to mention these cases in a rubric or in a supplement to one. Otherwise a situation will be created in which, in terms of grades, the 4 and the 10 are pushed completely out of the picture, and that is not the objective. It is precisely the extremes, in both the negative and positive sense, that require extra attention.

3. Further study

- Wikipedia,
  https://en.wikipedia.org/wiki/Rubric_(academic)

  The section *Steps to create a scoring rubric* (27 February 2015) includes an inspiring way to create a rubric based on student input.

- Do a quick Google search for images of rubrics. You will get a quick impression of the enormous diversity and perhaps some useful ideas.

- Stevens, D. & Levi, Antonia J. (2013). *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning*. Sterling, VA: Stylus Publishing.

- van den Berg, B.A.M.; van de Rijt, B.A.M.; Prinzie, P. (2014). *Beoordelen van academische schrijfvaardigheden met digitale rubrics.* Onderzoek van Onderwijs, volume 43, pp. 6 – 14
  http://dspace.library.uu.nl/handle/1874/293956

- Huisman, W. (2015). *Het beoordelen van klinisch redeneren. Mondeling tentamineren van coassistenten.* Latest version via www.willibrordhuisman.nl/onderwijsadvies/65

  This memorandum prepares examiners and students of the Medicine program at Radboud University for the assessment of clinical reasoning skills using a rubric.

- Huisman, W. (2015) *Toetsing, terugkoppeling, beoordeling.*
  Latest version via www.willibrordhuisman.nl/onderwijsadvies/67

  This self-study module in the form of annotated PowerPoint diagrams will give you a framework of concepts consistent with that used in this inspiration document. In Dutch.

==========