

Dilemmas in Designing Rubrics for the Assessment of Student Teachers in Teacher Education

Stories of a Design and Evaluation Process in the Context of Biology Teaching

Dineke Tigelaar,^{1*} Fred Janssen²

ABSTRACT

In the current paper, we present stories on the design and evaluation of scoring rubrics for the assessment of student teachers in teacher education, in particular, biology teaching. We describe dilemmas related to developing scoring rubrics for assessment of student teachers. Questions related to these dilemmas concern the level of detail in criteria, how to integrate theoretical and practical notions on effective teaching in criteria, how to decide on key aspects of teacher knowledge and functioning that should be assessed, and how the ideas and input of different stakeholders can be balanced. We also describe how we developed the criteria and standards for assessing our student teachers, taking these questions into account. We present stories on the design and evaluation process of our rubrics, which are illustrated by examples from assessment of two student teachers in biology.

KEYWORDS: teacher assessment, scoring rubrics, teacher education, biology teaching

Introduction

In the current paper, we present stories on the design and evaluation of scoring rubrics for the assessment of student teachers in teacher education, in particular, biology teaching.

In line with current notions on teaching, teacher assessment is seen as a longitudinal process involving various methods in order to gain a rich picture of teachers' knowledge and performance and to measure whether construction of meaning has taken place (Darling-Hammond & Snyder, 2000). Darling-Hammond and Snyder (2000) stress that teacher assessment should be contextualized and authentic. They mention four conditions for assessment of teachers (pp. 527, 528). First, teacher assessments should sample the actual knowledge, skills, and dispositions desired of teachers as they are used in teaching and learning contexts, rather than relying on more remote proxies. Second, teacher assessments require the integration of multiple kinds of knowledge and skills as

they are used in practice. Third, in teacher assessments, multiple sources of evidence are collected over time and in diverse contexts. Fourth, assessment evidence is evaluated by individuals with relevant expertise against criteria that matter for performance in the field of teacher education. Uhlenbeck, Verloop and Beijaard (2002) further outlined requirements for assessment of student teachers by reviewing various studies on conceptions of teaching, and on teacher assessment. Drawing on their findings, they formulated several specific requirements for assessment of teachers, such as, that both teachers' actions and their cognitions should be assessed, that assessments should measure what is actually demanded on the job, and that various sources of evidence should be collected over time, such as videotapes of lessons, written reflections by teachers on their performance and samples of students' work. In this paper, we focus on Uhlenbeck et al. (2002)'s requirements with regard to establishing criteria and standards on which teachers are judged. The requirements for setting criteria and standards for the assessment of teachers as summarized by Uhlenbeck, et al. (2002) are influenced by the widely accepted viewpoint that teaching is highly contextual and that there are numerous ways of good teaching (Dellandshere & Petrosky, 1998). This means that clear consensus as to what constitutes good teaching may not be possible, which implies that good teaching can never be fully defined and that there are no absolute criteria that can be used to evaluate teaching. Nevertheless, as is stressed in the review by

^{1*} Corresponding author: ICLON, Leiden University Graduate School of Teaching, PO Box 905, 2300 AX Leiden, The Netherlands

E-mail: DTigelaar@iclon.leidenuniv.nl

Tel.: +31 71 527 6552, Fax: +31 71 527 5342.

² ICLON, Leiden University Graduate School of Teaching, PO Box 905, 2300 AX Leiden, The Netherlands.

E-mail: FJanssen@iclon.leidenuniv.nl

Tel.: +31 71 527 6590, Fax: +31 71 527 5342.

Uhlenbeck, et al., (2002), it is acknowledged in the field of teacher evaluation that teacher assessments should be based on a conceptual framework that provides some sort of definition of what teachers should know and be able to do. Uhlenbeck et al. (2002), on the basis of what is known on good teaching, formulate some requirements for defining the content of criteria for the assessment of teachers, such as including criteria for assessing the different phases of teaching (preparing, and evaluating teaching as well as the actual teaching process), and including criteria for assessing reflection, and interaction with colleagues. In this paper, we focus on a number of the requirements as formulated by Uhlenbeck that may lead to dilemmas and questions regarding the design (form) of criteria and standards. In the following, we outline these dilemmas and questions.

First, since criteria should allow for a range of acceptable ways to teach, a careful balance must be struck between prescribing every detail of what will be assessed and leaving it completely open-ended (Darling, 2001). The question here is what should be the level of detail in criteria in order to allow for a range of acceptable ways to teach and yet to define what is unacceptable. Second, criteria need to be based on what is known about effective teaching, both from a theoretical perspective, as well as from the input of expert teachers' practical knowledge and theoretical notions need to be formulated and used in a way that is meaningful for practice. The question here is how theoretical and practical notions on effective teaching can be integrated in criteria. Third, with regards to standards, there is discussion on how standards are formulated: whether they describe in general terms what teachers should know and be able to do or whether they are subject- and level-specific and aligned to curriculum goals for students. In other words, the question here is how different standards should be used teachers in different levels, such as for licensed teachers and student teachers, and whether different standards should be used for teachers in different subject areas. Fourth, various stakeholders should be involved in defining teaching standards, and different stakeholders may have different ideas on what is important to assess. The question here is how the ideas and input of different stakeholders can be balanced in standards on how to decide on the key aspects of teacher knowledge and functioning that should be assessed.

In the current paper, we focus the development of criteria and standards for assessing student teachers in science teacher education. We developed scoring rubrics in which these criteria and standards were described. Andrade (2005) proposes that rubrics should be used above all for the assessment of students' work. Rubrics can also be used to give feedback, both during and at the end of the course. Feedback based on rubrics and information of student performance (e.g., a lesson observation, or a video-taped lesson) indicates the degree of success achieved and provide students with insight into what is expected of them as professionals. Rubrics must never be seen and used as a check-list with which an absolute score of

a student can be determined. Rather, rubrics can assist monitoring of students and can provide input for adjusting instruction to students' needs. In a rubric, the evaluation criteria that an assessor should consider when determining the quality of a student teacher 's functioning are made explicit (Reddy & Andrade, 2010). Furthermore, for each criterion, standards are described that provide a detailed explanation of what a student teacher must do to demonstrate attainment of a particular level of achievement, for example, poor, fair, good, or excellent. Research on scoring rubrics (Johnsson & Svingby, 2007; Reddy & Andrade, 2010) shows that reliable scoring in assessments can be supported by the use of rubrics, especially if they are analytic, topic-specific, and complemented with exemplars and/or assessor training. However, too much standardization of the rubrics and procedures for use would trivialize the assessment and would detract from validity. Johnson & Svingby (2007) stress that the use for rubrics does not enhance valid judgments in assessments per se. However, rubrics do seem to promote positive educational consequences by improving instruction and supporting learning by making expectations explicit and facilitating feedback and self-assessment.

In the following, we describe how we developed the criteria and standards for assessing our student teachers. We present stories of our design and evaluation process, illustrated by examples from assessment of two student teachers in biology.

Stories of a design and evaluation process: developing criteria and standards for the assessment of biology teachers in teacher education

Context

The criteria and standards were developed in the context of the teacher education program of the Leiden University Graduate School of Teaching. The teacher education program of Leiden University, the Netherlands, is a one year master program which is open for students with an appropriate master degree and for students with a bachelor degree students with teaching experience. The students spend half of their time in the schools where they work as teachers. All student teachers have a *school supervisor*, an experienced subject-teacher at the school who is trained for this task. Each student teacher also has a *university supervisor*, a lecturer who visits them at their school at least twice a year and observes a number of their lessons during the course. The students spend the other half of their time on activities that are related to course work at the university, where they come on one specific day in the week. On this day they discuss their experiences in a fixed group of fellow student teachers led by two university supervisors. On this day, there is also time for individual discussions with supervisors and for subject-related courses, which are supervised and assessed by specialized lecturers. Examples are subject methodology, psychology of adolescents, and classroom-management, and the academic spe-

cialization (a research project in a student teacher's own teaching context). The remaining one-and-a-half days is reserved for the preparation of courses and meetings and for carrying out various assignments.

Our student teachers are assessed on the basis of a number of assignments for the different modules (e.g., classroom management, teaching a specific subject, adolescent psychology and the academic specialization). A portfolio is used to keep track of the student teachers' competence development.

In order to assess our student teachers, formal discussions for formative assessment are scheduled at least two times during the year and are focused on the progress of the student teacher's competence and personal plan for further development. The summative assessment at the end of the course is discussed in an assessment meeting in which the final appraisal of the student teacher's progress is discussed. A number of people are involved in assessment during the course.

- The *supervisor who is responsible for the different modules*: This supervisor judges whether the student has fulfilled the demands of a specific module. The supervisors responsible for this are the supervisor for the modules in teaching a specific subject (in this specific case: the subject specialist for biology teaching), the supervisor for adolescent psychology, and the supervisor for the academic specialization.
- The *school supervisor and/or job coach*: co-assessor at the final assessment, advises interim and final assessments in relation to the practical assessment. Whenever a student gives classes at school, it is the supervisor of the last school who is involved in the end assessment.
- The *university supervisor*: responsible for the Go/no go-assessment, the middle evaluation and, together with the school supervisor, the end assessment. The institute supervisor can seek further advice from others in making these assessments, which could take the form of a further assessment.
- The *exams' commission*: in cases of disagreement between the student and supervisors, or between supervisors, concerning the assessment, the exams' commission has the last word.

Dealing with central issues in the development of criteria and standards

Below, we describe how we dealt with the central issues as derived from Uhlenbeck et al. (2002) in developing our criteria and standards.

Level of detail in the formulation of criteria (neither too broad; neither too narrow)

In our teacher education program, the required teaching competence is worded terms of in teacher roles: a professional being able to direct his/her own development, a subject teacher, a classroom manager, an adolescent psychologist, a

member of the school organization, and an academic specialist in a certain topic. These six roles are derived from national requirements for teacher competence. The national criteria were formulated in such a way, that room was provided for using them in various contexts of teaching and teacher education. In our study, we aimed at contextualizing the criteria to the culture in our institute in order to make them useful for application in our teacher education context. Furthermore, we aimed at formulating the criteria in such a way, that these would provide our student teachers as well as the supervisors/assessors with enough support for setting learning goals and making judgments, however, without making them a 'prisoner' of the criteria. For each role, we developed a rubric with four to seven criteria (see example in Appendix I).

Integrating research-based and practice-based perspectives in criteria

The national competence requirements for assessment of teachers are formulated based on what is currently known about good teaching as well as on the opinions of teachers themselves. The descriptions of the criteria focus both on what teachers should be able to do, as well as on the knowledge that teachers should possess. The competence requirements are actualized regularly. At least every six years, the national association for the profession of teaching in The Netherlands provides the secretary of education with a proposal for actualizing the competence requirements, both from evidence-based perspectives on what is known about good teaching, as well as from input by expert practitioners in the field. In our study, we aimed at integrating theoretical and practical notions on effective teaching in our criteria by regularly evaluating our criteria. These evaluations were performed both involving a research-based perspective on good teaching, provided by researchers knowledgeable on the literature on good teaching as well as by involving the supervisors/assessors in our institute and in our schools, and asking them for their experiences with using our criteria. In this way, both theoretical notions and practical notions were taken into account in our criteria for assessment of teachers.

Degree of specificity of standards

In the national competence requirements for assessment of teachers, for each competence, a number of standards were formulated in terms of what teachers know and are able to do. These standards were formulated on only one level, and they were not subject-specific, e.g., 'the teacher checks whether pupils are knowledgeable of the learning content and how the pupils carry out their assignments' and 'the teacher designs learning activities that pupils are able to carry out, from which they may choose, and which support them to self-direct their learning processes'. In our study, we aimed at developing descriptions of standards in four levels, ranging from insufficient to excellent. We aimed at formulating the standards in such a way, that these were applicable to student teachers in all subjects (see Example in Appendix I).

Key aspects of teacher knowledge and functioning; balancing input of different stakeholders

In order to decide on the key aspects of teacher knowledge and functioning as formulated in standards, we needed to balance the input of different stakeholders involved in the regular evaluation processes of our criteria and standards (e.g., researchers knowledgeable on the literature on good teaching as well as the supervisors/assessors in our institute and in our schools). We aimed at doing this by regularly reformulating our criteria and standards in development groups consisting of both researchers and supervisors/assessors in our institute and by asking our students as well as the supervisors/assessors in the school for feedback on our criteria and standards.

Evaluation of the developed criteria and standards

The evaluation of the developed criteria and standards was done in a cyclic process. Two cycles of design, evaluation and re-design were carried out. During each cycle, a new set of criteria and standards was developed, evaluated, and re-designed in the development groups mentioned in the former paragraphs. The developed criteria and standards were evaluated against the extent to which we succeeded satisfactorily in dealing with the central questions that we derived from Uhlenbeck et al. (2002). The evaluations were carried out by the members of the development group (consisting of both researchers and supervisors/assessors in our institute). Participants in the evaluations were students as well as supervisors/assessors in our institute and in the schools. Below, on the basis of our field notes and minutes from the evaluation processes, we describe the outcomes of the evaluation cycles. The outcomes are illustrated by examples from assessment of two student teachers in biology (Mary and Tom). To illustrate the outcomes of the evaluation of our rubrics, we use a lesson taught by Mary and Tom on combustion. More information on the lessons taught by Mary and Tom, is given in Appendix II.

Evaluation level of detail in the formulation of criteria (neither too broad; neither too narrow)

As an example, Appendix I provides the criteria for the role of 'Subject teacher' that were developed in the first development phase. Five main criteria were formulated for this role (i.e., learning conditions, subject matter, learning process, form of learning, and accountability for and development of subject-teaching choices). Each criterion consisted of a number of sub criteria; twelve in total. The educational vision behind the 'first' rubric (Appendix I) was quite specific, and the criteria in the rubric were specified according to this vision (i.e., the rubric was based on the vision of question-driven education, departing from the idea that good teaching leads to questions in pupils minds, which are subsequently answered, either by the pupil –pupil-centered- or by the teacher –teacher-centered-). The specific educational vision easily would make assessors judge student teachers who teach in line with the specific vision as better than colleague student

teachers who may have other educational visions (i.e., Tom's lesson would be judged as better than Mary's lesson, since Tom introduces a new topic by using a case that leads to questions in pupils' minds, whereas Mary provides answers to unposed questions by the pupils). The disadvantage of the 'first' rubric is that it departs from a particular educational vision, with specific criteria in line with this vision. The rubric thus may exclude forms of good teaching that depart from different educational visions. Therefore, the developed criteria were evaluated in the first cycle as being too narrowly defined for allowing a range of acceptable ways to teach, and the number of criteria as too much for being able to judge our student teachers in a fair and responsible way. As a result of the first evaluation, the criteria were re-designed and formulated more broadly, and the number of criteria was reduced. An example of how this was done for the role of Subject teacher can be found in Appendix III. In the 'new' rubric (Appendix III), the criteria for good education have been defined in a way that goes beyond a certain educational vision. The most important criterion for good education was defined as an alignment between teaching goals, pupils' learning processes, and pupil assessment. Such a criterion provides room for teaching according to different educational visions. However, a disadvantage of the 'new' rubric is, that it doesn't provide enough possibilities to assess an important quality of a student teachers' teaching, i.e., the cognitive level of lessons. Appendix II shows that that Mary has formulated a clear and feasible but low level cognitive goal for her lesson, and her explanations and the assignments are in line with this low level cognitive goal. By contrast, although Tom formulates his cognitive goals on a low level as well, his assignments are not in line with this goal, but on a much higher cognitive level (i.e., applying knowledge to new situations). The evaluation in the second cycle thus revealed that the criteria were now formulated too broadly, not providing enough information to students on what is expected of them and not providing supervisors/assessors with sufficient support for providing feedback and making judgments.

Evaluation integration research-based and practice-based perspectives in criteria

The developed criteria were evaluated in the first cycle as being too much focused on research-based perspectives, without translating them in a satisfactory way to teaching practice. Appendix I, and the description in the former paragraph shows that a number of theoretical notions (i.e., based on the vision of question-driven education) was used in the criteria for 'Subject teacher' which students and supervisors/assessors need to internalize before being able to apply them in practice. The 'old' rubric was based on a particular educational vision, which was further specified in the research of Janssen and De Hullu (2008). In their research, they developed a rationale for developing education from different subject-specific perspectives. This can be illustrated by Mary and Tom's lessons on combustion. The topic of combustion can be ana-

lyzed from different perspectives, and each perspective leads to specific questions, e.g., the mechanistic perspective (how does it work?), the functional perspective (what is it meant for?), the evolutionary perspective (how did it evolve?), or the comparative perspective (what are the differences and similarities?), etc. In the 'old' rubric, these perspectives are included in the criteria for assessment of the student teacher's lessons. However, judgment of lessons on this criterion requires assessors to be knowledgeable of these perspectives and how they can become manifest in teaching. Mastering the notions of these perspectives is not easy. As a result of the evaluations, the number of theoretical notions in the criteria was reduced in the re-formulated criteria. As outlined in the former paragraph, the most important criterion for good education was defined as an alignment between teaching goals, pupils' learning processes, and pupil assessment (see also: Appendix III). However, the evaluation of the re-formulated criteria in the second cycle, revealed that the criteria were now formulated too broadly (see also: the description in the former paragraph).

Evaluation degree of specificity of standards

The degree of specificity in standards was evaluated in the first cycle as unsatisfactory, since the standards did not provide students and supervisors/assessors with enough information on how to move on from one level to another in the range from insufficient to excellent, or for providing meaningful feedback on how to make such progress. Appendix I shows that although some of the standards for 'Subject specialist' were formulated in terms of detailed explanations, some of them were formulated only in terms of 'not satisfactory' to 'almost always satisfactory', while other standard descriptions were formulated in a very abstract sense. The evaluation results made the development group decide to reformulate the standards in a more specific way. Appendix III shows that the standards were formulated in such a way, that is easier to see how a student teacher may move from one level to another. Tom, for example, may move from level two to level three on the criterion teaching methods, by aligning teaching methods with the goals in his lesson. In the second evaluation cycle, students as well as supervisors/assessors were in the opinion that the new standards provided more support for moving from one level of functioning to another and for providing feedback on such progress. However they contended that the standards did not provide enough support for assessment of the student teachers since translating the standards to a particular subject was complex. They suggested that subject-specific exemplars may support their judgment processes.

Evaluation key aspects of teacher knowledge and functioning; balancing input of different stakeholders

From the results of the first evaluation cycle of the standards, it appeared that for our students, as well as the supervisors in the school, not all standards were considered useful and

meaningful for getting a grasp on what is expected of teachers and how to improve functioning. These results are in line with the evaluation of the degree of specificity of our standards (see description above): some of these were formulated in a too abstract sense, while other standards were not specific enough for practical use (see also Appendix I). After the first evaluation cycle, the development group re-formulated the standards by putting more emphasize on standard descriptions in terms of concrete behavior. Appendix II shows that the standards were formulated for each level in terms of actual behavior that a student teacher shows in classroom practice. In addition, for each role, standards were formulated for the knowledge that a student teacher needed to show in reflections on their functioning as a teacher. The 'new' rubric (Appendix III) indicates that a student teacher should not only be able to design lessons and to carry out these lessons accordingly, but should also be able to provide rationales for self-designed lessons, on what happened during a lesson, and on alternative approaches that may be possible. In the second evaluation cycle, students as well as the supervisors/assessors in the institute and in the schools commented that the standard descriptions in terms of actual behavior did provide more insight into what is expected from student teachers in the actual classroom setting, and for providing feedback on progress and functioning in classrooms. However, the standards did not provide enough support feedback on the knowledge that student teachers need to acquire and how they can make their knowledge explicit in reflections on their functioning. The standards in relation to knowledge aspects of functioning were evaluated as too vague for assessing, monitoring and providing feedback on student teachers' knowledge acquisition or the application of knowledge in reflections on functioning in classroom practice. What was lacking, according to the evaluators, was an overview of theoretical notions that may be used for providing feedback on designing lessons, and on reflecting on how a lesson was carried out. The assessment of Tom's lesson design may be supported by notions from task-centered-instruction (Merill, 2007), and motivation theory, e.g., the expectancy-value theory (Wigfield & Eccles, 2009). When assessors use such theoretical notions in their feedback, this may assist Tom in re-designing his lessons and reflecting on his lessons.

Discussion

In our stories that illustrate the development process of the rubrics, we haven't discussed the sources of evidence that were used to assess our student teachers. In our case, a lesson observation was used. A video of a lesson may be a useful alternative, in particular this enables discussion of specific video fragments during an assessment meeting, and this may result in meaningful feedback. Feedback by pupils and/or colleagues may be used as well. We also haven't discussed the procedures for using the rubrics and scoring the knowledge and performance of our student teachers. We have focused our stories on the criteria and standards (i.e., reference levels)

that are needed in order to enable pronunciation of judgment, to monitor progress, and to provide feedback to our student teachers. In further research, procedures for the use of rubrics may be a worthwhile topic of investigation.

From the stories on the results of the evaluation cycles, it becomes clear that the developers of the criteria and standards as well as their users, remained confronted with the issues in relation to the level of detail in standards and criteria for functioning, as well as with issues on integrating theoretical and practical perspectives, and balancing input of various stakeholders. We might thus conclude that the issues in relation to developing standards and criteria for assessment of teachers will remain dilemmas that developers and users of rubrics for the assessment of teachers are faced up to with. However, our stories show that dilemmas in relation to the level of detail and specificity of criteria and standards may be resolved by providing more (subject-specific) benchmarks as exemplars of performance for each of the criteria and levels of functioning as was also asked for by some of our supervisors/assessors, and which is also contended in the review on rubrics by Johnsson and Svingby (2007). However, Johnsson and Svingby (2007) also stress that 'benchmarks should be chosen with care since the scoring depends heavily on the benchmarks chosen to define the rubric' (p. 133). Integrating theoretical and practical notions may be improved by putting effort in using the results from the evaluation cycles for further the defining the constructs that we are assessing and by developing a framework of theoretical notions and rationales that underlie the rubrics for scoring. We are currently developing a theoretical framework in which general criteria for good teaching are formulated which are specific enough for doing justice to differences between cognitive levels of pupils, which are easy to interpret by student teachers and their supervisors/assessors, and which provide enough room for subject-specific and method-specific teaching approaches. This theoretical framework is meant to serve as a point of departure for re-formulating the criteria for our rubrics. Subsequently, the criteria will be illustrated by subject-specific exemplars and benchmarks. The theoretical framework may be discussed regularly with developers and users of the criteria and standards and will be updated regularly. A continuous discussion on the theoretical framework as well as the formu-

lation of criteria and standards for functioning may support the selection of key aspects from teacher knowledge and functioning that need to be represented in the criteria and standards. Focusing on the opinions of different stakeholders and to the cultural context in which the criteria and standards are used, may be complemented with research-based attempts to further define the constructs underlying good teaching that we aim to assess. Such attempts may also be supportive for balancing the input of various stakeholders. And, above all, analyzing assessment information and updating the underlying theoretical frameworks and vice versa, may inform the theoretical notions on good teaching.

References

- Andrade, H.G., Teaching with rubrics: The good, the bad, and the ugly, *College Teaching*, **53**(1), 27–30, 2005.
- Darling-Hammond, L., & Snyder, J., Authentic assessment of teaching in context, *Teaching and Teacher Education*, **16**, 523–545, 2000.
- Delandshere, G., & Petrosky, A., Assessment of complex performances: Limitations of key measurement assumptions, *Educational Researcher*, **27**(2), 14–24 1998.
- Janssen, F. J. J. M., & Hullu, A.E. de, A toolkit for productive thinking, *Journal of Biological Education*, **43**, 21–28, 2008.
- Johnsson, A., & Svingby, G., The use of scoring rubrics: Reliability, validity and educational consequences, *Educational Research Review*, **2**, 130–44, 2007.
- Reddy, Y. M., & Andrade, H. G. (2009). A review of rubric use in higher education. *Assessment and Evaluation in Higher Education*, **35**(4), 435–448.
- Uhlenbeck, A. M., Verloop, N. & Beijaard, D., Requirements for an assessment procedure for beginning teachers: Implications from recent theories on teaching and assessment, *Teacher College Records*, **104**(2), 242–272, 2002a.
- Merrill, M. D., First Principles of Instruction. In: C. M. Reigeluth & A. Carr (eds.), *Instructional Design Theories and Models III* (Vol. III). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers, 2009.
- Wigfield, A., Tonks, S. & Klauda, S. L., Expectancy-value theory. In: K. R. Wentzel & A. Wigfield (eds.), *Handbook on motivation at school* (pp. 55–77). New York: Routledge, 2009.

Appendix I: Subject teacher (version: 20 April 2006)

	Level 1	Level 2	Level 3	Level 4
Subject-teaching choices with ref. to the following dimensions plus corresponding aspects.	No or unclear	Yes; clear	+ Selection of main goals: approach agreed in consultation with pupils.	+ Teacher sets out framework; pupils flesh it out
<i>Learning conditions</i> Clear formulation of learning goals and procedures (what, how, with whom, why, by when, where, product) Relevance of learning activities to the learning goals and the starting situation. Response to input and pupils' learning problems.	Often no	Activities are often linked to the goals and to subject matter covered earlier	Activities are often linked to the goals and the actual starting situation.	Ditto; achieved because taking part in activities motivates participation in follow-on activities
<i>Subject matter</i> Use of perspectives Use of textbook and other sources of information	Not intentionally	Sometimes satisfactory For selection of subject material and/or for presentation of topic	Often satisfactory + for organisation of activities.	Almost always satisfactory + for use by pupils
<i>Learning process</i> Type of learning goals with ref. to nature of subject matter. Use of type of learning process: 1. motivating 2. presenting the topic (question) 3. answering the question 4. application 5. evaluation and feedback With variations in how these are steered.	Mainly follows chapter in textbook, possibly also with own material ("from cover to cover") Specific knowledge from textbook 1. n/a 2. n/a 3. offers knowledge (answer) 4. application exercises (to reproduce question) 5. no	+ Own material for motivation and presenting topic. Thereafter mainly follows chapter in textbook (possibly with some selection and re-organisation) + Examples and links chosen by student 1+2+starts by presenting situation which motivates and raises central question 3. offers knowledge (answer) and demonstrates method 4. followed by productive application exercises 5. evaluation of learning goals and feedback at class level	+ Own material for motivation and presenting topic. In addition to chapter other sources needed to deal with the topic. +General understanding and/or skills. 1+2+starts by presenting situation which motivates and raises central question. 3+ allows pupils to contribute to finding the answer 4. followed by productive application exercises 5. evaluation of learning goals and feedback at individual level and stimulation of self-evaluation	+ Justification of understanding and skills

Appendix I: Subject teacher (version: 20 April 2006)

	Level 1	Level 2	Level 3	Level 4
Subject-teaching choices with ref. to the following dimensions plus corresponding aspects.				
Use of group formation and learning materials	- teaching and individual work - choice not deliberate and execution frequently not effective - only uses the book	+ various (i.e. diverse) simple activating group forms and selected learning materials - choice usually functional and sometimes not yet implemented effectively.	+ various (i.e. diverse) complex activating group forms and selected learning materials - choice functional and usually effective.	+ projects where pupils help to determine the form and content. - choice functional and effective.
Form of learning	No	Uses some parts of the model with help often deliberately and purposefully, in several lessons.	Uses the whole model with some help; deliberately, purposefully and regularly.	Uses the whole model independently; deliberately, purposefully and continuously.
Use of subject teaching model when devising and analysing lessons.	No	Often, with some help.	Often and independently and can justify basic rules.	+ Examines basic rules critically.
Formulating basic rules and applying them in the lesson.	Sometimes	Sometimes with help.	Often and independently.	+ Critical examination of several theories.
Use of (subject-based) teaching theory	No	Can give a general outline and sometimes attempts to implement/improve them in practice.	Can give a specific description and often attempts to implement/improve them in practice.	+ Can explain and justify them (at least in part).
Identification of personal strengths and weaknesses and (subject-based) teaching vision.	No (mostly not deliberate)			

+ indicates that a teacher not only possesses the qualities described for that level, but also those given for the previous level. It therefore indicates an extension of his capabilities.

Appendix II: Mary and Tom's lessons on combustion

Below, a summary is provided from a lesson, given by two student teachers in biology (Mary and Tom). The topic of the lessons is combustion, with a specific emphasis on warm- and cold-bloodedness. The lesson was prepared for pupils in the lower level of high school. Both lessons take 50 minutes.

Mary and Tom formulated the following goals for their 50 minutes lessons:

- Pupils know the formula of combustion.
- Pupils know the differences between warm- and cold-blooded animals.

Mary's lesson:

- Discusses the homework that results from the former lesson.
- Explains what combustion is and demonstrates with a candle what is needed for combustion.
- Explains the differences between warm-blooded and cold-blooded animals.
- Pupils make the (reproductive) assignments in the textbook.

Tom's lesson:

- Discusses the homework that results from the former lesson.
- Introduces a case on a lizard and a fat man. The lizard challenges the fat man to join in a foot race with him. The day of the foot race, it happens to be rather cold. Tom asks the pupils: who do you think is going to win the foot race, the fat man or the lizard?
- The pupils note their prediction and their rationale on a work sheet.
- Tom, by using a candle, explains what combustion is, and how combustion works with human beings.
- In an instructional dialogue, Tom aims to let the pupils discover the differences between warm- and cold-blooded animals with respect to combustion.
- The pupils are being confronted with the case on the lizard and the fat man again. On the same work sheet, they fill out their prediction and their rationale. In addition, the pupils are being asked to rate on a ten point scale whether the case was motivating, clear, and challenging to them.
- Tom discusses the case plenary.
- The pupils make a number of (reproductive) assignments in the textbook.

Appendix III: Rubric Subject teacher (new)

	Level 1	Level 2	Level 3	Level 4
Learning goals	Sets no goals, or goals that are infeasible for the pupils.	Sets goals that are feasible for the pupils.	Sets goals that are feasible for the pupils and that can be assessed, and takes into account what learning activities are expected from pupils in order to reach the goals.	Sets goals that are feasible for the pupils and that can be assessed, and takes into account what learning activities are expected from pupils in order to reach the goals. Takes differences between pupils into account.
Teaching the learning content	Doesn't provide meaning to subject-specific concepts and doesn't provide links between these concepts and the pupils' environment or their prior knowledge.	Connects subject-specific concepts to pupils' prior knowledge and their environment.	Connects subject-specific concepts to pupils' prior knowledge and their environment. Is aware of typical student learning problems with regard to specific learning contents and takes these pitfalls into account in the lesson designs.	Connects subject-specific concepts to pupils' prior knowledge and their environment. Is aware of typical student learning problems with regard to specific learning contents and takes these pitfalls into account in the lesson designs. Checks during the lesson whether pupils develop misconceptions and, in such cases, uses various strategies in order to respond to these issues flexibly, fast, and adequately.
Learning process	Organizes pupils' learning processes insufficiently. Doesn't evaluate whether the learning goals have been realized by the learning processes that have been carried out.	Organizes learning processes which, in principle, may lead to realizing the learning goals. Evaluates whether learning processes have promoted that learning goals were being reached.	Organizes learning processes which, in principle, may lead to realizing the learning goals. Is aware of situations in which learning goals may not be reached, and adjusts the teaching in case necessary. Evaluates whether learning processes have promoted that learning goals were being reached, chooses the most appropriate evaluation methods.	Organizes learning processes which, in principle, may lead to realizing the learning goals. Uses strategies to discover as soon as possible when learning goals may fail to be reached and adjusts the teaching in case necessary. Evaluates whether learning processes have promoted that learning goals were being reached, chooses the most appropriate evaluation methods. Is able to design evaluation methods.
Teaching methods	Uses a limited amount of teaching methods, even when other teaching methods are more applicable.	Uses teaching methods which, in principle, may lead to a realization of the learning goals. The application of the teaching methods is not always satisfactory.	Uses various teaching methods that are suitable for the learning goals and applies these teaching methods in an appropriate way.	Has a large repertoire of teaching methods, and applies these teaching methods in a suitable way for reaching the learning goals. Is able to flexibly change to another teaching method in case necessary, and is able to argument for these changes.
Argumentation	Is unable to argument for the learning goals and teaching methods that have been chosen, as well as for the results of the teaching process.	Is able to argument for the learning goals and teaching methods that have been chosen, as well as for the results of the teaching process.	Is able to account for choices in the learning goals, the teaching methods, and the results of the teaching process by using theoretical rationales. Is aware of alternative approaches.	Has an overview of relevant theoretical notions and is able to use these notions when accounting for choices in the learning goals, the teaching methods, and the results of the teaching process. Has an overview of alternative approaches and is able to explain why these are suitable to the teaching situation or not.