Model-Based Course Assessment
– Principles and Practice

Per Runeson, *ETP*, and Björn Regnell

Abstract—**This paper presents the concepts of model-based course assessments. This involves a mathematical model to define the grading scheme of a course, based on a set of items to be scored. We report lessons learned from using this approach. An im**

TABLE II
GRADING MODELS FOR THE RE AND SE COURSES

Scale values RE	Scale values SE	Final grade	Description
< 0.5	< 0.5	Fail	Extra assignment needed
>= 0.5	>= 0.5	3	Pass
>= 0.6	>= 0.7	4	Pass with distinction
>= 0.8	>= 0.85	5	Pass with special distinction

Iprinciples in a course, we have recently introduced *model-based course assessments*. Thereby we refer to that the principles for the course grading are given in the format of a model, expressed in mathematical formulae. Each item of the course, which impacts on the final grade, is assessed on a scale between 0 and 1. The item scores are then weighted with a percentage figure and summed up to a total value, which finally is transformed into a grade.

Using the model-based approach, we communicate in quantifiable terms the main learning goals of the course, e.g. by giving the course project a weight of 35% and the written exam 65%. We have found that the model makes the grading criteria more transparent, and that it is a vehicle for students in their prioritization of learning efforts in relation to the learning objectives. If the model is combined with clear learning objectives, it adds to the clarity of what is expected from students. It also enables a modularization of the assessment into different items, which can be assessed independently by different teachers and combined by the course coordinator.

Experienced drawbacks are that the students tend to consider the last judged course part as the most influential part. Further, the model may create some threshold problems when transforming the total value into the final grade.

The paper introduces the principles of model-based assessment, and provides examples of its use in two courses, as well as lessons learned from the use. We want to initiate a discussion on the issues raised with the use of the model.

II. PRINCIPLES

The main principle of the model-based assessment approach is that the criteria for grading in a course are defined in a mathematical model. The course is assessed in the following steps:

1. Each item of the course that contributes to the grading is assessed with a score on a scale from 0 to 1, either in defined steps or on a continuous scale.
2. The items are multiplied by weights for each item and summed up to a score for the whole course.
3. This score is finally mapped onto a grade.

The items are graded on a ratio scale to enable correct transformation from a measurement-theoretic point of view [1]. The weights can be defined for each item separately, or in

TABLE III
ASSESSMENT MODEL FOR THE SE COURSE

Group	Item	Item weight	Group weight
Process	Status review 1	25%	30%
	Status review 2	35%	
	Status review 3	40%	
Product	Level	20%	30%
	Compliance	30%	
	Robustness	50%	
Project report	Level	50%	15%
	Quality	50%	
Individual report	Level	50%	15%
	Quality	50%	

$$\lfloor c <= Score \to 5$$

III. EXAMPLES

We have applied the model-based assessments in two courses, the course Requirements Engineering (RE)[1] and the course for Large-Scale Software Engineering (SE)[2].

A. Assessment model in the RE course

The model used in the RE course is defined as follows. The grading system is based on the scores on a scale from 0.0 to 1.0 of 11 course items that are weighted according to the scheme in Table I. The scores are given on a continuous scale, but mostly steps of 0.1 are used. Each item must have a minimal score of 0.1, otherwise rework is required.

The final course score is mapped onto grades according to the scale in Table II, in the first column.

B. Assessment in the SE course

---

[1] Course code ETS671, "Kravhantering"
[2] Course codes ETS032 and ETS311, "Programvaruutveckling för stora system (PUSS)" see further at http://serg.telecom.lth.se/education/

TABLE I
ASSESSMENT MODEL FOR THE RE COURSE

Item	Weight	Explanation
SRS	25%	Quality of the Software Requirements Specification
PFR	20%	Quality of Project Experience Report
PFP	5%	Quality of oral Project Final Presentation
EXA	3%	Quality of written hand-in of exercise A
EXB	15%	Quality of written hand-in of exercise B
EXC	3%	Quality of written hand-in of exercise C
EXD	3%	Quality of written hand-in of exercise D
EXE	3%	Quality of written hand-in of exercise E
EXF	3%	Quality of written hand-in of exercise F
RAP	5%	Quality of reading assignment oral presentation
PFP	5%	Quality of oral Project Final Presentation

The model used in the SE course has two levels of weights, item weights and group weights see Table III. Four different areas are graded and within each area there are two or three items that are scored between 0.0 and 1.0 in steps of 0.25. Specific scoring criteria are defined for each item, and mean:

0.00 – fail, rework of item needed
0.25 – fail, rework of item not needed
0.50 – pass
0.75 – pass with distinction
1.00 – pass with special distinction

The course score is calculated by first multiplying the item score with the item weight, then summing the items for each group, multiplying the sum with the group weight and adding the group scores to a course score, see equation (2). The final course score is mapped according to the scale in Table II, in the second column.

IV. LESSONS LEARNED

The model-based assessment approach has been used at four occasions, twice for each course, and we have collected experiences regarding a number of issues.

A. Transparency

We consider the models-based approach be very transparent to the students. Discussing the motivation for each item score can explicitly motivate a given grade, and then the grading is pure math.

However, many students are skeptical. In an operative assessment of the SE course of fall 2003, the criteria for assessment were given an average score of 3.4 on a 5-level Likert scale (Question: I know the criteria for grading; 1=agree; 5=disagree). On the other hand, a CEQ [2] questionnaire in the SE course in spring 2004 gives some positive indications as the Appropriate-Assessment-scale was given +37.

In a CEQ of the RE course that was handed out in fall 2003, many positive comments were made about the model-based assessment, but there were also several complaints regarding a perceived unbalance in some of the item weights.

B. Support in goal communication

The model-based assessment model is assumed to support communication of course goals. In the SE course, students tend to focus too much on the product group of items, while the course goals have their main emphasis on the process and report parts. This is clearly reflected in the assessment model, by the product weight of 30% and the sum of process and report weights of 70%. However, the CEQ questionnaire does not support that the message was perceived by the students, as the Clear-Goals-scale was measured –5.

In the RE course there is also a great emphasis on the project experience report (20%) and the course objectives where clearly stated in the course program. The perceived advantage in goal communication is not supported by the CEQ questionnaire, as the Clear-Goals-scale was measured +2, but on the other hand the Appropriate-Assessment-scale was given +52 and the Good-Teaching-scale was given +43, which is rather high[3].

C. Timing

The timing of the assessment of the separate items introduces some problems. Although the importance of each item for the final grade is given only by the weights, the students feel like the last assessed item impacts the most on the final grade. This is particularly visible in the SE course, where the three first groups are based on the assessment of the project in the course, while the final group is an individual assessment. Some students experience the latter group of items being crucial to the final grade, although it represents only 15% of the final grade.

This was not as apparent in the RE course as the grades were continuously given to students. However, some students that failed the course seemed to find the grading system unbalanced, as they realized that if other weight on items were given they would have passed the course.

D. Thresholds in grading model

The thresholds in the grading model are somewhat arbitrary. They are derived by using the model with different sets of scores for the individual items. In the SE course we transformed into this assessment system from an earlier system, where a judgment was done of the whole project according to some qualitative criteria. This judgment corresponds to the three first groups of the new assessment model. Hence we wanted a smooth transformation from the earlier assessment system; the grade 4 in the former system should correspond to the grade 4 in the new system.

Except for defining the thresholds, the thresholds as such causes some problems. As they are arbitrary cut-off values, there is always a risk that the grading of one single item causes the final grade being 3 or 4. In the SE course of fall 2003, the final score for 18 out of 76 students were within the range of +/- 0.01 points around the grade thresholds (4 students [0.69;0.71] and 14 students [0.84;0.86], cf. Table II).

---

[3] The entire CEQ analysis is available at:
http://www.telecom.lth.se/Kurser/kram/
CEQ-ETS671_2003_HT_LP1_arbetsrapport.pdf

E. Distributed assessment

If the assessment criteria of the model are reasonably unambiguous, it is fairly easy to have many different assessors involved in the course and use the mathematical model to combine the assessments. In the RE course, 3 different assessors were involved (2 PhD students, and the course coordinator). The model with its associated criteria helped communicating a grading standard and we believe that the grades were more consistent among assessors due to the model-based assessment.

V. CONCLUSIONS

The model-based assessment presented in this paper has the following potential benefits in relation to the CEQ scales [2].

- *Good teaching*. If the model-based assessment is used as a tool for continuous feedback on student work it may help in making it clear to students how they perform I relation to the course goals.
- *Clear goals*. As the items in the model are connected to the general goals of the course it may help to make the goals clearer if the assessment criteria them selves are clear.
- *Appropriate workload*. The percentage figure of the models help students to understand how the teacher wants students to prioritize their work. If students follow this recommendation it may be more likely that they can maintain an appropriate workload as they give less time to less important parts of the course.
- *Appropriate assessment*. The model-based assessment itself does not prevent criteria that are directed towards memory knowledge rather than deeper knowledge. But if the criteria themselves are directed towards deep learning, the model-based assessments give a powerful tool for emphasizing and measuring achievements in this direction.
- *Emphasis on independence*. One principal feature of the model-based assessment is that students can make their own choices about where to put their learning efforts, as it is clear what are the consequences in the assessment of different prioritizations.

There are, however, a number of challenges and open issues.

- *Objective assessment criteria*. The figures of the presented model give an impression that the results are objective, but this is actually depending on the criteria behind the assessment. It is important to put a lot of effort in the construction of objective assessment criteria in order to benefit from the advantages of model-based assessment.
- *Standard mapping to grades*. Should we have a consensus at LTH on how "tough" grades 4 and 5 should be? Is it desirable to have an absolute mapping, when we may have to comply with international grading systems such as European Credit Transfer System (ECTS) [3] that are relative?
- *Student acceptance*. Students are more familiar with existing assessment systems, such as written exams with a certain scale of grading. New systems may be received with some skepticism and it is important to explain the model-based assessment system to students and to gain acceptance for its usage.

REFERENCES

[1] N. E. Fenton and S. L: Pfleeger, *Software Metrics, - A Rigorous and Practical Approach*, Thomson Computer Press, 1996.
[2] P. Ramsden, *Learning to teach in Higher Education*, Rutledge, 1992.
[3] European Credit Transfer System, homepage visited April 2004: http://europa.eu.int/comm/education/programmes/socrates/ects_en.html