Roles and forms of assessment of modelling in secondary physics education in school practice

Onne van Buuren[1,2], André Heck[1], and Ton Ellermeijer[3]

[1]*University of Amsterdam, the Netherlands,* [2]*Montessori Lyceum of The Hague, the Netherlands,*

[3]*Foundation CMA, the Netherlands.*

Abstract

A learning path on modelling and experimenting with ICT has been developed for lower secondary physics education. For monitoring the students' progress on this learning path, several forms of assessment have been used. In this paper, advantages and disadvantages of several forms of assessment of modelling are discussed. Modelling offers possibilities for self-correction by students, especially if modelling is combined with animation. It is recommended to assess computer modelling and ICT-supported experimenting not only hands-on, but also by means of pencil-and-paper tasks, whether the purpose is formative or summative.

1 Introduction

For both computer modelling and ICT-supported experimenting in physics, many competencies are required. To mention a few: modellers and experimentalists must be able to use the software tools, they must be able to analyse and interpret graphs, they must have a sound understanding of the formulas that are involved, they must have sufficient understanding of the physics concepts that are involved, and modellers must understand their modelling approach. As a consequence, the cognitive load of computer modelling and of ICT-supported experimenting can be high. The required competencies cannot be mastered in just a few lessons by a novice student but require a learning path distributed over a long period of time.

Recently, such a learning path on computer modelling, combined with ICT-supported experimentation, has been developed for the Dutch lower secondary curriculum. This learning path is completely integrated into the physics curriculum and has been tested in school practice (Van Buuren, 2014). One of the goals of this learning path is that students can build simple quantitative computer models themselves at the end of lower secondary physics education. Currently, this learning path is extended into the first year of upper secondary education.

The development of the competencies of students on such a learning path must be monitored carefully by the developers of the learning path and by the teacher, in order to adapt instructional materials and teaching to student difficulties or to take advantage of opportunities for learning. The scale of such adaptations ranges from a small scale—a discussion between an individual student and the teacher—to the large scale of the entire curriculum. Preferably, the development of the students' understanding is also monitored by themselves: they must be able to correct themselves. The process of monitoring and adapting or correcting requires formative assessment. Modelling competencies must be assessed for summative reasons too.

The question is how modelling competencies can be tested, both for summative and formative purposes, in an effective way in school practice.

The ultimate summative tests are the exams. In Holland, the examination programme consists of two parts: a nationwide written 'final exam' and a 'school exam'. The school exam is an internal exam, designed by a teacher or a team of teachers at school. It consists of both written tests and more open, practical assignments. These assignments include practical investigations by students. Since 1991, computer modelling is part of the Dutch secondary physics examination program at pre-university level, but modelling competencies have not been tested in the final exam until 2013. As a consequence, many Dutch physics teachers did not pay much attention to computer modelling. Exceptions were the teachers at about two hundred schools participating in the 'compex exams', experimental computer examinations in which students' modelling competencies were tested hands-on (Boeijen & Uylings, 2004). The same holds for publishers of educational materials (Lijnse, 2008).

In 2013, new curricula for the upper levels of Dutch secondary science education have started. Computer modelling is now part of the programmes for both physics and biology, not only at pre-university level, but also at havo-level (havo is a five years senior general secondary education program preparing for higher vocational education). According to Savelsbergh et al. (2008), modelling should mainly be tested in the school exams because 'modelling is an iterative process for which creativity, reflection and deliberation are needed' and therefore must be tested in an open setting; only certain competencies, such as the competency to explore a given model, might also be tested in the nationwide final exams. In accordance with Savelsberg's advice, modelling is now part of the school exams only. The only exception is the program for physics at pre-university physics level. At this level, modelling will also be assessed in the nationwide final exam, by means of pencil-and-paper tests. Big question is whether modelling should not also be tested in *all* nationwide final exams.
A first reason to test modelling competencies in the final exams too, is that teachers tend to consider topics that are not a part of the final exams as less important. A second reason follows from a comparison of modelling with practical investigations by students. Assessment of practical investigation competencies is known to be difficult. It depends on what is considered to be the learning goal and there are a multitude of competencies that have to be dealt with (Gott & Duggan, 2002; Etkina, Karelina, & Ruibal-Villasenor, 2008). Results of different ways of assessment depend strongly on learning style: some students may perform better with hands-on practical tasks, others with pencil-and-paper tasks (Gott & Duggan, 2002; Roberts & Gott, 2006). Furthermore, the response of a student to a task may be a measure for a variety of competencies and its validity is therefore easily contaminated (cf., Wiliam & Black, 1996; Millar, 2010). For example, in school practice, students' written accounts of an investigation are often used as a surrogate for direct observation of students' actions because direct observation requires too much time, but students' writings skills do not necessarily correlate with their practical investigations skills (Gott & Duggan, 2002).

Because of the multitude of competencies required for modelling, similar problems can be expected with the assessment of computer modelling. The validity of the assessment may be increased when modelling competencies are assessed not only by means of open investigations in the school exams, but also in a more closed form in written exams. Assessment in a more closed form makes it possible to focus on specific competencies, which are difficult to measure in an open setting, because of the contaminating effect of other competencies.

The importance of focussing on specific competencies holds even more for formative assessment. Many competencies are essential for computer modelling. Not mastering one of these essential competencies can be a cause for serious student blockages (cf., Van Buuren, 2014). In order to monitor and adjust the development of a single competency of a student, assessment must be focussed to this competency.

In this paper, we present and discuss some of the forms of assessment that we have used while we developed our modelling learning path.

2 Method and setting

This paper can be considered as spin-off of an educational design research project. The main purpose of this research project is to establish characteristics of an effective learning path on graphical modelling in lower secondary education and in the first year of upper secondary education. In educational design research, educational materials are designed, tested in classroom, and redesigned in several cycles (Van den Akker, Gravemeijer, McKenney, & Nieveen, 2006). As a modelling approach, we have used the graphical version of Forrester's system dynamics (Forrester, 1961). We have used Coach 6 as an educational tool, because in this computer learning environment, modelling can be combined with doing and analysing

measurements. In addition, in Coach 6, modelling can be combined with animation (Heck, Kedzierska, & Ellermeijer, 2009).
The learning path has been developed for secondary physics education in general, but is mostly tested on a school for secondary Montessori education. Within the limits posed by the Dutch government to secondary education, this school strives to work according to the principles of the Italian educator Maria Montessori (1870-1952). A special feature of this school is that students are used to go over their own exercises.

For research purposes, we have made classroom observations, audio-recordings and computer screen recordings of multiple student groups. The classroom observations often lead to dialogues between students and the researcher. These dialogues had the character of small scale in-depth interviews. In addition, written materials and assessments have been collected. Although these data have not been collected with the purpose of studying the effects of forms of assessment, they did provide us with many indications about these effects. Often, these data were used to further develop questions and tasks that served formative assessment and to further develop tests.

3 Graphical modelling

The graphical version of Forrester's system dynamics is often referred to as 'graphical modelling' (Forrester, 1961). In a graphical model, variables and relationships between variables are represented by means of a system of icons in a diagram. Figure 1 shows an example of a graphical model.
From a mathematical viewpoint, a graphical model is a system of one-dimensional difference equations and direct relations. Running the model boils down to numerical integration of this system.
The direct relations must be entered by the modeller. An advantage of the diagrams is that they provide a clear overview over the main structure of the model. A disadvantage is that formulas and values are not directly visible. As a consequence, it takes more time for a teacher to inspect the formulas and values in the model in order to provide feedback.
It leads to far to explain graphical modelling in more detail in this paper. For more information, we refer to another publication (Van Buuren, Heck, and Ellermeijer, 2015).

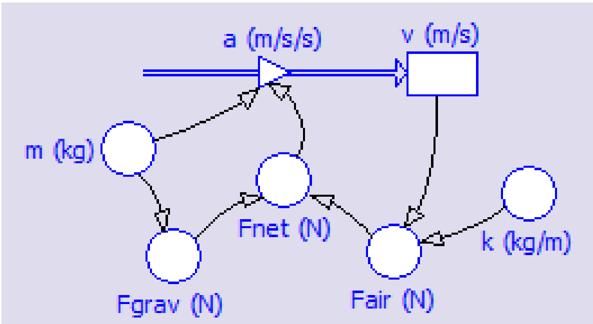

Fig. 1. Graphical model for the velocity v of an object falling through air. The motion of the body is governed by the difference equation $\Delta v = a \cdot \Delta t$, in which the acceleration a is defined as $a = F_{net}/m$, where the net force F_{net} equals the force of gravity F_{grav} minus the air resistance F_{air}. The air resistance is defined as $F_{air} = k \cdot v^2$, in which k is a constant. All quantities are depicted by means of icons; arrows indicate the presence of formulas. If necessary, the formulas can be made visible. in Coach 6, this can be done by double-clicking the icons.

4 Outline of the modelling learning path

One of the predominant principles of the learning path is that modelling is systematically combined with experimenting and (video-)measuring. The main purposes of the experiments and measurements are to familiarise students with the situations that must be modelled and to provide data that are used for evaluation of the models. On the learning path, the development of each modelling competency is tuned carefully to the development of other competencies and to the entire curriculum, and vice versa. We did not merely add modelling tasks to the curriculum, but, whenever necessary, adapted the whole curriculum, including the textbook.
The modelling learning path starts in the first year of physics education. In Holland, this is the second year of secondary education (age: 13-14 years). Currently, the learning path is distributed over the first two and a half years of physics education. For a more detailed description of the learning path and the principles that have been used to develop it, refer to Van Buuren (2014). Here, only a brief outline can be given.

On the learning path, after only four weeks of physics education, the concept of a model is introduced in a module on geometrical optics. After four months, students start to use simple graphical models in a module on kinematics. In this module, graphs are introduced too. At the start of the second year, students create a part of a model for the first time, by adding a formula to the model. During the second year, the main structures of graphical models are introduced. At the end of this year, students create simple models and complete a more complicate model for the first time. In the first month of the next year, students start to build and work with more complicated one-dimensional models in a module on dynamics. To do so, they must, amongst other things, understand the relation between the directions and the signs of physical quantities. Also, they learn to use conditional statements (if…then….else-statements).

5 Assessment of modelling

Several forms of assessment have been used to monitor the learning processes. In this paper we distinguish between five dimensions of assessment:

1. monitoring can be done by the teacher or by the student;
2. feedback can be provided almost immediately (fast) or delayed (as is the case with written assessments that have to be gone over by the teacher);
3. assessment can be done hands-on, with a modelling program, or with pencil and paper;
4. assessment can be done by means of open, practical tasks or by means of written tests;
5. the purpose of assessment can be formative or summative.

In the following subsections, we limit ourselves to the discussion of certain aspects, in particular possibilities for self-correction (dimensions 1 and 2) and the roles of pencil-and-paper questions and pencil-and-paper tests versus questions and tests with ICT (dimensions 3, 4, and 5).

5.1 Possibilities for self-correction

Key feature of formative assessment is feedback. Based on the students' feedback to his teaching, a teacher can adapt his teaching or his instructional materials. On the other hand, students can use constructive feedback to correct themselves. This feedback must be constructive and not judgemental, because judgemental feedback may have a negative effect on learning. Preferably, students assess themselves. This was already recognised by Montessori (1912), who developed educational materials that were self-correcting, in order for children to be less dependent on the feedback—and judgements—of adults. Recent research confirms the value of self-correction for learning (cf., Lillard, 2007; Black & Harrison, 2010).
Carefully designed modelling tasks offer possibilities for self-correction. Important aspects of modelling are interpretation and, subsequently, evaluation of model output. When students are able to evaluate the output of their models, they may themselves detect their errors and correct their models. In this way, the assessment of the students' modelling competencies is in the evaluation of the model output. A necessary condition is that students are able to interpret the model output. Therefore, they must (1) understand the way in which the output is represented and (2) have a fair idea what to expect.

Usually, model output is represented by means of tables and graphs. Of these two, graphs provide the better overview. Though, a graph is not yet a sufficient means for novice students to correct themselves, because students can have severe problems understanding and interpreting graphs (cf., Beichner, 1994). Even if students are able to read graphs properly, this is not yet sufficient. Experienced physicists recognise important features of graphs, such as parabolic, sinusoidal or exponential shapes, and automatically draw relevant conclusions from these features. As we observed in classroom, many students cannot yet draw such conclusions, even if they possess the essential knowledge and skills. We observed this phenomenon with first year upper secondary students, who were analysing video-measurements or were modelling a fall under influence of gravity and air resistance. If the shape of the resulting position-time graph is parabolic, it can be concluded that the net force on the moving object is constant, but many students were not able to draw this conclusion, even if they had the essential knowledge, as appeared from their answers to written tests that they had made earlier. Another example is the effect of an error in the sign of a quantity. This can be considered a minor error, but the consequences for the shape of the graph are immense. As we observed in classroom, this is often not recognised by novice modellers.

If we want to enable self-correction in modelling tasks, we need to do more. A possibility is to use additional representations that are more comprehensible for novices, such as animations. In Coach 6, models can drive animations. An example, taken from the first year of upper secondary physics, is shown in Figure 2. The model in this figure drives an animation. In this animation, also the vectors of the forces are drawn. The combination of model, animation and graphs provide students with more comprehensible feedback. In interviews, students stated that they considered these animations as very useful for improving their understanding of the variable forces that are involved in this type of movement. Another example of self-correction by means of a combination of animation and graphs is described by Van Buuren (2014). In this example, we observed how lower secondary students switched between the animation and the graphs in order to correct calculation errors and to increase their understanding of the graphs.

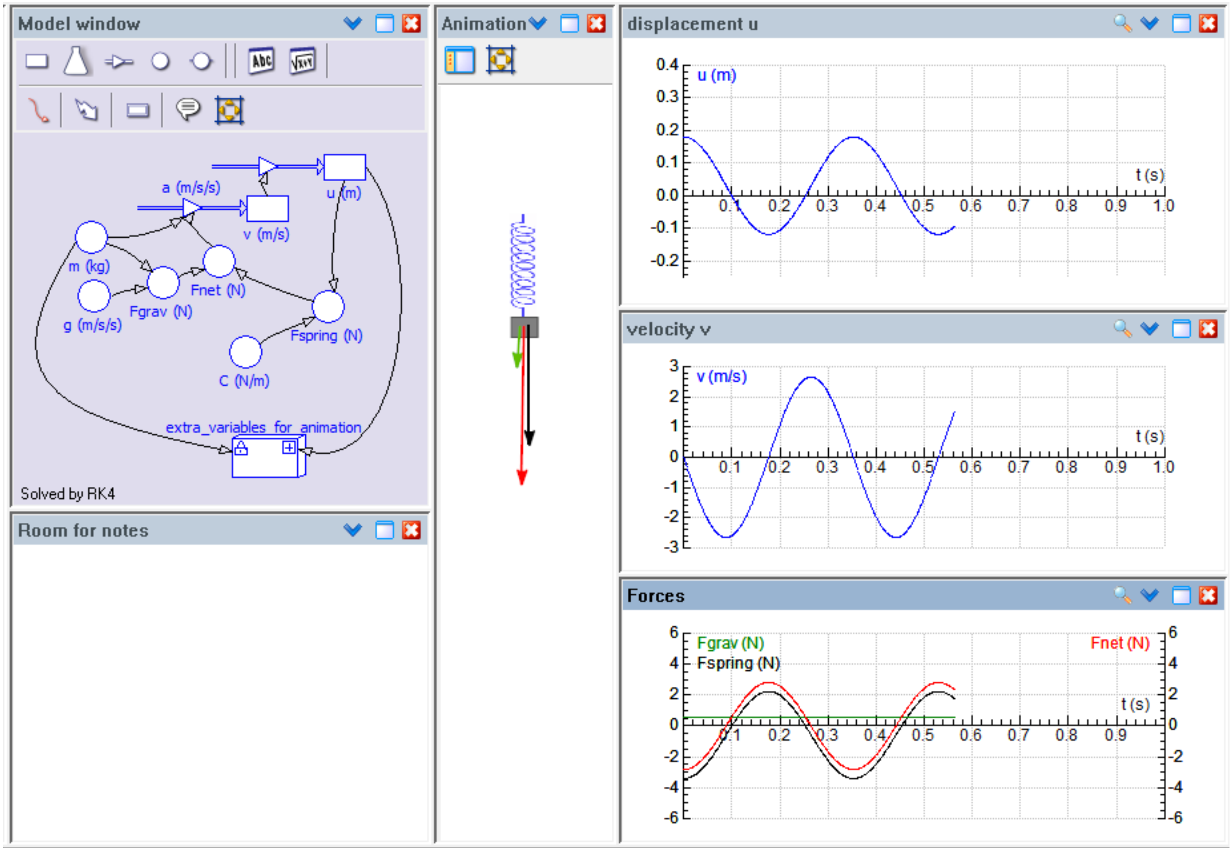

Fig. 2. Shot of part of the screen from a modelling activity in the first year of upper secondary education. A graphical model for a mass attached to a string in the upper left window drives an animation in the window in the middle. The vectors for the force of gravity F_{grav} ,the spring tension F_{spring} and the net force F_{net} are also animated. The graphs in the windows on the right are drawn simultaneously. The 'suitcase' in the model contains variables that are necessary of the animation, but not for the model itself. Such suitcase can also be used to store and hide correct models that students can use to evaluate their own models.

Another way to enable self-correction is to prepare students before the start of the modelling task, so that they know what to expect. For this reason, we combine modelling with experimenting. By doing experiments first, students can get acquainted with the behaviour of the real system. The experiments also provide data that can be used as a target result for the model. In Coach, these data can be presented in the form of a background graph.

Target results can also be created by letting students do some calculations beforehand. One way is to do a few iterations of the calculation process manually and create a table. If the output of the model is also presented in a table, students can recognise the values in the table. A more sophisticated way is to let student calculate specific properties that can be expected from the model output. An example is the constant velocity that is reached by an object falling a long time with air resistance and that can be calculated beforehand.
A special way of creating a target result is by providing students with a hidden correct model which uses the same initial values and constants as the model that students build themselves. In Coach 6, a 'locked suitcase' can be used to hide this correct model. Such a suitcase is shown in the model in Figure 2.
There is a drawback to the use of target results: they may stimulate trial-and-error behaviour. In addition, an incorrect models sometimes can create 'correct' output. Therefore, students must know the learning goals of the task and must be stimulated to reflect on their own work, as is advised to us by our students in classroom-discussions.

5.2 Hands-on versus pencil and paper

A first reason to assess ICT-related competencies not only hands-on but also by means of pencil-and-paper exercises is given in the introduction section of this paper. In a pencil-and-paper exercise, it is easier to focus on a specific competency, without the complexity and the contaminating effects of other competencies, such as the competency to use the software. This was confirmed by our students. They added to this that computer modelling tasks were often more complex than the tasks they use to perform without a computer.
But there are at least four other reasons. The first is a theoretical reason, drawn from the work of the Russian educator Gal'perin. With a concrete object (the computer model) at hand, the actions of a student tend to stay at a concrete level. In order for these actions to become mental, the concrete object must be removed (cf., Haenen, 2001). Therefore, after a modelling task, students are given written exercises to stimulate reflection and to support the process of internalisation. Three other reasons were given to us by students on several occasions. Firstly, students tend not to rehearse ICT-activities before a test because they tend to rehearse only the textbook (Van Buuren, 2014). Secondly, it can be cumbersome for students to get and start a computer to rehearse and practice modelling, especially in case the learning goal is a single competency. A recent example is the ability to construct conditional statements (if…then…else…-statements). Students explicitly advised us to add pencil-and-paper exercises on this subject. Finally, as students spontaneously explained in interviews, practical work is considered less important by students because it is usually not tested in school practice in most sciences. According to these students, this argument holds for modelling tasks as well. They advise and even warn us to test practical work, ICT-competencies, and modelling in regular tests because this stresses their importance. If this is the purpose of a test, written tests are in school practice a less cumbersome alternative for hands-on testing.

For summative purposes, we developed both completely written tests and tests that were partly hands-on: students had to use a modelling program on the computer. Comparing the ways students worked with these tests, we occasionally found a noticeable difference. In completely written tests, students can easily leave errors in their answers unnoticed. A teacher, going over these answers, can detect such errors but can also see the other, correct steps the students have made while answering the question. In case students have used a modelling program, students more easily detect their own errors because they can evaluate their answers by running the model. As a consequence, they may correct their errors, but we also observed students who completely ruined answers that contained only a minor error after running their model. These students realised that there was a flaw in their model, but sought in a wrong direction to correct it. This can have a demotivating effect on these students.

6 Conclusions

As we have shown, computer modelling offers possibilities for self-correction by students if the output of the model is represented in a comprehensible way. For this purpose, animations and target results can be useful. The possibility of self-correction can have a demotivating effect in case of summative tests in which students work with a modelling computer program if students are not able to detect their errors. We recommend to assess computer modelling and ICT-supported experimenting not only hands-on, but also by means of pencil-and-paper tasks, whether the purpose is formative or summative, because this makes it possible to focus on specific single competencies. Another recommendation is to assess modelling not only in the internal school exams but also in the nationwide final exams, because for both teachers and students, this stresses the importance of modelling.

References

Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics*, *62*(8), 750–762.

Boeijen, G. & Uylings, P. (2004). Exams of tomorrow. In E. Mechlová (ed.), Teachning and Learning Physics in New Contexts, Proceedings of GIREP 2004 (pp. 153-154). Ostrava: University of Ostrava.

Black, P., & Harrison, C. (2010). Formative assessment in science. In J. Osborne & J. Dillon (Eds.), *Good Practice In Science Teaching: What Research Has To Say* (p. 183). McGraw-Hill, Open University Press.

Etkina, E., Karelina, A., & Ruibal-Villasenor, M. (2008). How long does it take? A study of student acquisition of scientific abilities. *Physical Review Special Topics - Physics Education Research*, *4*(2), 020108.

Forrester, J. W. (1961). *Industrial Dynamics*. Cambridge, MA.: Mit Press.

Gott, R., & Duggan, S. (2002). Problems with the Assessment of Performance in Practical Science: Which way now? *Cambridge Journal of Education*, *32*(2), 183–201.

Haenen, J. (2001). Outlining the teaching–learning process: Piotr Gal'perin's contribution. *Learning and Instruction*, *11*(2), 157–170.

Heck, A., Kedzierska, E., & Ellermeijer, T. (2009). Design and implementation of an integrated computer working environment for doing mathematics and science. *Journal of Computers in Mathematics and Science Teaching*, *28*(2), 147–161.

Lijnse, P. (2008). Models of / for Teaching Modeling. In E. Van den Berg, T. Ellermeijer, & O. Slooten (Eds.), *Modelling in Physics and Physics Education.* (pp. 20–33). Amsterdam: University of Amsterdam.

Lillard, A. (2007). *Montessori: The Science Behind the Genius*. Oxford University Press, USA.

Millar, R. (2010). Practical work. In J. Osborne & J. Dillon (Eds.), *Good Practice In Science Teaching: What Research Has To Say* (p. 108). McGraw-Hill, Open University Press.

Montessori, M. (1912). *The Montessori Method*. (A. Everett George, Trans.). New York: Frederick A. Stokes Company. Retrieved from http://digital.library.upenn.edu/women/montessori/method/method.html

Roberts, R., & Gott, R. (2006). Assessment of performance in practical science and pupil attributes. *Assessment in Education: Principles, Policy & Practice*, *13*(1), 45–67.

Savelsbergh, E. (Ed.). (2008). *Modelleren en computermodellen in de β-vakken: Advies aan de gezamenlijke β-vernieuwingscommissies*. Utrecht: FISME.

Van Buuren, O. (2014). *Development of a Modelling Learning Path* (Doctoral thesis, University of Amsterdam). University of Amsterdam, Amsterdam.

Van Buuren, O., Heck, A., & Ellermeijer, T. (2015). Understanding of Relation Structures of Graphical Models by Lower Secondary Students. *Research in Science Education*, 1–34.

Van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (Eds.). (2006). *Educational Design Research*. London and New York: Routledge.

Wiliam, D., & Black, P. (1996). Meanings and consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, *22*(5), 537–548.