Chapter 40
Formative Assessment in Everyday Teaching of Mathematical Modelling: Implementation of Written and Oral Feedback to Competency-Oriented Tasks

Michael Besser, Werner Blum, and Malte Klimczak

Abstract The interdisciplinary research project Co²CA investigates how assessing and reporting students' performances in mathematics can be arranged in every-day teaching in such a way that teachers are able to analyse students' outcomes appropriately and organise further learning as well targeted as possible. In this context, 39 classes of German middle track schools were observed for several weeks while dealing with mathematical tasks focusing on technical and modelling competencies. Based on the assumption that assessing and reporting students' outcomes regularly will foster learning processes, students from some classes were given individual, task-related feedback, in some classes several times in a written form, in some classes in addition permanently accompanying the students' solution processes. In this chapter, we describe the study and report some preliminary results.

1   Introduction

Following ideas of the Danish KOM-project (Blomhoj and Jensen 2007; Niss 2003) and of activities in the context of the development of national education standards for mathematics in several countries (Deutsche Kultusministerkonferenz 2003; National Council of Teachers of Mathematics 2000), the discussion of how to improve competency-oriented teaching and learning of mathematics is of central interest in

M. Besser (✉)
Fachbereich Mathematik, Institute of Mathematics, University of Kassel, Kassel, Germany
e-mail: mbesser@mathematik.uni-kassel.de

W. Blum
Department of Mathematics, University of Kassel, Kassel, Germany

M. Klimczak
Department of Educational Quality and Evaluation, Deutsches Institut für Internationale Pädagogische Forschung, Frankfurt am Main, Germany
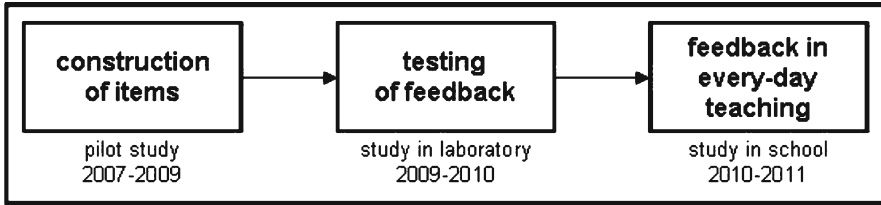
Fig. 40.1 Stages of the research project Co²CA (2007–2011)

mathematics education. Considering the tension between 'unguided learning' on the one hand and 'instructional learning' on the other hand (DeCorte 2007; Hoops 1998; Kirschner et al. 2006; Mayer 2004), several studies have tried to find out how every-day teaching of mathematics could be arranged so as to foster students' learning as well targeted as possible (among many others, see e.g. Dekker and Elshout-Mohr 2004; Leiss 2010; Teong 2003).

The interdisciplinary research project Co²CA (*Conditions and Consequences of Classroom Assessment*)[1] aims at investigating the impact of different kinds of feedback in competency-oriented mathematics teaching on students' performances, emotions and attitudes. In a first step, starting in 2007, competency-oriented tasks (modelling tasks and technical tasks) that were to assess students' outcomes reliably have been constructed successfully. In a second step, special kinds of feedback to students' responses on the constructed items have been developed and tested in the laboratory (Besser et al. 2010; Bürgermeister et al. 2011; Klieme et al. 2010). Here the effect of feedback on performance tests based on marks has been compared to criteria-based feedback (students who are as good as you are generally able to deal with the following topics) and feedback directly based on students' working processes (as can be seen from your answers to the test, you are able/not able to deal with the following topics). In a third step, from October 2010 to March 2011, the items as well as the feedback that had been developed were implemented in a 13 lesson teaching unit in 39 Year 9 classes of German middle track schools (see Fig. 40.1 for a timetable of the Co²CA-project; for a short overview of this special part of the study see Besser et al. 2011). In relation to this last step, one of the main research questions is: Will students in classes with an optimized kind of written and oral feedback outperform their counterparts who are not given such feedback, especially concerning their modelling competency?

In this chapter we will present the design of the Co²CA study in school as well as some very first results of this study that hint at challenges we have to deal with in further steps.

2 Implementation of Feedback in Every-Day Mathematics Teaching: Design of the Co²CA Study in School

According to results of pedagogical and psychological research (Hattie and Timperley 2007), it is reasonable to assume that assessing and reporting students' outcomes regularly in short intervals will foster students' learning. Such so called "formative assessment" (in contrast to ideas of "summing up" students' results only once at the end of a unit; for a general discussion about formative assessment see for example Black and Wiliam 2009) is said to be even more successful if the students are continuously offered feedback that is informative, individual and task-related (Deci et al. 1999; Kluger and DeNisi 1996) and if the assessment tries to answer some central questions concerning the students' learning processes: "Where am I going?", "How am I going?", and "Where to next?" (Hattie and Timperley 2007, p. 88).

The Co²CA project tries to implement written and oral feedback into teaching that sticks closely to the above-mentioned principles, that means it is given individually to students in short intervals (written feedback: three times during the 13 lessons; oral feedback: on the fly whenever possible), refers to students' solution processes, points out students' strengths as well as difficulties and offers strategies for students on how to improve themselves – especially feedback that helps students to concentrate on individual weaknesses and strengths on their own. In contrasting three different groups of students, the main question of how such feedback influences students' performances is pursued by the following research design (see Fig. 40.2).

2.1 A Teaching Unit Dealing with Pythagoras' Theorem

Altogether 39 Year 9 classes from 23 middle track schools (Realschule) in the state of Hessen (Germany) with 978 secondary students participated in this study. This sample can be regarded as fairly representative for this ability and age group. The classes were assigned randomly to either a control group CG: no special kind of feedback is given to the students, or one of two experimental groups, that is EG 1: students are given
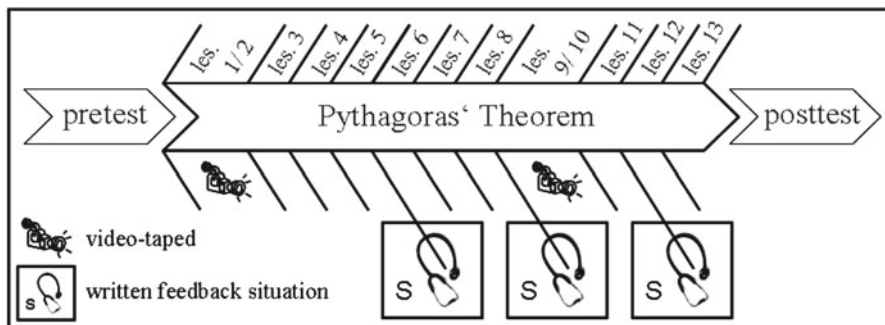

Fig. 40.2 Design of the Co²CA study in school (2010–2011)

written feedback three times within the 13 lessons and EG 2: in addition to written feedback students are supported by a special kind of oral feedback. Before starting the study, all teachers participated in half a day training to conduct a 13 lesson unit dealing with the topic area of Pythagoras' theorem. These 13 lessons comprised an introduction to Pythagoras' theorem (including a proof of the theorem), a phase with technical items, a phase with dressed up word problems and finally a phase with more demanding modelling problems. Referring to Kaiser (1995) and Maaß (2010) these modelling problems can be characterized in such a way that students' have to pass through the whole modelling cycle but that they only have to hark back to standardized, familiar ways of calculating. To control for the quality of teaching, every teacher was given a so called "logbook" with obligatory and optional tasks to use during the lessons. In addition, 4 of the 13 lessons were video-taped in all the classes.

2.2 Written Feedback in Both Experimental Groups

In the classes of the two experimental groups (EG 1 and EG 2) the students had to work on special short tasks on three occasions (at the end of lessons 5, 8 and 11). At the beginning of the next lesson, all students got back their solution, corrected by the teacher, together with an individual, process-oriented, written feedback and a suitable exercise to work on. The teachers were prepared to do so on a second half day of training. To ensure that all participating students worked on the aforementioned special short tasks, these were integrated into the regular lessons of the control group. An example of such a written feedback can be seen in Fig. 40.3. This example shows feedback given to the following modelling item that students were given at the end of lesson 11.

The rope of the cable car *Ristis* has to be replaced. 1 m of the rope costs 8 €. How much does a new rope cost approximately? Write down your solution process.
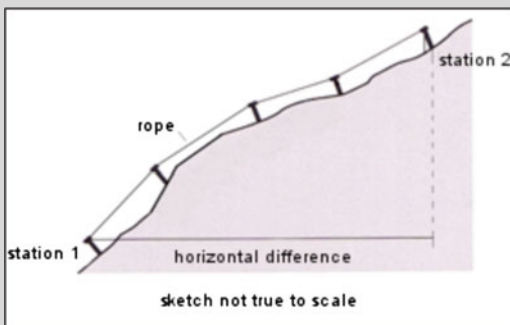
Name:	Cable car "Ristis"	Weight capacity:	132 × 3 persons
Station 1:	1,600 m above sea level	Haul capacity:	1,200 pers. per hour
Station 2:	1,897 m above sea level	Speed:	1.5 m/s
Horizontal difference:	869 m	Time of travel:	10 min

$$869^2 + 267^2 = c^2$$
$$c^2 \quad = 826470$$
$$c^2 = \approx 909$$
$$909 \cdot 8 = 7272$$

A: The rope costs about 7272 €.

YOUR PERSONAL FEEDBACK

You are already quite good at dealing with the following topics:
You are able to work on Pythagoras' theorem within a real situation.
You are able to write down an answer to your solution

You can still imporve at dealing with the following topics if concentrating on my hints:	*Hints on how you can improve:*
you have problems to understand the real situation	think about the real situation Concretely and have a closer look at the picture
you have problems working on equations	if extracting squareroot do so on both sides of the equation

Fig. 40.3 Example of written feedback

2.3 Oral Feedback in One of the Two Experimental Groups

In addition to the written feedback, the teachers of experimental group 2 (EG 2) were trained on a third half day to implement a special kind of oral feedback that copes with the requirements of competency-oriented tasks in every-day teaching of mathematics, similar to the so-called "operative-strategic" teaching method developed in the DISUM project (here students mainly have to deal with mathematical modelling tasks in groups and with only little support by the teacher; for details see Blum 2011 and Schukajlow et al. 2011). According to ideas of the DISUM project, the teachers were trained to orally intervene into students' working processes only by minimal-adaptive support in order to let the students work on their own as much as possible (Leiss 2005). The participating teachers were informed about different ways of intervening and supporting. Here we distinguish between four categories of teacher interventions: metacognitive interventions that give hints on a

meta-level (such as 'Imagine the situation'), interventions related to the special content of a problem, affective interventions (such as 'Well done so far'), and interventions referring to the organizational context in the classroom (Leiss 2007; Leiss and Wiegand 2005).

2.4 Pre-test and Post-test

To control for students' prior mathematical knowledge there was a pre-test immediately before the study and, to find out differences between students' mathematical performances, a post-test at the end of the study. Both tests only consisted of items that have been empirically identified as technical items (TI) or modelling items (MI) as a result of the pilot study (pre-test: 13 TI, 6 MI; post-test: 9 TI, 8 MI). Since students normally cannot solve items dealing with the topic of Pythagoras' theorem before this topic is explicitly taught, here only 'prior knowledge' – elements that were necessary to work on Pythagoras' theorem in the following weeks – was asked for (e.g., finding the square root of a number or naming characteristics of a triangle). Both tests could be linked by the item-parameters known from the pilot study. Examples of a pre-test item testing prior knowledge, a technical post-test item and a modelling post-test item are given below.
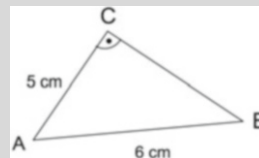
Prior-knowledge pre-test item:

A broom is rested against a wall as shown below.
Broom, wall and bottom form a triangle. Mark the triangle in the picture and give names to the sides.

Technical post-test item:

Calculate the length of the side a = |BC|.
a = _____

Modelling post-test item:

On May 1st people in Bad Dinkelsdorf dance around a so called "Maibaum". This is a tree which has a height of 8 m. While dancing, the people hold bands in their hands. These bands are 15 m long. How far away from the "Maibaum" are the people at the beginning of the dance?

3   Some Preliminary Results of the Field Study

Both pre-test and post-test have been rated and first analyses can be reported concerning the test results. The reported results are deduced from scores which have been given to the students' answers by trained raters, and these scores have been used for scaling the tests based on the Rasch model.

3.1   Test Results

Inter-rater reliability: The rating has been successful since the inter-rater reliability for the five trained raters can be said to be very good (pre-test: Cronbach's alphas between 0.829 and 1.000; post-test: Cronbach's alphas between 0.947 and 1.000).

Test reliability: Whether linked to the results of the pilot study or not, the *wle* (weighted likelihood estimation) and *eap* (expected a-posteriori) reliability of the tests[2] (as a one-dimensional mathematical construct) are acceptable (0.571–0.735). However, a two-dimensional scaling – separately for TI and MI – points to some problems concerning the MI dimension of the pre-test. First factor analyses hint at two out of six items not fitting sufficiently to this dimension.

Difficulties of the tests: One-dimensional as well as two-dimensional scaling illustrate bigger differences in the difficulty of the pre-test depending on whether the item-parameters of the pre-test are linked to the pilot study or not. If linked, the pre-test becomes much harder. Further analyses highlight that these differences

---

[2]Both *wle* reliability and *eap* reliability are computed in ConQuest as indicators for the reliability of a latent variable/construct. In general values greater than 0.6 are expected to be acceptable. For more details see also Rost (2004).

Table 40.1 Results of pre-test and post-test separated for CG, EG 1 and EG 2

Test	Test-results (one-dimensional-scaling)			Significance in differences	
	CG	EG 1	EG 2	CG vs. EG 1	CG vs. EG 2
Pre-test	−0.152	0.205	−0.420	No	Yes
	$SD=1.106$	$SD=0.929$	$SD=0.918$		($p<0.01$)
	$n=340$	$n=303$	$n=206$		
Post-test	−0.140	−0.230	−0.332	No	No
	$SD=1.274$	$SD=1.211$	$SD=1.219$		
	$n=334$	$n=302$	$n=220$		

seem to be caused by differences in technical abilities between the populations of the pilot study and of the field study. Since there were also higher track (Gymnasium) students in the population of the pilot study, these differences are apparently caused by these higher ability students (interestingly, there are no such differences concerning the modelling dimension of the pre-test and the TI or MI dimension of the post-test).

Differences in performance: One of the main questions of the study obviously is: Are there significant differences in the post-test performance between the three groups (control group and two experimental groups)? Unfortunately, this question cannot be answered satisfactorily yet – too many variables are not yet evaluated, and too little control for appropriate treatment implementation was possible to date. Nevertheless, some very first results concerning the students' performances in the control group and the two experimental groups shall be reported here, taking into account that these results have to be dealt with very carefully (here we only refer to results of a one-dimensional scaling of the tests since the reliability of the MT-dimension of the post-test is not really acceptable) (see Table 40.1).

Table 40.1 shows that there are no significant differences between CG and EG 1 or CG and EG 2 in the post-test. The control group performed significantly better in the pre-test than experimental group 2 (−0.152 vs. −0.420) and these differences are no longer visible in the post-test. Since analyses of covariance do not show any influences of the experimental condition either, we have to know a lot more about the quality of the implementation of the treatment to explain these effects – especially we have to know in detail what really happened in the 13 lessons.

3.2   Challenges for the Future

The main research question of Co^2CA is whether special kinds of formative assessment – theoretically based and optimised forms of written or oral feedback – can help teachers to improve students' learning processes when dealing with competency-oriented mathematical tasks (here: with technical and modelling tasks) and whether an implementation in everyday teaching can foster students' performances. Except for one special case (the reliability of the MI dimension of the pre-test), the performance

tests have worked quite well. Within the next few months, further analyses have to be carried out in order to answer the main question stated above, that is to find out whether there are differences in students' outcomes between different groups and whether these differences are really caused by our treatments. Therefore, the big challenge is to control both for the overall quality of teaching (by analysing about 160 h of video-taped lessons; see Lipowsky et al. 2009 for some relevant variables) and for the quality of written and oral feedback given by the teachers (by developing adequate coding schemes for both written and oral feedback). We will report about these analyses in the near future in particular at the next ICTMA.

References

Besser, M., Leiss, D., Harks, B., Rakoczy, K., Klieme, E., & Blum, W. (2010). Kompetenzorientiertes Feedback im Mathematikunterricht: Entwicklung und empirische Erprobung prozessbezogener, aufgabenbasierter Rückmeldesituationen. *Empirische Pädagogik, 24*(4), 404–432.

Besser, M., Klimczak, M., Blum, W., Leiss, D., Klieme, E., & Rakoczy, K. (2011). Lernprozessbegleitendes Feedback als Diagnose- und Förderinstrument: Eine Unterrichtsstudie zur Gestaltung von Rückmeldesituationen im kompetenzorientierten Mathematikunterricht. In *Beiträge zum Mathematikunterricht* (pp. 103–106). Münster: WTM Verlag.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Blomhoj, M., & Jensen, T. H. (2007). What's all the fuss about competencies? Experiences with using a competence perspective on mathematics education to develop the teaching of mathematical modelling. In W. Blum, P. L. Galbraith, H. W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education. The 14th ICMI study* (pp. 45–56). New York: Springer.

Blum, W. (2011). Can modelling be taught and learnt? Some answers from empirical research. In G. Kaiser, W. Blum, R. Borromeo Ferii, & G. Stillman (Eds.), *Trends in teaching and learning of mathematical modelling. ICTMA 14* (pp. 15–30). New York: Springer.

Bürgermeister, A., Klimczak, M., Klieme, E., Rakoczy, K., Blum, W., Leiss, D., Harks, B., & Besser, M. (2011). Leistungsbeurteilung im Mathematikunterricht – Eine Darstellung des Projekts "Nutzung und Auswirkunen der Kompetenzmessung in mathematischen Lehr-Lern-Prozessen". http://www.schulpaedagogik-heute.de/index.php/artikel-331.html. Accessed 12 Sept 2011.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668.

DeCorte, E. (2007). Learning from instruction: The case of mathematics. *Learning Inquiry, 1*(1), 19–30.

Dekker, R., & Elshout-Mohr, M. (2004). Teacher interventions aimed at mathematical level raising during collaborative learning. *Educational Studies in Mathematics, 56*, 39–65.

Deutsche Kultusministerkonferenz. (2003). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss*. München: Luchterhand.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hoops, W. (1998). Konstruktivismus. Ein neues paradigma für didaktisches design? *Unterrichtswissenschaft, 3*, 229–253.

Kaiser, G. (1995). Realitätsbezüge im Mathematikunterricht – Ein Überblick über die aktuelle und historische Diskussion. *Materialien für einen realitätsbezogenen Mathematikunterricht, 2*, 66–84.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experimental, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86.

Klieme, E., Bürgermeister, A., Harks, B., Blum, W., Leiss, D., & Rakoczy, K. (2010). Projekt Co²CA. Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht. *Zeitschrift für Pädagogik, 56*, 64–74.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.

Leiss, D. (2005). Teacher intervention versus self-regulated learning? *Teaching Mathematics and Its Applications, 24*(2–3), 75–89.

Leiss, D. (2007). *"Hilf mir es selbst zu tun". Lehrerinterventionen beim mathematischen Modellieren*. Hildesheim: Franzbecker.

Leiss, D. (2010). Adaptive Lehrerinterventionen beim mathematischen Modellieren – empirische Befunde einer vergleichenden Labor- und Unterrichtsstudie. *Journal für Mathematik-Didaktik, 31*(2), 197–226.

Leiss, D., & Wiegand, B. (2005). A classification of teacher interventions in mathematics teaching. *ZDM, 37*, 240–245.

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean theorem. *Learning and Instruction, 19*, 527–537.

Maaß, K. (2010). Classification scheme for modelling tasks. *Journal für Mathematik-Didaktik, 31*(2), 285–311.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case of guided methods of instruction. *The American Psychologist, 59*(1), 14–19.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

Niss, M. (2003). Mathematical competencies and the learning of mathematics: The Danish KOM project. In A. Gagatsis & S. Papastavridis (Eds.), *Mediterranean conference on mathematical education* (pp. 115–124). Athens: 3rd Hellenic Mathematical Society and Cyprus Mathematical Society.

Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. Bern: Huber.

Schukajlow, S., Leiss, D., Pekrun, R., Blum, W., Müller, M., & Messner, R. (2011). Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educational Studies in Mathematics, 79*(2), 215–237.

Teong, S. K. (2003). The effect of metacognitive training on mathematical word-problem solving. *Journal of Computer Assisted Learning, 19*, 46–55.