ASSESSMENT FOR LEARNING

Malmö Studies in Educational Sciences No. 68

ANDREIA BALAN
ASSESSMENT FOR LEARNING
A case study in mathematics education

The publication is also available online,
see www.mah.se/muep

To my parents

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

ABSTRACT

The aim of this study was to introduce a formative-assessment practice in a mathematics classroom, by implementing the five strategies of the formative-assessment framework proposed by Wiliam and Thompson (2007), in order to investigate: (a) if this change in assessment practices had a positive influence on students' mathematical learning and, if this was the case, (b) which these changes were, and (c) how the teacher and students perceived these changes in relation to the new teaching-learning environment.

The study was conducted in a mathematics classroom during the students' first year in upper-secondary school. A quasi-experimental design was chosen for the study, involving pre- and post-tests, as well as an intervention group and control group. The intervention was characterized by: 1) making goals and criteria explicit by a systematic use of a scoring rubric; 2) making students' learning visible by a use of problem-solving tasks and working in small groups; 3) providing students with nuanced information about their performance, including ways to move forward in their learning; 4) activating students as resources for each other through peer-assessment and peer-feedback activities; and 5) creating a forum for communication about assessment, involving both the students and the teacher.

The findings indicate an improvement in problem-solving performance for the students in the intervention group, for instance regarding how well they are able to interpret a problem and use appropriate mathematical methods to solve it. The students also show improvements in how to reason about mathematical solu-

tions, how to present a solution in a clear and accessible manner, and how to appropriately use mathematical symbols, terminology, and conventions. The findings also indicate a change in students' mathematical-related beliefs during the intervention, towards beliefs more productive for supporting learning in mathematics. The changes in students' beliefs include mathematical understanding, mathematical work, and the usefulness of mathematical knowledge. During interviews, the students expressed how they perceived the new teaching-learning environment. Students' responses indicate that they recognized and appreciated the different components of the formative-assessment practice as resources for their learning. Responses from both students and the teacher also indicate that the components of the formative-assessment practice were linked in complex ways, often supporting and reinforcing each other. Furthermore, most components had other effects as well, besides supporting the formative strategies they were intended to.

The findings from this study deepens our understanding of how the components of a formative-assessment practice may influence students and their learning in mathematics, but also how these components co-exist in an authentic classroom situation and influence each other.

LIST OF FIGURES

14

LIST OF TABLES

16

INTRODUCTION

Swedish students' results in mathematics and mathematics education as a whole have lately been a subject of debate in both the media and the educational community. Results from international studies like TIMMS (and PISA) demonstrate a negative trend for Swedish students, both in performance and in the motivation to study mathematics (Skolverket, 2010b). This tendency is confirmed by national test results showing a growing number of students in upper secondary school not reaching a passing level (Skolverket, 2010a). Furthermore, there are reports showing that Swedish students, regardless of their abilities, find mathematics boring (Skolverket, 2004).

This trend may have several explanations, such as the teaching, teacher competency (or lack thereof), the curriculum, the school system as such, segregation, or even the society at large. Mathematics as a science has undergone profound changes, which to a certain degree, is mirrored in the curriculum, but less so in everyday teaching practices. Mathematics is no longer viewed as a body of infallible and objective truths, but rather as a set of human sense-making activities, a product of human inventiveness and as problem-solving activities based on the modeling of reality (De Corte, 2004; Ernest, 1991). The changes are internationally reflected in documents such as Curriculum and Evaluation Standards for School Mathematics (National Council of Teachers of Mathematics, 2000, US), A National Statement on Mathematics for Australian Schools (Australian Education Council, 1991), and Syllabuses and Grading Criteria (Swedish National Agency for Education,

2000). All of these curricula have a strong emphasis on the acquisition of mathematical problem-solving skills, reasoning and communicating skills, and the application of mathematical knowledge in "real-life situations". These changes ought to influence the teaching and the assessment of students.

The change of perspective about mathematics as a science also has implications for how researchers define mathematical competency. In a review based on research from the last 25 years, Erik De Corte (2004) summarizes five aptitudes that students need to acquire in order to be competent in mathematics: 1) domain-specific knowledge that involves facts, symbols, rules, concepts, and algorithms that are well organized and flexibly accessible; 2) heuristics methods (i.e. a systematic approach to the representation, analysis and transformation of mathematical problems (see Koichu, Berman, & Moore, 2007); 3) meta-knowledge; 4) self-regulatory skills, which involves the self-regulation of the cognitive process (i.e. students are expected to be "meta-cognitively, motivationally and behaviorally active participants in their own learning process", De Corte, 2004, p. 290); and 5) beliefs about mathematics, mathematical learning, and the self in relation to mathematics.

Starting from the review, De Corte observes a mismatch between the above demands of mathematical competence and the current teacher-made tests. Traditional techniques of educational evaluation focus on the assessment of memorized knowledge, and the mastery of low-level skills, instead of giving information with regard to students' mathematical dispositions, problem-solving skills, and ability to communicate mathematical ideas. Since assessment influences learning, it can be an important factor in ordinary classrooms in explaining the declining results in mathematics; a factor which, however, is seldom discussed in this context. The fact that assessment has implications for students' learning is demonstrated by a number of research studies. For example, researchers (Dysthe, 2008; Shepard, 2000) are pointing to the gap between current theories of learning and the common assessment forms; the latter primarily having a focus on assessing what has been learned or not learned after the teaching has ended. Researchers argue that new forms of assessment are needed to match current theories of learn-

ing. Such assessment practices need to focus not only on the summative assessment *of* learning, but also on assessment *for* learning – that is assessment used for formative functions (Havnes & McDowell, 2008; Jönsson, 2008).

The role played by assessment in both forming the way the subject is taught, and in the way it is perceived by students and teachers, is also well documented. Several studies have, for example, investigated the relationship between students' perception of assessment and the learning outcomes (Brown & Hirschfeld, 2008; Entwistle & Entwistle, 1991; Gijbels, Segers, & Struyf, 2008; Nijhuis, Segers, & Gijselaers, 2005; Sambell, McDowell, & Brown, 1997; Struyven, Dochy, & Janssens, 2003). Students' perceptions of the learning environment seem, in fact, to affect students' approaches to the learning task more than the task itself (Entwistle & Entwistle, 1991), and students, who see assessment as a means of taking responsibility for their own learning, are reported to achieve more (Brown & Hirschfeld, 2008).

Furthermore, students' perceptions of the purpose of the assessment also seem to affect their achievement. Assessments that are perceived as "inappropriate" tend to encourage surface approaches to learning, while innovative assessment methods, which are perceived as "fair", may help students to learn in a deep way (Struyven et al., 2003). A surface approach to learning is characterized by memorization and reproduction, without seeking to understand what is being learned. Learning in a deep way, on the other hand, is associated with a search for meaning. Students who are acquainted with the expectations, have been shown to experience more freedom of learning and to a larger extent tend to adopt a deep approach to learning (Gijbels et al., 2008), whereas students who, among other things, do not see the goals clearly, experience less freedom and more often adopt a surface approach to learning (Entwistle & Ramsden, 1983; Trigwell & Prosser, 1991).

Another aspect of assessment that has also been shown to influences students' approaches to learning, besides their perceptions of the purpose of assessment and students' familiarity with the learning goals, concerns the demands set by the assessment as experienced by the students. For example, if the assessment was per-

ceived as requiring passive acquisition and reproduction, the students adopted a surface-learning approach with low level cognitive strategies. If, however, the assessment was perceived to require understanding, integration, and application of knowledge, the students were more likely to develop deep approaches to learning (Nijhuis et al., 2005). The ways in which students perceive the learning task and the assessment are further influenced by their experiences, their social-economic background, their motivation, and their study orientation (Sambell & McDowell, 1998).

A problem in relation to the discussion above, which appears clearly in the debate in the educational community, is that in spite of a number of research studies demonstrating the need for changes in the way assessment is applied in the classroom, assessment in school is still used mainly for summative purposes (e.g. Lindberg, 2007). Classroom assessment thus often contributes to the conservation of practice by using assessment as a means for grading and the possibility of using assessment as a tool for learning does not seem to be part of the ordinary classroom practice. Jesper Boesen (2006), for instance, highlights that the focus in mathematical education in Sweden, rests primarily on the procedural and algorithmic aspects of mathematical activity. He further notes that teacher-made tests consist mostly of tasks requiring imitative reasoning, meaning the type of reasoning where students copy or follow a model or an example without any attempt at creativity. This is in line with the conclusion by De Corte (2004), about the mismatch between the visions of the mathematical competence needed, and what is currently assessed in the ordinary classroom.

Many researchers of today (e.g. Black & Wiliam, 1998a; Dysthe, 2008; Sadler, 1989) assume that the use of formative assessment might be a valuable tool for both students and teachers in focusing and enhancing the learning processes. Assessment used formatively, it is argued, helps students to focus on what has been mastered, on difficulties experienced, and on strategies adopted. Furthermore, when applied as self-assessment, it has a potential to foster meta-cognitive thinking; when used in groups as peer assessment, it may help students to discover alternative ways of solving problems; and when used in combination with transparent goals, students may

gain a sense of meaning and control. Together these steps may help both student learning and student motivation. To teachers, assessment used for formative purposes offers a possibility to experience (at least part of) the students' learning processes, thereby identifying both strengths and needs for further development; information that can then be used to support students' learning.

In conclusion, a negative trend has been observed for Swedish students' results in mathematics as well as for their motivation to study mathematics. One explanation for this trend may be sought in the assessment of mathematical competency. For one thing, there seems to be a discrepancy between the competencies that students need to develop and what is actually assessed, which may steer student learning in the wrong direction. Furthermore, if the ordinary classroom-assessment practice in mathematics education is mainly summative, focusing on grades, this may not be optimal for student learning. Instead, there are reasons to believe that the use of formative assessment might be a valuable tool for both students and teachers in focusing and enhancing the learning processes and that the introduction of formative assessment is likely to affect students' learning of mathematics and their mathematical performance. The following chapter will therefore look more deeply into the research on formative assessment.

FORMATIVE ASSESSMENT

Towards the end of the 1980s new perspectives on knowledge assessment began to make a breakthrough and several researchers speak of the emergence of a new assessment culture. The changes were both about what is to be assessed and how and why it is assessed, and about who does the assessing and how the results are used. Assessment was thus no longer only about how the assessment of individuals could be made fairer and more efficient, neither did it only encompass the final product of learning, but rather was just as much about how new methods and systems for assessment could be developed in order to support students' learning (Korp, 2011). This new assessment culture had its foundations in several different (theoretical, empirical, and ideological) perspectives, which are often interlinked in practice. However, for analytical purposes, a division into different orientations can be made.

One such orientation is, for example, that of a changed view of knowledge. Lorrie Shepard (2000) maintains that there is a distinct connection between psychometric testing (sometimes called "traditional assessment") and the behaviorist view of knowledge and learning, while the new assessment culture is primarily based on constructivist and socio-cultural theory formation. In this, a division arises partly due to the view of knowledge, upon which psychometric assessment is based, as something that is objective and directly transferable, for example from the teacher to a more or less passive student, while learning, from a constructivist perspective, consists of a process of meaning creation; an active course in which the learner needs to be involved and in which earlier experi-

ences play an important part in determining what is learnt. One conclusion of a constructivist view of knowledge is thus that knowledge cannot be measured or investigated in an objective way, rather all such measurements are dependent on, and must be interpreted in relation to, their contexts. The method of working within the psychometric tradition with "de-contextualized" measurements (i.e. where the measurement is not connected to any particular situation, but where knowledge is instead thought of as something general and applicable to a number of different contexts), is thus incompatible with a constructivist view of knowledge (Jönsson, 2011).

Another orientation within the new assessment culture questions the breaking down of complex processes and knowledge into smaller units, or "items" (which are often multiple-choice or short-answer questions), in order to be able to test them. Thus, this critique is based on the limitations of instruments and the methods of analyzing results (i.e. analytical methods that require multiple-choice formats in order to be able to provide reliable results) rather than on the view of knowledge. The principle argument for changing prevailing assessment practices is that we need to be able to carry out assessments of complex knowledge and processes, since these are called for in society, while the instruments and analyses of today are not capable of capturing such knowledge. Rather, multiple-choice and short-answer questions are regarded as primarily being good for testing simple factual knowledge learnt by rote. This orientation has thus focused on the development of methods for assessing complex knowledge in applied contexts (e.g. Segers, Dochy, & Cascallar, 2003).

Yet another orientation within the new assessment culture has proceeded from the steering effect that assessment seems to have on certain students' learning (see e.g. Biggs, 1999; Struyven et al. 2003). Since it is often in the interest of the students to achieve as well as possible on assessments, they tend to adapt their learning to the assessment. For example, it has been shown that it is relatively easy, with questions of a reproductive nature, to get students to adopt a superficial approach to learning. On the other hand, it seems to be more difficult to do vice-versa, that is, to get students

to acquire an in-depth approach to learning (e.g. Gijbels & Dochy, 2006; Marton & Säljö, 1984; Wiiand, 1998). Thus, the principle use of multiple-choice and short-answer questions heightens the risk of students aiming to learn simple factual knowledge learnt by rote, instead of focusing on the actual learning goals according to the curriculum. Certain researchers (e.g. Frederiksen & Collins, 1989) thus argue that assessments should be constructed and used in ways that lead to improvements of the knowledge they are intended to assess, rather than leading to the narrowing of the breadth of the curricula.

The fourth (and last) orientation within the new assessment culture to be mentioned here, is an orientation that proceeds from the possibility of using the assessment as a tool for improving the teaching and for providing better conditions for students' learning. It is this orientation that is most often referred to when speaking of "formative assessment". Since this is also the orientation that forms the foundation of this thesis, it will be discussed in more detail.

Assessment with a formative purpose

Of fundamental importance for assessments with a formative purpose is that they support students' learning in some way. In order for them to do so, they need to provide nuanced information about the students' performances in relation to (predetermined) goals and criteria. This is necessary in order for strengths and weaknesses to be identified and used as a basis for the students' continued development towards the goals.

Even though assessments that are conducted for summative purposes should also be based on detailed information about the students' performances in relation to goals and criteria, the difference lies in that the information cannot be summarized in brief statements, such as "Passed" or "Achieves the goals", if they are to be used formatively. This is because such statements do not reveal anything about the students' strengths or needs for further development. Neither do they provide any guidance as to how the student students can develop further. However, for summative purposes, such statements can be fully sufficient.

Thus, the fundamental difference between formative and summative assessments lies in how the information is used – not in how it is collected. This means that assessments with different purposes do not necessarily have to be different; rather the same assessment can be used for both formative and summative purposes. For example, the results from several formative assessments can be compiled to form a summative judgment (like a report).

The information provided by the assessment for a formative purpose can be used in different ways: a teacher can either use the information to give feedback directly to the students, for example "You've done it this way. Try to do it like this next time", or the teacher can change the teaching on the basis of this information, for example: "The teaching doesn't seem to have given the results I'd expected. I'll try to make another exercise, where instead we can …". Furthermore, it does not have to be the teacher who uses the information; it can also be the students themselves. Royce Sadler (1989), whose theory about formative assessment has had a great impact, maintains that there are certain conditions that have to be fulfilled in order for the assessment to support students' learning in the long run. Basically, these conditions are about the need for the students themselves to develop the ability to assess the quality of what they do and then to utilize this information in order to improve their performances and to further develop towards the goals. If the students are always served feedback and suggestions for improvement by the teacher, they run the risk of becoming individuals who lack independence, in contrast to the case when they are forced to work actively with goals and criteria in relation to what they themselves (or their peers) achieve. Thus, practice in self and peer assessment is a central component of formative assessment.

Although the concept of "formative assessment" was introduced by Michael Scriven already in 1967, it was a relatively unknown concept up until the time when Paul Black and Dylan Wiliam (1998a) published their research review about formative assessment. By going through a large number of books and scientific journals, they found approximately 600 scientific articles which all, in some way, touched upon formative assessment. They then pro-

ceeded to compile a comprehensive research review based on these articles. Since there were several studies which showed that teaching, that in some way included formative assessment, led to statistically significant – and many times considerable – positive effects on the students' learning, the researchers came to the conclusion that there is evidence that formative assessment may work.

When attempts are made, for research purposes, to evaluate the effects of various changes in teaching practices, the measurement of effect size[1] is often used. For most of the studies about formative assessment this effect size lay in the interval between .4 and .7. This might seem small, but the fact is that these values are higher than for most changes that have been carried out in teaching practices. Black and Wiliam (1998b) themselves illustrate with the example that an effect size of .7 would lift England from a position in the middle to a position among the five best, in an international test such as TIMSS.

The conclusion that there exists scientific evidence that formative assessment works is of course open to discussion. On the one hand, the fact is that it is very difficult to achieve definite proof for anything within educational-science research. This is partly due to the complexity of the situations studied, in which several different factors concur (such as age, school subject, point in time, teacher, students' background, etc.). The fact that positive results have been observed under certain circumstances does therefore not necessarily imply that the same results would be reached under different conditions. Furthermore, since it is people who are the object of study – in contrast to lifeless things such as, for example, atoms or fossils – they do not always behave as expected. Thus, the same instruction might work differently for different people (for example depending on different experiences or interests), but also differently for the same person on different occasions (due to factors such as group dynamics or that the individual has previously been receptive, but has now tired).

In other words, there are an enormous number of various factors that can have an effect on a study in educational sciences, which

---

[1]The formula for computing the effect size, Cohen's *d,* is: $d = (\bar{x}_2 - \bar{x}_1)/s$, where $\bar{x}$ = mean (average of treatment or comparison conditions) and s = pooled standard deviation.

means that results need to be interpreted with great care. It can always be the special circumstances in the particular study that have contributed to the results (no matter whether they are positive or negative).

On the other hand, it is possible, once results have been gathered from a large number of studies, to see patterns that allow for the drawing of general conclusions, since the uncertainty present in the individual studies tends to even out. This also seems to apply to a lacking of quality in individual studies (Hattie, 2009), which means that research reviews provide a considerably more robust basis of conclusions than individual studies do. Since the studies in the review by Black and Wiliam stretched over several ages (from 5-year olds to university level), over several subjects, and over several countries and still pointed in very much the same direction, it is reasonable to come to the same conclusion as they did – even though it has been pointed out that it is difficult to draw such conclusions on the basis of the individual studies (see e.g. Dunn & Mulvenon, 2009).

No matter which stance is taken, more research in the area is required. This is important for several reasons, not least due to the fact that Black and Wiliam (1998b), on the basis of their research, point to a number of strategies that seem to be particularly effective in supporting students' learning (e.g. clear goals, task-related and constructive feedback, as well as practice in self-assessment), which would then be worth investigating further.

The review by Black and Wiliam (1998a) thus provided stimulation for a large amount of new studies. For example, they themselves carried out a project with a number of teachers in England, in which they tried, together with 24 teachers, to implement formative assessment in six different schools. The effect size in this study was on average .32 (Black, Harrison, Lee, Marshall, & Wiliam, 2003; Wiliam, Lee, Harrison, & Black, 2004).

Another study was carried out in 35 schools in Scotland, in which the teachers tried to develop a formative way of assessing, here too on the basis of the research reviewed by Black and Wiliam. The results of this study indicate that formative assessment can lead to the students taking more responsibility for their learning

and that they become more motivated, become more self-confident and, also, perform better. This applies particularly to low achieving students, which Black and Wiliam also showed. The work also had an effect on the teachers and their teaching, for example due to the teachers gaining a better understanding of the idea of formative assessment and that they, to a greater extent, abandoned a more teacher-centered way of working, in favor of a more student- and learning-centered teaching (Kirton, Hallam, Peffers, Robertson, & Stobart, 2007).

One possible weakness of the review by Black and Wiliam was that they wrote about formative assessment in very general terms, while formative assessment can be carried out in very many different ways and can thus lead to very different results. Jeffrey Nyquist (2003) therefore placed 185 studies, in which the effects of formative assessment (in higher education) had been investigated, along a gradient, starting from what he called weak feedback (in which, for example, the students are only told whether they had been right or wrong) and that ends in strong formative assessment (in which the students are told what is good and what can be further developed, as well as how the students should go about moving forward).

What Nyquist shows in his study, is that a strong formative-assessment practice has a decidedly greater effect on students' learning (.56) compared to weaker forms (.14-.29). This indicates that students cannot simply be told what they have done and then be expected to know, by themselves, how to go forward with this information. Rather, it is necessary to help them by pointing out a direction forwards. The same results were arrived at in a research review about feedback by John Hattie (2009), drawing from the results from 23 meta-studies (i.e. studies which, in turn, had re-analyzed the results of individual studies, in this case a total of 287 studies). On the one hand he shows that feedback can give positive effects on students' learning (effect size = .73), and, on the other hand, that feedback can be given in different ways and thus be more or less effective.

Just like in Nyquist's study, it has been shown to be less effective to give feedback on what is right or wrong, while it is considerably more effective to give what John Hattie and Helen Timperley (2007)

call "feedforward" – that is, forward-looking feedback that shows how the students should go about performing better next time.

Another research review that can be of interest is the review of scoring rubrics by Anders Jönsson and Gunilla Svingby (2007). In this review all available research about rubrics was examined, which, added together, included 75 studies. For those not yet acquainted with scoring rubrics, a rubric is an instrument for the assessment of qualitative knowledge, which partly contains criteria for what it is that is to be studied, and partly a number of standards for each criterion (for further information on assessment criteria see, for example, Jönsson, 2011; Wiggins, 1998). An interesting finding in this review was that the use of scoring rubrics could support students' learning by making the criteria and expectations clearer to them, which also facilitates feedback and self-assessment. This implies that the use of rubrics has a potential for the facilitation of that which Black and Wiliam sought after, in their review of formative assessment, that is, clear goals, feedback, and self-assessment. Since scoring rubrics also contain several levels of quality, they might be able to create more advantageous conditions for the possibility of giving forward-looking feedback, by pointing to a conceivable next level for the student.

The connection to self-assessment is particularly interesting, since there were a number of studies that showed very large effect sizes when rubrics were combined with self or peer assessment, for example a study by Gavin Brown, Kath Glasswell, and Don Harland (2004) (effect size = 1.6) and Heidi Andrade (1999b) (effect size = .99). These results are to be interpreted with some care, since these kinds of studies are relatively rare, but they do provide a hint that this might be something interesting to investigate further. This has also, to a certain extent, been done, for example in a study where the use of a rubric was combined with concrete examples of students' answers in teacher education (Jönsson, 2010).

All in all, the quantitative measures of efficiency found in studies on assessments with a formative purpose should, indeed, be interpreted with some care (see Bennett, 2009), but, at the same time, it is undeniable that there is good reason to believe that such assessments can give positive effects on students' learning, motivation,

and self-esteem (e.g. Black et al., 2003). However, since there are many different factors involved and, what is more, it is human beings that are being studied, there will never be any guarantees that corresponding results will be observed in different contexts. Furthermore, it is difficult to say exactly how the formative assessment should be carried out in order to give the optimal effect, precisely because all studies are influenced by factors other than the ones that are being studied, which means that it cannot be concluded with certainty whether or not it was the formative-assessment practice that had an effect on students' learning. However, formative assessments may have a positive effect on student learning, depending on how the assessment is designed, how it is used, how it is received by the students, etc.

Formative assessment vs. assessment for learning

The two concepts "formative assessment" and "assessment for learning" are sometimes used interchangeably, but have also been given different meanings. Already in 2004 Paul Black, Christine Harrison, Clare Lee, Bethan Marshall, and Dylan Wiliam discussed the distinction between the concepts stressing the purpose of using assessment data to adapt the teaching to students' needs:

> Assessment for learning is any assessment for which the first priority in its design and practice is to *serve the purpose* of promoting students' learning. It thus differs from assessment designed primarily to serve the purposes of accountability, or of ranking, or of certifying competence. An assessment activity can help learning if it provides information that teachers and their students can use as feedback in assessing themselves and one another and in modifying the teaching and learning activities in which they are engaged. Such assessment becomes 'formative assessment' when the *evidence is actually used* to adapt the teaching work to meet learning needs. (p. 10, emphasis added)

As is evident in the citation above (see also Wiliam, 2011), the term "formative assessment" – as used by these authors – refers to the function of the assessment, while "assessment for learning" re-

fers to the purpose of the assessment. This means that even an assessment that is not first and foremost designed with a formative purpose (such as national tests) can be used formatively, provided that the information is used to change the instruction or in other ways support student learning. As a consequence, whether an assessment is to be considered formative does not primarily depend on the assessment format or the intent, but on how the assessment information is actually used:

> Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited. (Black & Wiliam, 2009, p. 9)

In the study at hand, the terminology above will be adopted by using "assessment for learning" as an overarching concept, in order to express the formative purpose of the classroom-assessment practice. However, most of the assessments are also formative in the sense that they involve the engagement of various agents (teacher, students, or peers) in the seeking of evidence of the student's learning process in relation to goals and criteria, as well as using that information to support student learning.

A framework of assessment for learning

In a number of articles, Black and Wiliam (2006, 2009) and Wiliam (2007, 2011) search to provide a theoretical grounding for formative assessment. For instance, in the article published in 2007, the purpose is to draw together ideas developed in earlier publications in order to provide a unifying basis for the diverse practices that are said to be formative. One step was taken by going back to the early work on formative assessment, identifying the main types of activity that are inherent to effective formative-assessment practices, such as sharing success criteria with learners and involve the students in peer- and self-assessment (Black et al.,

2003; Wiliam, 2007). A second step was to propose a comprehensive framework of formative assessment (Wiliam & Thompson, 2007). In this framework, the authors build on earlier research, including meta-studies from a breadth of different subjects and school systems, in order to guide future research in this area. The framework is based on three "key processes" on learning and teaching, as outlined for instance by Royce Sadler (1989). These processes are basically to establish: (1) where the learners are in their learning, (2) where they are going, and (3) what needs to be done to help them get there. Since there may be three different agents involved in each of these processes (i.e. teacher, peer, learner), there is a total of nine (3×3) combinations of process versus agent. In the framework, however, a couple of these combinations have been merged (see Figure 1), resulting in a framework with five "strategies":

1. Clarifying and sharing learning intentions and criteria for success,
2. Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding,
3. Providing feedback that moves learners forward,
4. Activating students as instructional resources for one another, and
5. Activating students as the owners of their own learning.

As can be seen, the strategies above go beyond changing merely the assessment practice. In fact, they involve changing the whole teaching-learning environment of the classroom. For instance, the teacher has traditionally been responsible for both teaching and assessing, but according to the research on formative assessment the involvement of the learners themselves and their peers may be crucial for productive learning. In the following, each of the strategies is presented in a little more detail.

	Where the learner is going	*Where the learner is right now*	*How to get there*
Teacher	Clarifying and sharing learning intentions and criteria for success	Engineering effective classroom discussions, and other learning tasks that elicits evidence of student understanding	Providing feedback that moves learners forwards
Peer	Understanding and sharing intentions and criteria for success	Activating students as instructional resources for one another	
Learner	Understanding learning intentions and criteria for success	Activating students as owners of their own learning	

Figure 1. A framework for assessment for learning as proposed by Dylan Wiliam and Marnie Thompson (2007).

Making goals and criteria explicit and understandable

Several researchers (e.g., Good & Brophy, 2003; Natriello, 1987) have pointed to the problem that goals and criteria are often unknown to the students. As a result, students do not understand the meaning of the activities in the classroom or what is being evaluated by the assessment. Mary Alice White (1971) makes an analogy in saying that the situation for the student is like "sailing on a ship

across an unknown sea, to an unknown destination. ... The chart is neither available nor understandable to him...The daily chore, the demands, the inspections, become the reality, not the voyage, nor the destination." (ibid, p. 340).

The students' understanding of the assessment criteria seems to be an important factor behind the positive effects reported by studies where students are involved in the assessment process (Black & Wiliam, 2006; Boud, 1995; Dochy, Segers, & Sluijsmans, 1999; Stefani, 1998). In strong contrast to conventional summative assessments, where goals and success criteria may not always be shared with the students, transparent goals and assessment criteria can help students to focus their energy on what is being evaluated. By clarifying criteria and learning intentions, students have an opportunity to get a perception of different qualities in their own work and that of others. This is especially valuable for students who do not share the dominant school culture (Bourdieu, 1985) and for low-achievers (White & Frederiksen, 1998). A central intention of the strategy of making goals and criteria understandable to students is, thus, to create a common ground for the communication between teachers and students (Sadler, 1989).

Creating situations that make learning visible

The function of the strategy "making learning visible" is to reveal different aspects of students' thinking and/or understanding; information that can help teachers to rethink their instruction or to offer additional support to what they are already doing (Wiliam, 2007). By engaging students in tasks, discussions, and/or activities in which it is not the answer but the thinking behind the answer that is the primary target, different aspects of students' thinking and different perspectives are revealed, not only for the teacher but also for the students.

Another way of gathering information about what leads students to react in a certain way to instruction is by gaining awareness of students' beliefs about learning and themselves in the context of learning. These beliefs function as a filter or lens through which students process their classroom experiences. When teachers find out which beliefs students hold, part of students' behavior and re-

actions to instruction can be explained. At the same time, this information can also suggest actions needed to be taken in order to influence those of students' beliefs that are not supporting learning.

In the research literature, beliefs are considered to be an individuals' own constructions, or part of the "tacit knowledge" (Pehkonen, 2001), which are rarely explicated but involved in every learning situation. For example, under the influence of different factors from the surrounding world, such as social, economic, and cultural factors, an individual builds her/his own personal knowledge and values and her/his beliefs (Pehkonen, 1995). When an individual is confronted with new experiences or another individual's beliefs, the old beliefs may be reconsidered. In this way, new beliefs can be adopted and integrated into a larger structure of an individual's personal knowledge (i.e. the individual's beliefs system; Green, 1971).

Introducing feedback that promotes students' learning

Feedback is a central concept in assessment for learning. In their well-known article "The power of feedback", Hattie and Timperley (2007) show that feedback can be a powerful tool for improving learning. The term "feedback" is not, however, defined in a similar way across studies, sometimes making comparisons problematic. As an example, the definition used by Hattie and Timperley is quite inclusive and focus on the transfer of information: "information provided by an agent (e.g. teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding" (p. 81). Sadler (1989), on the other hand, holds that the information has to be used in order to qualify as feedback – otherwise it is only "dangling data" – and the usage should be in order to "close the gap" between current performance and the performance strived for. Consistent with the definition used by Hattie and Timperley, the term "feedback" will be used here to denote information transfer. The information provided does not, however, only indicate where the student is (and how the student is doing) in relation to standards, but also what can be done in order to improve student performance; two dimensions of

feedback that are called "feedback" and "feedforward" respectively by Hattie and Timperley.

Activating students as resources to each other

A number of research studies have demonstrated that working in collaboration on a learning task can have positive effects on students' learning (Kagan, 1992; Slavin, 1995; Webb, 2007). Such research often refers to Vygotsky's theory of development and learning (Vygotsky, 1978). According to Vygotsky, a student can reach higher levels of development with the help of an expert (teacher) or through collaboration with more knowledgeable peers. To reach higher levels of cognition, the student must also attach personal meaning and value to the activity taking place. Building on Vygotsky's theory, researchers have studied collaborative learning by shared activity, common goals, continuous communication, and co-construction of understanding through exploring each other's reasoning and views (Goos, Galbraith, & Renshaw, 2002). Interpreted in the field of mathematics education, the function of collaborative learning is to produce mutually acceptable solution methods and interpretations that require students to present and defend their ideas and to ask their peers to clarify and justify their own ideas. Thus the peers share and explore ideas with one another and this process is not only about cooperation and agreement, but also about disagreement and conflict.

Activating students as owners of their own learning

The key element in the strategy of activating students as owners of their own learning is that of self-regulation (Wiliam, 2007). As defined by Monique Boekaerts, Stan Maes, and Paul Karoly (2005), self-regulation is "a multilevel, multicomponent process that targets affect, cognition, and actions, as well as features of the environment for modulation in the service of one's goals" (p. 1078). Three broad areas are brought together in the concept of self-regulation: cognition, motivation, and behavior.

Results from empirical and theoretical research have identified several characteristics of self-regulated learners, such as that: they manage study time well, analyze more frequently and accurately,

set higher specific and proximal goals, have increased self-efficacy, and persist in spite of obstacles (De Corte, 2004). In order to become self-regulatory, learners adopt several strategies. Barry Zimmerman and Manuel Matinez-Pons (1986, 1988) have identified ten such strategies, one of them being self-evaluation. According to Areti Panaoura and George Philippou (2005) "self-evaluation refers to the subjects' appraisal of the difficulty of the various tasks and the adequacy or success of the solutions they give to the tasks." (p. 2). Furthermore, they believe that self-evaluation is not just one strategy among ten, but a first step in the process of becoming a self-regulating learner. When instruction is adopted to introduce and support students in their self-evaluation, the engagement and ownership of their learning in enhanced (Wiliam, 2007). Student engagement and ownership, in turn, have effects on their willingness, capacity (regulation of cognitive tactics and strategies), and their desire to learn.

Assessment for learning in mathematics education

Two different orientations may be identified in research on mathematical education regarding assessments with a formative purpose.

The first orientation is characterized by a focus on the teachers, for instance teachers' experiences of using and developing scoring rubrics (e.g. Lehman, 1995; McGatha & Darcy, 2010; Meier, Rich, & Cady, 2006; Saxe, Gearhart, Franke, Howard, & Crockett, 1999; Schafer, Swanson, Bene, & Newberry, 1999). Studies belonging to this category are often concerned with the training of teachers in assessment literacy (e.g. Black et al., 2003; Elawar and Corno, 1985; Even, 2005). As suggested by the abovementioned research, when it comes to assessment, many teachers are in need of extensive in-service training. Such training is needed not only in order to change teachers' ways of providing feedback (Lee, 2006), but also to change their views on mathematical learning. Examples of the latter may be to accept "non-conventional" student answers (Even, 2005); to value how students process mathematical problems, as opposed to only focusing on the final answer (Black et al., 2003), or to reconsider the importance of providing students with

high-quality comments about both strengths and ways to improve, instead of only providing numerical scores (Butler, 1988).

The second orientation in research on mathematical education is primarily directed towards students' mathematical learning. In the majority of studies, the focus is on one dimension of mathematical competency. For instance, by working with co-assessment, some studies focused on the development of students' mathematical reasoning (Lauf & Dole, 2010) or on mathematical communication (Summit & Venables, 2011). Other studies, working with peer assessment and peer feedback (Lingefjärd & Holmquist, 2005; Tanner & Jones, 1994), were concerned with students' development of mathematical modeling, mathematical proofs (Zerr & Zerr, 2011), and problem-solving skills (Carnes, Cardella, & Diefes-Dux, 2010). What is common to many of the studies included in this category is the concern with rethinking how to organize instruction. In the majority of studies, the primary interest has been students' development of mathematical skills by the use of one or several formative tools. What seems to be lacking, however, is an attention to students' perceptions of the use of formative-assessment practices to support learning, but also an attention to how to implement a systematic change in classroom-assessment practices, which encompass and affect the instruction as a whole. There are reasons to believe that such a change could affect students' development of all the dimensions included in mathematical competency.

Conclusions

The problem outlined in the introduction was the negative trend observed for Swedish students' results in mathematics, as well as their motivation to study mathematics. As suggested, part of the explanation for this trend may be found in current classroom-assessment practices in mathematics education. Therefore a change of this practice, towards a more formative-assessment practice, may influence students' learning of mathematics positively.

The research presented above lends support to the assumption that an assessment practice, with the intention of supporting students' learning (and teachers' instruction), may indeed be a power-

ful way of raising students' performances. A problem, however, when trying to establish the effectiveness of formative-assessment practices is that it is difficult to clearly define which practices are to be considered as formative. As a way to guide research in this area, Wiliam and Thompson (2007) have developed a framework for formative assessment, in which five strategies of assessment for learning are outlined. Although these strategies can be seen as indicators of a formative-assessment practice, the connections between each of the strategies still remains largely unclear. Is it, for instance, enough to implement one of the strategies (in isolation) in order to define the practice as assessment for learning, or does all strategies have to be combined? Furthermore, are some of the strategies more fundamental, and therefore more essential, than the others? Can, for instance, students become self-regulated without understanding the goals and criteria? Or what about the other way around: Can students understand goals and criteria without being self-regulating?

As becomes evident, the framework by Wiliam and Thompson may aid in identifying important dimensions of assessment for learning, which can be used for instance when implementing a formative-assessment practice in a classroom, but how these strategies interact by strengthening and interfering with each other in the classroom is still not very well known. This is also apparent in the way the strategies in the framework are supported by research evidence one by one and not as a coherent whole. It may therefore be argued that studies investigating how the strategies co-exist, as opposed to investigating each of the strategies in isolation from the others, are needed. This can be done for instance by implementing assessment for learning, with all strategies present, in a classroom and follow the classroom-assessment practice over time, investigating the interactions between changes made in assessment practices on the one hand and the teacher's and students' perceptions of the teaching-learning environment, as well as possible changes in student learning, on the other. In the next chapter, such a research design is outlined.

A MIXED-METHOD, QUASI-EXPERIMENTAL INTERVENTION STUDY

As was argued in the previous chapter, the framework by Wiliam and Thompson (2007) can be used when implementing a formative-assessment practice in a classroom. The interactions between changes made in assessment practices and the teacher's and students' perceptions of the teaching-learning environment, as well as possible changes in student learning, can then be investigated in order to find out whether – and how – assessment for learning can be used to influence student learning in a positive way. In this chapter, such a research design is outlined, while the description of the classroom intervention is given in the following chapter.

Aim of the study

The aim of this study is to introduce a formative-assessment practice in a mathematics classroom, by implementing the five strategies of the formative-assessment framework proposed by Wiliam and Thompson (2007), in order to investigate: (a) if this change in assessment practices has a positive influence on students' mathematical learning and, if this is the case, (b) which these changes are, and (c) how the teacher and students perceive these changes.

Intervention group

Control group

Figure 2. The overall organization of the different components in the study for the intervention- and control groups: (A) Self-reported questionnaire (Appendix 1), (B1) Problem-solving tasks (Appendix 2), (B2) Problem-solving tasks (Appendix 3), (C) National test in mathematics, and (D) Interviews (all students in the intervention group).

Overall research design

Since the five strategies of the formative-assessment framework are implemented in a classroom, the study may be classified as an intervention study. However, since pre- and post-tests, as well as a control group, were used in order to evaluate the effects of implementing formative-assessment practices, the research design may also be described as "quasi-experimental". Furthermore, since the quantitative data used for evaluating the effects are complemented with qualitative data about teacher and student perceptions, the study design may be classified as a mixed-method design. Taken together, therefore, the study may be called a mixed-method, quasi-experimental intervention study. The overall organization of the different components in the study is given in Figure 2.

In the intervention, both the intervention group and the control group started the first semester in upper secondary school by (A) completing a questionnaire on mathematical beliefs, followed by a mathematical test (B1). At the end of the first semester, the questionnaire (A) was answered by the students once again along with a new mathematical test (B2). After that the National test in mathematics was administrated (C). For the intervention group, individual interviews concerning the students' experiences of the intervention were added (D), which ends the first phase of the study.

The second phase comprised the whole of the second semester. In both groups the teaching went on as planned. Since it is well documented that changing beliefs takes time (Furinghetti & Pehkonen, 2002), the questionnaire (A) was again administered to both groups at the end of the second phase.

Setting and participants

An upper-secondary school situated in the south of Sweden was chosen for the study. To validate changes in students' perceptions and performance, the study was conducted over one whole year (2 semesters) and the effects were compared to those of a control group. The students were in their first year and took the basic mathematics courses (Course A and B), which occupies 86 lessons (a lesson is 80 minutes) during Year 1. The first phase (semester), comprising four months with three lessons a week, consisted of the subject matter included in the A-course: Numbers, Geometry, Functions, and Statistics. In the second phase, comprising five months, the content of the B-course was covered: Functions, Geometry, Algebra, and Probability.

Forty-five students in two classes were randomly assigned to the experiment/intervention group or control group. The intervention group had twenty-one students: ten girls and eleven boys. The control group had twenty-four students: twelve girls and twelve boys.

Two teachers were randomly assigned to the two conditions. The teacher in the intervention group was a female teacher in her twenties and recently graduated, while the teacher in the control group was a male teacher with twenty years' experience.

Ethical considerations

Students in both groups were informed about the research study at the beginning of the school year. The conditions of the two groups were made clear for the students as well as the intention of collecting data by using a problem-solving test and a questionnaire. Students were also informed that their participation was important for the study, but nonetheless voluntary, which meant that they could refuse to participate further at any time.

The confidentiality of the material, namely the students' performances in the tests and their answers in the questionnaire, was guaranteed. No-one other than the researcher would know the identity of the students. In the analysis of the material and in the case that reference would be made to any individual student's answer or solutions, a fictive name would be used.

For the students in the intervention group, a consent form was sent to their guardians. In this form they received information about the study, about the confidentiality of the material, and about the intention to only use the material only for research purposes. They were also reassured that everything would be destroyed after the dissertation. On this form, the students' guardians were given the option to give consent. For all the participating students in the intervention group, this form was signed in the affirmative.

The instruments for data collection

In order to investigate whether the formative-assessment practice had a positive effect on students' mathematical learning, data on students' mathematical performance was collected. However, as was described in the previous chapter on the strategy of "making learning visible", there are also other ways of gathering information about student learning, such as investigating students' beliefs about learning and themselves in the context of learning. Information about changes in students' beliefs can, therefore, be used as an indicator of student learning in much the same way as student performance. Other information collected is interview data on teacher and student perceptions of the teaching-learning environment. Below, the instruments for data collection will be presented

along with the arguments for choosing them. An overview of the data collected, the instruments used, and the analyses made, is given in Table 1.

Students' mathematical performance

Two types of tests were used to gather information about the level of students' mathematical performance. The first one was a problem-solving test, which was administrated as a pre- and post-test, and the second one was a National test in mathematics. The problem-solving test consisted of tasks chosen from earlier National tests in mathematics, where the tasks in the pre-test where chosen from national tests for the ninth grade in compulsory school (see Appendix 2) and the tasks in the post-test from national tests for the A-course in upper secondary school (see Appendix 3).

Table 1. An overview of the data collected, the instruments used, and the analyses made.

Data	Instrument	Main analysis
Data on student mathematical performance	Problem-solving test	Comparison between students' results before and after phase I (*t*-test)
		Comparison between intervention- and control group (ANOVA)
	National test in mathematics (Course A)	Comparison between intervention- and control group (ANOVA)
Data on students' beliefs	Epistemological beliefs questionnaire	Comparison between students' answers before and after the intervention (t-test)
	Mathematical self-concept questionnaire	
	Questionnaire on beliefs about assessment	Comparison between intervention- and control group (ANOVA)
Data on perception of teaching-learning environment	Semi-structured interviews	Qualitative content analysis

National tests in mathematics

All Swedish students take the National test in mathematics at the end of compulsory school (ninth grade) and at the end of the A-course in upper-secondary school. The two tests have much in common, but are not identical. Both tests include a set of "ordinary" short-answer items and one more comprehensive problem-solving task. The mathematical content covered by the two tests is nearly the same.

Students in the intervention group took the National test in mathematics for the ninth grade prior to the intervention, and, at the end of the first phase, the National test in mathematics for the upper secondary A-course. The test for the A-course consists of two parts. In part I the questions only require short answers and have to be solved without a calculator. In part II students are required to present both the answer and a complete solution. The last task on the test is a comprehensive problem-solving task in which the students have to apply their mathematical knowledge in a "real-world setting" and where a scoring rubric is used for assessment.

The main reason for using students' results from the National test is because all the students have to take the same test, but also because it facilitates a reliable assessment. Furthermore, since the tasks included in the (pre- and post-) problem-solving tests are chosen from earlier national tests, students' results on the problem-solving tasks can be compared to their mathematical performance in general (as measured by the national test as a whole).

The problem-solving tests

In order to get information about students' mathematical performance before and after the first phase of the intervention, a problem-solving test was used. The intention with the test was both to reveal changes in students' mathematical performance during the intervention, as well as to compare the performances of the intervention- and the control group. As was mentioned previously, the problem-solving tests consisted of three problem-solving tasks chosen from earlier versions of the National test in mathematics. According to the information given by the test developers, these tasks

are designed to make students apply their mathematical knowledge in "real-world situations".

The assessment of the problem-solving tasks was performed in agreement with the instructions for the National test, with the aid of a scoring rubric. The rubric includes criteria for assessing three different dimensions of students' problem-solving skills in mathematics: "Method and Execution", "Mathematical Reasoning", and "Presentation and Mathematical Language". The criteria for "Method and Execution" concern the degree to which a student is able to interpret and solve a mathematical problem, but also how well the student can use the appropriate mathematical methods and approaches to solve the problem. The criteria for "Mathematical Reasoning" concern, for instance, how well students are able to evaluate and reflect upon their solutions to mathematical problems. The criteria for "Presentation and Mathematical Language" concern the clarity and completeness of students' solutions, as well as their use of mathematical symbols, terminology, and conventions. By using these three dimensions, changes in students' mathematical performance can be described in a nuanced way.

When using problem-solving tests to estimate students' performance in mathematics, two important questions need to be considered: "Are problem-solving tests valid measures of students' mathematical performance?" and "How to guarantee that the tasks included in the tests are indeed problem-solving tasks?"

In the first case, there are two reasons for using students' performances on problem-solving tests as a measure of their mathematical performance. Firstly, it is largely acknowledged in research literature (see e.g., Hembree, 1992; Kilpatrick, Swafford & Findell, 2001; Lester & Lambdin, 2004; Stein, Boaler, & Silver, 2003) that in order to solve problems, students have to use several aspects of mathematical competency (such as the ability to use and understand mathematical concepts and procedures, as well as to communicate and use mathematical symbols). As a consequence, problem-solving is emphasized in Swedish school curricula as both a means and an end. Secondly, problem-solving tasks that are assessed by the use of a scoring rubric – encompassing the different dimensions of the problem-solving process – provides nuanced in-

formation, which may in turn be used to analyze and draw inferences about changes in students' mathematical performances.

In the second case, even though different researchers may use different definitions of what they consider to be a problem-solving task, there is a wide consensus in the standpoint that a "task" becomes a "problem" when the student cannot apply a ready-made method to solve the task (Stanic & Kilpatrick, 1989). For instance, if the student has to reflect on the complex and problematic relation between mathematics and "reality", and on the difficulties that can emerge when using mathematics to solve "real-world situations", then such a ready-made method is seldom available (Verschaffel, De Corte, Lasure, Van Vaerenbergh, Bogaerts, & Ratinckx, 1999). It may therefore be argued that the tasks used in the problem-solving tests can be regarded as problem-solving tasks, since the tasks are designed for making use of mathematical knowledge in "real-world situations".

Students' beliefs

The measurement of students' beliefs about mathematics was carried out with the aid of a questionnaire. This has been practiced by several researchers in the domain of mathematics (e.g. Kloosterman & Stage, 1992; Op't Eynde, De Corte, & Verschaffel, 2006; Mason & Scrivani, 2004; Steiner, 2007), which facilitated the construction of the questionnaire.

The main advantage of using a questionnaire to investigate students' beliefs about mathematics, lies in the possibility to aggregate students' answers and being able to compare them to another group. Disadvantages concern the lack of information about the underlying reasons or connections between different subgroups of beliefs. Furthermore, for a researcher aiming to describe students' mathematical beliefs, the fact that they are open to subjectivity and context sensitivity as well as being influenced by emotions and values, makes great carefulness necessary in the analysis of data in this field. Even if several researchers have used questionnaires to collect data on students' beliefs in the domain of mathematics (Kloosterman & Stage, 1992; Mason & Scrivani, 2004; Op't Eynde et al., 2006; Steiner, 2007), the use of a questionnaire to investigate be-

liefs needs to be commented on and complemented with a qualitative method.

Since beliefs are thought to occur in sets of groups, researchers often investigate several groups in the same questionnaire, such as beliefs about mathematics as a science and as a school subject, beliefs about the self, and beliefs about the context in which mathematics education occurs (Op't Eynde, De Corte & Verschaffel, 2002; Malmivuori, 2001). The same has been done in this study, which is why the questionnaire is divided into three "subscales": Epistemological beliefs (16 items), Mathematical self-concepts (8 items), and Beliefs about assessment (12 items). The subscales are presented in detail in the following.

Epistemological beliefs questionnaire
Sixteen items aim to assess students' epistemological beliefs, namely the students' views of mathematics as a discipline, the nature of knowing and the learning of mathematics, as well as the usefulness of mathematics (Muis, 2004). The items are based on the Mathematics Belief Scales used by Steiner (2007), which is a revision of the Indiana Mathematics Belief Scales (Kloosterman & Stage, 1992) and the Usefulness of Mathematics Scale (Fennema-Sherman, 1976). The items were translated and adopted to the Swedish language and ask for students' views on the importance of understanding concepts in mathematics (four items), on the usefulness of mathematics in daily life (four items), on the character of mathematical problems, that is the complexity and the time consumption (four items), and on the strategies/steps used when solving problems (four items) (see Appendix 1).

Mathematical self-concept questionnaire
Eight items assess students' self-concept in mathematics. The items chosen for inclusion in the questionnaire are based on the Mathematics Belief Scales used by Steiner (2007), who revised the Self-Description Questionnaire (Marsh, 1989). As was done with the items for epistemological beliefs, these items were translated into, and adopted to, the Swedish language. The items ask students about how easy and challenging they perceive mathematics, about

how they deal with mathematic tasks in terms of avoiding or not avoiding, and whether or not they think it is necessary to have talent in order to be able to learn mathematics (see Appendix 1).

Questionnaire on beliefs about assessment
Students' perceptions of the assessment system have been shown to influence the outcomes of the teaching-learning situation. The questionnaire thus included items on assessment in mathematics. The items are primarily based on the review presented by Struyven et al. (2003) and the questionnaire comprise items intended to measure students' views on instruments used for assessment (four items), what is meant by a "fair" assessment, (four items), and strategies students use when preparing for examinations (four items) (see Appendix 1).

Reliability issues
In order to investigate the reliability of the beliefs questionnaire, the items were answered by 166 students. All students had just started their first semester at the same upper-secondary school and were attending a mathematical A-course. Students rated the items on a 4-point Likert-type scale (ranging from 1 = totally disagree to 4 = totally agree).

The three scales of beliefs (epistemological, self-concept, and beliefs about assessment) were analyzed with factor analysis, and the reliability of each of the scales was computed. The analysis led to the removal of 12 of the 36 original items in the questionnaire. From the scale on epistemological beliefs, initially with 16 items, 6 were excluded, resulting in a Cronbach's alpha of .70. From the scale on self-concept, initially with 8 items, one item was removed (Cronbach's alpha = .77) and from the scale on beliefs about assessment in mathematics, initially with 12 items, 5 items were excluded (Cronbach's alpha = .69). According to George and Mallery (2003) Cronbach's alpha values that are greater than .7 are regarded as being good, while values greater than .6 are acceptable. Considering this, and the fact that other studies such as Mason and Scrivani (2004) and Gijbles, Segers, and Struyf (2008) have used

scales (in their questionnaires) with values below .6, the reliability found in the belief questionnaire is considered to be sufficient.

Students' and the teacher's perceptions

The semi-structured interview (Patton, 1990) is a commonly used qualitative method for data collection. The interview may be either the primary data source or it can be used in combination with other methods, such as observation, video recording, or document analysis. In this study, the interviews were used in order to gain insight into how the students perceived the intervention by allowing them to provide rich descriptions of their experiences. An advantage with interviews in this case is that the interviews gave an opportunity to explore affective as well as cognitive aspects of students' responses. Furthermore, as the subjects were young adults the interview situation allowed the interviewer to explain or help clarify questions in order to make sure the responses were useful for the study (cf. Frechtling & Sharp, 1997). On the other hand, the interviewer may at the same time influence the answers given, which limits the conclusions that can be drawn from the interview data. As a consequence, the results from the interviews must be interpreted with great care and it should be kept in mind that the interview data is a complement to the other data collected.

In the interviews an interview guide was used (see Appendix 6). The main advantage with using such a guide is that it gives the interviewer freedom to follow up and probe deeper into respondents' answers within the predetermined areas of inquiry, without being tied to a standardized interview protocol that has to be the same for all respondents. At the same time, the interview guide may give enough structure for not losing focus, leaving out important issues, or lack time to cover all areas of interest. Furthermore, the interview guide to some extent guarantee that similar information is obtained from each respondent, thereby making interviews with multiple subjects (as was the case with the students in this study) becomes more systematic and comprehensive (Flick, 2002).

During the interviews, the students were encouraged to reflect on their experiences of the intervention and each interview lasted for

approximately fifteen minutes. A mini-disc was used to record the interviews.

For the teacher an interview was also prepared. However the teacher preferred to get the questions, think them over and deliver her answers in the form of a written essay.

Analyses

Two types of analyses were performed: quantitative and qualitative.

Quantitative analyses

When analyzing data on students' mathematical performance, as well as students' mathematics-related beliefs, ANOVA (i.e. "analysis of variance") and t-tests were most frequently used. The ANOVA is a statistical test used to estimate whether or not there is a statistically significant difference in means between two groups. The t-test, on the other hand, may be used to estimate whether there is a statistically significant difference between means measured at different times in the same group. An exploratory analysis was carried out before using the ANOVA and the t-test in order to examine the necessary prerequisites for using these tests. For ANOVA, these requirements are "normality", "homogeneity of variance", and "independence" (Norusis, 2008). For the t-test it is mainly "normality".

Normality was tested with the Shapiro-Wilks Test. Results from this test indicate that the scores on the problem-solving test, as well as the scores on the beliefs questionnaire, were normally distributed in both groups. The Levene Test was then used to investigate the homogeneity of variance between the two groups. The test results indicate homogeneity of variance for students' scores on the problem-solving test and for their scores on the scales "Epistemological beliefs" and "Self-concept" in the beliefs questionnaire. The scores on the scale "Beliefs about assessment" did not, however, meet the condition of homogeneity ($p = .025 < .05$), which means that a non-parametric test was used when these scores are analyzed. With regard to independent random sampling (i.e. the requirement of independence), the two groups were two heterogene-

ous classes that were not taking the same course and the students were distributed randomly. Taken together, the results from the abovementioned tests indicates that ANOVA can be used in order to compare the two groups (i.e. the intervention group and the control group), except for the "Beliefs about assessment" scale in the beliefs questionnaire. In this particular case, the Mann-Whitney test – a non-parametrical test – was used instead. Furthermore, the *t*-test was used when comparison were made within the same group at different occasions (i.e. pre- and post-tests).

Qualitative analyses

The data on students' perceptions of the teaching-learning environment was analyzed with a focus on students' experiences in relation to the different aspects of the intervention, such as their perception of learning and understanding mathematics, of the assessment system, of the emphasis on problem-solving, and of their mathematical self-concepts and beliefs. In this analysis, the aim was to identify and describe a breadth of different perceptions (Marton & Booth, 1997), not the number of students expressing each particular perception or the relative strength of the different aspects of the intervention.

The qualitative analysis was performed in two steps. In the first step, the transcribed interviews were arranged according to themes that originated from the framework of assessment for learning. The themes were: the use of a scoring rubric, the use of peer assessment and feedback, working with problem-solving, and students' mathematics-related beliefs. This is a theory-driven analytic approach (Boyatzis, 1998), meaning that the themes are generated – driven by – the theoretical framework and research relating to each of the strategies in the framework. The first three themes are directly related to the instruments and methods used in the intervention. The fourth theme relates to students' general perceptions of the intervention and its effects on their views about mathematical teaching and learning.

In a second step, students' answers were read a number of times and categorized by the researcher. By using a meaning concentration method (Kvale, 1997), a number of sub-themes were identi-

fied. The process was iterative in order to verify that meaning associated with the main theme was not misinterpreted in the analytic process. Furthermore, the original data was continuously consulted to make sure that quotations from the students were not misinterpreted when taken from the context of the interview.

THE CLASSROOM INTERVENTION

In this chapter, the classroom intervention will be presented. The presentation is organized according to the formative-assessment strategies previously introduced, although in practice there is considerable overlap between these strategies. The intervention as a whole was designed by the researcher in collaboration with the teacher of the intervention group.

Making goals and criteria explicit and understandable

Several methods may be used in order to clarify goals and criteria. For example, when students are asked to assess other students' work and motivate their assessment by identifying qualities they also get an opportunity to explore notions of quality and to reflect upon them. Students can also be engaged in the design of their own tests, or in creating posters of key words regarding the purpose, the description and the evaluation of the subject matter.

An assessment instrument that has received a lot of attention lately is the so called "scoring rubric". The rubric is an assessment instrument that contains assessment criteria as well as standards specifying levels of attainment for each criterion (Jönsson, 2008; Perlman, 2003; Wiggins, 1998). Although the most common use of rubrics might be for enhancing reliability in summative assessments, for instance when teachers assess national tests in Sweden (see e.g. Kjellström & Pettersson, 2005), the rubric can also be used for formative purposes (see Jönsson & Svingby, 2007). Research literature (Jönsson, 2008, Reddy & Andrade, 2010) highlights several positive effects of using rubrics. For instance, by clari-

fying goals and criteria, the use of rubrics can be used to enhance transparency. These research results, together with the fact that rubrics are already commonly used in Swedish schools (and therefore to some extent familiar to the students), provide the basis for choosing a rubric as the means for "making goals and criteria explicit and understandable".

Needless to say, the mere fact that a rubric is used does not guarantee any positive effects on student learning. The question, therefore, is under which conditions the use of rubrics is more likely to produce positive effects. The reviews on research by Jönsson and Svingby (2007) and Reddy and Andrade (2010) give some guidance, but also clearly indicate that there is no straightforward answer to the question of the effectiveness of the use of rubrics.

As mentioned above, the existence of a rubric does not automatically improve students' performance. Instead, whether the students are trained to use the rubric or not, seems to have an impact on students' work. For instance, in a study from middle school (Andrade, 2001), but also in a study conducted in higher education (Green & Bowser, 2006), a rubric was provided to students before performing a task (writing an essay and a literature review respectively). Results showed no significant difference in either of the two studies between the treatment group using the rubric on the one hand, and the control group working without a rubric on the other. However, in the study by Andrade (2001), data collected from a written questionnaire revealed that students in the treatment group could refer to a greater variety of academic criteria relevant for writing. The researcher therefore concluded that by using the rubric students got a broadened conception of effective writing, although this was not associated with improved performance.

This finding implies that students may indeed be helped by using a rubric, but that this may not necessarily manifests itself in improved performance during short-time interventions. The time students have for learning to use the rubric, therefore, seems to be an important factor to consider when trying to improve student performance (Jönsson, 2008). For instance, when Andrade (1999a) used a rubric and self-assessment in order to improve students' writing performance, less than 40 minutes were used for introduc-

ing the rubric. In this study, the results were not positive for all students involved. On the other hand, when Brown et al. (2003) implemented a training program on writing an average of approximately 27 hours was allocated to using the rubric. The results from this study showed a large improvement in students' performance (effect size = 1.6).

Besides time allocated to use the rubric, another important factor to consider is rubric design, such as whether the rubric is "generic" or "task-specific". While generic rubrics aim at general skills and can span over a whole subject area, task-specific rubrics are directed towards a specific task. According to Wiliam (2007), students have less problems interpreting and applying criteria and standards in the case of task-specific rubrics. On the other hand, the criteria and standards for a specific task cannot easily be transferred and task-specific rubrics may therefore be of limited value when supporting student progression across tasks. This means that generic rubrics may be more appropriate for a formative-assessment practice, since they "focus on qualities that transcend the immediate task" (p.1077).

Another factor, which may influence the effects of using rubrics, is the provision of examples of student work. Such "examplars" may help overcome some of the difficulties with understanding and applying the criteria and standards in the rubric and the use of exemplars can therefore contribute to a common understanding of the concepts of quality used for assessing student work (Sadler, 1989). Furthermore, when students in a study by Paul Orsmond, Stephen Merry, and Kevin Reiling (2002) were allowed to use exemplars, as well as being engaged in the construction of criteria, an increased understanding of criteria and standards could be observed.

The last factor potentially influencing the effects of using rubrics is the combination of rubrics and peer-, self- or co-assessment. This combination is recommended by several researchers (e.g. Jönsson, 2008; Orsmond et al., 2002; Reitmeier, Svendsen, & Vrchota, 2004) since it may help students not only assessing where they currently are in relation to instructional goals, but also how they can move on and improve. By creating opportunities for students to

discuss their work, and/or that of other students, with their peers, they are provided with a forum for negotiating what counts as quality and also for reflecting on possible actions for improvement (William, 2007). Since the use of peer-, self- or co-assessment relates to other strategies in the formative-assessment framework, this will be further discussed later on.

To summarize, several methods may be used to make goals and criteria explicit and understandable, such as students participating in the creation of tests or involving students in formulating assessment criteria. In this study, scoring rubrics has been chosen as the primary tool for this strategy. In order to promote student learning through the use of rubrics, time need to be allocated so that students are thoroughly introduced to the criteria and standards in the rubric. Furthermore, although task-specific rubrics may be easier for students to understand, generic rubrics may actually be more appropriate for formative-assessment practices, since they can be used at several occasions with different tasks addressing the same skills. Also, student understanding of criteria and standards may be aided by the use of exemplars, or by the use of self-, peer-, or co-assessment, in combination with the rubric.

Making goals and criteria explicit and understandable in the classroom

In the intervention, the scoring rubric (see Appendix 7) was introduced by using the problem shown in Figure 3. Students were asked to solve the problem in groups. Then the teacher collected the solution from each group and the differences between the solutions were analyzed during a whole-class discussion. The rubric was introduced after the discussion and each assessment criterion was discussed. Then the students were asked to assess the solutions by identifying the level of quality for each criterion in the rubric. For each criterion that was discussed, the teacher asked the students to justify their judgments.

The length of a rectangle increases by 10 % and the width decreases by 10 %.

One of the following statements is true. Find out which one it is. Motivate your choice with the help of calculations and/or diagrams.
- The area does not change.
- If the area becomes smaller or larger depends on the original lengths of the sides.
- The area will always become smaller.
- The area will always become larger.

Figure 3. Problem used when introducing the scoring rubric.

Following this introduction, the rubric was used systematically in the day-to-day class-work in order to create a shared language for the communication between students and teacher. The Students had the rubric lying on their desks and the teacher made references to it when different tasks were discussed, both in whole-class discussions and in discussions with individual students. When, for instance, students worked by themselves in their exercise books, they could call for the teacher's attention in order to discuss the assessment of a problem. They could also ask the teacher to help them to further develop their solution in order to reach a higher level according to a certain criterion.

Creating situations that make learning visible
At the heart of the strategy of making learning visible, is the understanding of where students are in their learning. Learning is, however, nothing that can be observed directly. Instead, what have to be sought for are observable indicators (or signs) of learning. In order to find such signs of learning, it is often recommended to use challenging tasks, which do not only call for a correct answer, but require students to explicate their thinking. As a consequence, problem-solving tasks are often used in the field of mathematics education in order to "make learning visible" and researchers have since long advocated methods of teaching mathematics that focus

on problem solving and discovery learning (Brownell, 1942; Polya, 1957). Recent research also shows that mathematics teaching by problem solving may enhance students' understanding, motivation, and meaning making in mathematics. Furthermore, since students have to connect, extend, and elaborate their prior knowledge when working with a mathematical problem, the problem-solving process offers a possibility to deepen student understanding of mathematics (Lester & Lambdin, 2004).

Schoenfeld (1985) has done substantial research on students' behaviors when working with mathematical problem solving, and he claims that student mathematical behavior is determined by four factors: resources, heuristics, control and beliefs. Resources stand for what De Corte (2004) calls "domain-specific knowledge", such as theorems, definitions, proofs, and algorithms. Heuristics consists of rules of thumbs, strategies, and techniques used to solve a problem. Control is about managing and employing resources and heuristics, which in the terminology of De Corte can be referred to as "cognitive self-regulation". A student's system of beliefs consists of beliefs about oneself and about mathematics and mathematical learning. Schoenfeld (1985) demonstrated that if students possess beliefs that are "anti-mathematical", they can fail to be successful problem-solvers even if they have the necessary resources and the appropriate heuristics and control. Schoenfeld (1992) also added an additional category, called "practice", to his original classification. This category concerns students' enculturation into the mathematical (community of) practice and refers to students' habits and ways to think about mathematics.

Schoenfeld's clarification of the elements included in problem-solving activities has been used by several researchers when trying to explain the effects of working with problem solving. Three categories of focus can be distinguished: heuristics, systems of beliefs, and practices. In the first category, a focus on heuristics has been shown to improve students' achievement in mathematics, as well as their self-concept and metacognition (Barak & Mesika, 2007; Hohn & Frey, 2002; Koichu et al., 2007). For instance, in an intervention directed towards developing middle-school students' "heuristics literacy" Boris Koichu, Abraham Berman, and Michael

Moore (2007) found that ¬– besides positive effects on students' achievement, especially for low-achievers – students developed an increased confidence in mathematics. The heuristic-orientated activities gave low-achievers more opportunities to participate in whole-class discussions, which enhanced their self-confidence. They also made more attempts to produce own solutions to mathematical problems, instead of just trying to understand solutions produced by others.

Studies included in the second category, concerning system of beliefs, show that an instruction focusing on problem-solving activities tends to affect students' beliefs systems, while at the same time students' beliefs systems influence how the students approach problem-solving tasks. For instance, if a student believes that a problem that is not solved within five minutes cannot be solved at all, or that there is only one correct way to solve each problem, then the use of mathematical problem solving as a way to elicit students' understanding may become problematic. On the other hand, several researchers have shown that an instruction focusing on problem-solving activities may have positive effects on students' learning, but also on their beliefs about mathematics and problem solving (Higgins, 1997; Mason & Scrivani, 2004; Verschaffel et al., 1999).

As an example, Terry Wood and Patricia Sellers (1997) performed a longitudinal study where they analyzed students' achievements and mathematical-related beliefs in three groups of students in elementary school. One group received textbook instruction, while the second and the third group received instruction focusing on problem-solving activities during one and two years respectively. In this study two types of achievement tests were used. The first one was a standardized achievement test (i.e. a test designed to measure an individual's level of knowledge in a particular area) and the other was an arithmetic test designed to measure students' conceptual understanding of arithmetic. In addition, a personal goal- and beliefs questionnaire was administered in order to measure students' beliefs and motivation for learning mathematics. The results show that the students experiencing the two-year instruction focusing on problem-solving activities performed signifi-

cantly higher at both of the achievement tests as compared to the other groups. These students were also shown to hold different beliefs about mathematics, as compared to the group that had received textbook instruction. Furthermore, the students who experienced the two-year instruction focusing on problem-solving activities were more motivated to understand than to compete with others. These students recognized the importance of persistence and the value of collaborating with others in order to understand. They also believed that it was important to find their own way to solve problems and not only to replicate the teacher's method. These effects remained even after two years, when the students had returned to a textbook-centered instruction. Interestingly, the same results were not observed for the students with only one-year experience of instruction focusing on problem-solving activities. This study therefore indicates that instruction focusing on problem-solving activities may have positive effects on student achievement, as well as on students' beliefs, but that it may take time for these changes to occur.

The third category includes studies concerned with the arrangement of the teaching-learning context (Dawkins, 2009; Hershkowitz & Schwarz, 1999; Tatsis & Koleza, 2008; Yackel & Cobb 1996). These studies have shown that by directing attention to students' habits, approaches, and socialization (or "socio-mathematical norms") when working with mathematical problem-solving, may have an influence on students' performances and beliefs. Socio-mathematical norms encompass an understanding of what counts as mathematically different (such as identifying which of several presented solutions to a problem are to be considered as different solutions), mathematically sophisticated, mathematically efficient, and what constitutes an acceptable explanation.

As an example, Yackel, Rasmussen, and King (2000) used data from a classroom experiment with high school students in order to investigate how socio-mathematical norms were constructed and how they came into play when the students worked with problem-solving activities. In the experiment, students worked with problem-solving tasks in small groups, followed by whole-class discussion. The primary focus of these activities was students' explana-

tions, and the justifications of their thinking, to peers and to the teacher (cf. the strategy of using students as resources for learning). The results from the study showed that the students developed both sophisticated conceptual understandings of the subject matter (i.e. differential equations) and important mathematical skills. During the experiment students had to write electronic journals every week. Apart from revealing students' thinking, these journals also provided information about changes in students' beliefs. This information shows that the socio-mathematical norms are connected to meaningful learning, to student's beliefs about their own (and other students') role in the classroom, and to students' beliefs about the nature of mathematical activities.

The focus on socio-mathematical norms is associated with a view on learning, which presumes that students can advance in their learning by working in social contact with others (cf. Vygotsky's concept "Zone of Proximal Development", Vygotsky, 1978). Such an example is working with problem-solving in group, activities that have also been shown to create favorable conditions for student learning, as well as positively influencing students' beliefs about the nature of mathematical activity (Dahl, 2004). For example, in the study of Yackel et al. (2000) a key feature was student collaboration during problem-solving activities, where the frame of collaboration and dialogue was dictated by the socio-mathematical norms. The positive effects of working in small groups, or in pairs of students, are further confirmed by a meta-analysis from mathematics in compulsory school (Slavin, Lake, & Groff, 2009).

To summarize, a common method for "making learning visible" is using open questions and problem-solving tasks, which require students to explicate their thinking. Studies in the field of mathematical education also support the use of problem-solving activities as a means to improve students' achievement. In order to support student learning through problem-solving activities, instruction may focus on heuristics, beliefs and/or socio-mathematical norms. Furthermore, working with problem-solving activities in small groups has also been shown to create favorable conditions for student learning, as well as to influence students' beliefs.

Creating situations that make learning visible in the classroom

Since problem-solving activities are complex, and involve the use of different mathematical concepts and strategies, it is important that the concepts involved are introduced to the students, but also that different contexts for their application are presented, prior to the problem-solving activities. In line with this, the important concepts of every part of the course were introduced by the teacher using a historical perspective of how the concept had developed. For example, the teacher may have started by presenting the current needs of society at the particular time in history and the driving forces for the development of concepts such as numbers, percentage, or algebra. This historical picture was then complemented by adding information about how the concept is used today.

When the concepts had been introduced, students were given time to work in their exercise books using the criteria in the scoring rubric. Meanwhile, the teacher walked around in the classroom, observing, discussing, and helping the students. During this time she collected information, which she later used in order to clarify possible misunderstandings. When the teacher judged the students ready to work with a larger problem-solving task, they were given a group assignment.

The group assignments consisted of carefully designed realistic, open-ended, and complex problems, similar to the problem-solving tasks used in the National test in mathematics in Sweden. Figure 4 shows an example of a problem-solving task used as a group assignment. The first (a) and second (b) questions can be found in similar tasks in the textbook and they require only "imitative reasoning" (Boesen, Lithner & Palm, 2010), which means that these questions can be answered by using algorithms without having an understanding of the concepts involved. The main reason for using such routine questions is to introduce the students to the task. These questions may also be used as a scaffolding structure, guiding the students in how to solve the non-routine questions. In this way, students who are low-achieving or unconfident can still participate when working with the assignment. The third (c) and fourth (d) questions are not found in the textbook and in order to answer

these questions students have to use "creative reasoning", which means that they have to reflect upon the interpretation and meaning of the mathematical results and apply them to a real-life situation.

> Children's sleep needs
>
> A child's need for sleep can be approximately calculated with the formula $S = 5 - \frac{n}{2}$ where S is the number of hours of sleep per day and n is the age of the child in years.
>
> a) Anton is 4 years old.
> How many hours of sleep does he need according to the formula? *Only the answer is needed.*
> b) With your starting point in the formula draw a diagram that can be used in order to derive a child's need for sleep.
> c) Within which age range can the formula apply? Motivate your choice.
> d) Describe in everyday language what is meant by the formula.

Figure 4. A problem-solving task used as a group assignment.

Introducing feedback that promotes student learning

In order to be effective, feedback should first and foremost help students answer the following questions (Wiliam & Thompson, 2007):

- Where am I going?
- How am I going?
- Where to next?

By concentrating on reducing the discrepancy between what is understood and what is aimed to be understood, this type of feedback has been shown not only to enhance understanding, but also to increase students' effort, motivation, and engagement (Hattie & Timperley, 2007).

Even if the three questions above may be considered fundamental for formative feedback (i.e. feedback that is used to support student learning), this is not the only factor influencing the efficiency of feedback for student learning. Other important factors identified in the research literature are the "focus of feedback", the

timing, the "context of delivery", and the "processing of feedback". These factors are explained further below.

The focus of feedback

In their meta-analysis on feedback research, Hattie and Timperley (2007) directed a special attention to what they referred to as the "focus of feedback", which is either at the task-, process-, self-regulation, or self-level. Different effects on student performance are observed, depending on which level the feedback is focused on, but at the same time some levels may be more appropriately used in different situations.

Feedback at the task level is about how well a task is being performed, for instance whether an answer is correct or incorrect. According to Hattie and Timperley (2007), 90 percent of teacher questions in the classroom can be classified as asking about this type of right-or-wrong information. In an earlier meta-analysis, Avraham Kluger and Angelo DeNisi (1996) found that feedback is more effective when it provides information on correct rather than incorrect responses (effect size = .43 and .25 respectively). However, feedback at task level seems to be most powerful when it provides students with information on misunderstandings (Hattie & Timperley, 2007). This can help them to generalize and use the information when confronted with other tasks requiring the same understanding (Thompson, 1998).

In her review about formative feedback, Valerie Shute (2008) offers a number of guidelines for the design of task-level feedback. In order to enhance learning, feedback should be focused on the what, how, and why of a problem, rather than on the correctness of the answer. Furthermore, the feedback should be presented in manageable units and formulated "as simple as possible", which means that the feedback should not offer too much detailed and specific information. Such detailed and specific information may instead prevent students' own thinking and their learning (cf. Jönsson, 2013).

Feedback at the process level is about the processes underlying tasks (Hattie & Timperley, 2007) and is primarily directed towards students' strategies to detect errors or overcome impediments.

When an error is detected, the student needs to reconsider the strategy adopted and feedback at this level may help students to provide themselves with feedback (Hattie & Timperley, 2007). Feedback at the process level appears to be more effective than at the task level for enhancing deeper learning.

Feedback at the self-regulation level focuses on the way students monitor, direct, and regulate actions towards the learning goals. This means that feedback at this level is largely aimed at students' meta-cognitive skills, so that students may learn to assess their progress and regulate their learning in relation to goals and criteria (Hattie & Timperley, 2007). According to Deborah Butler and Philip Winne (1995), this may be one of the important features distinguishing effective learners, who are able to formulate internal feedback while they are engaged in academic tasks, from less effective learners, who depend more on external feedback from teachers or peers.

Unlike the three levels of feedback discussed above, feedback that is directed towards the self-level (such as students' personal characteristics and praise) rarely seems to be effective for student learning (Hattie & Timperley, 2007). One explanation for this is that feedback at this level do not address the three essential feedback questions (see above) and does therefore not contain enough relevant information to support student learning. Another explanation is that feedback at this level tends to direct students' attention away from the task, sometimes discouraging students and threatening their self-esteem. Feedback at this level, including praise, should therefore be used judiciously or not at all (Shute, 2008).

The timing of feedback

According to Wiliam (2007), the timing of feedback is crucial. For instance, feedback that is given before students have had a chance to work and think about a task may actually inhibit learning. Hattie and Timperley (2007) make a distinction between feedback at the task- and process level in relation to timing. In the first case, immediate feedback may be beneficial while delayed feedback may be more appropriate at the process level. This may, however, also be affected by the difficulty of the task, since effect sizes for de-

layed feedback varies markedly between easy (-.06) and difficult tasks (1.17). Still, whether a task is easy or difficult ultimately depends on the proficiency of the student and Shute (2008) suggests that the effects of timing are associated with student capability. If the student is a novice and the task is beyond her capability, then immediate feedback may be most beneficial, but if the task is perceived as simple and within the student's capability, then delayed feedback may be more appropriate. Shute therefore recommends delayed feedback for high-achievers and immediate feedback to low-achievers. She also recommends a facilitative use of feedback for high-achievers, by using hints, prompts, and cues, since they benefit from feedback that challenges them. For low-achievers, on the other hand, she recommends more immediate and explicit feedback, since they need more support and guidance.

The context of delivery

Hattie and Timperley (2007), but also Wiliam (2007), stress the importance of considering the classroom context in relation to feedback. If, for instance, error and disconfirmation are not considered as part of the learning process, then the feedback at any level may not be welcomed and used by the students. If students are responding only when they are sure that they can answer correctly, then the learning opportunity may be lost.

Wiliam (2007) further argues that even if the feedback delivered does not provide answers to the three feedback questions, it may still have a formative effect if there are established classroom norms that support the use of feedback information. For example, if the teacher asks a student to "give more detail", this could be considered formative practice in a classroom with certain norms established, while it may not be formative practice in other classrooms.

The processing of feedback

Already in 1989, Royce Sadler wrote that to be able to improve their performance, students have to possess some strategies for using the information provided by the feedback. Research on students' use of feedback, however, indicates that many students do

not use the feedback they receive and that one important explanation for this is that students lack such strategies (Jönsson, 2013). Consequently, Clara Lee (2006) suggests that students must be provided with opportunities to act on the information they get from feedback. She also argues that teachers should demand a response from the student and clearly indicate what the student must do to respond. Shute (2008) also suggests that teachers deliver partial feedback and ask the students to complete the task with the help of the feedback received, or deliver the whole feedback and ask students to perform another similar task.

In order to support students in developing strategies necessary for the processing of feedback, some researchers (e.g. Dochy et al., 1999; Gipps, Hargreaves & McCallum, 2001) argue that students need to be involved in the assessment process, for instance by engaging them in activities such as peer-, and co-assessment, which will be further discussed below.

To summarize, feedback may be delivered by different agents, such as peers or the teacher, and in different ways, such as written or oral. In order to be effective for student learning, feedback needs to provide information about where the students are going, how they are doing, and how to move on. Furthermore, feedback should preferably be directed at the task-, process-, or self-regulation level, but not at the self-level. The timing of feedback (i.e. immediate or delayed) seems to be important for supporting student learning, but since this factor interacts with several other factors, such as the focus of feedback, task difficulty, and student proficiency, it is difficult to give general recommendations. However, immediate feedback may be more beneficial when focusing on task level and for difficult tasks, while delayed feedback may be more appropriate for the process level and for tasks perceived as easy by the students. Perhaps self-evident, feedback needs to be integrated in instruction and aligned with the established classroom norms. And lastly, teachers need to teach students strategies for processing the feedback they receive, so that students are not left to figure this out on their own.

Providing feedback that promotes learning in the classroom

In order to provide students with formative feedback, the teacher in the intervention group had to collect information about student learning. This was done continuously during instruction, but a rich source of information was when students worked with problem-solving in groups. During these activities, the teacher gained information about students' understanding of the subject matter and the concepts involved, the quality of students' thinking, and also, how well the students could apply different concepts in the context of specific problems. The teacher then used this information when giving feedback to the groups. The group solutions were not only discussed from a perspective of strengths and weaknesses, but also from the perspective of alternative solutions. To communicate where the students were in relation to goals and criteria, and what they needed to do in order to improve, the teacher used the rubric. In this way the discussion was structured around different aspects of the problem-solving process and different levels of quality.

In order to further integrate the use of formative feedback into instruction, a set of "mini-rubrics" was used in the assessment of individual written tests (see Figure 5). In this example, the mini-rubric shows for example that the aspects of "Reason about" and "Give Account" are assessed in this task (the white parts). Moreover, it shows that the level of difficulty for these aspects is not higher than level 2 (i.e. V = well done). Using the mini-rubrics was a way of moving away from simple numeric scores, towards assessing different aspects in a qualitative manner. Instead of scoring the task, the teacher marked if the student's solution displayed a reasoning corresponding to the standard for the first level (G = passed) or the second level (V = well done) in the mini-rubric. In this way the student was given a more differentiated assessment about the strengths in his/her solution to the problem and what needed to be improved.

Suppose that a square of side *a* cm is cut along the diagonals and that three of the parts are subsequently put together again to form a pyramid. Draw a diagram and determine the height of the pyramid.

G1	V1 V5	M1	*Solv problems*
G2	V2	M3	*Reason about*
G3	V4	M1	*Give account*
G4	V3	M2 M4	*Interpret*

Figure 5. An example of a task assessed with a "mini rubric".

The results from the group assignments, from the individual written assignments, and from the pair-tests were collected by the teacher and used as a basis for formative feedback. This feedback was delivered to the students on two occasions before grading. Below, two examples of how the teachers in the study documented students' results are given. One example (Figure 6) is from a student that participated in the intervention group and the other (Figure 7) is from a student that participated in the control group. The samples show that in the first case, the documentation provides information about the student's strengths and what she/he needs to improve. The second case shows a total score for each test and a grade.

It should also be noted that the first example of documentation not only provides information about the individual student's performance, but also informs the teacher about the success of her teaching, for instance which parts of the curriculum that the students had problems with and therefore may need further attention. This information was used in the intervention group to direct and design the instruction to further support students' learning. The other way to document student achievement does not provide nuanced information to the teacher about the different competences (included in the curricula) that students have developed (or not developed) during the course. Basically, the only information that the teacher (and the students) get is the score on each test.

Name	Criteria	Group work Func.	Max	Test 1 Alg. Func.	Max	Test 2 Geo. Prob.	Max
Solve problems	G1	1	1	7	7	8	8
	V1	1	1	6	6	6	8
	V5	0	1	3	3	4	5
	M1	0	1	1	1	2	4
Reason about	G2	1	1	7	7	8	8
	V2	1	1	6	7	6	8
	M3	1	1	2	3	2	4
Give account	G3	1	1	7	7	8	8
	V4	0	1	5	7	7	8
	M1	0	1	1	1	2	4
Interpret	G4	1	1	7	7	8	8
	V3	1	1	6	7	6	8
	M2	1	1	2	3	1	2
	M4	0	0	2	3	0	1

Abbreviations: Max. = Maximum score;; Alg. = Algebra; Func. = Functions;; Geo. = Geometry; Prob. = Probability;

Figure 6. Documentation of a student's result typical for the intervention group.

Name:	Test 1 Algebra	Test 2 Functions	Test 3 Geometry	Test 4 Statistics	Grade
	19/1	11/1	10/3	9/2*	G
	20/G	12/G	13/VG	11*/G	

Figure 7. Documentation of a student's result typical for the control group.

Activating students as resources for each other

Translated into the terminology of assessment for learning, the activities taking place when students act as resources for each other, are for example peer assessment and peer feedback. Peer assessment is defined by Topping (2009, p. 20) as "an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners". Peer-assessment activities may involve different student constellations, such as working in pairs or groups, and a variety of student performances, such as oral presentations, portfolios, tests, or group assignments. Peer-assessment activities may also be performed in different subjects or in different areas of the curriculum.

Peer feedback may be considered an integral part of peer assessment, in the same way as teacher feedback is associated with formative-assessment practices: "Peer assessment is an educational arrangement where students judge a peer's performance quantitatively, by providing a peer with scores or grades, and/or qualitatively, by providing the peer with written or oral feedback" (Topping, 1998, p. 266). Unlike teacher feedback, peer feedback can be plentiful (since there are more students than teachers in a classroom), immediate, and individualized (Hattie, 2009).

In their review on self-, peer-, and co-assessment in higher education, Filip Dochy, Mien Segers, and Dominique Sluijsmans (1999) conclude that peer assessment can be a valuable tool for student learning. The gains of using peer assessment are associated with increased time on task and practice, identification of knowledge gaps, and greater metacognitive awareness. Dochy et al. make no distinction between peer assessment and peer feedback, but researchers focusing particularly on peer feedback (e.g. Cho & MacArthur, 2010; Liu & Carless, 2006; Nicol & MacFarlane-Dick, 2006) acknowledge that when students engage in peer-feedback activities their learning is enhanced in several ways. For example, according to David Nicol and Debra MacFarlane-Dick (2006), students may sometimes be more able than the teacher to explain in a language that is more accessible to their peers; during peer discussions alternative strategies and ways of thinking are explicated for

the students, which can help students when revising their own work; through revising the work of others, students may develop skills that aid them in the assessment of their own work. Nicol (2010) goes even further in this reasoning and argues that while engaged in peer-feedback activities, students have to process and re-process feedback from different sources, which provides multiple perspectives and invokes multiple opportunities for scaffolding. He further suggests that is not only the receiving of feedback, but the formulation of a response that is most beneficial for student learning, since the construction of feedback amplifies the level of students' engagement. In order to formulate a response, students have to recognize criteria of quality in an assignment and they "learn that quality does not come in a pre-defined form, rather there is a spectrum of possibilities" (p. 510). By learning to give feedback, as well as seek and connect feedback to their work, students may at the same time improve their skills in perceiving and utilizing teacher feedback.

The effects of peer assessment on students learning have been further investigated in a review by Marjo Van Zundert, Dominique Sluijsman, and Jeroen Van Merrienboer (2010). These authors provide an overview of how conditions, methods, and outcomes of peer assessment are related. Four categories of outcome variables are identified: (1) psychometric qualities of peer assessment, which relate to validity and reliability of peer marking; (2) domain-specific skills, which focus on the learning gains from peer assessment; (3) peer-assessment skills, which refer to the quality of students' feedback; (4) students' attitudes toward peer assessment, which refer to students' confidence in assessing their peers and the perceived benefits from peer assessment. According to Van Zundert et al. (2010), the first, third, and forth category of outcomes can be improved by training and experience, while the second category can be improved by revising tasks based on peer feedback, working in small groups, sufficient time on task and a willingness to follow instructional guidelines.

It is interesting to note that the first outcome identified above (i.e. the validity and reliability of peer marking) is mentioned in a number of studies on peer assessment and peer feedback. As an ex-

ample, Dochy et al. (1999) identifies the accuracy of peer assessment as a problem. Of course, if the peer-assessment results are to be used for "high-stakes" decisions, then the accuracy of this assessment would be important. This problem of accuracy, however, is basically rooted in a notion of peer assessment used for summative purposes, while it is of limited interest in a context of assessment for learning. On the contrary, in such a context it is more important to elicit evidence about students learning, than to sacrifice this for the sake of reliability (Bennett, 2011).

As an example of research relating to the second outcome identified above (i.e. the gains of domain-specific skills), Sarah Gielen, Ellen Peeters, Filip Dochy, Patrick Onghena, and Katrien Struyven, (2010) investigated the relationship between domain-specific skills and different types of peer feedback in writing. As a starting point, they used a synthesis of seven peer feedback key characteristics identified by Minjeong Kim (2005), Dominique Sluijsmans, Saskia Brand-Gruwel, and Jeroen Van Merrienboer (2002), and Miao Yang, Richard Badger, and Zhen Yu (2006). The participants in this quasi-experimental study were 43 seventh-grade students from secondary school and the intervention took place during two semesters. The students received three assignments in a course on the Dutch language, where each assignment had a pretest (draft) and posttest (final text). In between, the students received peer feedback formulated through the use of a scoring rubric. This feedback was then analyzed according to the following criteria: Appropriateness, Specificity, Justification, Suggestion, and Formulation. Results show that feedback including justification significantly improved performance, but only for student with low performance on the pre-test. Feedback complying with the other criteria showed no significant effect on students' performance.

The accuracy of students' feedback did have a positive impact on performance, but analyses indicate (once again) that is more important for peer feedback to contain a justification than to be accurate. Instead, peer feedback that is not correct may have a great pedagogical value, since the receiver has to evaluate the feedback; a process that requires the student to actively engage with the feedback (Zerr & Zerr, 2011).

That peer feedback can indeed make a difference for student performance is further supported by Kwangsu Cho and Christian Schunn (2007). In their study they compared teacher feedback with peer feedback, where the peer feedback was either from a single peer or from multiple peers. These researchers found that the quality of students' work improved the more when students received feedback from multiple peers, as compared to receiving feedback from a single expert (effect size = 1.23). In a follow-up study, Kwangsu Cho and Charles MacArthur (2010) showed that students receiving feedback from a single expert made simpler amendments to their work, which often did not improve the quality, as compared to students receiving feedback from peers. Again, those students who received feedback from multiple peers did more complex changes and improved the quality of their work. The authors suggest that these effects can be explained by the fact that it is easier for the students to understand (and therefore to use) feedback from peers. Furthermore, the feedback from multiple peers could contain three times as much "non-direct feedback", which is general observations that can be applied to other tasks as well, as compared to the other types of feedback (cf. Carnes et al., 2010).

As an example of research relating to the third outcome (i.e. peer-assessment skills and the quality of peer feedback), Orsmond et al. (2002) found that when students are actively engaged with the understanding of standards and criteria, for instance by engaging them in the generation of criteria or working with exemplars, their feedback will be better founded (cf. Sadler, 2002). Peer-assessment skills can also be improved by giving teacher feedback on students' peer feedback (Nicol, 2010).

As an example of research relating to the fourth outcome (i.e. students' attitudes toward peer assessment and peer feedback), Ngar Fun Liu and David Carless (2006) have shown that in order to engage in peer feedback, students need to be aware of the purpose of this activity and be able to relate it to the goals of instruction (see also Lingefjärd & Holmquist, 2005). Furthermore, if students can see the benefits of peer assessment and peer feedback, in terms of learning, this will have a positive impact on their motivation and their attitudes (Boud, 2000). In research investigating stu-

dents' perceptions of peer assessment (in higher education), it can be seen that students are indeed able to recognize the benefits of such activities, but they are also aware of the problems that may occur if the activities are not appropriately integrated into the course design (Hanrahan & Issacs, 2001). These findings are confirmed by Philip Vickerman (2009), who also shows that peer assessment can help students structure their work, as well as becoming more active, engaged, and independent learners.

To summarize, in the assessment-for-learning literature two methods are characteristic in order to activate students as resources for each other: peer assessment and peer feedback. Research supports the claim that the use of peer assessment and peer feedback may have both cognitive and motivational gains. It is noteworthy, that in order to support student learning, it does not seem necessary that peer feedback is accurate. Instead, it seems more important that students justify their assessment and feedback, but also that students understand the standards and criteria they are supposed to use and that the peer-feedback activities are appropriately integrated in the course context and aligned with the goals of instruction. Furthermore, feedback from multiple peers has been shown to improve student performance to a greater extent than feedback from a single expert.

Activating students as owners of their learning

Several methods can be used to activate students as owners of their learning, such as self-, peer-, and collaborative assessment (or "co-assessment", which means that student and teacher assess together). However, Dochy et al. (1999) recommend that peer-, or co-assessment is used as a go-between, before starting to self-assess, in order to help students develop skills that are required when self-assessing. Since peer assessment has already been discussed in the previous section, the focus here will be on co-assessment as a first step towards activating students as owner of their learning.

Co-assessment, or "assessment in partnership" (Stefani, 1998), offers a middle way between assessment performed by the teacher and assessment performed by students themselves, or by groups of peers, which means that co-assessment creates an opportunity for

students and teacher to share the assessment process. In an environment characterized by co-assessment, students therefore participate actively in the assessment process and students and teacher collaborate in order to clarify goals and criteria, negotiate details of the assessment, and discuss any misunderstandings that exist (Gouli, Gogoulou, & Grigoriadou, 2010).

Co-assessment is thought to promote a sense of ownership on part of the students (Stefani, 1998), but the use of co-assessment has also been shown to improve student learning and motivation (Dochy et al., 1999; McConnel, 2002; Stefani, 1992). For instance, Raymond Summit and Anne Venables (2011) found that co-assessment activities improved students' communication skills and use of mathematical symbols. According to Evangelina Gouli et al. (2010), co-assessment may even help students to develop generic skills, such as decision-making and collaboration. In most studies, however, co-assessment is combined with self-, or peer assessment (Dochy et al., 1999; Falchikov, 2001; Hall, 1995; Orpen, 1982; Stefani, 1992), making it difficult to identify the specific contribution of co-assessment. According to David McConnell (2002), what co-assessment adds to the other forms of assessment is the "openness" of the assessment process. This transparency is brought about by the collaborative process between teacher and student, which permits the negotiation of criteria and the mutual understanding of different qualities. In order to secure the success of co-assessment, a classroom culture building on trust and acceptance of criticism has to be fostered (McConnell, 2002, Lauf & Dole, 2010; Summit & Venables, 2011).

Weaker points of co-assessment are connected to organizational issues, such as time constrains and stress, but also to deeper issues, such as the teachers role in the assessment process. As noted by Dominique Sluijsmans, Filip Dochy, and George Moerkerke (1999) the "idea that teachers do the teaching and marking is hard to change" (p.314).

To summarize, several methods can be used for activating students as owners of their learning, such as self-, peer-, and co-assessment. Research has shown that co-assessment can be effective in improving students' learning of both domain-specific and gener-

ic skills, as well as increasing student motivation. An important contribution from co-assessment activities, is the transparency provided, although this may not come easy, since the collaboration of students and the teacher in the assessment process depart from the traditional view of (summative) assessment, where the teacher takes all responsibility.

Activating students as resources for each other and as owners of their learning in the classroom

In the sections above, it is argued that involving the students in the assessment process may have gains for their learning. Accordingly, a combination of peer-, and co-assessment was used when students in the intervention group worked with group assignments. The intention was to bring the teacher and students together in a way that optimized student learning. Peer- and co-assessment activities were therefore also combined with the receiving and formulation of feedback. In order to accomplish this, students were asked to assess and give feedback through the use of the scoring rubric. These activities were carried out in a sequence described in detail in Figure 8.
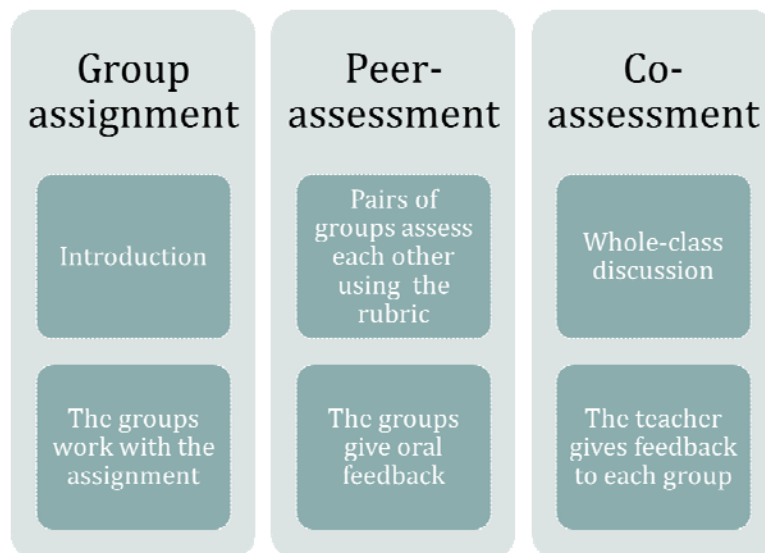
Figure 8. Working with a group assignment in the intervention group. First the students received an assignment, which they solved in groups (Group assignment). Then pairs of groups exchanged and assessed each other's solutions using the rubric and gave each other oral feedback (Peer assessment and feedback). Then the teacher conducted a whole-class discussion, comparing and discussing the qualities of different solutions (Co-assessment). Afterwards, the teacher gave individual feedback to each group (Teacher feedback).

Firstly, students were divided into small groups (3-4 students) by the teacher. Then the assignment (for some examples, see Appendix 5) was introduced. In order to avoid misunderstandings, the students were given a couple of minutes to read through the assignment and formulate questions to the teacher. The groups then worked by themselves for about 40-60 minutes. At the end of this session, they switched solutions with another group (which was decided by the teacher beforehand) and assessed the other group's work using the scoring rubric. This activity took about 10-15 minutes. During this time, students within a group would sometimes have different opinions about the quality of the other group's

work. In order to make their voices heard, the students needed to have solid arguments to persuade the others. Thereafter, the groups presented their assessments to each other and gave each other oral feedback. By formulating feedback, students had to verbalize their thoughts and justify their judgments about the other group's work. The purpose of using the rubric was, besides clarifying goals and criteria, to give structure to students' feedback and assessment. You could say that the use of the rubric was intended not only to increase transparency regarding criteria, but also to increase transparency in students' thinking. The feedback was directed towards the task and the processes involved when working with the assignment. At this stage, the groups compared their solutions and commented upon differences in reasoning and presentation. They also gave suggestions for how the other group could make improvements to their work, in order to reach higher levels in the scoring rubric. Finally, the groups handed over the solutions to the teacher. This was the end of the lesson. The teacher would then assess the work done by each of the groups.

The next lesson (which could be after a couple of days) started with a whole-class discussion about the assignment and the problem was solved on the whiteboard. In this way, several solutions were presented. The differences between the different solutions, as well as the differences in quality related to the assessment criteria in the rubric, were discussed by the teacher together with the students. The teacher then talked to each group separately, where the teacher's assessment of the group assignment was presented together with oral feedback to the group. The feedback focused on strengths as well as areas in need of improvement in the assignment and the teacher also justified her judgments. The feedback was complemented by a general discussion where references were made to the peer feedback and to the whole-class discussion.

The integration of peer-, and co-assessment was also reflected in the summative assessment by using "pair tests". The pair tests were performed in pairs. First they received four problems to be solves and then each pair of students chose one problem to present to the class. The presentation was followed by questions from other pairs of students, where one particular pair of students was told to assess

and give oral feedback to the presenting pair. As before, the students used the scoring rubric for assessing and giving feedback. Finally the pair-work was collected by the teacher.

The intervention-group teacher

An important condition for the success of the empirical study was the relation between the researcher and the intervention-group teacher. Although the present thesis does not go into any detail in this matter, the teacher's views about learning (especially views about mathematical learning), as well as the teacher's willingness to make changes in her teaching, were crucial for the intervention. That the intervention-group teacher followed the agreements made with the researcher was imperative, not only for the success of the instruction, but also for being able to interpret the effects of the intervention. Since the study took place during a whole school year, a personal and open communication between the researcher and the teacher was necessary in order to assure that the intervention was maintained during the whole period and that no deviations were made from the design.

In the study at hand, the intervention-group teacher proved to be very keen on adopting the spirit of "assessment for learning" even though this meant a lot of additional work for her. The researcher and the teacher had weekly meetings and daily contact though e-mail. They planned and discussed the lessons and they created the group assignments and the individual tests together. This close co-operation made it possible to implement the intervention as it was intended by the researcher and also to have an insight into what was happening in the classroom. For the teacher it meant a constant support and a way of receiving in-service training.

The control group

The teaching in the control group was conducted in a "traditional" manner typical for the school where the intervention took place. This means, for example, that the students were given an overall planning at the beginning of the semester, covering the pages in the textbook and the tasks students were supposed to work with. In this way, students could work individually and at their own pace. The teaching of a new unit was conducted in the following way: The teacher introduced the new concepts by lecturing at the whiteboard and then the students worked individually in their exercise books. When students worked on their own, the teacher helped those who needed assistance. At the end of a unit, for example Geometry or Functions, the students had to do an individual written assessment. The tasks in the test were very much like those in the textbook. Each task was scored and the total score was used to inform the students how well they had performed at the test by translating the scores into a letter grade. More differentiated feedback was given to the students when they were working on their own.

In relation to the idea of formative assessment, it is important to note that neither students' results on the summative tests, nor the teacher's observations in the classroom, were used to adjust the initial planning or to make other changes in the instruction. Rather, the teacher gave recommendations to students with low test results to do some of the exercises again and/or to work with additional exercises. The goals from the curriculum were presented at the beginning of the course, but after that the teacher did not return to them. The teaching did not involve working in groups or working with problem-solving tasks.

THE INFLUENCE OF "ASSESSMENT FOR LEARNING" ON STUDENTS' MATHEMATICAL LEARNING

In this chapter, the effects of the intervention on student learning in mathematics are reported. The changes in students' performances and students' beliefs are reported separately, first after one semester and then after two semesters. Comparisons with the control group are be made throughout the chapter.

Pre-test

Before the intervention began, students' mathematical problem-solving performances, as well as their mathematics-related beliefs, were compared between the groups using a problem-solving test and the beliefs' questionnaire. These comparisons were made by ANOVA. Means and standard deviations from the comparisons are presented in Table 2.

According to the comparisons, there were no significant differences between the intervention group and the control group, either with respect to students' performances on the problem-solving test or to their answers on the beliefs questionnaire.

Table 2. Means and standard deviations of the intervention group and control group in all the dependent variables at the beginning of the first semester

	Intervention group		Control group	
	M	SD	M	SD
Epistemological beliefs	34.00	3.40	33.60	3.88
Beliefs about assessment	20.85	2.37	21.22	3.58
Self-concept	21.30	2.95	21.81	3.43
Problem-solving	6.80	2.67	6.70	2.28

Besides the comparisons, a correlation analysis was performed on the items in the beliefs questionnaire, which showed that students' initial epistemological beliefs had a positive correlation with their beliefs of themselves as mathematical learners. This means that those who see themselves as good mathematical learners also have conceptions about mathematical learning that are more availing (i.e. more productive in relation to learning). The results from the problem-solving test also correlated positively, albeit at a lower level, with students' self-concept, which indicates that the students (to some degree) were aware of their mathematical skills. These correlations are shown in Table 3.

In conclusion, as far as the instruments show, there were no significant differences between the control group and the intervention group with regard to their performance in mathematical problem-solving or their mathematics-related beliefs.

Table 3. Inter-correlations between variables at the start of the study (N = 45).

	Epistem. beliefs	Beliefs assessment	Self-concept	Probl. solving
Epistem. beliefs	1	.224	.599**	.213
Beliefs assessment		1	-.004	.000
Self-concept			1	.298*
Probl. solving				1

**Correlation is significant at the .01 level (2-tailed); *Correlation is significant at the .05 level (2-tailed). Abbreviations: Epistem. beliefs = Epistemological beliefs; Probl. solving = Problem solving

Results after one semester (Phase I)

How did students' performances and mathematics-related beliefs change in the two groups during the first semester? Data for answering this question was collected with the help of three instruments: two mathematical tests and the beliefs questionnaire. An analysis of variance was performed in order to compare students' results on the mathematical tests, as well as their answers to the beliefs questionnaire.

Mathematical performance

Students' results on the problem-solving test and the National test in mathematics are presented here. Students' results on the problem-solving test are first compared between the groups and then within the intervention group.

Total scores on the mathematical problem-solving test

Table 4 presents the descriptive statistics (i.e. means and standard deviations) for the comparison between the problem-solving test at the start and at the end of the semester for both the intervention group and the control group. Several remarks can be made from the table.

Table 4. Means and standard deviations from the comparison between the intervention group and the control group on the problem-solving test.

	Pre-test		1st post-test	
	M	SD	M	SD
Intervention group	6.80	2.67	7.60	2.09
Control group	6.70	2.28	4.95	1.58

Firstly, after one semester the results on the problem-solving (post-) test were significantly higher for students in the intervention group, as compared to the results for students in the control group (p < .001). As the table illustrates, this difference is quite large and has an effect size[2] of 1.43 (Cohen's *d*).

Secondly, it can be noted that the students in the intervention group improved their problem-solving performance (increase from 6.80 to 7.60) during the semester. The improvement, however, was not statistically significant, which was confirmed by a *t*-test.

Thirdly, the performance of the students in the control group did not improve during the semester. In fact they performed less well in the post-test than in the pre-test (decrease from 6.70 to 4.95). A *t*-test confirmed that the difference was statistically significant (p < .001).

Analytical scoring of the mathematical problem-solving test
The results of the intervention group and the control group on the problem-solving test were analyzed using the three assessment criteria in the scoring rubric (i.e. "Method and Execution", "Mathematical Reasoning", and "Presentation and Mathematical Language") by estimating students' scores with respect to each of the criteria. Thus the total amount of points awarded (in relation to each criterion) was divided by the maximum score (see Table 5).

---

[2] Hattie (2009, p. 8-9) suggests that values 0.2 to 0.3 might be a "small" effect, around 0.5 a "medium" effect, and above 0.8 a "large" effect.

This was done for both the intervention group and the control group. For example: the control group attained (as a group) 101 points with respect to the "Method and Execution" criterion, of a maximum of 216 points. Thus they achieved 47 percent of the highest possible score for this criterion.

As can be seen in Table 5, the relative scores of the intervention group are higher than those of the control group with respect to all criteria. The largest difference between the groups is with respect to the "Presentation and Mathematical Language" criterion. Within the control group, "Method and Execution" has the highest proportion of points awarded, while the relative scores for the other criteria are considerably lower. In the intervention group, the criterion "Mathematical Reasoning" has the lowest proportion of points awarded, but as compared to the control group the scores are more evenly distributed among the three criteria.

Table 5. Group results on the problem-solving test for each of the three criteria. The results are presented as relative scores (i.e. attained score / maximum score) for each group.

	Method and Execution	Mathematical Reasoning	Presentation and Mathematical Language
Intervention group	69 %	53 %	60 %
Control group	47 %	30 %	32 %

Changes within the intervention group

A closer look at the results on the problem-solving test shows that some of the students in the intervention group improved their performance, while others did not. In fact, most of the students who improved their performance, was students who performed less well on the pre-test (i.e. attained scores in the lower half of the rating scale). The students were therefore divided into two groups: one group that attained scores in the lower half of the rating scale on the pre-test ("low-achievers") and one that attained scores in the upper half of the rating scale ("high-achievers"). There were ten students in the first category, of which three performed equally well on the post-test as on the pre-test. The other seven students improved their performance. These seven students are referred to as "Group Low" below.

In the group of high-achievers, there were ten students, of which three performed at a higher level on the post-test as compared to the pre-test, while the others did not. These three students are referred to as "Group High" below.

Table 6 presents the changes in results (from pre-test to post-test) for the two groups with respect to the three criteria in the rubric. It should be noted that these changes are not comparable across the groups, since there are seven students in Group Low and only three students in Group High. The aim of Table 6, however, is not to compare the absolute values between the groups, but rather to determine in relation which criteria the two groups improved the most. As can be seen in Table 6, Group Low made significant improvement with respect to the criteria of "Mathematical Reasoning", closely followed by "Method and Execution", while Group High improved most with respect to "Presentation and Mathematical Language". This means that the students in Group Low improved the way they interpret problems and how well they use appropriate mathematical methods and reasoning in order to solve problems. The students in Group High, on the other hand, improved the clarity and completeness of their solutions, as well as the use of mathematical symbols, terminology, and conventions.

Table 6. Changes in the results for Group Low and Group High with respect to the three assessment criteria.

	Method and Execution	Mathematical Reasoning	Presentation and Mathematical Language
Group Low	+14	+15	+10
Group High	+ 6	+2	+9

Samples from the performances of the intervention group students in the pre- and the post-test are used below to illustrate the above-mentioned changes. For this purpose, three students were selected: Two students from Group Low (L1 and L2) and one from Group High (H1). One item from the pre-test and the corresponding item in the post-test were selected for each student.

For student L1, a task called "Assembly halls" was selected from the pre-test (Appendix 2), together with the corresponding task from the post-test (Appendix 3). In both tasks, the students had to work with numbers and find a pattern. In order to guide the students, the first questions of the task asked them to calculate some special cases. For example, the students had to calculate the number of chairs in the rows of an assembly hall. In the end, however, the students had to find a general formula that mathematically proves that the number of chairs follow a recurring pattern.

On the pre-test, student L1 applied an incorrect interpretation of the formula presented in the problem (in the second part of the problem) and used an incorrect method for solving part three (Figure 9). On the post-test, however, the same student selected and used appropriate methods for solving a similar problem. As a consequence, the performance of student L1 improved in relation to the criterion "Method and Execution".

```
I a)    31 seats
        row 1 – 10 seats
        row 2 = 13 seats
        row 3 = 16
        row 4 = 19
        row 5 = 22
        row 6 = 25
        row 7 = 28
        row 8 = 31.    Answer: row 6 has 25 seats


        b) Answer 3 rows


        c) 10 + 3n


II      12 + 5n
        row 1 = 12
        every row increases with 5 seats


III     10 + 3n
        19 · 8 = 152 seats
        Kalle is wrong


        10 + 13 + 16 + 19 + 22 + 25 + 28 + +31 = 162 seats
```

Figure 9. Solution to the problem "Assembly halls" (Appendix 2) by student L1. The solution in the figure is transcribed and translated, but the original solution can be found in Appendix 4.

Student L1 also showed improvement in relation to the criterion "Mathematical Reasoning". While this student failed to find a correct general formula, and presented no mathematical reasoning, on the pre-test, on the post-test the same student presented a correct general formula and proved that the formula applied for a special case.

Regarding the criterion "Presentation and Mathematical Language", the performance of student L1 did not change dramatically. On the pre-test, this student's presentation was clear, but the

use of mathematical language and symbols was quite poor. On the post-test, this student still used inadequate mathematical language and the geometrical representation was not very well developed.

For students L2 and H1, a task called "Currency exchange" was chosen from the pre-test (Appendix 2) and a task called "The inheritance" from the post-test (Appendix 3). In both tasks, the students had to work with functions and numbers. Three alternatives of currency exchange and/or inheritance were presented and the students had to choose the most appropriate one.

On the pre-test, student L2 used a mathematical method that was partly incorrect. On the post-test, however, the same student used correct methods for the calculation of the three alternatives in the inheritance problem. As a consequence, the performance of student L2 improved in relation to the criterion "Method and Execution".

B.   0.5 $500 - 250 = 250$

     ·500 ₂

       00          $100 = 7.7$ pounds

       00

        5          $7.\not{7}^{1} = 3.80$

   +2   2

   250.0

             $200 \, kr = 2$ $= 15.4 \, £$

                  ·7.7 ₁

                  15.4

             $50 \, kr = 3.80 \, £$

             $250 = 15.4 + 3.8 \, £ = 18.2 \, £$

          Answer: $250 \, kr = 18.2 \, £$.

          $18.2 \cdot 2 = 36.4 \, £ = 500 \, kr$

Figure 10. Part of the solution to the problem "Currency exchange" (Appendix 2) by student L2. The solution in the figure is transcribed and translated, but the original solution can be found in Appendix 4.

The performance of student L2 also improved in relation to the criterion "Mathematical Reasoning". On the pre-test, no attempts were made to reason about the different possibilities for currency exchange and the wrong alternative was selected. On the post-test, however, the same student presented general formulas for two of the alternatives, in the form of linear and exponential functions. Yet, the reasoning about the most appropriate alternative was still limited (Figure 11).

Student L2 also showed improvement in the use of mathematical symbols and language. As compared to the pre-test, when student L2 used poor mathematical language (Figure 10), on the post-test the same student communicated his mathematical reasoning and made attempts to use mathematical symbols (Figure 11).

The relationship between the amount of money and number of years in alternative A.

The fixed amount is 500 kr, he receives this every year. So the total amount increases by 500 kr/year
$550 \cdot x$
x = years

The relationship between amount of money & number of years in alternative C.
Robert receives a sum of 2000 kr that increases by 11 % every year. After (x) years he thus has
1.11 = exchange factor

$2000 \cdot 1.11^{x}$

Figure 11. Part of the solution to the problem "The inheritance" (Appendix 3) by student L2. The solution in the figure is transcribed and translated, but the original solution can be found in Appendix 4.

Student H1 showed improvement in relation to the criterion "Method and Execution". The choice and execution of the mathematical method was more complete on the post-test as compared to the pre-test. On the post-test, this student not only calculated the three inheritance alternatives correctly, but also selected an appropriate method to compare them with each other (which was not the case on the pre-test).

In relation to the criterion "Mathematical Reasoning", student H1 did not improve considerably on the post-test as compared to the pre-test. Although this student selected an appropriate method and performed the calculations, student H1 gave no evidence of a more developed mathematical reasoning.

A

1000kr = 70 £
500kr = 35 £

100 sek= 8.6 £ C

8.6 · 5 = 43p = 43 £
43 – 5 = 37 £

B   100 sek = 7.7 £

7.7 · 5 = 39.5
38.5
·0.05 24
19.25   500·0.05 = 25

7.7 · 4.75 = 36.58
500 – 25 = 475

Figure 12. Part of the solution to the problem "Currency exchange" (Appendix 2) by student H1. The solution in the figure is transcribed and translated, but the original solution can be found in Appendix 4.

The greatest improvement for student H1 was in the use of mathematical language and symbols. On the pre-test this student made use of a quite poor mathematical language and the communication

96

of the reasoning was not easy to follow (Figure 12). On the post-test, however, student H1 used a more appropriate mathematical language. By presenting two inheritance alternatives as functions, and providing a clear description of the variables x and y, the student demonstrated a well-developed mathematical language. The graph used may have been more precise, but the use of the graphical representation was appropriate for the problem (Figure 13).
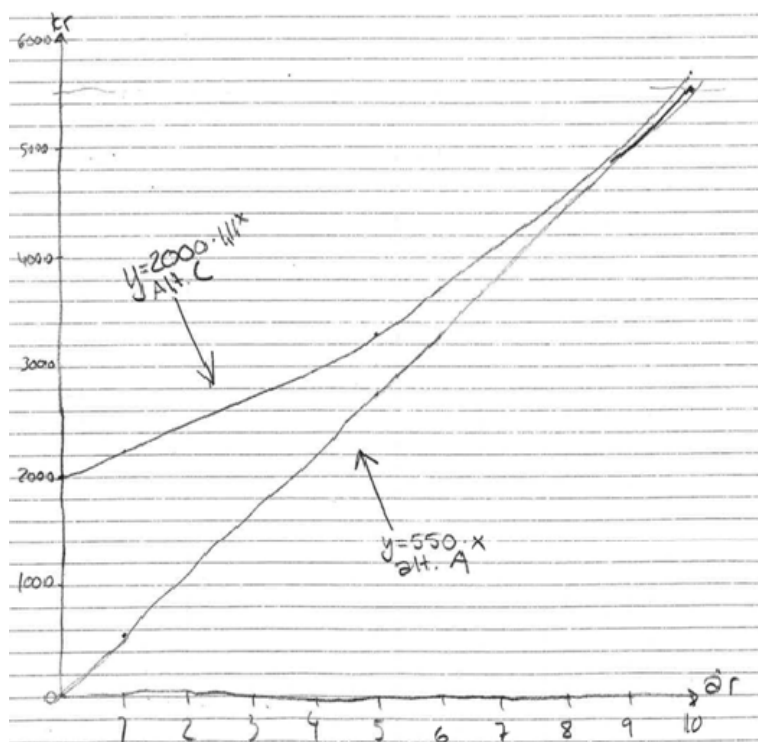
Figure 13. Part of the solution to the problem "The inheritance" (Appendix 3) by student L2. The solution in the figure is transcribed and translated, but the original solution can be found in Appendix 4.

Results on the National Test in Mathematics
At the end of the first semester, students' mathematical performance was also measured by the ordinary National test in mathe-

matics for upper-secondary schools. The test consisted of 25 questions and the students were allowed 180 minutes to solve these questions. The solutions were assessed according to national assessment criteria. As can be seen in Table 7, the students in the intervention group performed slightly better than the students in the control group (43.73 compared to 38.91 points) on the test as a whole. This difference was, however, not statistically significant.

Besides performing somewhat better on the test as a whole, the intervention group also performed somewhat better on the problem-solving task included in the test. An ANOVA confirmed that the students in the intervention group performed at a significantly higher level as compared to the students of the control group ($p < .05$; Cohen's $d = .6$).

Students' beliefs

The beliefs questionnaire consisted of three scales, intended to measure students' beliefs about: (a) Mathematics as such (the "Epistemological beliefs" scale), (b) themselves as students of mathematics (the "Self-concepts" scale), and (c) the assessment of mathematics in school (the "Beliefs about assessment" scale). The results from the questionnaire are presented in Table 8 and 9, which show means and standard-deviation for both groups at the beginning and at the end of the first semester.

Epistemological beliefs
Students' answers to the items in the epistemological beliefs scale did not change much in the intervention group from the beginning to the end of the first semester.

Table 7. Means and standard deviations for the two groups on The National Test and the problem-solving task included in the test. The highest possible score on The National Test as a whole was 60 and on the problem-solving task 11.

	National Test as a whole		Problem-solving task	
	M	SD	M	SD
Intervention group	43.73	9.00	5.73	3.06
Control group	38.91	9.12	4.04	2.48

Table 8. Means and standard deviations of the intervention group on the beliefs variables at the beginning and end of the first semester.

	Pre-test		1st Post-test	
	M	SD	M	SD
Epistemological beliefs	34.00	3.40	33.85	3.83
Self-concept	21.30	2.95	21.20	2.28
Beliefs about assessment	20.85	2.37	21.00	3.11

Table 9. Means and standard deviations of the control group on the beliefs variables at the beginning and end of the first semester.

	Pre-test		1st Post-test	
	M	SD	M	SD
Epistemological beliefs	33.60	3.88	32.41*	3.98
Self-concept	21.81	3.43	20.72*	3.83
Beliefs about assessment	21.22	3.58	20.95	3.14

* Significant at the .05 level, as compared to the results on the pre-test.

The answers of the control group, on the other hand, changed significantly as compared to the pre-test (p < .05), towards becoming less availing. This means that, at the end of the first semester, students in the control group to a greater extent answered that mathematics is about applying rules and methods, that it is more about delivering the right solution than understanding the reasoning behind the solution, that mathematical problems can be solved in only one way, and that it is more important to learn mathematics for future education than for use in everyday life. Still, there were no significant differences between the intervention group and the control group on the epistemological beliefs scale.

By comparing the answers on individual items, however, it can be seen that there are some interesting differences between the intervention group and the control group. This is especially true for one particular item, which asks the students to agree/disagree as to whether a person, who does not understand why an answer is correct, has indeed solved the problem. In relation to this item, students in the intervention group to a larger extent thought that you need to understand why the answer is correct, in order to have indeed solved the problem.

Mathematical self-concept
When comparing the answers from the two groups at the end of the first semester, no significant differences could be identified on the self-concept scale. However, when comparing the answers from the pre- and post-test, the answers from the control-group students indicated less availing self-concept beliefs at the end of the semester, as compared to the pre-test ($p < .05$). This means that after a semester at upper-secondary school, the students in the control group were less positive about their mathematical skills and their ability to learn mathematics.

Beliefs about assessment
In both groups, students' answers on the assessment scale were basically unchanged from the beginning to the end of the first semester and, as a whole, the differences between the groups were small. On individual items, however, the answers sometimes differed. For instance, it was noted that the control group to a greater extent thought that it was important to know the assessment criteria in mathematics.

Relations between the variables
A correlation analysis was performed in order to investigate the relations between the four variables (i.e. the results in the problem-solving test and the three categories of students' beliefs) at the end of the first semester. Any correlation between mathematical performance, on the one hand, and beliefs about learning mathematics in school, as well as self-concept in relation to mathematics, on the other, would be of interest.

Table 10. Correlations between different variables in the control group at the end of the first semester.

	NT	PS post-test	PS pre-test	Epist. post-test	Self. post-test
NT	1	.691**	.651**	.592**	.686**
PS post-test		1	.430*	.221	.280
PS pre-test			1	.617**	.731**
Epist. post-test				1	.887**
Self. post-test					1

**Correlation is significant at the .01 level (2-tailed). *Correlation is significant at the .05 level (2-tailed). Abbreviations: NT = National Test in Mathematics; PS = Problem-solving test; Epist. = Epistemological beliefs; Self. = Self-concept.

In the case of the control group, the correlation analysis showed that students' total scores on the National test were strongly correlated both with students' self-concept (r = .686, p < .001) and with their epistemological beliefs (r = .592, p = .004) (see Table 10). This means that students with good results on the achievement tests expressed more availing beliefs both about their self-concept and about mathematics. Compared to the analysis made at the start of the semester, the correlations between students' self-concept and their epistemological beliefs (r = .887; p < .001) were stronger after one semester. These results indicate that after the first semester in upper-secondary school, students with more availing beliefs about mathematics also have greater self-confidence and enjoy working with mathematics.

Furthermore, students' self-concept in the control group was shown to be positively correlated with their initial results in problem-solving (r = .731, p < .001), but not significantly correlated to their final results. Instead, changes in problem-solving performance

were reversed in relation to the changes in self-concept in the control group (r = -.483, p = .023). This means that students who were confident about their capacity to learn mathematics, and expressed an interest in mathematics, actually performed less well on the problem-solving tests at the end of the first semester.

In the intervention group, a correlation was found between students' results on the National test and their results on the problem-solving post-test[3] (r = .519, p = .015) (see Table 11). There was, however, no significant correlation between students' results on the problem-solving test before and after the first semester, which indicates that those who performed well on the first test did not necessarily perform well on the post-test. This indicates that changes in students' rank order may have taken place during the semester. A correlation analysis between changes in students' results on the problem-solving test and their results on the pre-test confirmed this by showing a significant negative correlation (r = -.703, p < .001). These results mean that students with low results on the pre-test performed higher on the post-test.

---

[3] When an outsider was excluded.

Table 11. Inter-correlations between the results on the National test and the problem-solving test in the intervention group.

	NT	PS post-test	PS pre-test	Diff. PS
NT	1	.519*	.484*	-.290
PS post-test		1	.353	.438
PS pre-test			1	-.703**
Diff. PS				1

**Correlation is significant at the .01 level (2-tailed). *Correlation is significant at the .05 level (2-tailed). Abbreviations: NT = National Test in Mathematics; PS = Problem-solving; Diff. PS = Differences in the results on the problem-solving tests.

Additional analyses show no significant correlations between the National test results, the problem-solving pre- and post-test, and the beliefs variables. However, if only analyzing the beliefs variables, there is a correlation ($r = .563$; $p < .01$) between students' self-concept and their epistemological beliefs (Table 12). A closer look shows that the difference in self-concept (i.e. the difference in self-concept scores on the questionnaire between the beginning and the end of the semester) correlates negatively with students' self-concept at the start of semester ($r = -.699$, $p < .01$). This means that students, who – at the beginning of the semester – expressed low self-confidence in mathematics, as measured by the self-concept scale, expressed a greater self-confidence at the end of the semester.

Table 12. Inter-correlations between the different beliefs variables in the intervention group at the end of the first semester.

	Self. post-test	Epist. post-test	Diff. self.	Self. pre-test
Self. post-test	1	.563**	.380	.396
Epist. post-test		1	-.011	.446*
Diff. self.			1	-.699**
Self. pre-test				1

**Correlation is significant at the .01 level (2-tailed). *Correlation is significant at the .05 level (2-tailed). Abbreviations: Self. = Self-concept; Epist. = Epistemological beliefs; Diff. self. = Difference in Self-concept.

With regard to students' beliefs about assessment in mathematics in the intervention group, no significant correlation was found in relation to the other beliefs variables.

Summary

The analyses of the results after the first semester show a clear positive change in mathematical problem-solving performance for the students in the intervention group, as compared to the students in the control group. This was demonstrated both on the specific problem-solving test and on the National test in mathematics. Furthermore, more detailed analyses of students' results in the intervention group showed that the students with low results on the pre-test were the ones who improved the most after one semester. This improvement was especially pronounced with regard to how well they interpreted problems and how well they used appropriate mathematical methods and reasoning in order to the solve problems. Students with high results on the pre-test, on the other hand, mostly improved on how clearly and completely they presented solutions and how well they used mathematical symbols, terminology, and conventions.

Another interesting finding was the lack of correlation between students' performance in the intervention group on the one hand, and their answers to the beliefs scales on the other. In the control group, however, there was a positive correlation between students' results on the National test and their beliefs about self-concept and mathematical learning.

Yet another difference between the two groups was that the beliefs of the students in the intervention group did not change during the first semester. The control group, on the other hand, displayed less availing beliefs in relation to the epistemological beliefs scale and the self-concept scale.

In the intervention group, results also showed that students with initial low scores on the self-concept scale had higher scores at the end of the first semester.

After the second semester (Phase II)

After the second semester, students' mathematical beliefs were investigated once more.

Beliefs questionnaire

The beliefs questionnaire was administrated again after the second semester. The descriptive statistics are presented below (Table 13 and 14).

Table 13. Means and standard deviations for the beliefs variables in the intervention group at the start of the first semester and at the end of the second semester.

	Pre-test		2nd post-test	
	M	SD	M	SD
Epist. beliefs	34.00	3.40	37.05**	2.62
Self-concept	21.30	2.95	21.05	3.14
Beliefs about assessment	20.85	2.37	22.10*	2.77

**Significant at the .01 level as compared to the control group. *Significant at the .05 level as compared to the control group. Abbreviations: Epist. = Epistemological.

Table 14. Means and standard deviations for the beliefs variables in the control group at the start of the first semester and at the end of the second semester.

	Pre-test		2nd post-test	
	M	SD	M	SD
Epist. beliefs	33.60	3.88	33.46	4.69
Self-concept	21.81	3.43	19.96*	3.83
Beliefs about assessment	21.22	3.58	19.92	3.92

*Significant at the .05 level as compared to pre-test results. Abbreviations: Epist. = Epistemological.

Epistemological beliefs

Table 13 shows that students' answers to the items in the epistemological-beliefs scale changed in the intervention group, as compared to nine months earlier, when the intervention began. A *t*-test showed that these changes were statistically significant (p < .001). According to students' answers on the second post-test, they

thought it was time-consuming to solve mathematical problems, that it was important to understand the mathematical concepts involved, that problems could be solved in different ways, and that mathematics was important not only for future studies, but also for everyday life.

No changes could be observed in students' answers to the epistemological scale in the control group, which indicates that the views of these students about mathematics and mathematical learning had not changed during their first year at upper-secondary school. An ANOVA showed that the difference between the answers of the control group and intervention group is statistically significant ($p < .01$).

The answers from the intervention group differed from those of the control group on individual items. Particularly, three items stand out where students' in the intervention group expressed significantly more availing beliefs, as compared to the control group. One item has already been commented upon (after the first semester), which asks the students to agree/disagree as to whether a person, who does not understand why an answer is correct, has indeed solved the problem. The other two items are about usefulness. The first one is about the importance of knowledge in mathematics for the future and the other about how useful knowledge in mathematics is in everyday life.

Mathematical self-concept

As opposed to epistemological beliefs, students' answers to the items about how they perceive themselves as learners in mathematics, and if they enjoy working with mathematics, did not change in the intervention group (Table 13).

Students in the control group, on the other hand, changed their answers in a way that indicates less positive feelings towards learning mathematics, lower confidence in their own competence, and less interest in learning mathematics as compared to their answers when they started upper-secondary school. A *t*-test showed that this change was statistically significant ($p < .05$).

The answers from the intervention group differed slightly from those of the control group on individual items. However one ob-

servation was noted in the answers from the intervention group. For instance, students in the intervention group answered significantly more positively to the question that asked whether they found it difficult to understand mathematics than their answers to the question that asked whether they felt inadequate if they did not understand mathematics.

Beliefs about assessment

In relation to beliefs about assessment, a change in students' beliefs, towards more availing beliefs, could be observed in the intervention group, while a change in the opposite direction was noted for the control group. For instance, students' answers in the intervention group indicated that they preferred being assessed in different ways, to receive feedback, and to understand the criteria for assessment, as opposed to only being assessed in a "traditional" summative way. A non-parametric test showed that this difference between the groups was statistically significant ($p < .05$).

Again, the answers from the intervention group differed from those of the control group on individual items. For example, it was obvious that the students in the intervention group were more positive about oral assessments in mathematics. They also found it more important to receive feedback from the teacher, as compared to the control group. Furthermore, and in opposition to the results presented after the first semester, the students in the intervention group, as compared to the students in the control group, to a larger extent agreed on that it is important to know the assessment criteria in mathematics.

Correlations between variables

The changes in students' answers to the beliefs questionnaire were analyzed in detail, in order to reach a better understanding of the processes that took place during the second semester.
The results from correlation analyses show that there was a weaker correlation between students' epistemological beliefs and self-concept in the control group at the end of the second semester, as compared to the first post-test ($r = .887$, $p < .001$). These results indicate that changes have occurred and correlations between these

changes in epistemological beliefs (i.e. between the pre-test and the second post-test) shows that there was a positive change for students who held less availing beliefs about mathematics from the beginning, as compared to those who held availing beliefs (r = .639, p < .001). The same development was observed regarding students' self-concept. Still, these changes do not seem to be related to students' mathematical performance. Instead, as can be seen in Table 15, there were no correlations between changes in students' beliefs and changes in students' performance on the problem-solving test.

Table 15. Inter-correlations between different beliefs variables in the control group at the end of the second semester.

	Epist. beliefs	*Self-con.*	*Diff. epist.*	*Diff. self.*	*Diff. PS*
Epist. beliefs	1	.628**	.639**	.040	-.288
Self-con.		1	.329	.578**	-.236
Diff. epist.			1	.227	-.027
Diff. self.				1	.022
Diff. PS					1

**Correlation is significant at the .01 level (2-tailed). *Correlation is significant at the .05 level (2-tailed). Abbreviations: Epist. = Epistemological; Self-con. = Self-concept; Diff. epist. = Difference in Epistemological beliefs; Diff. self. = Difference in Self-concept; Diff. PS = Difference in Problem-solving.

The results from the correlation analysis showed a stronger connection between students' epistemological beliefs and their self-concept in the intervention group, as compared to the first post-test (r = .652, p < .01, as compared to r = .563, p < .01). This means

that the students who viewed themselves as being good at mathematics, and who enjoyed working with mathematics, also were the ones who held more availing beliefs about what working with, and learning, mathematics means. Furthermore, the correlation analyses indicated that among the students, whose mathematical performance improved during the intervention, were also the students whose conceptions about mathematical learning and self-concept changed the most. Table 16 shows a medium positive correlation between changes in epistemological beliefs and self-concept related to changes in performance on the problem-solving test ($r = .454$, $p < .05$ and $r = .544$, $p < .05$ respectively).

With regard to students' beliefs about assessment, Table 16 shows a strong correlation between students' beliefs on the second post-test and the changes that occurred between the beginning of the intervention and two semesters later. This means that the students, who viewed assessment in mathematics as being authoritative and one-way, with no variation or any interaction from the students, were the ones who changed their views the most.

Table 16. Inter-correlations between different beliefs variables in the intervention group at the end of the second semester.

	Beliefs assess.	*Diff. epist.*	*Diff. assess.*	*Diff. self-con.*	*Diff. PS*
Beliefs assess.	1	.174	.735**	.071	.107
Diff. epist.		1	.388	.505*	.454*
Diff. assess.			1	.127	.182
Diff. self-con.				1	.544*
Diff. PS					1

**Correlation is significant at the .01 level (2-tailed). *Correlation is significant at the .05 level (2-tailed). Abbreviations: Beliefs assess. = Beliefs about assessment in mathematics; Diff. epist. = Difference in epistemological beliefs; Diff. assess. = Difference in beliefs about assessment in mathematics; Diff. self-con. = Difference in self-concept; Diff. PS = Difference in problem-solvning.

Summary

The results after the second semester showed that the intervention group significantly changed their beliefs about mathematics and about assessment in mathematics, towards more availing beliefs. A similar change could not be observed in the control group. Furthermore, the changes in students' beliefs in the intervention group are not matched by changes in self-concept with respect to mathematics. However, even if the mean value of students' self-concept remained unchanged, some individual changes occurred. These changes had a positive relation to changes in students' views about mathematics and mathematical learning, as well as their problem-solving performance. As an example, some students who improved

their mathematical performance also displayed a change in self-concept and their views about mathematics, towards more availing beliefs. Conversely, results from students in the control group on the second post-test indicated less availing self-concept beliefs. These changes do not seem to be linked to changes in problem-solving performance.

The results from the intervention as a whole clearly indicate positive changes regarding students' problem-solving performance and their mathematical-related beliefs. In addition to a significant improvement in the mean value, results on the problem-solving test show that the students who performed less well on the pre-test, were the ones who improved most during the intervention. The main improvements of these students were associated with skills in how to apply mathematical methods and use mathematical reasoning when solving problems.

With regard to students' beliefs in the intervention group, results show that the different aspects of students' beliefs systems were more closely linked to each other, and also to students' performance, after two semesters. This can, for instance, be seen by the fact that those students who changed their beliefs, towards more availing beliefs, were also the students who improved their problem-solving performance. Even though a statistically significant change in mean values could not be established for all variables, the analyses indicate that several students with less availing beliefs about assessment did indeed change towards more availing beliefs.

For the students in the control group, the results show no mean development, either in relation to their beliefs about mathematics, mathematical learning, or assessment. Furthermore, the results indicate a decline in students' problem-solving performance and self-concept. In this group too, changes occurred in the relationship between the variables, showing a weakening of the connection between the beliefs variables. No connection between changes in beliefs and changes in problem-solving performance could be observed.

THE STUDENTS' AND THE TEACHER'S PERCEPTIONS

In the previous chapter, results from quantitative analyses were presented, which will be used to answer the questions if – and in what ways – students' learning was affected by the intervention. However, in order to gain a deeper understanding of these results, for instance how the intervention brought about changes in student performance and beliefs, the results from the quantitative data will be set in relation to the various ways in which the students and the teacher perceived the intervention. Individual interviews with the students from the intervention group, and an essay written by the teacher after the first semester, provide the data of this analysis. The interview guide can be found in Appendix 6.

Results from student interviews

As described in the Analyses section, the transcribed interviews were arranged according to themes that originated from the framework of assessment for learning. These themes were: (a) the use of a scoring rubric, (b) the use of peer assessment and peer feedback, (c) working with problem-solving, and (d) students' mathematics-related beliefs.

Table 17. Main themes from student interviews with sub-themes.

Main theme	Sub-theme
The use of a rubric	Transparency
	Working with mathematics
	Assessment in mathematics
	The use of mini-rubrics for individual assessment
The use of peer assessment and peer feedback	Learning by receiving feedback
	Learning by giving feedback
	Learning from peer assessment
	Peer feedback vs. teacher feedback
Working with problem solving	Problem-solving tasks vs. textbook tasks
	Usefulness of mathematics
	Mathematical learning
Mathematic-related beliefs	Beliefs about the self
	Beliefs about mathematical learning
	Beliefs about assessment in mathematics

The use of a rubric

During the interviews, the students were asked about their experiences of working with rubrics, which resulted in a number of qualitatively different categories of perceptions. In the following, these different categories are described and exemplified with student responses. For the student responses, each individual student has been assigned a specific number, which is given in parenthesis after the quote.

Transparency

A number of students expressed that using a rubric helped them to understand the assessment criteria, which made the goals of instruction and the teacher's expectations more transparent to them. Students referred to using the rubric in the daily classroom work and when they worked with group assignments. Some of the students expressed difficulties with understanding the rubric in the beginning, but also that it became easier by practicing.

> It feels better when you know which goals you're supposed to fulfill. (3)

> Then we knew what was required for specific grades. (4)

> It's quite clear what you have to know: that, that, that. (2)

> The drawback is that I didn't understand from the beginning, but now it's easy. I tried to reach the "Passed with distinction" level, so that's helped me. (2)

> In the beginning it was quite difficult with the rubric. But it gets easier after you learn more. (7)

Working with mathematics

When students worked with tasks, especially complex tasks, they used the rubric, which was always lying on their desks. By using the rubric, students claim that they have learned to handle mathematical tasks in a different way and to think more systematically.

I think more about what I'm doing and how I write it. (8)

It's easier to see the different steps in a task. (1)

… to see the tasks in a different way. (10)

In the beginning it was quite tough. It felt like you really had to learn a hell of a lot, but now I feel like it's taught me. Cos when you need to solve more comprehensive tasks it's much easier. (9)

Assessment in mathematics

By using a rubric that includes criteria for several aspects of the problem-solving process, students express that they have become more aware of these different dimensions in mathematical solutions. Through the rubric, students were reminded to think of these other aspects, instead of just giving the answer. The rubric thus changed students' focus from merely delivering an answer to how to present and reason about the mathematical solution.

I'm not used to explaining when I do math tasks so it's really a big difference. I have to reflect, write everything down, and I haven't done that before. Before, all we had to do was give an answer. (8)

Before, on all the tests we just had to write the answer, no calculations, but now we've learnt to do that and to write down how you thought and that's good. (4)

It [the rubric] has made me think more about the answer. (12)

The use of mini-rubrics for individual assessment

The use of mini-rubrics, as opposed to numerical scoring, meant that the individual assignments were not marked by the teacher in the way students were used to. Instead, the mini-rubrics indicated which aspects were assessed in a task and which level of quality that the student had reached. Students expressed that this new way

118

of marking was not straightforward for them. However, even if students claimed to experience difficulties initially, they gradually learnt to appreciate the use of the qualitative categories, instead of the numerical scores. The students expressed that the mini-rubrics helped them to understand what was expected of them and to see in relation to which aspects they had to improve. Still, some students said that they missed the feeling of "security" that the scores had provided.

> Sure, if you see that you have too few points in relation to the "Give account" [criteria], you know you have to practice some more. (17)

> … if you get assessed in that way then you know what you were good at and what you didn't do well in and why. (16)

> But it's also good to have grades because then it's easier to get a feeling for how things went. (3)

> I still think it's quite good to, like, get a grade, to know what grades you got so that you know more exactly. With this assessment you didn't really know how things went. (7)

Peer assessment and peer feedback

The organization of group assignments was designed to facilitate peer assessment and peer feedback. During group assignments, two different groups assessed each other's work and then gave each other feedback. Students expressed both positive and negative experiences of peer assessment and peer feedback. The comments they made apply both to the receiving and to the giving of feedback. For instance, when assessing and giving feedback, students learned that other students may understand the task quite differently. Another experience was that assessing other's work is a difficult and strenuous task and that being able to justify your solution is of great importance.

Learning by receiving feedback

Students expressed that receiving feedback was very effective for their understanding and for discovering different ways of presenting and solving problems. Some students, however, were more reserved about the experience.

> You learnt about how others interpreted a task, solved a problem. You learn something new, maybe how to interpret a task in a different way. (7)

> The peer feedback helps you understand what you need to spend more time on, that you need to make more of an effort. (17)

> It's not bad but I didn't get much out of it. But it's good anyway. (13)

Learning by assessing and giving feedback

Peer assessment was not appreciated to the same extent as peer feedback, partly because some students thought it was difficult to assess other's work, and partly because they did not see the meaning of it in the beginning. Nevertheless, most of the students appreciated the activity since it gave them greater insight into what it means to assess and a better understanding of the importance of presenting a solution to others.

> You learn by assessing another group. You see other ways of writing it and thinking about it. (1)

> … and by seeing what other people have done then you learn a lot more and then you see some mistakes and you know how to avoid making the same mistake next time. (8)

> Sometimes you don't really know what to say. (13)

> I realized how hard it is to perform assessments. But then I also realized, oh so that's what you should write more. Then it was good. (7)

120

Peer feedback vs. teacher feedback

When asked whether teacher feedback or peer feedback was the most valuable source of feedback, most students answered that they learned more from peer feedback. The teacher feedback came afterwards and contained less information about the task. Still, teacher feedback was desirable because it was perceived as a guarantee of what was right or wrong.

> It always feels a bit more special when it comes from the teacher. …That's probably the only difference I noticed. (15)

> I learnt more from the group's assessment. They explained how they had seen it and how they had solved the problem. I think maybe you learnt more from that. (10)

Working with problem-solving

Students expressed positive feelings about their work with problem-solving tasks, as opposed to working with the (routine) exercises in the textbook. Furthermore, their positive feelings were linked to perceptions of how problem-solving activities contributed to their learning and their view of school mathematics. Students referred to experiences of working in small groups with a mathematical problem, but also to the individual and pair tests where they were confronted with similar problems.

Problem-solving tasks vs. textbook tasks

Working with one extensive problem for a whole lesson, instead of solving several routine tasks in the textbook, was appreciated by the students. Students claimed that working with complex and contextual problems was challenging and fun.

> You had to think more about those tasks. It was a bit more of a challenge. (10)

> It's a bit more of a challenge so maybe that's why I find it more rewarding. (15)

Usefulness of mathematics

Students expressed that the mathematical problems, which were chosen to reflect familiar settings, had showed them the usefulness of mathematical knowledge in everyday life.

> It also makes math feel more meaningful and you can see the use for math in everyday life. (1)

> You learn to apply the math you've learnt to a real situation. It feels more meaningful. (5)

Mathematical learning

Problem-solving activities were appreciated as a valuable preparation for the National test. Students also appreciated the opportunity to apply their knowledge: a process that they claimed gave a whole new perspective, and a deeper understanding, of mathematics.

> Problem-solving is good because you have to think more and use what you've learnt in the whole chapter. You can see a connection. (6)

> It feels like when you have more difficult tasks, you need to have understood what you are supposed to do. (3)

> It was a major task covering one whole chapter, which meant that you learnt about the whole chapter and in that way you were maybe more prepared for the National test. (10)

Mathematics-related beliefs

When asked about how they perceived the new teaching-learning situation, with regard to their beliefs about themselves as learners, of mathematical learning, and of assessment in mathematics, students' experiences were mostly positive.

Students' self-image

When students expressed their feelings regarding their mastery of mathematics, a predominant perception was that learning mathematics had become easier. Students brought up several reasons for this, for example the way of working with mathematics, which made learning more meaningful and more fun.

> I think that math is easier now. (1)

> This way of working has made me more alert and able to think more quickly. Don't know why, it just feels like that. (2)

> …more fun and it feels like I have more use for it [mathematics] because I know what I can use it for in the real world. (16)

> It's more fun working with math now than it was in compulsory school, because it's been easier for me to take part in everything. (3)

Beliefs about mathematical learning

Students' responses indicate that their beliefs about mathematical learning have changed. For instance, a number of students acknowledged that it is important to have an understanding of mathematical activity, as opposed to learn formulas by rote. The students also claimed to have gained a greater insight into school mathematics and how it could be useful for them.

> It's influenced me because I now realize that it's all so much about how you express yourself. Not just getting an answer, instead you have to show how you got that answer. (5)

> [when you work with problem-solving] it doesn't help to just have learnt it by rote, you have to be able to understand it. (3)

> Working with both oral and written assessments have shown me another perspective of math. It's given me more of a feeling of math as a whole rather than just doing arithmetic in the book. (9)

Now we have more examples of how we can use math in every-day life and now it's no longer important to memorize formulas, but rather to be able to use them. (16)

Beliefs about assessment in mathematics
Students expressed that their experiences with the new teaching-learning environment have shown them the importance of feedback and peer assessment for mathematical learning. They also recognize the importance of clear instructions for a better performance.

Feedback is good because then it's someone else who's looked at your work and it's easier for them to see what you've done wrong and what you could do better. You can learn a lot from that. (4)

I think it's quite good to assess other people's work and to think about what's logical and that. (15)

I do better at tests because I think of those [rubrics]. You know how she [the teacher] will assess my work, so that's been a help. I realize that there're several things that get assessed and since you can see other people's errors when you assess their work, it's easier to think more in the right way yourself. (16)

The teacher's perceptions

Data on the teacher's perception of the intervention was collected by the use of a "written interview", which complements the continuous dialogue between teacher and researcher during the two semesters.

In many ways, the intervention represented a new way of thinking about teaching, assessing, and learning mathematics to the teacher. As a consequence, even though the teacher collaborated with the researcher, she had to work hard in order to understand and develop the design needed for a formative-assessment practice. According to the teacher, the intervention truly changed her way of conceptualizing teaching and learning in mathematics. After one semester she summarized her experiences in the "written inter-

view" and her responses were analyzed and categorized in the same way as the students' responses. The categories are the teacher's perception of the assessment system, the emphasis on problem-solving, and students' learning.

The assessment system
The teacher stressed that focusing on formative assessment was extremely stimulating, even if it was quite time consuming initially. On the whole, she admits that it changed her views about instruction in mathematics.

> I have found working in this new way to be very stimulating and instructive. Also, the new way of working has been time-consuming and, to a certain extent, it has increased the workload.
> [But] I think the advantages of working this way outweighs the time it takes.

The combination of all the formative tools: the rubric, the problem-solving tasks, and the peer-assessment activities helped the teacher gain a better understanding of students' knowledge and gave her new ideas about learning in mathematics.

> It's been beneficial and it's given me another view of both the teaching and how you can make math more stimulating for the students.
> …and, I think that I've gotten a better feeling for what the students are actually thinking.

Problem-solving
On the whole, the teacher could observe positive effects on students' learning. The problem solving in small groups was seen as stimulating and enjoyable and as a way to create new possibilities for learning. The teacher noticed that these activities were of particular value for low-achieving students.

> The students have found the problem-solving part fun and it's given them a chance to practice their ability to think logically and practically. This way of working has led to many discussions in math and therefore it also becomes a more sociable way of working. When they work together with the problems in groups they complement each other and often get further than they would have on their own, which, at least, benefits the weaker students.

The teacher observed that, by means of the problem-solving tasks, students were provided with opportunities to use their mathematical knowledge in more varied and more relevant situations, as opposed to only working with the exercises in the textbook.

> It's easy to make them relevant to society, can integrate some physics (my other subject) in them and everyday problems that the students are familiar with are often brought up.

Students' learning

Taken together, the teacher claims that the formative tools have helped students to improve their problem-solving skills, as well as becoming better at structuring and regulating their work. Furthermore, the teacher noticed that the intervention changed students' ways of thinking about mathematics. For example, by discussing mathematical tasks, the students recognized the importance of reasoning in mathematics and the quality of their arguments gradually improved. As a consequence, discussions about whether the solution was "correct or wrong" were replaced by a reasoning about different solutions and interpretations.

> With this way of working I have found that the students have become a lot better, especially at problem-solving. They think differently, have become better at interpreting tasks, but most of all at discussing math and how to reason.

> What's more they've become much better at justifying their solutions and at structuring and planning tasks.

> When they ask about tasks, they almost just as often ask about interpretations and whether their reasoning is good as they ask about whether or not their calculations are correct.

According to the teacher, the students also become more self-conscious and critical about their mathematical knowledge and how it was assessed. The scoring rubric, in particular, helped the students to change focus from merely giving an answer to interpreting and reasoning about various solutions. In combination with peer assessment and feedback, the scoring rubric seemed to help students understand their own strengths and weaknesses.

> Apart from the assessment having made the students more aware of what is assessed, I think they've become slightly more self-critical, but also more self-conscious.

> By assessing their classmates' solutions they have to understand more than their own solution and, at the same time, learn to see from more than one perspective and more ways of thinking and interpreting.

The teacher confirms that the assessment process was initially difficult for the students. However, she noticed that students' assessments successively became more similar to her own assessments, which may be due to the influence of the rubric.

> In the beginning they found this difficult and probably did a lot of guessing, but as we worked with this more and more they have found it easier and easier and now I think that they almost always assess the same way as I do.

DISCUSSION

The aim of this study was to introduce a formative-assessment practice in a mathematics classroom, by implementing the five strategies of the formative-assessment framework proposed by Wiliam and Thompson (2007), in order to investigate: (a) if this change in assessment practices had a positive influence on students' mathematical learning and, if this was the case, (b) which these changes were, and (c) how the teacher and students perceived these changes in relation to the new teaching-learning environment. In this chapter, the findings of the study are discussed and possible lines of future research are outlined.

The influence on students' mathematical learning from the change in assessment practices

In order to answer the research questions, we first need to consider how mathematical learning has been operationalized in this study. Since learning is not directly observable, two specific indicators of learning have been chosen: student performance on problem-solving tasks and students' mathematical beliefs. These indicators are (of course) not perfect mirrors of student learning. The students may have learned a lot of things during the intervention, which are not captured by the tests used to estimate students' performances and beliefs. Or the students may have improved their performance, or changed their beliefs, without learning the things that they were intended to learn. What we can tell from these tests is therefore whether students' performances and beliefs have changed and in what direction – not what students' have actually learnt. Similarly,

we cannot tell for sure whether it is in fact the changes in assessment practices that have influenced students' performances and beliefs, or if they have been influenced by other factors as well. This problem will, however, be discussed in depth later on. For now, we will first continue by discussing whether students' mathematical performance has indeed improved during the intervention and, in that case, which direction these changes have taken. After that, a similar discussion will be held in relation to students' beliefs.

Have students' performances in problem-solving improved?

Over the intervention period, the intervention-group students improved their problem-solving performances both in relation to their own performances prior to the intervention and in relation to the control group. As shown in the results section, the scores on the problem-solving test increased by roughly 10 percent for the intervention group. Compared to the control group, the difference was even greater, since scores in the control group were actually lower on the post-test as compared to the pre-test. That the control group performed less well on the post-test might indicate that this test was more difficult for the students than the pre-test, even though care was taken to select similar tasks[4]. Despite the potential difference in difficulty, however, the students in the intervention group improved their performance.

How have students' performances in problem-solving changed?

The use of a scoring rubric in the assessment of the problem-solving tasks in the pre- and post-tests allowed for a nuanced interpretation of students' results. The findings show that the intervention-group students improved in relation to all of the aspects included in the rubric although, as compared to the control group, the most evident improvement was with regard to how they presented their solutions and how they made use of mathematical symbols.

---

[4] As described previously, the tasks in the pre-test were chosen from the National test for year 9 in compulsory school, while the tasks in the post-test were chosen from the National test for the A-course in upper-secondary school.

Even if the intervention group as a whole performed better on the post-tests, all students in the intervention group did not improve equally much. Instead, it was the "low-achievers" (i.e. the students who performed less well on the pre-test) who improved the most. As compared to the initial performances of the intervention-group students, the most evident improvement for "low-achievers" was with regard to how well they interpreted a problem and used the appropriate mathematical methods, but also with regard to their reasoning about mathematical solutions. For the "high-achievers", improvements could be seen in the clarity and completeness of their presentations of solutions, as well as the appropriateness in their use of mathematical symbols, terminology, and conventions.

Taken together, it can be seen that even though the students in the intervention group improved their performance as a group, the improvement is neither unidimensional in relation to problem solving nor is it homogeneous within the group. Instead, the students made improvements along several different dimensions of problem solving and there seems to be a difference in which dimensions are mostly affected by the formative-assessment practice, depending on students' initial level of achievement. This means, among other things, that a formative-assessment practice does not necessarily favor any particular dimension of performance, but rather that improvement may be facilitated along different dimensions. Students may therefore start their "journey" at different locations, and have different needs in terms of improvement, but still develop in the same direction. This is seen in the findings, where the students who performed less well initially had their greatest improvements in dimensions different from the students who performed well from the beginning.

Have students' mathematical beliefs changed?

The findings of this study indicate that it may take time (two semesters) to observe changes in students' beliefs. For example, the epistemological beliefs, which barely changed after one semester, had changed in a positive direction at the end of the school year. Students' beliefs about assessment in mathematics also showed a

significant change as compared to those of the students in the control group, but only after two semesters. Since these beliefs mainly belong to the cognitive domain, rather than the affective one (Muis, 2004), an explanation for the delay might be that the students needed time to "expand their knowledge base" before their beliefs could change.

As opposed to the cognitively oriented beliefs, students' self-concept in mathematics remained unchanged. This result is surprising and needs further attention. When the correlation between beliefs variables and students' results in the problem-solving test was analyzed, it was noted that among the students who improved their problem-solving performance the most, were also the students who changed their beliefs about mathematics and the self in a positive direction (i.e. towards more availing beliefs). This is an interesting finding because the mean value of the self-concept scale did not change in the intervention group during the intervention. Consequently, the self-concept scores of some other students' must have decreased. This seems paradoxical, however, since nearly all of the students claimed to have experienced increased enjoyment during the intervention, together with feelings of being capable to keep up with instruction, as well as improved understanding and meaningfulness. The students attribute these experiences to the new teaching-learning environment, but at the same time their self-concept scores in the questionnaire do not reflect these experiences. As an example, student (8) says in the interview that:

> Math has become more difficult and more fun. It's more difficult because I'm not used to this way of teaching. But I learn a lot more. And when I understand something it becomes more fun. If I just sit there and don't understand anything it gets boring. It's been more difficult, but it gets easier and therefore more fun.

Despite the positive experiences expressed by the student, on the self-concept scale in the questionnaire this student's score goes from 22 to 21 and then to 17 at the three measuring points. Taking into account that the interview took place between the second and

third administration of the questionnaire, these results may seem confusing. The same discrepancy can be seen for several other students as well and it raises questions about the validity of the self-concept scale in the questionnaire.

When comparing with the other scales in the questionnaire, the interpretations of the mean values of students' responses to these scales are more in line with students' answers in the interviews. This observation indicates that it is not necessarily the function or the construction of the questionnaire that fails, but instead the affective nature of the self-concept beliefs may give this scale low stability and an inclination for changes influenced by external factors (Op't Eynde et al., 2002; Pehkonen, 1995). If these beliefs are indeed sensitive to emotional events, such as receiving test results or grades, then the occurrence of such an event just before either of the data-collection situations, might have affected the results. Unfortunately, events of this kind have not been documented in the study, making it impossible to verify this explanation.

The analysis of the relation between epistemological beliefs and self-concepts in the intervention group further indicates that the questionnaire functions as intended. This analysis shows a positive development over the intervention period, which is in line with results from previous research, stressing that these variables are intrinsically linked (Steiner, 2007). Changes in motivational beliefs follow changes in epistemological beliefs and in achievement, although at a lower rate. The analysis confirms the developmental process undertaken by the students in the intervention group.

In the control group, a different change of students' beliefs was observed. While there was a development towards a stronger correlation between students' performances and their mathematics-related beliefs in the intervention group, the correlation between the same variables developed in the opposite direction in the control group and weakened at the end of the intervention.

How have students' mathematics-related beliefs changed?
Significant changes in students' beliefs about learning mathematics in the intervention group are: (a) the importance of understanding why an answer is correct, in contrast to merely delivering a correct

answer; (b) that there may be different solutions to a problem and that these can be presented in different ways; and (c) that it is necessary to understand the solution and not just to apply methods mechanically in order to solve mathematical problems. Students also (d) recognized the importance of knowledge in mathematics for their future and the usefulness of mathematics in everyday life.

Furthermore, in comparison to the control group, the students in the intervention group to a greater extent recognized the importance of: (a) receiving feedback from the teacher on assessments, (b) being aware of the assessment criteria in mathematics, and (c) performing oral assessments in mathematics. The students in the intervention group also displayed insight into the assessment process and of the importance of variation in the assessment formats.

Even though students' views about themselves as learners barely changed on the questionnaire, students expressed a more positive self-concept in the interviews. These findings indicate a more positive perception of mathematical learning, such as a greater ease in understanding and enjoying working with mathematics.

The formative-assessment practice

So far, improvements in student performance and changes in mathematically-related beliefs, towards more availing beliefs, in the intervention group have been verified. However, in order to understand how the formative-assessment practice contributed to these changes, the teacher's and students' perception of the new teaching-learning environment have to be consulted.

Working with rubrics

In the current study, a scoring rubric was chosen as a way to make goals and criteria explicit and understandable. According to previous research on rubrics, in order to promote student learning through the use of rubrics, time needs to be allocated so that students are thoroughly introduced to the criteria and standards in the rubric. Another finding is that generic rubrics may be more appropriate for formative-assessment practices, since they can be used at several occasions with different tasks addressing the same skills,

even if task-specific rubrics may be easier for students to understand.

Much in line with previous research, the students in the intervention group did experience a deeper understanding of what was expected of them through the use of a rubric. They also expressed improved communication with the teacher, which led to a feeling of security: their "journey" had become more understandable and pleasant. Admittedly, there were initial difficulties in understanding and using the rubric, but the students stated that using this instrument became easier and more meaningful after a while. This is confirmed by the teacher, who observed that the assessment process was difficult for the students in the beginning. However, she also noticed that the students soon developed a "connoisseurship" for what is considered quality in mathematical problem solving. This development of a "sense of quality" can be interpreted as a consequence of internalizing the criteria in the rubric (see e.g. Nicol & MacFarlane-Dick, 2006).

In most studies where rubrics have been used in school settings without an extensive implementation process (thus leaving the interpretation of the formulations in the rubric to the students), only small or partial improvements in student performance have been observed (e.g. Andrade, 2001). Students' "sense of quality" in this study is therefore most probably not entirely a consequence of the formulations in the rubric, but instead builds on experience. By applying the rubric in a variety of different situations, thereby simulating the use of exemplars that has been shown to be effective for student learning in other studies, the students can develop a sense for how quality can be represented in actual student work.

What the present study thus indicates is that generic criteria may be used with success, even if they are abstract, at least if the criteria are thoroughly introduced and used systematically during instruction. In spite of the sometimes elusive formulations in the rubric, the students managed to use the criteria when planning and solving mathematical problems, when monitoring their performance, and when assessing other students' work. This finding is of special interest in the context of the latest curriculum in Sweden, in which

the assessment criteria may be – for a number of reasons – difficult to interpret (see e.g. Lundahl, 2011).

Besides giving the students a deeper understanding of expectations, the use of a scoring rubric also provided the students with a scaffolding structure for how to go about solving – and thinking about – mathematical problems. The regular use of the scoring rubric made the assessment criteria an integral part of instruction in mathematics, which led to new ways of working with mathematics and the majority of students claimed that the use of a rubric influenced the way they tackled mathematical tasks. They said that it helped them structure their work, and present mathematical problems, in ways that was different from before the intervention. This was confirmed by the teacher, who claimed that the students became more self-conscious and critical about their mathematical knowledge and how it was assessed. The scoring rubric, in particular, helped the students to change focus from merely giving an answer, to interpreting and reasoning about various solutions. Furthermore, in combination with peer assessment and feedback, the scoring rubric seemed to help students understand their own strengths and weaknesses.

Problem solving in small groups
In the intervention group, problem-solving tasks had a central position and the choice of working with such tasks was twofold. On the one hand, problem-solving tasks have frequently been argued to be advantageous for students' learning (Lester & Lambdin, 2004) and, on the other hand, problem-solving tasks may give teachers information about students' understanding (i.e. "make learning visible"), which in turn can help them rethink instructions or to get support for what they are already doing (Wiliam, 2007). Still, there are a number of studies that have failed to find empirical support of these assumptions. For example, in a recent meta-analysis of 26 quasi-experimental studies in middle- and high-school mathematics, Robert Slavin et al. (2009) found that the effects of focusing on problem solving were negligible on test performance. The lack of positive effects is explained by the discrepancy between the type of tests used to measure students' perfor-

mance on the one hand, and the focus on problem solving in the teaching process, on the other. It is argued that the traditional test items (such as multiple-choice questions) were not able to capture the more sophisticated skills developed by the students (Schoenfeld, 2006). Studies that report positive effects of using problem-solving tasks in instruction (Mason & Scrivani, 2004; Verschaffel et al., 1999), however, have typically evaluated students' performances by the use of problem-solving tasks instead of traditional test items. One explanation for the positive results in the current study may therefore be methodological, since there was an alignment in instruction and assessment (including the National test in mathematics), both focusing on problem-solving tasks.

Another explanation for the positive results may be the motivating effect of the realistic problems. The problem-solving tasks gave the students an opportunity to apply their knowledge, which was described by the students as challenging, but also as more fun to work with as compared to ordinary textbook exercises – an observation also shared by the teacher.

A possible downside of emphasizing problem solving is that the extra attention to problem solving could potentially "steal" time from other parts of the curriculum, which in turn might result in less well performance on the National test. In this case, however, such negative effects were not observed and students' results at the National test were instead better than the results in the control group. On the other hand, there are no indications of transfer from the problem-solving activities to the other parts of the mathematics curriculum.

The students mainly worked in small groups or pairs, since problem-solving activities in small groups have been shown to create favorable conditions for student learning (e.g. Schoenfeld, 1992), as well as to influence students' beliefs (Yackel et al., 2000). According to the students, working in small groups added to the positive experience, by offering them opportunities to discuss on "equal terms". This was confirmed by the teacher, who noted that the problem-solving activities in small groups were stimulating and enjoyable for the students. By discussing mathematical tasks, the students recognized the importance of reasoning in mathematics

and the quality of their arguments gradually improved. As a consequence, discussions about whether the solution was "correct or wrong" were replaced by reasoning about different solutions and interpretations. The teacher noticed that these activities were of particular value for low-achieving students.

Nonetheless, there are some challenges associated with working in small groups, such as creating well-functioning constellations of students, and some students reported dissatisfaction with the composition of the groups. Obviously, working with others is a delicate matter, especially when being assessed together (Goos et al., 2002; Freeman, 1995). Still, focusing too much on the social interaction of the group, while neglecting the meta-cognitive activities (such as peer assessment), often fails to show positive effects on students' learning (Seidel & Shavelson, 2007). This indicates that, although challenging, the social interactions in group work may not be as vital for student learning as an instruction that provides support through feedback and the regulation of learning.

Focusing on socio-mathematical norms has also been suggested as a way to facilitate student learning. In this case, however, it is difficult to distinguish the effects of fostering such norms, since they were mediated through group assignments, in combination with peer assessment and peer feedback. What seems clear, however, is that students were confronted with the need to identify sophisticated or efficient solutions, as well as acceptable explanations, which are manifestations of the socio-mathematical norms.

Feedback
According to the research previously reviewed, in order to be effective for student learning feedback needs to provide information about where the students are going, how they are doing, and how to move on. In line with this research, feedback delivered by the teacher in the intervention group was designed to provide information about where the students were going, how they were doing, and how to move on. The scoring rubric was used to facilitate the feedback process, since the levels in the rubric can be used to communicate both the current position and the next step. The interviews, along with students' answers in the beliefs questionnaire,

confirmed that the use of a rubric to support the feedback process gave them a better understanding of what they needed to work with in order to improve.

One of the most innovative arrangements may have been the marking of individual tests with mini-rubrics instead of scores. Even if students' performances were still documented in a quantitative manner, this change resulted in nuanced teacher feedback about students' performances. Some students, however, missed being given total scores. This type of summative feedback was something that they were used to and which gave them a feeling of security, even if the scoring with mini-rubrics provided them with more information about their performance.

Besides providing information about where the students were going, how they were doing, and how to move on, feedback should preferably be directed at the task-, process-, or self-regulation level, but not at the self level. Accordingly, the delivery of feedback from the teacher was directed towards the tasks and the work process. The focus was on students' performance in relation to criteria and not on students' diligence or on other personal characteristics.

Another important factor for supporting student learning is the timing of feedback. For instance, there are indications that immediate feedback may be more beneficial when focusing on task level and for difficult tasks, while delayed feedback may be more appropriate for the process level and for tasks perceived as easy by the students.

Unfortunately, a rigorous consideration of the timing of feedback was not permitted by the class schedule. There were situations in which either direct or delayed feedback would have been more appropriate, but due to the timetable students had to leave for another class. In the case of group assignments, where the organization of activities was carefully planned beforehand, this problem was not as pronounced. On such occasions students received extensive feedback from both peers and the teacher.

Last, but not least, the norms established in the classroom culture created a good basis for the receiving of feedback since these norms included the acknowledgement of error and misunderstanding as part of the learning process, and of argumentation as a nec-

essary complement to a mathematical solution. These norms contributed to the students being more receptive to what was identified by the teacher as being either strengths or weaknesses in their work. The findings indicate that the students saw the feedback as opportunities for getting information about their learning progression.

Peer assessment and peer feedback
The intention with using peer assessment and peer-feedback activities was to engage students in supporting each other's learning by exploring and sharing ideas with each other. Such arrangements have been shown to have both cognitive and motivational gains (see Dochy et al., 1999; Gielen et al., 2010).

In the present study, students reported both positive feelings and perceived positive learning effects of being engaged in peer assessment and peer feedback. For instance, in the interviews students referred to the assessment of their peers as a means of deepening their own mathematical understanding. By seeing the different solutions presented by other groups, the students claimed to understand that the same task could be solved in several different ways. Furthermore, students' awareness of the importance of reasoning and communicating in mathematical learning was enhanced by formulating feedback to peers, since this required them to justify their assessment. Students claimed that formulating feedback to peers made them more conscious about of their own strengths and weaknesses. As compared to teacher feedback, students considered peer feedback to be more immediate and more closely related to the task, which made them perceive peer feedback as more helpful for their learning.

Even if most students obviously saw the positive effects of peer assessment and peer feedback, some students initially expressed that they did not like to assess their peers, since they did not see the meaning of it and also experienced it as uncomfortable. However, as the students became more used to this new way of working (cf. Falchikov & Boud, 1989), they began to see the meaning of peer assessment and, eventually, accepted it. On the whole it appears that the activity of peer assessment was experienced as less dra-

matic and became more natural since it was integrated in the teaching-learning environment.

As indicated by previous research, whether peer feedback is accurate seems to be of lesser importance for supporting student learning. Instead, what seems more important is that students justify their assessment and feedback. As a consequence, the accuracy of students' assessment was not judged by the teacher in the intervention group. However, students could compare the peer assessment with the teacher's assessment. This seemed to trigger the students to put more effort into their assessments and explanations, by trying to perform assessments that were as good as the teacher's assessment.

The fact that the students acted as "teachers" themselves is an important aspect of the intervention. Traditionally, the teacher is the one who feeds back information about student performance. However, even if the intervention students appreciated and needed teacher feedback as a confirmation of their progress, they reported that they learnt more from peer feedback since this was more extensive and more directed to the task. The exchange of meaning and understanding that occurred between the students was regarded as a rich source for learning. What distinguishes this discussion from other peer-to-peer discussions is the framework created by the use of a scoring rubric. The rubric provided a structure for the assessment- and feedback process, preventing students' judgments and arguments from becoming arbitrary. Furthermore, the rubric provided the students with a common language to communicate their assessments and their feedback.

Co-assessment
Research has shown that co-assessment can be effective in improving students' learning of both domain-specific and generic skills, as well as increasing student motivation (see Gouli et al., 2010; McConnell, 2002). It is not possible to distinguish the individual effect of the co-assessment activities on students' learning and their motivation, but a specific contribution of co-assessment (or "whole-class discussion" as the students called it) is the collabora-

tion of students and the teacher in the assessment process, which the students claimed added to their learning.

The whole-class discussions provided a forum where students, together with the teacher, could establish a shared language and common norms for working with mathematics. It also made students' ideas available to others, and tangible to themselves, which may increase mathematical learning. These conditions may also help students make connections across the mathematical systems and to see mathematics as a way of explaining, justifying, and reasoning (Lee, 2006). Accordingly, the students in the intervention group claimed that when solutions were presented on the whiteboard, and when they were being verbalized, this facilitated a deeper understanding of mathematics.

A critical point, which has been discussed among teachers, is whether allowing the students to express their thinking might be confusing for the other students. If non-correct ideas are presented, students may have difficulties knowing what is right or wrong. The findings from the present study, however, indicate no such negative experiences.

Conclusions

The implementing of the five strategies of the formative-assessment framework proposed by Wiliam and Thompson (2007) did have a positive influence on students' mathematical learning. Firstly, students' performance improved both in relation to the pre-test and in relation to the control group. The most evident improvement was with regard to how the students in the intervention group presented their solutions and how they made use of mathematical symbols. The "low-achievers" in the intervention group improved the most during the intervention. These students improved their performance with regard to how well they interpreted a problem and used the appropriate mathematical methods, but also with regard to their reasoning about mathematical solutions. The "high-achievers" made improvements in the clarity and completeness of their presentations of solutions, as well as the appropriateness in their use of mathematical symbols, terminology, and conventions. Secondly, students' mathematical beliefs changed during the inter-

vention, towards more availing beliefs. The students' and the teacher's perceptions of the intervention suggest that all of the strategies implemented made contributions towards improved learning. However, the different components also affected and reinforced each other, making the evaluation of each individual strategy impossible. For instance, the scoring rubric did provide transparency to the assessment, but not on its own. Instead, in order to clarify goals and criteria the rubric depended on the peer assessment, peer feedback, and co-assessment, at the same time as these activities all depended on the use of a rubric, since students needed to be familiar with the criteria in order to perform the activities. This relation shows clearly the interdependence of the different components in the intervention. In a similar way, the problem-solving tasks did "make learning visible", but they did so in combination with working in small groups and with a focus on socio-mathematical norms, since the group assignments and the classroom norms made the students explicate their thinking. Furthermore, teacher feedback was dependent upon the rubric, but also on the norms established in the classroom, which made students receptive to the feedback.

Another characteristic of the components of the formative-assessment practice was that they not only provided support for the intended strategy, but tended to have other positive effects as well. For instance, the positive effects of using a scoring rubric were not restricted to clarifying goals and criteria, but also provided the students with a scaffolding structure for how to address problem-solving tasks and besides making students' learning visible, the problem-solving tasks also affected students' motivation positively.

Taken together, this study suggests that a formative-assessment practice, encompassing the five strategies of assessment for learning, may have a positive influence on students' learning in mathematics. Furthermore, the findings suggest that such a teaching-learning environment may also affect students' motivation, making learning in mathematics a fun and enjoyable experience.

Methodological limitations

There are, of course, a number of restrictions to the conclusions above. Firstly, the main performance targeted in this study is problem-solving. This means that there are indications of students improving their skills in problem-solving, but it is not possible to draw any conclusion about other mathematically related skills. However, the intervention did not seem to counteract the learning of other skills, since the students in the intervention group did not perform less well on non-problem-solving tasks on the National test in mathematics. The fact that the intervention-group students did not fail to solve non-problem-solving tasks on the National test, may also be used as an argument against the intervention only being a way to "teach to the test" (i.e. drilling the students to perform well on a specific assessment), since their performance was not restricted to only problem-solving tasks.

Secondly, students' problem-solving skills on the post-test were evaluated with different tasks, as compared to the pre-test. Although care was taken to select similar tasks, the findings indicate that the post-test tasks may have been more difficult for the students to solve, which makes the comparison between pre- and post-tests problematic. A way to avoid this problem would have been to use the same tasks in both tests. In that case, however, students might have recognized the task and – if the students had discussed the solution with each other – the comparison would have been misleading.

Thirdly, the two groups in the study have experienced different instruction, but they were also taught by different teachers. This is problematic, since different teachers may vary very much in teaching proficiency. For instance, Steven Rivkin, Eric Hanushek, and John Kain (2005) have shown that students with an effective teacher can learn as much in six months, as other students with a less effective teacher learn in two years! In the current design, it is not possible to exclude the possibility that differences in teacher proficiency have had an influence on the results, when comparing the two groups. Still, there are also other findings, such as when the students themselves attribute changes in their learning to the components of the new teaching-learning environment. For in-

stance, the students claimed that the systematic use of a rubric helped them to identify the different dimension of mathematical problem-solving and to structure their work.

Fourthly, a number of problems have been encountered when trying to measure students' beliefs. Two aspects have been particularly problematic. First of all, it does not seem to be sufficient to rely solely on a questionnaire. In order to gain a deeper insight into students' beliefs, the questionnaire may need to be complemented with interviews or other qualitative data. Even if such a combination was used in the present study, the focus of the interviews was not to clarify students' answers to the items in the questionnaire, but on retrieving information about students' experiences of the intervention. By explicitly focusing on students' beliefs in the interviews, possible misunderstandings of the items in the questionnaire can be discovered. Additional information about their answers, as well as students' reasoning, can also be attained.

The second problematic aspect of measuring students' beliefs is that some beliefs may be sensitive to external factors. This means that contextual factors have to be considered, for instance so that students do not answer questions about their self-concept directly after receiving test results or grades.

Consequently, care has to be taken when interpreting measures of students' beliefs. In particular, a measure of the affective dimensions may be considered an indicator of students' beliefs at a certain moment, while cognitively-oriented beliefs may be more consistent across time.

Lastly, the mode of assessment may also affect the interpretation of the findings. In the present study, a generic rubric was used in order to assess the different aspects of students' problem-solving skills. Such qualitative assessments cannot be performed with precision, since the criteria have to be interpreted and applied to cases that are not straight forward. This means that although the use of a rubric may facilitate a fine-grained interpretation of students' results and the delivery of nuanced feedback, the assessment is less precise, as compared to assessments of whether students' answers are correct or not.

Lessons learnt

The findings of the study point towards some areas for improvement, when embedding assessment for learning in instruction. Firstly, it appears that the introduction of the assessment criteria (in this case the rubric), has been difficult both for the students and for the teacher. Even if they eventually became familiar with, and learnt to understand and apply, the rubric, the initial difficulties might be discouraging. By providing students and teachers with several exemplars, and by showing them how to interpret the criteria in concrete situations, the introduction period might be shortened.

Secondly, more attention could be given to the organization of specific activities, where students have the possibility to actually use the feedback they receive from either the teacher or from peers. To really use the feedback is important for closing "the feedback loop" (Black et al., 2003). When students process the feedback, and are given the opportunity to act upon it, this could be considered one more step towards realizing the full potential of a formative-assessment practice.

Thirdly, in order to foster student self-regulation, self-assessing activities could be introduced. However, introducing self-assessment needs to be done with care. Before students can become owners of their own learning, they need to be familiar with the assessment criteria. Furthermore, they need to be able to formulate constructive feedback, including actions for improvement. Both of these requirements can be met, as in the study at hand, by working with peer assessment and peer feedback, or with co-assessment.

Lastly, closer attention needs to be given to the assessment literacy of the teacher. In the present study, the researcher spent a considerable amount of time collaborating with the teacher. This, in turn, required a lot of extra time beyond the teacher's regular working hours. Such conditions are not realistic in regular school settings. Previous research (Kirton et al., 2007; Lundahl, 2011; Wiliam et al., 2004) has shown that an extensive introduction to the principles of assessment for learning is an important condition for positive effects. However, introducing teachers can probably not be accomplished merely through participation at lectures. In-

stead, a close tutorial is more likely to be effective. By providing direct support with the classroom work, a supervisor could guide teachers in embedding assessment for learning in their own instruction. A number of teachers could also provide support for each other, much in the same way as the students did in the current study.

Concluding remarks

This thesis started with noticing the proposed decline in Swedish students' mathematical knowledge. The inconsistency between changes in the definition of mathematical competency, which is reflected in the latest school curricula in Sweden, and the way classroom assessment is mainly focused on rote learning, was suggested as one possible explanation for the abovementioned decline. Recent research in assessment suggests that assessment performed with the purpose of supporting students' learning (i.e. "assessment for learning"), can have powerful effects on students' performances. Still, a need for studies was identified, in which the general principles emerging from research are transformed into guidelines for professional practice (Black & Wiliam, 1998b). Such studies are needed in order to embed and sustain the practice of assessment for learning in instruction. Moreover, there exist only a small number of studies investigating formative-assessment practices in mathematics. The current study therefore makes a contribution to this field, by performing an empirical study on a formative-assessment practice in mathematics, where a clear account is given of the actual classroom methods used.

It may not seem surprising that the new teaching-learning environment introduced was shown to have a positive impact on students' performances and their mathematics-related beliefs, considering the amount of research that supports the use of assessment for learning. What may be considered unexpected, however, is the interactions between the different components used. Both the students and the teacher expressed that each of these components contributed to the learning process. Although the specific contribution of each of the components is not possible to evaluate in this study, it is important not only to note that, for instance, the use of rubrics

supports learning, but also to get a detailed description of how rubrics can be used effectively in a mathematical classroom. Furthermore, it is important to note that such a use is strongly connected to how the use of rubrics is combined with activities that "makes learning visible" or allow students to act as resources for each other.

In other words, an important observation is that none of the strategies included in the framework of assessment for learning is best on its own. Instead, all the components of the formative-assessment practice were shown not only to contribute to the positive effects, but to reinforce each other. A more productive question is therefore not which of the strategies to use, but how to combine them and how to adapt them to a specific group of students and to the context of the course. As a consequence, the awareness of the teacher becomes an essential factor, for instance in relation to under which conditions the components of a formative-assessment practice may be used in order to optimize for student learning.

The lessons learnt from this study can be used to advance the knowledge about the use and function of assessment for learning in the mathematical classroom and hopefully this study may function as a model for inspiration for teachers in mathematics, as well as for others with an interest in mathematical education. Furthermore, the study has the ambition to highlight the importance of beliefs held by students about their learning. This was shown to be a difficult variable both to address in instruction and to investigate. Nonetheless, the study indicates that students' beliefs are expressions of how the students experience their learning and (consequently) that students' beliefs are variables that need to be considered to a greater extent in the planning and evaluation of instruction. If teachers are more aware of students' beliefs, they may gain a better understanding of what needs to be adjusted in instruction, including assessment. Such information may contribute to teachers being better prepared to meet students' needs and to understand students' reactions to classroom experiences.

Finally, this study shows that by a systematic use of assessment for learning, improvements can be made on the cognitive, as well

as on the emotional/motivational, level. This study can therefore be a starting point for a change in perspective and practice with regard to how instruction and assessment are orchestrated.

Further research

Two different directions are suggested for further research. The first one concerns the possibility of making the results of this study generalizable, while the other suggests an exploration of how to combine the different components of the formative-assessment practice, in order to optimize the effects on students' learning.

Generalizability

One of the major limitations of this study is the low number of participants. This was a conscious choice from the beginning, in order to make the intervention design and the qualitative data collection possible. However, since the intervention was seen to produce positive effects on students' performances in mathematical problem solving, and on their mathematics-related beliefs, it would be of interest to investigate how consistent these findings are if the intervention were to be implemented on a larger scale. A study that includes a larger number of teachers and students, who work with mathematics in the context of the Swedish school system, would be an interesting and useful area of further research. In such a study, a key condition would be to thoroughly introduce the teachers to the principles of assessment for learning (cf. Kirton et al., 2007; Wiliam et al., 2004). The teacher's mathematical knowledge also plays an important part in how successful the implementation of assessment for learning is (Bennett, 2011).

Another design could be to implement the same intervention in different subjects. Such a design could give indications of important differences in how assessment for learning works with different subjects and in the perception of the teaching-learning environment between students engaged in different subjects. Furthermore, close attention could be given to the subject-specific norms and behaviors.

Deeper understanding

An investigation of how students actually use the different tools in the intervention, could contribute to a better understanding of the processes in assessment for learning. This could be accomplished by investigating what students really do, for instance when they use the rubric or when they negotiate and collaborate with others during peer-assessment activities. In addition to students' actions, students' perceptions of their motivation and learning could also be investigated. Such an approach is in line with the recommendations formulated by Tina Seidel and Richard Shavelson (2007). In their meta-analysis about the effects of teaching on students learning, they highlight the importance of investigating the motivational-affective and learning-process outcomes in addition to the cognitive outcomes. These recommendations build on recent theories of teaching and learning, which characterize learning as a self-regulated and constructive process. Studies conducted in this way, these authors maintain, could refine the process of teaching, in order to optimize for students' learning.

REFERENCES

Andrade, Heidi Goodrich (1999a). The role of instructional rubrics and self-assessment in learning to write: A smorgasbord of findings. Annual meeting of the American Educational Research Association, Montreal, Canada.

Andrade, Heidi Goodrich (1999b). Student self-assessment: At the intersection of metacognition and authentic assessment. Annual meeting of the American Educational Research Association, Montreal, Canada.

Andrade, Heidi Goodrich (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education, 4*(4).

Australian Education Council (1991). *A National Statement on Mathematics for Australian Schools*. Canberra: Australian Education Council.

Barak, Moshe & Mesika, Pnina (2007). Teaching methods for inventive problem-solving in junior high school. *Thinking Skills and Creativity. 2*(1), 19-29.

Bennett, Randy, E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy and Practice 18*(1)*,* 5-25.

Biggs, John (1999). What the student does: teaching for enhanced learning. *Higher Education Research & Development, 18*(1), 57-75.

Black, Paul, Harrison, Christine, Lee, Clare, Marshall, Bethan, & Wiliam, Dylan (2003). *Assessment for learning: putting it into practice.* Berkshire, England: Open University Press.

Black, Paul J., Harrison, Christine, Lee, Clare, Marshall, Bethan, & Wiliam, Dylan (2004). Working inside the black box: Assessment for learning in the classroom*. Phi Delta Kappan, 86*(1), 9-21.

Black, Paul J. & Wiliam, Dylan (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7-73.

Black, Paul J. & Wiliam, Dylan (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80 (2),* 139-148.

Black, Paul & Wiliam, Dylan (2006). Developing a theory of formative assessment. In John Gardner (Ed.), *Assessment and learning* (pp. 81-100). London: Sage.

Black, Paul J. & Wiliam, Dylan (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5-31.

Boekaerts, Monique, Maes, Stan & Karoly, Paul (2005). Self-regulation across domains of applied psychology: Is there an emerging consensus? *Applied Psychology: An International Review, 54*(2), 149-154

Boesen, Jesper (2006). *Assessing Mathematical Creativity.* Doctoral thesis. Umeå: Umeå University, Sweden.

Boesen, Jesper, Lithner, Johan, & Palm, Torulf (2010). The relation between types of assessment tasks and the mathematical reasoning students use. *Educational Studies in Mathematics, 75,* 89-105.

Boud, David (1995). *Enhancing Learning through Self-Assessment.* London: Kogan Page.

Boud, David (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22(2),* 151-167.

Bourdieu, Pierre (1985). The social space and the genesis of groups. *Social Science Information, 24*(2), 195-220.

Boyatzis, Richard E. (1998). *Transforming qualitative information: Thematic analysis and codes.* London: Sage.

Brookhart, Susan M., Andolina, Marissa, Zuza, Megan, & Furman, Rosalie (2004). Minute math: an action research study of student self-assessment. *Educational Studies in Mathematic, 57*(2) 213-227.

Brown, Gavin T. L., Glasswell, Kath, & Harland, Don (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing, 9*, 105-121.

Brown, Gavin T. L. & Hirschfeld, Gerrit H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice, 15*(1), 3-17.

Brownell, William A. (1942). *"Problem Solving" – The Philosophy of Learning, Part II.* The National Society for the Study of Education, Chicago, IL.

Butler, Ruth (1988). Enhancing and undermining intrinsic motivation; the effects of task-involving and ego-involving evaluation on interest and performance, *British Journal of Educational Psychology, 58***,** 1-14.

Butler, Deborah & Winne, Philip (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245-274.

Carnes, Mark T., Cardella, Monica E., & Diefes-Dux, Heidi A. (2010). Progression of Student Solutions over the Course of a Model-Eliciting Activity (MEA). 40th ASEE/IEEE Frontiers in Education Conference, Washington DC, USA.

Cho, Kwangsu & MacArthur, Charles (2010). Student revision with peer and expert reviewing. *Learning and Instruction, 20,* 328-338.

Cho, Kwangsu & Schunn, Christian D. (2007). Scaffolded writing and rewriting in the discipline. *Computers and Education, 48*, 409-426.

Dahl, Bettina (2004). Analysing cognitive learning processes through group interviews of successful high school pupils: development and use of a model. *Educational Studies in Mathematics, 56*, 129-155.

Dawkins, Paul C. (2009). *Non-Traditional Socio-Mathematical Norms in Undergraduate Real Analysis.* The University of Texsas at Arlington, TX.

De Corte, Erik (2004). Mainstreams and perspectives in research on learning (mathematics) from instruction. *Applied Psychology: An International Review, 53*(2), 279-310.

Dochy, Filip, Segers, Mien, & Sluijsmans, Dominique (1999). The use of self-, peer and co-assessment in higher education: a review. *Studies in Higher Education, 24(3),* 331-350.

Dunn, Karee & Mulvenon, Sean (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research and Evaluation, 14*(7).

Dysthe, Olga (2008). The challenges of assessment in a new learning culture. In Anton Havnes & Liz McDowell (Eds.) *Balancing Dilemmas in Assessment and Learning in Contemporary Education (pp. 15-28).* New York, NY: Routledge.

Elawar, Maria C. & Corno, Lyn (1985). A factorial experiment in teachers' written feedback on student homework: Changing teacher behavior a little rather than a lot. *Journal of Educational Psychology, 77*(2), 162-173.

Entwistle, Noel J. & Entwistle, Abigail (1991). Contrasting forms of understanding for degree examination: The student experience and its implication. *Higher Education, 22*(3), 205-227.

Entwistle, Noel & Ramsden, Paul (1983). *Understanding student learning.* London: Croom Helm.

Ernest, Paul (1991). *The philosophy of mathematics education.* London: Falmer Press.

Even, Ruhama (2005). Using assessment to inform instructional decisions: How hard can it be? *Mathematics Education Research Journal, 17*(3), 45-61.

152

Falchikov, Nancy (2001). *Learning together: Peer tutoring in higher education*. London: Routledge Falmer.

Falchikov, Nancy & Boud, David (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59,* 95-430.

Fennema, Elisabeth & Sherman, Julia A. (1976). *Fennema-Sherman mathematics attitudes scales.* Instrument designed to measure attitudes towards the learning of mathematics by woman and man. Madison, WI: Wisconsin Center for Educational Research.

Flick, Uwe (2002). *An Introduction to Qualitative Research*. London: Sage.

Frechtling, Joy & Sharp, Laure (1997). *User-Friendly Handbook for Mixed Method Evaluations.* Arlington, VA: Diane Publishing.

Frederiksen, John & Collins, Allan (1989). A systems approach to educational testing. *Educational Researcher, 18*, 27-32.

Freeman, Mark (1995). Peer assessment by groups of group work. *Assessment & Evaluation in Higher Education, 20*(3).

Furinghetti, Fulvia & Pehkonen, Erkki (2002). Rethinking characterizations of beliefs. In Gilah Leder, Erkki Pehkonen, & Günter Törner (Eds.), *Beliefs: A hidden variable in mathematics education?* (pp. 39-58). Dordrecht: Kluwer Academic Publication.

George, Darren & Mallery, Paul (2003). *SPSS for Windows step by step: A simple guide and reference.* 11.0 update (4th ed.). Boston: Allyn & Bacon.

Gielen, Sarah, Peeters, Elien, Dochy, Filip, Onghena, Patrick, & Stryven, Katrien (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*, 304-315.

Gijbels, David & Dochy, Filip (2006). Students' assessment preferences and approaches to learning: can formative assessment make a difference? *Educational Studies, 32*, 399-409.

Gijbels, David, Segers, Mien, & Struyf, Elke (2008). Constructivist learning environments and the (im)possibility to change students' perceptions of assessment demands and approaches to learning. *Instructional Science*, *36,* 431-443.

Gipps, Caroline, Hargreaves, Eleanore, & Mccallum, Bet (2001). *What makes a good primary school teacher? Expert classroom strategies*. London: Taylor & Francis.

Good, Thomas, & Brophy, Jere (2003). *Looking in classrooms*. Boston, MA: Allyn and Bacon.

Goos, Merrilyn, Galbraith, Peter, & Renshaw, Peter (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Educational Studies in Mathematics 49,* 193-223.

Gouli, Evangelina, Gogoulou, Agorista, & Grigoriadou, Maria (2010). *Supporting Self- Peer- and Collaborative -Assessment through a Web-based Environment.* Department of Informatics & Telecommunications, University of Athens, Greece.

Green, Thomas (1971). *The activities of teaching.* New York, NY: McGraw-Hill.

Green, Rosemary, & Bowser, Mary (2006). Observations from the field: Sharing a literature review rubric. *Journal of Library Administration, 45*, 185-202.

Hall, Kathy (1995). Co-assessment: participation of students with staff in the assessment process. A report of work in progress, paper given at the 2nd European Electronic Conference on Assessment and Evaluation, European Academic & Research Network (EARN):

http://listserv.surfnet.nl/archives/earli-ae.html.

Hanrahan, Stephanie J. & Issacs, Geoff (2001). Assessing self- and peer-assessment: the students' views. *Higher Education Research & Development, 20(1),* 53-70.

Hattie, John (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* London and New York: Routledge.

Hattie, John & Timperley, Helen (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Havnes, Anton & McDowell, Liz (2008). *Balancing dilemmas in assessment and learning in contemporary education.* New York, NY: Routledge.

Hembree, Ray (1992). Experiments and relational studies in problem solving: A meta-analysis. *Journal for Research in Mathematics Education, 23*(3), 242-273.

Hershkowitz, Rina & Schwarz, Baruch (1999). Reflective processes in a mathematics classroom with a rich learning environment. *Cognition and Instruction, 17*, 66-91.

Higgins, Karen (1997). Effects of a year-long instruction in mathematical problem solving on middle-school students' attitudes, beliefs and abilities. *Journal of Experimental Education 66*(1)*,* 5-24.

Hohn, Robert & Frey, Bruce (2002). Heuristic training and performance in elementary mathematical problem solving. *Journal of Educational Research, 95*(6), 374-380.

Kagan, Spencer (1992). *Cooperative learning.* San Juan Capistrano, CA: Resources for Teachers, Inc.

Kilpatrick, Jeremy, Swafford Jane, & Findell Bradford (2001) *Adding It Up: Helping Children Learn Mathematics.* Washington, DC: National Academy Press.

Kim, Minjeong (2005). The effects of the assessor and assessee's roles on pre-service teachers' metacognitive awareness, performance, and attitude in a technology- related design task. Unpublished doctoral dissertation, Florida State University, Tallahassee, FL.

Kirton, Alison, Hallam, Susan, Peffers, Jack, Robertson, Pamela, & Stobart, Gordon (2007). Revolution, evolution or a Trojan horse? Piloting assessment for learning in some Scottish primary schools. *British Educational Research Journal, 33*(4), 605-627.

Kjellström, Katarina & Pettersson, Astrid (2005). The curriculum's view of knowledge transferred to national tests in mathematics in Sweden. *ZDM The International Journal on Mathematics Education, 37*(4), 308-316.

Kloosterman, Peter & Stage, Frances K. (1992). Measuring beliefs about mathematical problem-solving. *School Science and Mathematics*, *92*(3), 109-115.

Kluger, Avraham N. & De Nisi, Angelo (1996). The effects of feedback interactions on performances: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2)*,* 254-284.

Koichu, Boris, Berman, Abraham, & Moore, Michael (2007). Heuristic literacy development and its relation to mathematical achievement of middle school students. *Instructional Science, 35*, 99-139.

Korp, Helena (2011). *Kunskapsbedömning – hur, vad och varför* [Assessment – how, what, and why]. Stockholm: Skolverket.

Kvale, Steiner (1997). *Den kvalitativa forskningsintervjun* [The Qualitative Research Interview].Lund: Studentlitteratur

Jönsson Anders (2008). *Educative assessment for/of teacher competency.* Doctoral dissertation. Malmö: Malmö University, Sweden.

Jönsson, Anders (2010). The use of transparency in the "Interactive examination" for student teachers. *Assessment in Education: Principles, Policy & Practice*, *17*, 185-199.

Jönsson, Anders (2011). *Lärande bedömning.* [Educative assessment]. Malmö: Gleerups.

Jönsson, Anders (2013). Facilitating productive use of feedback in higher education, *Active Learning in Higher Education*, *14*.

Jönsson, Anders & Svingby, Gunilla (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*, 130-144.

Lauf, Lorraine & Dole, Shelley (2010). Assessment for learning tasks and the peer assessment process. In L. Sparrow, B. Kissane, & C. Hurst (Eds.), *Shaping the future of mathematics education: Proceedings of the 33rd annu-*

al conference of the Mathematics Education Research group of Australia. Freemantle: MERGA, Australia.

Leder, Gilah, Pehkonen, Erkki, & Törner, Günter (2002). *Beliefs: A hidden variable in mathematics education?* Dordrecht: Kluwer Academic Publication.

Lee, Clare (2009). *Language for Learning Mathematics: Assessment for Learning in Practice*. New York, NY: Open University Press.

Lehman, Michael L. (1995). Assessing mathematics performance assessment. *National Center for Research on Teacher Learning, paper* 9, 5-7.

Lester, Frank K. & Lambdin, Diana V. (2004). Teaching mathematics through problem solving. In Barbara Clarke, Doug` M. Clarke, Göran Emanuelsson, Bengt Johansson, Diana V. Lambdin, Frank K. Lester, Anders Wallby, & Karin Wallby (Eds.), *International perspectives in learning and teaching mathematics* (pp. 189-207). Göteborg University, Sweden.

Lindberg, Viveca (2007). Långtgående slutsatser trots få lärare och elever i studierna [Far-reaching conclusions, although few teachers and students in the studies]. In Agneta Petterson (Ed.). *Sporre eller otyg – om bedömning och betyg* [Spur or nuisance – about assessment and grading] (pp. 131-153). Malmö: Elanders Berling.

Lingefjärd, Thomas & Holmquist, Mikael (2005) To assess students' attitudes, skills and competencies in mathematical modeling. *Teaching Mathematics and Its Applications, 24*(2-3), 123-133.

Liu, Ngar-Fun & Carless, David (2006). Peer feedback: the learning element of peer assessment. *Studies in Higher Education, 11(3)* 279-290.

Lundahl, Christian (2011). *Bedömning för lärande* [Assessment for Learning]. Stockholm: Norstedts.

Malmivuori, Marja-Liisa (2001). *The dynamics of affect, cognition, and social environment in the regulation of personal learning processes: The case of mathematics*. Helsinki University, Finland: Research Report 172.

Marsh, Herbert W. (1989). *The self-description questionnaire (SDQ) III manual*. The University of Western Sydney, Australia.

Marton, Ference & Booth, Shirley (1997). *Om lärande* [About learning]. Lund: Studentlitteratur.

Marton, Ference & Säljö, Roger (1984). Approaches to learning. In Ference Marton, Dai Hounsell, & Noel Entwistle (Eds.), *The Experience of Learning* (pp. 39-48). Edinburgh: Scottish Academic Press.

Mason, Lucia & Scrivani, Luisa. (2004). Enhancing students' mathematical beliefs: an intervention study. *Learning and Instruction, 14*, 153-176.

McConnell, David (2002). The experience of collaborative assessment in e-learning. *Studies in Continuing Education, 24*(1), 74-92.

McGatha, Maggie B. & Darcy, Peg (2010). Rubrics at play. *Mathematics Teaching in the Middle School. 15*(6), 329-336.

Meier, Sherry L., Rich, Beverly, S., & Cady, Joann (2006). Teachers' use of rubrics to score non-traditional tasks: factors related to the discrepancies in scoring. *Assessment in Education: Principles, Policy and Practice, 13*(1), 69-95.

Muis, Krista R. (2004). Personal epistemology and mathematics: A critical review and synthesis of research. *Review of Educational Research, 74*(3), 317-377.

National Council of Teachers of Mathematics (2000). *Curriculum and Evaluation Standards for School Mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Natriello, Gary (1987). The impact of evaluation processes on students. *Educational Psychologist, 22*(2), 155-175.

Nicol, David (2010). From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment and Evaluation in Higher Education, 35*(5), 501-517.

Nicol, David J. & MacFarlane-Dick, Debra (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199-218.

Nijhuis, Ian F. H., Segers, Mien, & Gijselares Wim H. (2005). Influence of redesigning a learning environment on student perceptions and learning strategies. *Learning Environments Research, 8*, 67-93.

Norusis, Marija (2008). *SPSS 16.0 Guide to Data Analysis.* New York, NY: Prentice Hall Press.

Nyquist, Jeffrey (2003). The benefits of reconstruing feedback as a larger system of formative assessment: a meta-analysis. Unpublished master's thesis, Vanderbilt University, Nashville Tennessee, TN.

Op't Eynde, Peter, De Corte, Erik, & Verschaffel, Lieven (2002). Framing students mathematics-related beliefs. A quest for conceptual clarity and a comprehensive categorization. In Gilah Leder, Erkki Pehkonen, & Günter Törner (Eds.), *Beliefs: A hidden variable in mathematics education?* (pp. 13-38). Dordrecht: Kluwer Academic Publication.

Op't Eynde, Peter, De Corte, Erik, & Verschaffel, Lieven (2006). Epistemic dimensions of students' mathematics-related belief systems. *International Journal of Educational Research, 45*(1-2), 57-70.

Orpen, Christopher (1982) Student versus lecturer assessment of learning: a research note. *Higher Education, 11*, 567-572.

Orsmond, Paul, Merry, Stephen, & Reiling, Kevin. (2000). The use of student derived marking criteria in peer and self-assessment, *Assessment & Evaluation in Higher Education, 25*(1), 23 28.

Orsmond, Paul, Merry, Stephen, & Reiling, Kevin (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education, 27*(4), 309 323.

Panaoura, Areti & Philippou, George (2005). The measurement of young pupils' metacognitive ability in mathematics: the case of self-representation and self-evaluation. *Proceedings of CERME 4*, Sant Feliu de Guíxols, Spain.

Patton, Michael Q. (1990). *Qualitative Evaluation and Research Methods* (2nd ed.). Newbury Park, CA: Sage Publications.

Pehkonen, Erkki (1995). *Pupils`views of mathematics: Initial report for an international comparison project.* Research Report 152. Department of Teacher Education, University of Helsinki, Finland.

Pehkonen, Erkki (2001). A hidden regulating factor in mathematics classroom: Mathematics-related beliefs. In Maija Ahtee, Ole Björkquist, Erkki Pehkonen, & Virpi Vatanen (Eds.), *Research on Mathematics and Science Education: From Beliefs to Cognition, from Problem Solving to Understanding* (pp. 11-35*).* Jyvaskyla University, Finland.

Perlman, Carol C. (2003). Performance assessment: Designing appropriate performance tasks and scoring rubrics. In Janet E. Wall & Gerry R. Walz (Eds.), *Measuring up: Assessment Issues for Teachers, Counselors, and Administrators.* Greensboro, NC: ERIC Counseling and Student Services Clearinghouse.

Polya, George (1957). How *To Solve It – A New Aspect Of Mathematical Method.* Princeton University, Princeton.

Reddy, Malini & Andrade, Heidi (2010). A review of rubric use in higher education. *Assessment & Education in Higher Education 35*(4), 435-448.

Reitmeier, Cheryll A., Svendsen, Lisbet. K., & Vrchota, Denise A. (2004). Improving oral communication skills of students in food science courses. *Journal of Food Science Education,3,* 15-20.

Rivkin, Steven, Hanushek, Eric, & Kain, John (2005). Teachers, schools and academic achievement. *Econometrica, 73*(2), 417-458.

Sadler, Royce D. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119-144.

Sadler, Royce (2002). Learning dispositions: Can we really assess them?. *Assessment in Education: Principles, Policy and Practice, 9*(1), 45-51.

Sambell, Kay & McDowell, Liz (1998). The construction of the hidden curriculum: messages and meaning in the assessment of student learning. *Assessment and Evaluation in Higher Education, 23*(4), 391-402.

Sambell, Kay, McDowell, Liz, & Brown, Sally (1997). "But is it Fair?": An exploratory study of students' perceptions of the consequential validity of assessment. *Studies in Educational Evaluation, 23*(4), 349-371.

Saxe, Geoffrey, B., Gearhart,Meryl, Franke, Megan L., Howard, Sharon, & Crockett, Michele (1999). Teachers' shifting assessment practices in the context of educational reform in mathematics. *Teaching and Teacher Education, 15,* 85-105.

Schafer, William, D., Swanson, Gwenyth, Bene, Nancy, & Newberry, George (1999). Effects of teacher knowledge of rubrics on student achievement in four content areas. Paper presented at the Annual Meeting of the American Educational Research Association. Montreal, Canada.

Schoenfeld, Alan (1985). *Mathematical Problem Solving.* New York, NY: Academic Press.

Schoenfeld, Alan (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In Douglas Grouws, (Ed.), *The handbook of research on the teaching and learning of mathematics* (pp. 334-370). New York, NY: Macmillan.

Schoenfeld, Alan H. (2006). What doesn't work: The challenge and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher, 35*(2), 13-21.

Scriven, Michael (1967). The methodology of evaluation. In Ralph W. Tyler, Robert M. Gagné, & Michael Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago, IL: Rand McNally.

Segers, Mien, Dochy, Filip, & Cascallar, Eduardo (2003). *Optimizing new modes of assessment: In search of qualities and standards.* Dordrecht: Kluwer Academic Publishers.

Seidel, Tina & Shavelson, Richard J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4)*,* 454-499.

Shepard, Lorrie A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.

Shute, Valerie J. (2008). Focus on formative feedback. *Review of Educational Research, 78(1),* 153-189.

Skolverket (2004). *PISA 2003. Svenska femtonåringars kunskaper och attityder i ett internationellt perspektiv* [Swedish fifteen-year olds' knowledge and attitudes in an international perspective]. Stockholm: Skolverket.

Skolverket (2010a). *Gymnasieskolans kursprov vårterminen 2009* [The national tests in upper-secondary school, spring semester 2009]. Stockholm: Skolverket.

Skolverket (2010b). *Svenska elevers kunskaper i TIMSS Advanced 2008 och 1995. Rapport 336* [Swedish students' results at TIMSS Advanced 2008 and 1995]. Stockholm: Skolverket.

Slavin, Robert E. (1995). *Cooperative learning: Theory, research and practice* (2nd ed.). Boston, MA: Allyn & Bacon.

Slavin, Robert E., Lake, Cynthia, & Groff, Cynthia (2009). Effective programs in middle and high school mathematics: Best-evidence synthesis. *Review of Educational Research, 79(2)*, 839-911.

Sluijsmans, Dominique, Dochy, Filip, & Moerkerke, George (1999). Creating a learning environment by using self-, peer- and co-assessment. *Learning Environments Research, 1,* 293-319.

Sluijsmans, Dominique, Brand-Gruwel, Saskia, & Van Merriënboer, Jeroen (2002). Peer assessment training in teacher education: effects on performance and perceptions. *Assessment and Evaluation in Higher Education, 27*, 443-454.

Stanic, George & Kilpatick, Jeremy (1989). Historical perspectives on problem solving in the mathematics curriculum. In R. Charles & E. Silver (Eds.) *The teaching and assessing of mathematical problem solving* (pp. 1-22). Reston, VA: National Council of Teachers of Mathematics.

Stefani, Lorraine A. J. (1992). Comparison of collaborative, self, peer, and tutor assessment in a biochemistry practical. *Biochemical Education, 20*, 148-151.

Stefani, Lorraine A. J. (1998). Assessment in partnership with learners. *Assessment & Evaluation in Higher Education, 23*(4), 339-351.

Stein, Mary K., Boaler, Jo, & Silver, Edward A. (2003). What does the research say? In Harold L. Schoen & Randall I. Charles (Eds.), *Teaching mathematics through problem solving: "It's about learning mathematics."* (pp. 245–256). Reston, VA: National Council of teachers of mathematics.

Steiner, Lorraine (2007). The Effect of Personal and Epistemological Beliefs on Performance in a College Developmental Mathematics Class. Unpublished doctoral dissertation. College of Education, Kansas State University, KA.

Struyven, Katrien, Dochy, Filip & Janssens, Steven (2003). Student's perceptions about new modes of assessment in higher education: a review. In Mien Segers, Filip Dochy, & Eduardo Cascallar (Eds.), *Optimizing new modes of assessment: In search of qualities and standards* (pp. 171-223). Dordrecht: Kluwer Academic Publishers.

Struyven, Katrien, Dochy, Filip, Janssens, Steven, & Gielen, Sarah (2006). On the dynamics of students' approaches to learning: The effects of the teaching/learning environment. *Learning and Instruction, 16*(4), 279-294.

Summit, Raymond & Venables, Anne (2011). The use of co-assessment in motivating tutorial attendance, preparation and participation in engineering subjects. *Australian Journal of Engineering Education, 17*(2), 79-91.

Swedish National Agency for Education (2000). *Kursplaner och betygskriterier* [Syllabuses and grading criteria]. Stockholm: Skolverket.

Tanner, Howard & Jones, Sonia (1994). Using peer-and self-assessment to develop modeling skills with students aged 11 to 16: a socio constructive view. *Educational Studies in Mathematics, 27*(4), 413-431.

Tatsis, Konstatinos & Koleza, Eugenia (2008). Social and sociomathematical norms in collaborative problem solving. *European Journal of Teacher Education, 31*(1), 89-100.

Thompson, Burt (1998). Metamemory accuracy: Effects of feedback and the stability of individual differences. *American Journal of Psychology, 111*(1), 33-42.

Topping, Keith (1998). Peer assessment between students in colleges and universities. *Review of Educational Research 6,* 249-276.

Topping, Keith J. (2009). Peer assessment. *Theory Into Practice, 48*(1), 20-27.

Trigwell, Keith & Prosser, Michael (1991). Relating approaches to study and the quality of learning outcomes at the course level. *British Journal of Educational Psychology, 61,* 265-275.

Van Zundert, Marjo, Sluijsmans, Dominique, & Van Merriënboer, Jeroen (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20,* 270-279.

Venables, Anne & Summit, Raymond (2003). Enhancing scientific essay writing using peer assessment. *Innovations in Education and Teaching International, 40,* 281-290.

Verschaffel, Lieven, De Corte, Erik, Lasure, Sabien, Van Vaerenbergh, Griet, Bogaerts, Hedwig, & Ratinckx, Elie (1999). Learning to solve mathematical application problems: A design experiment with fifth grades. *Mathematical Teaching and Learning, 1*(3), 195-229.

Vickerman, Philip (2009). Student perspectives on formative peer assessment: an attempt to deepen learning?. *Assessment and Evaluation in Higher Education, 34*(2), 221-230.

Vygotsky, Lev (1978). *The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Webb, Noreen M. (2007). Teacher practices and small-group dynamics in cooperative learning classrooms. In Robyn M. Gillies, Adrian Ashman & Jan

Terwel (Eds.), *The Teacher's Role in Implementing Cooperative Learning in the Classroom.* (pp. 199-219). New York, NY: Springer.

White, Mary Alice (1971). The view from the student's desk. In Melvin L. Silberman (Ed.), *The experience of schooling* (pp. 337-345). New York, NY: Rinehart and Winston.

White, Barbara & Frederiksen, John (1998). Inquiry, modeling and metacognition: Making science accessible to all students. *Cognition and Instruction 6*(1). 3-118.

Wiggins, Grant (1998). *Educative Assessment.* San Francisco, CA: Jossey-Bass.

Wiiand, Towe (1998). *Examinationen i fokus. Högskolestudenters lärande och examination – en litteraturöversikt* [Focusing on the examination. University students' learning and examination – a literature review]. Rapportserie från enheten för utveckling och utvärdering. Rapport nr. 14. Uppsala: Uppsala universitet, Sweden.

Wiliam, Dylan (2006). Formative assessment: getting the focus right. *Educational Assessment, 11*(3-4)*,* 283-289.

Wiliam, Dylan (2007). Keeping learning on track. In Frank Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 1053-1098). Charlotte, NC: Information Age.

Wiliam, Dylan (2011). What is assessment for learning? *Studies in Educational Evaluation, 37,* 3-14.

Wiliam, Dylan, Lee, Clare, Harrison, Christine, & Black, Paul (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education, 11*(1)*,* 49-65.

Wiliam, Dylan & Thompson, Marnie (2007). Integrating assessment with learning: What will it take to make it work? In Carol A. Dwyer (Ed.), *The Future of Assessment* (pp. 53-82). New York, NY: Routledge.

Wood, Terry & Sellers, Patricia (1997). Deepening the analysis: longitudinal assessment of a problem-centered mathematics program. *Journal for Research in Mathematics Education, 28*(2), 163-186.

Yackel, Erna & Cobb, Paul (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education,* 27, 458-477.

Yackel, Erna & Rasmussen, Chris (2002). Beliefs and norms in the mathematics classroom. In Gilah Leder, Erkki Pehkonen, & Günter Törner (Eds.), *Beliefs*: *A hidden variable in mathematics education*? (pp. 313-330). Dordrecht: Kluwer Academic Publication.

Yackel, Erna, Rasmussen, Chris & King, Karen (2000). Social and sociomathematical norms in an advanced undergraduate mathematics course. *The Journal of Mathematical Behaviour, 19* (3), 275-287.

Yang, Miao, Badger, Richard, & Yu, Zhen (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15*, 179-200.

Zerr, Jessica M. & Zerr Ryan J. (2011). Learning from their mistakes: using students' incorrect proofs as a pedagogical tool. *PRIMUS, 21*(6)*,* 530-544.

Zimmerman, Barry J. & Martinez-Pons, Manuel (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal, 23*, 614-628.

Zimmerman, Barry J. & Martinez-Pons, Manuel (1988). Construct validation of a strategy model of student selfregulated learning. *Journal of Educational Psychology, 80*, 284-290.

APPENDICES

Appendix 1. Beliefs questionnaire

Epistemilogical beliefs
Time
–Maths problems don't take much time to solve. (excluded)
–If I don't manage to solve a maths problem in five minutes then I won't manage to solve it at all.
+ Understanding maths can sometimes take a long time. (excluded)
+I can solve difficult maths problems if I devote a lot of time to them.

Strategies
–To solve mathematical problems all you need to do is learn the method. (excluded)
–It's important to be able to memorise in order to solve problems in maths. (excluded)
+Maths problems can be solved in different ways.
+Maths problems can be solved with logics and common sense instead of with rules and methods. (excluded)

Understanding
-When I solve a maths task it's enough for me to get the answer right; I don't need to understand why the answer's correct.
+ A person who doesn't understand why an answer is correct hasn't really solved the problem. (excluded)
- Understanding maths means that you learn about methods.
+It's important to investigate whether or not the solution has been correctly worked out even if the answer's right.

Usage
+Maths is good to know for the future.

-Studying maths is a waste of time.
-I'm only studying maths because it's important for my further studies.
+It's useful to be able to do maths.

Beliefs about assessment
Instrument
+ I want oral tests in maths.
+I like tests that contain tasks where I have to reason. (excluded)
-I don't want to be assessed in a group in maths. (excluded)
+I want to be assessed in other ways than by tests in maths.

Fairness
+It's important for me to get feedback from the teacher on a maths test.
+It's important for me to understand the criteria for assessment in maths.
+ I appreciate when maths tests include tasks that I can recognise from my daily life.
+ A maths test should contain different types of tasks so that all skills are tested.

Strategies for repetition
- I try learning different methods off by heart before a maths test. (excluded)
-I only swot for the parts that I know the teacher will grade. (excluded)
+ For me, it's important to understand the whole picture when I swot for a maths test.
-I can pass a test even if I only know a part of the course contents. (excluded)

Mathematical self-concept
+Maths is easy for me to understand.
+ I've always done well in maths.
- I feel inadequate when I don't understand maths.
+ Other people come to me for help in maths.
+ I think that maths problems can be interesting and challenging.
-I have difficulties understanding maths.
-I think that you have to have a special aptitude in order to be good at maths. (excluded)
-I avoid anything to do with maths.

Appendix 2. The three tasks used in the pre test
Task no. 1 - Circumference
In this task you are to work with four different geometrical figures.
All of the figures must have the same circumference, 12 cm.
You are to work with the following geometrical figures:
- a rectangle, in which the length is twice as long as the breadth,
- a square,
- an equilateral triangle, and
- a circle.

You shall investigate and compare the different areas.
What conclusions can you make?

In assessing your work the teacher will take into consideration:
- how correctly and clearly you have drawn your figures
- whether or not you have calculated correctly
- how well you have accounted for your calculations and methods
- how well you have motivated your conclusions.

Task no. 2 – Assembly halls
When assessing your work the teacher will consider the following:
- which mathematical skills you have demonstrated,
- how well you have presented your work,
- which descriptions and which conclusions you have reached.

I   An assembly hall is to be built in the new school in which the first row has 10 seats and the second row has 13 seats. Row 3 has 16 seats and so each row continues to increase by 3 seats all the way to the last row which has 31 seats.
- a) How many seats are there on row 6?
- b) How many rows are there in the hall?
- c) Describe in words or with a formula how you calculate the number of seats in row n.

II   In another assembly hall the number of seats in row n can be calculated by using the formula $12 + 5n$. Describe how this hall is organised.

III   Kalle claims that one can always calculate the total number of seats in an assembly hall, that has been built in the same way, by multiplying the number of seats in the middle row by the number of rows.
Work out whether or not Kalle is right.

Task no. 3 - Currency exchange
When you exchange money to another currency the exchange office charges for its services.
Some exchange offices take out the premium by providing a lower exchange rate while others charge an exchange premium.

You are going to exchange Swedish kronor (SEK) for British pounds (£) and you naturally want to receive as many pounds as possible. You compare three different exchange offices with each other.

Exchange office A
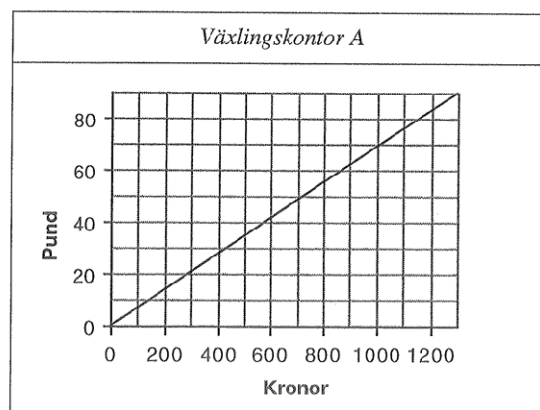
Exchange office B	Exchange office C
100 SEK = £ 7.7	100 SEK = £ 8.6
Exchange premium 5 % of	Exchange premium £ 5 per exchange
the exchanged amount.	occasion.

- You have 500 Swedish kronor. These are to be exchanged into pounds. Which exchange office gives you the most pounds? Account for your calculations.
- Is this exchange office always the best alternative no matter how much money you need to exchange? Find out when you should exchange at each office respectively.   (6/7)

When assessing your work the teacher will consider the following:
- which mathematical knowledge you have demonstrated,
- how well you have accounted for and carried out your calculations,
- how well you have motivated your conclusions.

168

Appendix 3. The three tasks used in the post test

Task no. 1 - The inheritance

When assessing your work the teacher will consider the following:

- how well you have presented your work,
- which methods you have used when you have compared the various alternatives, and
- which conclusions you have arrived at as well as how well you have motivated these.

Robert has a rich aunt. She wrote this letter to Robert:

Sifferby,
6 January 1999

Dear Robert!

Alas, time passes and I'm getting on in age (I still feel healthy and alert, but I've nonetheless just turned 75, as you know). I've been thinking of giving you some of my savings. I'll put aside a sum of money to you every year, and will start with this in January, 2000. You can choose among these alternatives which one you would prefer me to use.

A. SEK 550 on 1 January 2000 and thereafter SEK 550 on 1 January every year, etc.

B. SEK 1,000 on 1 January 2000, SEK 900 on 1 January the year after, SEK 800 1 January the year after that, etc.

C. A lump-sum of SEK 2000. You will receive an annual interest of 11 % from 1 January 2000.

Naturally, this is only for as long as I'm still alive. The money will be paid to you at the time of my decease. I look forward to hearing from you and seeing which of these alternatives you would prefer and why.

With love from,
Aunt Hulda

- Investigate and compare the different alternatives depending on how long Aunt Hulda will live. Suggest which alternative Robert should choose. Also, motivate why you chose this alternative.
- Describe in words or a formula the connection in Alternative A between the sum Robert receives and the number of years.

- Describe with words or a formula the connection in Alternative C between the sum Robert receives and the number of years.

Task no. 2 - The bed stairs

Moa and Martin are going to build a stairway up to the bed (see the illustration)
   a) They decide that the height between the steps should be 20 cm. How many steps do they need to saw?
   b) Decide the distance out from the bed (x) that you want their stairs to be and motivate your decision. Decide how long the planks need to be for the sides. (modified from the original)
   c) Help Moa and Martin to calculate the other measurements needed to saw the side pieces to the right size.(modified from the original)
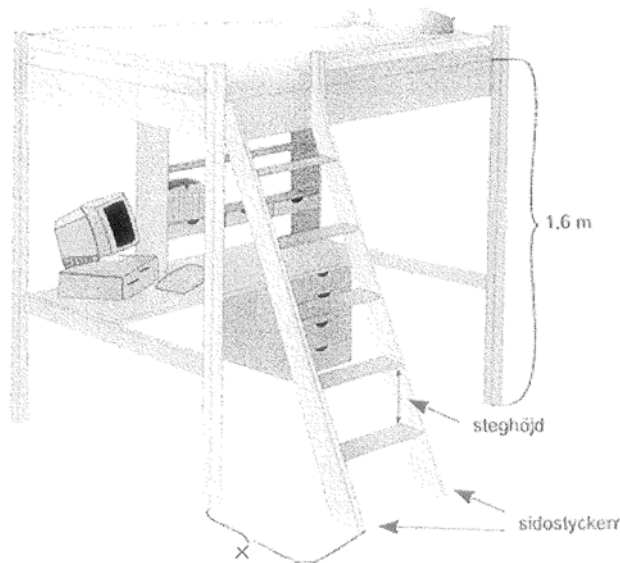
NB! You cannot measure in the diagram in order to solve the problem.

Task no. 3 – deleted because of privacy

Appendix 4. Excerpts from the original work of student L1, L2 and H1 in the pre- and post tests

Extract from Student L1 – pre test – from the problem Assembly halls

I a) 31 platser

rad 1 - 10 platse
rad 2 - 18 platser
rad 3 - 16
rad 4 - 19
rad 5 = 22
rad 6 = 25
rad 7 = 28
rad 8 = 31   Svar: rad 6 har 25 platser

b) Svar 8 rad

c) $10 + 3n$

II $12 + 5n$
rad 1 = 12
Varje rad ökar platser med 5

III $10 + 3n$
$19 \cdot 8 = 152$ platser

Kalle har fel

$10 + 13 + 16 + 19 + 22 + 25 + 28 + 31 = 162$ platser

Extract from Student L2 – pre test – from the problem Currency exchange

(J)

A. 500 kr = 45 pund

B.

05
.500 £
00
00
5
+2
250,0

500 - 250 = 250

100 = 7,7 pund

7,7 = 3,80
 2

200 kr = 2
        . 7,7   = 15,4£
        15,4

50 kr = 3,80£

250 = 15,4 + 3,8£ = 18,2£

Svar: 250 kr = 18,2£

18,2 · 2 = 36,4£ = 500 kr

**Extract from Student L2 – post test – from the problem
The inheritance**

Samband i alt A mellan summa och
antal år.

Den fasta summan är 550 kr, det får
han varje år.
Så summan ökar med 550 kr/år

$$550 \cdot x$$
$$x = \text{år}$$

Samband i alt C mellan summa &
antal år.
Robert får en summa på 2000 kr
som sedan ökar med 11% varje år.
alltså så får han efter (x) år   1,11 = förändring!
$$2000 \cdot 1,11^{x}$$ faktor.

**Extract from Student H1 – pre test – from the problem
Currency exchange**

174

Extract from Student H1 – post test – from the problem The inheritance

Alternativ C:

$y = 2000 \cdot 1,11^x$

$x = $ antal år

$y = $ summan på kontot

år 1: $2000 \cdot 1,11^1 = 2220$ kr på kontot

år 5: $2000 \cdot 1,11^5 \approx 3370$ kr på kontot

år 10: $2000 \cdot 1,11^{10} \approx 5678$ kr på kontot
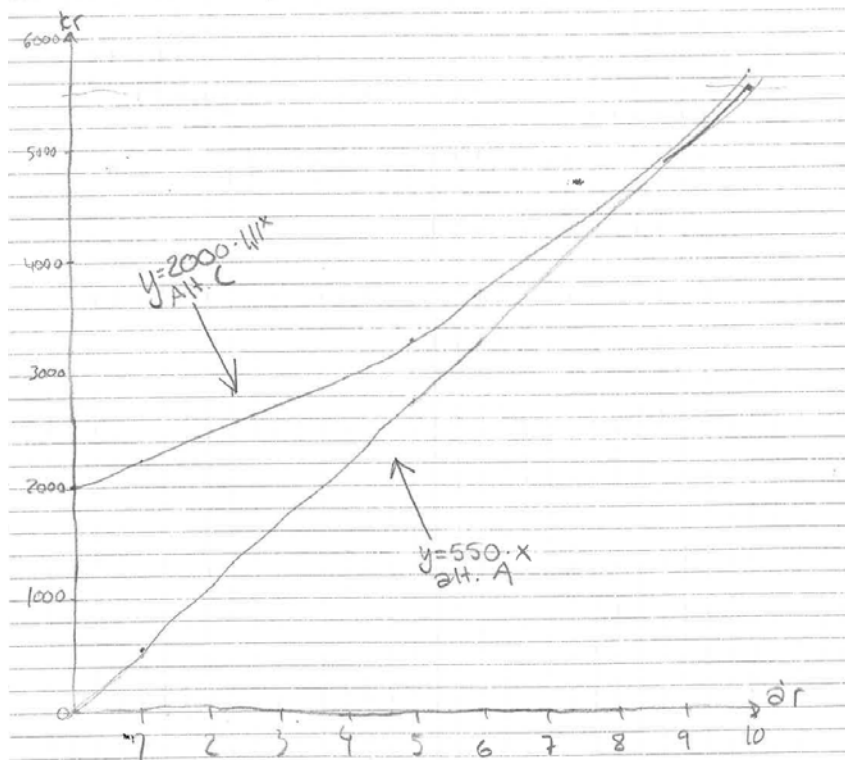
Appendix 5. Two examples of group assignments

Bridges and quadratic curves

Certain bridge spans can be described with the help of quadratic functions. In this task we shall take a look at four well-known bridges and the functions with which they can be described:

The Victoria Falls:
$$y = \frac{116 - 21x^2}{120}$$

The Langwies Viaduct:
$$y = 2 - \frac{2x^2}{9}$$

Tower Bridge:
$$y = \frac{9x^2}{80}$$

The Royal Tweed Bridge:
$$y = 1 - \frac{2x^2}{37}$$

- Make a sketch of what the span of the bridge can look like.
- The function of Tower Bridge describes one of the cables on the side of the bridge. The length of the bridge is 20 units of length. At what height is the cable attached?
- Which of the above bridge spans is the shortest?
- Which of the above bridge spans is the highest?

Children's sleep needs

A child's need for sleep can be approximately calculated with the formula $S = 15 - \dfrac{n}{2}$ where S is the number of hours of sleep per day and n is the age of the child in years.

- *a)* Anton is 4 years old.
  How many hours of sleep does he need according to the formula? *Only the answer is needed.*

- b) With your starting point in the formula draw a diagram that can be used in order to derive a child's need for sleep.

- c) Within which age range can the formula apply? Motivate your choice.

- d) Describe in everyday language what is meant by the formula.

Appendix 6. Interview guide

Interviews – students
Questions that are directly related to answers in the questionnaire

1. If you think back to the maths that you had in compulsory school and compare with how you've found the maths in this course, do you think there are any differences
2. Has your view of tests/assessments in maths changed and, if so, in what way?
3. Has your interest in maths been affected? Do you find it easier/more difficult, more fun/more boring, better/worse?
4. Do you work in any other way at school? At home?

Questions related to the teaching

1. If, for example, we take a rubric [show the rubric], what would you say is the point of using one?
2. Are there any other advantages with a matrix? … Any disadvantages? … Has it been of any help to you? … When and in which way?
3. What did it feel like when your classmates were allowed to give feedback on a group assignment? Why do you think that you did this exercise?
4. What did you get out of it?
5. Did you get anything out of assessing someone else's solution…? … If so, what?
6. Did you get anything out of working with more comprehensive tasks? If so, what?
7. What did it feel like working in a group?
8. Was there any difference between getting feedback from the teacher and your classmate? If so, what was the difference?
9. How was your experience of the pair-tests? How was it in relation to the written tests?
10. What was your experience of the accounts/presentations? Did you learn anything?
11. Did you have any use for the reflections in the logbook?
12. Is there anything you would like to change?

Interview – teacher

1. What was your experience of working in a new way?
2. What has this entailed with regards to workload?
3. The new method of working has been characterised by four elements: a scoring rubric, peer- and joint assessment, feedback, and problem solving. Can you say which the advantages and disadvantages are that you've experienced with each element?
4. Is there anything in the method that can be improved?
5. Is there anything that can be taken away and replaced with something else?
6. What is your experience of the effect the method has had on the pupils?
7. What words of advice would you like to give to other teachers who would also like to change their way of teaching?